5-2022

Genomic tools and models for investigating the role of germline diversity in mouse antibody repertoire development.

Justin T. Kos
University of Louisville

GENOMIC TOOLS AND MODELS FOR INVESTIGATING THE ROLE OF GERMLINE DIVERSITY IN MOUSE ANTIBODY REPERTOIRE DEVELOPMENT

By

Justin T. Kos

B.S., Saint Louis University, 2012

M.S., University of Missouri – St. Louis, 2015

A Dissertation

Submitted to the Faculty of the

School of Medicine of the University of Louisville

In Partial Fulfillment of the Requirements

For the Degree of

Doctor of Philosophy in Biochemistry and Molecular Genetics

Department of Biochemistry and Molecular Genetics

University of Louisville

Louisville, Kentucky

May 2022

GENOMIC TOOLS AND MODELS FOR INVESTIGATING THE ROLE OF GERMLINE DIVERSITY IN MOUSE ANTIBODY REPERTOIRE DEVELOPMENT

By

Justin T. Kos

B.S., Saint Louis University, 2012

M.S., University of Missouri – St. Louis, 2015

A Dissertation Approved on

April 13, 2022

by the following Dissertation Committee:

_____

Ronald G. Gregg, Ph.D.

_____

Thomas C. Mitchell, Ph.D.

_____

David J. Samuelson, Ph.D.

_____

Corey T. Watson, Ph.D.

_____

Jun Yan, Ph.D.

ACKNOWLEDGEMENTS

Thank you to Corey T. Watson for your mentorship, and for the opportunity to pursue a Ph.D. in your lab. I am also thankful for many collaborators who I have worked with and learned from over the years, particularly Andrew M. Collins, William D. Lees, and Katherine J. L. Jackson. Their feedback and guidance throughout my degree have contributed to my development as a scientist. I would also like to acknowledge my department chair, Dr. Ronald Gregg, for his mentorship and help creating congenic mouse lines.

ABSTRACT

GENOMIC TOOLS AND MODELS FOR INVESTIGATING THE ROLE OF GERMLINE DIVERSITY IN MOUSE ANTIBODY REPERTOIRE DEVELOPMENT

Justin T. Kos

April 13, 2022

Given the diversity and complexity within immunoglobulin (IG) loci, effective mouse models first require characterization of intra-strain differences and construction of high-quality reference assemblies for IG loci in several representative strains. To understand light chain germline diversity across biomedically significant mouse strains, we profiled the expressed IGK and IGL repertoires of 18 commonly used laboratory mouse strains using AIRR-seq. Across strains, we observed germline IGKV sequences shared by three different IGK haplotypes and a more conserved IGLV germline repertoire among common laboratory strains. Pacific Biosciences (PacBio) Single-Molecule Real-Time (SMRT) sequencing was used to sequence and assemble bacterial artificial chromosomes (BAC) clones spanning the IGH locus in BALB/cByJ and the IGK

locus in NOD/LtJ, which represented divergent IG haplotypes. We assembled the BALB/cByJ-IGH assembly into five independent contigs containing 192 functional and 135 non-functional IGHV genes, 30 IGHD genes, 4 IGHJ genes, and 8 IGHC genes. The NOD/ShiLtJ-IGK assembly was assembled into two independent contigs, which contained 82 functional and 31 non-functional IGKV genes. These data guided construction of congenic strains on a C57BL/6 background that carried divergent BALB/cByJ or NOD/ShiLtJ IGH or IGK loci from, respectively. In addition, bulk AIRR-seq data from the BALB/cByJ-IGH congenic strain showed that divergent IGH haplotype influenced usage frequencies of germline IGKV and IGLV repertoire. Overall, this work revealed significant unexplored IG haplotype diversity through AIRR-seq, generated new IG reference assemblies, identified incomplete germline gene databases that lacked haplotype diversity, and provided evidence that heavy and light chain pairing frequencies are likely influenced by underlying IG haplotype variation.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1 : INTRODUCTION

Antibodies (Abs), encoded by the immunoglobulin (IG) loci, are critical components of the immune system that function as cell-surface receptors for antigens and soluble effector molecules[1]. Antibodies can directly recognize a virtually unlimited array of molecular patterns on antigens and have many downstream effector functions[2]. Abs are formed from two pairs of identical heavy and light kappa or lambda genes, encoded by genes at three different loci in the mouse genome[1,2]. IG loci are among the most diverse and structurally complex regions of mammalian genomes, with high levels of single nucleotide polymorphisms (SNPs) and copy number variation (CNV), in both coding and non-coding regions[3,4]. In mice, genes from the IG heavy chain locus (IGH) are encoded by genes at a single locus on chromosome 12, whereas IG light chain genes are either encoded at the IG kappa (IGK; chromosome 6) or IG lambda (IGL; chromosome 16) locus[1,2,5]. Abs are separated into two domains according to their function: variable (V) domains that bind antigens, and constant (C) domains with effector or structural functions such as complement activation and Fc receptor binding[6]. V domains are created through a series of complex gene rearrangements that occur in developing B cells, while C domains are germline encoded. The entire

expressed component of antibodies circulating within an organism is referred to as the antibody (Ab) repertoire.

IG loci comprise duplicated variable (V), diversity (D, IGH only), joining (J), and constant (C) genes that recombine in B cells to generate the expressed Ab repertoire. The IGH locus comprises 97 variable (V), 14 diversity (D; IGH only), 4 joining (J), and 8-9 constant (C) functional/open reading frame genes[7]. The IGK locus is similarly complex and spans 3.2 Mb, with 101 functional V segments, 4 functional J segments, and 1 C segment, encoding approximately 95% of the mouse Ab light chain molecules[8,9]. In contrast, the IGL locus is smaller, containing only 3 functional V segments, 3 functional J segments, and 3 C segments spanning 240 Kb[7]. Structurally, genes within the loci are often described by their position relative to the constant region. Genes opposite the constant region are referred to as “distal”, whereas as genes closer to the constant region are “proximal”. Besides the V, (D), and J gene segments throughout the loci, the loci contain many pseudogenes, allelic variants, copy number variants (CNVs), and structural variants (SVs) that add to the already complex gene structure[10,11]. To better understand the recombination potential for different V, (D), and J genes throughout the loci, we must first properly understand the genomic organization of this unique family of genes and characterize both the coding and regulatory sequences of Ab heavy and light chain loci.

IMMUNOGLOBULINS – MEMBERS OF A MULTIGENE FAMILY

Immunoglobulins (Igs), members of the immunoglobulin superfamily (IgSF), are composed of 2 heavy (H) and 2 light (L) chains, where the L chain is either a kappa or lambda chain[12]. Both heavy and light chain Ig molecules are composed of multiple Ig domains, which are compactly folded regions of protein structure of approximately 110 amino acids. The L chains contain two Ig domains, while the H chain contains four. There are five different classes, or isotypes, of Igs: IgM, IgD, IgG, IgA, and IgE, which are distinguished from each other according to their C domains. The N-terminal domain of the protein is commonly referred to as the V region, and the C-terminal domain of the protein, which is used to distinguish different isotypes, is known as the constant (C) region. H chains contain 3 or 4 C domains, whereas the L chains contain only 1 C domain. Structures of both B cell receptors and antibodies are identical, except for the B cell receptor containing a hydrophobic C domain to anchor the molecule in the plasma membrane of the B cell.

The gene families encoding for Ab molecules share features with other multigene families such as the tRNA genes, DNA satellites, and ribosomal RNA[13]. Mammalian germ cells contain at least three families of antibodies, corresponding to the heavy, kappa, and lambda loci. Hood et al. elegantly described the three different types of multigene families that exist in the genome[13]. First, simple-sequence families are composed of DNA segments derived from 10^3 to 10^7 repeats of a short fundamental sequence of approximately 6-15 nucleotides[13]. Second, multiplicational families are 10-10,000 copies of an 80-1,000 nucleotide gene, with

identical repeats[13]. Lastly, informational families can have individual family members that differ in sequence composition from one another, although they are homologous and share ancestry[13]. Ab genes are members of the informational multigene families since they have evolved to generate information from multiple germlines and somatic hypermutation, express this information on each lymphocyte, and can expand its information through clonal selection and amplification[13,14]. The three different multigene families, including antibodies, share four characteristics that make them challenging to interrogate: multiplicity, close linkage, sequence homology, and overlapping functions[13].

VARIABLE REGION GENE FAMILIES

The variable region of an Ab is what gives the Ab its specificity in binding to antigen. To produce highly variable V regions capable of recognizing a large array of antigens, the mouse immune system has evolved to house a large number of highly polymorphic IGHV and IGKV gene segments in its germline. These highly polymorphic gene segments have been classified into different V-gene families and V-region protein groups[15]. A total of 16 different IGHV families have been described in the C57BL/6 IGH V region[16]. The initial gene family groupings were originally based on amino acid sequence similarity, but later revised to include classification based on nucleotide sequence similarity[17–20]. The IGHV gene family known as J558, or commonly known as the IGHV1 subgroup by the International ImMunoGeneTics information system (IMGT) and Honjo et al., is the largest V gene family, with 52 V genes, 37 pseudogenes, and it spans approximately 1.53

Mb of the 2.5 Mb mouse IGH locus[7,21]. The IGHV1 family is encoded in the distal part of the IGH locus along with the 3609 family, which is also commonly referred to as the IGHV8 family. The Q52 and 7183 families, or IGHV3 and IGHV5, respectively, contain fewer genes and are located at the proximal region of the C57BL/6 IGH locus.

The mouse IGK locus contains 20 different IGKV families, which have been described by multiple groups throughout the late 1980s and early 1990s[22–25]. Like the IGHV gene families, IGKV gene families also vary in size. The largest IGKV gene family, VK4/5, also known as IGKV4, was found to contain 15 germline IGKV genes[26]. IGKV11, IGKV17, and IGKV15 are the smallest VK gene families, each containing just one IGKV germline gene. Compared to the mouse IGH locus, the IGK locus appears to lack a defined proximal and distal gene family positional bias. While the IGKV3 family is the most proximal IGKV gene family, there is still some order to the location of gene families that is absent from the human IGK locus. Extending from the IGKV3 family is a clan of related IGKV8, IGKV6, and IGKV7 families, followed by a region with IGKV12 and IGKV5, and then a region of IGKV4 genes. At the distal region of the IGK locus VK genes from different families appear intermingled with no clear order[27].

Evolutionarily, it has been hypothesized that the mouse variable light chain repertoire is almost entirely composed of IGKV genes over IGLV genes. This has been observed directly from serum data, which shows that the peripheral Ab repertoire is 95% or more IGKV, compared to the approximately 60% IGKV light chain repertoire in humans[28]. Seeking to understand this discrepancy,

comparisons of the human and mouse structural IGKV repertoire revealed that the mouse germline IGKV genes are inherently more diverse than human IGKV genes[26]. Hypervariable loop structural comparisons between mouse and human IGKV germline genes showed that the mouse IGKV repertoire was capable of forming more canonical structure classes than human, and that the repertoire was more structurally diverse. Almagro et al. suggest that this could be a way for the mouse to compensate for the small size of the IGL locus[26]. In fact, phylogenetic trees comparing the VK germline genes of mice and humans show that the mouse IGKV repertoire contains more diversity than human because of the small size of the IGL locus[26]. Therefore, the increased structural diversity of the mouse IGKV repertoire compensates for the smaller mouse IGLV repertoire.

DEVELOPMENT OF A DIVERSE ANTIBODY REPERTOIRE

Many elements contribute to the diversity of the naïve and memory V gene repertoires. First, the mouse IGH locus is structurally complex, consisting of repeated, highly homologous sets of approximately 97 variable (V), 14 diversity (D; IGH only), 4 joining (J), and 8-9 constant (C) functional/open reading frame genes. The IGK locus is also complex, with approximately 94-96 variable (V), 4 joining (J), and 1 constant (C) functional/open reading frame genes. In addition, the loci contain numerous pseudogenes, allelic variants, copy number variants (CNVs), and structural variants (SVs) that add to the already complex gene structure. V(D)J recombination is the process through which the variable region of antigen receptor genes is constructed from somatic rearrangements of individual

V, D, and J gene segments[29]. During this process 1 V gene segment, 1 D gene segment (only IGH), and 1 J gene segment are somatically rearranged to create a single V(D)J exon that will code for the Ab V region. DNA sequences termed recombination signal sequences (RSSs), located 3' of each V gene segment, 5' and 3' of each D gene segment, and 5' of each J gene segment, help drive V(D)J recombination. RSS sites consist of a highly conserved heptamer, adjacent to the coding sequence, followed by a 12 or 23 nucleotide spacer, then an AT-rich nonamer. The length of the spacer sequence defines either the 12-RSS or 23-RSS[30,31]. The kappa and lambda loci, which lack D segments, undergo a single rearrangement that joins a V segment to a J segment.

The overall rearrangement process for the heavy chain locus occurs sequentially during the pro-B cell stage of B cell development[32], whereas the light chain loci rearrangements occur at the pre-B cell stage following heavy chain rearrangement, with the first joining of a D and J segment, followed by the joining of a V segment to the joined DJ segment. Combinatorial diversity is the process that tandemly rearranges variable (V), diversity (D; IGH only), and joining (J) germline gene segments through somatic recombination at the DNA level on one of two chromosomes in a B cell[1]. The process is termed combinatorial diversity since different combinations of gene segments are joined somatically to generate unique B cell receptors. Further diversity is generated by junctional diversity, where nucleotides are added or removed at the junctions of the V and D, D and J, or V and J segments when the segments are joined[33–35]. Since an Ab is composed of two identical heavy and light chains, and these loci are encoded in different regions

of the genome, different heavy and light chain pairings further help increase the diversity of the repertoire. Lastly, upon encounter with an antigen, a B cell can be stimulated to undergo somatic hypermutation (SHM), which introduces mutations within the V region of the Ab[36]. Combined, these mechanisms will generate B cells that express a unique combination of V_H, D_H, and J_H segments as the V_HDJ_H gene encoding the V region of the heavy chain, and a unique combination of V_L and J_L segments as the V_LJ_L gene encoding the V region of the light chain[37,38]. The unique Ab repertoires formed during V(D)J recombination can also be affected by extrinsic factors, such as pathogens or vaccinations, and by intrinsic factors occurring within the organism such as aging and autoimmunity[39]. While combinatorial and junctional diversity are often described as random events, studies have shown that the Ab repertoires generated from these processes are not a result of random events but rather a result of genetic biases[32].

GENOMIC RESOURCES FOR IG LOCI

Early studies on mouse IG genetics utilized deletion mapping and yeast artificial chromosome (YAC) contig assembly to map the position of functional V genes[40,41]. While constant, joining, and diversity sequences had already been identified and sequenced, the mouse IGK locus was the first IG locus to be properly sequenced, with Zachau et al. identifying 118 IGKV sequences derived from BAC and YAC clones. They described 18 different IGKV gene families, with 4 of the families only containing a single sequence, dispersed throughout the 3.2 Mb IGK locus. In 2006, the C57BL/6 mouse IGH locus was sequenced from BAC clones

and elucidated the presence of IGHV, IGHD, and IGHJ genes spanning 2.5 Mb of genomic sequence. Johnston and colleagues generated an annotated map of the IGH locus V region, which included both V genes, pseudogenes, repeats, and nonrepetitive intergenic sequences[16].

In mice, single nucleotide polymorphism (SNP) studies performed across the genome have informed us that significant diversity exists within inbred mouse populations[42,43]. Presently, the ImMunoGeneTics (IMGT) database contains 114 functional C57BL/6 IGHV genes, which were catalogued by validating sequences from the 2006 reference genome with functional RSS sequences[7]. IMGT primarily contains genes and alleles for C57BL/6, therefore it is unlikely to represent the immunoglobulin genetic variation documented from SNP studies. Recently, the Adaptive Immune Receptor Repertoire (AIRR) Community has established a new database named Open Germline Reference Database (OGRDB)[44]. This database was designed to help evaluate inferred allele generated through AIRR-seq experiments and provide an updated record of affirmed germline sequences. It currently hosts inferred IGHV germline genes for BALB/c and C57BL/6 generated from long-read SMRT sequencing of VDJ-C amplicons from F1 hybrid animals[45].

GENETIC ORIGINS OF THE LABORATORY MOUSE

The mouse has served as an invaluable model organism in the biomedical sciences for over a century, used across multiple biomedical disciplines to increase the quality of human life and help prevent and treat disease[46,47]. While mice are mammals and phylogenetically close to humans, their small size, minimal

maintenance, and short generation time make them very useful in science[48]. During the first decade of the twentieth century, the breeding of laboratory mice gained popularity in the sciences, as researchers such as C. C. Little began studying coat color inheritance using mouse models[49,50]. Introduced into the laboratory over a century ago, the mouse has been used in genetic studies, random mutagenesis experiments, the development of inbred lines, and the direct engineering of the genome through knock-in, knockout, and transgenic techniques[46]. These uses have made *Mus musculus* one of the most well-studied models in mammalian genetics[51]. Though the mouse is invaluable to science, the origins of the laboratory mouse are complex and point to underlying genetic diversity in laboratory mice.

Early classifications on the subdivisions of *Mus musculus* were completed using biochemical and molecular markers[49]. These markers revealed that the subdivisions of *Mus musculus* resulted from a radiation that occurred in Northern India about 0.5 million years ago (MYA)[52]. From Northern India, the subdivisions of *Mus musculus* migrated into Eurasia, leading to the present sub-species we have today. *M. m. domesticus* emigrated to western Europe and the Mediterranean basin, *M. m. musculus* to central Europe and northern China, and *M. m. castaneus* to southeast Asia. *Castaneus* and *muscuclus* subspecies groups also formed a hybrid subspecies in Japan, molossinus, which still exists today. The *domesticus* subspecies group in Western Europe entered the European 'fancy' mouse trade, whereas the *musculus and molossnius* subspecies groups entered the East Asian 'fancy' mouse trade. Therefore, the laboratory mouse we have today results from

the European and East Asian 'fancy' mice, and the radiation of the different subspecies led to a mosaic evolution of different parts of the mouse genome.

In the late 1960s, work began to better understand the polymorphisms and genetic heterozygosity present in *M. m. musculus* and *M. m. domesticus* subspecies living in Denmark. The Danish mouse populations were unique since they existed in a zone of allopatric hybridization yet appeared to have little gene flow between them. To determine the heterozygosity between the two Danish subspecies, Selander et al. examined the electrophoretic variation in 36 proteins controlled by 41 genetic loci[53]. They found that 56% of the proteins (20/36) were monomorphic and identical across sub species. 16 of the 36 proteins (44%) were polymorphic for two or three alleles in one or more of the samples they tested. Overall, the frequency of polymorphic loci was higher in *M. m. musculus* compared to *M. m. domesticus*. Although this data was the first of its kind detailing biochemical differences between the two mouse subspecies, it lacked a detailed description of the genetic variation due to the lack of high-resolution genomic data[54]. It wasn't until genome-wide shotgun SNP discovery data from the Mouse Genome Sequencing Consortium became available that it was possible to describe the genetic variation among inbred laboratory mouse strains.

The first genome-wide comparisons across different mouse strains were made possible in the early 2000s when mouse genomic sequences became available[46,55]. Wade et al. compared assembled genomic sequence from C57BL/6 to genomic sequence generated from other laboratory and wild-derived strains[54,56]. The polymorphism rates they calculated ranged from long segments (greater than

1 Mb) of high polymorphism (40 SNPs per 10 kb) or low polymorphism (0.5 SNPs per 10 kb). Out of all the strain-to-strain comparisons they made, roughly one third of the genome fell into long regions of high polymorphism, which was consistent with previous data on the estimated divergence rates between *Mus musculus domesticus* and either *M. m. musculus* or *M. m. castaneus*[54]. Comparison of inbred laboratory strains to C57BL/6 revealed two-thirds of the genome to have a very low polymorphism rate, which was indicative of genomic regions in which the two strains shared a recent common subspecies origin. In contrast, the remaining one-third of the genome had a much higher polymorphism rate, which suggested that these regions were areas of the genome in which the two strains inherited the region from different subspecies. Overall, the bimodal distribution Wade et al. described represented regions of the genome that had different or similar subspecies origins and revealed the mosaic model common to mouse haplotype analyses. However, the lack of available reference databases made it challenging to assign a subspecies origin to genomic regions of a given laboratory strain.

The mosaic model of mouse genetic variation became more apparent after whole-genome SNP arrays were used to characterize haplotype diversity. In 2004, Frazer et al. produced a high-resolution map of genetic variation for 11 classical strains and 4 wild-derived strains[57]. Using an oligonucleotide array they identified 8.27 million SNPs throughout the genome, generating a genome-wide haplotype map. Their map revealed the genetic contributions for classical strain haplotypes to be 68%, 6%, 3%, and 10% for *M. m. domesticus*, *M. m. musculus*, *M. m.*

castaneus, and *M. m. molossinus*, respectively. The remaining 13% of haplotypes were of unknown origin.

IG HAPLOTYPE VARIATION IN INBRED MOUSE STRAINS

Laboratory mouse strains are derived from *M. musculus*, which is a species with multiple lineages, including three major subspecies, *M. m. domesticus*, *Mus musculus musculus*, and *Mus musculus castaneus*[52]. It was not until researchers began exploring the genetic basis of Ab diversity in mouse strains that IG haplotype diversity became known. Seeking to characterize mouse germline IGHV genes, Brodeur and Riblet used Southern blot analysis across eighteen inbred mouse strains to define IGHV gene families, estimate the total number of mouse IGHV genes, define haplotypes across strains, and compare IGHV polymorphisms across strains[17]. The twenty-four IGHV probes used in their study revealed over one hundred IGHV genes present in the mouse IGH locus. Genes within one of the 15 IGHV gene families shared more than 80% sequence homology, whereas sequence homology between gene families was less than 70%. Brodeur and Riblet's data was the first of its kind to reveal haplotype assignments for the IGHV region of the IGH locus. They used the IGHV DNA probes from all IGHV gene families to determine which mouse strains had the same or different IGHV gene. Their work demonstrated that strains with the same IGHC region haplotype generally, though not always, shared the same IGHV haplotype. Their work revealed a total of thirteen different IGHV region haplotypes, with four of the

thirteen haplotypes represented by more than one strain[17]. Though this Southern blot experiment lacked the level of resolution possible with today's technology, it was an early glimpse into IGH locus diversity across mouse strains.

Similar early haplotype RFLP experiments were also performed on the light chain loci. Thirteen different IGKV probes used to detect RFLP patterns revealed the presence of seven different IGK haplotypes among 55 inbred mouse strains[28]. The largest haplotype group, IGK^C, contained 38 different classical strains, including BALB/cByJ and C57BL/6N[28]. There is also haplotype variation for the IGL locus, though it is not as diverse as the IGH and IGK loci.

INTERROGATING ADAPTIVE IMMUNE RECEPTOR REPERTOIRES

To effectively study and characterize Ab repertoires, Adaptive Immune Receptor Repertoire Sequencing (AIRR-seq) was developed utilizing high throughput sequencing technology. AIRR-seq studies utilize large-scale sequencing of DNA libraries that are prepared by amplifying either genomic DNA (gDNA) or mRNA coding for the B-cell receptor (BCR) using PCR[58]. Library prep procedures can vary depending on the biological question. 5' rapid amplification of cDNA ends (5'RACE) is common in the field since it does not bias the sequencing library with V segment primers that are specific to V genes or V gene families. Since BCR sequences are not encoded directly in the genome, the data generated by AIRR-seq must also be processed differently than standard DNA- or RNA-seq data. First, the set of V, (D), and J segments used by each BCR must be inferred.

Second, SHM must be considered, which can change more than 5% of the bases in BCRs[58]. If the goal of the experiment is to infer germline sequences rather than characterize the Ab repertoire, then the length of the complete library must fall within the technical limits of the sequencing technology, and the libraries must have captured the entire recombined V(D)J molecule internal of the amplification primers[59].

GENOMIC SEQUENCING TECHNOLOGIES

ILLUMINA SHORT READ SEQUENCING

Next generation DNA sequencing technology has made it possible to interrogate and categorize Ab repertoires. The most prominent sequencing technology used for AIRR-seq is 5'RACE with 2x300 Illumina sequencing. This method can accurately sequence molecules up to approximately 500 bp in size using paired-end sequencing. Currently, commercial kits following a core methodology are available that can generate sequencing libraries starting directly from tissue. Briefly, RNA is extracted from the source, which for mouse AIRR-seq is usually spleen but can also be isolated B cells. Next, 5'RACE converts RNA into cDNA using a dT primer and reverse transcriptase (RT). This approach uses a template-switch step, which adds non-templated nucleotides to the end of the cDNA once the RT reaches the 5' end of the mRNA template. An oligonucleotide is added that anneals to the non-templated nucleotides, which functions as a template for addition of additional nucleotides into the first-strand cDNA. The first-

strand cDNA is then primed using isotype primers and a universal oligonucleotide that allows template switching for the reverse transcriptase (RT) reaction. This method is successful at generating a significant amount of sequence depth for the library. One potential limitation to the technique, however, is it does not capture the entire V(D)JC molecule, but rather just the V(D)J portion due to the 2x300 limitation. This method also requires that the paired-end reads be assembled to generate the complete V(D)J amplicon, which can be challenging when the Illumina R2 reads can have lower base quality as the read lengths approach 300 bp.

PACIFIC BIOSCIENCES SINGLE MOLECULE REAL-TIME SEQUENCING

Single-molecule real-time (SMRT) sequencing, a third-generation sequencing technology developed by Pacific Biosciences (PacBio), offers an alternate method of sequencing that can sequence through complex genomic regions, which pose a significant challenge to sequencing platforms that rely on short reads[60]. SMRT sequencing offers considerable advantages over second-generation DNA-sequencing technologies like Illumina, such as longer read lengths for *de novo* assembly, haplotype detection, and phasing, and higher consensus accuracy[61]. Briefly, the technology works by directly observing a molecule of DNA polymerase as it synthesizes DNA. The DNA polymerase is confined in a zero-mode waveguide, and nucleotide base additions are measured with real-time fluorescence detection. The long reads generated by the PacBio

platform also makes it a valuable resource for generating new reference assemblies.

Although Illumina sequencing is very popular for AIRR-seq, Pacific Biosciences (PacBio) SMRT sequencing has also been used to characterize mouse Ab repertoires[4]. Because sequencing on the PacBio platform is not limited by sequence length, AIRR-seq with PacBio is able to capture the complete V(D)JC molecule, depending on primer placement. Another advantage of using the PacBio system is the high-quality data that is generated due to the circular consensus algorithm used to generate circular consensus reads, which are usually much higher quality than reads from Illumina. The extremely high read accuracy, combined with the elucidation of the complete V(D)JC molecule, makes the PacBio platform ideal for germline gene inference. However, the PacBio platform remains more expensive than an Illumina MiSeq run, which may be cost prohibitive depending on the goal of the experiment.

OXFORD NANOPORE TECHNOLOGIES LONG READ SEQUENCING

Oxford Nanopore Technologies (ONT) is another third-generation sequencing technology that can generate long sequencing reads with high scalability. This technology relies on the transit of a DNA molecule through a pore, and the measurement of the change in electric current across the pore[61]. Because this technology does not rely on imaging equipment to detect nucleotides, the entire system can be scaled down to a much smaller footprint than either Illumina

or PacBio sequencers[62]. In addition, because ONT directly detects the input DNA without PCR amplification, there is no limit to the length of DNA that can be sequenced[62]. The long reads produced by the sequencer can help with *de novo* genome assembly, and sequence through genomic regions containing long stretches of repetitive DNA. A disadvantage to ONT sequencing, however, is the relatively high error rate compared to short read technology like Illumina. As of 2019, ONT error rates ranged from 5% to 20% depending on input material and library preparation methods, and the errors included both insertions and deletions[62]. One of the key components to ONT sequencing is the nanopore itself. The biological nanopore is constructed from modified alpha-hemolysin pore with an exonuclease attached on the extracellular face of the pore. On the inside surface of the nanopore there is a covalently attached synthetic cyclodextrin sensor. The entire system is within a synthetic lipid bilayer so that when DNA is loaded onto the exonuclease-containing face and voltage is applied, the exonuclease cleaves off individual nucleotides. Once cleaved, the nucleotides are detected by the change in current through the nanopore.

AIRR-SEQ ANALYSIS

Many different programs exist for the analysis of AIRR-seq data, but they all share a similar core of features. During sequencing read pre-processing, samples that were multiplexed using unique molecular identifiers (UMIs) are demultiplexed to sets of reads for each sample. Next, read quality is assessed

using quality control software such as FastQC, with the goal of obtaining Phred-like scores greater than 30 for the majority of the read, meaning 1 error per 1000 base pairs. At this stage, sequences with low sequence quality can be discarded from the dataset and excluded from downstream analysis. Then, primer sequences are masked, and the sequencing reads are assembled with their mate pairs if sequenced on the Illumina platform. AIRR-seq performed on the PacBio platform does not require mate pair assembly as the complete V(D)J amplicon is sequenced. Once the mate pairs are assembled the reads are then collapsed to remove duplicate sequences and reduce the size of the dataset. Finally, the collapsed sequences are aligned to a germline gene database that contains either heavy or light germline IG sequences for the strain of interest. Once germline gene assignments have been made, clonal clusters can be assigned to the sequences. The Kleinstein lab has developed and published a suite of tools that can process AIRR-seq reads from raw data to a refined set of IG repertoire sequences collapsed by clone, and plotted using their R package Alakazam[63,64].

GERMLINE GENE INFERENCE

Many bioinformatic tools are available to infer germline IG genes from next-generation sequencing (NGS) data, but they all have a core function of identifying novel IG germline genes. To accurately examine the expressed Ab repertoires in mice, AIRR-seq data must be compared to expressed Ab sequences with reference databases of variable (V) germline genes to calculate metrics like gene usage, expression frequency, and degree of somatic hypermutation[65]. Because

the mouse germline gene databases are mostly incomplete for strains other than C57BL/6, novel germline genes must be inferred with bioinformatic tools like TIgGER[66] or IgDiscover[65], with IgDiscover best suited to analyze unmutated, naïve germline sequences.

There are four critical steps during germline gene inference. First, processed sequences must be aligned against a starting gene database that contains germline IG sequences for the locus and strain of interest. The initial alignment of input sequences is performed with IgBLAST[67], and any novel germline genes are assigned to the most similar database sequence. Next, sequences are clustered to identify potential germline sequences, with the option to set different thresholds for the number of supporting sequences per novel germline sequence cluster. Once potential novel germline sequences have been clustered, filters are applied to keep only true germline sequences. For example, IgDiscover employs a germline filter to help remove false positives that may result from PCR bias. The filter functions by ensuring that true germline sequences contain an identical V region and are present in multiple unique V(D)J rearrangements, which is observed by many unique CDR3 sequences. Sequences that have undergone SHM and that are overrepresented from PCR bias will be filtered out of the inferred germline gene database. Finally, the inference process is repeated using the newly constructed germline gene database.

SUMMARY

While the research presented in the introduction provides foundational knowledge necessary to understand the work completed throughout this dissertation, the research presented in the following chapters will advance the understanding of how genomic variation within the IGH and IGK loci influence Ab repertoire diversity and dynamics and address gaps in knowledge due to incomplete gene databases and reference assemblies for a single IG haplotype. In Chapter 2, since germline light chain Ab diversity has not been explored and characterized, we hypothesized that profiling the germline light chain repertoires across multiple strains of mice would reveal extensive haplotype diversity. In Chapter 3, additional IG reference assemblies were added to the immunogenomics field by sequencing, assembling, and annotating BAC clones spanning the IG loci of mouse strains containing IG haplotypes divergent from the current C57BL/6 reference assembly. Finally, in Chapter 4 we constructed congenic IG mouse lines with controlled IG haplotype diversity and showed that the germline light chain Ab repertoire can be influenced by haplotype variation in the heavy chain loci.

CHAPTER 2 : CHARACTERIZATION OF EXTENSIVE DIVERSITY IN IMMUNOGLOBULIN LIGHT CHAIN VARIABLE GERMLINE GENES ACROSS BIOMEDICALLY RELEVANT MOUSE STRAINS

INTRODUCTION

Antibodies (Abs), encoded by the immunoglobulin (IG) loci, are critical components of the immune system that function as cell-surface and soluble receptors for antigens[1]. The process of somatic V(D)J recombination within B cells governs the formation of a diverse Ab repertoire capable of recognizing a vast array of antigens through interaction with Ab variable domains[37]. Abs are formed from two pairs of identical heavy and light (kappa or lambda) chains, encoded by genes at three different loci in the mouse genome. In mice, the IG heavy chain is encoded by genes at a single locus on chromosome 12 (IGH), whereas IG light chain genes are encoded at the IG kappa (IGK; chromosome 6) and IG lambda (IGL; chromosome 16) loci[1,5]. Abs are separated into two functional domains: variable (V) domains that bind antigen; and constant (C) domains that carry out effector functions such as complement activation and Fc receptor binding[6]. The V

domain is encoded by variable (V), diversity (D, IGH only), and joining (J) genes, while constant (C) genes encode the C domain. Together, V, D, J, and C genes somatically recombine in B cells to generate the Ab repertoire, the entire expressed component of Abs circulating within an organism.

The mouse IG loci are structurally complex and consist of repeated, highly homologous gene segments[7,16,68]. For example, the IGH locus in the C57BL/6 strain comprises 102 variable (V), 9 diversity, 8 of which are unique (D; IGH only), 4 joining (J), and 8-9 constant (C) functional/open reading frame genes[7,45]. The IGK locus of C57BL/6 is similarly complex in this strain and spans 3.2 Mb, with 91 functional V segments, 4 functional J segments, and 1 C segment, representing approximately 95% of the C57BL/6 germline light chain genes[8,9]. In contrast, the C57BL/6 IGL locus spans 240 kb and includes only 3 functional V segments, 3 functional J segments, and 3 C segments[7].

At the genomic level, the mouse IG loci have only been comprehensively characterized in the C57BL/6 strain. However, the C57BL/6 mouse does not fully represent variation within the mouse IG loci. In 2007, Retter et al. sequenced and assembled bacterial artificial chromosome (BAC) clones spanning the IGH constant region and part of the variable region in the 129S1 mouse strain[69], which was predicted by restriction fragment length polymorphism (RFLP) to carry a divergent IGH haplotype compared to C57BL/6[17]. They showed that the IGH^A haplotype of the 129S1 strain is genetically different from the IGH^B haplotype of the C57BL/6 strain, containing major germline gene duplications present in the IGH^A haplotype that are absent in the IGH^B haplotype. Though the light chain loci

have only been characterized in C57BL/6, early RFLP experiments reported the existence of 9 IGK haplotypes in commonly used inbred mouse strains[15], and more recent Sanger sequencing identified significant IGKV polymorphisms in NOD mice[70,71]. In addition, early isoelectric focusing experiments of mouse light chains highlighted IGK haplotype differences by showing that SWR/J, C3H/HeJ, DBA/1J, A/J, CBA/J, and C57BL/6J had identical focusing bands, whereas AKR/J and C58/J had observed differences[72]. More recently, genome-wide high-throughput single nucleotide polymorphism (SNP) studies have revealed inter-strain diversity across all three IG loci[42,43].

The more recent application of high-throughput Adaptive Immune Receptor Repertoire Sequencing (AIRR-seq) studies has also led to the discovery of extensive variation in the germline IGHV genes. For example, AIRR-seq studies of C57BL/6 and BALB/c mice demonstrated that the BALB/c IGHV germline set consisted of >160 genes, only 4 of which overlapped with those found in C57BL/6. In a subsequent study of 5 additional inbred strains, including 4 wild-derived strains representing diverse sub-species origins, we observed significant inter-strain variation in IGHV germline sequences, and catalogued 247 germline alleles unaccounted for in existing reference databases[4]. However, despite evidence of potentially similar levels of diversity within the IGKV and IGLV coding regions, these loci have not been comprehensively explored across inbred strains.

In this study, to better understand mouse light chain germline diversity, we conducted AIRR-seq analysis and germline inference in 18 different inbred mouse strains, again including 4 wild-derived strains from diverse sub-species origins, as

well as an additional 14 strains commonly used in biomedical research. Consistent with our observations in the IGH locus, we observe significant germline sequence variation between strains. In addition, inferred IGLV genes across the classical and wild-derived strains reveal the presence of fewer germline genes in classical strains than wild-derived strains, which may result from the breeding history of classical laboratory mice. This level of germline diversity is unexplored in the study of immune phenotypes and unaccounted for in existing gene databases. Despite the diversity observed, we uncover evidence for the presence of shared germline IGKV/LV gene sets and haplotypes among subgroups of classical laboratory strains, indicating that the light chain loci reflect shared sub-species origins.

MATERIALS AND METHODS

AIRR-SEQ LIBRARY PREPARATION AND SEQUENCING

Whole dissected spleens, preserved in RNAlater (Thermofisher, Cat. No. AM7020; Waltham, MA, USA), were obtained from female mice from Jackson Laboratories (Bar Harbor, ME, USA; https://www.jax.org) for eighteen inbred strains [BALB/cByJ (Jax stock #001026), n = 1; NOR/LtJ (Jax stock #002050), n = 1; 129S1/SvlmJ (Jax stock #002448), n = 1; MRL/MpJ (Jax stock #000486), n = 1; A/J (Jax stock #000646), n = 1; AKR/J (Jax stock #000648), n = 1; CBA/J (Jax stock #000656), n = 1; C3H/HeJ (Jax stock #000659), n = 1; C57BL/6J (Jax stock #000664), n = 1; DBA/1J (Jax stock #000670), n = 1; DBA/2J (Jax stock #000671),

n = 1; NZB/BlNJ (Jax stock #000684), n = 1; SJL/J (Jax stock #000686), n = 1; CAST/EiJ (Jax stock #000928), n = 1; NOD/ShiLtJ (Jax stock #001976), n = 1; LEWES/EiJ (Jax stock #002798), n = 1; MSM/MsJ (Jax stock #003719), n = 1; PWD/PhJ (Jax stock #004660), n = 1].

We extracted total RNA from 30 mg of spleen tissue using the RNeasy Mini kit (Qiagen, Cat. No. 74104; Germantown, MD, USA). For each sample, IGK and IGL 5'RACE AIRR-seq libraries were generated using the SMARTer Mouse BCR Profiling Kit (Takara Bio, Cat. No. 634422; Mountain View, CA, USA), following the manufacturer's instructions. Individual indexed IGK and IGL AIRR-seq libraries were assessed using the Agilent 2100 Bioanalyzer High Sensitivity DNA Assay Kit (Agilent, Cat. No. 5067-4626) and the Thermofisher Qubit 3.0 Fluorometer dsDNA High Sensitivity Assay Kit (Thermofisher, Cat. No. Q32851). Libraries were pooled to 10 nM and sequenced three times on the Illumina MiSeq platform using the 600-cycle MiSeq Reagent Kit v3 (2x300 bp, paired-end; Illumina, Cat. No. MS-102-3003); per sample read depth is provided in Table 2.1.

DATA PROCESSING AND GERMLINE GENE INFERENCE

IgDiscover v0.12[65] was used to construct a germline IGK database for each strain. Briefly, we combined FASTQ reads from each MiSeq run and processed using IgDiscover v0.12[65] using the following parameters: (1) "barcode_consensus" set to false since samples did not have barcodes; (2) "race_g" set to "true" to account for the run of G nucleotides present at the start of the sequence; (3)

"stranded" set to "true" since the forward primer was always at the 5' end of the sequence; (4) "limit" set to false to process all reads; (5) "merge_program" set to flash; and (6) "ignore_j" set to "true" to ignore whether a joining (J) gene had been assigned to an inferred IGKV or IGLV gene. We used IGKV and IGKJ mouse [20]sequences downloaded from the ImMunoGeneTics Information System (IMGT) (downloaded August 2021) as the starting database for IGKV inference.

Germline IGLV sequences were manually inferred using our previously established procedure[4]. Briefly, IGL sequences were processed using the Immcantation Pipeline[63,64] with the IMGT IGL gene database serving as the starting IgBLAST[67] database for germline gene/allele assignment. First, IGL primer sequences (Supplemental Table 1) were identified using maskPrimers align. Since primer sequences were not provided with the SMARTer Mouse BCR Profiling Kit, we manually determined the primer sequences by performing a multiple sequence alignment of the first 30 base pairs of the Illumina R1 reads. Next, read pairs were assembled using assemblePairs align, then duplicate reads were collapsed using collapseSeq, with the duplicate count of each collapsed sequence recorded as "dupcount". Downstream processing required that all sequences have a dupcount ≥ 2. Initial assignments to germline IGLV and IGLJ genes were performed using IgBLAST, with the resulting output parsed with Change-O MakeDb. Clones were identified by defineClones, with the clonal thresholds determined independently for each strain using distToNearest function in SHazaM[63,73]. Next, clones were clustered based on IGLV gene assignment and the percent identity to the nearest mm10 reference sequence. Finally, we determined consensus IGLV gene

sequences using CD-HIT (cd-hit-est v4.6.8)[74], requiring that a given cluster sequence be represented by at least 0.1% of the total number of clones identified per strain. This process was repeated across all strains to create a unique inferred germline IGLV gene set for each strain.

We validated our inferred IGKV and IGLV germline sequences using TIgGER[63,66]. Briefly, Presto-processed reads were input into IgBlast[67] and Change-O[63], with our inferred germline IGKV and IGLV sequences as the starting gene database for IgBLAST[67] alignment. Upon generation of a Change-O table for each strain, the inferGenotype function of TIgGER was performed, and sequences were considered validated if they were successfully identified.

IGKJ AND IGLJ GERMLINE GENE INFERENCE

J genes were analyzed independently of V genes for IGK and IGL germline repertoires. First, IgBLAST[67] was run on all sequences using the IMGT germline gene database (release 202209-1; 28 February 2022) for IGK and IGL, and we passed the resulting IgBLAST output tables to Change-O. Then, we manually inspected Change-O tables for frequently occurring sequences in the J calls that were not exact sub-sequences of the IMGT reference J alleles. Sequences present in a strain at a rate > 1% were added to a new J gene reference database that included novel J sequences and IMGT database J sequences. We then re-ran IgBLAST and Change-O's MakeDb function with the new reference J gene set and inspected MakeDb output to ensure that no valid reads failed the pipeline due to

missing J reference alleles. Lastly, we validated our candidate novel J alleles using OGRDBstats[75].

DATABASE AND INTER-STRAIN COMPARISONS OF GERMLINE GENE SETS

We compared our inferred germline sequences to the existing IMGT gene database using IMGT HighV-QUEST v1.8.3 (7 May 2021)[34,76]. We also compared inferred IGKV and IGLV germline gene sets between strains in a pairwise fashion using BLAT[77,78]. For each inter-strain comparison, the germline set of the strain with the smallest number of inferred sequences was used as the ”query”. For each sequence in the query set, the best match from the alternate strain was assigned based on percent identity and match alignment length, requiring a minimum alignment length of 275 bp. The mean sequence identity of the best matches for all sequences in the query set was computed and used to express the average similarity of sequences between two strains. The mean similarity between inferred sequences across all strains in relation to their predicted SNP haplotypes was visualized using the Pheatmap[79] package in R.

SNP-PREDICTED HAPLOTYPE AND SUB-SPECIES ORIGIN ANALYSIS

Each strain’s SNP-predicted haplotype and sub-species origin were determined using whole-genome SNP data from the Mouse Phylogeny Viewer[80]. Data were viewed and downloaded for mouse IGK and IGL loci (IGK,

chr6:67449994-70709994; IGL, chr16:19055093-19265093) to determine SNP-predicted haplotypes and sub-species origin. We performed a multiple sequence alignment of genotypes at SNP positions spanning the IGK and IGL loci to generate a SNP-predicted haplotype for each strain. We then used the multiple sequence alignment to construct a neighbor-joining phylogenetic tree to cluster strains into different shared haplotypes (Figures 2.1, 2.2). To ensure that the SNP-predicted haplotypes were assigned accurately, we required genotypes to be present for all strains at all positions for which there was SNP data. For example, if the SNP array produced an N for a position in a given strain, then the position was masked across all strains in the multiple sequence alignment to prevent the Ns from contributing to the topology of the neighbor-joining phylogenetic tree, and thus biasing the haplotype groupings. In total, we masked 23% (74/324) of SNP positions for IGK and 39% (12/31) for IGL. Haplotype groups were assigned to strains that clustered together in the phylogenetic trees, with groupings annotated in Table 2.1 and Figures 2.1 and 2.2.

Table 2.1. Subspecies origin and subspecies identity of the immunoglobulin kappa and lambda loci of classical laboratory and wild-derived mouse strains selected for IGKV and IGLV germline gene inference.

Strain	Subspecies Origin	Immunoglobulin Kappa Chain Subspecies identity by SNV Analysis	Immunoglobulin Lambda Chain Subspecies identity by SNV Analysis	Classical vs Wild-Derived	IGK SNP-Predicted Haplotype Group	IGL SNP-Predicted Haplotype Group	Number of IGK Illumina MiSeq Reads	Number of IGL Illumina MiSeq Reads
129S1/SvIm	N/A	*M. m. domesticus*	*M. m. domesticus*	Classical	C	E	1,145,322	1,599,660
A/J	N/A	*M. m. domesticus*	*M. m. domesticus*	Classical	C	E	1,286,740	1,278,216
AKR/J	N/A	*M. m. domesticus* and *M. m. castaneus*	*M. m. domesticus*	Classical	B	E	655,738	1,256,636
BALB/cByJ	N/A	*M. m. domesticus*	*M. m. domesticus*	Classical	C	E	1,461,573	1,717,669
C3H/HeJ	N/A	*M. m. domesticus*	*M. m. domesticus*	Classical	C	E	1,403,724	1,484,669
C57BL/6J	N/A	*M. m. domesticus*	*M. m. domesticus*	Classical	C	E	1,063,818	1,327,303
CAST/EiJ	*M. m. castaneus*	*M. m. castaneus*	*M. m. castaneus*	Wild-derived	B	C	1,125,129	2,020,994
CBA/J	N/A	*M. m. domesticus*	*M. m. domesticus*	Classical	C	E	1,003,013	1,795,640
DBA/1J	N/A	*M. m. domesticus*	*M. m. domesticus*	Classical	C	E	1,015,408	1,690,039
DBA/2J	N/A	*M. m. domesticus*	*M. m. domesticus*	Classical	C	E	707,388	1,600,720
LEWES/EiJ	*M. m. domesticus*	*M. m. domesticus*	*M. m. domesticus*	Wild-derived	C	D	1,822,863	2,386,907
MRL/MpJ	N/A	*M. m. domesticus* and *M. m. castaneus*	*M. m. domesticus*	Classical	B	E	1,777,206	2,001,406
MsM/MsJ	*M. m. molossinus*	*M. m. musculus*	*M. m. musculus*	Wild-derived	A	A	1,322,823	1,754,050
NOD/ShiLtJ	N/A	*M. m. castaneus*	*M. m. domesticus*	Classical	B	E	1,166,675	1,347,448
NOR/LtJ	N/A	*M. m. castaneus*	*M. m. domesticus*	Classical	B	E	1,081,493	1,505,559
NZB/BINJ	N/A	*M.m. castaneus* and *M. m. domesticus*	*M. m. domesticus*	Classical	C	E	669,641	1,475,349
PWD/PhJ	*M. m. musculus*	*M. m. musculus*	*M. m. musculus*	Wild-derived	A	B	1,455,566	2,162,093
SJL/J	N/A	*M. m. domesticus*	*M. m. domesticus*	Classical	C	E	954,850	1,597,060

PHYLOGENETIC AND SEQUENCE ANALYSIS OF IGHV AND IGKV GENES IN WILD-DERIVED STRAINS

A phylogenetic analysis of IG variable region genes among wild-derived strains was performed through a multiple sequence alignment of inferred IGHV[4] and IGKV sequences among LEWES/EiJ, CAST/EiJ, MSM/MsJ, and PWD/PhJ. These strains represented the four major *Mus* sub-species origins: *M. m. domesticus*, *M. m. castaneus*, *M. m. molossinus* (a natural hybrid of *M. m. musculus* and *M. m. castaneus*), and *M. m. musculus*, respectively. Phylogenetic trees, one for IGHV and one for IGKV, were generated for all sequences inferred in these strains. Clusters of sequences were identified and grouped across the two phylogenetic trees using a threshold of 0.1, defined as the fraction of the maximum branch length. A multiple sequence alignment was used to generate a consensus sequence for each cluster, the nucleotide divergence of each sequence in a given cluster was then computed relative to the consensus, with divergence defined as the fraction of diverged positions in the alignment.

RESULTS

SELECTING MOUSE STRAINS TO REPRESENT DIVERSE SUB-SPECIES ORIGINS AND IGK/L HAPLOTYPES

The mouse has been used in genetic studies, random mutagenesis experiments, the development of inbred lines, and the direct engineering of the genome through knock-in, knockout, and transgenic techniques for over one hundred years[46]. As a result, many mouse strains are available for use in biomedical research. For example, C57BL/6, the most common and best-studied classical laboratory strain today[81], has been a popular model organism in immunology, with various knockout lines available and its genome sequenced by the Mouse Genome Sequencing Consortium[46]. Another common strain, BALB/c, served as an early model organism used to induce plasmacytomas and monoclonal Ab production[82,83]. Other strains, such as NOD/ShiLtJ, SJL/J, and MRL/MpJ, are used to model autoimmune disorders such as autoimmune type 1 diabetes, experimental autoimmune encephalomyelitis, and systemic lupus erythematosus and Sjogren's syndrome[84,85]. In addition, wild-derived mouse strains are often used to incorporate wild mouse genetics into laboratory strains by creating F1 hybrids[48]. Given the diverse breeding history of these strains[51,52,57,86 41]we expected the genetic diversity of the IG light chains to resemble the diversity observed in the IG heavy chain[4,87].

To select strains that captured IG light chain germline diversity, we examined the predicted sub-species origins and SNP-based haplotypes of the

mouse IG light chain loci[42,43] (Table 2.1, Figures 2.1, 2.2) across classical and wild-derived strains. We leveraged early studies that reported the existence of alternate IGK and IGL haplotypes across mouse strains[15,28,88–90], as well as genome-wide SNP data available for 62 wild-derived laboratory strains and 100 classical strains[42]. To account for the different *Mus* subspecies, we included strains with IG loci predicted to represent the three major *Mus* subspecies, *M. castaneus, M. domesticus, and M. musculus,* which form the genetic background for classical inbred laboratory mouse strains[54,56]. In addition, we chose CAST/EiJ, LEWES/EiJ, MSM/MsJ, and PWD/PhJ to represent wild-derived mouse strains, in which we have previously inferred germline IGHV, IGHD, and IGHJ genes[4]. In total, we sequenced the IGK and IGL repertoires of 18 different mouse strains representing three SNP-predicted IGK haplotype groups and five SNP-predicted IGL haplotype groups[80] (Figures 2.1, 2.2).

MOUSE LIGHT CHAIN VARIABLE GENES ARE UNDERREPRESENTED IN GERMLINE GENE DATABASES

First, we inferred germline light chain repertoires was performed across our selected mouse strains (Figure 2.3). We used IgDiscover to infer each strain's IGKV germline sequences and our previous clustering method[4] to infer IGLV

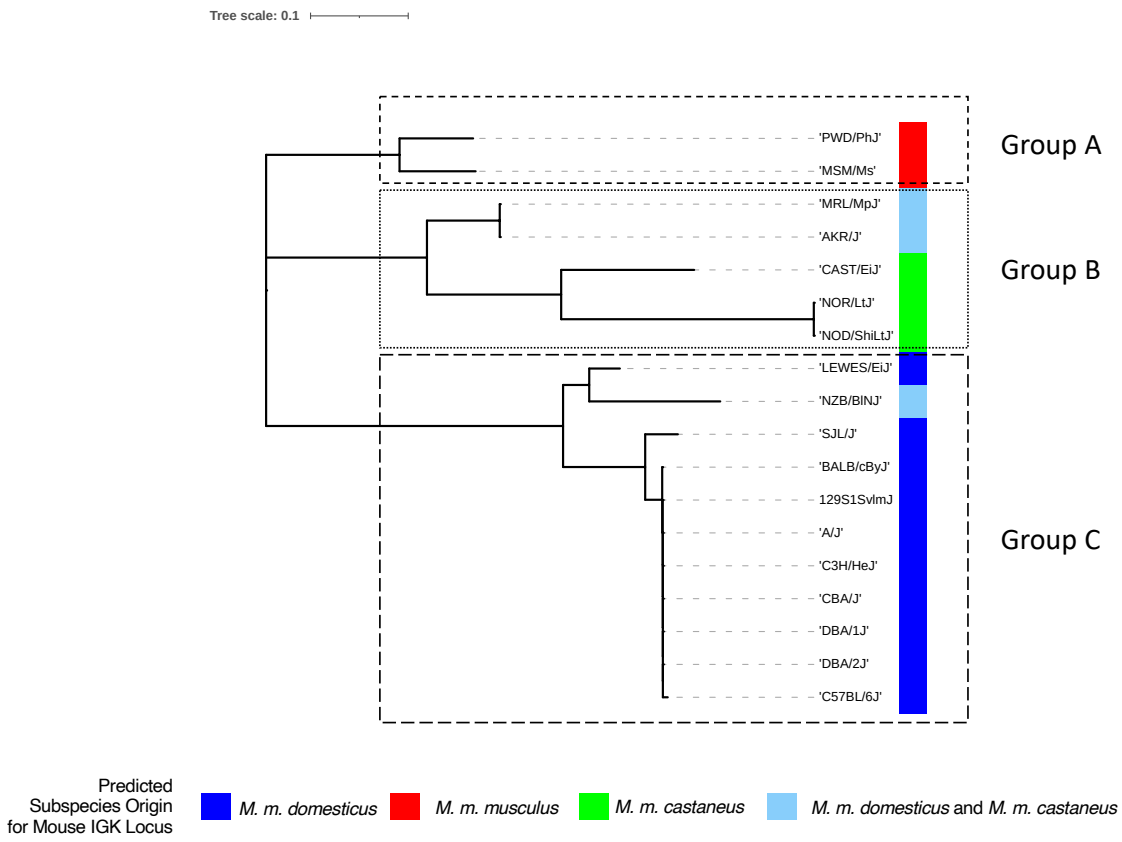


Figure 2.1. IGK haplotype phylogenetic tree generated from SNP data spanning the IGK locus[80].

Color next to strains reflects the predicted subspecies origin for each strain's IGK locus. Boxed clades represent three potential shared IGK haplotypes. Group A (PWD/PhJ, MSM/MsJ), group B (MRL/MpJ, AKR/J, CAST/EiJ, NOR/LtJ, NOD/ShiLtJ), and group C (LEWES/EiJ, NZB/BINJ, SJL/J, BALB/cByJ, 129S1/SvlmJ, A/J, C3H/HeJ, CBA/J, DBA/1J, DBA/2J, C57BL/6J).

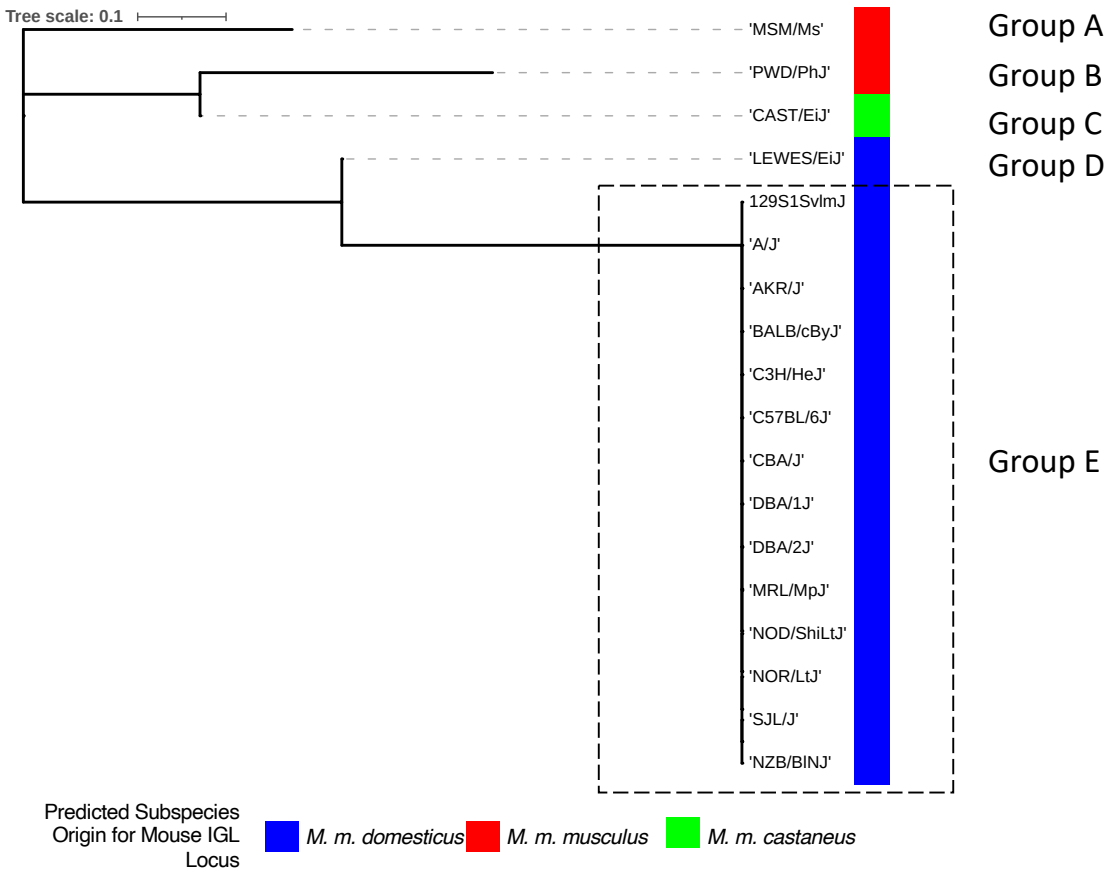


Figure 2.2. IGL haplotype phylogenetic tree generated from SNP data spanning the IGL locus[80].

Color next to strains reflects the predicted subspecies origin for each strain's IGL locus. Boxed strains (Group E) represent a potentially shared IGL haplotype.

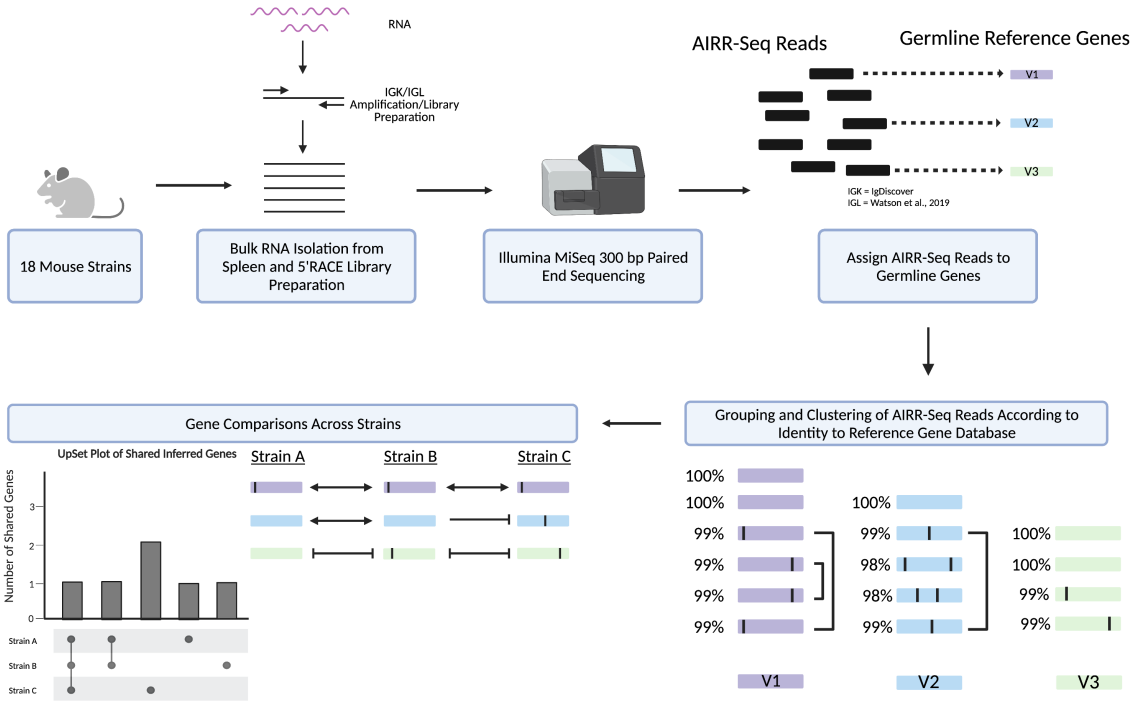


Figure 2.3. Experimental overview for IGKV/LV Germline Gene Inference.

Briefly, RNA was extracted from 18 different inbred mouse strains, then 5'RACE was used to generate IGK and IGL sequencing libraries. Next, libraries were sequenced on an Illumina MiSeq instrument (2x300 bp PE Sequencing), then the AIRR-seq reads were assigned to germline genes using either IgDiscover (IGKV) or our in-house pipeline (IGLV). After gene assignment, AIRR-seq reads were grouped and clustered according to their identity to genes in the IMGT reference gene database. Lastly, we visualized how the light chain germline gene repertoires were shared across strains using UpSet plots.

germline sequences. All sequences were also validated using TIgGER[66]. To benchmark the performance of our inference approach, we first assessed how well our C57BL/6J inferences compared to known IMGT C57BL/6 IGKV and IGLV germline sequences. 91/91 IGKV inferred sequences matched IMGT C57BL/6 IGKV germline sequences with 100% identity. Additionally, our 3 C57BL/6J IGLV inferences were 100% identical to IMGT IGLV sequences. However, since the three IMGT IGLV sequences that matched our inferences were derived from BALB/c[7], we also compared these three inferences to the mouse reference genome, mm10, derived from C57BL/6. The 3 IGLV C57BL6/J inferences matched mm10 with 100% identity.

Across the 18 mouse strains, 1582 IGKV and 63 IGLV sequences were inferred, representing 459 and 22 unique IGKV and IGLV sequences, respectively. The sizes of inferred IGKV germline gene sets varied across strains, from 105 in NZB/BlNJ to 62 in NOD/ShiLtJ (Figure 2.4). In contrast, the numbers of inferred IGLV germline genes were more conserved across strains. Three IGLV germline sequences were inferred for all classical laboratory strains and PWD/PhJ. However, LEWES/EiJ, MSM/MsJ, and CAST/EiJ had > 3 genes inferred from their repertoire data (Figure 2.5).

Of the 459 and 22 IGKV and IGLV unique sequences inferred across strains, 67.8% (n=311, IGKV) and 59% (n=13, IGLV) were undocumented in IMGT (Figures 2.6 A,B).

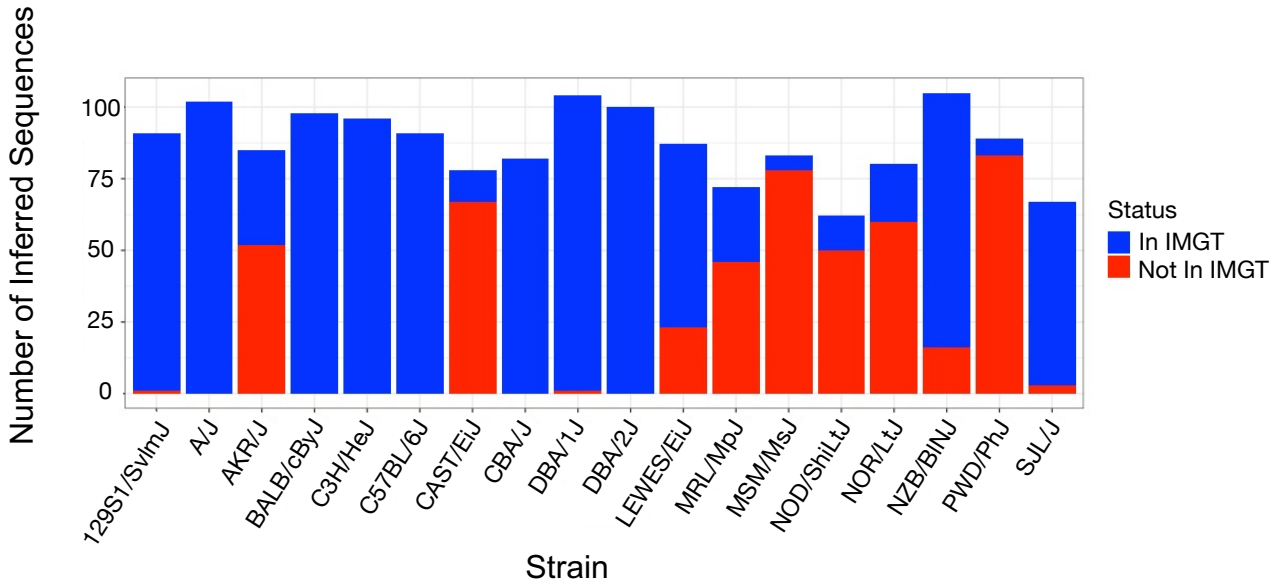


Figure 2.4. Inferred IGKV germline sequences present or absent in the IMGT Gene Database.

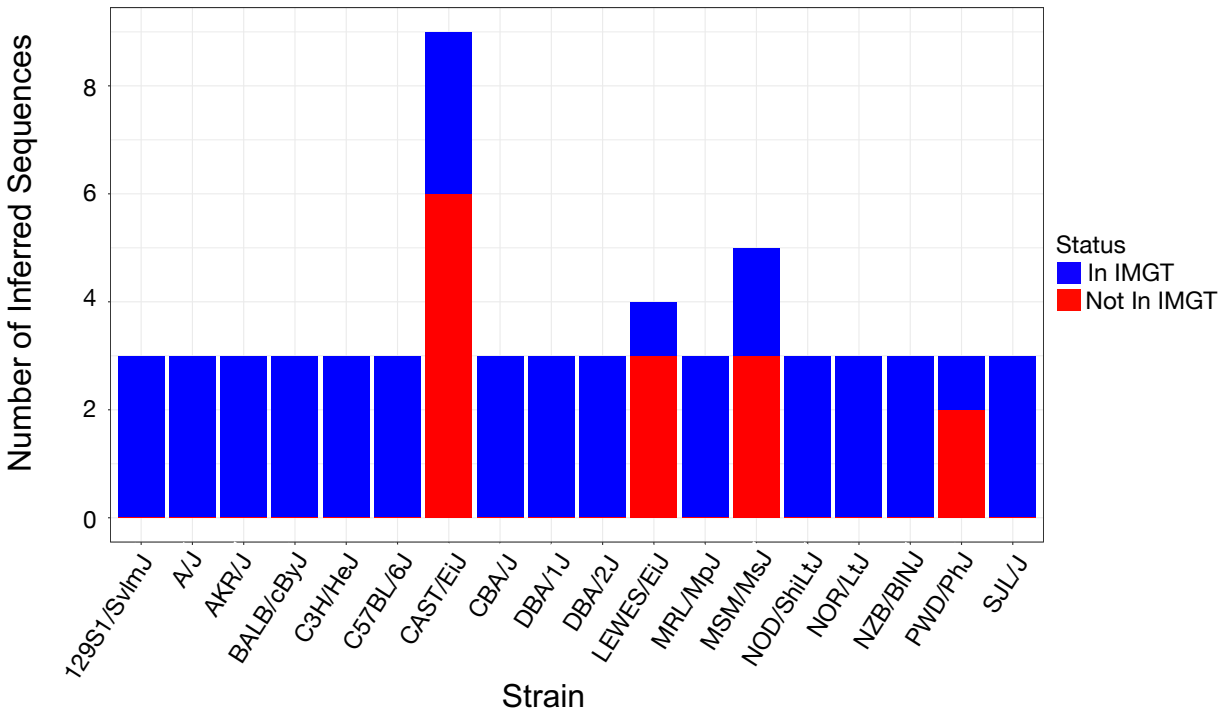


Figure 2.5. Inferred IGLV germline sequences present or absent from the IMGT Gene Database.

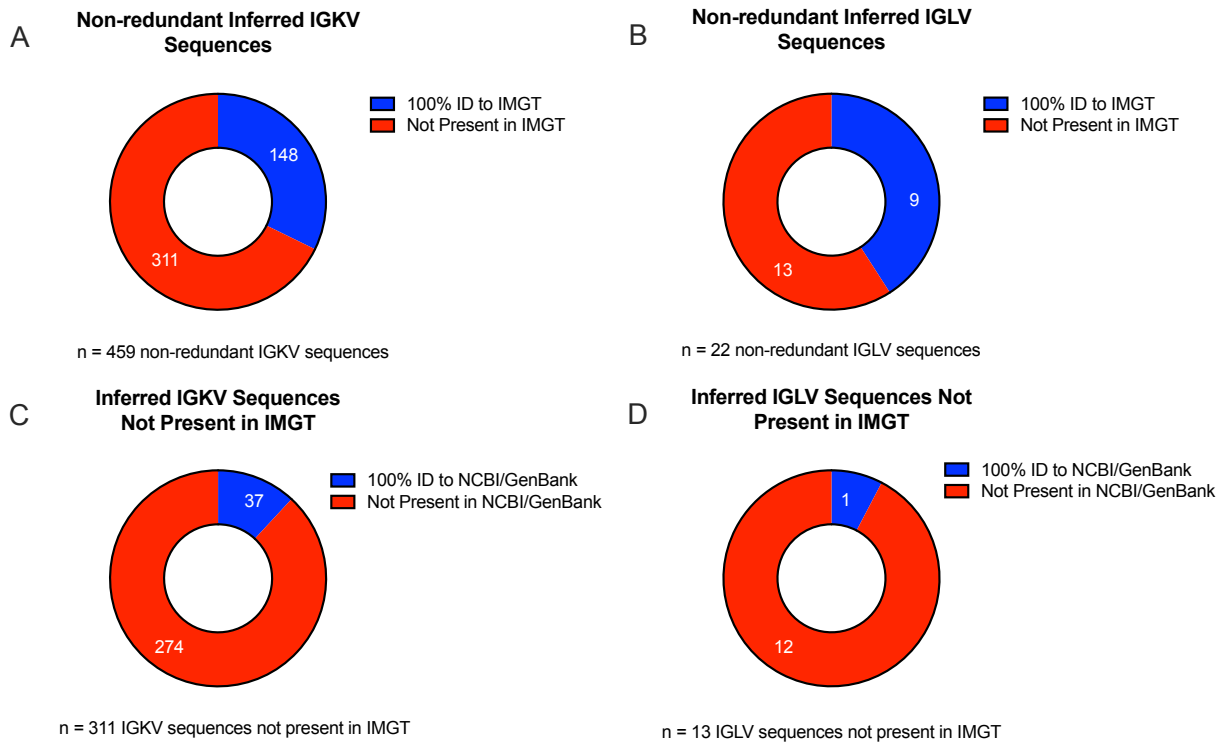


Figure 2.6. Presence and absence of non-redundant IGKV and IGLV inferred sequences in existing gene databases.

(A and B) Non-redundant IGKV (A) and IGLV (B) sequences present/absent in IMGT gene database. (C and D) Donut plots depicting the 311 IGKV (C) and 13 IGLV (D) inferences missing from IMGT and their presence/absence in NCBI/GenBank.

A fraction of these non-IMGT alleles were identified in NCBI GenBank with 100% identity: 12% (n=37) of IGKV and 8% (n=1) of IGLV (Figure 2.6 C,D). The number of undocumented (non-IMGT) alleles varied by strain (Figures 2.4, 2.5). For IGKV, while a significant fraction of non-IMGT alleles were inferred from wild-derived strains, there were several biomedically relevant classical strains in which the majority of sequences are not curated in IMGT (Figure 2.4), likely reflecting divergence from the C57BL/6 haplotype, as has been previously noted[70]. Strains such as NOD/ShiLtJ, NOR/LtJ, AKR/J, and MRL/MpJ, commonly used to model autoimmune disorders, have poor IGKV germline representation in IMGT (Figure 2.4)[84,91,92]. Sequence alignment of these strains' inferred IGKV germline sequences to the IMGT database yielded percent identities ranging from 100% to 89.61% for IGKV, and 100% to 93.88% for IGLV (Figures 2.7, 2.8). Alignment percent identity was strain-dependent, as we observed significant IGKV sequence variation for the four wild-derived strains, and AKR/J, MRL/MpJ, NOD/ShiLtJ, NOR/LtJ, and NZB/BINJ. Of all the strains investigated, MSM/MsJ have the fewest IGKV germline sequences documented in IMGT (5/83), and CAST/EiJ have the fewest IGLV germline sequences (3/9). Collectively, we observe high levels of diversity currently unaccounted for in the IMGT gene database.

IGKJ germline inference across strains revealed a novel IGKJ2 allele, with a single T to C transition shared between PWD/PhJ and MSM/MsJ (Figure 2.9), strains of *M. m. musculus* subspecific origin, and members of group A in our IGK SNP-haplotype phylogeny (Figure 2.1).

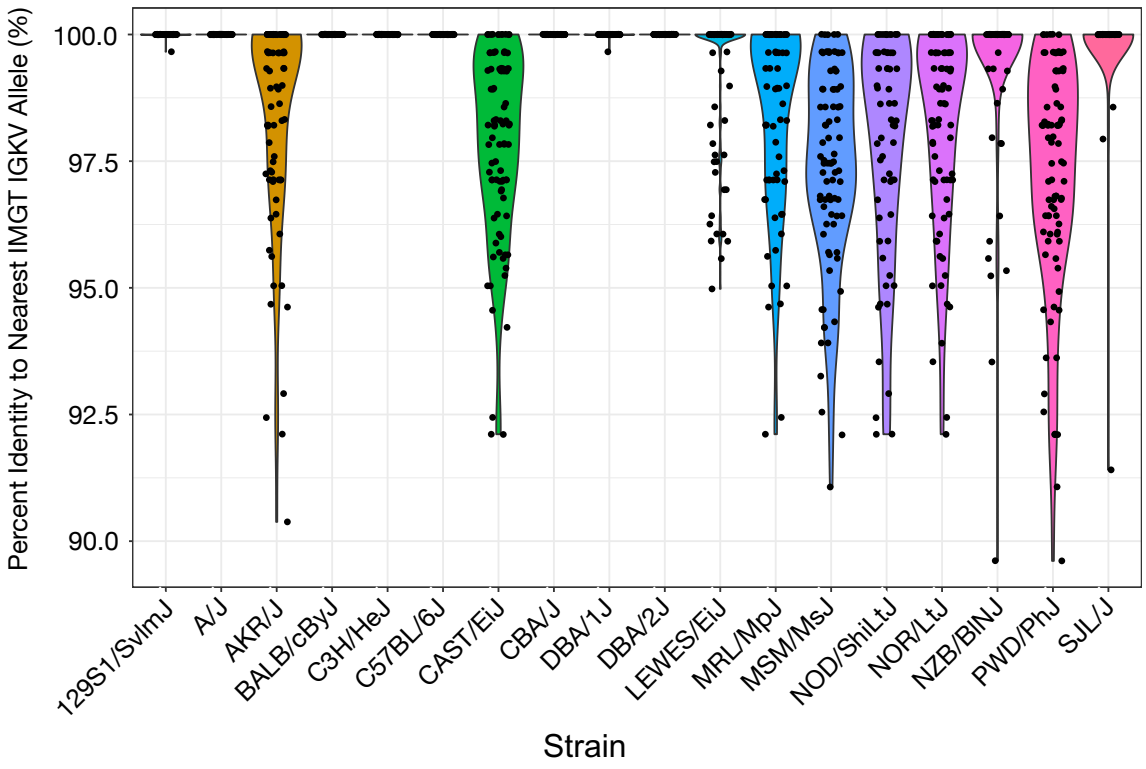


Figure 2.7. Sequence alignment percent identity of IGKV inferences to the IMGT Gene Database.

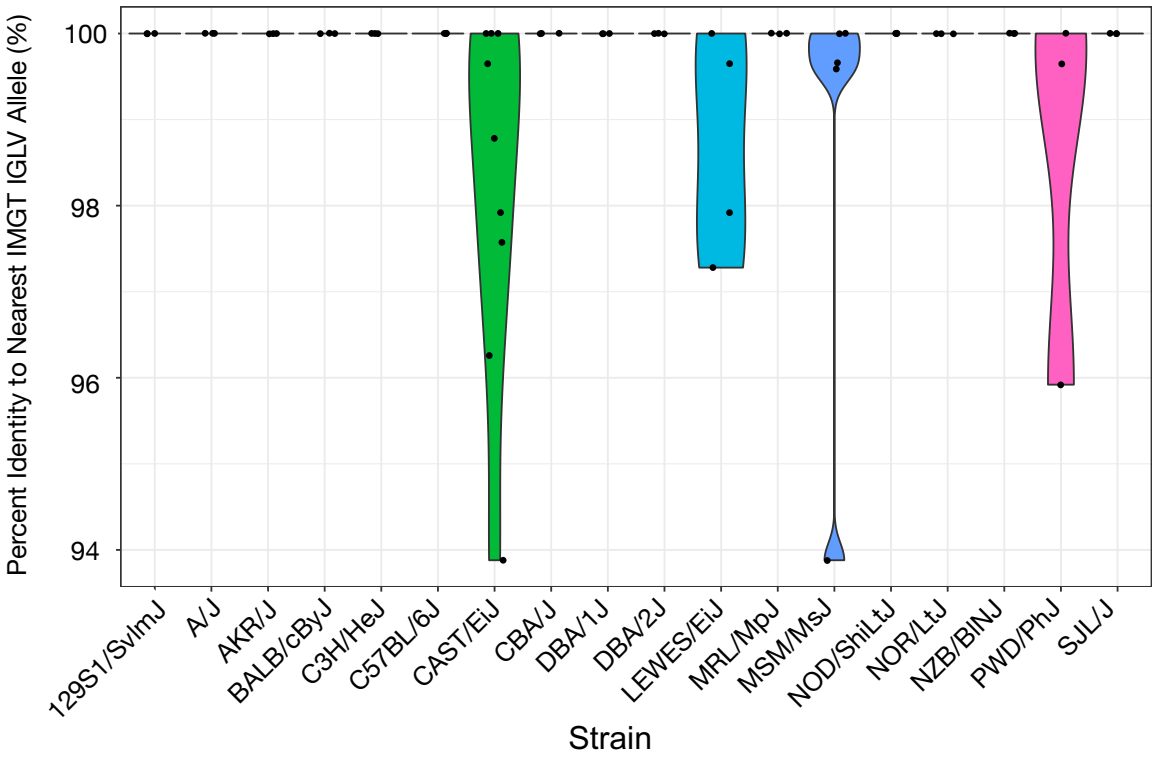


Figure 2.8. Sequence alignment percent identity of IGLV inferences to the IMGT Gene Database.

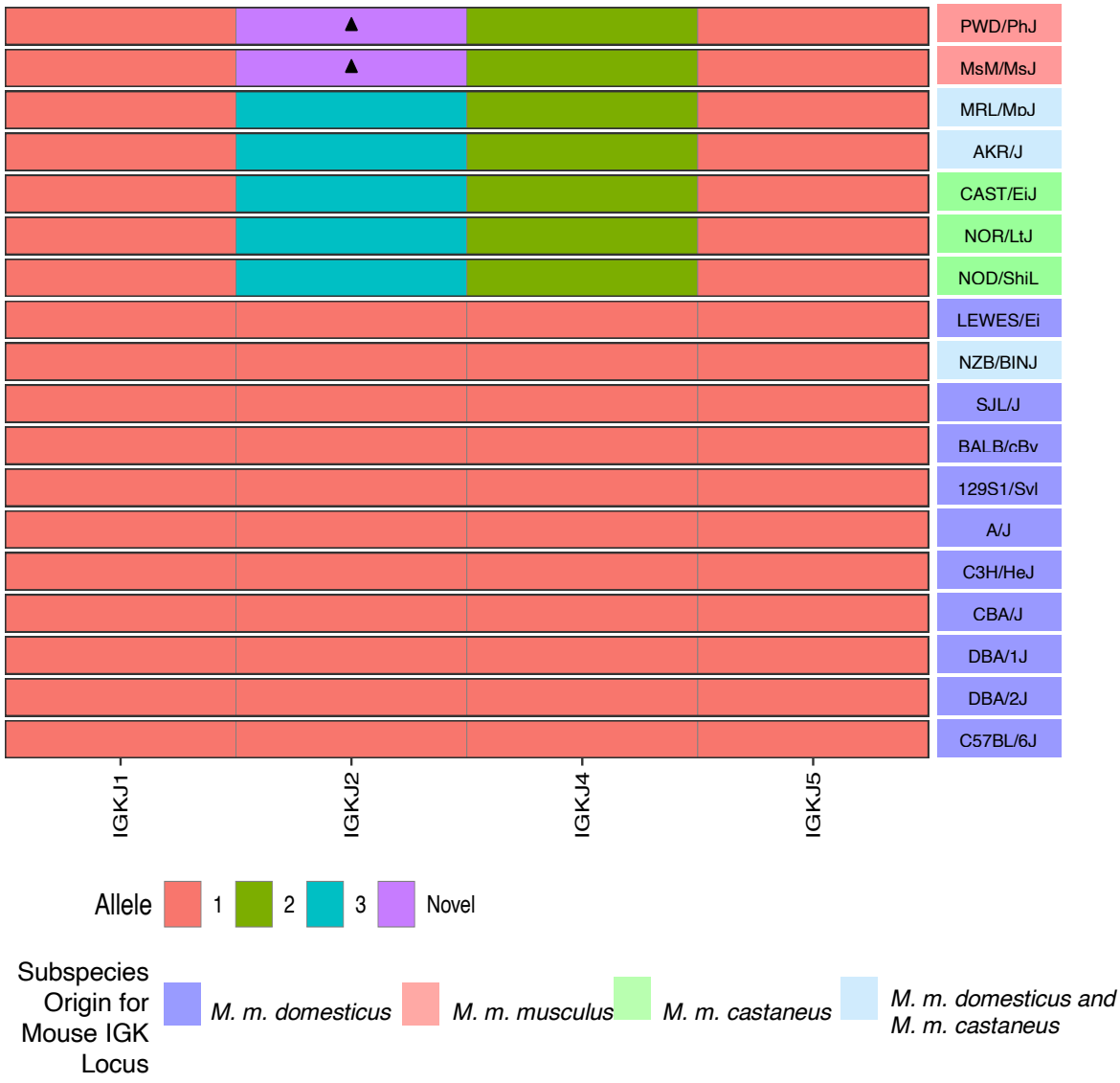


Figure 2.9. Inferred IGKJ novel alleles.

A single IGKJ2 novel allele was inferred for PWD/PhJ and MSM/MsJ (5'-TGTA**C**ACGTTCGGATCGGGGACCAAGCTGGAAATAAAAC-3'). Bolded position is novel SNP that diverges from IMGT reference sequence.

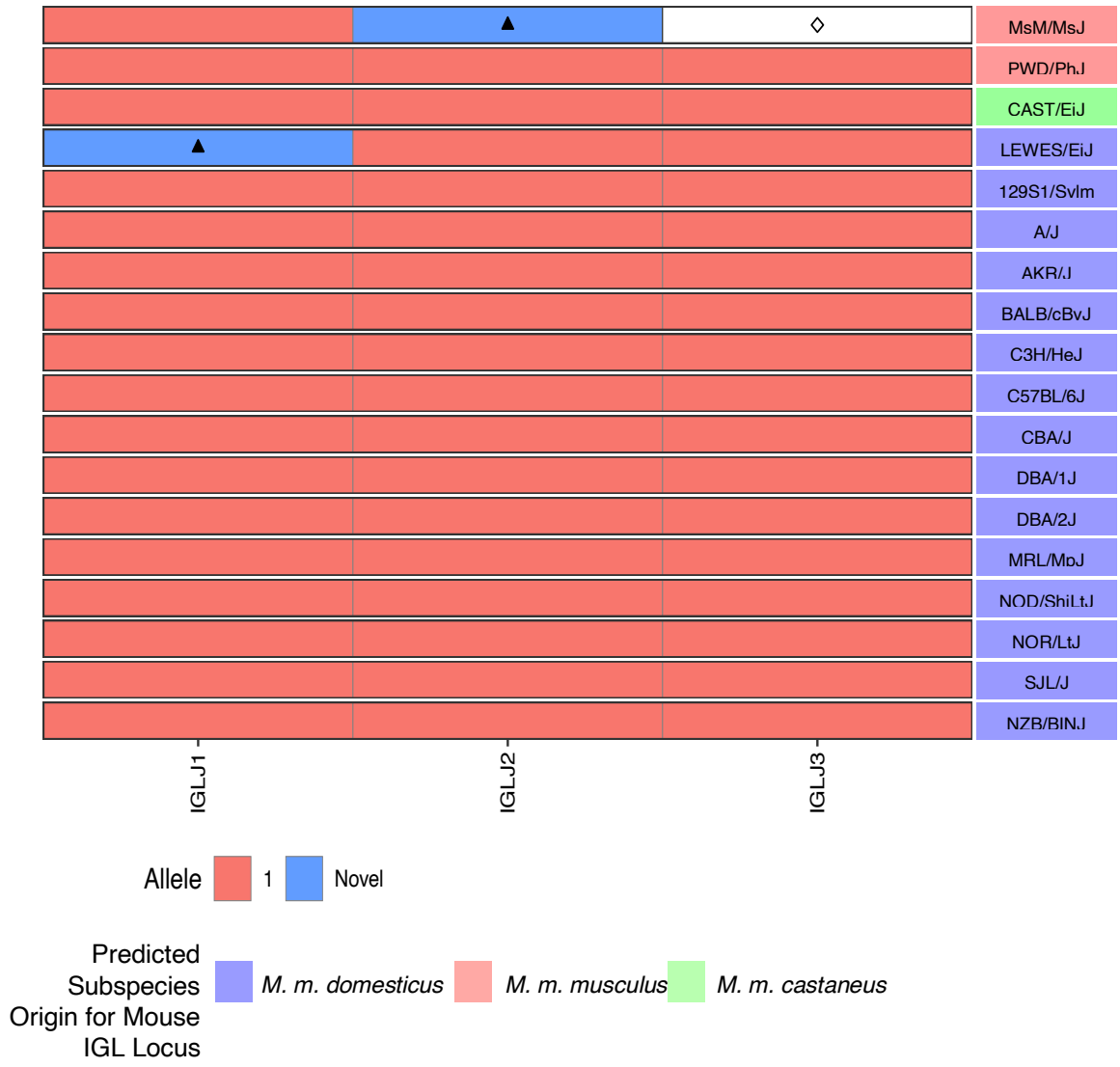


Figure 2.10. Inferred IGLJ novel alleles.

Two novel alleles were inferred for IGLJ: Novel IGLJ1 allele (5'-CTGGGTGTTCGGTGGAGGAACCAAA**T**TGACTGTCCTAG-3') and novel IGLJ2 (5'-TTATGTTTTCGGC**A**GTGGAACCAAGGTCACTGTCCTAG-3'). Bolded positions are novel SNPs that diverge from IMGT reference sequence.

We inferred two novel IGLJ alleles among the wild-derived strains (Figure 2.10). A novel IGLJ1 allele was inferred for LEWES/EiJ, and a novel IGLJ2 allele was inferred for MSM/MsJ. Both novel IGLJ alleles differ from the IMGT reference sequence by a single nucleotide. Interestingly, we did not infer any IGLJ3 alleles in MSM/MsJ; all MSM/MsJ J gene usage was restricted to IGLJ1*01 and the novel IGLJ2 allele. Overall, we found our SNP-haplotype groupings reflected in the IGKJ and IGLJ allelic variation across strains.

INTER-STRAIN IGKV/IGLV GERMLINE DIVERSITY

We next considered the extent to which inferred IGKV and IGLV sequences were shared among strains (Figures 2.11 and 2.12). Across IGKV germline gene sets, the most strain-specific sequences were observed among the wild-derived strains, CAST/EiJ (n=58), PWD/PhJ (n=57), MSM/MsJ (n=55), and LEWES/EiJ (n=21), with an additional 19 sequences uniquely common to PWD/PhJ and MSM/MsJ (Figure 2.11). There were fewer unique sequences in each of the classical laboratory strains. For example, only 9 unique sequences were seen in the NOD/ShiLtJ strain, which had the highest number of unique sequences amongst the classical inbred strains. Instead, we observed large sets of sequences that were identical across many strains (Figure 2.11). The most extensive shared sequence set comprised 27 inferred IGKV germline sequences inferred from 11 different strains (SJL/J, CBA/J, LEWES/EiJ, C57BL/6J, 129S1/SvImJ, C3H/HeJ,

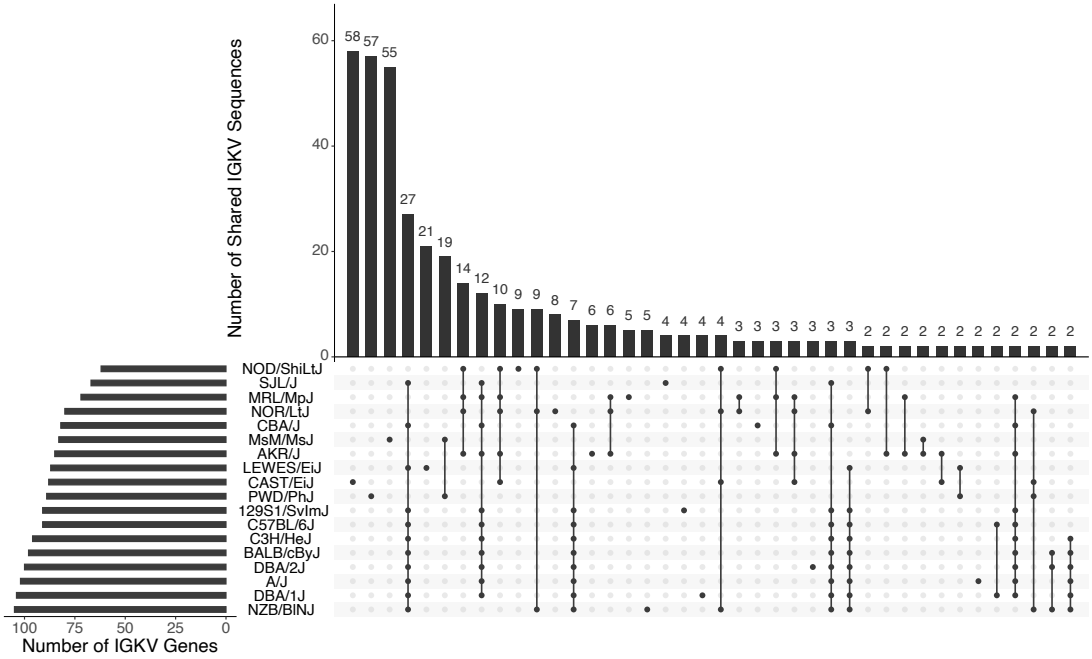


Figure 2.11. IGKV UpSet plot.

Depicts the size of the germline IGKV set from 18 mouse strains and the number of sequences that were unique to a given strain (dot) or shared among strains (connected dots).

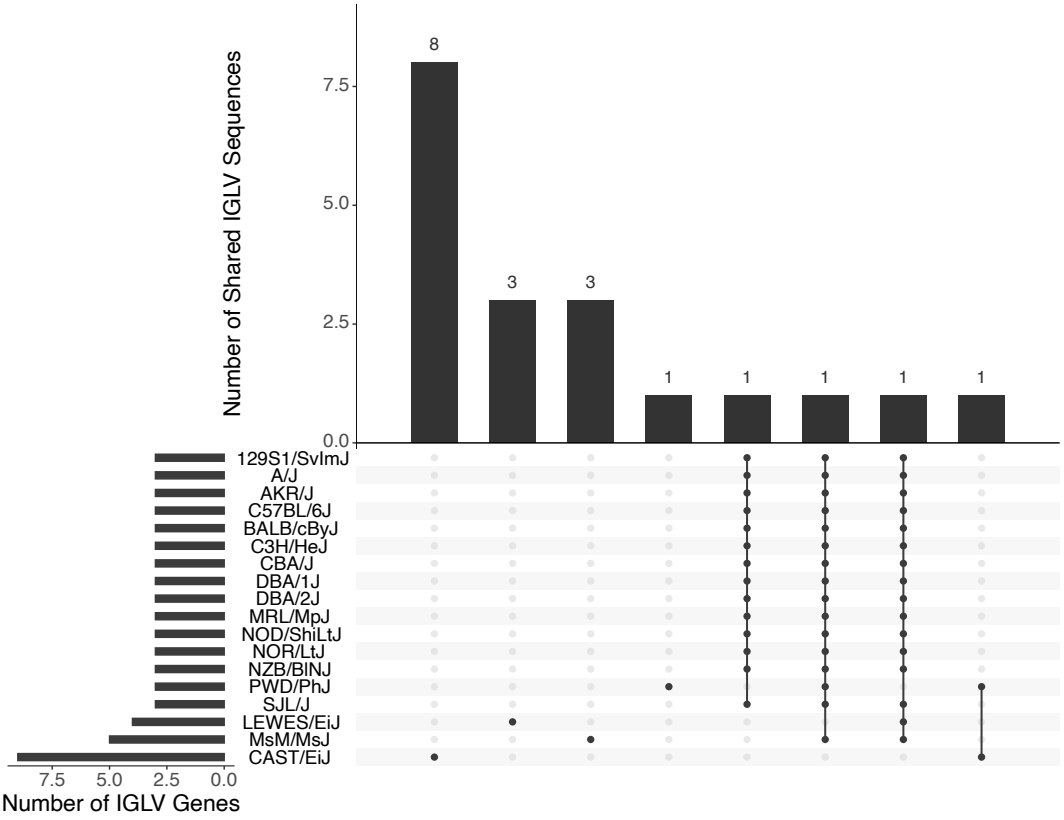


Figure 2.12. IGLV UpSet plot.

Depicts the size of the germline IGLV set from 18 mouse strains and the number of sequences that were unique to a given strain (dot) or shared among strains (connected dots).

BALB/cByJ, DBA1/J, DBA2/J, A/J, and NZB/BlNJ). This degree of allele sharing was suggestive of the presence of potentially shared haplotypes. We assessed our IGK SNP-Predicted haplotype phylogenetic tree (Figure 2.1) and found all 11 strains in Group C. NZB/BINJ was the only strain in Group C predicted to have a different sub-specific origin for the IGK locus. The Mouse Phylogeny Viewer[80] reports a *M. m. domesticus* and *M. m. castaneus* sub-specific origin for the NZB/BINJ IGK locus, which contrasts the other strains' *M. m. domesticus* IGK locus sub-specific origin.

Another group of strains, NOD/ShiLtJ, NOR/LtJ, MRL/MpJ, and AKR/J, formed a different cluster with 14 shared IGKV germline sequences (Figure 2.11). These four strains fall into Group B of our SNP-Predicted haplotype phylogenetic tree (Figure 2.1) and represent an IGK locus sub-specific origin of either completely *M. m. castaneus*, or a mixture of *M .m. castaneus* and *M. m. domesticus*. Finally, PWD/PhJ and MSM/MsJ, both wild-derived strains with a *M. m. musculus* sub-specific origin, were located in the Group A cluster, with 19 unique IGKV sequences shared between themselves (Figure 2.1, 2.11). We also examined inter-strain diversity at the level of IGKV gene family by comparing the number of inferred IGKV sequences for each gene family across strains (Supplemental Figure 1). Overall, the IGKV4 subfamily was most variable in size and the most abundant family across strains, whereas IGKV20 was the least abundant and only inferred in PWD/PhJ, MSM/MsJ, and CAST/EiJ (Supplemental Figure 1).

In each of the 14 classical strains, a total of 3 IGLV sequences were inferred, which is consistent with the number of genes found in the C57BL/6 mm10 genome[93]. These IGLV inferences were identical across the 14 classical strains (Figure 2.12), supporting the predicted sub-species origins and SNP-Haplotype phylogenetic tree clustering (Figure 2.2). In contrast to the classical strains, additional putative genes were inferred in three of the wild-derived strains, CAST/EiJ, LEWES/EiJ, and MSM/MsJ, totaling 9, 4, and 5 inferred genes for each strain respectively. Phylogenetic analysis of inferred IGLV sequences revealed that inferred genes unique to CAST/EiJ formed an additional outgroup to all other IGLV1, IGLV2, and IGLV3 gene sequences characterized in the classical strains (Figure 2.13). In addition to these IGLV paralogs, the IGLV2 sequences inferred in CAST/EiJ and PWD/PhJ, and the IGLV3 sequences inferred from CAST/EiJ, PWD/PhJ, and LEWES/EiJ, differed from those characterized in the classical strains, likely representing allelic variants. Thus, taken together, the majority of IGLV germline diversity observed in the animals studied here came from wild-derived strains.

To help validate the relationship between predicted IG haplotypes, sub-species origin, and the inferred germline sets across strains, we performed all-by-all pairwise comparisons of inferred IGKV and IGLV sequences across strains to group strains by sequence similarity. We reasoned that inferred IGKV/IGLV sequences within strains sharing predicted haplotypes would have higher sequence similarities. We found that mean pairwise sequence similarities varied

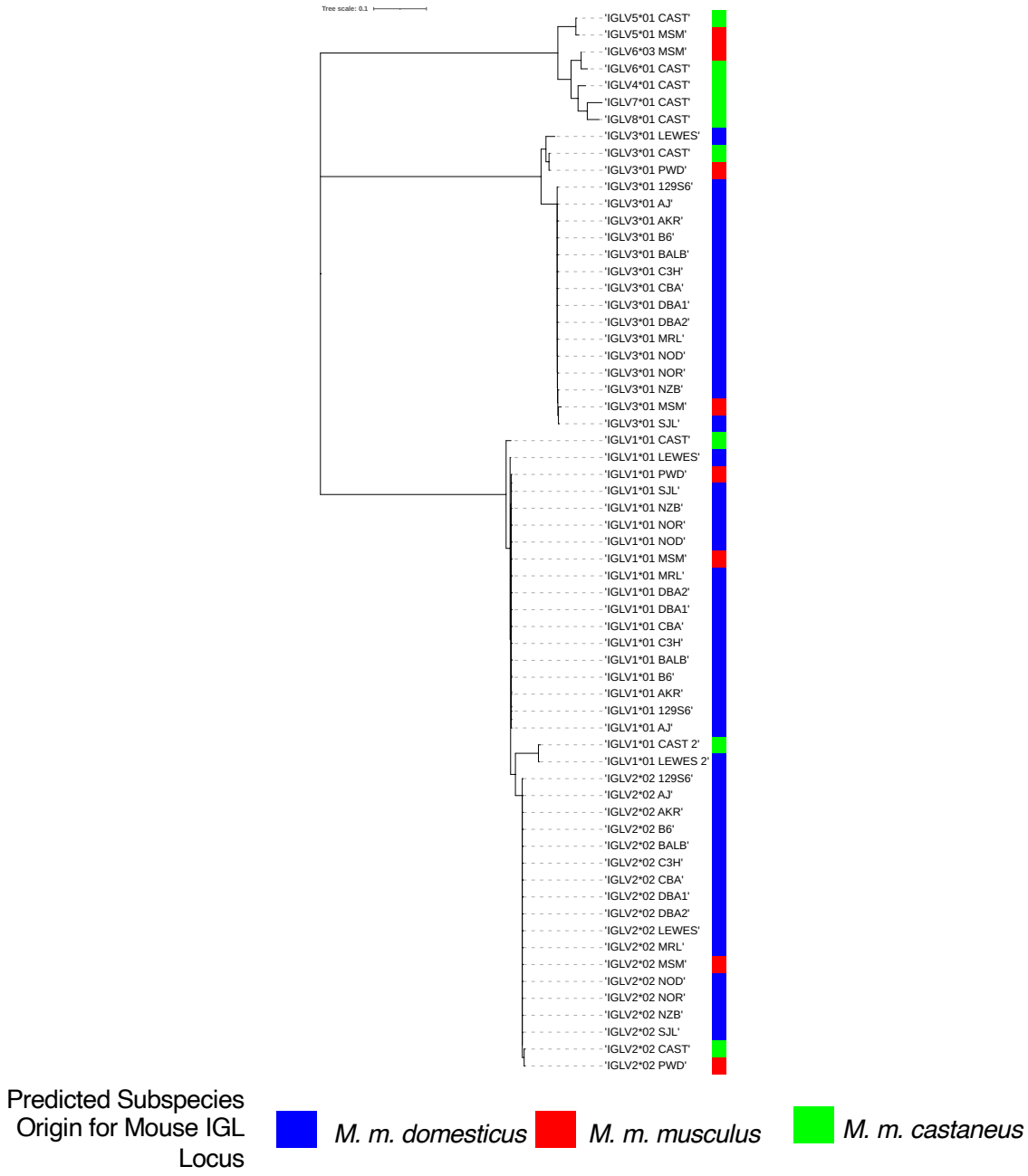


Figure 2.13. Phylogenetic tree of inferred IGLV germline sequences.

considerably. For IGKV, the highest similarity (99.99%) observed was between C3H/HeJ and A/J, whereas the divergent comparison was between NOR/LtJ and PWD/PhJ (95.30%). We used hierarchical clustering to group strains based on mean pairwise similarities, and the results corresponded to the three haplotype groups obtained from our SNP-Haplotype phylogenetic tree (Figure 2.14). IGLV inferences were much more conserved across strains. The 4 IGLV germline sequences inferred for LEWES/EiJ are consistent with Potter et al., who reported that wild *Mus musculus domesticus* had at least three IGLV genes[94].

DIVERSIFICATION OF IGHV AND IGKV SEQUENCES AMONG WILD-DERIVED MOUSE STRAINS

Gene evolution through duplication and diversification events have helped shape the diversity of the mouse immunoglobulin genes. In mice and humans, light chain gene rearrangement begins on the kappa chain, but it follows heavy chain rearrangement. If this initial rearrangement is auto-reactive, then the gene organization of the kappa chain locus permits additional gene rearrangements through a process known as receptor editing to form a rearrangement that is not autoreactive[5]. While both the heavy and kappa chain repertoires have many V genes that have evolved through these processes, the kappa chain repertoire

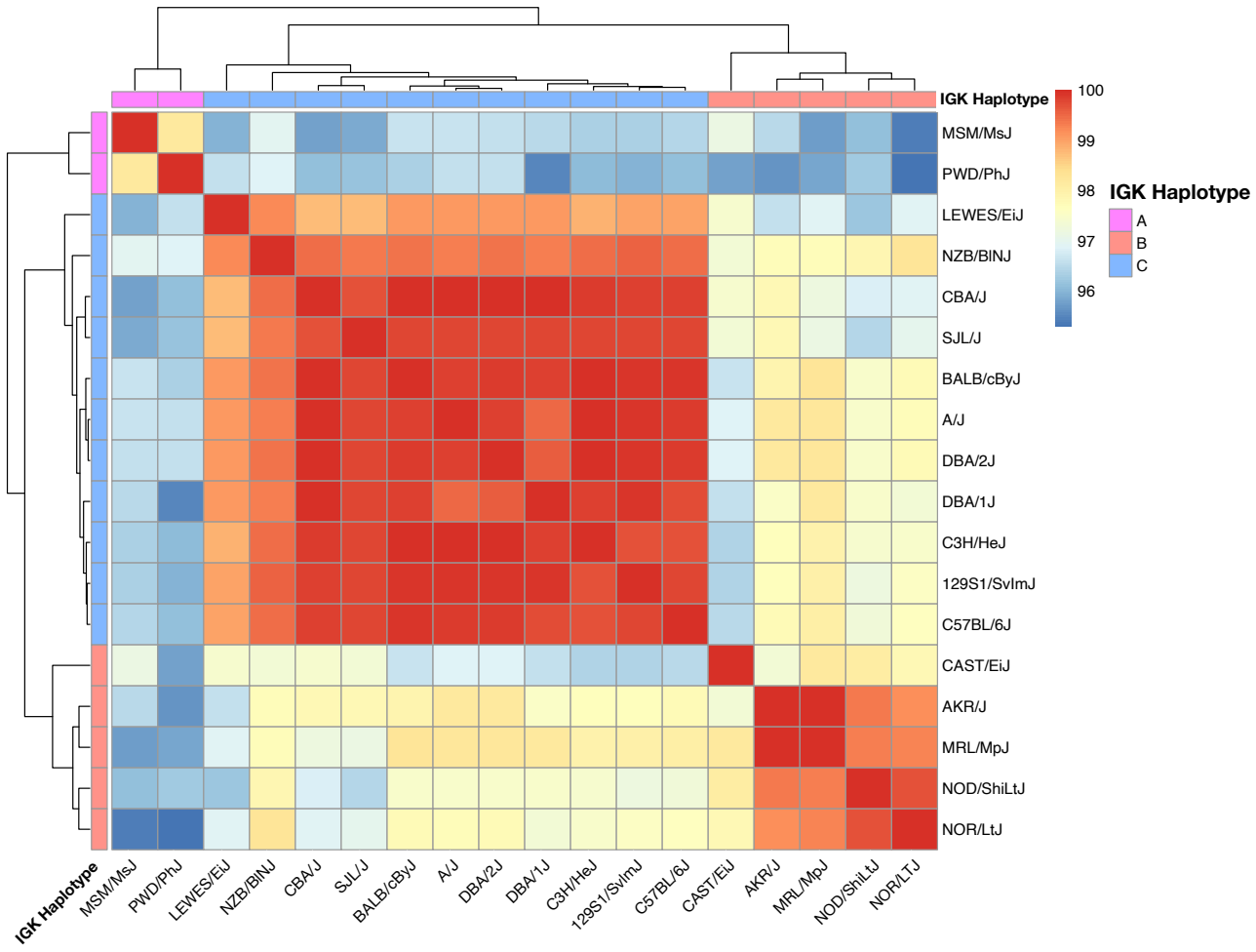


Figure 2.14. Inferred IGKV germlines pairwise comparison.

Mean percent identities from all-by-all BLAT pairwise comparisons of inferred IGKV germline genes across strains. Strains cluster according to the assigned IGK haplotype group from the SNP-Predicted Haplotype phylogeny.

contains less inherent germline diversity than the heavy chain[5]. It has been hypothesized that the reduced germline diversity of the kappa chain repertoire has evolved to limit self-reactivity, while the heavy chain repertoire has evolved to increase diversity. Support for this hypothesis stems mainly from human AIRR-seq data, in which there is some evidence that there is less allelic diversity in IGKV compare to IGHV[5]. Others have also suggested that the patterns of diversification and divergence are potentially different between heavy and light chain immunoglobulin genes. For example, Schwartz et al. analyzed amino acid sequences for human germline heavy and light chain genes and concluded that heavy and lambda V genes had higher diversities compared to kappa V genes[95]. Our AIRR-seq dataset provides us with the opportunity to compare diversification and divergence patterns between germline heavy and light chain gene sets across multiple mouse strains representing diverse mouse subspecies origins. Therefore, we hypothesized that the mouse kappa chain Ab repertoire would have decreased diversity in germline V genes compared to the heavy chain Ab repertoire as a way to minimize potential auto-reactivity.

One metric that can be used to compare germline sequence evolution of genes is the edit distance between two sequences, defined as the minimal number of mutations separating the two sequences[96]. In 2019, we inferred germline IGHV genes among wild-derived mouse strains representing the major *Mus* sub-species origins[4]. With the additional germline IGKV genes from the same wild-derived strains, we compared germline variable gene sequence divergence rates for both the IGH and IGK loci of the wild-derived strains using phylogenetic trees

constructed from multiple sequence alignments of germline IGHV and IGKV sequences. We observed that the inferences for LEWES/EiJ were more divergent compared to the other wild-derived strains for both the IGH and IGK loci, whereas PWD/PhJ had the lowest sequence divergence for both loci (Figures 2.15 and 2.16). Lastly, to estimate germline heavy and kappa chain repertoire diversity across the wild-derived strains, we compared sequence identity patterns across the two loci by performing sequence alignments within and between wild-derived strains, with the goal of detecting sequence variance patterns between the IGH and IGK loci. Overall, we observed that IGHV sequences had a lower mean percent identity both within and between strains than IGKV sequences and a greater variance than the IGK germline repertoire (Figure 2.17). This data suggests that the mouse IGH locus contains greater inherent germline sequence diversity than the IGK locus despite having a similar number of germline genes. In addition, the IGHV sequences had a greater variance than IGKV, indicating that the IGHV sequences are more variable across the wild-derived strains than the IGKV sequences.

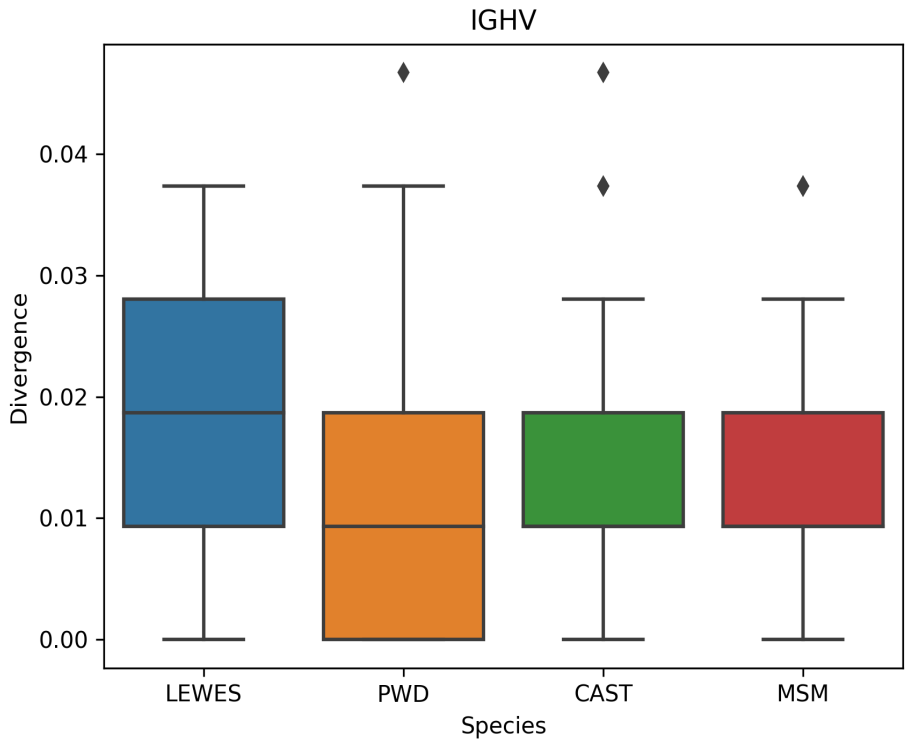


Figure 2.15. Germline IGHV gene divergence among wild-derived mouse strains representing the four major *Mus* sub-species origins.

Shown are the IGHV sequence divergences for LEWES/EiJ (*M. m. domesticus*), PWD/PhJ (*M. m. musculus*), CAST/EiJ (*M. m. castaneus*), and MSM/MsJ (*M. m. musculus*/*M. m. castaneus*).

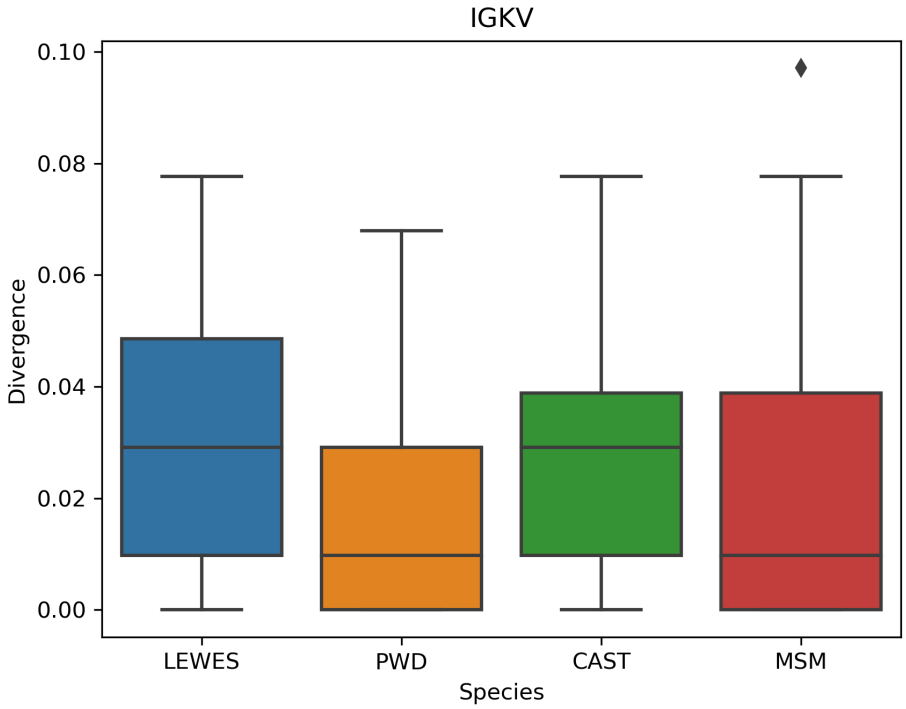


Figure 2.16. Germline IGKV gene divergence among wild-derived mouse strains representing the four major *Mus* sub-species origins.

Shown are the IGKV sequence divergences for LEWES/EiJ (*M. m. domesticus*), PWD/PhJ (*M. m. musculus*), CAST/EiJ (*M. m. castaneus*), and MSM/MsJ (*M. m. musculus*/*M. m. castaneus*).

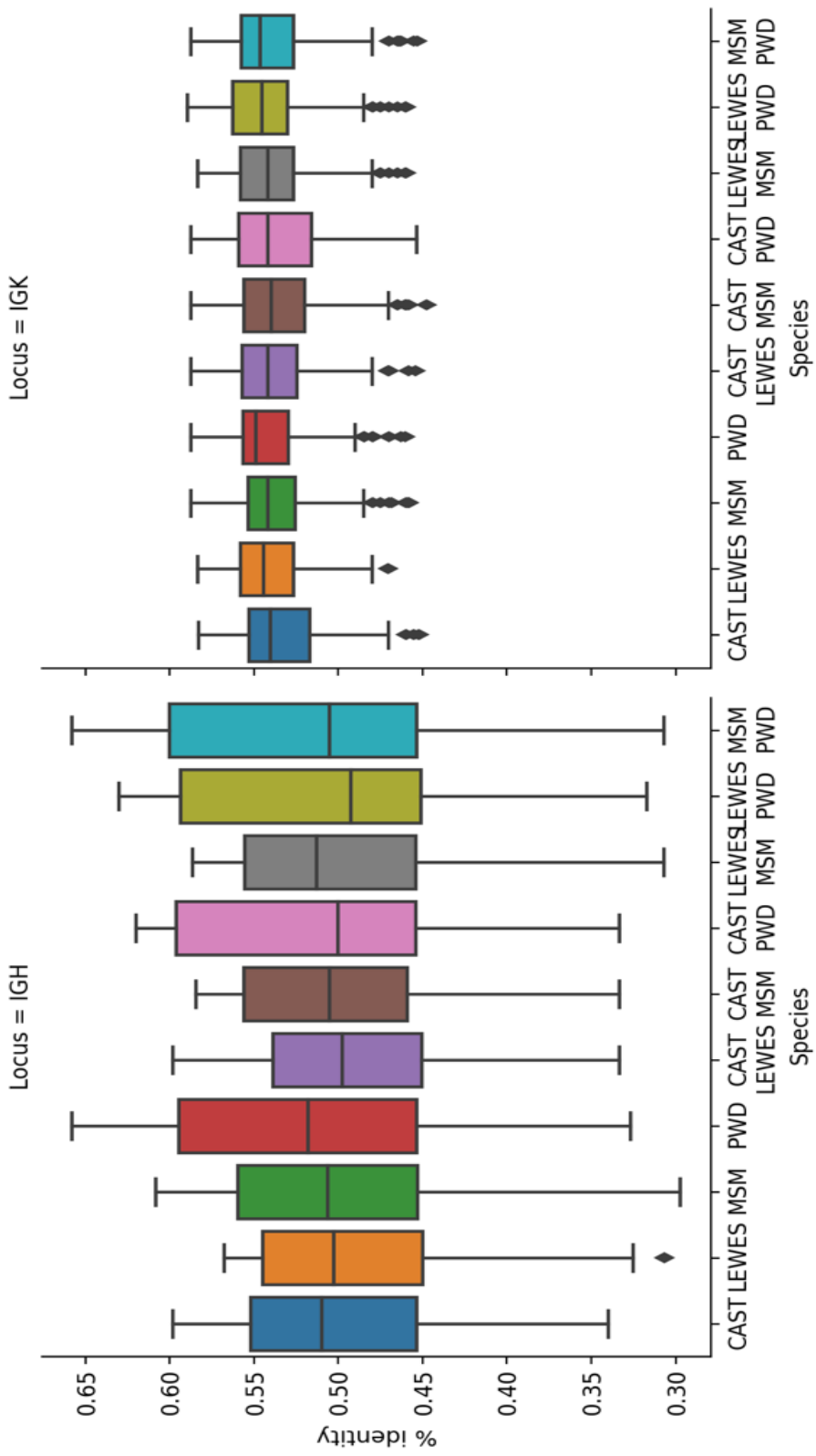

Locus = IGH
Locus = IGK
% identity
0.65
0.60
0.55
0.50
0.45
0.40
0.35
0.30
CAST
LEWES
MSM
PWD
Species

Figure 2.17. Germline gene divergence among wild-derived mouse strains representing the four major *Mus* sub-species origins.

Shown are within strain IGH (left panel) and IGK (right panel) germline gene variance.

DISCUSSION

This study was performed as a follow-up to our 2019 study in which we used AIRR-seq to infer IGHV, IGHD, and IGHJ genes of wild-derived strains representing each of the three major subspecies of the house mouse (CAST/EiJ: *M. m. castaneus*; LEWES/EiJ: *M. m. domesticus*; PWD/PhJ: *M. m. musculus*) and the *M. m. musculus/M. m. castaneus* hybrid strain MSM/MsJ. Overall, we found little overlap in germline IGHV repertoires among the wild-derived mouse strains and could not attribute all of the repertoire differences observed to variation in sub-specific origin. For example, while apparently *musculus*-derived strains did share IGHV sequences, many of these sequences were unique to each strain. Most importantly, the inferences were largely absent in existing gene databases such as IMGT. The diversity of germline IGHV sequences led us to hypothesize that extensive germline light chain variable genes would also be present. Therefore, in addition to sequencing the light chain variable genes for the wild-derived strains sequenced in 2019, we expanded the number of strains to include various strains used across the biomedical sciences that represented different SNP-predicted haplotypes for the light chain loci. The strains encompassed a variety of disease models to study infection, autoimmunity, diabetes, cancer, and regeneration[91,97].

One of the most fundamental tasks in analyzing AIRR-seq data is V, D, and J gene assignment, which is accomplished by performing alignment of AIRR-seq reads to germline sequences from a gene database. IMGT is the most commonly used database that curates the germline repertoires used for sequence alignment in AIRR-seq analysis. When comparing our inferred light chain germline

sequences to those curated in IMGT, we observed that many strains' germline sequences were absent from the database, and few undocumented sequences were found in NCBI/GenBank (Figures 2.4 and 2.5). In contrast to IGKV, the majority of genes characterized in IGLV among classical strains are curated in IMGT, including previously described wild-derived IGLV4 sequences detected in CAST/EiJ[98].

Critically, many of the inferred germline genes found to be absent from IMGT showed evidence of significant divergence from the closest curated allele in the database. For example, 59% (185/311) of non-IMGT IGKV sequences, and 62% (8/13) of non-IMGT IGLV sequences, had <98% identity to their nearest IMGT IGKV sequence. Thus, we would expect that these missing data could greatly impact the accuracy of germline gene assignment and somatic hypermutation (SHM) estimation in studies using these strains. Given that the IGK locus has similar complexity as the IGH locus, with different combinations of SNP-predicted sub-species origins and haplotypes, the IGK locus likely contains genetic variation similar to that observed in IGH between BALB/c, C57BL/6, and 129S1[69,87]. Our inferred IGKV germline sequences supported these three distinct haplotype clusters (Figures 2.1, 2.9, 2.11). For example, 11 strains shared 27 IGKV germline sequences, and all 11 strains belonged to Group C in the IGK SNP-Predicted haplotype phylogeny. We compared these strains to the historic IGK haplotypes identified using RFLP and observed that 9 of the 11 strains were previously designated the historical IGK^A haplotype[28]. We also observed 14 IGKV sequences shared among 4 strains in our dataset belonging to Group B (Figure 2.1),

containing the historical IGK^B haplotype[28]. Though CAST/EiJ does not share the 14 unique IGKV sequences with NOD/ShiLtJ, NOR/LtJ, MRL/MpJ, and AKR/J, the SNP-phylogeny data, and additional shared sequences between these strains, suggest that CAST/EiJ does indeed belong in haplotype Group B. Similar to the historic IGH^A and IGH^B haplotypes[40,68,69,99], our results suggest that the IGK loci of strains carrying the IGK^B haplotype are similar to one another and different from the loci of strains carrying the IGK^A haplotype.

Strain clustering according to the sub-species origin was also apparent after performing all-by-all pairwise sequence comparisons (Figure 2.14) and validated our SNP-phylogeny groupings. Of note, the 5 strains in Group B (AKR/J, MRL/MpJ, NOD/ShiLtJ, NOR/LtJ, and CAST/EiJ) exhibited significant sequence divergence from the IMGT alleles (Figure 2.7), highlighting the lack of representation for strains sharing this IGK haplotype. However, like the IGH locus, only a single complete IGK reference is available based on C57BL/6[9], illustrating an IGK haplotype not shared by all strains and only representative of a single sub-species origin.

Our cohort contained 5 SNP-predicted IGL haplotypes (Figure 2.2); however, only one haplotype had representation by more than one strain. All wild-derived strains, MSM/MsJ, PWD/PhJ, CAST/EiJ, and LEWES/EiJ were clustered into single-strain clades according to the SNP-predicted haplotype, while the 14 remaining classical laboratory strains clustered into a single haplotype. Our data supported the predicted haplotype clustering, with each wild-derived strain containing at least one unique IGLV allele not shared with other strains and classical laboratory strains sharing IGLV alleles amongst each other (Figure 2.12).

Of note, we did infer more than the 3 canonical IGLV germline sequences in CAST/EiJ, MSM/MsJ, and LEWES/EiJ, which was expected based on previous data hypothesizing the existence of additional IGLV genes in wild-derived strains[94]. A total of 13 novel sequences were identified among the wild-derived strains, with an allelic variant of IGLV2*02 shared between CAST/EiJ and PWD/PhJ. Inferred IGLV germline repertoires in classical strains were more conserved and composed of 3 IGLV genes.

Although expressed lambda chain genes account for only 3 to 5% of total serum IG, little is known regarding how the mouse lambda locus reduced in size[94]. Furthermore, though gene deletion events could cause a reduced IG lambda locus in mice, it is unclear where it occurred in mouse phylogenetic history. The rat IG loci, similar to the mouse IG loci, have not been extensively explored and characterized across strains. Reports suggest that the IGL locus in rat only contains a single IGLV gene and two IGLC genes making the rat IGL locus even smaller than mouse[100]. If we consider the rat and mouse to be distant relatives, this presents two possible scenarios that may have occurred during rat and mouse evolution. Either the mouse IGL locus expanded through duplications, or the rat IGL locus decreased in size through deletions.

Overall, in conjunction with other published studies, the data presented here demonstrate that the germline heavy and light chain repertoires are not conserved across biomedically relevant mouse strains[4,87]. Our phylogenetic analysis of heavy and kappa germline inferences for wild-derived strains revealed sequence divergence differences among the four major *Mus* sub-species origins.

Furthermore, it showed that the heavy chain locus contained more inherent germline sequence variability than the kappa chain locus, which had previously been hypothesized but not explored across multiple strains representing different subspecies origins. Suppose the function of the kappa chain locus is to limit inherent germline diversity to prevent B-cell receptor auto-reactivity as hypothesized. In that case, it is crucial to properly characterize IGK haplotypes for mouse strains used in autoimmune research. In humans, we know that the light chain repertoire in autoimmune diseases like Myasthenia Gravis can become perturbed and result in errors during receptor editing during B-cell development[101]. Although AIRR-seq data can help build germline gene databases for various mouse strains used to model complex biomedical diseases, these data lack critical information on noncoding elements and gene positions that could elucidate gene expression Although AIRR-seq data can help build germline gene databases for various mouse strains, this data lacks critical information on noncoding elements and gene positions that could elucidate gene expression differences between strains. Additional IG genome assemblies are required that reflect the haplotype and sub-specific diversity that we have presented in the mouse IG loci.

CHAPTER 3 : DEVELOPMENT OF NEW IG GENOMIC ASSEMBLIES TO STUDY THE EFFECT OF GERMLINE VARIATION ON THE AB REPERTOIRE

INTRODUCTION

One of the overarching goals of this work is to understand how haplotype variation helps shape the Ab repertoire. In most biomedical research, model organisms, like mice, allow scientists to study the Ab repertoire in a controlled manner. Unfortunately, one of the most significant obstacles to accurately and reliably studying repertoire diversity in the context of genomics is a lack of high-quality reference assemblies and model organisms. Currently, the C57BL/6 mouse is the only complete IG haplotype resource that exists, which was originally published in 2006 by Corcoran et al[16]. A partial assembly for the IGH^A haplotypewas generated shortly after by Retter et al., but it is only 1.6 Mb and only contains about 1/3 of the IGH locus.

As sequencing technology has advanced, high-throughput sequencing methods are being used for AIRR-seq to interrogate the function of B-cell receptors in the immune response. These technologies, however, rely on reference gene databases to make accurate and reliable gene calls. Due to the complex structure of the IG loci, the current reference assemblies are incomplete and can be

misleading. In addition, the assumption is made that IG loci are conserved across all strains of mice, despite evidence of single nucleotide polymorphisms (SNPs), structural variants (SVs), and repeats within this region[4,69,87].To overcome this obstacle and properly understand the effect of germline haplotype variation in the mouse Ab repertoire model, alternate reference genomes are necessary that account for known haplotype diversity existing across biomedically relevant strains. Given that complete IG haplotypes only exist for C57BL/6, existing genomic resources do not contain any IG genetic variation. Due to this barrier, no system exists that allows the research of the genetic mechanisms of repertoire development. To overcome these obstacles, we have characterized two divergent IG haplotypes; one for IGH and one for IGK, by selecting, growing, and sequencing bacterial artificial chromosome (BAC) clones spanning the NOD/ShiLtJ IGK locus and BALB/cByJ IGH locus. We also annotated functional and open reading frame V, D (IGH only), and J genes from each assembly. We assembled the BALB/cByJ-IGH assembly to five non-overlapping contigs containing 192 functional and 135 non-functional IGHV genes, 30 IGHD genes, 4 IGHJ genes, and 8 IGHC genes, totaling 4.86 Mb. The NOD/ShiLtJ-IGK assembly was assembled into two non-overlapping contigs containing 82 functional and 31 non-functional IGKV genes.

MATERIALS AND METHODS

SELECTION OF IGH AND IGK BACTERIAL ARTIFICIAL CHROMOSOME (BAC) CLONES

To build new reference databases for mouse IG loci we selected, grew, and sequenced BAC clones that spanned the entire NOD/ShiLtJ IGK and BALB/cByJ IGH locus. The CHORI-29 NOD/ShiLtJ mouse *Mus musculus* BAC library was generated from DNA isolated from kidney cells of the NOD/ShiLtJ mouse. 206,971 NOD BACs were viewed in the Ensembl browser and used to select BAC tiling paths spanning any locus of interest. A total of 37 overlapping CHORI-29 NOD/ShiLtJ clones were selected. The goal in selecting overlapping clones was to ensure that, once sequenced, the clones could be assembled into large, continuous contigs that serve as a novel IG reference haplotype.

To resequence the IGH locus in BALB/cByJ, we used the CHORI-28 *Mus musculus* BALB/cByJ mouse BAC library. This library was previously constructed from DNA isolated from kidney cells of a male *Mus musculus* BALB/cByJ mouse, that represented an 11-fold genome coverage. However, unlike the NOD/ShiLtJ BAC library, no BAC-end sequencing had been conducted for this library, thereby requiring the use of traditional BAC library screening. Using IGHV gene sequences from BALB/cByJ that spanned the IGHV gene families[87], we designed unique probes to screen filters of the CHORI-28 library. The probes were end-labeled using 32p-ATP and T4 Polynucleotide Kinase. BAC filtering was conducted by performing filter hybridization, washing filters, and then visualizing the hybridized

filters on a phosphoimager. Hybridization-positive BAC DNA was amplified using mouse IGH primers designed from C57BL/6 or 129S6 genomic DNA to confirm that they contained the locus of interest. Given that the BAC clones were on average approximately 100 Kb, tiling paths for the BALB/c clones were constructed by designing IGH PCR primers from 129S6 and C57BL/6 that spanned every 100,000 bases of IGH genomic sequence. Primers were tested against the BAC clones, and the first set of ten positive non-overlapping BAC clones that spanned the entire IGH locus were sequenced on the PacBio RSII platform. To iteratively fill gaps in the assembly, new primers were designed off the flanks of newly sequenced BAC clones that did not overlap with an already sequenced clone. In total, 71 BALB/cByJ BACs were grown, prepared, and sequenced.

BAC CLONE DNA ISOLATION

Single BAC clones from glycerol stocks were inoculated into starter cultures of 5 mL LB with 12.5 μg/μL chloramphenicol. 0.5 mL of the starter culture was inoculated into 100 mL of LB + 12.5 μg/μL chloramphenicol and grown at 37ºC for 14 hours with shaking (~250 rpm). Following growth, the 100 mL of culture was split into two 50 mL conical tubes and centrifuged at 4500 x g for 20 minutes. Bacterial pellets were resuspended in 10 mL of Qiagen Buffer P1, with RNase A added to Buffer P1 at 100 μg/mL. Next, 10 mL of Qiagen Buffer P2 was added to each tube, mixed thoroughly by inverting the tube 4-6 times, then incubated at room temperature for 5 minutes. Following incubation, 10 mL of chilled Qiagen Buffer P3 was added to each tube and mixed immediately by inverting 4-6 times.

Samples were incubated on ice for 15 minutes, then centrifuged at >= 20,000 x g for 30 minutes at 4ºC. Supernatant containing the plasmid DNA was removed immediately. A QIAGEN-tip 100 was equilibrated by applying 4 mL of Qiagen Buffer QBT, and allowing the column to empty by gravity flow. Next, BAC supernatants were applied to the QIAGEN-tips and allowed to enter the resin by gravity flow. QIAGEN-tips were washed two times with 10 mL of Qiagen Buffer QC after the BAC supernatants were completely passed through the columns. DNA was eluted from the columns using five, 1 mL aliquots of prewarmed (65ºC) Qiagen Buffer QF. Next, DNA was precipitated by adding 3.5 mL of room-temperature isopropanol to eluted DNA. DNA and isopropanol were mixed by gentle swirling, then immediately centrifuged at >= 15,000 x g for 30 minutes at 4ºC. Supernatant was decanted and the DNA pellet was washed with 2 mL of room-temperature 70% ethanol. Samples were centrifuged at 15,000 x g for 10 minutes, then the supernatant was decanted without disturbing the DNA pellet. The DNA pellet was air-dried for approximately five minutes until no longer wet, then redissolved in 100 µL of 10 mM TrisCl, pH 8.5.

BAC CLONE DNA SEQUENCING AND ASSEMBLY

Due to the large sizes of the BAC inserts and the structural complexity of the IG loci, PacBio long-read sequencing was used to sequence the BAC clones. Following BAC DNA isolation, individual BAC clones were barcoded, pooled, then sequenced on a PacBio RS II, Sequel, or Sequel II, to generate high-quality, long-read genomic sequences. Isolated BAC gDNA was sheared to 15-20kb using

Covaris GTubes (SKU: 520079, Covaris, Woburn, MA, USA) upon entering library prep. The PacBio SMRTbell Express Template Prep Kit 2.0 (PN: 100-938-900) was used to generate SMRTbell libraries for each BAC clone. Following library prep, molecules without SMRTbell adapters were removed using the PacBio Enzyme Clean Up kit (PN: 101-746-400). The concentration of completed libraries was determined using the Thermofisher Qubit 3.0 Fluorometer dsDNA High Sensitivity Assay Kit (Thermofisher, Cat. No. Q32851), and sized using the Agilent High Sensitivity Fragment Analyzer 5200 (Agilent, Santa Clara, CA). A total of 71 BALB/cByJ BAC clones were sequenced that generated 12 Mb of total sequence. BAC sequences were assembled using the Canu[102] assembler by creating a contiguous assembly of sequence encompassing the IGHV, IGHD, and IGHJ genes from overlapping clones.

BALB/cByJ AND NOD/ShiLtJ LONG-READ GDNA SEQUENCING

To facilitate the closure of gap regions that existed between BALB/cByJ and NOD/ShiLtJ BAC clones, we used Oxford Nanopore Technology's (ONT) long read sequencing and PacBio Continuous Long Read (CLR) sequencing to generate long sequencing reads from gDNA that could serve as linkers between non-overlapping clones. gDNA from an 8-week-old female BALB/cByJ and NOD/ShiLtJ mouse was isolated from a freshly harvested kidney using the Qiagen MagAttract HMW gDNA Kit (Cat. No. 67563). gDNA was assessed using the Thermofisher Qubit 3.0 Fluorometer dsDNA Broad Range Assay Kit (Thermofisher, Cat. No.

Q32850) and Agilent High Sensitivity Fragment Analyzer 5200 (Agilent, Santa Clara, CA).

The Genomic DNA by Ligation ONT protocol (Cat. No. SQK-LSK110) with adaptive sampling was performed on 4 µg BALB/cByJ and NOD/ShiLtJ input DNA and sequenced on a MinION MK1C. FAST5 ONT sequencing output was converted to high-accuracy FASTQ reads using ONT's Guppy basecaller software (v5.0.11) set to high-accuracy (https://community.nanoporetech.com). For adaptive sampling, a .bed file with the genomic coordinates of the IG loci was provided to the sequencing software with a FASTA file containing the genomic sequence of the target regions. Since the BALB/c IGH and NOD/ShiLtJ IGK loci represented divergent alternate haplotypes that are not currently characterized, we used sequence from contiguous, assembled BAC clones as a reference sequence for the adaptive sampling. During adaptive sampling, the software examined the first few hundred bases of a DNA molecule as it passed through a pore and determined if the molecule matched sequence provided in the reference FASTA file. If there was no match, the molecule was rejected from the pore by reversing the voltage. However, molecules that matched the reference sequence were retained, sequenced, and basecalled.

PacBio CLR libraries for BALB/cByJ and NOD/ShiLtJ were prepared from 5 µg of mouse kidney gDNA, then sheared to 30-40kb with Covaris g-TUBE (SKU: 520079, Covaris, Woburn, MA, USA). Sheared DNA was concentrated with AMPure PB Beads, then assessed with Thermofisher Qubit 3.0 Fluorometer dsDNA High Sensitivity Assay Kit (Thermofisher, Cat. No. Q32851) and the Agilent

High Sensitivity Fragment Analyzer 5200 (Agilent, Santa Clara, CA). The PacBio SMRTbell Express Template Prep Kit 2.0 (PN: 100-938-900) was used to generate a pooled SMRTbell library for both BALB/cByJ and NOD/ShiLtJ gDNA, followed by nuclease treatment (V1) (PN: 101-746-400) to remove unligated molecules.

BAC SEQUENCE ANALYSIS AND GENE ANNOTATION

Assembled BAC contigs were checked for quality on a BAC-by-BAC basis by aligning subread sequences to the assembled BAC contig using unimap[103]. Alignments were manually inspected using the Integrated Genomics Viewer[104,105] (IGV) for evidence of mis-assembly or vector contamination. In addition, SNV candidates were identified for each BAC clone using whatshap find_snv_candidates, with the find_snv_candidates output vcf file serving as input for whatshap genotype. Nucleotides were called as a SNV if there was a significant basecall discrepancy between the subreads and assembled contig. We estimated the accuracy of the assembled contigs by dividing the total number of SNVs by the size of the contig to calculate the percent accuracy of our assembly contigs.

Genes at the loci were annotated by aligning the sequenced BAC c[7]ones to the IMGT gene database[7] and BALB/cJ germline IGHV sequences[45,87]. We classified genes identified in the assembly as either functional or non-functional according to the presence of internal stop codons in gene exons and a productive recombination signal sequence (RSS). Sequences were considered functional if they were encoded in an open reading frame (ORF) without internal stop codons

and flanked by a productive RSS, whereas sequences with internal stop codons or a poor RSS were classified as non-functional. L-Part1, L-Part2, V-exon, heptamer, spacer, and nonamer were also annotated for each V-Gene. To be considered a functional V gene, the L-Part1, L-Part2, and V-exon were required to be encoded in an ORF with no internal stop codons. Nonamers, heptamers, and spacers were also annotated for D (IGH only) and J genes. RSS sequences in the assemblies were identified using the RSS prediction tool RSSsite[106]. The tool predicts RSS site functionality by calculating the recombination information content (RIC) for physiological RSSs[107]. Genes in our assemblies were required to have a "Pass" for the RIC score calculated for the RSS.

Dot plot comparisons to existing IG genome assemblies were generated with D-GENIES[108]. Phylogenetic comparisons were created with Geneious Prime 2022.0.2 (https://www.geneious.com), and annotated with strain, gene family, and the assembled BALB/cByJ or NOD/ShiLtJ contig they were derived using Interactive Tree of Life V5 software[109]. Briefly, multiple sequence alignments were performed on functional IGHV and IGKV sequences from the BALB/cByJ and NOD/ShiLtJ assemblies, respectively. We restricted phylogenetic comparisons to functional and ORF sequences, as pseudogenes cannot be confirmed with AIRR-seq data, and constructed neighbor joining phylogenetic trees from the multiple sequence alignment with the following parameters: global alignment with cost matrix identity (1.0/0.0) and the Jukes-Cantor genetic distance model. Sequence comparisons to pre-existing gene databases were performed using BLAT[78].

RESULTS

SEQUENCE ASSEMBLY OF THE BALB/cByJ IGH LOCUS AND NOD/ShiLtJ IGK LOCUS

Prior to 2006, the mouse IGH locus had been partially assembled by the Ensembl project[110], however, the assembly was incomplete, with both large and small gaps accounting for about 450 kb of genomic sequence. Seeking to close all gaps and assemble the entire locus, Johnston et al. sequenced and assembled BAC clones spanning the length o[16]the V region of the IGH locus[16]. Their data provided 2.75 Mb of the mouse IGHV region and generated an annotated map of IGHV genes, pseudogenes, repeats, and nonrepetitive intergenic sequences. However, this data did not represent the haplotype variation known to exist within the mouse IGH locus since it only accounted for the IGH^B haplotype. Later, Retter and colleagues sequenced and assembled a portion of the IGH^A haplotype for the 129S1 strain, but their assembly was incomplete since it only included the proximal region of the locus containing IGHC, IGHD, IGHJ, and the most proximal IGHV genes[69]. Similarly, the IGK locus has only been sequenced in C57BL/6, which only represents a single IGK haplotype despite AIRR-seq data indicating the presence of at least three distinct IGK haplotypes among biomedically relevant mouse strains. We have assembled the BALB/cByJ IGH locus manually using an iterative method of PCR screening, cloning, and sequencing, but used available BAC end sequences to identify candidate NOD/ShiLtJ clones spanning the IGK locus.

In C57BL/6, the IGH locus spans 3.3 Mb on chromosome 12 (chr12:113245000-116045000, GRCh38[46,111]). We assembled 12 Mb of BALB/cByJ BAC clones (n=73) that spanned the IGH locus into 4.86 Mb of non-redundant sequence, which captured the complete IGHC, IGHJ, and IGHD regions, as well as 3.8 Mb of the IGHV region. The 4.86 Mb assembly is distributed among five non-overlapping contigs with varying sizes. The positions of the five assembled contigs in relation to the C57BL/6 IGH locus is shown in Figure 3.1. The largest assembled contig is approximately 2.78 Mb (n=43 BAC clones) and includes the IGHC, IGHD, IGHJ, and most of the proximal IGHV region, whereas the smallest contig represents a single BAC clone of 148 kb and is in the distal region of the IGH locus. Contig 5 is the most distal contig that contains distal IGHV genes as well as the gene *Vipr2*[112], which is encoded just outside the mouse IGH locus. The assembled contigs and their contents are summarized in Table 3.1. A BALB/cByJ BAC clone tiling path for each contig is shown in Figures 3.2-3.6.

The IGK locus in C57BL/6 spans 3.2 Mb, with 101 functional V segments, 4 functional J segments, and 1 C segment, encoding approximately 95% of the C57BL/6 Ab light chain molecules[8,9]. While the IGK locus has been sequenced and assembled in C57BL/6[27,113–115], the assembly only represents a single IGK haplotype. This is a problem since AIRR-seq data and RFLP experiments have demonstrated haplotype differences across mouse strains[4,28,45]. 32 BAC clones spanning the NOD/ShiLtJ IGK locus have been sequenced and assembled into two non-overlapping contigs of 1.87 Mb and 763 kb. Due to clone and end-sequence availability, as well as the inability to locate additional BAC clones in the

constant region, we acknowledge that our BACs do not span the complete IGK locus. The assembled NOD/ShiLtJ contigs and their contents are summarized in Table 3.2. A NOD/ShiLtJ BAC clone tiling path for each contig is shown in Figures 3.7 and 3.8.

Table 3.1. Summary of BALB/cByJ Contigs Assembled from BAC Clones

Contig Name	Number of BAC Clones	Size	Location	Number of Functional Genes	Number of Non-Functional Genes
Contig 1	42	2.78 Mb	Proximal	232 IGHV, 31 IGHD, 4 IGHJ, 8 IGHC	112 IGHV
Contig 2	1	148 Kb	Distal	7 IGHV	2 IGHV
Contig 3	4	590 Kb	Distal	30 IGHV	12 IGHV
Contig 4	8	401 Kb	Distal	18 IGHV	3 IGHV
Contig 5	16	938 Kb	Distal	37 IGHV	6 IGHV

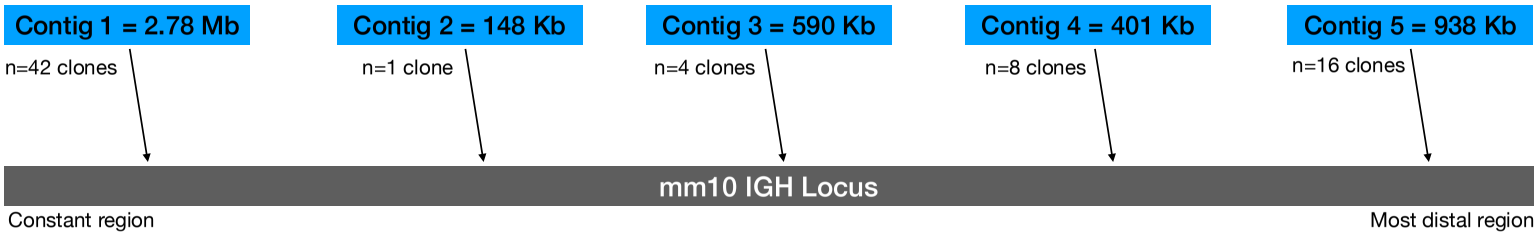
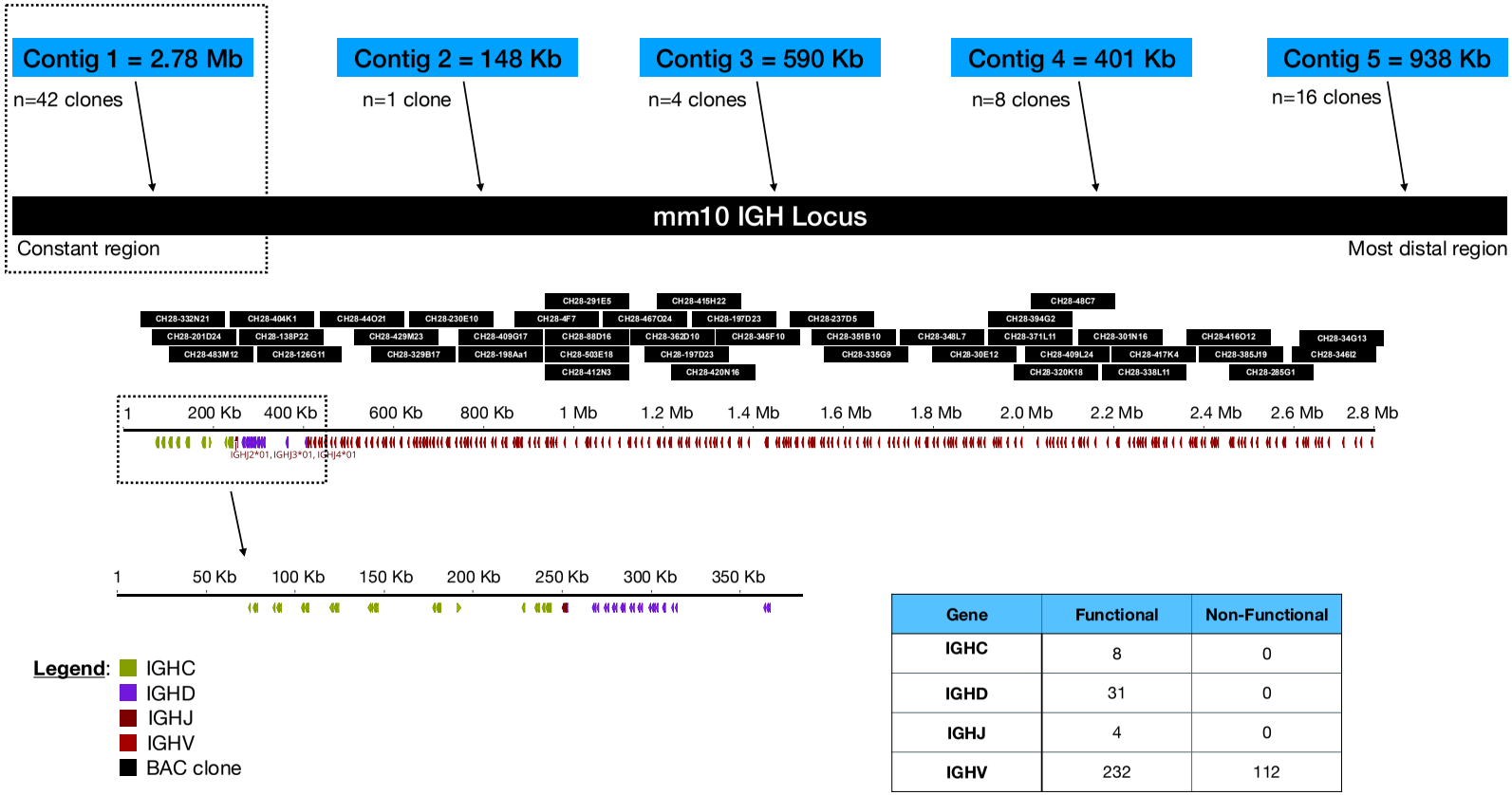

Contig 1 = 2.78 Mb
Contig 2 = 148 Kb
Contig 3 = 590 Kb
Contig 4 = 401 Kb
Contig 5 = 938 Kb
n=42 clones
n=1 clone
n=4 clones
n=8 clones
n=16 clones
mm10 IGH Locus
Constant region
Most distal region

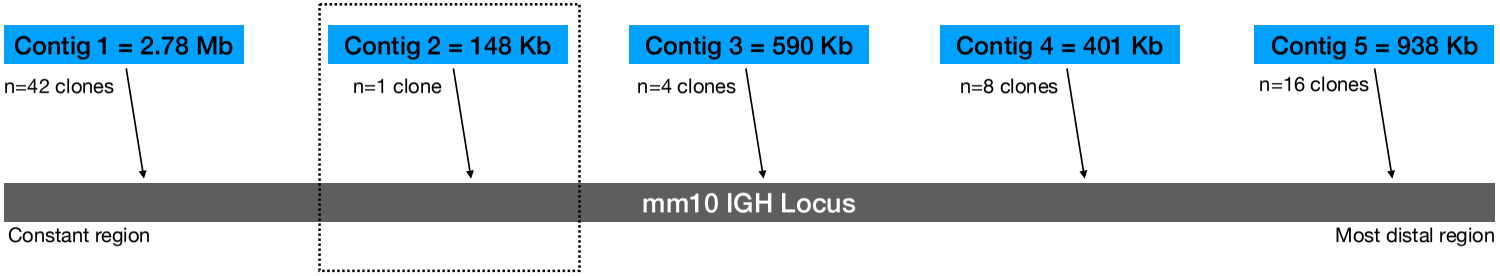
Figure 3.1. Assembled BALB/cByJ contigs in relation C57BL/6 IGH locus (mm10).

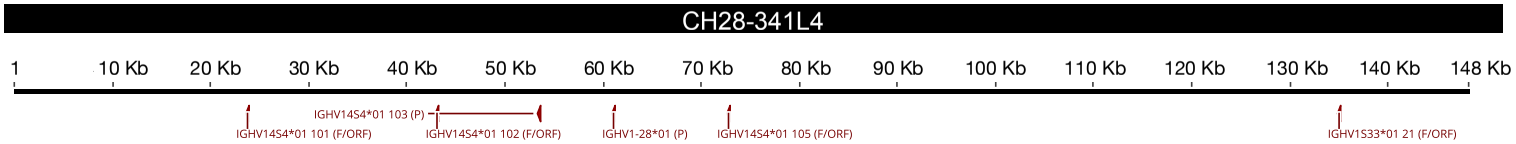


Gene	Functional	Non-Functional
IGHC	8	0
IGHD	31	0
IGHJ	4	0
IGHV	232	112

Figure 3.2. BALB/cByJ BAC clone tiling path and gene distribution for Contig 1 (2.8 Mb).

Individual BAC clones are colored black and named. Below the tiling path is a gene map, showing an approximation of gene content per BAC clone. For clarity, RSS sequences, introns, and leaders are not marked.



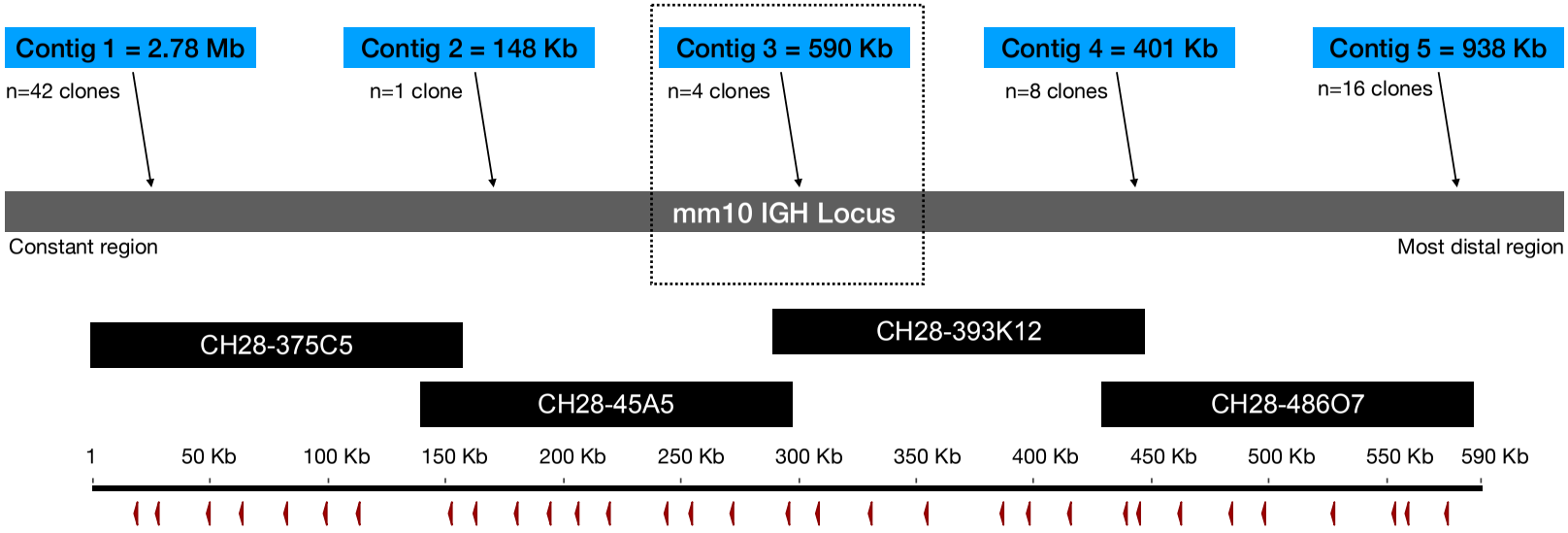


Legend:
- IGHC
- IGHD
- IGHJ
- IGHV
- BAC clone

Gene	Functional	Non-Functional
IGHC	0	0
IGHD	0	0
IGHJ	0	0
IGHV	7	2

Figure 3.3. BALB/cByJ BAC clone tiling path and gene distribution for Contig 2 (148 kb).

Each position on the map represents a single gene. For clarity, RSS sequences, introns, and leaders are not shown.



Legend:
- IGHC
- IGHD
- IGHJ
- IGHV
- BAC clone

Gene	Functional	Non-Functional
IGHC	0	0
IGHD	0	0
IGHJ	0	0
IGHV	30	12

Figure 3.4. BALB/cByJ BAC clone tiling path and gene distribution for Contig 3 (590 Kb).

Individual BAC clones are colored black and named. Each position on the map represents a single gene. For clarity, RSS sequences, introns, and leaders are not shown.

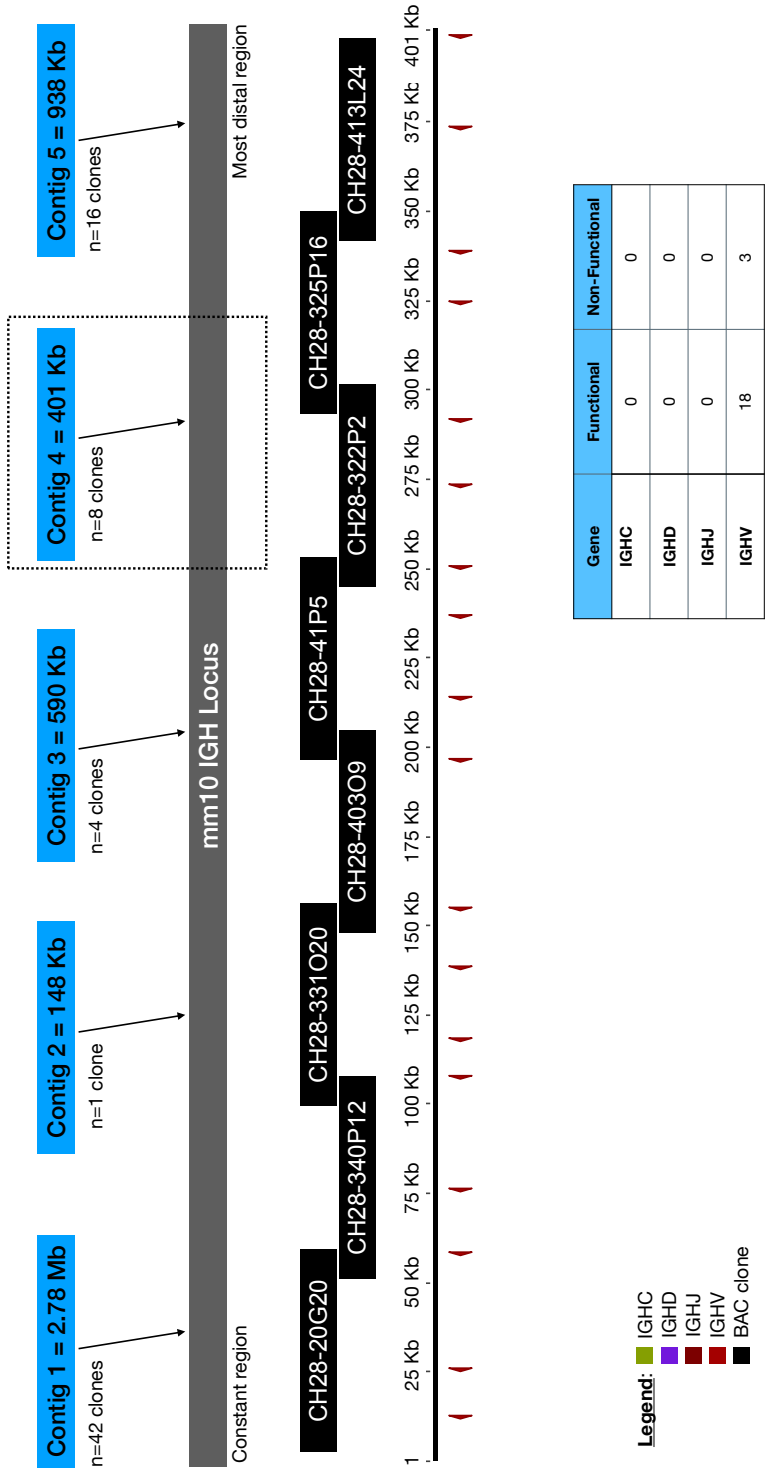
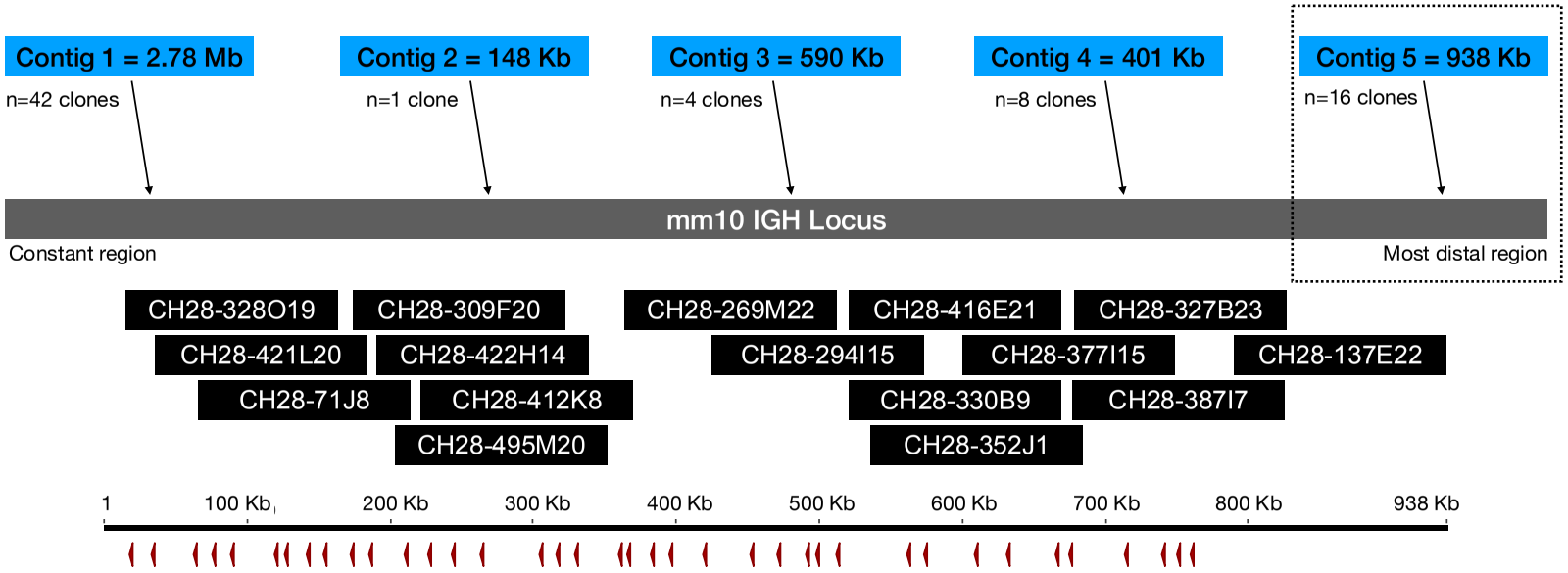


Figure 3.5. BALB/cByJ BAC clone tiling path and gene distribution for Contig 4 (401 Kb).

Each position on the map represents a single gene. For clarity, RSS sequences, introns, and leaders are not shown.



Legend:
- IGHC
- IGHD
- IGHJ
- IGHV
- BAC clone

Gene	Functional	Non-Functional
IGHC	0	0
IGHD	0	0
IGHJ	0	0
IGHV	37	6

Figure 3.6. BALB/cByJ BAC clone tiling path and gene distribution for Contig 5 (938 Kb).

Individual BAC clones are colored black and named. Each position on the map represents a single gene. For clarity, RSS sequences, introns, and leaders are not shown.

Table 3.2. Summary of NOD/ShiLtJ Contigs Assembled from BAC Clones

Contig Name	Number of BAC Clones	Size	Location	Number of Genes
NOD IGK A	22	1.87 Mb	Distal	56 IGKV
NOD IGK B	8	763 Kb	Proximal	31 IGKV

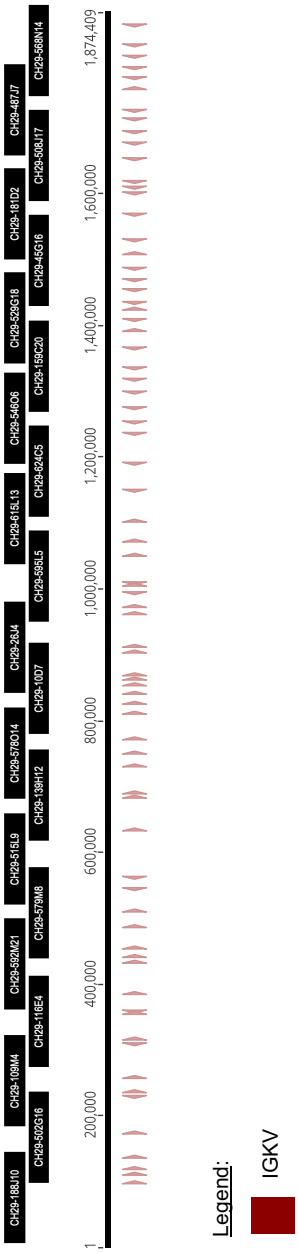


Figure 3.7. NOD/ShiLtJ BAC clone tiling path and gene distribution for NOD IGK A contig (1.87 Mb).

Individual BAC clones are colored black and named. Each position on gene map represents a single gene. For clarity, RSS sequences, introns, and leaders are not shown.

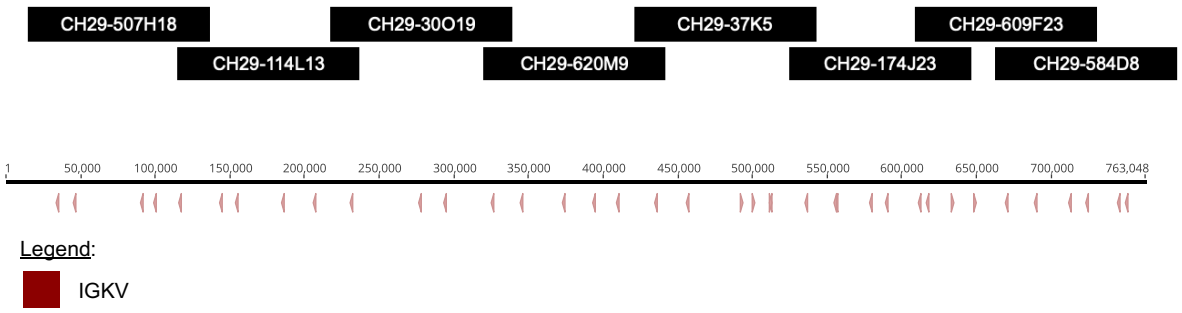


Figure 3.8. NOD/ShiLtJ BAC clone tiling path and gene distribution for NOD IGK B (763 Kb).

Individual BAC clones are colored black and named. Each position on gene map represents a single gene. For clarity, RSS sequences, introns, and leaders are not shown.

ANNOTATION OF BALB/cByJ IGHV, IGHD, IGHJ, AND IGHC GENES IN THE BALB/cByJ IGH ASSEMBLY

327 IGHV genes were annotated from the current BALB/c assembly. Of these, 192 (52%) are predicted to form functional antibodies, either encoded in open reading frames (ORFs) or verified by AIRR-seq data. According to the AIRR community, ORF genes are germline genes that have an open reading frame in their coding region, but may have altered splicing sites, different recombination signals or regulatory elements, or changes in conserved amino acids that could lead to incorrect folding [7,116]. The remaining 135 IGHV genes are not coded in an ORF and were classified as pseudogenes. Included with the pseudogenes were sequences with 100% identity to sequences described by Retter and colleagues as relics[69]. Retter et al. identified these sequences as pseudogenes but could not assign them to a gene family despite their proximity to IGHV5 sequences. The 327 IGHV sequences were found distributed among the fifteen known gene families, with the IGHV1 family containing the most annotated functional and pseudogene sequences, followed by IGHV5, IGHV2, and IGHV8 (Figure 3.9). Gene names were assigned to annotated assembly sequences by taking the gene family of the highest similarity IgBLAST alignment to known genes in the IMGT gene database, and then appending a number according to the gene's position in the assembly. The sequences for annotated IGHV genes, L-Part1, L-Part2, and RSS are provided in Supplemental Table 2.

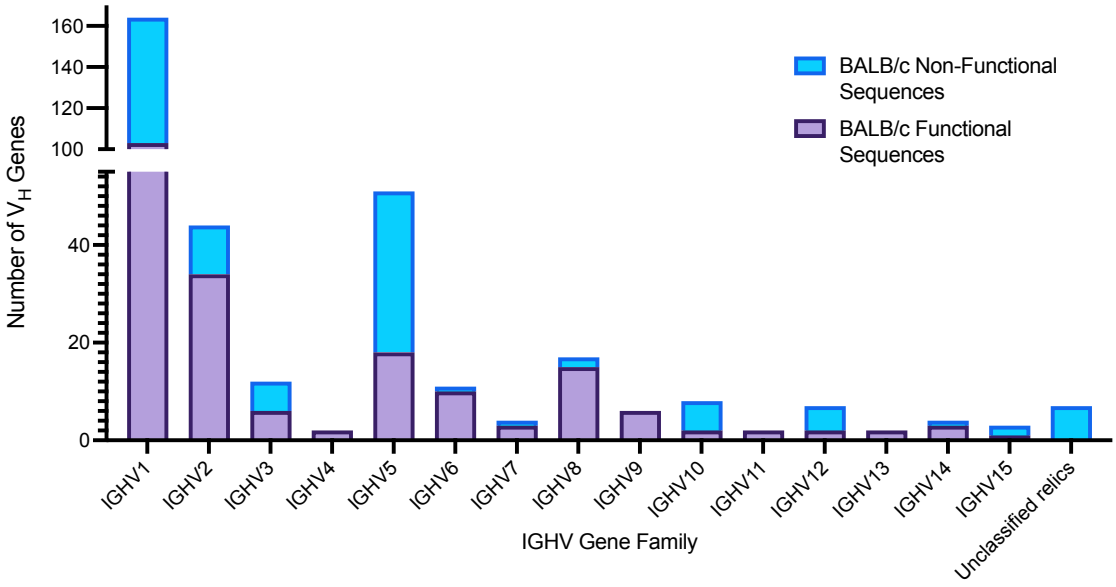


Figure 3.9. Distribution of functional and non-functional IGHV sequences among IGHV gene families.

30 D_H genes and their heptamer, spacer, and nonamer sequences were annotated from the current BALB/cByJ assembly and named according to their most similar IMGT match (Supplemental Table 3). These sequences were identified within a 1.6 kb region in the proximal region of a 2.8 Mb assembled contig (Figure 3.10). Previous reports using AIRR-seq had identified 10 unique IGHD sequences in BALB/c, and these 10 sequences corresponded to 12 IGHD genes since IGHD2-2*01 and IGHD2-7*01, and IGHD2-1*01 and IGHD2-8*01 are identical and have been previously reported[87,117]. All 12 inferred IGHD sequences from Collins et al. are present in our assembly with 100% identity. Further inspection of our sequences reveals that some IGHD genes contain copy number variation within the assembly. There are 7 copies of IGHD5, 3 copies of IGHD6, 3 copies of IGHD5-1*01, and 2 copies of IGHD2-4*01. Two IGHD genes listed in IMGT, IGHD1-3*01 and IGHD3-1*01, were not inferred from Collins' et al. AIRR-seq data but are present in our assembly with 100% identity and functional RSS sites[87,106]. In addition, Collins et al. did not infer IGHD2-9*01, IGHD2-11*01, and IGHD4-1*01. While we did not annotate IGHD2-11*01 or IGHD2-9*01 in our assembly, we did locate IGHD4-1*01. Of note, IGHD1-3*01 was previously identified in the most distal end of the D region of the IGH^A haplotype, but named DFL16.3[69].

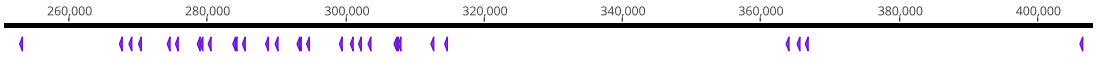


Figure 3.10. Portion of BALB/cByJ Contig 1 that spans the immunoglobulin heavy chain diversity (IGHD) gene locus.

Shown are the IGHD genes (purple) spanning the D-gene region, and numbering corresponds to position within Contig 1.

The four mouse IGHJ genes were identified 3' of the IGHD region but before the IGHC region (Figure 3.11). The four J_H genes are located within a short 1.3 kb adjacent to the most proximal IGHD gene, IGHD4-1*01. These sequences are 100% identical to the BALB/c IMGT IGHJ sequences[7]. In addition to the four IGHJ genes, we also assembled sequence containing all IGHC exons spanning 169kb in the BALB/cByJ Proximal and Distal 1 contig, which code for eight different isotypes of the Ab heavy chain constant region. This includes IGHM, IGHD, IGHG3, IGHG1, IGHG2B, IGHG2A, IGHE, and IGHA (Supplemental Table 5).

ANNOTATION OF NOD/ShiLtJ IGKV GENES IN THE NOD/ShiLtJ IGK ASSEMBLY

113 IGKV sequences were annotated in the NOD/ShiLtJ IGK assembly, with 82 functional sequences present in an open reading frame. While this is fewer than the 101 functional IGKV genes reported for C57BL/6 mice[8], we acknowledge that our assembly is incomplete at its current state due to the inability to find additional BAC clones for the most proximal part of the IGK locus. 66% (75/113) of all annotated sequences were located in the more distal 1.8 Mb NOD/ShiLtJ Contig 1, with the remaining 38 IGKV sequences in the proximal NOD/ShiLtJ Contig 2. We predict that 73% (82/113) of all annotated sequences to be functional since they are in an open reading frame with no internal stop codons. We annotated the remaining 31 sequences as pseudogenes since they contained

internal stop codons or had an RSS site predicted to be non-functional according to RSSsite[106].

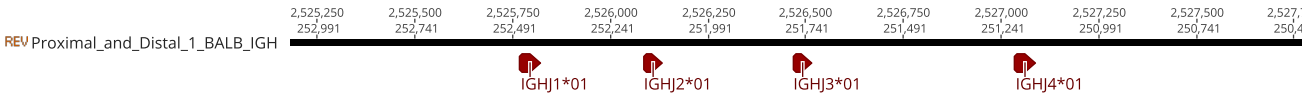


Figure 3.11. Portion of BALB/cByJ Contig 1 that spans the immunoglobulin heavy chain diversity (IGHJ) gene locus.

Shown are the IGHJ genes (red) spanning the J-gene region.

The non-functional sequences were excluded from downstream analyses. The annotated NOD/ShiLtJ IGKV sequences and RSS sites are provided in Supplementary Table 6.7.

The number of IGKV sequences per family varied. IGKV4 had the largest representation with 16 functional sequences and 4 non-functional sequences, followed by IGKV1, IGKV8, and IGKV12 (Figure 3.12). Only one functional IGKV sequence was identified for IGKV7, IGKV11, IGKV13, IGKV15, IGKV16, IGKV18, IGKV19, and IGKV20, with no functional sequences annotated for IGKV3. Despite the underlying IGK haplotype differences between C57BL/6 and NOD/ShiLtJ, we used the C57BL/6 IGK reference as a proxy to estimate the completeness of our assembly. When comparing our NOD/ShiLtJ IGK assembly to the C57BL/6 IGK locus, no NOD IGKV sequences were inferred in the C57BL/6 genomic region spanning IGKV8-19 to the constant region, which is approximately 389 Kb. However, this estimate is based on the C57BL/6 reference, and we know from the Chapter 2 AIRR-seq data that the NOD IGKV repertoire is very different from the C57BL/6 IGKV repertoire.

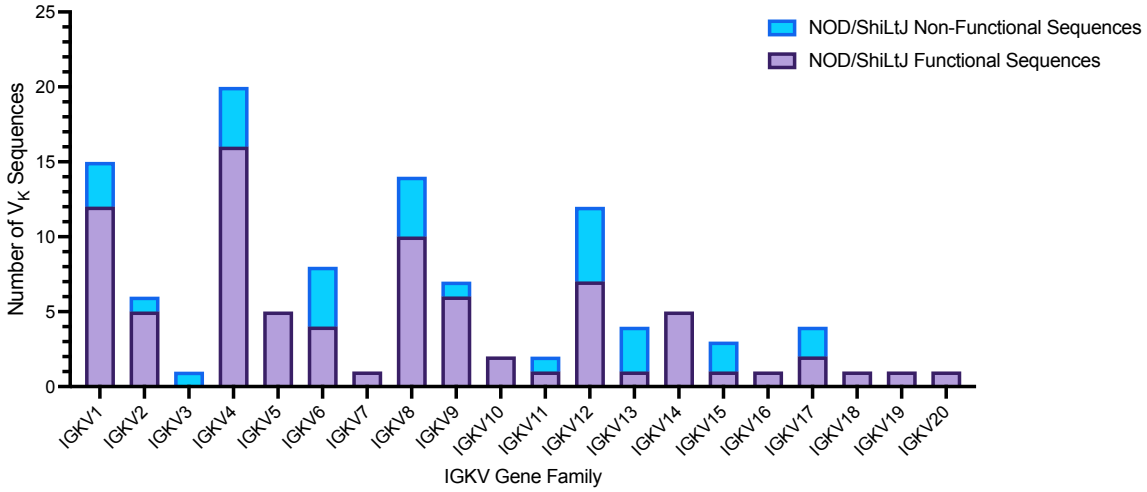


Figure 3.12. Distribution of functional and non-functional IGKV sequences among IGKV gene families.

COMPARISON OF THE IGHV GENES OF BALB/cByJ TO C57BL/6 AND EXISTING DATASETS

The polymorphisms in the mouse IG loci have been described as the result of strong evolutionary pressures on the loci, with IGHV genes subject to divergent evolution from diversifying selection and evolution[118]. To assess how well pre-existing datasets represent the sequences we identified in our assemblies, we compared our sequences to pre-existing gene databases using BLAT[78]. Our first comparison was between our inferred functional BALB/cByJ IGHV genes and functional C57BL/6 IGHV genes inferred by Jackson et al[45], which are now deposited on the AIRR community's website OGRDB. We chose this dataset as it is the most current and complete collection of inferred BALB/c and C57BL/6 germline IGHV sequences. We hypothesized that we would see a strong divergence between the two IGHV repertoires, like what was previously reported for BALB/c and C57BL/6 from AIRR-seq data. When Collins et al. performed a similar comparison in 2015 using inferred BALB/c and C57BL/6 IGHV sequences, they reported that only five IGHV sequences were common to both strains: IGHV5-6*01, IGHV1-69*01, IGHV2-3*01, IGHV2-5*01, and IGHV5-2*01[87]. As expected, our data displayed a similar relationship between the BALB/cByJ and C57BL/6 germline IGHV repertoires, with only 5 IGHV sequences shared between the two strains with 100% identity. Our 5 BALB/c IGHV sequences that were shared with C57BL/6 were identical to Collins et al. except for IGHV1-69*01. BALB/c IGHV1-62-2*01 was shared with C57BL/6 with 100% identity, whereas our BALB/c IGHV

sequence IGHV1.47 was 99.32% identical to C57BL/6 IGHV1-69*01. Given that family 1 IGHV genes are abundant in the distal region of the IGH locus, and that there are still assembly gaps remaining in this part of the locus, it is likely that the BALB/c equivalent of IGHV1-69*01 has not been sequenced. The remaining 187 BALB/c IGHV sequences had a wide range of percent identities when compared to C57BL/6 germline IGHV genes (Figure 3.13). 143 BALB/cByJ IGHV sequences were less than 98% identical to their nearest C57BL/6 sequence, with the remaining 44 sequences between 98% and 99.9% identical.

Next, to validate our manually annotated sequences, we compared our sequences to the inferred BALB/c germline IGHV sequences from Jackson et al.[45] (Figure 3.13). 77% (148/192) of our functional BALB/cByJ IGHV sequences matched the database sequences with 100% identity. Of the remaining 44 IGHV sequences, five IGHV sequences had percent identities between 98% and 99.9%, while 39 BALB/cByJ IGHV sequences had percent identities < 98%. Of the database sequences that our assembly does not yet capture, the IGHV1 family was the largest, with 30 sequences absent from the assembly. This was likely due to the location of the IGHV1 family in the distal part of the IGH locus, which still contains assembly gaps. The gene family distribution of the remaining OGRDB BALB/c germline IGHV sequences missing from our

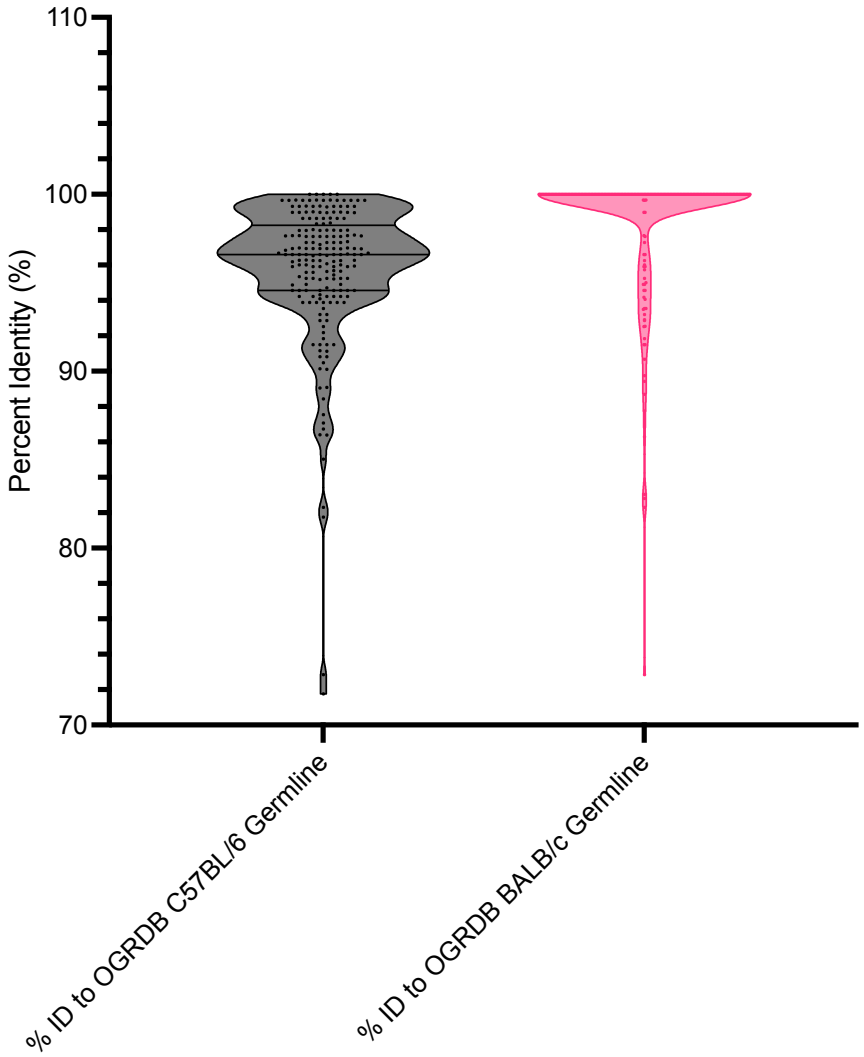


Figure 3.13. Comparison of functional BALB/cByJ IGH assembly sequences to OGRDB C57BL/6 and BALB/c germline sequences.

Table 3.3. Summary of Germline BALB/cByJ IGHV Sequences Absent From BALB/cByJ-IGH Assembly

Number of IGHV Sequences	IGHV Gene Family
30	IGHV1
6	IGHV6
3	IGHV8
2	IGHV2
1	IGHV3
1	IGHV7
1	IGHV13

assembly is summarized in Table 3.3. Lastly, since we believe our assembly to be complete except for the proximal half of the BALB/cByJ IGH locus, we also compared our sequences against the IGHV sequences published from the 129S6 mouse, which contains the same IGH haplotype as BALB/cByJ. As expected, 100% of the 129S6 sequences are also present in the BALB/cByJ assembly, further highlighting the IGH haplotype similarity between these two strains.

Finally, we compared each assembled BALB/cByJ-IGH contig to the C57BL/6 IGH locus by dot plot (Figure 3.14 and 3.15). This plot shows the overall sequence concordance between the two loci. If the two loci are genetically similar to one another, then a diagonal line will be visible in the dot plot output. Frameshifts, direct repeats, and inverted repeats can all alter the topology of the dotplot. From the dotplots we observed that the proximal end of Contig 1 has a high similarity match percentage represented by the diagonal green line spanning approximately 560 Kb into Contig 1 (Figure 3.14). However, the distal region of Contig 1 contains regions with much lower match percentages, with evidence of numerous frameshift mutations. Given that the proximal region of the locus contains the more conserved IGHC and IGHJ genes, we expected this region to have a higher match percentage compared to the distal region.

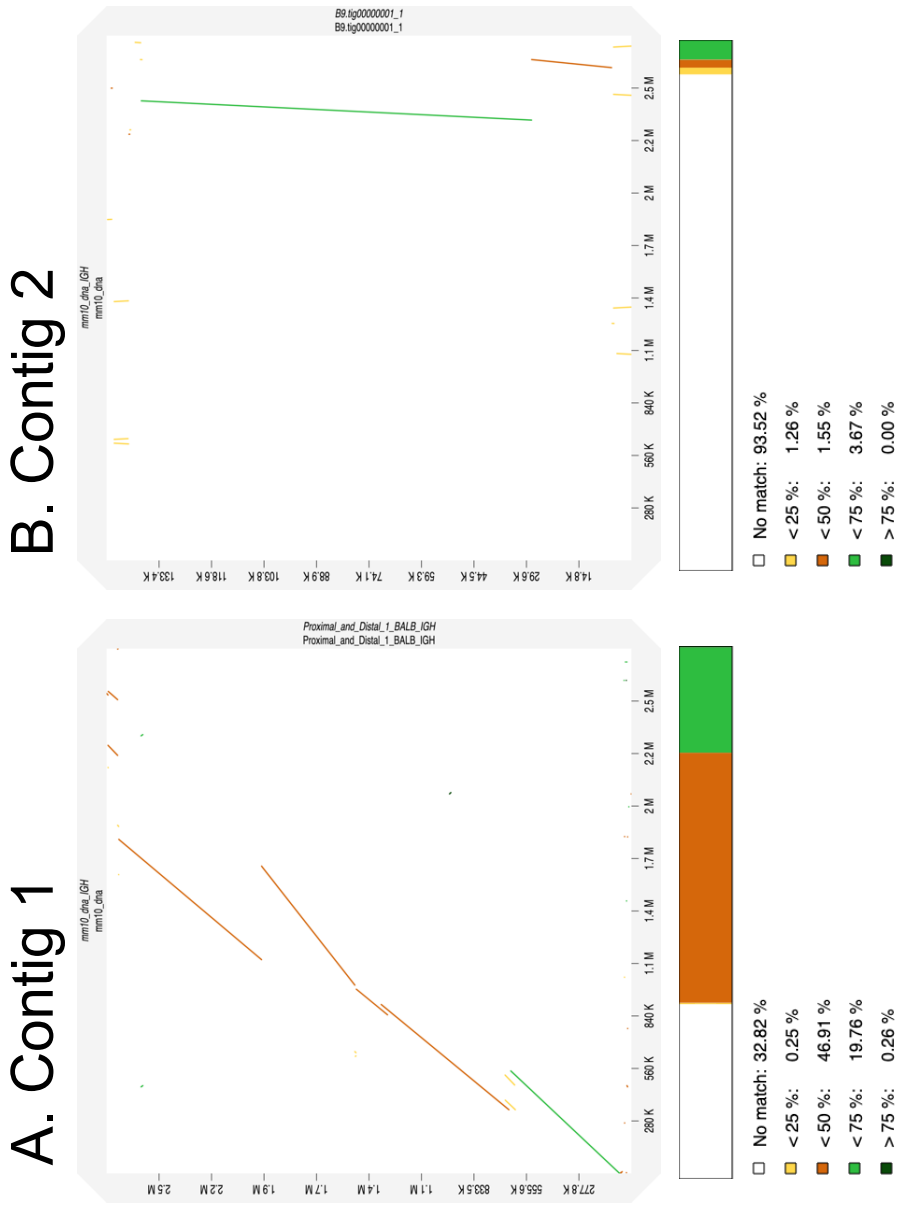


Figure 3.14. Dot plot comparison of BALB/cByJ assembled contigs to C57BL/6 IGH locus.

Diagonal line represents alignment match, with the color of the line indicating the quality of the match.

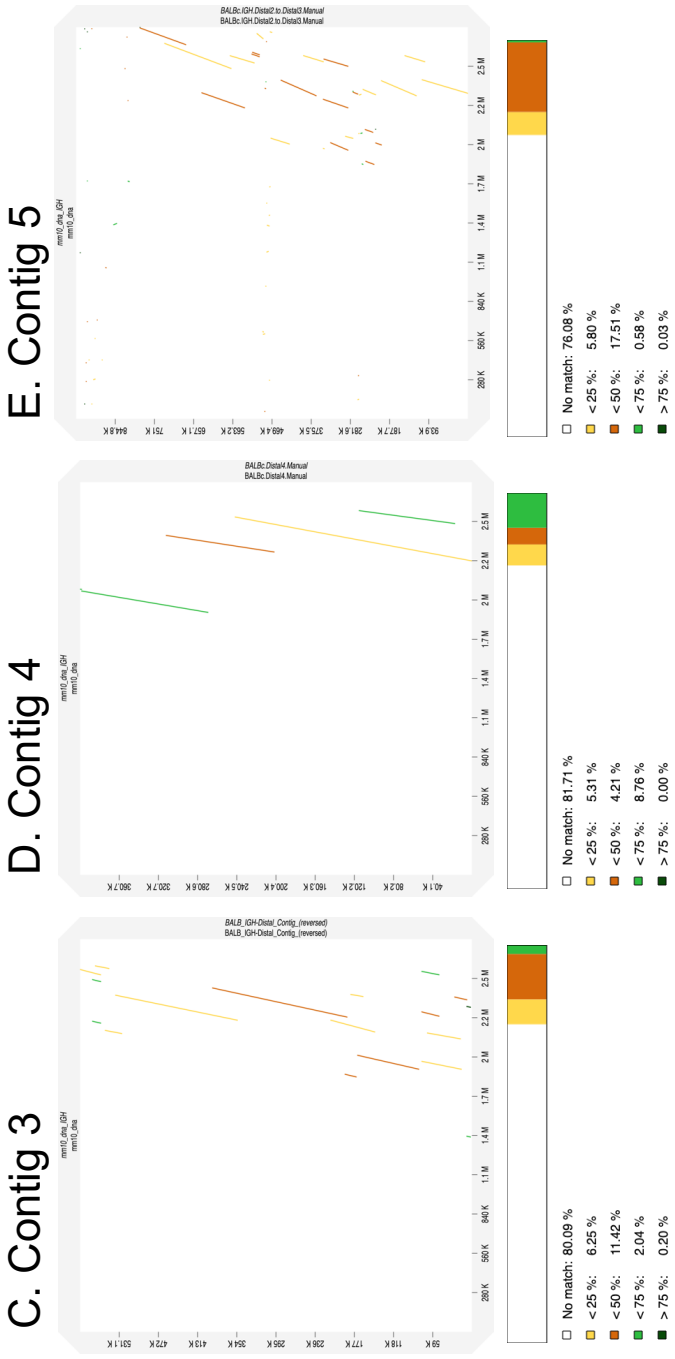


Figure 3.15. Dot plot comparisons of BALB/cByJ assembled contigs to C57BL/6 IGH locus.

Diagonal line represents alignment match, with the color of the line indicating the quality of the match.

DIVERSIFICATION OF BALB/cByJ and C57BL/6 IGHV SEQUENCES

To better understand the evolutionary relationships of functional IGHV genes in C57BL/6 and BALB/cByJ, or simply the evolutionary relationship of the IGH locus in the IGH^{A} and IGH^{B} haplotypes, we constructed a neighbor joining phylogenetic tree from a global alignment. The most expanded gene family for BALB/cByJ strains was IGHV1. However, as the IGHV1 gene family is in the most distal part of the locus which has remaining assembly gaps, we looked for additional gene families contained in a single contig that appeared to have been expanded or contracted in BALB/cByJ. The IGHV2 family was restricted to a single assembled contig in the BALB/cByJ-IGH assembly (Contig 1) and consisted of 17 functional IGHV2 genes; C57BL/6 contained 10 functional IGHV2 genes. This gene expansion was observed within a 680 Kb region in our BALB/cByJ IGH assembly and was consistent with others that had reported a similar expansion in 129S6[69]. A dot plot of this expansion in relation to the C57BL/6 IGH reference is shown in Supplemental Figure 2.

COMPARISON OF THE IGKV GENES OF NOD/ShiLtJ TO C57BL/6 AND EXISTING DATASETS

The NOD/ShiLtJ IGKV germline repertoire had not been comprehensively explored until our AIRR-seq experiment in Chapter 2. In fact, there are no NOD/ShiLtJ IGKV sequences present in the IMGT gene database. The first comparison we made was between our functional NOD/ShiLtJ IGKV assembly sequences and functional IMGT C57BL/6 IGKV sequences. Our AIRR-seq data in Chapter 2 informed us that the inferred germline IGKV repertoires between C57BL/6 and NOD/ShiLtJ were divergent, and that NOD/ShiLtJ and C57BL/6 represented two distinct IGK haplotypes. Therefore, we hypothesized that we would see a similar divergence when comparing our NOD/ShiLtJ assembly sequences to C57BL/6 IMGT IGKV sequences. Currently, there are 109 C57BL/6 IGKV sequences in the IMGT gene database. Of these sequences, 63 are functional or ORF, and the remaining 46 are pseudogenes. We performed a BLAT of the 82 functional assembly IGKV sequences to the 63 IMGT C57BL/6 functional and ORF IGKV sequences and observed that only 4 of the 82 functional NOD/ShiLtJ IGKV assembly sequences had a 100% match to an IMGT sequence (Figure 3.16). 59 of the non-IMGT IGKV sequences had <98% identity to their nearest IMGT IGKV sequence, whereas 18 sequences had percent identities between 99.68% and 98.01% (Figure 3.16). Of the 78 IGKV sequences with <100% identity to the nearest IMGT IGKV sequence, only 4 had a 100% match to a sequence deposited within NCBI.

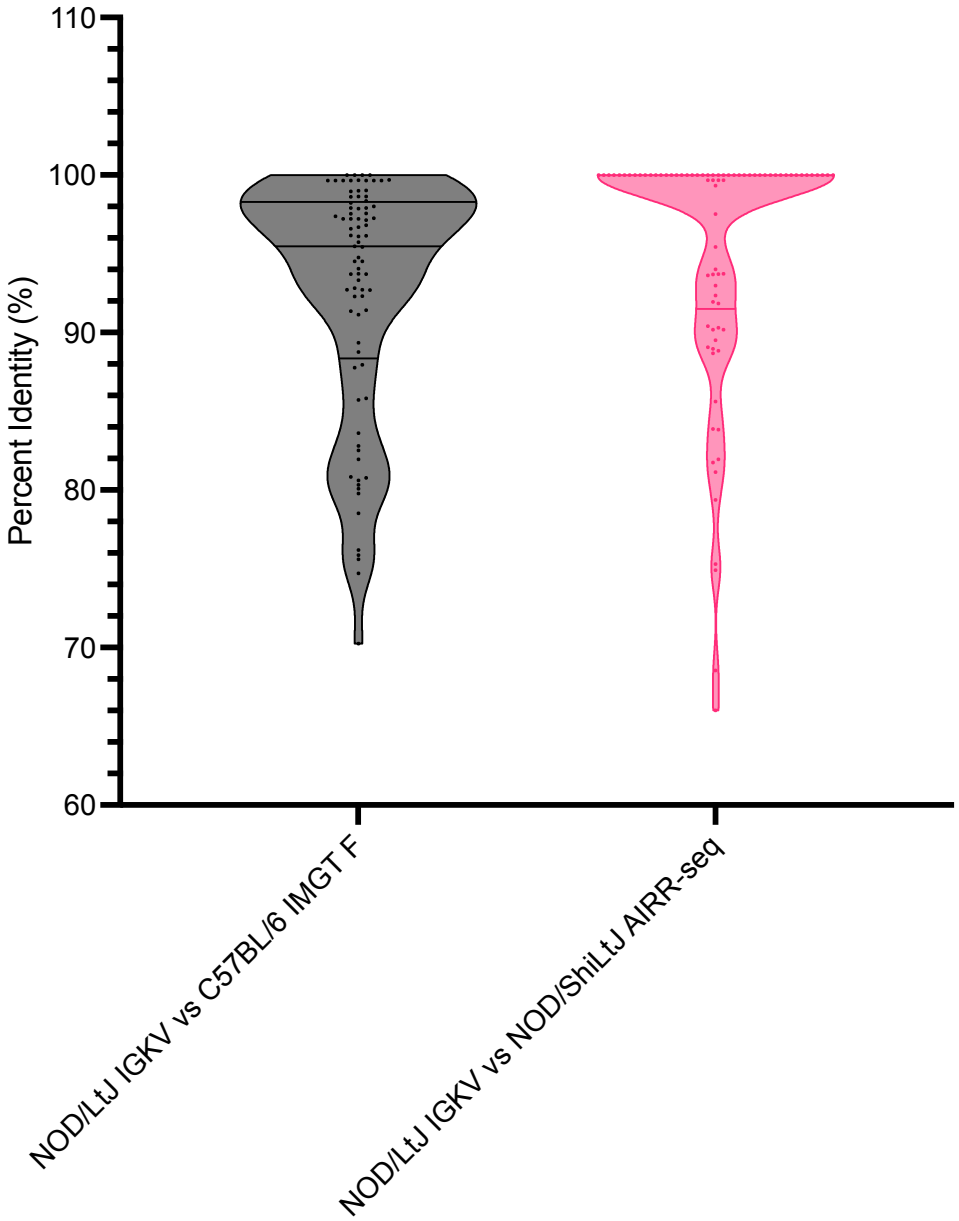


Figure 3.16. Alignment percent identities of NOD/ShiLtJ-IGK assembly IGKV sequences aligned to the IMGT C57BL/6 IGKV sequences and Chapter 2 NOD/ShiLtJ AIRR-seq IGKV inferences.

Next, to help validate our assembly sequences, we compared the NOD/ShiLtJ assembly sequences against the inferred NOD/ShiLtJ IGKV germline sequences from Chapter 2. The AIRR-seq data from Chapter 2 showed significant gaps in the IMGT gene database. In that experiment we inferred a total of 62 IGKV sequences for NOD/ShiLtJ, however, only 12 of those sequences were documented in IMGT. Because there were fewer NOD/ShiLtJ IGKV inferences compared to sequences identified in the assembly, we allowed inferences to be compared against multiple assembly sequences. 46 NOD/ShiLtJ IGKV assembly sequences have a 100% match to an inference sequence, with 31 NOD/ShiLtJ IGKV assembly sequences <98% identical to their nearest AIRR-seq inference (Figure 3.16). It is important to note that assembly sequences with lower percent identity alignments to inferences are not necessarily incorrect annotations. Sequences from the NOD/ShiLtJ IGK assembly that did not have a match to an inference could just be present at very low frequencies in the spleen and were missed during the AIRR-seq experiment and inference.

We also compared each assembled NOD/ShiLtJ-IGK contig to the C57BL/6 IGK locus by dot plot (Figure 3.17). From the dot plots we observed that the proximal region contig (NOD IGK B) has a high similarity match percentage to the proximal region of the C57BL/6 IGK locus, represented by the diagonal green line spanning approximately the length of the contig. The NOD IGK A contig has much

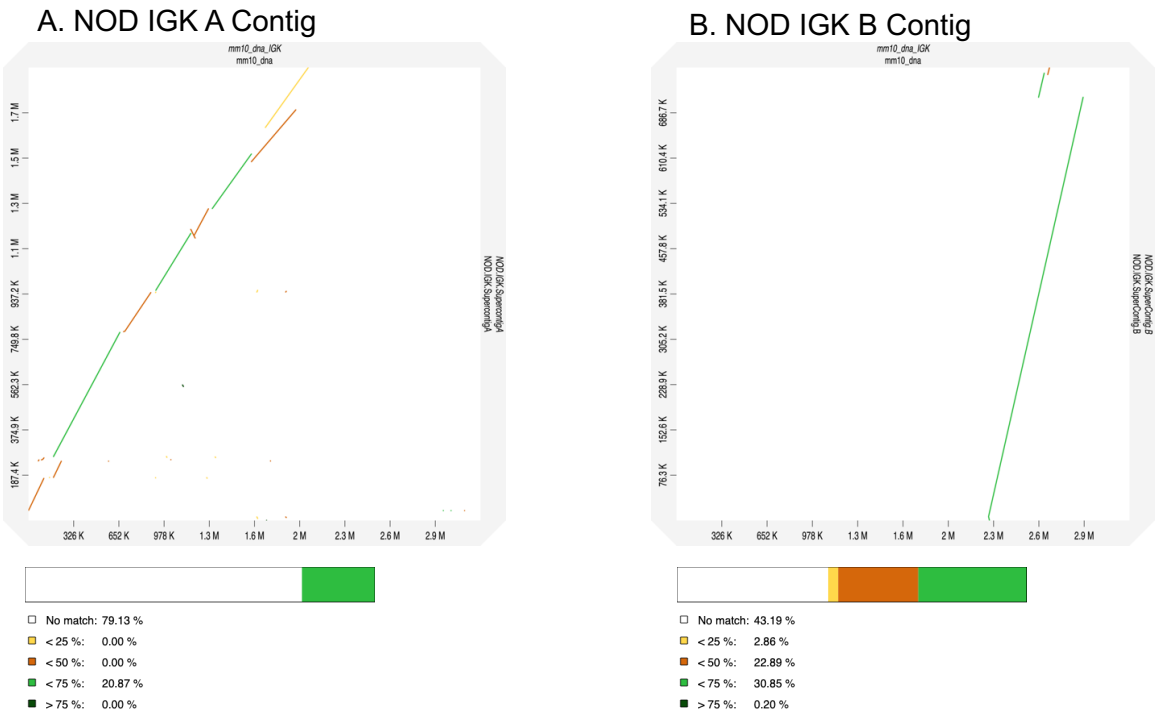


Figure 3.17. Dot plot comparison of NOD/ShiLtJ assembled contigs to C57BL/6 IGK locus.

Diagonal line represents alignment match, with the color of the line indicating the quality of the match.

greater sequence divergence from the C57BL/6 haplotype, shown by the numerous frameshift mutations in the dot plot.

Lastly, in 2010, Henry et al. published a set of germline wild-type NOD IGKV genes they obtained when trying to determine whether insulin-binding B-cells contain polymorphic V genes. In their analysis they noted polymorphisms throughout the NOD IGK locus, and that the polymorphisms represented 43 distinct IGKV genes belonging to 14 different IGKV gene families[70]. 20 germline sequences from this study were deposited into GenBank, and we compared these sequences against our annotated functional NOD/ShiLtJ IGKV sequences. 50% (10/20) of the GenBank sequences were 100% identical to our NOD/ShiLtJ IGKV assembly sequences, with the remaining 10 sequences ranging in percent identity from 99.65% to 91.96%. Although we would have expected greater overlap between the two gene sets, there are two possible explanations for this observation. First, our NOD/ShiLtJ assembly is not complete and may be missing germline genes that Henry et al. had identified. Second, the NOD mouse line that Henry et al. used in their experiment was not a NOD/ShiLtJ mouse like what our assembly was generated from, but rather a transgenic line constructed by introducing IGHV genes related to anti-insulin mab125 into the germline of a wild-type NOD mice[70,119,120].

DISCUSSION

The generation of new IG reference assemblies from BAC clones for divergent mouse IGH and IGK loci has revealed significant diversity when comparing the new assemblies to those that have already been generated for C57BL/6. The data generated thus far have supported early RFLP and SNP hypotheses that multiple IG haplotypes exist across commonly used mouse strains. From the comparisons we made to existing databases and to other strains, it is evident that the IGH and IGK loci from BALB/cByJ and NOD/ShiLtJ, respectively, diverge significantly from their respective C57BL/6 counterpart in both their coding and non-coding regions.

One future direction for this work will be the assessment of remaining gaps in the assembly followed by gap closure. Since these new IG assemblies diverge significantly from C57BL/6 and include large repetitive regions, we will use Bionano Saphyr platform (Bionano Genomics, San Diego, CA, USA) optical mapping to detect variation and genome architectural changes on a larger scale, further improving the quality of our assemblies. Briefly, Bionano is a scaffolding technology that can be used to bridge gaps between our individual contigs, ideally linking all five individual contigs together into a single contig[121]. Optical mapping is the backbone of the technology, and it works by tagging sequence motifs along unbroken DNA strands > 100 Kb in length, preserving large structural variations that exist within the sample[122]. Once the DNA is tagged with a fluorophore, the DNA is linearized, and single molecules are imaged and digitized using the Saphyr instrument. Finally, the DNA is uniquely identifiable through a distinct distribution

of sequence motif labels, which are then assembled by pairwise alignment into *de novo* genome maps. We will then compare our BALB/cByJ-IGH and NOD/ShiLtJ-IGK assembled contigs to the Bionano optical maps to correct errors in the assemblies, determine the size of gaps between contigs, and assess methods to close gaps depending on size.

Additional work will also be required to examine the non-coding regions of these assemblies. Studies have shown that the IGH locus contains six conserved sites that help orchestrate long-range chromosomal interactions[123]. These sites are critical for chromosomal looping that enables V(D)J recombination, but it is not known if the sites are conserved across all strains of mice. More importantly, three of the six sites fall within the distal region of the IGH locus. Our data shows that the distal region of the BALB/cByJ IGH locus diverges significantly from C57BL/6 when viewed as a dot plot. Therefore, the function and conservation of these sites in divergent haplotypes remains to be explored.

CHAPTER 4 : CONGENIC MOUSE MODEL FOR IG HAPLOTYPE VARIATION

INTRODUCTION

Other than infectious diseases, diseases that are under multi-factorial and multigenic or polygenic control are a significant public health threat in terms of the number of individuals affected. These diseases can result from the complex and subtle interactions between genetic and non-genetic factors, and include arteriosclerosis and hypertension, insulin-dependent diabetes mellitus, rheumatoid arthritis, asthma, systemic lupus erythematosus, multiple sclerosis, and cancer[124]. Most autoimmune diseases are under complex genetic control and are subject to strong interactions between genetics and the environment[125]. A key challenge to understanding the genetic basis of complex autoimmune diseases is the genetic and clinical heterogeneity that exists in human populations. Inbred mouse strains are valuable since they can reduce the genetic heterogeneity and have a standard environment and diet, thus reducing environmental variables. However, as previously showed in Chapters 2 and 3, considerable undocumented genetic variation exists within the IG loci of a variety of mouse strains.

The mouse has been an invaluable model organism in the immunogenetics field. Shortly after C. C. Little generated the first mouse lines in 1910, the mouse

played a significant role in immunology research, where it was involved in tumor studies and established the basic principles of tissue transplantation[47,92,126]. From there, the mouse model continued to gain traction in the biomedical sciences. During the mid-1950s, the mouse was used to study Ab-Ag interactions, lymphoid differentiation, and the response to infectious agents[127]. Inbred, congenic, and recombinant congenic mice made it possible to determine how the polymorphic MHC genes regulate the mammalian immune system[92]. Regulation of V(D)J recombination has also been studied extensively in mice and has explained basic molecular mechanisms underlying heavy and light chain IG gene regulation. Since an Ab is composed of two identical heavy and light chains, and these loci are encoded in different regions of the genome, various heavy and light chain pairings help increase the diversity of the Ab repertoire. Despite detailed knowledge of V(D)J rearrangement and IG loci architecture, this research is limited to C57BL/6 mice, which only represent a single haplotype for the IGH and IGK locus. Therefore, these models do not represent the IG variation shown to exist across mouse strains and human populations.

Besides combinatorial diversity, junctional diversity, somatic hypermutation, and allelic diversity, Ab heavy and light chain pairing is an essential aspect of the diversification of the Ab repertoire. Unique combinations of the complementarity-determining regions (CDRs) of the heavy and light chains create the Ag binding site in the variable domain of the Ab. The supposed random assembly of three gene segments for the variable heavy chain and two gene segments for the variable light chain occurs during V(D)J recombination and form the variable

domain of the Ab. However, not all combinations of V_H and V_L are beneficial. Some combinations of V_H and V_L may form autoantibodies, which can be avoided through secondary rearrangement of the light chain, deletion, or anergy. In human, the light chain repertoire in autoimmune diseases like Myasthenia Gravis can become perturbed and result in errors during receptor ed[101]ng during B-cell development[101]. Vander Heiden and colleagues noted reduced V-J segment distance in the light chain repertoire, which suggested decreased receptor editing during B-cell development[101]. It has been hypothesized that the function of the light chain is to help limit auto-reactivity, whereas the function of the heavy chain has evolved to increase repertoire diversity[5]. By limiting germline diversity in the light chain repertoire, the immune system may help prevent auto-reactivity that could arise if both the heavy chain and light chain loci contained significant germline diversity.

Support for this hypothesis stems from human AIRR-seq data, where most human IMGT IGKV alleles have no documented allelic variation, whereas only a single human IMGT IGHV gene lacked allelic variation[5]. We observed a similar trend in mice with diverse subspecific origins, with less inherent germline diversity in the kappa chain repertoire compared to the heavy chain repertoire (Chapter 2). While studies using IGHV and IGKV/IGLV transgenic mice have shown that some IGHV and IGKV/IGLV chain combinations can prevent B-cells from properly developing *in vivo*, no studies have comprehensively investigated the effect of germline variation on heavy and light chain pairing[128].

Recent high-throughput heavy and light chain repertoire sequencing in humans has provided insights to several medically and immunologically important concepts. Since light-chain rearrangements are more limited because of their lack of a diversity segment and restricted CDR lengths, they are expected to be less diverse and pair with multiple heavy chains[129]. These repeated light chains, often referred to as promiscuous light chains, can pair with a wide range of heavy chains, creating biased gene usage of heavy and light chain pairs in the Ab repertoire. However, only now with single-cell sequencing is it possible to identify which heavy chain genes pair with the promiscuous light chain genes in a high throughput manner. This is important since identification of heavy and light chain pairs can help identify Abs with broadly neutralizing Ab (bNAb)-like features[129]. Existing research has shown that there is a lack of understanding on the mechanisms that regulate the heavy and light chain loci, thus hindering our ability to better understand heavy and light chain pairing biases[130]. For example, while AIRR-seq studies show that divergent IG haplotypes have different Ab repertoires[4,45,87], it is not possible to discern without single-cell sequencing which heavy chains pair with a given light chain, and how haplotype variation affects heavy and light chain pairing and the immune response.

One of the overarching goals of this work was to better understand how haplotype variation shapes the Ab repertoire. Like most biomedical research, the use of model organisms, specifically mouse, allows the study of the Ab repertoire in a controlled manner. Unfortunately, one of the greatest obstacles in accurately and reliably studying repertoire diversity in genomics is the lack of high-quality

reference assemblies and model organisms. Until the new BALB/cByJ IGH and NOD/ShiLtJ IGK assemblies from BAC clones were created, the only full haplotype IG resource that existed was for C57BL/6 mice, which prevented researchers from asking questions related to the effect of haplotype diversity on the Ab repertoire[16]. If we are to understand the role of germline haplotype variation in the mouse Ab repertoire model and pursue research on IG genetic diversity, then we also need model organisms with controlled haplotype variation for both the heavy and the light chain loci.

The preexisting C57BL/6 mouse models and genome assembly do not allow segregation and characterization of the key genomic variants that contribute to Ab expression and function. While not genetically altering the background genome or other immunoglobulin loci, we have generated two congenic mouse lines on a C57BL/6 background genome that contain either a divergent IGH locus or a divergent IGK locus. Congenic mice are a special type of inbred mouse strain in which a defined chromosomal segment of one mouse strain is introgressed onto another by backcrossing a donor mouse strain to a recipient strain. Congenic strains are beneficial in that they help standardize potential genetic background effects and allow different haplotypes to be examined in the context of a different genetic background. Speed congenics can be generated in less time than traditional congenic lines by positively selecting for the desired chromosomal segment and selecting against the rest of the donor genome [131]. Therefore, the animals with the desired chromosomal segment, but little donor-strain genetic material, are selected for breeding.

To construct our congenic lines, we identified two mouse strains with predicted IGH and IGK haplotypes that are divergent from the IGH and IGK haplotypes for C57BL/6. The C57BL/6 mouse, whose germline IG genes make up most of the IMGT reference set[132], is predicted to contain an IGH locus derived from *M. m. musculus*, whereas its IGK locus originates from *M. m. domesticus*[42,43]. Therefore, the two strains chosen for the congenic lines must contain a divergent IGH haplotype with conserved IGK haplotype, and a conserved IGH haplotype with divergent IGK haplotype. The strain BALB/cByJ contains an IGH haplotype divergent from C57BL/6 (*M. m. domesticus* compared to *M. m. musculus*) but a shared IGK haplotype (*M. m. domesticus*). Conversely, the strain NOD/ShiLtJ contains an IGH haplotype similar to C57BL/6 (*M. m. musculus*) but a divergent IGK locus (*M. m. castaneus* compared to *M. m. domesticus*) (Figure 4.1). Since SNP-inferred haplotype studies have suggested that NOD/ShiLtJ and C57BL/6 have divergent IGK loci, we have created a NOD/ShiLtJ-IGK congenic line that contains a C57BL/6 genomic background and a NOD/ShiLtJ IGK locus. Abs made by these animals that use the IGK locus have a NOD/ShiLtJ light chain IG and a C57BL/6J heavy chain IG. Conversely, since BALB/cByJ and C57BL/6 are predicted to be divergent at the IGH locus, we have created a BALB/cByJ-IGH congenic line that contains a C57BL/6J genomic background and produces Abs with a BALB/cByJ heavy chain IG and a C57BL/6 light chain IG (Figures 4.2 and 4.3). Our C57BL/6 control mouse for these experiments is the B6(Cg)-*Tyr*$^{c\text{-}2J}$/J mouse, which is a C57BL/6J mouse with a mutation in the tyrosinase gene that makes the animals' skin, hair, and eyes absent of pigment[133].

For this experiment we compared productive and unproductive naïve gene usage frequencies between the BALB/cByJ-IGH congenic line, BALB/cByJ control mice, and B6(Cg)-*Tyr*$^{c\text{-}2J}$/J control mice for the naïve IGK and IGL germline repertoire. Since the BALB/cByJ-IGH congenic line and B6(Cg)-*Tyr*$^{c\text{-}2J}$/J control mice are genetically identical except for their IGH locus, initial comparisons were to investigate the effect of a divergent IGH haplotype on the light chain repertoire. Subsequent comparisons included the BALB/cByJ control line, which has a genome that is completely BALB/cByJ.

Sequences were grouped by whether they were predicted to be productive or unproductive, with productive sequences defined as those with an open reading frame and no stop codon, no defect in the initiation codon, and proper splicing[134]. The rationale for separately comparing gene usage frequencies for productive and unproductive sequences was to help identify significant IGKV genes that contributed to differences in the Ab repertoire. Unproductive sequences would not

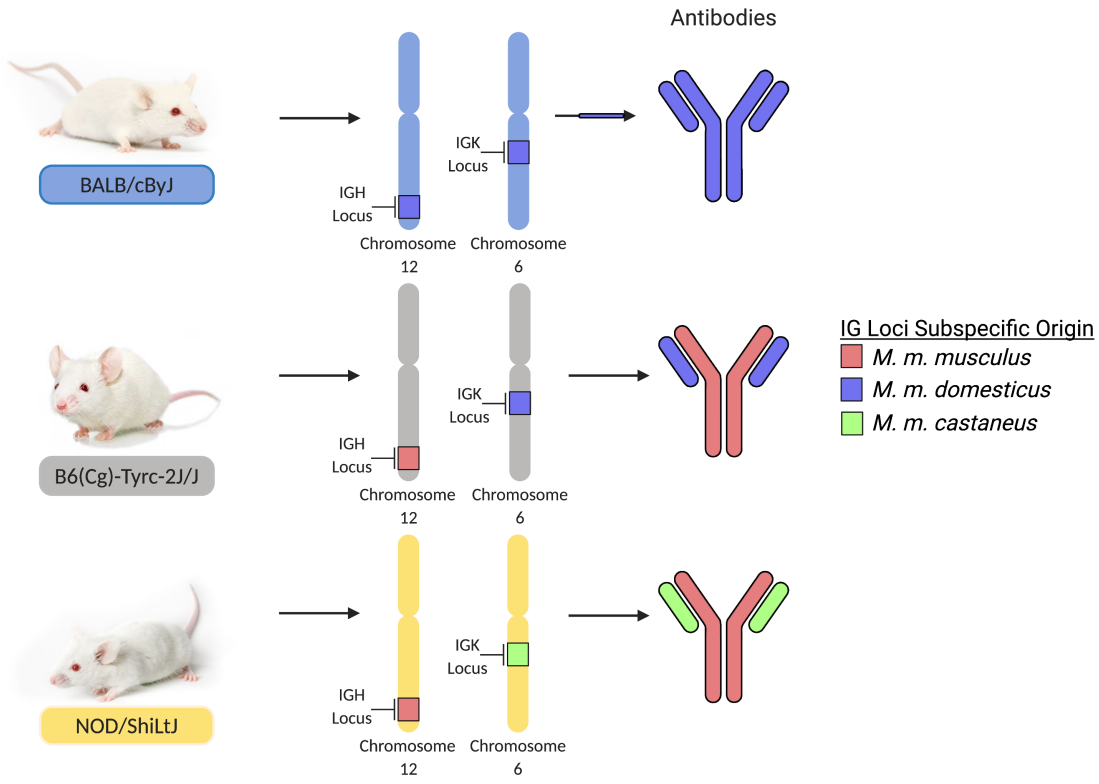


Figure 4.1. Depiction of the three mouse strains used to create congenic IGK and IGK mouse lines.

The color behind the strain name and of the chromosomes represent each strain's background genome. The colors of the IGH and IGK loci correspond to the predicted subspecific origins of each locus. Finally, haplotype diversity at the Ab-level is represented by the subspecific origin-colored antibodies.

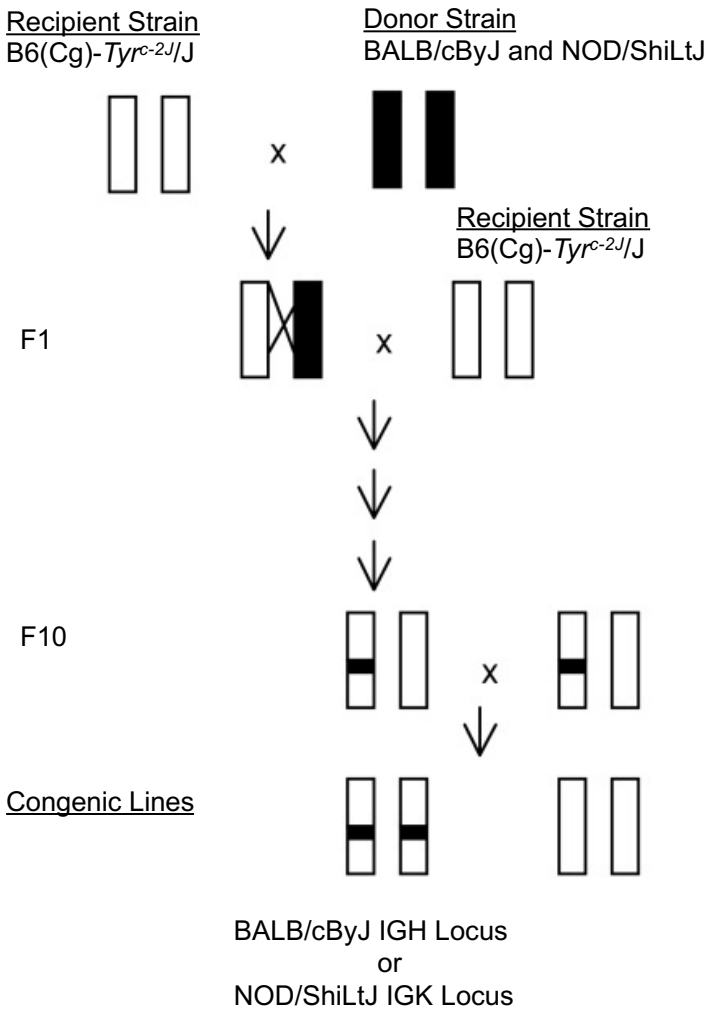


Figure 4.2. Generation of congenic mouse model.

Adapted from Kim et al. (2010)[135]. Mating of two inbred mouse strains, C57BL/6 x BALB/cByJ, and C57BL/6 x NOD/ShiLtJ, results in a heterozygous F1 generation. F1 animals are then backcrossed to the recipient strain, B6(Cg)-*Tyr*$^{c-2J}$/J , and this cycle is repeated for 6 generations before the Jax Labs genome scan. Mice are genotyped at each generation to ensure the donor IGH or IGK locus is present in the region of interest. At the end of congenic line construction, mice heterozygous for IGH or IGK markers are crossed to produce offspring homozygous for

BALB/cByJ IGH or NOD/ShiLtJ IGK (depicted at bottom with black donor loci on recipient strain genomic background).

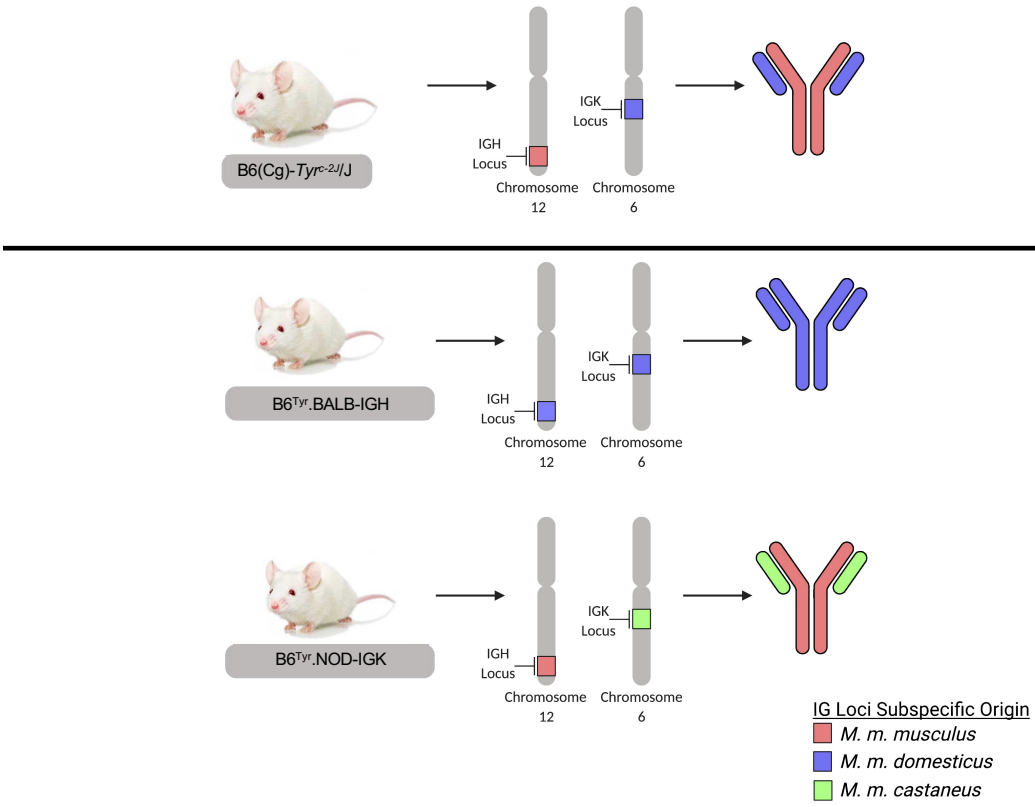


Figure 4.3. Congenic BALB/cByJ-IGH and NOD/ShiLtJ-IGK relation to control B6(Cg)-*Tyr*$^{c\text{-}2J}$/J.

The two congenic lines and how their antibodies compare to a B6(Cg)-Tyr$^{c\text{-}2J}$/J control mouse background genome (top), IGH and IGK loci subspecific origin, and Ab haplotype diversity. Both congenic lines will contain a B6(Cg)-Tyr$^{c\text{-}2J}$/J background genome, which is represented by the grey name and chromosomes, noting the similarity to the B6(Cg)-Tyr$^{c\text{-}2J}$/J control mouse (top). However, the congenic lines will contain either a divergent IGH or IGK locus, which is represented by the subspecies origin-colored IGH and IGK loci. BALB-IGH congenic mice will be genetically similar to the control mouse except for the IGH locus. NOD-IGK congenic mice will be genetically similar to the control mouse

except for the IGK locus. The colored Ab molecules represent the effect of IG loci subspecific origin.

be translated into functional antibodies; therefore, they would not undergo selection and would not pair with a heavy chain IG molecule. Productive sequences would be under selection and would be required to pair with a heavy chain IG molecule. We hypothesized that the underlying IGH haplotype difference between the BALB/cByJ-IGH congenic and B6(Cg)-*Tyr*$^{c\text{-}2J}$/J control would drive usage frequency differences of productive sequences but not unproductive sequences. Using a principal component analysis (PCA) we observed changes in the IGKV repertoire that were an effect of the divergent IGH locus in the congenic line, and this effect was still present in the data after refining the dataset to include only genes with significant usage frequency differences.

The new IGH and IGK assemblies (Chapter 3), combined with these new congenic IG mouse models, address the lack of genomic resources and models that account for mouse haplotype diversity in the IG loci. The congenic mouse lines leveraged the newly characterized genomic haplotype resource, such that future research can investigate the Ab repertoire in a controlled way. Combined, these resources and models will enable the field to accurately and reliably investigate the effects of IG haplotype variants on repertoire development.

MATERIALS AND METHODS

BREEDING SCHEME

Two different crosses generated congenic mouse lines with controlled IG genetic diversity (Figure 4.3). The first cross was a B6(Cg)-*Tyr*$^{c\text{-}2J}$/J to a NOD/ShiLtJ, and the second was a BALB/cByJ to a B6(Cg)-*Tyr*$^{c\text{-}2J}$/J. Nomenclature for the congenic strains is provided in Table 4.1. Strain abbreviations will be used throughout the remainder of the text. Primers amplified regions containing known SNPs between the two strains across the two loci, besides regions upstream and downstream of the loci[42,43] (Table 4.1). The UCSC genome browser *in silico* PCR tool was used to exclude off target primers[111]. Each animal was identified and a biopsy taken for genotyping with IGH or IGK SNP markers, and only heterozygous individuals were selected for breeding. Breeding continued until the F6 generation, when a complete genome scan at The Jackson Laboratory[133] was performed using 143 NOD vs. B6 SNP markers (147 BALBc vs. B6) to determine the background B6 genomic content. At this stage, the congenic mice with the highest percentage of B6 SNPs, and still maintained the congenic IGH or IGK locus, were selected for additional breeding. Once animals were identified that genotyped as homozygous B6 for the markers tested on the genome scan, a final Jax genome scan was performed. Once the final cohort of animals were selected, they underwent brother-sister mating to obtain either a homozygous NOD IGK locus, or homozygous BALBc IGH locus.

Table 4.1. Nomenclature and abbreviations for mouse strains used in Chapter 4.

Strain	**Type**	**Abbreviation**
B6(Cg)-*Tyr*$^{c\text{-}2J}$/J	Control	B6
BALB/cByJ	Control	BALBc
NOD/ShiLtJ	Control	NOD
B6(Cg)-Tyrc-2J/J. BALB/cByJ-IGH	Congenic	B6^{Tyr}.BALB-IGH
B6(Cg)-*Tyr*$^{c\text{-}2J}$/J. NOD/ShiLtJ-IGK	Congenic	B6^{Tyr}.NOD-IGK

Table 4.2. B6^{Tyr}.BALB-IGH and B6^{Tyr}.NOD-IGK Congenic Mouse Line Genotyping Primers.

Locus	SNP Position	Primer Name	Forward Primer Sequence	Reverse Primer Sequence	Variant (Balb/B6, NOD/B6)
IGH	Chr12: 114495014	IGH 5'	CCAAGGCTCCTTAGGTCTCAGTCC	CTGGTTTCTAGCAACTTCCCTCTG	G/A
IGH	Chr12: 114547640	IGH VDJ	GCCTATGTTCCTGCTCAATCGGAC	AGGATGTGCAGGATCTAGAAGCTCA	G/C
IGH	Chr12: 117267709	IGH 3'	TGTGGAAATCCTTTGGATGAG	ACAGGGCCAAGGAGAGTCTT	A/G
IGK	Chr6: 067287072	IGK 5'	GCAAAGGCAATGGGTTCTTA	CAGGCTGACAGAGACATGGA	C/T
IGK	Chr6: 067483466	IGK VDJ	GCACTTGAGCAGAATGTGGA	AAGCATGGGAGAATGAGGTG	A/G
IGK	Chr6: 070754483	IGK 3'	TTAACCGGGAGAACAACCAC	CACCCCCCACAGAAAACAAAC	T/C
IGK	Chr9: 122337931	rs3672156	GAAACCAGGAAGGGTGTGTC	AGGAGAACTCTGGAGATCCATT	G/A
IGK	Chr6: 053482462	rs3680652	CACACACACACAACCCAAGA	GACAGATTACAGGGCGCAAT	A/G
IGK	Chr6: 090142535	rs3708822	GCCTTGGTTGAGGAGGTGAT	TGAGGCTCTGAGTGGAGGTT	A/C
IGK	Chr6: 114232644	rs3693392	GTGGTAGAACTGCCCAGAGC	CATCCAATCTCCCTGGAACT	A/G
IGK	Chr6: 038230055	rs6266537	CAAATGCCAATTAGCAAGCA	CTGCCATCAAAGGGTGACTA	G/A
IGH	Chr17: 089136255	rs3659009	TCTCCATGTTTTTCCTTCAATG	TGGGTTAATAGAGTGCATGAGG	C/T
IGH	Chr10: 083144168	rs3717445	GCAGGTTCACATGTGTGCAG	TTCCTGATTACCAAGCCGGG	G/C
IGH	Chr11: 023035271	rs4228627	TGGCAGAGGTCATTTTGTGC	TCCCCTTTGTCATTTCCTCAGA	C/T
IGH	Chr12: 112276179	rs3686631	ACGAATGCCTGAATGTCACCT	GCAACCCTTTTCTTTTCTCTGCT	T/A

GENOTYPING

Briefly, mouse tail biopsies were incubated in 100 µL of DirectPCR Lysis Reagent plus 1 µL of proteinase K (1µg/1mL) (Viagen Biotech, Cat. No. 302-C, Los Angeles, CA, USA) at 50ºC for 9 hours then 45 minutes at 85ºC to inactivate proteinase K. DNA extract (1µL) was used for PCR genotyping. Initial genotyping used three primer pairs for each congenic strain (Table 4.2). Since the goal of this project was to introgress the divergent IGH or IGK locus onto the B6 background, we identified three unique SNP positions for each line that were positioned 5', within, and 3' of the IG loci of interest. These positions were generated from the Mouse Phylogeny Viewer[42,136], which is a database of SNP-predicted variation across a variety of classical and wild-derived mouse strains. After 7 generations of backcrossing, litters were also genotyped with primers that were internal to the IGH and IGK locus, which were designed using the BALBc IGH and NOD IGK reference assemblies (Chapter 3). The Mouse Phylogeny Viewer[136] and the Jax Genome Scan[133] both use the mm9 reference genome. All primer sequences were designed from this reference. If a sample was genotyped to be heterozygous for the V(D)J marker, it was then genotyped for the upstream and downstream markers to ensure it contained the complete IGH or IGK locus from BALBc and NOD, respectively. After 7 generations of backcrossing, a Jax genome scan was completed for both lines, which uses a custom SNP panel to confirm the amount of B6 background present. Since our goal was to have the entire genome B6 except for the IGH or IGK locus in BALBc and NOD, respectively, any SNP positions revealed to be a genotype other than B6 were eliminated through

additional backcrossing. A standardized PCR reaction was used for all genotyping, with 1 µL of crude DNA extract used with Qiagen AllTaq Master Mix Kit (Cat. No. 203144; Germantown, MD, USA). DNA was PCR amplified at cycling conditions of 94ºC for 30 seconds, 32 cycles of 94ºC for 30 seconds, 60ºC for 30 seconds, and 72ºC for 1 minute, and 72ºC for 10 minutes. PCR samples were Sanger sequenced using an ABI PRISM 3130XL Genetic Analyzer and interpreted with Geneious Prime 2022.0.2 (https://www.geneious.com).

IG LOCI INSERT VALIDATION

To ensure that the complete IG locus was crossed onto the B6 background, and to estimate how much additional donor strain genome was present, we designed primers using SNP positions downloaded from the Mouse Phylogeny Viewer that flanked the loci and localized the crossover key events. Primers were designed both upstream and downstream of the BALBc IGH (Table 4.3) and NOD IGK (Table 4.4) loci that were introgressed onto the B6 background.

Table 4.3. B6^{Tyr}.BALB-IGH Congenic Line Insert Validation Primers.

Locus	SNP Position	Forward Primer Sequence	Reverse Primer Sequence	Variant (Balb/B6)
IGH	Chr12: 96534266	GCTCTTGCCTCTCAGGTCTG	TTTGATCACCCAGGAGGCAC	A/T
IGH	Chr12: 99208162	TGCAGTGGGAACTTGATGCA	TATACACCATGGGGGGAGCCA	C/A
IGH	Chr12: 100579781	ACATCTCTAGAGGCAGGCCA	TCTACTCGGGGCAAAGAGGA	A/G
IGH	Chr12: 101959821	CCCCAGCAACCCTTCATCTT	CTGTGGCGTGTGGTAGATGT	C/T
IGH	Chr12: 102902264	GCTCTCTAGCCCCCATCAAGC	TCTCCGCCCCCAGTATATGT	G/A
IGH	Chr12: 105862390	ATCAGGCCCCAACAGGAATG	TAGACACTCCCTGCCCTCTC	C/T
IGH	Chr12: 108360092	ATGGCCCTGGTGACATCTTG	CTGTCCTTGCCTTCTCTGGG	T/G
IGH	Chr12: 110726860	AGGCTAGCGTGGTCTACAGA	GTCTGACTGGCCTGGAACTC	C/T
IGH	Chr12: 114388138	GGGGCAATGGATGGAAGGAA	TCACTGTCCGTGGAAGAAGC	T/C
IGH	Chr12: 118008045	CAAACCTTCAAGCCCAGGTG	GGTGGGTGTCAAGTTCTCCA	A/G
IGH	Chr12: 119267006	ACTCCGCAGACACTGAAAACA	GCAGATTGCGCGTGTTCTTT	T/C
IGH	Chr12: 119437920	TCCAGGACTAGTTAGCGGCT	CAGCATCAGCAGTGATTGGC	C/T
IGH	Chr12: 121005208	AGAGATCTCCAGGGGCACTT	GAGGTGTGGGATGTGAAGCA	C/T

Table 4.4. B6^{Tyr}.NOD-IGK Congenic Line Insert Validation Primers.

Locus	SNP Position	Forward Primer Sequence	Reverse Primer Sequence	Variant (NOD/B6)
IGK	Chr6: 53575131	CCAATGGCCACTTCCCTTC	CCACATCCCTCCGCTTTTC	T/C
IGK	Chr6: 57060038	TGCAAGTCCATCCTCAACC	TGTCCACTGTCAACAACAC	A/G
IGK	Chr6: 61587593	GCTTCCTCTCTGCTCGCTA	GGGACATCCACTGAAGGG	G/A
IGK	Chr6: 65836389	GGCTTAAGCAGTAGACGG	CGTAAGCTACAGGTGGGTC	T/C
IGK	Chr6: 67005882	ATGAGAGTCAGAGCCAGG	TTTCCAGTGTACGAACCCC	C/G
IGK	Chr6: 70712600	TCCATGGGGTGATGTCGT	CATAGCAATGCGGCAGACA	A/T
IGK	Chr6: 71769288	GGTTCCTGACGCCAGATT	TCTCCCCCTCCACTTTCCA	A/G
IGK	Chr6: 73235717	ATGAAGCTTGGCTCTGCT	ACCTGGACCCCTGAGATCT	C/T
IGK	Chr6: 77383011	GGATGGGGTGGAAATCAG	GCACAGCTCTGTAGATGC	T/G
IGK	Chr6: 84797590	GCAGCACTAGAAGCCAGG	CGGGAGTAGGAGTCAGCA	G/T

We used NCBI Primer BLAST to generate unique primers that would not amplify multiple regions within our IG reference assemblies and the B6 background genome. Since the Jax genome scan is not a fine map of the size of the congenic insert, these primers were used to provide greater resolution to the location of the crossover event. In addition, since the crossover was not specific to just the IGH or IGK locus, genes upstream and downstream of the locus were investigated to determine if they were BALBc, NOD, or B6, and whether those genes would influence V(D)J recombination or the immune system. To assess which BALBc or NOD genes were present on the B6 genomic background, the genomic regions from the insert validation were uploaded to UCSC Table Browser[137].

BULK AIRR-SEQ LIBRARY PREPARATION AND SEQUENCING OF CONGENIC IG MOUSE MODELS AND CONTROLS

AIRR-seq was performed on $B6^{Tyr}$.BALB-IGH to assess the congenic animals' light chain Ab repertoire. Experimental samples are summarized in Table 4.5. Spleens were dissected from 10-week-old mice and preserved in RNAlater (Thermofisher, Cat. No. AM7020; Waltham, MA, USA). Total RNA was extracted from 30 mg of spleen tissue using the RNeasy Mini kit (Qiagen, Cat. No. 74104; Germantown, MD, USA). For each sample, IGK and IGL 5'RACE AIRR-seq libraries were generated using the SMARTer Mouse BCR

Table 4.5. Experimental samples for bulk IGM, IGK, and IGK Illumina AIRR-seq on congenic lines.

Sample ID	Short Name	Strain	Sex
CW42_01_BALB-IGH_F2	IGH-F2	$B6^{Tyr}$.BALB-IGH	F
CW42_01_BALB-IGH_F3	IGH-F3	B6Tyr.BALB-IGH	F
CW42_01_BALB-IGH_F4	IGH-F4	B6Tyr.BALB-IGH	F
CW42_01_BALB-IGH_F5	IGH-F5	B6Tyr.BALB-IGH	F
CW42_01_BALB-IGH_F6	IGH-F6	B6Tyr.BALB-IGH	F
CW42_02_BALB-IGH_M1	IGH-M1	B6Tyr.BALB-IGH	M
CW42_02_BALB-IGH_M2	IGH-M2	B6Tyr.BALB-IGH	M
CW42_02_BALB-IGH_M3	IGH-M3	B6Tyr.BALB-IGH	M
CW42_02_BALB-IGH_M4	IGH-M4	B6Tyr.BALB-IGH	M
CW42_02_BALB-IGH_M5	IGH-M5	B6Tyr.BALB-IGH	M
CW42_06_BALB JAX_F1	JAX-F1	BALBc	F
CW42_06_BALB JAX_F2	JAX-F2	BALBc	F
CW42_06_BALB JAX_F3	JAX-F3	BALBc	F
CW42_06_BALB JAX_F4	JAX-F4	BALBc	F
CW42_06_BALB JAX_F5	JAX-F5	BALBc	F
CW42_07_BALB JAX_M1	JAX-M1	BALBc	M
CW42_07_BALB JAX_M2	JAX-M2	BALBc	M
CW42_07_BALB JAX_M4	JAX-M4	BALBc	M
CW42_07_BALB JAX_M5	JAX-M5	BALBc	M
CW42_09_B6 TYR_F5	TYR-F5	B6	F
CW42_08_B6 TYR_M1	TYR-M1	B6	M
CW42_08_B6 TYR_M2	TYR-M2	B6	M
CW42_08_B6 TYR_M3	TYR-M3	B6	M
Light.Chain.B6	LC.B6	B6	F

Profiling Kit (Takara Bio, Cat. No. 634422; Mountain View, CA, USA), following the manufacturer's instructions. Individually indexed IGK and IGL AIRR-seq libraries were assessed using the Agilent Fragment Analyzer 5200 (Agilent, Part No. M5310AA; Santa Clara, CA, USA) and the Thermofisher Qubit 3.0 Fluorometer dsDNA High Sensitivity Assay Kit (Thermofisher, Cat. No. Q32851). Libraries were pooled to 10 nM and sequenced on the Illumina MiSeq platform using the 600-cycle MiSeq Reagent Kit v3 (2x300 bp, paired-end; Illumina, Cat. No. MS-102-3003). Bulk AIRR-seq was only performed for the B6^{Tyr}.BALB-IGH, BALBc, and B6 because insufficient NOD/ShiLtJ-IGK mice were available. In addition, because at the time of this writing we are waiting for additional data for B6 samples, we have included the C57BL/6J IGK and IGL data from Chapter 2 as an additional sample to increase the number of control samples (noted in Table 4.5). This additional sample was prepared using the same SMARTer Mouse BCR Profiling Kit (Takara Bio, Cat. No. 634422; Mountain View, CA, USA), and sequenced on the Illumina MiSeq platform.

DATA PROCESSING FOR BULK AIRR-seq OF CONGENIC IG MOUSE MODELS AND CONTROLS

Bulk AIRR-seq reads were processed using the Immcantation[63,64] suite of tools to characterize B-cell repertoires, with the IMGT mouse IGK and IGL reference directory sets[132] serving as the starting IgBLAST[67] database for germline gene and allele assignment. First, sequences were filtered to a quality score

greater than or equal to Q25 using FilterSeq. Next, paired-end reads were assembled using the align command of AssemblePairs. Then, IGK and IGL primer sequences (Supplemental Table 1) were identified using a local alignment and cut from the assembled reads using MaskPrimers align, with a max error rate of 0.2. Because the primer sequences were not provided with the SMARTer Mouse BCR Profiling Kit, we manually determined the primer sequences by performing a multiple sequence alignment of the first 30 base pairs of the Illumina R1 reads. Next, duplicate reads were collapsed using collapseSeq, with the duplicate count of each collapsed sequence recorded as “dupcount”. Downstream processing required that all sequences have a dupcount ≥ 2. Assignments to germline IGKV and IGLV were performed using IgBLAST[67], with the resulting output parsed with the Change-O module MakeDb[63]. Clones were identified using defineClones, with clonal thresholds determined independently for each strain using the distToNearest function in SHazaM[63].

IDENTIFYING DIFFERENTIAL USAGE OF IGKV GENES BETWEEN LINES

We performed a PCA on productive and unproductive sequences to test the hypothesis that haplotype variation in the IGH locus will affect the light chain repertoire The built-in R function prcomp was used to calculate the principal components of the dataset[138], and the R package factoextra[139] was used to construct PCA plots. In addition, a one-way ANOVA was performed on IGKV and IGLV gene usage frequency to determine if there were usage frequency

differences. A principal component analysis was then performed on genes with a P-value <0.05 to observe if the underlying haplotype variation contributed to the usage frequency differences. Finally, P-values were adjusted for multiple comparisons by controlling for false discovery rate.

RESULTS

$B6^{Tyr}$.BALB-IGH AND $B6^{Tyr}$.NOD-IGK CONGENIC LINES

The size of the congenic insert varied for each line, and neither line had a precise introgression of the IG locus of interest. Using PCR primers flanking both sides of the IGH locus, we estimated the $B6^{Tyr}$.BALB-IGH congenic insert to extend from Chr12: 112,276,179 to at least Chr12: 121,005,208; approximately 8.7 Mb (Figure 4.4). We could not resolve genotypes between the telomere and Chr12: 121,005,208 due to lack of SNPs in that region. For the $B6^{Tyr}$.NOD-IGK congenic line, we estimated the congenic insert to be approximately 39 Mb of NOD/ShiLtJ genome extending from Chr6: 38,230,055 to Chr6: 77,383,011 (Figure 4.5). In the $B6^{Tyr}$.BALB-IGH congenic line, we estimate that 89 BALBc genes (Supplemental Table 7) were introgressed onto the B6 background genome in addition to the IGH locus. Of the 89 genes, none were enriched for any biological processes as determined by the GO Enrichment from The Gene Ontology resource[140–142]. In the $B6^{Tyr}$.NOD-IGK congenic line, 398 genes were introgressed onto the B6 background genome (Supplemental Table 8). Three GO biological processes,

including cellular responses to copper ion, response to pheromone, and embryonic skeletal system morphogenesis represented these NOD genes[140–142].

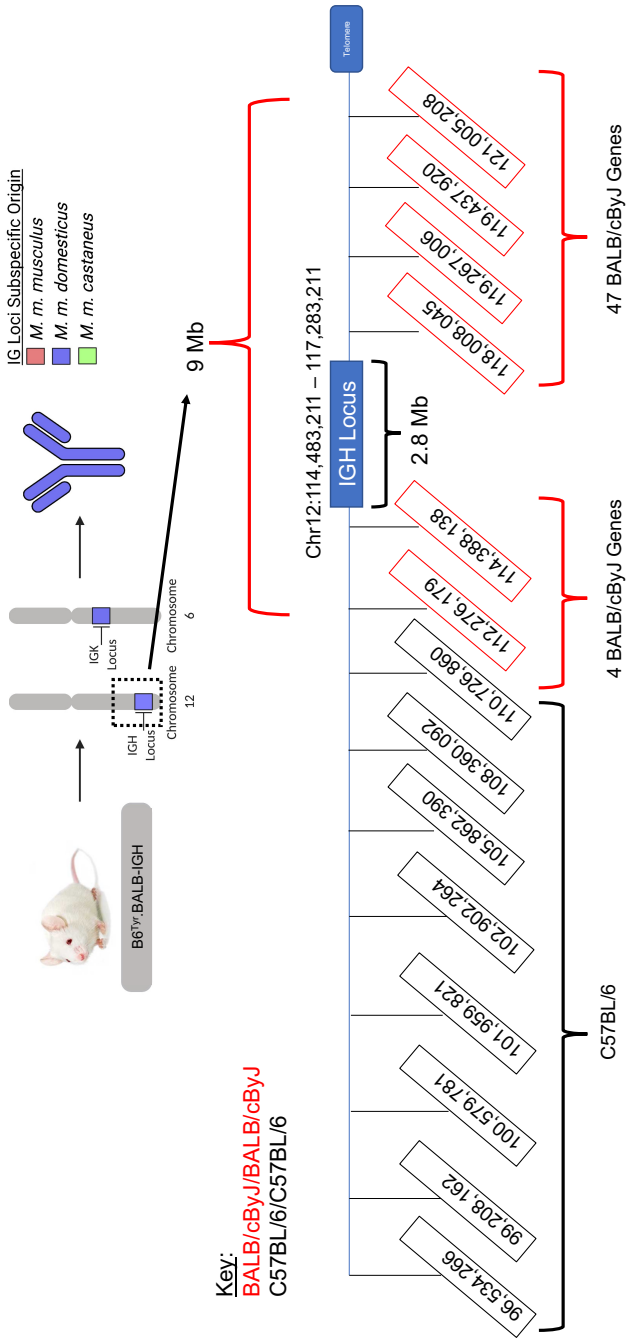


Figure 4.4. B6^{Tyr}.BALB-IGH congenic insert validation.

Validation of the BALB/cByJ congenic insert was performed using primers flanking both the 5' and 3' regions of the IGH locus. Positions were selected that had SNP-predicted nucleotide differences between BALB/cByJ and C57BL/6. See Table 6.8 for full list of introgressed BALB/cByJ genes.

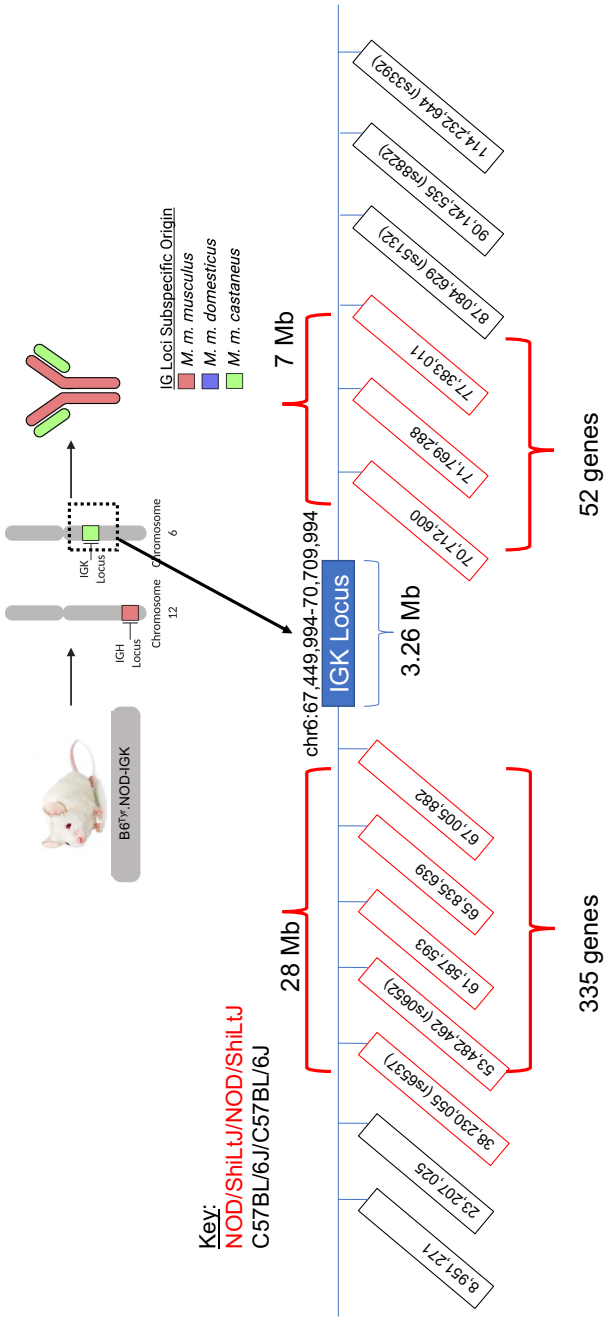


Figure 4.5. B6^{Tyr}.NOD-IGK congenic insert validation.

Validation of the NOD/ShiLtJ congenic insert was performed using primers flanking both the 5' and 3' regions of the IGK locus. Positions were selected that had SNP-predicted nucleotide differences between NOD/ShiLtJ and C57BL/6J. See Table 6.9 for full list of introgressed NOD/ShiLtJ genes.

EFFECTS OF IGH HAPLOTYPE VARIATION ON LIGHT CHAIN AB REPERTOIRE – B6^{Tyr}.BALB-IGH VS. B6

One of the key objectives for this experiment was to determine the effect that haplotype variation has on the Ab repertoire. This initial experiment on the newly generated congenic lines was to test the hypothesis that, due to heavy and light chain pairing, haplotype variation within the IGH locus could cause usage frequency changes in the light chain Ab repertoire. To test this hypothesis, we leveraged the B6^{Tyr}.BALB-IGH congenic line and the control B6 line. The only genetic difference between these two lines was the 8.7 Mb introgressed BALBc donor genome that contained the IGH locus. Therefore, the only difference between these two animals was that the B6^{Tyr}.BALB-IGH congenic animals would produce antibodies with a divergent heavy chain IG compared to the B6 control (Figure 4.2). From my data in chapter 2 and published literature[28], the IGK haplotype was the same for BALBc and B6. Thus, to determine the effect that the divergent IGH locus had on the light chain Ab repertoire, comparisons were made between the sample groups' germline IGKV repertoires.

We first separated sequences according to whether they were productive unproductive. The rationale for separating sequences according to their productivity was because unproductive sequences would not make a functional Ab and would therefore not undergo selection. Productive sequences, however, would produce an IG transcript that could be translated into an IG protein, and thus undergo heavy and light chain pairing to form an Ab that could undergo selection.

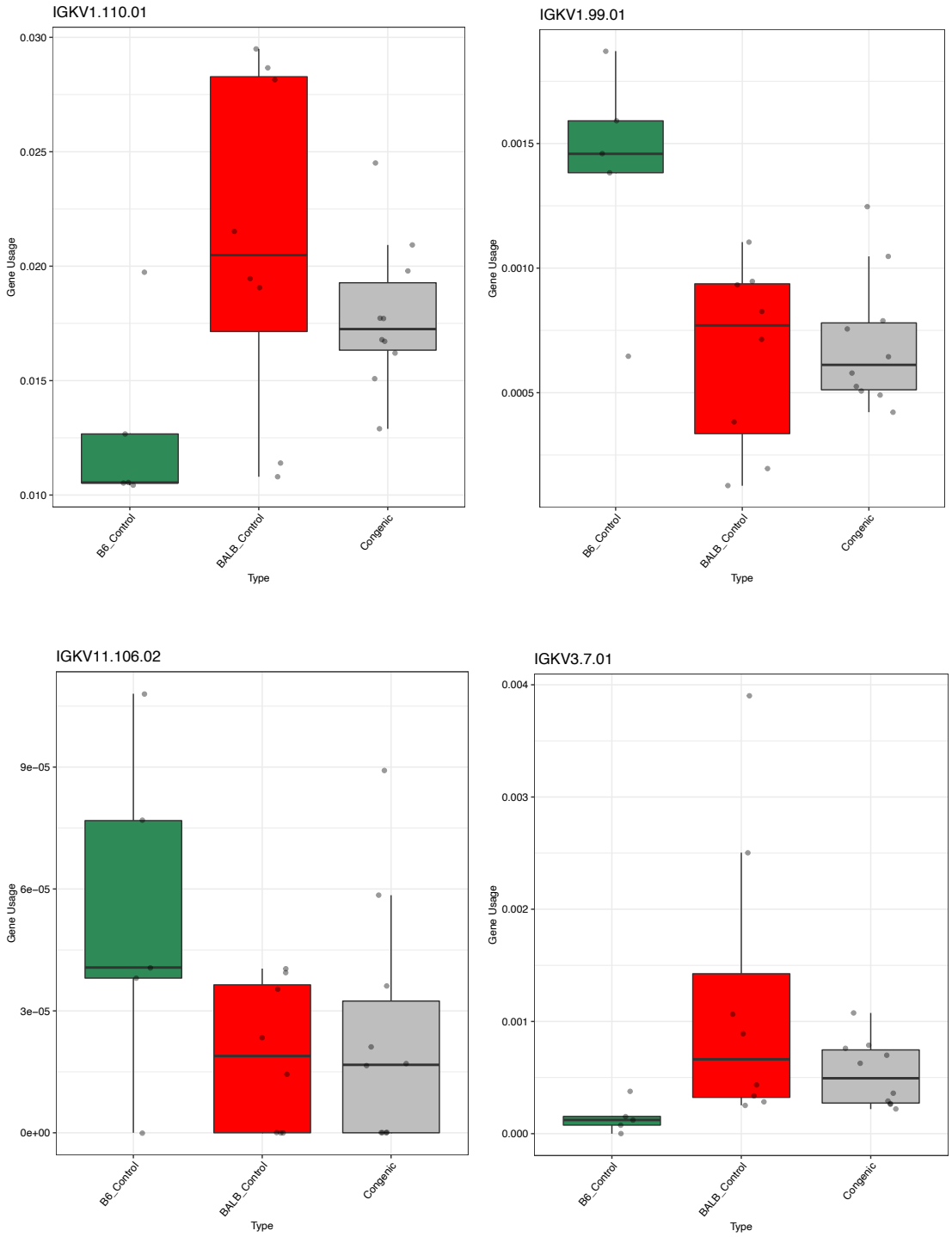
A one-way ANOVA performed on IGKV usage frequencies in the B6^{Tyr}.BALB-IGH congenic and B6 control animals identified IGKV sequences with significant usage frequency differences. Only 10 IGKV genes (10/127) had P-values < 0.05 (Figure 4.6, Table 4.6). A PCA analysis of these 10 significant genes showed distinct clusters for the B6^{Tyr}.BALB-IGH congenic animals and the B6 control animals (Figure 4.7A). As expected, a PCA of these same 10 genes, but in unproductive sequences, showed no difference between the B6^{Tyr}.BALB-IGH congenic and B6 control animals (Figure 4.7B). After adjusting the P-values for multiple comparisons by applying the method of Benjamini and Hochberg to control the false discovery rate[143], only IGKV 1.110.01 had a significant usage difference between B6^{Tyr}.BALB-IGH and B6.

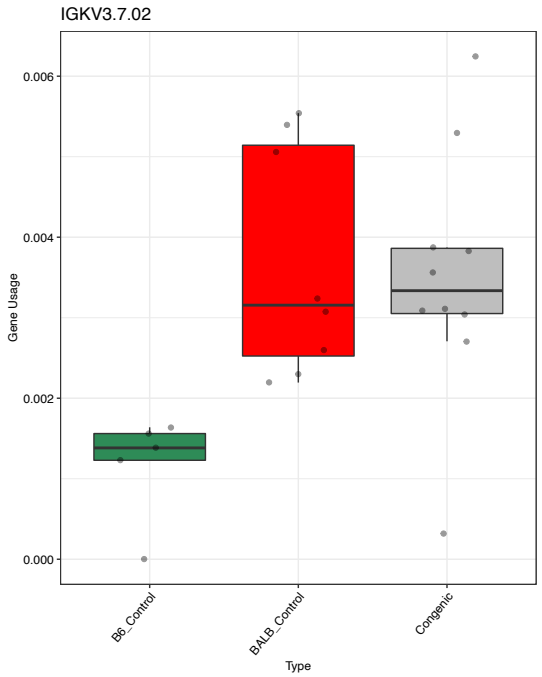
In addition, a PCA calculated according to sex showed that both male and female samples overlapped, indicating that the IGH haplotype clustering observed in the data was not influenced by sex (Figure 4.8). Last, we used a PCA biplot to determine whether particular IGKV sequences could have been overrepresented in the sequencing data and affected the distribution of the PCA clusters (Supplemental Figure 3). IGKV8-21 and IGKV8-30 were visible on the biplot, however, gene usage plots across all groups showed that IGKV8-21 had a greater variance in the B6 control animals compared to BALBc control and B6Tyr.BALB-

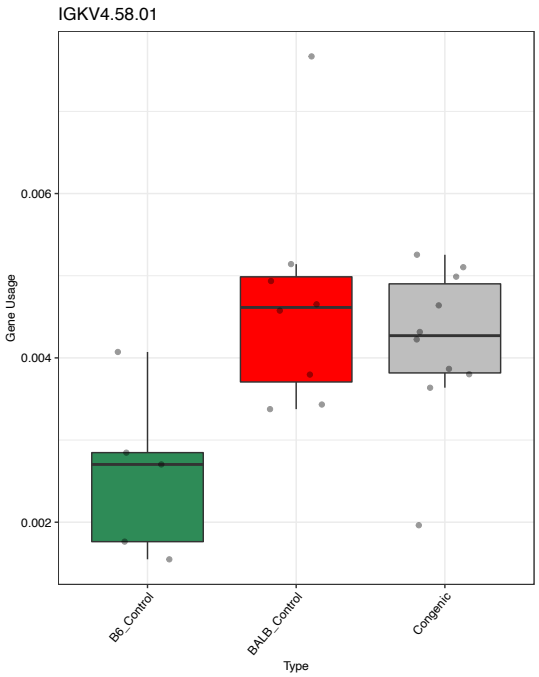
Table 4.6. IGKV genes with significant usage frequency differences between $B6^{Tyr}$.BALB-IGH and B6.

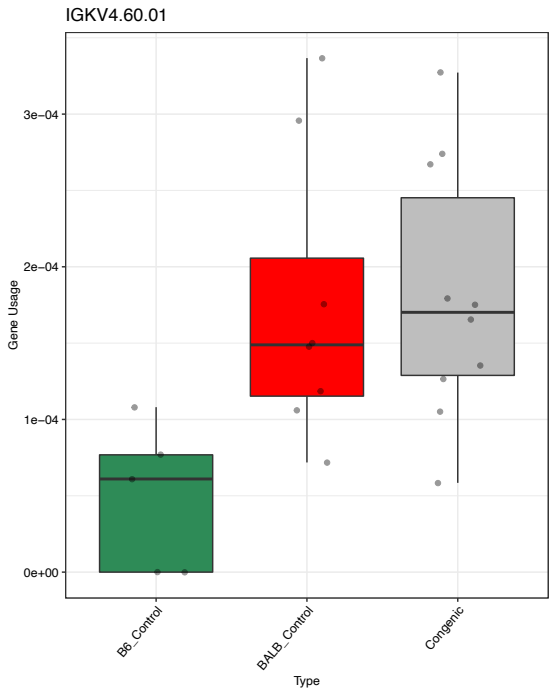
P-values calculated using one-way ANOVA and FDR correction for multiple comparisons performed using Benjamini, Hochberg, and Yekutieli method[143,144].

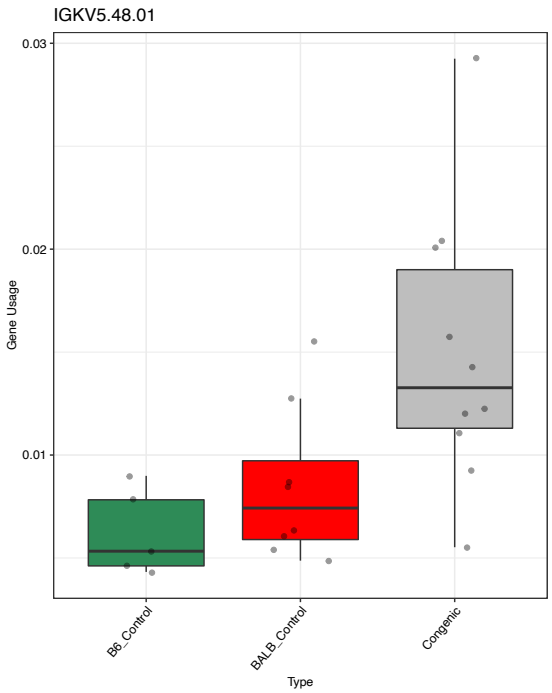
Gene	P-Value	FDR-Corrected P-Value
IGKV1.110.01	0.00173364	0.04403439
IGKV1.99.01	0.00843187	0.17847453
IGKV11.106.02	0.03890488	0.41370267
IGKV3.7.01	0.04309839	0.41370267
IGKV3.7.02	0.02595984	0.36632217
IGKV4.58.01	0.03410535	0.41370267
IGKV4.60.01	0.02083779	0.36632217
IGKV5.48.01	0.02329526	0.36632217
IGKV6.14.01	0.0472502	0.41370267
IGKV6.15.01	0.0435350	0.41370267


IGKV1.110.01
IGKV1.99.01
IGKV11.106.02
IGKV3.7.01
Gene Usage
Type
B6_Control
BALB_Control
Congenic


IGKV3.7.02
Gene Usage
0.006
0.004
0.002
0.000
B6_Control
BALB_Control
Congenic
Type


IGKV4.58.01
Gene Usage
0.006
0.004
0.002
B6_Control
BALB_Control
Congenic
Type


IGKV4.60.01
Gene Usage
3e−04
2e−04
1e−04
0e+00
B6_Control
BALB_Control
Congenic
Type


IGKV5.48.01
Gene Usage
0.03
0.02
0.01
B6_Control
BALB_Control
Congenic
Type

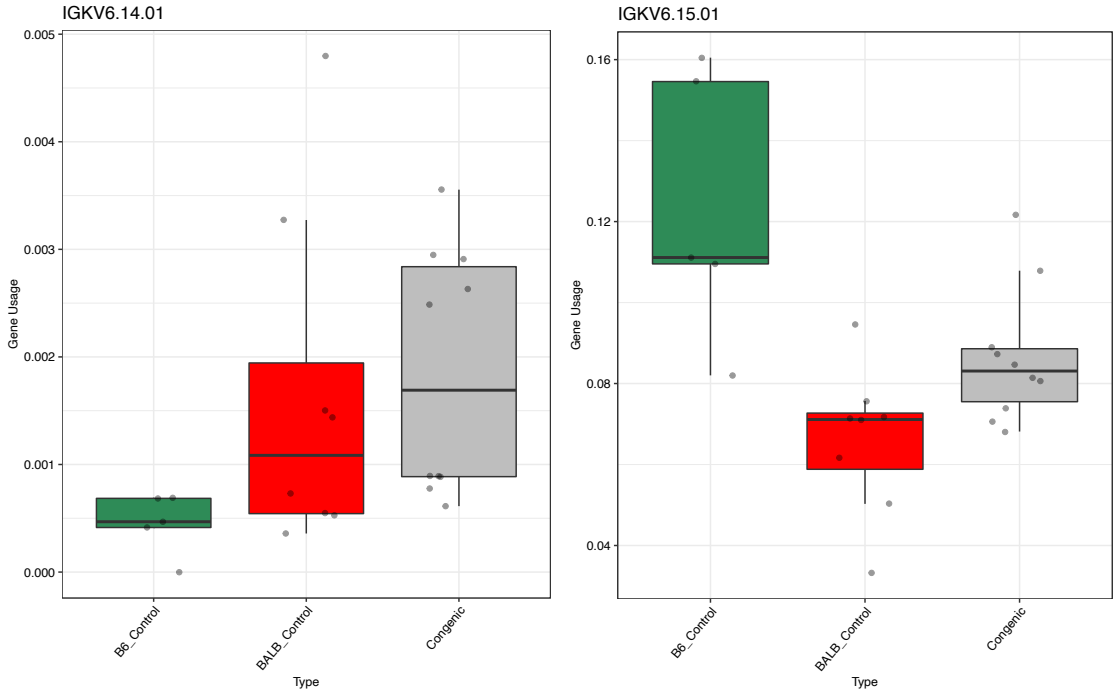


Figure 4.6. IGKV genes with significant usage frequency differences between B6^{Tyr}.BALB-IGH (grey) and B6 (green).

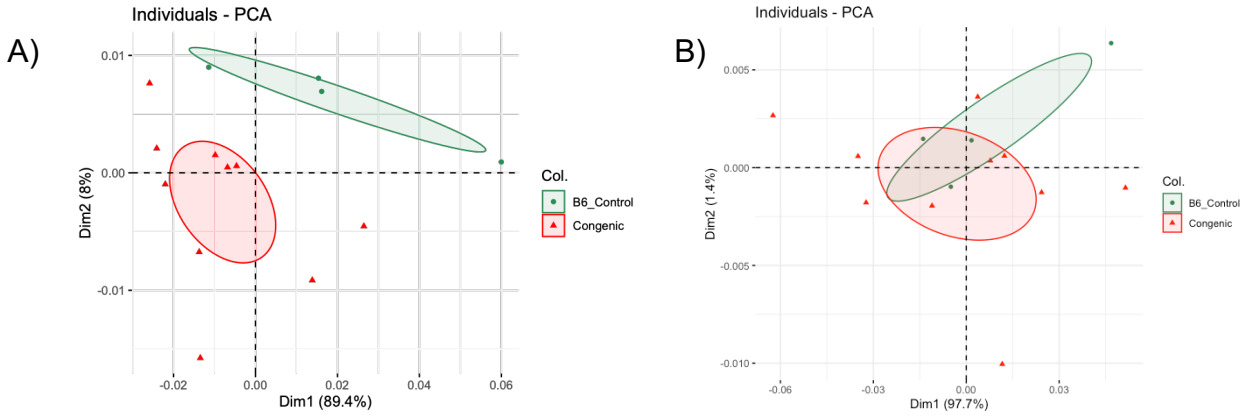


Figure 4.7. PCA analysis of the 10 IGKV genes with significant usage frequency differences between B6^{Tyr}.BALB-IGH congenic line and B6 control animals.

A) Productive sequences B) Unproductive sequences.

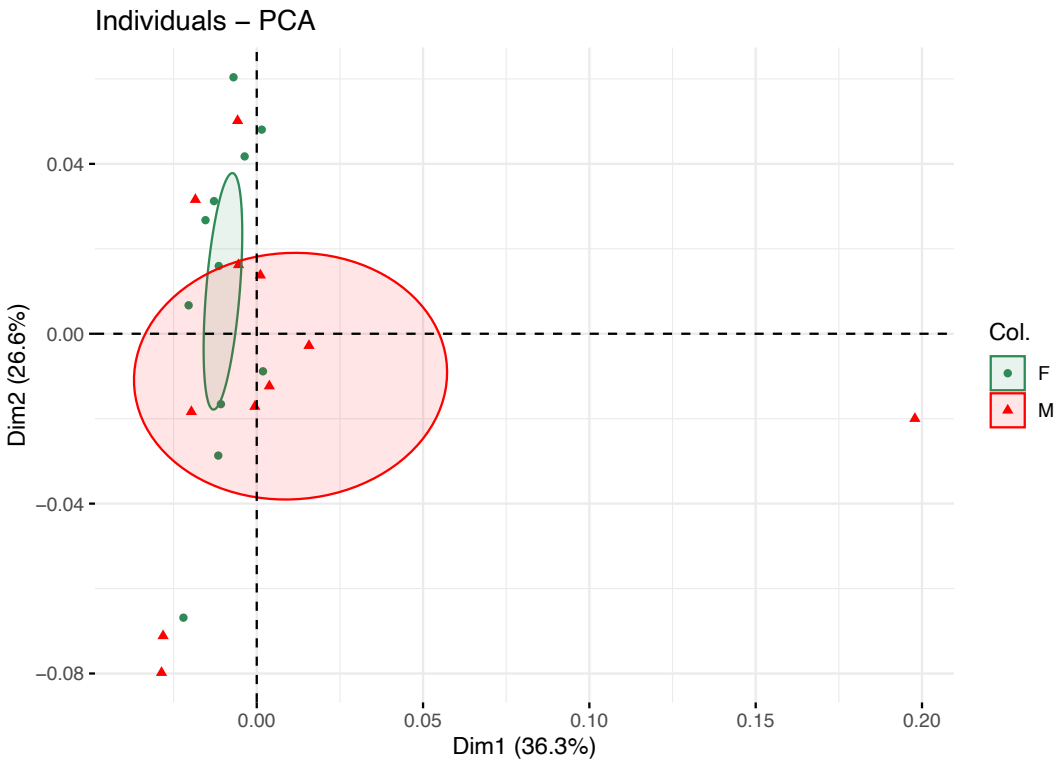


Figure 4.8. PCA clustering on productive IGKV sequences according to sex of the animal.

The sex of the animal does not appear to be responsible for the PCA clustering observed in the congenic animals.

IGH congenic animals. Removal of these sequences did not affect PCA clustering. Thus, we concluded that IGKV8-21 and IGKV8-30 were not responsible for influencing the observed clustering according to IGH haplotype.

We also examined the effect that a divergent IGH locus has on the germline IGL repertoire. Since IGKV germline genes account for 95% or more of the mouse germline light chain repertoire[26], we hypothesized that the haplotype effect observed for IGKV germline genes would not be as significant for IGLV genes as it was for IGKV genes. To test this hypothesis, we performed another one-way ANOVA and PCA analysis, but this time used IGLV genes with P-values < 0.05. IGLV1 and IGLV2 were significantly different between the B6 control animals and B6^{Tyr}.BALB-IGH congenic animals (Figure 4.9, Table 4.7). When plotted in a PCA, B6 and B6^{Tyr}.BALB-IGH formed distinct, non-overlapping clusters (Figure 4.10).

EFFECTS OF BALBc BACKGROUND GENOME ON LIGHT CHAIN AB REPERTOIRE – B6^{Tyr}.BALB-IGH VS. BALBc

To determine if the genetic background of B6 vs BALBc had an impact on IG gene usage I compared BALBc to B6^{Tyr}.BALB-IGH. These lines have the same IGH and IGK locus, thus difference in usage would suggest other regions in the genome are having an impact. Besides the comparison between B6^{Tyr}.BALB-IGH and B6, we also performed a one-way ANOVA on IGKV usage frequencies that compared BALBc to B6^{Tyr}.BALB-IGH. The rationale for this comparison was to see if there were any significant genes in this comparison that were also significant in

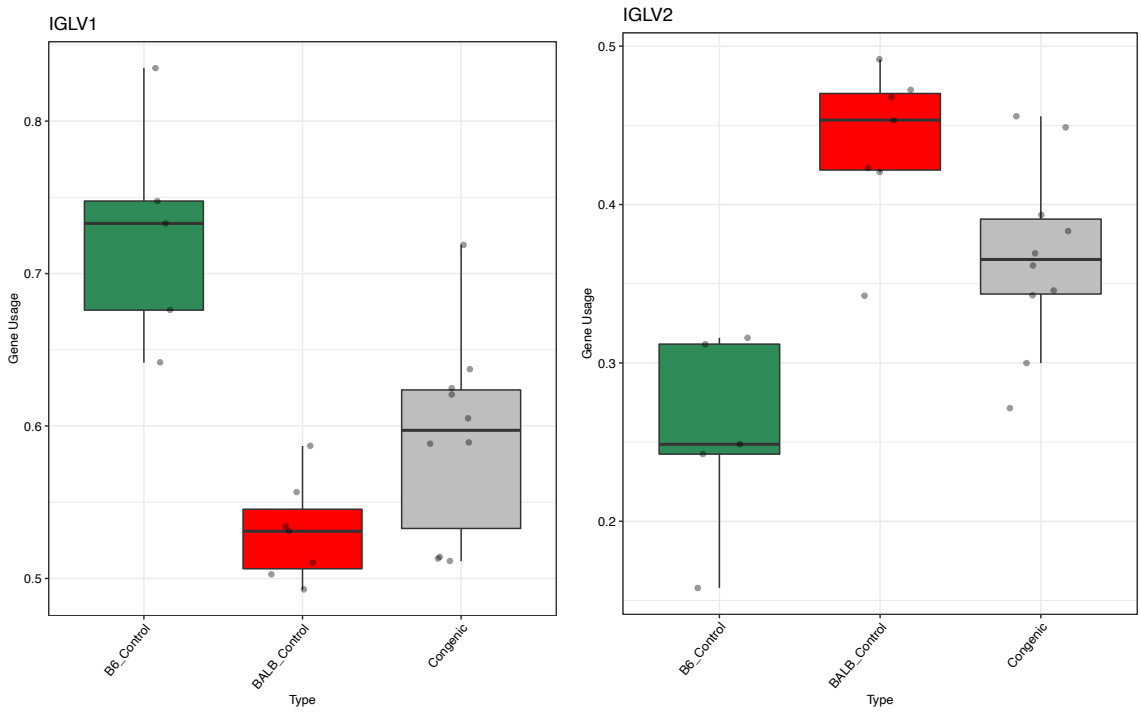


Figure 4.9. IGLV genes with significant usage frequency differences between B6^{Tyr}.BALB-IGH (grey) and B6 (green).

Table 4.7. IGLV genes with significant usage frequency differences between $B6^{Tyr}$.BALB-IGH and B6.

P-values calculated using one-way ANOVA and FDR correction for multiple comparisons performed using Benjamini, Hochberg, and Yekutieli method[143,144].

Gene	P-Value	FDR-Corrected P-Value
IGLV1	0.00335286	0.007074139
IGLV2	0.00471609	0.007074139
IGLV3	0.24478423	0.244784226

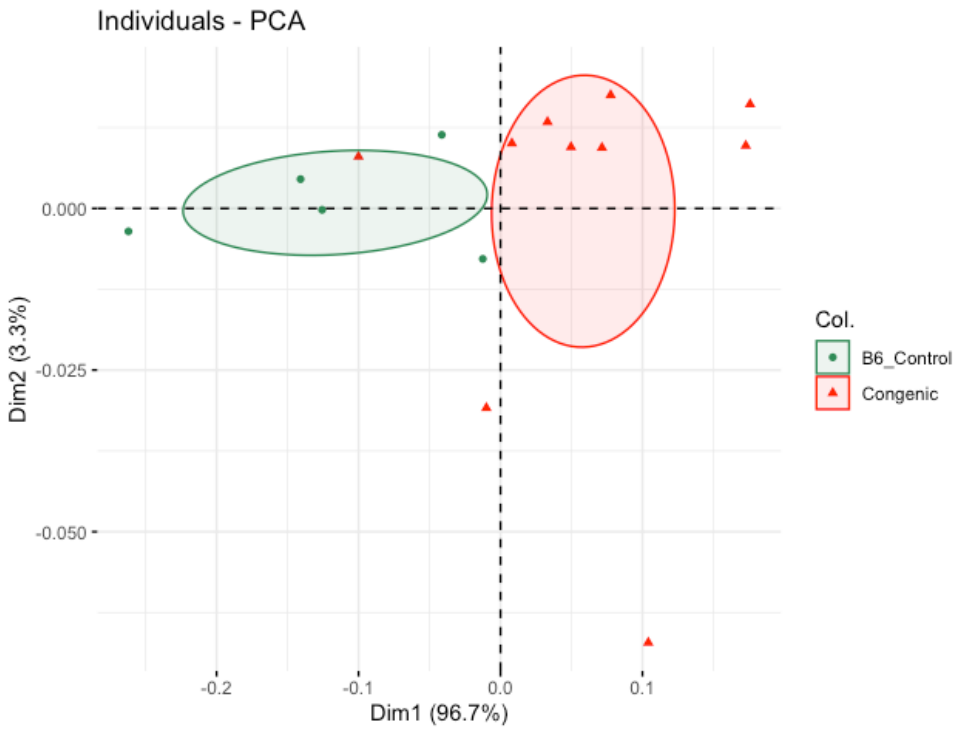


Figure 4.10. PCA analysis of IGLV1 and IGLV2 usage between B6^{Tyr}.BALB-IGH and B6 (productive sequences).

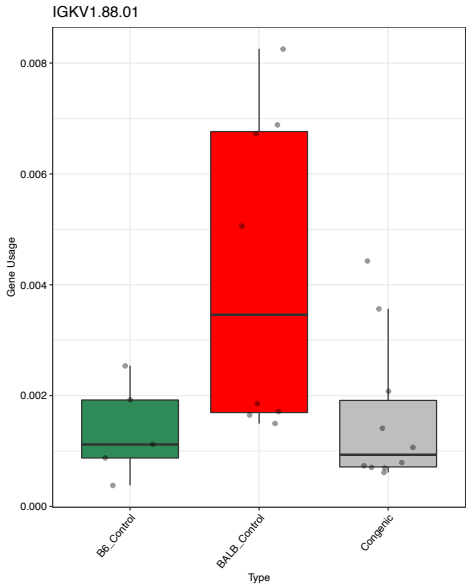
the B6^{Tyr}.BALB-IGH congenic vs. B6 control animals comparison. The new comparison accounted for the difference in the lines' background genomes. The B6^{Tyr}.BALB-IGH congenic line is genetically similar to the B6 control line, except for the IGH locus; its background genome is B6. The BALBc control line, however, contains a BALBc background genome, but has the same IG loci haplotypes as the B6^{Tyr}.BALB-IGH congenic line. Therefore, it was expected that would produce similar naïve antibodies. 14 IGKV genes had P-values < 0.05 (Table 4.8, Figure 4.11). A PCA analysis of the 14 significant genes showed distinct clusters formed by BALBc and B6^{Tyr}.BALB-IGH (Figure 4.12). Interestingly, 2 of the 14 significant IGKV genes for this group, IGKV5.48.01 and IGKV6.15.01, were also identified as significant when comparing B6^{Tyr}.BALB-IGH to B6.

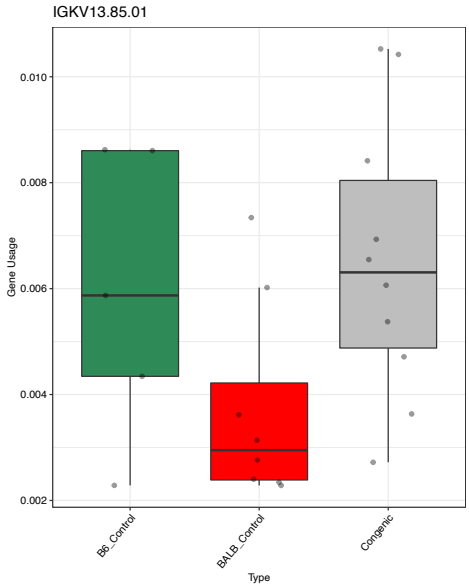
Like the B6^{Tyr}.BALB-IGH and B6 comparison, only IGLV1 and IGLV2 had significantly different usage frequencies. The PCA for the two significant IGLV usage frequencies clustered like the B6^{Tyr}.BALB-IGH congenic and B6 control comparison (Figure 4.13).

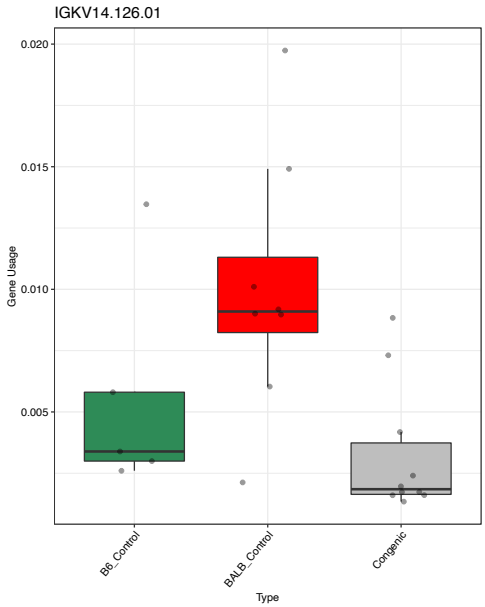
Table 4.8. IGKV genes with significant usage frequency differences between BALBc control line and $B6^{Tyr}$.BALB-IGH congenic animals.

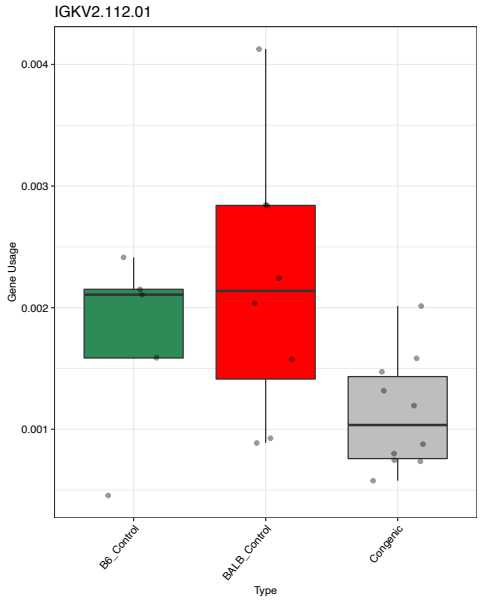
P-values calculated using one-way ANOVA and FDR correction for multiple comparisons performed using Benjamini, Hochberg, and Yekutieli method[143,144].

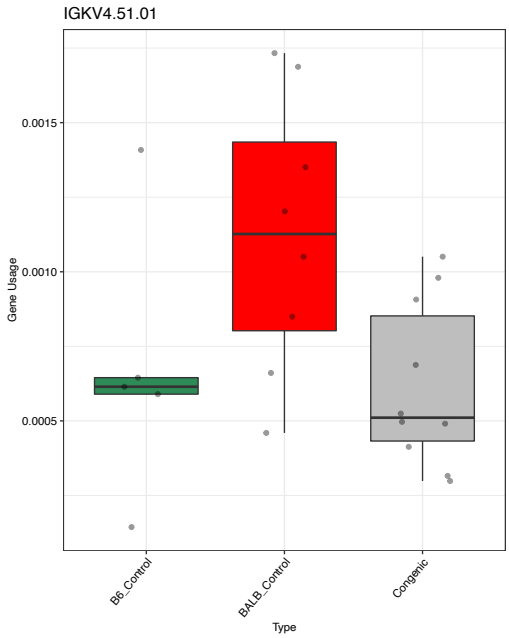
Gene	P-Value	FDR Corrected P-Value
IGKV1.88.01	0.020759226	0.03109692
IGKV13.85.01	0.022971004	0.03109692
IGKV14.126.01	0.002981774	0.01569471
IGKV2.112.01	0.013392019	0.03109692
IGKV4.51.01	0.010201255	0.02856351
IGKV4.53.01	0.022468209	0.03109692
IGKV4.57.01	0.02715749	0.03109692
IGKV4.70.01	0.001666825	0.01569471
IGKV4.92.01	0.032646074	0.03264607
IGKV5.48.01	0.028660305	0.03109692
IGKV6.13.01	0.003363152	0.01569471
IGKV6.15.01	0.025348179	0.03109692
IGKV8.28.02	0.008836773	0.02856351
IGKV9.119.01	0.028875713	0.03109692

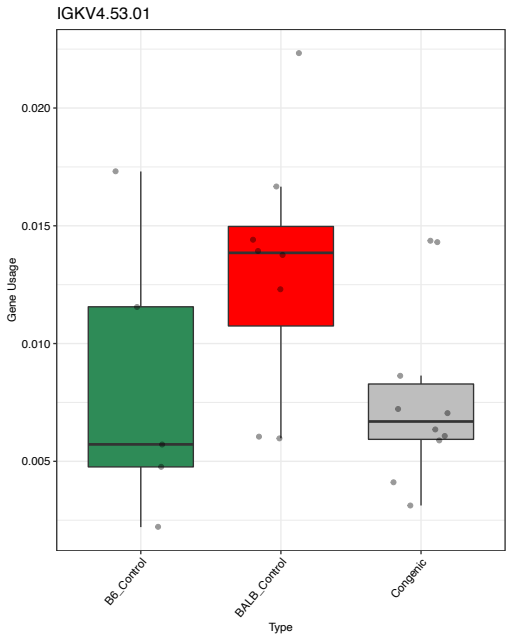

IGKV1.88.01
Gene Usage
0.008
0.006
0.004
0.002
0.000
B6_Control
BALB_Control
Congenic
Type

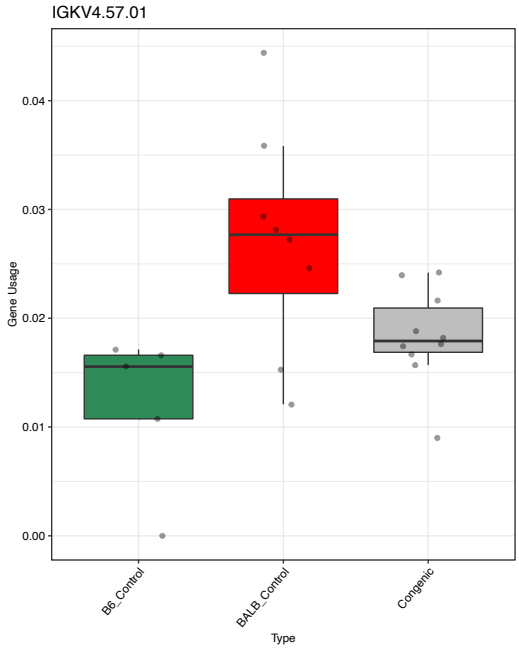

IGKV13.85.01
Gene Usage
0.010
0.008
0.006
0.004
0.002
B6_Control
BALB_Control
Congenic
Type

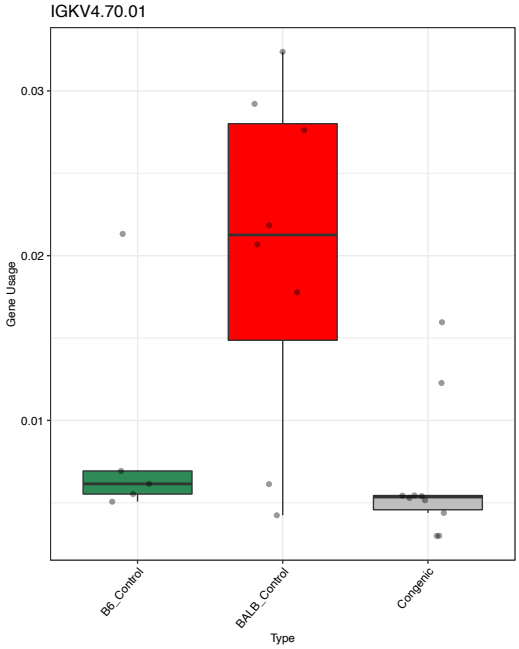

IGKV14.126.01
Gene Usage
0.020
0.015
0.010
0.005
B6_Control
BALB_Control
Congenic
Type

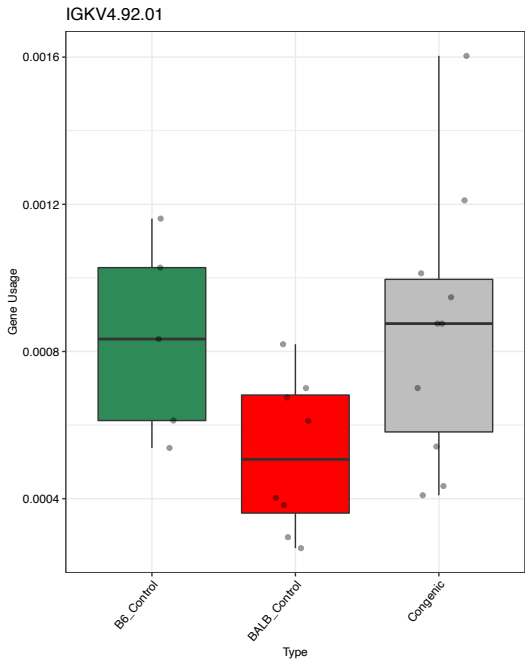

IGKV2.112.01
Gene Usage
0.004
0.003
0.002
0.001
B6_Control
BALB_Control
Congenic
Type

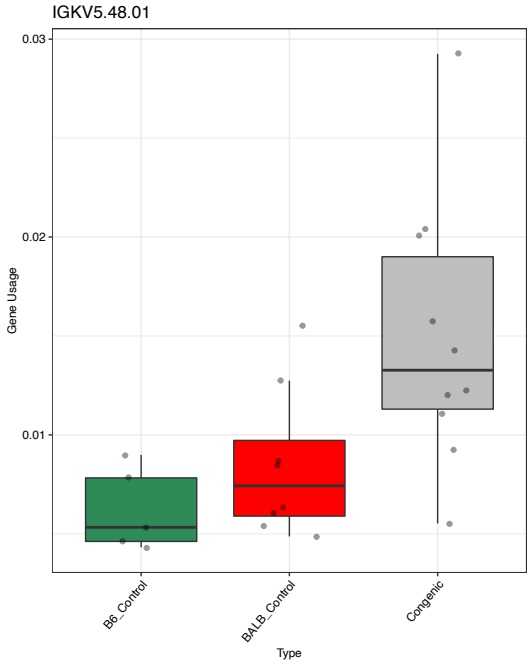

IGKV4.51.01
Gene Usage
0.0015
0.0010
0.0005
B6_Control
BALB_Control
Congenic
Type

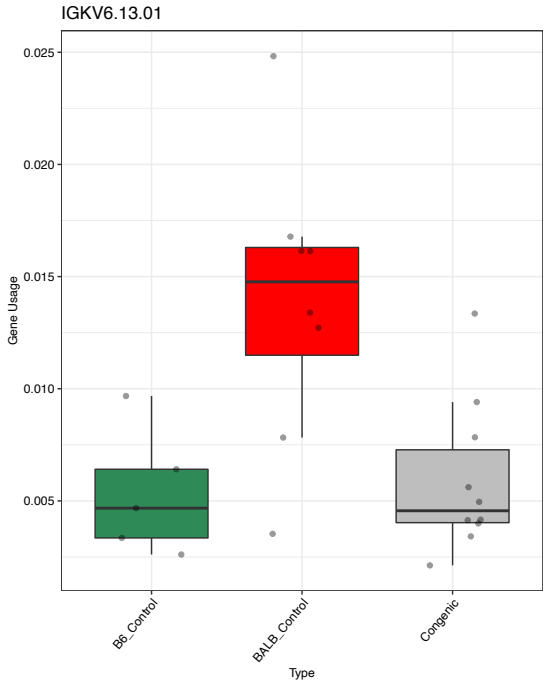

IGKV4.53.01
Gene Usage
0.020
0.015
0.010
0.005
B6_Control
BALB_Control
Congenic
Type

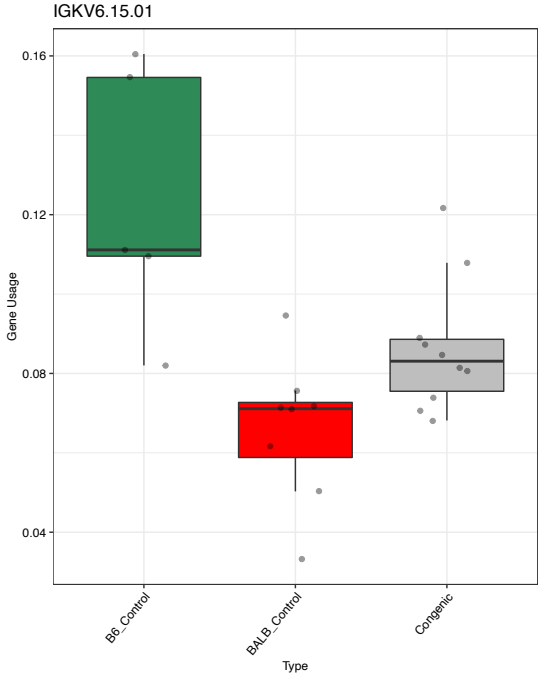

IGKV4.57.01
Gene Usage
0.04
0.03
0.02
0.01
0.00
B6_Control
BALB_Control
Congenic
Type


IGKV4.70.01
Gene Usage
0.03
0.02
0.01
B6_Control
BALB_Control
Congenic
Type


IGKV4.92.01
0.0016
0.0012
0.0008
0.0004
Gene Usage
B6_Control
BALB_Control
Congenic
Type


IGKV5.48.01
0.03
0.02
0.01
Gene Usage
B6_Control
BALB_Control
Congenic
Type


IGKV6.13.01
0.025
0.020
0.015
0.010
0.005
Gene Usage
B6_Control
BALB_Control
Congenic
Type


IGKV6.15.01
0.16
0.12
0.08
0.04
Gene Usage
B6_Control
BALB_Control
Congenic
Type

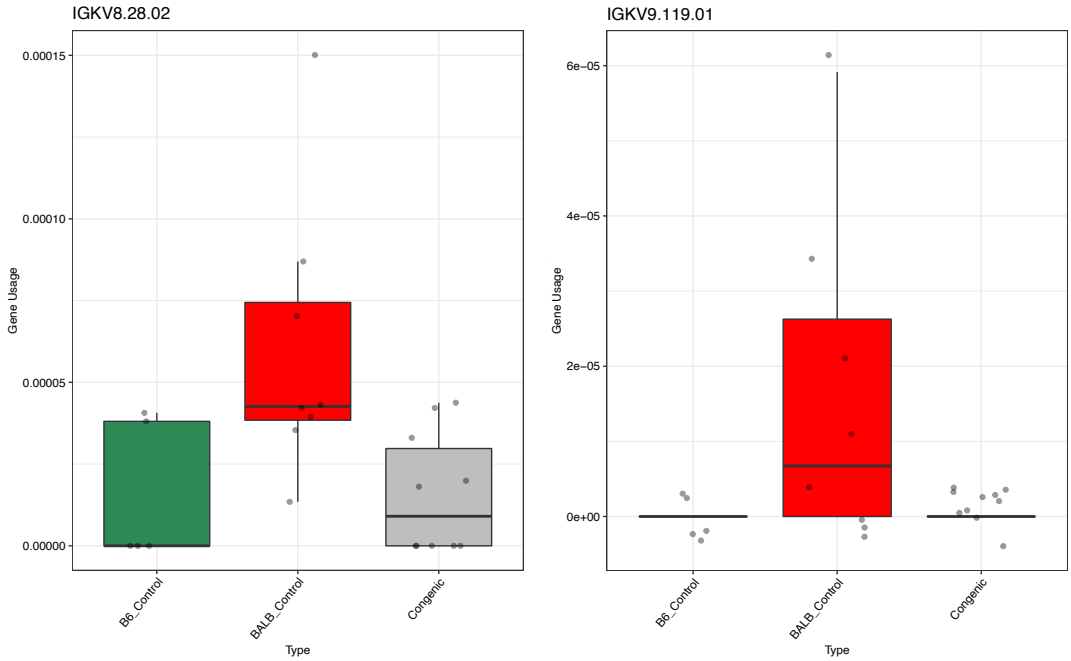


Figure 4.11. IGKV genes with significant usage frequency differences between B6^{Tyr}.BALB-IGH (grey) and BALBc (red).

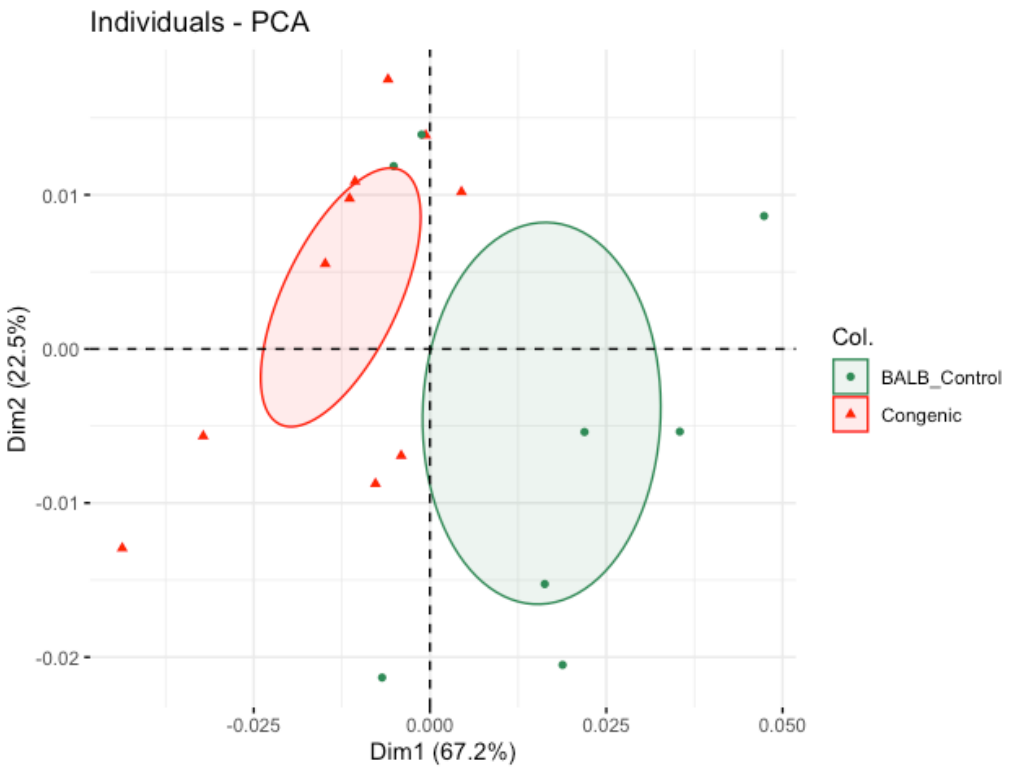


Figure 4.12. PCA analysis of the 14 IGKV genes with significant usage frequency differences between B6^{Tyr}.BALB-IGH and BALBc (productive sequences).

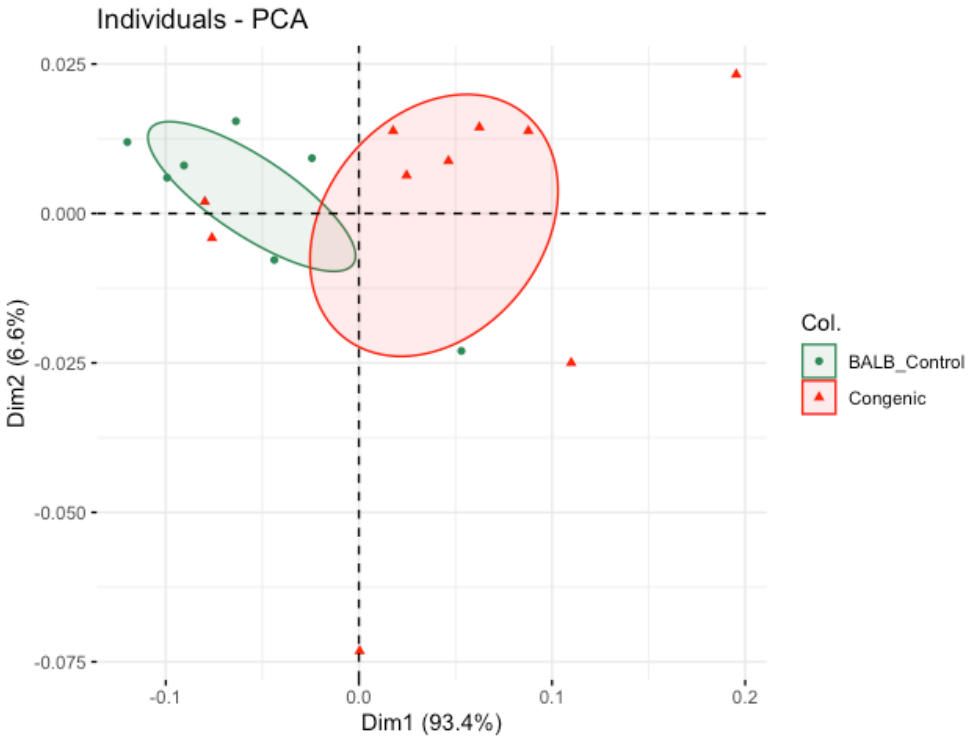


Figure 4.13. PCA analysis of IGLV1 and IGLV2 usage frequency differences between B6^{Tyr}.BALB-IGH and BALBc (productive sequences).

P-value < 0.05.

DISCUSSION

The two congenic mouse strains created in this experiment will serve as valuable models for future experiments investigating the effect of haplotype diversity on the Ab repertoire. They are the first model organisms with controlled IG loci haplotype variation on a B6 background. The next experiment will be a similar bulk AIRR-seq on the $B6^{Tyr}$.NOD-IGK congenic line once sufficient numbers are available. An additional future experiment for both congenic lines will be to decrease the size of the introgressed regions to contain just the IGH or IGK locus through additional crossing. Currently, the BALBc congenic insert is approximately 9 Mb, and the size of the IGH locus is approximately 2.8 Mb. The NOD congenic insert, however, is significantly larger at 39 Mb. From the gene ontology analysis for the $B6^{Tyr}$.NOD-IGK congenic line, we know that the congenic insert contains 398 genes that are homozygous for NOD/ShiLtJ. Even though the gene ontology analysis did not predict the non IGH BALBc genes and NOD genes to interact with the immune system, it is possible that the extra genetic material could disrupt gene expression through changes in long-range chromosomal conformation.

Chromatin conformation is an important mechanism that regulates gene expression and recombination[145]. During V(D)J rearrangement, over 100 functional V_H genes across the 2.8 Mb mouse IGH locus and 3.2 Mb IGK locus must recombine with either a rearranged DJ_H element or a J_K gene. To accomplish this task, the IGH and IGK locus in pro-B cells contracts to place the distal V_H and V_K genes next to proximal D_H and J_K segments, which facilitate V(D)J

recombination. One way that the congenic insert could affect long-range chromosomal interactions is through disruption of topologically associated domains (TADs).

TADs are large, megabase-sized local chromatin interaction domains that are stable across different cell types and highly conserved across species[146]. CCCTC-binding factor (CTCF) sites at sub-TAD boundaries are also involved in TAD formation and facilitate long-range chromatin looping through interaction with cohesin[147–149]. A TAD is formed when a pair of CTCF-cohesin dimers move in opposite directions along DNA, extruding a DNA loop behind them until they reach a convergent pair of CTCF binding sites[150,151]. Studies have shown that the mouse IGH locus is within a 2.9 Mb TAD in pro-B cells, and that this TAD can be divided into sub-TADs, A, B, and C. sub-TADs A and C correspond to the D_H-proximal and D_H-distal V_H gene families, and sub-TAD B contains intermediate V_H gene segments[123]. The IGK TAD structure is not identical to IGH. The IGK locus is organized into six TADs, commonly referred to as loops 1-6 (L1-L6)[152]. Studies on mouse IGH TAD structure have helped reveal regulatory elements involved in forming this architecture of the mouse IGH locus during V(D)J recombination[123,153,154]. The B-cell commitment factor Pax5 is one regulatory element that has been shown to facilitate distal V_H-DJ_H rearrangements by binding to PAIR elements throughout the loci[155,156], which induces long-range chromatin looping[146].

TAD domain regulatory sites across the C57BL/6 IGH locus have shed light on gene usage patterns, such as why some V gene families have greater usage

frequencies than others[157]. Within the IGK locus, there is a preference for transcribing VK genes that are proximate to, and oriented away from, CTCF-bound sites[150,152]. While these experiments have provided insights about basic rules regarding the mechanics of V(D)J rearrangement, they are still limited to C57BL/6 and provide no insight into how the germline variation in our congenic lines could affect this process.

An additional future experiment on the congenic lines could be the use of chromosome conformation capture (3C). 5C, a high-throughput derivative of 3C that involves selective amplification of chromatin interactions captured by proximity ligation, can examine chromatin interactions at the molecular level in each cell population and define chromatin structure at the level of single restriction fragments. Combined, these techniques can define long-range chromatin looping to reveal large-scale changes in chromatin organization that result from haplotype variation within the IG loci. We have observed that IG locus haplotype variation effects the Ab repertoire. In addition, the congenic lines have shown that haplotype variation restricted to just the IGH locus affects gene usage frequencies of the light chain repertoire. Given the complexity of V(D)J recombination and the IG loci, it is likely that the haplotype variation within the congenic lines causes significant changes in locus topology. Significant structural variation and polymorphisms in the NOD IGK and BALBc IGH assemblies are visible in sequence-level comparisons to B6. This variation can lead to changes in the loci's key regulatory elements such as CTCF sites, PAIR elements, and enhancers, which can cause changes in IG gene usage. To better understand the mechanisms by which

haplotype variation within the IGH locus causes changes in the light chain repertoire, future experiments can examine the chromosomal topological changes that occur during V(D)J recombination.

CHAPTER 5 : FUTURE DIRECTIONS

This dissertation research was born out of an interest sparked from a 2015 study that examined the mouse Ab heavy chain repertoire in C57BL/6 and BALB/c mice and a recently published study on the germline IGHV repertoires between F1 C57BL/6 x BALB/c mice. Those studies demonstrated significant germline IGHV repertoire divergence between the C57BL/6 and BALB/c mouse strains[45,87]. The striking conclusion from Collins et al. was the observation of only 5 germline IGHV sequences shared between the two mouse strains. Prior to that experiment, divergent IGH and IGK haplotypes had been demonstrated using RFLP and large SNP panels[28,54,56,57,158], but none had comprehensively cataloged IGHV differences between the two strains in a high-throughput manner. Even today, the mouse IMGT germline gene database is largely comprised of C57BL/6 sequences, and the only full haplotype IG resource that exists is for C57BL/6 mice. This prevents us from asking questions related to the effect of genomic diversity on the Ab repertoire. Fortunately, with new germline gene database resources like OGRDB from the AIRR Community, updating and sharing germline gene databases on a strain-by-strain basis will be possible.

The goal of Chapter 2 was to profile the germline light chain Ab repertoires since the loci are structurally similar to the IGH locus yet largely unexplored. The findings revealed that the germline IGKV repertoires of the biomedically important

mouse strains are largely shared by strains within the same IGK haplotype group. For example, the IGK haplotype of C57BL/6 contains 10 other strains, including BALB/cByJ. Two additional haplotype groups not present in the IMGT database were identified in our dataset, with one haplotype containing MRL/MpJ, NOR/LtJ, and NOD/ShiLtJ. Future experiments related to this work could increase germline gene databases by inferring IG genes from additional strains to account for additional IGK haplotypes. In addition, the study could also be expanded to include the inference of IGHV genes across multiple IGH haplotypes. This would provide greater resolution to the germline gene differences across haplotype groups, but also update existing gene databases with new germline genes that can be used for AIRR-seq analysis.

Although germline gene inferences from AIRR-seq are valuable resources that shed light on IG haplotype differences across strains, they do not replace the locus-wide resolution of coding and noncoding elements that is possible with new IG reference assemblies. The BALB/cByJ-IGH and NOD/ShiLtJ-IGK reference assemblies from Chapter 3 are new IG reference assemblies for divergent IGH and IGK haplotypes. Large-scale comparisons of the new assemblies to the existing C57BL/6 IGH and IGK references show significant structural differences between the haplotypes. While these assemblies have shown large structural differences visible by dotplot comparisons, the assemblies are still incomplete and must be further refined. Currently, the BALB/cByJ-IGH assembly exists as five separate contigs that vary in size, while the NOD/ShiLtJ-IGK assembly is composed of two separate contigs. Because we cannot simply align our

assemblies to C57BL/6 to get an estimate of the size of the gaps, we will use Bionano optical mapping on gDNA samples for both strains to determine the correct size of the gaps in the current assembly. Based on the results from the Bionano optical mapping we will assess the best way to close the remaining gaps in the assembly, either using additional BAC clones or PacBio HiFi Whole Genome Sequencing. Given the rapid advancement in whole genome sequencing using the PacBio platform, it will be possible to sequence and assemble the IG loci of other strains using HiFi Whole Genome Sequencing to build additional reference assemblies that account for additional haplotype variation.

The key goal in constructing the congenic animals was to create a resource so we could determine if the underlying IGH locus had an impact on light chain Ab gene usage. the animals so that the only genetic differences between the congenic animal and C57BL/6 control mouse was the IG loci; the congenics would have either a divergent IGH or IGK locus on a C57BL/6 background. While we were successful in reducing the B6^{Tyr}.BALB-IGH congenic line to 9 Mb of introgressed BALBc genome, the B6^{Tyr}.NOD-IGK congenic line still has a much larger 39 Mb NOD IGK region. The 39 Mb region is predicted to contain at least 298 different genes, and while GO analysis of these genes did not reveal interactions with the immune system, we are not able to rule out the effect that these genes could have on the Ab repertoire given that the IG loci can participate in long-range chromatin interactions. In our B6^{Tyr}.BALB-IGH mice, we observed significant changes in IGKV gene usage frequency compared to B6. This suggests that the composition of the IGH locus impacts the possible Ab repertoire. The introgressed BALBc

genome is still approximately 9 Mb, therefore we cannot completely rule out the effect that these other genes and noncoding elements may be having on the germline IGKV repertoire. Once the B6^{Tyr}.NOD-IGK congenic insert is refined to a smaller size, the bulk AIRR-seq will be performed on these animals to assess if there are IGK haplotype effects on the germline IGHV repertoire. In addition, gene usage may have also been different since the cells used for bulk AIRR-seq were B cells from the spleen that had already undergone selection. A future experiment could investigate gene usage changes that take place during the different stages of B cell development.

Since our bulk AIRR-seq experiment does not allow for heavy and light chain pairing information, a significant future experiment is single-cell sequencing. This is important because it will allow for us to determine which heavy chains are pairing with a given light chain, and determine whether IGKV and IGHV genomic/allelic variation leads to IGHV/IGKV repertoire shifts. This is important because an Ab uses both its heavy chain and light chain variable regions to contact antigen. Therefore, when a studies identifies a particular IGHV allele that strongly neutralize an antigen, the light chain V allele must also be considered. Heavy and light chain pairing combinations can be observed by using the 10X Genomics Chromium Single Cell VDJ Sequencing platform, which will preserve the heavy and light chain pairings of the Ab since interactions between heavy and light chains have been shown to affect the binding kinetics of peptides[128]. We expect that these experiments will establish heavy and light chain pairing trends for the congenic lines, and if there are any promiscuous light chain genes that pair with multiple

heavy chain genes. An alternative approach to the 10X platform could be the droplet-based single-cell mRNA capture developed by Georgiou and colleagues[159].

In conclusion, this study has overcome key obstacles to allow for a greater understanding of the effect of haplotype variation on the Ab repertoire. First, this work has overcome the lack of IG reference assemblies for alternate IG haplotypes by generating new IG reference assemblies for divergent IG loci by sequencing, assembling, and annotating BAC clones for the BALB/cByJ IGH locus and NOD/ShiLtJ IGK locus. Another challenge to understanding the effect of haplotype variation on the Ab repertoire is the incomplete germline gene databases, which can directly impact the way we use animals to study the immune response. We overcame this challenge by inferring novel IGKV and IGLV germline genes for 18 biomedically important mouse strains. Lastly, a significant obstacle in the field of IG genetics is the lack of animal models to study the genetics contributing to observed repertoire diversity. The congenic mouse lines with divergent IG loci on a C57BL/6 background directly addresses this problem. Though the sample size must be increased in future experiments, preliminary AIRR-seq data from these lines shows that haplotype variation affects the Ab repertoire. Specifically, changing just the haplotype of the IGH locus affects the usage frequency of light chain Ab genes. This can have significant consequences for antigen recognition, as it shows that even if an individual contains a light chain gene capable of recognizing a specific antigen, the genetics of the heavy chain must also be considered.

REFERENCES

1. Kenneth, M. & Casey, W. *Janeway's immunobiology*. (Garland science, 2016).

2. Abbas, A. K., Lichtman, A. H. & Pillai, S. Cellular and Molecular Immunology (Sixth Edition). v (2010) doi:10.1016/b978-1-4160-3123-9.50003-6.

3. Avnir, Y. *et al.* IGHV1-69 polymorphism modulates anti-influenza antibody repertoires, correlates with IGHV utilization shifts and varies by ethnicity. *Scientific Reports* 6, 20842 (2016).

4. Watson, C. T. *et al.* A comparison of immunoglobulin IGHV, IGHD and IGHJ genes in wild-derived and classical inbred mouse strains. *Immunol Cell Biol* (2019) doi:10.1111/imcb.12288.

5. Collins, A. M. & Watson, C. T. Immunoglobulin Light Chain Gene Rearrangements, Receptor Editing and the Development of a Self-Tolerant Antibody Repertoire. *Frontiers in Immunology* 9, 2249 (2018).

6. Herold, E. M. *et al.* Determinants of the assembly and function of antibody variable domains. *Sci Rep-uk* 7, 12276 (2017).

7. Lefranc, M.-P. IMGT, the international ImMunoGeneTics database. *Nucleic Acids Res* 29, 207–209 (2001).

8. Aoki-Ota, M., Torkamani, A., Ota, T., Schork, N. & Nemazee, D. Skewed Primary Igκ Repertoire and V–J Joining in C57BL/6 Mice: Implications for Recombination Accessibility and Receptor Editing. *J Immunol* 188, 2305–2315 (2012).

9. Brekke, K. M. & Garrard, W. T. Assembly and analysis of the mouse immunoglobulin kappa gene sequence. *Immunogenetics* 56, 490–505 (2004).

10. 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. *The American Journal of Human Genetics* 92, 530–546 (2013).

11. Watson, C. T. & Breden, F. The immunoglobulin heavy chain locus: genetic variation, missing data, and implications for human disease. *Genes Immun* 13, 363 (2012).

12. Schroeder, H. W. & Cavacini, L. Structure and function of immunoglobulins. *J Allergy Clin Immun* 125, S41–S52 (2010).

13. Hood, L., Campbell, J. H. & Elgin, S. C. R. The Organization, Expression, and Evolution of Antibody Genes and Other Multigene Families. *Annu Rev Genet* 9, 305–353 (1975).

14. 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).

15. Kofler, R., Geley, S., Kofler, H. & Helmberg, A. Mouse Variable-Region Gene Families: Complexity, Polymorphism and Use in non-Autoimmune Responses. *Immunol Rev* 128, 5–21 (1992).

16. Johnston, C. M., Wood, A. L., Bolland, D. J. & Corcoran, A. E. Complete Sequence Assembly and Characterization of the C57BL/6 Mouse Ig Heavy Chain V Region. *J Immunol* 176, 4221–4234 (2006).

17. Brodeur, P. H. & Riblet, R. The immunoglobulin heavy chain variable region (Igh-V) locus in the mouse. I. One hundred Igh-V genes comprise seven families of homologous genes. *Eur J Immunol* 14, 922–930 (1984).

18. Winter, E., Radbruch, A. & Krawinkel, U. Members of novel VH gene families are found in VDJ regions of polyclonally activated B-lymphocytes. *Embo J* 4, 2861–2867 (1985).

19. Brodeur, P. H., Osman, G. E., Mackle, J. J. & Lalor, T. M. The organization of the mouse Igh-V locus. Dispersion, interspersion, and the evolution of VH gene family clusters. *J Exp Medicine* 168, 2261–2278 (1988).

20. Rathbun, G. A., Otani, F., Milner, E. C. B., Capra, J. D. & Tucker, P. W. Molecular characterization of the AJ J558 family of heavy chain variable region gene segments. *J Mol Biol* 202, 383–395 (1988).

21. Honjo, T. & Matsuda, F. Immunoglobulin Genes (Second Edition). *Part Ii Organization Rearrange Immunoglobulin Genes* 145–171 (1995) doi:10.1016/b978-012053640-5/50010-0.

22. Strohal, R., Helmberg, A., Kroemer, G. & Kofler, R. MouseVk gene classification by nucleic acid sequence similarity. *Immunogenetics* 30, 475–493 (1989).

23. Valiante, N. M. & Caton, A. J. A new Igk-V gene family in the mouse. *Immunogenetics* 32, 345–350 (1990).

24. Shefner, R. *et al.* Identification of a new V kappa gene family that is highly expressed in hybridomas from an autoimmune mouse strain. *J Immunol Baltim Md 1950* 145, 1609–14 (1990).

25. D'Hoostelaere, L. A. & Klinman, D. Characterization of new mouse V kappa groups. *J Immunol Baltim Md 1950* 145, 2706–12 (1990).

26. Almagro, J. C., Hernández, I., Ramírez, M. C. & Vargas-Madrazo, E. Structural differences between the repertoires of mouse and human germline genes and their evolutionary implications. *Immunogenetics* 47, 355–363 (1998).

27. Kirschbaum, T. *et al.* The central part of the mouse immunoglobulin κ locus. *Eur J Immunol* 29, 2057–2064 (1999).

28. D'Hoostelaere, L. A., Huppi, K., Mock, B., Mallett, C. & Potter, M. The Ig kappa L chain allelic groups among the Ig kappa haplotypes and Ig kappa crossover populations suggest a gene order. *J Immunol Baltim Md 1950* 141, 652–61 (1988).

29. Jung, D., Giallourakis, C., Mostoslavsky, R. & Alt, F. W. MECHANISM AND CONTROL OF V(D)J RECOMBINATION AT THE IMMUNOGLOBULIN HEAVY CHAIN LOCUS. *Annu Rev Immunol* 24, 541–570 (2006).

30. Schatz, D. G. & Ji, Y. Recombination centres and the orchestration of V(D)J recombination. *Nat Rev Immunol* 11, 251–263 (2011).

31. Schatz, D. G. & Swanson, P. C. V(D)J Recombination: Mechanisms of Initiation. *Annu Rev Genet* 45, 167–202 (2011).

32. Collins, A. M. & Jackson, K. J. L. On being the right size: antibody repertoire formation in the mouse and human. *Immunogenetics* 70, 143–158 (2018).

33. Meek, K. Analysis of junctional diversity during B lymphocyte development. *Science* 250, 820–823 (1990).

34. Alamyar, E., Duroux, P., Lefranc, M.-P. & Giudicelli, V. Immunogenetics, Methods and Applications in Clinical Practice. *Methods Mol Biology* 882, 569–604 (2012).

35. Feeney, A. J. V(D)J Recombination. *Adv Exp Med Biol* 73–81 (2009) doi:10.1007/978-1-4419-0296-2_6.

36. Álvarez-Prado, Á. F. *et al.* A broad atlas of somatic hypermutation allows prediction of activation-induced deaminase targets. *J Exp Med* 215, jem.20171738 (2018).

37. Rajewsky, K., Forster, I. & Cumano, A. Evolutionary and somatic selection of the antibody repertoire in the mouse. *Science* 238, 1088–1094 (1987).

38. Tonegawa, S. Somatic generation of antibody diversity. *Nature* 302, 575–581 (1983).

39. Csepregi, L., Ehling, R. A., Wagner, B. & Reddy, S. T. Immune Literacy: Reading, Writing, and Editing Adaptive Immunity. *Iscience* 23, 101519 (2020).

40. Mainville, C. A. *et al.* Deletional mapping of fifteen mouse VH gene families reveals a common organization for three Igh haplotypes. *J Immunol Baltim Md 1950* 156, 1038–46 (1996).

41. Chevillard, C., Ozaki, J., Herring, C. D. & Riblet, R. A Three-Megabase Yeast Artificial Chromosome Contig Spanning the C57BL Mouse Igh Locus. *J Immunol* 168, 5659–5666 (2002).

42. Yang, H. *et al.* Subspecific origin and haplotype diversity in the laboratory mouse. *Nat Genet* 43, 648 (2011).

43. Yang, H., Bell, T. A., Churchill, G. A. & Villena, F. P.-M. de. On the subspecific origin of the laboratory mouse. *Nat Genet* 39, 1100–1107 (2007).

44. Ohlin, M. *et al.* Inferred Allelic Variants of Immunoglobulin Receptor Genes: A System for Their Evaluation, Documentation, and Naming. *Front Immunol* 10, 435 (2019).

45. Jackson, K. J. *et al.* A BALB/c IGHV Reference Set, defined by haplotype analysis of long-read VDJ-C sequences from F1 (BALB/c / C57BL/6) mice. *Biorxiv* 2022.02.28.482396 (2022) doi:10.1101/2022.02.28.482396.

46. Consortium, M. G. S. Initial sequencing and comparative analysis of the mouse genome. *Nature* 420, 520 (2002).

47. Paigen, K. A miracle enough: the power of mice. *Nat Med* 1, 215–220 (1995).

48. Viney, M., Lazarou, L. & Abolins, S. The laboratory mouse and wild immunology. *Parasite Immunol* 37, 267–273 (2015).

49. BERRY, R. J. & SCRIVEN, P. N. The house mouse: a model and motor for evolutionary understanding. *Biol J Linn Soc* 84, 335–347 (2005).

50. Castle, W. E. & Little, C. C. On a Modified Mendelian Ratio Among Yellow Mice. *Science* 32, 868–870 (1910).

51. Din, W. *et al.* Origin and radiation of the house mouse: clues from nuclear genes. *J Evolution Biol* 9, 519–539 (1996).

52. Boursot, P., Auffray, J. C., Britton-Davidian, J. & Bonhomme, F. The Evolution of House Mice. *Annu Rev Ecol Syst* 24, 119–152 (1993).

53. Selander, R. K., Hunt, W. G. & Yang, S. Y. Protein Polymorphism and Genic Heterozygosity in Two European Subspecies of the House Mouse. *Evolution* 23, 379 (1969).

54. Wade, C. M. *et al.* The mosaic structure of variation in the laboratory mouse genome. *Nature* 420, 574–578 (2002).

55. Mural, R. J. *et al.* A Comparison of Whole-Genome Shotgun-Derived Mouse Chromosome 16 and the Human Genome. *Science* 296, 1661–1671 (2002).

56. Wade, C. M. & Daly, M. J. Genetic variation in laboratory mice. *Nat Genet* 37, 1175–1180 (2005).

57. Frazer, K. A. *et al.* A sequence-based variation map of 8.27 million SNPs in inbred mouse strains. *Nature* 448, 1050–1053 (2007).

58. Yaari, G. & Kleinstein, S. H. Practical guidelines for B-cell receptor repertoire sequencing analysis. *Genome Med* 7, 121 (2015).

59. Bernat, N. V. *et al.* High-Quality Library Preparation for NGS-Based Immunoglobulin Germline Gene Inference and Repertoire Expression Analysis. *Front Immunol* 10, 660 (2019).

60. Rhoads, A. & Au, K. F. PacBio Sequencing and Its Applications. *Genom Proteom Bioinform* 13, 278–289 (2015).

61. Schadt, E. E., Turner, S. & Kasarskis, A. A window into third-generation sequencing. *Hum Mol Genet* 19, R227–R240 (2010).

62. Kono, N. & Arakawa, K. Nanopore sequencing: Review of potential applications in functional genomics. *Dev Growth Differ* 61, 316–326 (2019).

63. Gupta, N. T. *et al.* Change-O: a toolkit for analyzing large-scale B cell immunoglobulin repertoire sequencing data. *Bioinformatics* 31, 3356–3358 (2015).

64. Heiden, J. A. V. *et al.* pRESTO: a toolkit for processing high-throughput sequencing raw reads of lymphocyte receptor repertoires. *Bioinformatics* 30, 1930–1932 (2014).

65. Corcoran, M. M. *et al.* Production of individualized V gene databases reveals high levels of immunoglobulin genetic diversity. *Nature Communications* 7, 13642 (2016).

66. Gadala-Maria, D., Yaari, G., Uduman, M. & Kleinstein, S. H. Automated analysis of high-throughput B-cell sequencing data reveals a high frequency of novel immunoglobulin V gene segment alleles. *Proc National Acad Sci* 112, E862–E870 (2015).

67. Ye, J., Ma, N., Madden, T. L. & Ostell, J. M. IgBLAST: an immunoglobulin variable domain sequence analysis tool. *Nucleic Acids Res* 41, W34–W40 (2013).

68. Tutter, A. & Riblet, R. Evolution of the immunoglobulin heavy chain variable region (Igh-V) locus in the genusMus. *Immunogenetics* 30, 315 (1988).

69. Retter, I. *et al.* Sequence and Characterization of the Ig Heavy Chain Constant and Partial Variable Region of the Mouse Strain 129S1. *The Journal of Immunology* 179, 2419–2427 (2007).

70. Henry, R. A., Kendall, P. L., Woodward, E. J., Hulbert, C. & Thomas, J. W. Vκ polymorphisms in NOD mice are spread throughout the entire immunoglobulin kappa locus and are shared by other autoimmune strains. *Immunogenetics* 62, 507–520 (2010).

71. Woodward, E. J. & Thomas, J. W. Multiple Germline κ Light Chains Generate Anti-Insulin B Cells in Nonobese Diabetic Mice. *J Immunol* 175, 1073–1079 (2005).

72. Gibson, D. Genetic polymorphism of mouse immunoglobulin light chains revealed by isoelectric focusing. *J Exp Medicine* 144, 298–303 (1976).

73. Yaari, G. *et al.* Models of Somatic Hypermutation Targeting and Substitution Based on Synonymous Mutations from High-Throughput Immunoglobulin Sequencing Data. *Frontiers in Immunology* 4, 358 (2013).

74. Huang, Y., Niu, B., Gao, Y., Fu, L. & Li, W. CD-HIT Suite: a web server for clustering and comparing biological sequences. *Bioinformatics* 26, 680–682 (2010).

75. Lees, W. *Standardised reporting of genotype statistics as used by OGRDB*.

76. Aouinti, S., Malouche, D., Giudicelli, V., Kossida, S. & Lefranc, M.-P. IMGT/HighV-QUEST Statistical Significance of IMGT Clonotype (AA) Diversity per

Gene for Standardized Comparisons of Next Generation Sequencing Immunoprofiles of Immunoglobulins and T Cell Receptors. *Plos One* 10, e0142353 (2015).

77. Schneider, V. A. *et al.* Evaluation of GRCh38 and de novo haploid genome assemblies demonstrates the enduring quality of the reference assembly. *Biorxiv* 072116 (2016) doi:10.1101/072116.

78. Kent, W. J. BLAT—The BLAST-Like Alignment Tool. *Genome Res* 12, 656–664 (2002).

79. Raivo, K. & Raivo, K., Maintainer. *Package 'pheatmap.'*

80. Wang, J. R., Villena, F. P.-M. de & McMillan, L. Comparative analysis and visualization of multiple collinear genomes. *Bmc Bioinformatics* 13, S13 (2012).

81. Sarsani, V. K. *et al.* The Genome of C57BL/6J "Eve", the Mother of the Laboratory Mouse Genome Reference Strain. *G3 Genes Genomes Genetics* 9, g3.400071.2019 (2019).

82. Potter, M. Antigen-Binding Myeloma Proteins of Mice. *Adv Immunol* 25, 141–211 (1978).

83. Potter, M., Wax, J. S., Anderson, A. O. & Nordan, R. P. Inhibition of plasmacytoma development in BALB/c mice by indomethacin. *J Exp Medicine* 161, 996–1012 (1985).

84. Vandamme, T. F. Use of rodents as models of human diseases. *J Pharm Bioallied Sci* 6, 2–9 (2014).

85. Swearengen, J. R. Choosing the right animal model for infectious disease research. *Animal Model Exp Medicine* 1, 100–108 (2018).

86. Moulia, C. *et al.* Wormy mice in a hybrid zone: A genetic control of susceptibility to parasite infection. *J Evolution Biol* 4, 679–687 (1991).

87. Collins, A. M., Wang, Y., Roskin, K. M., Marquis, C. P. & Jackson, K. J. L. The mouse antibody heavy chain repertoire is germline-focused and highly variable between inbred strains. *Phil Trans R Soc B* 370, 20140236 (2015).

88. Solin, M. L. & Kaartinen, M. Immunoglobulin constant kappa gene alleles in twelve strains of mice. *Immunogenetics* 37, 401–407 (1993).

89. Kindt, T. J., Gris, C., Guenet, J. L., Bonhomme, F. & Cazenave, P. Lambda light chain constant and variable gene complements in wild-derived inbred mouse strains. *Eur J Immunol* 15, 535–540 (1985).

90. Scott, C. L., Mushinski, J. F., Huppi, K., Weigert, M. & Potter, M. Amplification of immunoglobulin λ constant genes in populations of wild mice. *Nature* 300, 757–760 (1982).

91. Kikutani, H. & Makino, S. The Murine Autoimmune Diabetes Model: NOD and Related Strains. *Adv Immunol* 51, 285–322 (1992).

92. Masopust, D., Sivula, C. P. & Jameson, S. C. Of Mice, Dirty Mice, and Men: Using Mice To Understand Human Immunology. *J Immunol* 199, 383–388 (2017).

93. Schneider, V. A. *et al.* Evaluation of GRCh38 and de novo haploid genome assemblies demonstrates the enduring quality of the reference assembly. *Genome Res* 27, 849–864 (2017).

94. Scott, C. L. & Potter, M. Variation in V lambda genes in the genus Mus. *J Immunol Baltim Md 1950* 132, 2638–43 (1984).

95. Schwartz, G. W. & Hershberg, U. Conserved variation: identifying patterns of stability and variability in BCR and TCR V genes with different diversity and richness metrics. *Phys Biol* 10, 035005 (2013).

96. Barak, M., Zuckerman, N. S., Edelman, H., Unger, R. & Mehr, R. IgTree©: Creating Immunoglobulin variable region gene lineage trees. *J Immunol Methods* 338, 67–74 (2008).

97. Heber–Katz, E., Leferovich, J., Bedelbaeva, K., Gourevitch, D. & Clark, L. The scarless heart and the MRL mouse. *Philosophical Transactions Royal Soc Lond Ser B Biological Sci* 359, 785–793 (2004).

98. Amrani, Y. M., Voegtlé, D., Montagutelli, X., Cazenave, P.-A. & Six, A. The Ig light chain restricted B6.κ-λSEG mouse strain suggests that the IGL locus genomic organization is subject to constant evolution. *Immunogenetics* 54, 106–119 (2002).

99. Tutte, A. & Riblet, R. Duplications and deletions of Vh genes in inbred strains of mice. *Immunogenetics* 28, 125–135 (1988).

100. Steen, M.-L., Hellman, L. & Pettersson, U. The immunoglobulin lambda locus in rat consists of two Cλ genes and a single Vλ gene. *Gene* 55, 75–84 (1987).

101. Heiden, J. A. V. *et al.* Dysregulation of B Cell Repertoire Formation in Myasthenia Gravis Patients Revealed through Deep Sequencing. *J Immunol* 198, 1460–1473 (2017).

102. Nurk, S. *et al.* HiCanu: accurate assembly of segmental duplications, satellites, and allelic variants from high-fidelity long reads. *Genome Res* gr.263566.120 (2020) doi:10.1101/gr.263566.120.

103. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34, 3094–3100 (2018).

104. Robinson, J. T. *et al.* Integrative genomics viewer. *Nat Biotechnol* 29, 24–26 (2011).

105. Thorvaldsdóttir, H., Robinson, J. T. & Mesirov, J. P. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform* 14, 178–192 (2013).

106. Merelli, I. *et al.* RSSsite: a reference database and prediction tool for the identification of cryptic Recombination Signal Sequences in human and murine genomes. *Nucleic Acids Res* 38, W262–W267 (2010).

107. Cowell, L. G., Davila, M., Kepler, T. B. & Kelsoe, G. Identification and utilization of arbitrary correlations in models of recombination signal sequences. *Genome Biol* 3, research0072.1-research0072.20 (2002).

108. Cabanettes, F. & Klopp, C. D-GENIES: dot plot large genomes in an interactive, efficient and simple way. *Peerj* 6, e4958 (2018).

109. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res* 49, gkab301- (2021).

110. Hubbard, T. *et al.* The Ensembl genome database project. *Nucleic Acids Res* 30, 38–41 (2002).

111. Lee, B. T. *et al.* The UCSC Genome Browser database: 2022 update. *Nucleic Acids Res* (2021) doi:10.1093/nar/gkab959.

112. Mackay, M. *et al.* Chromosomal Localization in Mouse and Human of the Vasoactive Intestinal Peptide Receptor Type 2 Gene: A Possible Contributor to the Holoprosencephaly 3 Phenotype. *Genomics* 37, 345–353 (1996).

113. Kirschbaum, T., Jaenichen, R. & Zachau, H. G. The mouse immunoglobulin ϰ locus contains about 140 variable gene segments. *Eur J Immunol* 26, 1613–1620 (1996).

114. Kirschbaum, T. *et al.* The 3′ part of the immunoglobulin κ locus of the mouse. *Eur J Immunol* 28, 1458–1466 (1998).

115. Röschenthaler, F. *et al.* The 5′ part of the mouse immunoglobulin κ locus. *Eur J Immunol* 29, 2065–2071 (1999).

116. Giudicelli, V. & Lefranc, M.-P. IMGT-ONTOLOGY 2012. *Frontiers Genetics* 3, 79 (2012).

117. Feeney, A. J. & Riblet, R. DST4: a new, and probably the last, functional DHgene in the BALB/c mouse. *Immunogenetics* 37, 217–221 (1993).

118. Ota, T. & Nei, M. Divergent evolution and evolution by the birth-and-death process in the immunoglobulin VH gene family. *Mol Biol Evol* 11, 469–482 (1994).

119. Hulbert, C., Riseili, B., Rojas, M. & Thomas, J. W. Cutting Edge: B Cell Specificity Contributes to the Outcome of Diabetes in Nonobese Diabetic Mice. *J Immunol* 167, 5535–5538 (2001).

120. Rojas, M., Hulbert, C. & Thomas, J. W. Anergy and not Clonal Ignorance Determines the Fate of B Cells that Recognize a Physiological Autoantigen. *J Immunol* 166, 3194–3200 (2001).

121. Chan, S. *et al.* Copy Number Variants, Methods and Protocols. *Methods Mol Biology* 1833, 193–203 (2018).

122. Bocklandt, S., Hastie, A. & Cao, H. Single Molecule and Single Cell Sequencing. *Adv Exp Med Biol* 1129, 97–118 (2019).

123. Montefiori, L. *et al.* Extremely Long-Range Chromatin Loops Link Topological Domains to Facilitate a Diverse Antibody Repertoire. *Cell Reports* 14, 896–906 (2016).

124. Becker, K. G. *et al.* Clustering of non-major histocompatibility complex susceptibility candidate loci in human autoimmune diseases. *Proc National Acad Sci* 95, 9979–9984 (1998).

125. Rogner, U. C. & Avner, P. Congenic mice: cutting tools for complex immune disorders. *Nat Rev Immunol* 3, 243–252 (2003).

126. Seok, J. *et al.* Genomic responses in mouse models poorly mimic human inflammatory diseases. *Proc National Acad Sci* 110, 3507–3512 (2013).

127. Morse, H. C. The Mouse in Biomedical Research (Second Edition). 1–11 (2007) doi:10.1016/b978-012369454-6/50013-3.

128. Jayaram, N., Bhowmick, P. & Martin, A. C. R. Germline VH/VL pairing in antibodies. *Protein Eng Des Sel* 25, 523–530 (2012).

129. 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 (2015).

130. Brezinschek, H. P., Foster, S. J., Dörner, T., Brezinschek, R. I. & Lipsky, P. E. Pairing of variable heavy and variable kappa chains in individual naive and memory B cells. *J Immunol Baltim Md 1950* 160, 4762–7 (1998).

131. Markel, P. *et al.* Theoretical and empirical issues for marker-assisted breeding of congenic mouse strains. *Nat Genet* 17, 280–284 (1997).

132. Marie-Paule, L. & Gérard, L. *The Immunoglobulin Factsbook*. (Academic press, 2001).

133. The Jackson Laboratory. https://www.jax.org/.

134. Lefranc, M.-P. Immunoglobulin and T Cell Receptor Genes: IMGT® and the Birth and Rise of Immunoinformatics. *Front Immunol* 5, 22 (2014).

135. Kim, H. Y. *et al.* Gene expression profiles of a mouse congenic strain carrying an obesity susceptibility QTL under obesigenic diets. *Genes Nutrition* 5, 237–250 (2010).

136. Mouse Phylogeny Viewer. http://msub.csbio.unc.edu/.

137. Karolchik, D. *et al.* The UCSC Table Browser data retrieval tool. *Nucleic Acids Res* 32, D493–D496 (2004).

138. Team, R. C. *Environment for statistical computing*. (R Foundation for Statistical Computing, 2018).

139. Kassambara, A. & Mundt, F. *factoextra: Extract and Visualize the Results of Multivariate Data Analyses*. (2020).

140. Mi, H., Muruganujan, A., Ebert, D., Huang, X. & Thomas, P. D. PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Res* 47, D419–D426 (2019).

141. Ashburner, M. *et al.* Gene Ontology: tool for the unification of biology. *Nat Genet* 25, 25–29 (2000).

142. Carbon, S. *et al.* The Gene Ontology resource: enriching a GOld mine. *Nucleic Acids Res* 49, D325–D334 (2020).

143. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J Royal Statistical Soc Ser B Methodol* 57, 289–300 (1995).

144. Benjamini, Y. & Yekutieli, D. The control of the false discovery rate in multiple testing under dependency. *Ann Statistics* 29, (2001).

145. Dostie, J. *et al.* Chromosome Conformation Capture Carbon Copy (5C): A massively parallel solution for mapping interactions between genomic elements. *Genome Res* 16, 1299–1309 (2006).

146. Kenter, A. L. & Feeney, A. J. New insights emerge as antibody repertoire diversification meets chromosome conformation. *F1000research* 8, F1000 Faculty Rev-347 (2019).

147. Haarhuis, J. H. I. *et al.* The Cohesin Release Factor WAPL Restricts Chromatin Loop Extension. *Cell* 169, 693-707.e14 (2017).

148. Wang, Y. *et al.* TAD boundary and strength prediction by integrating sequence and epigenetic profile information. *Brief Bioinform* 22, (2021).

149. McArthur, E. & Capra, J. A. Topologically associating domain boundaries that are stable across diverse cell types are evolutionarily constrained and enriched for heritability. *Am J Hum Genetics* 108, 269–283 (2021).

150. Karki, S., Banerjee, S., Mclean, K., Dinner, A. & Clark, M. R. Transcription factories in Igκ allelic choice and diversity. *Advances in immunology* 141, 33–49 (2018).

151. Alipour, E. & Marko, J. F. Self-organization of domain structures by DNA-loop-extruding enzymes. *Nucleic Acids Res* 40, 11202–11212 (2012).

152. Karki, S. *et al.* Regulated Capture of Vκ Gene Topologically Associating Domains by Transcription Factories. *Cell Reports* 24, 2443–2456 (2018).

153. Hill, L. *et al.* Wapl repression by Pax5 promotes V gene recombination by Igh loop extrusion. *Nature* 1–6 (2020) doi:10.1038/s41586-020-2454-y.

154. Kumar, S. *et al.* Flexible ordering of antibody class switch and V(D)J joining during B-cell ontogeny. *Gene Dev* 27, 2439–2444 (2013).

155. Fuxa, M. *et al.* Pax5 induces V-to-DJ rearrangements and locus contraction of the immunoglobulin heavy-chain gene. *Gene Dev* 18, 411–422 (2004).

156. Ciccone, D. N. *et al.* The murine IgH locus contains a distinct DNA sequence motif for the chromatin regulatory factor CTCF. *J Biol Chem* 294, 13580–13592 (2019).

157. Kenter, A. L., Watson, C. T. & Spille, J.-H. Igh Locus Polymorphism May Dictate Topological Chromatin Conformation and V Gene Usage in the Ig Repertoire. *Front Immunol* 12, 682589 (2021).

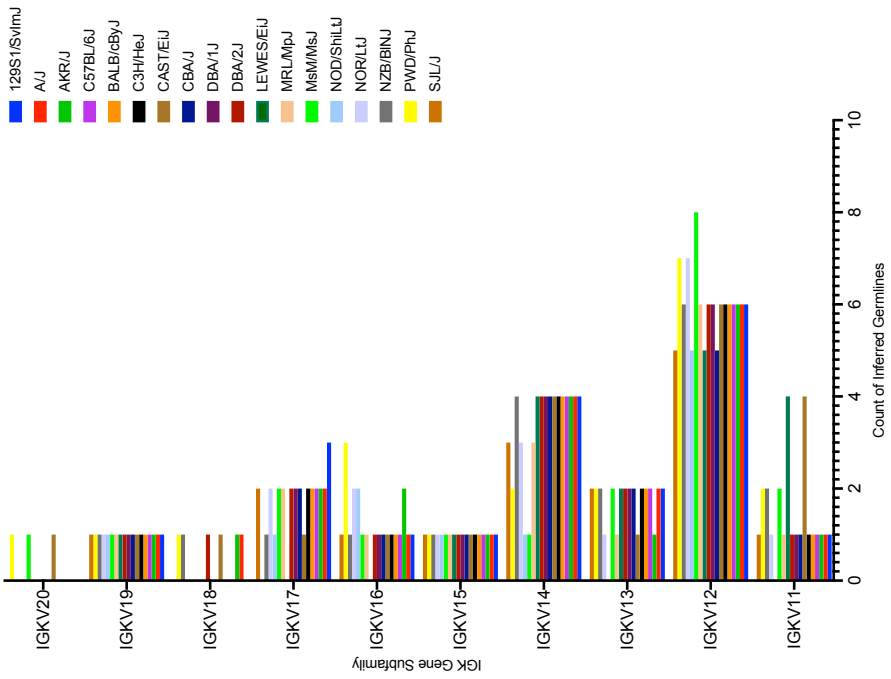
158. Fitzsimmons, S. P., Clark, K. J. & Shapiro, M. A. Evolutionary relationships and polymorphisms among Vκ10 genes from inbred and wild-derived inbred mice. *Immunogenetics* 54, 9–19 (2002).

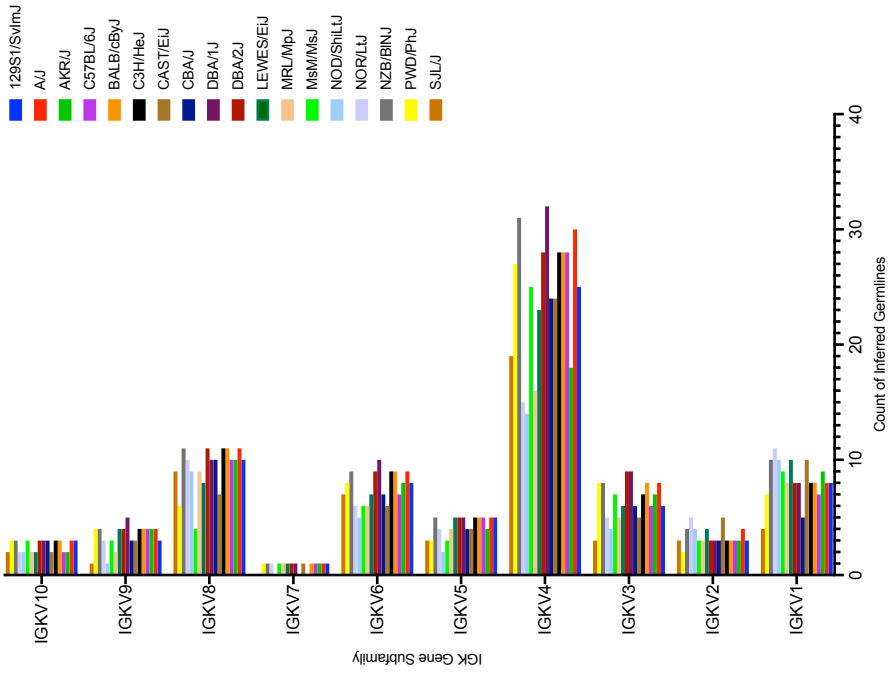
159. Georgiou, G. *et al.* The promise and challenge of high-throughput sequencing of the antibody repertoire. *Nat Biotechnol* 32, 158–168 (2014).

APPENDIX: SUPPLEMENTAL DATA

Supplemental Table 1. IGKV and IGLV AIRR-seq Primer Sequences.

Name	Sequence
IGLC1	AGCTCTTCAGAGGAAGGTGG
IGLC_var1	AGCTCTTCAGGGGAAGGTGG
IGLC2	AGCTCCTCAGAGGAAGGTGG
IGLC3	AGCTCCTCAGGGGAAGGTGG
IGKC	GATGGTGGGAAGATGGATAC


129S1/SvImJ
A/J
AKR/J
C57BL/6J
BALB/cByJ
C3H/HeJ
CAST/EiJ
CBA/J
DBA/1J
DBA/2J
LEWES/EiJ
MRL/MpJ
MsM/MsJ
NOD/ShiLtJ
NOR/LtJ
NZB/BINJ
PWD/PhJ
SJL/J
IGKV20
IGKV19
IGKV18
IGKV17
IGKV16
IGKV15
IGKV14
IGKV13
IGKV12
IGKV11
IGK Gene Subfamily
0
2
4
6
8
10
Count of Inferred Germlines


129S1/SvImJ
A/J
AKR/J
C57BL/6J
BALB/cByJ
C3H/HeJ
CAST/EiJ
CBA/J
DBA/1J
DBA/2J
LEWES/EiJ
MRL/MpJ
MsM/MsJ
NOD/ShiLtJ
NOR/LtJ
NZB/BINJ
PWD/PhJ
SJL/J
IGKV10
IGKV9
IGKV8
IGKV7
IGKV6
IGKV5
IGKV4
IGKV3
IGKV2
IGKV1
IGK Gene Subfamily
0
10
20
30
40
Count of Inferred Germlines

Supplemental Figure 1. IGKV gene subfamily comparison across biomedically significant mouse strains.

Supplemental Table 2. Annotated BALB/cByJ IGHV sequences, RSS sites, and leader sequences.

Name	Type	Sequence
388419	RSS	CACAGTGCTGTAACCACATCCAGAGTGTGTCTGAAACCC
IGHV1.64	IGHV	CAGGTCCAGCTGCAGCAGTCTGGAGCTGAGCTGGTAAGGCCTGGGACTTCAGTGAAGGTGTCCTGCAAGGCTTCTGGATACGCCTTCACTAATTACTTGATAGAGTGGGTAAAGCAGAGGCCTGGACAGGGCCTTGAGTGGATTGGAGTGATTAATCCTGGAAGTGGTGGTACTAACTACAATGAGAAGTTCAAGGGCAAGGCAACACTGACTGCAGACAAATCCTCCAGCACTGCCTACATGCAGCTCAGCAGCCTGACATCTGATGACTCTGCGGTCTATTTCTGTGCAAGA
IGHV1-54*01	Leader	GTGTTCACTCC
IGHV1-54*01	Leader	ATGGAATGGAGCGGAGTCTTTATCTTTCTCCTGTCAGTAACTGCAG
IGHV1.49	IGHV	CAGGTGCAACTGCAGCAGTCTGGGCCTCAGCTGGTTAGGCCTGGGGCTTCAGTGAAGATATCCTGCAAGGCTTCTGGTTACTCATTCACCAGCTACTGGATGCACTGGGTGAAGCAGAGGCCTGGACAAGGTCTTGAGTGGATTGGCATGATTGATCCTTCCGATAGTGAAACTAGGTTAAATCAGAAGTTCAAGGACAAGGCCACATTGACTGTAGACAAATCCTCCAGCACAGCCTACATGCAACTCAGCAGCCCGACATCTGAGGACTCTGCGGTCTATTACTGTGCAAGA
IGHV1-12*01	Leader	GTGTCCACTCC
IGHV1-69*02 3	Leader	ATGGGATGGAGCTGTATCATCCTCTTCTTGGTATCAACAGCTACAG
918668	RSS	CACAGTGTTACAACCACATCCTGAGTGTGTCAGTAACCC
IGHV1.78	IGHV	CAGGTCCAGCTGCAGCAGTCTGGACCTGAGCTGGTGAAGCCTGGGGCCTCAGTGAAGATTTCCTGCAAAGCTTCTGGCTACGCATTCAGTAGCTCTTGGATGAACTGGGTGAAGCAGAGGCCTGGACAGGGTCTTGAGTGGATTGGACGGATTTATCCTGGAGATGGAGATACTAACTACAATGGGAAGTTCAAGGGCAAGGCCACACTGACTGCAGACAAATCCTCCAGCACAGCCTACATGCAGCTCAGCAGCCTGACCTCTGTGGACTCTGCGGTCTATTTCTGTGCAAGA
IGHV1-77*01 2	Leader	GTGTCCATTGC
IGHV1-82*01	Leader	ATGGGATGGAGCCGGATCTTTCTCTTCCTCCTGTCAATAACTGCAG
23628	RSS	CACAGTGCTGCAACCACATCCTGAGTGTGTCAGAAACCC

IGHV1.45	IGHV	CAGGTTCAGCTGCAGCAGTCTGGGGCTGAGCTGGTGAGGCCTGGGTCCTCAGTGAAGATTTCCTGCAAGGCTTCTGGCTATGCATTCAGTAGCTACTGGATGAACTGGGTGAAGCAGAGGCCTGGACAGGGTCTTGAGTGGATTGGACAGATTTATCCTGGAGATGGTGATACTAACTACAATGGAAAGTTCAAGGGTAAAGCCACACTGACTGCAGACAAATCCTCCAGCACAGCCTACATGCAGCTCAGCAGCCTAACATCTGAGGACTCTGCGGTCTATTTCTGTGCAAGA
IGHV1-12*01	Leader	GTGTCCACTCC
IGHV1-80*01	Leader	ATGGAATGGCCTTGTATCTTTCTCTTCCTCCTGTCAGTAACTGAAG
324511	RSS	CACAGTGTTGCAACCACATCCTGAGAGTGTCTGAAAACC
IGHV1.65	IGHV	CCGGTCCAACTGCAGCAGCCTGGGGCTGAGCTTGTGAAGCCTGGGGCTTCAGTAAAGCTGTCCTGCAAGGCTTCTGGCTACACCTTCACCAGCTACTGGATGCACTGGGTGAAGCAGAGGCCTGGACGAGGCCTTGAGTGGATTGGAAGGATTGATCCTAATAGTGGTGGTACTAAGTACAATGAGAAGTTCAAGAGCAAGGCCACACTGACTGTAGACAAACCCTCCAGCACAGCCTACATGCAGCTCAGCAGCCTGACATCTGAGGACTCTGCGGTCTATTACTGTACAAGA
IGHV1-12*01	Leader	GTGTCCACTCC
IGHV1S36*01	Leader	ATGGGATGGAGCTGTATCATGCTCTTCTTGGCAGCAACAGCTACAG
IGHV1.42.P	IGHV	CAGGTCCAGTTTCAACAGTCTGGATCTGAGGTGGGGAGGCCTGGTGCCTCAGTGAAGATGTCCTGCAAGGCTTCTGGCTACACATTCACTAACTACTATATGTATTGATTAAAGCAGAGTCATGGAGATAGCCTAGAGTCAATTGGATATATTTATCCTGGAAATGTTCTTGCTAGATATGCCAAGAAGTTCAAAGGCAAGGCCATATTGACTGTAGACAATTCAGCCAGCAGAGCTTACATACAGCTCAGCAGCCTGACATCTGAAGCCTCTGCTGACTATTACTGTGCTAGACAAGTG
IGHV1-60*01	Leader	GTGTCTACTCC
IGHV1-60*01	Leader	ATGGAATGGAGCTGTATCCTCTCTCTATGTGCAGCATCTACAGGTA
903413	RSS	CACAGTGATACAACCACATCCTGAGTGTGTCACAAAACC
IGHV1.56.P	IGHV	TCTCAGTGAAGATTTCCTGCAAGGGTTCCGGCTACACATTCACTGATTATGCTATGCACTGGGTGAAGCAGAGTCATGCAAAGAGTCTAGAGTGGATTGGAGTTATTAGTACTTACTCTGGTAATACAAACTACAACCAGAAGTTTAAGGGCAAGGCCACAATGACTGTAGACAAATCCTCCAGCACAGCCTATATGGAACTTGCCAGATTGACATCTGAGGATTCTGCCATCTATTACTGTGCAAGA

IGHV1S137*01	Leader	GAGGCCTGGGG
IGHV1-67*01	Leader	ATGGGTTGGAGCTGTATCATCTTCTTTCTGGTAGCAACAGCTACAG
42934	RSS	CACATTGTTGTAACCACATCCTGAGTGTGTCAGAAATCT
IGHV1.46	IGHV	CAGGTCCAGCTGCAGCAGTCTGGAGATGATCTGGTAAAGCCTGGGGCCTCAGTGAAGCTGTCCTGCAAGGCTTCTGGCTACACCTTCACCAGCTACTGGATTAACTGGATAAAACAGAGGCCTGGACAGGGCCTTGAGTGGATAGGACGTATTGCTCCTGGAAGTGGTAGTACTTACTACAATGAAATGTTCAAGGGCAAGGCAACACTGACTGTAGACACATCCTCCAGCACAGCCTACATTCAGCTCAGCAGCCTGTCATCTGAGGACTCTGCTGTCTATTTCTGTGCAAGA
IGHV1S132*01 2	Leader	GTGTCTACTGT
IGHV1S132*01 2	Leader	ATGGACTGGAGTTGGGTCTTTCTCTTCCTCCTGTCAGTAAATGAAG
IGHV1.43.P	IGHV	CAGGTCCAGCTGCAGCAGTCTGGAGCTGAGCTGGTAAGGCCTGGGACTTCAGTGAAGAAGTCCTGCAAGGTTTCTGGATACACCTTCGCTAACTACTGGATAGGTTGGGTAAAGCAGAGGCCTGGACATGGCCTTGAGTGGATTGGAGATATTTACCCTGGAGACGGTGTTACTAACTACAATGAGAAGTTCAAGGGCAAGGCCACACTGACTGCAGACAAATCCTCCAGCACAGCCTACATGTAGCTCAGCAGCCTGACATCTGAGGACTCTGCGGTCTATTACTGTTCGAGA
IGHV1S28*01	Leader	GTGTCCACTCG
IGHV1-63*02	Leader	ATGGAATGGAGCGGGGTCTTTATCTTTCTCCTGTCAGTAACTGCAG
53229	RSS	CACAGTGGTTCAATCACATCCTGAGTGTGTCAGAAACCC
IGHV1.40.P	IGHV	CGGGTCCAGCTGCAGCAGTCTGGGGCTGAGCTGGTGAGGCCTGGGGTCTCAGTGAAGATTTCCTGCAAGGGTTCTGGCTACACATTCACTGATTATGCTATGCACTGGGTGAAGCAGAGTCATTTAAAGAGTCTAGAGTGAATTGGAGTTATTAGTCCTTACTATGGTGATACTAGCTACAGTCAGAAGTTCAAAGGCAAGGCCACAATGACTGTAGACAAATCCTCCAGCACAGCCTATATGGAACTTGCCAGATTGACATCTGAGGATTCTGCCGTCTATTACTGTGCAAGA
IGHV1S137*01 2	Leader	GTGTGTACTCC
IGHV1S137*01 5	Leader	ATGGGTTGGAGTCGTATCATATTTCTCGTTGCAACAGCTACAGGTA
342358	RSS	CACAGTGTTGCAACCACATCCTGAGTGTGTCAGAAACCC

IGHV1.66	IGHV	CAGGTCCAGCTGCAGCAGTCTGGAGCTGAGCTGGTAAGGCCTGGGACTTCAGTGAAGATGTCCTGCAAGGCTGCTGGATACACCTTCACTAACTACTGGATAGGTTGGGTAAAGCAGAGGCCTGGACATGGCCTTGAGTGGATTGGAGATATTTACCCTGGAGGTGGTTATACTAACTACAATGAGAAGTTCAAGGGCAAGGCCACACTGACTGCAGACACATCCTCCAGCACAGCCTACATGCAGCTCAGCAGCCTGACATCTGAGGACTCTGCCATCTATTACTGTGCAAGA
IGHV1-12*01	Leader	GTGTCCACTCC
IGHV1-63*02	Leader	ATGGAATGGAGCGGGGTCTTTATCTTTCTCCTGTCAGTAACTGCAG
60839	RSS	CAAAGGGTTATAATCTCATCCTGAGTGTGTCAGAATCCC
IGHV1.41.P	IGHV	TCCCACATATGTTAACTCTCTGGTTCAGCTGCAGCAGTGTGGAGGTGAGGTTGTGAGACTCAGTGATGGTGTCCTGCTCAGCTTCTGTCTACATATTTAGCAACTGCTATGAACTGTATAAAGCAGAGGCCTGGACAGGCTCTTGAGTGGATTGGATGAATTGATTCTTGAAGTTTTGGTACTACCTACAATCAGAAGTTCAAAGGCCAGGCCACATCTACTGTAGACAAATCCTCCAGCACAGCCTACCTGAAACTAATATCCTGACATCTGAGAACTCTGAAGCCTATTACTGTGAAGA
IGHV1-34*02 3	Leader	GTGCCTTTCCC
IGHV1-34*02 3	Leader	AGGGAATGGAGCTGCATCTTTCTCCTCCTCCTGTCAATAAAAACAG
IGHV1.50	IGHV	CAGGTCCAACTGCAGCAGCCTGGGGCTGAACTGGTGAAGCCTGGGGCTTCAGTGAAGATGTCCTGCAAGGCTTCTGGCTACACATTCACCAGCTACTGGATGCAATGGGTGAAGCAGAGGCCGGGACAAGGCCTTGAGTGGATTGGAAATATTAATCCTAATAGTGGTAGTACTAACTACAATGAGAAGTTCAAGAGCAAGGCCACACTGACTGTAGACAAATCCTCCAGCACAGCCTACATGCAACTCAGCAGCCTGACATCTGAGGACTCTGCAGTCTATTACTGTACAAGA
IGHV1-12*01	Leader	GTGTCCACTCC
IGHV1S17*01	Leader	AAGGGATGGAGCTGTATCATCCTCTTCTTGGTAGCAACAGCTACAG
518355	RSS	CACAGTGTTGCAACCACATCCTGAGAGTGTCAGAAACCC
IGHV1.79	IGHV	CAGGTCCAACTGCAGCAGCCTGGGGCTGAGCTGGTGAGGCCTGGGGCTTCAGTGAAGCTGTCCTGCAAGGCTTCTGGCTACACGTTCACCAGCTACTGGATGAACTGGGTTAAGCAGAGGCCTGAGCAAGGCCTTGAGTGGATTGGAAGGATTGATCCTTACGATAGTGAAACTCACTACAATCAAAAGTTCAAGGACAAGGCCATATTGACTGTAGACAAATCCTCCAGCACAGCCTACATGCAACT

		CAGCAGCCTGACATCTGAGGACTCTGCGGTCTATTACTGTGCAAGA
IGHV1-12*01	Leader	GTGTCCACTCC
IGHV1-52*01	Leader	ATGGGATGGAGCTATATCATCCTCTTCTTGTTAGCAACAGCTACAT
72694	RSS	CACAGTGTTGCAACCACATCCTGAGAGTGTCAGAAACCC
IGHV1.47	IGHV	CAGGTCCAACTGCAGCAGCCTGGGGCTGAGCTTGTGAAGCCTGGGGCTTCAGTGAAGCTGTCCTGCAAGGCTTCTGGCTACACCTTCACCAGCTACTGGATGCACTGGGTGAAGCAGAGGCCTGGACAAGGCCTTGAGTGGATCGGAGAGATTGATCCTTCTGATAGTTATACTAACTACAATCAAAAGTTCAAGGGCAAGGCCACATTGACTGTAGACAAATCCTCCAGCACAGCCTACATGCAGCTCAGCAGCCTGACATCTGAGGACTCTGCGGTCTATTACTGTGCAAGA
IGHV1-12*01	Leader	GTGTCCACTCC
IGHV1-69*02 3	Leader	ATGGGATGGAGCTGTATCATCCTCTTCTTGGTATCAACAGCTACAG
324511	RSS	CACAGTGTTGCAACCACATCCTGAGAGTGTCTGAAAACC
IGHV1.67	IGHV	CAGGTCCAACTGCAGCAGCCTGGGGCTGAACTGGTGAAGCCTGGGGCTTCAGTGAAGCTGTCCTGCAAGGCTTCTGGCTACACCTTCACCAGCTACTGGATGCACTGGGTGAAGCAGAGGCCTGGACAAGGCCTTGAGTGGATTGGAGAGATTAATCCTAGCAACGGTCGTACTAACTACAATGAGAAGTTCAAGAGCAAGGCCACACTGACTGTAGACAAATCCTCCAGCACAGCCTACATGCAACTCAGCAGCCTGACATCTGAGGACTCTGCGGTCTATTACTGTGCAAGA
IGHV1-63*02 2	Leader	ATGTCCACTCC
IGHV1S81*02	Leader	ATGGGATGGAGCTATATCATCCTCTTTTTGGTAGCAACAGCTACAG
518355	RSS	CACAGTGTTGCAACCACATCCTGAGAGTGTCAGAAACCC
IGHV1.80	IGHV	CAGGTCCAACTGCAGCAGCCTGGGGCTGAGCTTGTGAAGCCTGGGACTTCAGTGAAGCTGTCCTGCAAGGCTTCTGGCTACAACTTCACCAGCTACTGGATAAACTGGGTGAAGCTGAGGCCTGGACAAGGCCTTGAGTGGATTGGAGATATTTATCCTGGTAGTGGTAGTACTAACTACAATGAGAAGTTCAAGAGCAAGGCCACACTGACTGTAGACACATCCTCCAGCACAGCCTACATGCAACTCA

		GCAGCCTGGCATCTGAGGACTCTGCTCTCTATTACTGTGCAAGA
IGHV1-14*01	Leader	GTGTCCACTCT
IGHV1-55*01 2	Leader	ATGGGATGGAGCTATATCATCCTCTTCTTGGTAGCAACAGCTACAG
IGHV1.51	IGHV	CAGGTCCAACTGCAGCAGCCTGGGGCTGAGCTGGTGAGGCCTGGGGCTTCAGTGAAGCTGTCCTGCAAGGCTTCTGGCTACACATTCACCAGCTACTGGATGCACTGGATTAAGCAGAGGCCTGGACAAGGTCTTGAGTGGATTGGAACGATTCATCCTTACGATAGTGAAACACACTACAATCAAAAGTTCAAGGACAAGGCCACATTGACTGTAGACAAATCCTCCAGCACAGCCTACATGCAACTCTGCAGCCTGACATCTGAGGACTCTGCGGTCTATTACTGTGCAAGA
IGHV1-12*01	Leader	GTGTCCACTCC
IGHV1-61*01	Leader	ATGGGATGGAGCTATATCATCCTCTTCTTGGTATCAACAGCTACAG
848501	RSS	CACAGTGCTGCAACCACATCCTGAGAGTGTCCGAAAATC
IGHV1.81	IGHV	CAGGTCCAACTGCAGCAGCCTGGGGCTGAACTGGTGAAGCCTGGGGCTTCAGTGAAGTTGTCCTGCAAGGCTTCTGGCTACACCTTCACCAGCTACTGGATGTACTGGGTGAAGCAGAGGCCTGGACAAGGCCTTGAGAGGATTGGAGAGATTAATCCTAGCAATGGTGGTACTAACTACAATGAGAAGTTCAAGAGCAAGGCCACACTGACTGTAGACAAATCCTCCAGCACAGCCTACATGCAACTCAGCAGCCTGACATCTGAGGACTCTGTGGTCTATTACTGTACAAGA
IGHV1-12*01	Leader	GTGTCCACTCC
IGHV1S81*02	Leader	ATGGGATGGAGCTATATCATCCTCTTTTTGGTAGCAACAGCTACAG
IGHV1.44.P	IGHV	CAGATCCAGCTTCCTCAGTCTGGATCTGAGGTGGGGCGGACTGGTGCCTCAGTGAAGATGTTCTGCAAGGCTTCTGGCTACACATTCACTAACTACTATATGTATTGATTAAAGCAGAGTCATGGAGATAGCCTAGAGTGGATTTGATATATTTATCCTGGAAATGGTCTTACTAGCTATTCCAAGAAGTTCAAAGGCAAGGACTCATTGACTATAGACAATTCAGCCAGCACAGCCTACATGCAGCTCAGCAGCATGACATCTGAAGCCTCTGATGACTATTACTGTGCTAGACAAGTG

IGHV1-60*01 2	Leader	GTATCTACTCC
IGHV1-60*01 2	Leader	ATGGAATGGAGCTGTATCCTCTTTCTATGGGCAGCATCTACAGGTA
164306	RSS	CACAGTGTTGTAACCACATCCTGAGTGTGTCAGAAACCC
IGHV1.68	IGHV	CAGGTCCAGCTGCAGCAATCTGGACCTGAGCTGGTGAAGCCTGGGGCTTCAGTGAAGATATCCTGCAAGGCTTCTGGCTATACCTTCACAAGCTACTATATACACTGGGTGAAGCAGAGGCCTGGACAGGGCCTTGAGTGGATTGGATATATTTATCCTAGAGATGGTAGTACTAATTACAATGAGAAGTTCAAGGGCAAGGCCACACTGACTGCAGACACATCCTCCAGCACAGCCTACATGCAGCTCAGCAGCCTGACATCTGAGGACTCTGCAGTCTATTTCTGTGCAAGA
IGHV1-77*01 2	Leader	GTGTCCATTGC
IGHV1S12*01	Leader	ATGGGATGGAGCTGGATCTTTCCCTTCCTCCTGTCAGGAACTGCAG
IGHV1.52	IGHV	CAGGTCCAACTGCAGCAGCCTGGGGCTGAGCTGGTGAGGCCTGGGGCTTCAGTGAAGCTGTCCTGCAAGGCTTCTGGCTACACCTTCACCAGCTACTGGATGAACTGGGTGAAGCAGAGGCCTGGACAAGGCCTTGAATGGATTGGTATGATTGATCCTTCAGACAGTGAAACTCACTACAATCAAATGTTCAAGGACAAGGCCACATTGACTGTAGACAAATCCTCCAGCACAGCCTACATGCAGCTCAGCAGCCTGACATCTGAGGACTCTGCGGTCTATTACTGTGCAAGA
IGHV1-12*01	Leader	GTGTCCACTCC
IGHV1-61*01 2	Leader	ATGGGATGGAGCTGTATCATCCTCTTCTTGGTAGCAACAGCTACAG
IGHV1.53.P	IGHV	TCTCAGTGAAGATTTCCTGCAAGGGTTCTGGCTACACATTCACTGATTATGCTATGCACTGGGTGAAGCAGAGTCATGCAAAGAGTCTAGAGTGGATTGGAGTTATTAGTACTTACTATGGTGATGCTAGCTACAACCAGAAGTTCAAGGGCAAGGCCACAATGACTGTAGACAAATCCTCCAGCACAGCCTATATGGAACTTGCCAGATTGACATCTGAGGATTCTGCTGTTTATTACTGTGCAAGAAACAGTGGTTCAACAACATCCTGAATGTGTCACAAAACCTGGAGGAGAAG
IGHV1S137*01	Leader	GAGGCCTGGGG
IGHV1-67*01	Leader	ATGGGTTGGAGCTGTATCATCTTCTTTCTGGTAGCAACAGCTACAG
817090	RSS	CACAGTGTTGCAACCACATCCTGAGAGTGTCAGAATCCC

IGHV1.82	IGHV	CAGGTCCAAATACAGCAGCCTGGGGCTGAGCTGGTGAGGCCTGGGGCTTCAGTGAAGATGCCCTGCAAGGCTTCTGGCTATACCTTCACCAGCTACAGGATGAACTGGGGGAAGCAGAGGCCTGGACAAGGCCTTGTGTGGATTGGATATATTTATGCTGGTAGTGGTAGTACTGACTACAATGAGAAGTTCAAGAGCAAGGCCACACTGACTGTAGACAAATCCTCCAGCACAGCCTACATGCAGCTCAGCAGCCTGACATCTGAGGACTCTGGGATCTATTATTGTGCAAGA
IGHV1-12*01	Leader	GTGTCCACTCC
IGHV1-55*01 3	Leader	ATAGGATGGAGCTATATCATCCTCCTTTTGGTAGCAACAGCTACAG
570603	RSS	CACATTGACACAGCTTCAGTTTTCAGCTGTACAGTAATT
IGHV8.10	IGHV	CAAGTTACTCTAAAAGAGTCTGGCCCTGGGATATTGAAGCCCTCACAGACCCTCAGTCTGACTTGTTCTTTCTCTGGGTTTTCACTGAGCACTTCTGGTATGGGTGTAGGCTGGATTCGTCAGCCTTCAGGGAAGGGTCTGGAGTGGCTGGCACACATTTGGTGGGATGATGATAAGTACTATAACCCATCCCTGAAGAGCCAGCTCACAATCTCCAAGGATACCTCCAGAAACCAGGTATTCCTCAAGATCACCAGTGTGGACACTGCAGATACTGCCACTTACTACTGTGCTCGAAGAG
IGHV8-8*01 5	Leader	ATGTCTTGTCC
IGHV8-8*01 5	Leader	ATGGACAGGCTTACTTCTTCATTCCTGCTGCTGATTGTCCCTGCAT
134841	RSS	CACAGTGTTCTAACCACATCCTGAGTGTGTCAGAAACCC
IGHV1.48	IGHV	CACGTCCAGCTGCAGCAATCTGGACCTGAGCTGGTGAGGCCTGGGGCTTCAGTGAAGCTGTCCTGCAAGGCTTCTGGCTATATCTTCATCACCTACTGGATGAACTGGGTGAAGCAGAGGCCTGGACAGGGCCTTGAGTGGATTGGACAGATTTTTCCTGCAAGTGGTAGTACTAACTACAATGAGATGTTCGAGGGCAAGGCCACATTGACTGTAGACACATCCTCCAGCACAGCCTACATGCAGCTCAGCAGCCTGACATCTGAGGACTCTGCGGTCTATTACTGTGCAAGA
IGHV1S40*01 2	Leader	GTGTCCACAGC
IGHV1S40*01	Leader	ATGGAATGGAGTTGGGTCTTTCTCTTCCTCCTGTCATTAACTTCAG
108972	RSS	CACAGTGTTGCAACCACATCCTGAGAGTGTCAGAAACCC

IGHV1.69	IGHV	CAGGTCCAACTGCAGCAGTCTGGGCCTGAGCTGGTGAGGCCTGGGGCTTCAGTGAAGATGTCCTGCAAGGCTTCAGGCTATACCTTCACCAGCTACTGGATGCACTGGGTGAAACAGAGGCCTGGACAAGGCCTTGAGTGGATTGGCATGATTGATCCTTCCAATAGTGAAACTAGGTTAAATCAGAAGTTCAAGGACAAGGCCACATTGAATGTAGACAAATCCTCCAACACAGCCTACATGCAGCTCAGCAGCCTGACATCTGAGGACTCTGCAGTCTATTACTGTGCAAGA
IGHV1-12*01	Leader	GTGTCCACTCC
IGHV1S127*01 2	Leader	ATGGGATGGAGCTGTATCATCCTCTTCTTGGCAGCAACAGCTACAG
197154	RSS	CACAGTGTTGTAACCACATCCTGAGTGTGTCAGAAACCC
IGHV1.83	IGHV	CAGGTTCAGCTGCAGCAGTCTGGACCTGAGCTGGTGAAGCCTGGGGCTTTAGTGAAGATATCCTGCAAGGCTTCTGGTTACACCTTCACAAGCTACGATATAAACTGGGTGAAGCAGAGGCCTGGACAGGGACTTGAGTGGATTGGATGGATTTATCCTGGAGATGGTAGTACTAAGTACAATGAGAAATTCAAGGGCAAGGCCACACTGACTGCAGACAAATCCTCCAGCACAGCCTACATGCAGCTCAGCAGCCTGACTTCTGAGAACTCTGCAGTCTATTTCTGTGCAAGA
IGHV1-77*01 2	Leader	GTGTCCATTGC
IGHV1S56*01 4	Leader	ATGGGATGGAGATGGATCTTTCTTTTCCTCCTGTCAGGAACTGCAG
IGHV1.45.P	IGHV	CAGGTCCAGCTGCAGCAGTCTGGACCTGAGCTGGTGAGGCCTGGGACTTCAGTGAAGATATCCTGCAAGGCTTCTGGCTATACCTTCCTCACCTACTGGATGAACTGGGTGAAGTAGATGCCTGGACAGGGCCTTGAGTGGATTGGACAGATTTTTCCTGCAAGTGGTAGTACTAACTACAATGAGATGTTCAAGGGCAAGGCCACATTGACTGTAGACACATCCTCCAGCACAGCCTACATGCAGCTAAGCAGCCTGACATCTGAGGACTCTGCGGTCTATTTCTGTGCAAGA
IGHV1-46*01	Leader	GTGTCCACTGC
IGHV1S10*01	Leader	ATGGAATGCAGCTGGGTCTTTCTCTTCCTCCTGTCATTAACTGCAG
783838	RSS	CACAGTGGTGCAACCACATCCTGAGTGTGTCACAAAACC
IGHV1.57.P	IGHV	TCTCAGTGAAGATTTCCTGCAAGGGTTCTGGCTACACATTCACTGATTATGCTATGCACTGGGTGAAGCAGAGTCATGCAAAGAGTCTAGAGTGGATTGGAGTTATTAGTACTTACTATGGTGATGCTAGCTACAACCAGAAGTTCAAGGGCAAGGCCACAATGACTGTAGACAAATCCTCCAGCACAGCCTATATGGAACTTGCCAGACTGACATCTGAGGATTCTGCCATCTATTACTGTGCAAGA

164306	RSS	CACAGTGTTGTAACCACATCCTGAGTGTGTCAGAAACCC
IGHV1.70	IGHV	CAGGTCCAGCTGCAGCAGTCTGGAGCTGAGCTGGTGAAGCCTGGGGCTTCAGTGAAGCTGTCCTGCAAGACTTCTGGCTACACCTTCACCAGCTACTGGATTCAGTGGGTAAAACAGAGGCCTGGACAGGGCCTTGGGTGGATTGGAGAGATATTTCCTGGAACTGGCACTACTTACTACAATGAGAAGTTCAAGGGCAAGGCCACACTGACTATAGACACATCCTCCAGCACAGCCTACATGCAGCTCAGCAGCCTGACCTCTGAGGACTCTGCTGTCTATTTCTGTGCAAGA
IGHV1S137*01	Leader	GAGGCCTGGGG
IGHV1-12*01	Leader	GTGTCCACTCC
IGHV1-67*01	Leader	ATGGGTTGGAGCTGTATCATCTTCTTTCTGGTAGCAACAGCTACAG
IGHV1S132*01	Leader	ATGGAATGGAGCTGGGTCTTTCTGTTCCTCCTGTCAGTAACTTCAG
IGHV1.53	IGHV	CAGGTCCAACTGCAGCAGCCTGGGTCTGTGCTGGTGAGGCCTGGAGCTTCAGTGAAGCTGTCCTGCAAGGCTTCTGGCTACACCTTCACCAGCTACTGGATGCACTGGGCGAAGCAGAGGCCTGGACAAGGCCTTGAGTGGATTGGAGAGATTCATCCTAATTGTGGTAATATTAACTACAATGAGAAGTTCAAGGGCAAGGCCACACTGACTGTAGACACATCCTCCAGCACAGCCTACGTGGATCTCAGCAGCCTGACATCTGAGGACTCTGCGGTCTATTACTGTGCAAGA
IGHV1-12*01	Leader	GTGTCCACTCC
IGHV1-61*01 3	Leader	ATGGGATGGAGCTCTATCATCCTCTTCTTGGTAGCAACAGCTACAG
518355	RSS	CACAGTGTTGCAACCACATCCTGAGAGTGTCAGAAACCC
IGHV1.84	IGHV	CAGGTCCAACTGCAGCAGCCTGGGGCTGAGCTGGTGAGGCCTGGGGTTTCAGTGAAGCTGTCCTGCAAGGCTTCTGGCTACACATTCACCAGCTACTGGATGCACTGGATTAAGCAGAGGCCTGAGCAAGGCCTTGAGAGGATTGGAGAGATTAATCCTAGCAATGGTGGTACTAACTACAATGAGAAGTTCAAGAGCAAGGCCACACTGACTGTAGACAAATCCTCCAGCACAGCCTACATGCAACTCAGCAGCCTGACATCTGAGGACTCTGCGGTCTATTACTGTGCAAGA
IGHV1-12*01	Leader	GTGTCCACTCC
IGHV1-61*01 2	Leader	ATGGGATGGAGCTGTATCATCCTCTTCTTGGTAGCAACAGCTACAG

IGHV8.4	IGHV	CAGGTTACTCTGAAAGAGTCTGGCCCTGGGATATTGCAGCC CTCCCAGACCCTCAGTCTGACTTGTTCTTTCTCTGGGTTTTCA CTGAGCACTTCTGGTATGAGTGTAGGCTGGATTCGTCAGCCT TCAGGGAAGGGTCTGGAGTGGCTGGCACACATTTGGTGGAA TGATGATAAGTACTATAACCCAGCCCTGAAAAGCCGGCTCAC AATCTCCAAGGATACCTCCAACAACCAGGTATTCCTCAAGAT CGCCAGTGTGGTCACTGCAGATACTGCCACATACTACTGTGC TCGAATA
IGHV8-8*01	Leader	ATGTCCTGTCC
IGHV8-8*01	Leader	ATGGGCAGGCTTACTTCTTCATTCTTGCTACTGATTGTCCCTG CAT
751041	RSS	CACAGTGTTGTAACCACATCCTGAGAGTGTCAGAAACCC
IGHV1.85	IGHV	CAGGTCCAACTGCAGCAGCCTGGGGCTGAACTGGTGAAGCC TGGGACTTCAGTGAAAATGTCCTGCAAGGCTTCTGGCTACAC CTTCACCAGCTACTGGATGCACTGGGTGAAGCAGAGGCCGG GACAAGGCCTTGAGTGGATTGGAGATATTTATCCTGGTAGTG ATAGTACTAACTACAATGAGAAGTTCAAGAGCAAGGCCACAC TGACTGTAGACACATCCTCCAGCACAGCCTACATGCAACTCA GCAGCCTGACATCTGAGGACTCTGCGGTCTATTACTGTGCAA GA
IGHV1-12*01	Leader	GTGTCCACTCC
IGHV1-55*01 4	Leader	TTGGGATGGAGCTGTATCATCCTCTTCTTGGTAGCAACAGCT ACAG
IGHV1.46.P	IGHV	CAGGTCCAGCTGCAGCAGTCTGGAGCTGAGCTGGTAAGGCC TGGGACTTCAGTGAAGAAGTCCTGCAAGGCTTCTGGATACAC CTTCGCTAACTACTGGATAGGTTGGGTAAAGCAGAGGCCTG GACATGGCCTTGAGTGGATTGGAGATATTTACCCTGGAGAC GGTGTTACTAACTACAATGAGAAGTTCAAGGGCAAGGCCACA CTGACTGCAGACAAATCCTCCAGCACAGCCTATATGGAGCTT AGTAGATTGACATCTGAGGACTCTGCAGTCTAATACTGTGCA A
IGHV1S11*01	Leader	GTGTCTACTCG
IGHV1-63*02	Leader	ATGGAATGGAGCGGGGTCTTTATCTTTCTCCTGTCAGTAACT GCAG
204332	RSS	CACAGTGTTGTAACCACATCCCGAGTGTGTCAGAAACCC
IGHV1.71	IGHV	AAGGTCCAGCTGCAGCAGTCTGGAGCTGAGCTGGTGAAACC CGGGGCATCAGTGAAGCTGTCCTGCAAGGCTTCTGGCTACA CCTTCACTGAGTATACTATACACTGGGTAAAGCAGAGGTCTG GACAGGGTCTTGAGTGGATTGGGTGGTTTTACCCTGGAAGT GGTAGTATAAAGTACAATGAGAAATTCAAGGACAAGGCCACA TTGACTGCGGACAAATCCTCCAGCACAGTCTATATGGAGCTT

		AGTAGATTGACATCTGAAGACTCTGCGGTCTATTTCTGTGCAAGACACGAAGA
IGHV1-12*01	Leader	GTGTCCACTCC
IGHV1-62-2*01	Leader	ATGGAATGGTGCTGGGTCTTTCTCTTCCTCCTGTCAGTAACTGCAG
IGHV1.54	IGHV	CAGGTGCAACTGCAGCAGCCTGGGGCTGAGCTGGTGAAGCCTGGGGCTTCAGTGAAGATATCCTGCAAGGCTTCTGGCTACACCTTCACCAGCTACTGGATGAACTGGGTGAAGCAGAGGCCTGGACAAGGCCTTGAGTGGATCGGAGAGATCGATCCTTCTGATAGTTATACTAACAACAATCAAAAGTTCAAGGACAAGGCCACATTGACTGTAGACAAATCCTCCAGCACAGCCTACATGCAGCTCAGCAGCCTGACATCTGAGGACTCTGCGGTCTATTACTGTGCAAGA
IGHV1-12*01	Leader	GTGTCCACTCC
IGHV1-12*01	Leader	ATGGGATGGAGTTGTATCATCCTCTTCTTGGTAGCAACAGCTACAG
726856	RSS	CACAGTGTTGCAACCACATCCTGAGAGTGGCCGAAAACC
IGHV1.86	IGHV	CAGGTCCAACTGCAGCAACCTGGGTCTGAGCTGGTGAGGCCTGGAGCTTCAGTGAAGCTGTCCTGCAAGGCTTCTGGCTACACATTCACCAGCTACTGGATGCACTGGGTGAAGCAGAGGCCTGGACAAGGCCTTGAGTGGATTGGAAATATTTATCCTGGTAGTGGTAGTACTAACTACGATGAGAAGTTCAAGAGCAAGGCCACACTGACTGTAGACACATCCTCCAGCACAGCCTACATGCAGCTCAGCAGCCTGACATCTGAGGACTCTGCGGTCTATTACTGTACAAGA
IGHV1-12*01	Leader	GTGTCCACTCC
IGHV1S5*01	Leader	ATGGGATGGAGCTCTATCATCCTCTTCTTGGTAGCAACAGCCTCAG
IGHV8.9	IGHV	CAGGTTACTCTGAAAGAGTCTGGCCCTGGGATATTGCAGCCCTCCCAGACCCTCAGTCTGACTTGTTCTTTCTCTGGGTTTTCACTGAGCACTTCTGGTATGGGTGTAGGCTGGATTCGTCAGCCTTCAGGGAAGGGTCTGGAGTGGCTGGCACACATTTGGTGGGATGATGACAAGCGCTATAACCCAGCCCTGAAGAGCCGACTGACAATCTCCAAGGATACCTCCAGCAACCAGGTATTCCTCAAGATCGCCAGTGTGGACACTGCAGATACTGCCACATACTACTGTGCTCGAATAG

IGHV8-8*01	Leader	ATGTCCTGTCC
IGHV8-8*01	Leader	ATGGGCAGGCTTACTTCTTCATTCTTGCTACTGATTGTCCCTGCAT
IGHV1.55	IGHV	CAGGTCCAACTGCAGCAGCCTGGGGCTGAACTGGTGAAGCCTGGGGCTTCAGTGAAGTTGTCCTGCAAGGCTTCTGGCTACACCTTCACCAGCTACTATATGTACTGGGTGAAGCAGAGGCCTGGACAAGGCCTTGAGTGGATTGGGGGGATTAATCCTAGCAATGGTGGTACTAACTTCAATGAGAAGTTCAAGAGCAAGGCCACACTGACTGTAGACAAATCCTCCAGCACAGCCTACATGCAACTCAGCAGCCTGACATCTGAGGACTCTGCGGTCTATTACTGTACAAGA
IGHV1-12*01	Leader	GTGTCCACTCC
IGHV1S81*02	Leader	ATGGGATGGAGCTATATCATCCTCTTTTTGGTAGCAACAGCTACAG
485460	RSS	CACATTGTCACAGCCTCAGTTATCAGCTGTACATTAATT
IGHV8.11	IGHV	CAGGTTACTCTGAAAGAGTCTGGCCCTGGGATATTGCAGCCCTCCCAGACCCTCAGTCTGACTTGTTCTTTCTCTGGGTTTTCACTGAGCACTTCTGGTATGGGTGTAGGCTGGATTCGTCAGCCTTCAGGGAAGGGTCTGGAGTGGCTGGCACACATTTGGTGGGATGATGATAAGTACTATAACACAGCCCTGAAGAGCGGGCTCACAATCTCCAAGGATACCTCCAAAAACCAGGTCTTCCTCAAGATCGCCAGTGTGGACACTGCAGATACTGCCACATACTACTGTGCTCGAATA
IGHV8-8*01	Leader	ATGTCCTGTCC
L-Part1	Leader	ATGGGCAGACTTACTTCTTCATTCTTGCTACTGATTGTCCCTGCAT
164306	RSS	CACAGTGTTGTAACCACATCCTGAGTGTGTCAGAAACCC
IGHV1.72	IGHV	CAGGTCCAGCTGCAGCAGTCTGGACCTGAGCTGGTGAAGCCTGGGGCTTCAGTGAAGATGTCCTGCAAGGCTTCTGGCTACACCTTCACAAGCTACTATATACACTGGGTGAAGCAGAGGCCTGGACAGGGACTTGAGTGGATTGGATGGATTTATCCTGGAGATGGTAGTACTAAGTACAATGAGAAGTTCAAGGGCAAGACCACACTGACTGCAGACAAATCCTCCAGCACAGCCTACATGTTGCTCAGCAGCCTGACCTCTGAGGACTCTGCGATCTATTTCTGTGCAAGG
IGHV1-77*01 2	Leader	GTGTCCATTGC
IGHV1S56*01 3	Leader	ATGCGATGGAGCTGGATCTTTCTCTTCCTCCTGTCAATAACTGCAG

IGHV1.47.P	IGHV	ATCTGTCCATCCATAGGTATTCACTCCCAGGTCTAGCTGCAG CAGTCAGGGGCTGTGAAGATGTCGTTCAAGGTTTCTGGTTAT ACCTTTACTGAATACTACATGCCCTGGTCAAGCAGAATCATG GAAAGACCCTTGACTGGATTGGCAATATTAATTCTTACAATG GTGGCATAACTACAATGAAAATTTCAAGGGCAAGGACACATT AACTGTAGACAAATCCTCCAGCACAGCCTATATGTTGCTTAG CAAATTGACATTTGAGGATTCTGTGGTCTATTATTGTGCAAGA
IGHV1-34*02	Leader	ACATCCATTCT
IGHV1-34*02	Leader	ATGGGATGGAGCTATATCATCTTCTTCCTTGTAGCAACAGCT ATAG
693268	RSS	CACATTGTTTTAACCACATCCTGAGTGTGTCAGAAACAC
IGHV1.87	IGHV	GAGGTCCAGCTTCAGCAGTCTGGAGCTGAGCTGGGGAGGC CTGGGTCCTCAGTGAAGCTGTCCTGCAAGACTTCTGGATATA CATTCACAAGCTATGGTATAAAGTGGGTGAAACAGAGGCCTG GACAGGGCCTGGAATGGATTGGATATATTTATCCTGGAAATG GTTATACTGTGTACAATGAGAAGTTCCAGGGCAAGGCCACAC TGACTTCAGACACATCCTCCAGCACAGCCTACATGCAGCTCA GAAGCCTGACATCTGAGGACTCTGCAATCTATTTCTGTGCAA GA
IGHV1-14*01	Leader	GTGTCCACTCT
IGHV1S134*01	Leader	ATGGAATGGAGCTGCATCTTTCTCTTCCTCCTGTCAGTAACT GCAG
150229	RSS	CACAGTGATACAACCACATCCTGAGTGTGTCACAAAACC
IGHV1.54.P	IGHV	AATCAGTGAAGATTTCCTGCAAGGGTTCCGGCTACACATTCA CTGATTATGCTATGCACTGGGTGAAGCAGAGTCATGCAAAGA GTCTAGAGTGGATTGGAGTTATTAGTATTTACTATGATAATAC AAACTACAACCAGAAGTTTAAGGGCAAGGCCACAATGACTGT AGACAAATCCTCCAGCACAGCCTATATGGAACTTGCCAGATT GACATCTGAGGATTCTGCCATCTATTACTGTGCAAGA
IGHV1S137*01	Leader	GAGGCCTGGGG
IGHV1-67*01	Leader	ATGGGTTGGAGCTGTATCATCTTCTTTCTGGTAGCAACAGCT ACAG
IGHV1.48.P	IGHV	TTCAGTCCTGGTCAGCTTCTTAATCTGCATTTGTCTGGCAAAC ACTGCTCTTGACTCTTTCTCAGGCAAAGGAACTGAGCGGGAA ATAGGAGATGGGGAATACAGGCAAGAGAGAGAGAGAGAGAG AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG AGATATTTCACAAAGATCAGATGTTTGAGGACAGAACCATAC TGGATGCAGATATAATGTTCAACACAGCTTGATGGAGCTCAG CAGCCTGACATTGGAAGGCGCTGTTATCTATTGAAGTGTTAG G
IGHV1-81*01	Leader	GTTACTTCTCC

IGHV1-81*01	Leader	CTGCTCTGCACAATTATTTTTTTCTGTTCTCTTACTATGTCTATAA
674185	RSS	CACATTGTCACAGCCTCAGTTATCAACTGTACATTAATT
IGHV8.12	IGHV	CAGGTTACTCTGAAAGAGTCTGGCCCTGGGATATTGCAGCCCTCCCAGACCCTCAGTCTGACTTGTTCTTTCTCTGGGTTTTCACTGAGCACTTCTAATATGGGTGTAGGCTGGATTCGTCAGCCTTCAGGGAAGGGTCTGGAGTGGCTGTTACACATTTTGTGGAATGATAGTAAGTACTATAACCCAGCCCTGAAGAGCCGGCTCACAATCTCCAAGGATACCTACAACAACCAGGTATTCCTCAAGATCGCCAATGTGGACACTGCAGATACTGCCACATACTACTGTGCTCGAATAG
L-Part2	Leader	ATGTCCTGTGC
L-Part1	Leader	ATGGGCAGGCTTACTTTTTCATTCTTGCTACTGATTGTCCCTGCAT
IGHV8.5	IGHV	CAGGTTACTCTGAAAGAGTCTGGCCCTGGGATATTGCAGCCCTCCCAGACCCTCAGTCTGACTTGTTCTTTCTCTGGGTTTTCACTGAGCACTTATGGTATAGGAGTAGGCTGGATTCGTCAGCCTTCAGGGAAGGGTCTGGAGTGGCTGGCACACATTTGGTGGAATGATAATAAGTACTATAACACAGCCCTGAAGAGCCGGCTCACAATCTCCAAGGATACCTCCAACAACCAGGTATTCCTCAAGATCGCCAGTGTGGACACTGCAGATACTGCCACATACTACTGT
IGHV8-11*01	Leader	ATGTCCTCTCC
IGHV8-11*01	Leader	ATGGGCAGGCTTACTTCTTCATTCCTGCTACTGATTGTTCCTGCAT
108972	RSS	CACAGTGTTGCAACCACATCCTGAGAGTGTCAGAAACCC
IGHV1.73	IGHV	CAGGTCCAACTGCAGCAGCCTGGGGCTGAGCTTGTGATGCCTGGGGCTTCAGTGAAGATGTCCTGCAAGGCTTCTGGCTACACATTCACTGACTACTGGATGCACTGGGTGAAGCAGAGGCCTGGACAAGGCCTTGAGTGGATCGGAGCGATTGATACTTCTGATAGTTATACTAGCTACAATCAAAAGTTCAAGGGCAAGGCCACATTGACTGTAGACGAATCCTCCAGCACAGCCTACATGCAGCTCAGCAGCCTGACATCTGAGGACTCTGCGGTCTATTACTGTGCAAGA
IGHV1-69*01	Leader	GTGTCAACTCC
IGHV1-69*01	Leader	ATGAGATGGAGCTGTATCATCCTCTTCTTGGTAGCAACAGCTACAG
RSS Pass	RSS	CATGGTGACATTGAATGGTACAGAAATA
RSS Pass	RSS	CACCAAGGTACAGACTCTAGAAGAATCA
108972	RSS	CACAGTGTTGCAACCACATCCTGAGAGTGTCAGAAACCC

IGHV1.74	IGHV	CAGGTCCAACTGCAGCAGCCTGGGGCTGAGCTTGTGAAGCCTGGGGCTCCAGTGAAGCTGTCCTGCAAGGCTTCTGGCTACACCTTCACCAGCTACTGGATGAACTGGGTGAAGCAGAGGCCTGGACGAGGCCTCGAGTGGATTGGAAGGATTGATCCTTCCGATAGTGAAACTCACTACAATCAAAAGTTCAAGGACAAGGCCACACTGACTGTAGACAAATCCTCCAGCACAGCCTACATCCAACTCAGCAGCCTGACATCTGAGGACTCTGCGGTCTATTACTGTGCAAGA
IGHV1-12*01	Leader	GTGTCCACTCC
IGHV1-61*01 2	Leader	ATGGGATGGAGCTGTATCATCCTCTTCTTGGTAGCAACAGCTACAG
IGHV1.49.P	IGHV	CAGGTCCAGCTGCAGCAGTCTGGACCTGAGCTGGTGAAGCCTGGGGCTTCAGTGAGGATATCCTGCAAGGCTTCTGGCTACACCTTCACAAGCTACTATATACACTGGGTGAAGCAGAGGCCTGGACAGGGACTTGAGTGGATTGGATGGATTTATCCTGGAGATGGTAATACTAAGTACAATTGAGAAGTTCAAGAGCAAGACCACACTGACTGCAGACAAATCCTCCAGCACAGCCTACATGCAGCTCAGCAGCCTGACCTCTGAGGACTCTGCGGTCTATTTCTGTGCAAGA
IGHV1-77*01 2	Leader	GTGTCCATTGC
IGHV1-82*01	Leader	ATGGGATGGAGCCGGATCTTTCTCTTCCTCCTGTCAATAACTGCAG
RSS Pass	RSS	CACAGTGGTACAGAGGAGTCTCAAATCC
RSS Pass	RSS	CACGGTGAGAGAGGCATGTACAGAATCC
632346	RSS	CACAGTGTTGTAGCCACATCCTGAGTGTGTCAGAAACAC
IGHV1.88	IGHV	CAGGTCCAGTTGCAGCAGTCTGGAGCTGAGCTGGTAAGGCCTGGGACTTCAGTGAAGATATCCTGCAAGGCTTCTGGATACGCCTTCACTAACTACTGGCTAGGTTGGGTAAAGCAGAGGCCTGGACATGGACTTGAGTGGATTGGAGATATTTACCCTGGAAGTGGTAATACTTACTACAATGAGAAGTTCAAGGGCAAAGCCACACTGACTGCAGACAAATCCTCGAGCACAGCCTATATGCAGCTCAGTAGCCTGACATCTGAGGACTCTGCTGTCTATTTCTGTGCAAGA
IGHV1-12*01	Leader	GTGTCCACTCC
IGHV1-63*01	Leader	ATGGAATGGAGTGGGGTCTTTATCTTTCTCTTGTCAGTAACTGCAG

IGHV1.56	IGHV	CAGGTCCAGCTGCAGCAGTCTGGGCCTGAGCTGGTGAGGC CTGGGGTCTCAGTGAAGATTTCCTGCAAGGGTTCCGGCTAC ACATTCACTGATTATGCTATGCACTGGGTGAAGCAGAGTCAT GCAAAGAGTCTAGAGTGGATTGGAGTTATTAGTACTTACTAT GGTAATACAAACTACAACCAGAAGTTTAAGGGCAAGGCCACA ATGACTGTAGACAAATCCTCCAGCACAGCCTATATGGAACTT GCCAGATTGACATCTGAGGATTCTGCCATCTATTACTGTGCA AGA
IGHV1-67*01	Leader	GTGTGCACTCC
IGHV1-67*01	Leader	ATGGGTTGGAGCTGTATCATCTTCTTTCTGGTAGCAACAGCT ACAG
620251	RSS	CACAGTGTTTCAACCACATCCTGAGAGTGTCAGAAACCC
IGHV1.89	IGHV	CAGGTGCAACTGCAGCAGCCTGGGGCTGAGCTGGTGAAGC CTGGGGCCTCAGTGAAGATGTCCTGCAAGGCTTCTGGCTAC ACATTTACCAGTTACAATATGCACTGGGTAAAGCAGACACCT GGACAGGGCCTGGAATGGATTGGAGCTATTTATCCAGGAAA TGGTGATACTTCCTACAATCAGAAGTTCAAAGGCAAGGCCAC ATTGACTGCAGACAAATCCTCCAGCACAGCCTACATGCAGCT CAGCAGCCTGACATCTGAGGACTCTGCGGTCTATTACTGTGC AAGA
IGHV1-12*01	Leader	GTGTCCACTCC
IGHV1-12*01	Leader	ATGGGATGGAGTTGTATCATCCTCTTCTTGGTAGCAACAGCT ACAG
388419	RSS	CACAGTGCTGTAACCACATCCAGAGTGTGTCTGAAACCC
IGHV1.75	IGHV	CAGGTCCAGTTGCAGCAGTCTGGAGCTGAGCTGGTAAGGCC TGGGACTTCAGTGAAGGTGTCCTGCAAGGCTTCTGGATACG CCTTCACTAATTACTTGATAGAGTGGGTAAAGCAGAGGCCTG GACAGGGCCTTGAGTGGATTGGGGTGATTAATCCTGGAAGT GGTGGTACTAACTACAATGAGAAGTTCAAGGGCAAGGCAAC ACTGACTGCAGACAAATCCTCCAGCACTGCCTACATGCAGCT CAGCAGCCTGACATCTGATGACTCTGCGGTCTATTTCTGTGC AAGA
IGHV1-54*01	Leader	GTGTTCACTCC
IGHV1-54*03	Leader	ATGGAATGGAGCAGAGTCTTTATCTTTCTCCTATCAGTAACTG CAG
IGHV1.57	IGHV	CAGGTCCAACTGCAGCAGCCTGGGGCTGAGCTGGTGAAGC CTGGGGCTTCAGTGAAGATGTCCTGCAAGGCTTCTGGCTAC ACCTTCACCAGCTACTGGATGCACTGGGTGAAGCAGAGGCC TGGACAAGGCCTTGAGTGGATCGGAACGATTGATCCTTCAG ATAGTTATACTAGCTACAATCAAAAGTTCAAGGGCAAGGCCA CATTGACTGTAGACACATCCTCCAGCACAGCCTACATGCAGC

		TCAGCAGCCTGACATCTGAGGACTCTGCGGTCTATTACTGTACAAGA
IGHV1-69*01	Leader	GTGTCAACTCC
IGHV1-69*01	Leader	ATGAGATGGAGCTGTATCATCCTCTTCTTGGTAGCAACAGCTACAG
607058	RSS	CACAGTGCTGCAACCACATCCTGAGAGTGTCCGAAAACC
IGHV1.90	IGHV	CAGGTCCAACTGCAGCAGTCTGGGGCTGAACTGGTGAAGCCTGGGGCTTCAGTGAAGTTGTCCTGCAAGGCTTCTGGCTACACCTTCACCAGCTACTATATGTACTGGGTGAAGCAGAGGCCTGGACAAGGCCTTGAGTGGATTGGAGAGATTAATCCTAGCAATGGTGGTACTAACTTCAATGAGAAGTTCAAGAGCAAGGCCACACTGACTGTAGACAAATCCTCCAGCACAGCATACATGCAACTCAGCAGCCTGACATCTGAGGACTCTGCGGTCTATTACTGTACAAGA
IGHV1-12*01	Leader	GTGTCCACTCC
IGHV1S81*02 5	Leader	ATGGGATGGAGCTATATCATCCTCTTTTTGGTAGCAACAGCAACAG
IGHV1.76	IGHV	CAGGTCCAACTGCAGCAGCCTGGGGCTGAGCTTGTGAAGCCTGGGGCTTCAGTGAAGATGTCCTGCAAGGCTTCTGGCTACACCTTCACCAGCTACTGGATAACCTGGGTGAAGCAGAGGCCTGGACAAGGTCTTGAGTGGATTGGAGATATTTATCCTGGTAGTGGTAGTACTAACTACAATGAGAAATTCAAGAGCAAGGCCACACTGACTGTAGACACATCCTCCAGCACAGCCTACATGCAGCTCAGCAGCCTGACATCTGAGGACTCTGCGGTCTAT
IGHV1-12*01	Leader	GTGTCCACTCC
IGHV1-55*01	Leader	ATGGGATGGAGCTGTATCATCCTCTTTTTTGGTAGCAGCAGCTACA
IGHV8.6	IGHV	CAGGTTACTCTGAAAGAGTCTGGCCCTGGGATATTGCAGCCCTCCCAGACCCTCAGTCTGACTTGTTCTTTCTCTGGGTTTTCACTGAGCACTTCTGGTATGGGTGTGAGCTGGATTCGTCAGCCTTCAGGAAAGGGTCTGGAGTGGCTGGCACACATTTACTGGGATGATGACAAGCGCTATAACCCATCCCTGAAGAGCCGGCTCACAATCTCCAAGGATACCTCCAGCAACCAGGTATTCCTCAAGATCACCAGTGTGGACACTGCAGATACTGCCACATACTACTGTGCTCGA
IGHV8-8*01	Leader	ATGTCCTGTCC

L-Part1	Leader	ATGGACAGGCTTACTTCCTCATTGCTGCTGCTGATTGTCCCT GCAT
518355	RSS	CACAGTGTTGCAACCACATCCTGAGAGTGTCAGAAACCC
IGHV1.9 1	IGHV	CAGGCCCAAATACAGCAGCCTGGGGCTGAGCTGGTGAGGC CTGGGGCTTCAGTGAAGATGCCGTGCAAGGCTTCTGGCTAT ACCTTCACCAGCTACAGGATGAACTGGGGGAAGCAGAGGCC TGGACAAGGCCTTGAGTGGATTGGATATATTTATGCTGGTAG TGGTAGTACTGACTACAATGAGAAGTTCAAGAGCAAGGCCAC ACTGACTGTAGACAAATCCTCCAGCACAGCCTACATGCAGCT CAGCAGCCTGACATCTGAGGACTCTGGGATCTATTATTGTGC AAGA
IGHV1-12*01	Leader	GTGTCCACTCC
IGHV1-55*01 3	Leader	ATAGGATGGAGCTATATCATCCTCCTTTTGGTAGCAACAGCT ACAG
570603	RSS	CACATTGACACAGCTTCAGTTTTCAGCTGTACAGTAATT
IGHV8.1 3	IGHV	CAAGTTACTCTAAAAGAGTCTGGCCCTGGGATATTGAAGCCC TCACAGACCCTCAGTCTGACTTGTTCTTTCTCTGGGTTTTCAC TGAGCACTTCTGGTATGGGTGTAGGCTGGATTCGTCAGCCTT CAGGGAAGGGTCTGGAGTGGCTGGCACACATTTGGTGGGAT GATGATAAGTACTATAACCCATCCCTGAAGAGCCGGCTCACA ATCTCCAAGGATACCTCCAGAAACCAGGTATTCCTCAAGATC ACCAGTGTGGACACTGCAGATACTGCCACTTACTACTGTGCT CGAAGAG
IGHV8-8*01 5	Leader	ATGTCTTGTCC
IGHV8-8*01 5	Leader	ATGGACAGGCTTACTTCTTCATTCCTGCTGCTGATTGTCCCT GCAT
27263	RSS	CACAGTGTTACAACCATATCCTGAGTGTGTCAGAAACAC
IGHV1.7 7	IGHV	CAGGTCCAGCTGCAGCAGTCTGGACCTGAGCTGGTGAGGCC TGGGGCTTCAGTGAAGATATCCTGCAAGGCTCCTGGCTACA CCTTCACCAGCCACTGGATGCAGTGGGTAAGACAGAGGCCT GGACAGGGCCTTGAGTGGATTGGAGAGATTTTTCCTGGAAG TGGTAGTACTTATTATAATGAGAAGTTCAAGGGCAAGGCCAC ACTGACTTTAGACACATCCTCCAGCACAGCCTACATGCAGCT CAGCAGCCTGACATCTGAGGACTCTGCGGTCTATTTCTGTGC AAGA
IGHV1-56*02	Leader	TGTTCCACTGC
IGHV1-56*02	Leader	ATGGAATGGAGTTGGGTCTTTCTCTTCCTCCTGTCAGTAACT GCAG
197154	RSS	CACAGTGTTGTAACCACATCCTGAGTGTGTCAGAAACCC

IGHV1.92	IGHV	CAGGTTCAGCTGCAGCAGTCTGGACCTGAGCTGGTGAAGCCTGGGGCTTTAGTGAAGATATCCTGCAAGGCTTCTGGTTACACCTTCACAAGCTACGATATAAACTGGGTGAAGCAGAGGCCTGGACAGGGACTTGAGTGGATTGGATGGATTTATCCTGGAGATGGTAGTACTAAGTACAATGAGAAATTCAAGGGCAAGGCCACACTGACTGCAGACAAATCCTCCAGCACAGCCTACATGCAGCTCAGCAGCCTGACTTCTGAGAACTCTGCAGTCTATTTCTGTGCAAGA
IGHV1-77*01 2	Leader	GTGTCCATTGC
IGHV1S56*01 4	Leader	ATGGGATGGAGATGGATCTTTCTTTTCCTCCTGTCAGGAACTGCAG
IGHV1.50.P	IGHV	CAGGTCCAGCTGCAGCAGTCTGGACCTGAGCTGGTGAGGCCTGGGACTTCAGTGAAGATATCCTGCAAGGCTTCTGGCTATACCTTCCTCACCTACTGGATGAACTGGGTGAAGTAGAGGCCTGSACAGGGCCTTGAGTGGATTGGACAGATTTTTCCTGCAAGTGGTAGTACTAACTACAATGAGATGTTCAAGGGCAAGGCCACATTGACTGTAGACACATCCTCCAGCACAGCCTACATGCAGCTAAGCAGCCTGACATCTGAGGACTCTGCGGTCTATTTCTGTGCAAGA
IGHV1-46*01	Leader	GTGTCCACTGC
IGHV1S10*01	Leader	ATGGAATGCAGCTGGGTCTTTCTCTTCCTCCTGTCATTAACTGCAG
541112	RSS	CACAGTGATACAACCACATCCTGAGTGTGTCACAAAACA
IGHV1.58.P	IGHV	CAGGTCCAGCTGCAGTCTGGGCCTGAGGTGGTGAGGCCTGGGGTCTCAGTGAAGATTTCCTGCAAGGGTTCCGGCTACACATTCACTGATTATGCTATGCACTGGGTGAAGCAGAGTCATGCAAAGAGTCTAGAGGATTGGAGTTATTAGTACTTACAATGGTAATACAAACTACAACCAGAAGTTTAAGGGCACACACAATGACTGTAGACAAATCCTCCAGCACAGCCTATATGGAACTTGCCAGATTGACATCTGAGGATTCTGCCATCTATTACTGTGCAAGA
IGHV1-67*01	Leader	GTGTGCACTCC
IGHV1-67*01 5	Leader	CCATGGGTTGGAGCGTATCATCTTCTTCTGGTAGCAACAGCTACAG
IGHV1.58	IGHV	CAGGTCCAACTGCAGCAGCCTGGGTCTGTGCTGGTGAGGCCTGGAGCTTCAGTGAAGCTGTCCTGCAAGGCTTCTGGCTACACCTTCACCAGCTCCTGGATGCACTGGGCGAAGCAGAGGCCTGGACAAGGCCTTGAGTGGATTGGAGAGATTCATCCTAATAGTGGTAATACTAACTACAATGAGAAGTTCAAGGGCAAGGCCACACTGACTGTAGACACATCCTCCAGCACAGCCTACGTGGATCTCAGCAGCCTGACATCTGAGGACTCTGCGGTCTATTACTGTGCAAGA

IGHV1-12*01	Leader	GTGTCCACTCC
IGHV1-61*01 3	Leader	ATGGGATGGAGCTCTATCATCCTCTTCTTGGTAGCAACAGCTACAG
IGHV1.55.P	IGHV	ATAAACATGCATCAGTCTGGGGCTGAGCTGGTGAAGCCTGGGATCTCTGTGAAGATGTTGTGCAAGGCTTCTGATTATACATTTACTGAATACTACATGCCCTGGTCAAGCAGAATCATGGAACGATCCTTGAGTGGATTGAAAATATTAATATTTACAATTGTGGTATAACTATAATGAAAATTTCAAGGGCAAGGACACATCAACTGTAGACTATTCCTCCAGTACAGCCTATATGTTGCTTAGCAAATTGACATCTGAGGATTCTAAGTTCTATTACTGTGCAAGA
IGHV1-34*02 2	Leader	GAATCCATCTC
IGHV1-34*02 2	Leader	AATGTGTGCATACTAAGATCATTCTTAGAAATGGGAATAAAATACC
RSS Pass	RSS	CACAGTGAGTGAATGTTACTGTGAGCTCAAACACAAACC
IGHV5.1.P	IGHV	AGCTGGTGGAGTCTGGGGGAGGCTTAGTGCAGCCTGGAGGGTCCCGGAAACTCTCCTGTGCAGCCTCTGGATTCACTTTCAGTAGCTATGCCATGTCTTGGGTCCGCCAGACTCCGGAGAAGAGGCTGGAGTGGGTCGCAGCCATTAGTACTGATGGTAGTTTCATCTACTAACCAGACACTGTAAAAGGCCGATTCACCATCTCCAGAGACAATGCCAAGAACACCCTGTTTCTGCAAATGAGCAGTCTAAGGTATGAGGACACGGCCATGTATTACTGTTTGAGACA
L-Part2	Leader	GTGTGAGTTGC
L-Part1	Leader	ATGGACTGCGGGATCAGCTTGGTTTTCCTTGTCCTTATATTAAAAA
RSS Fail	RSS	CACACACACAAACTGCCAAGGGTCCTCCATTTCTCTCAG
IGHV2.1.P	IGHV	GTGTCCTGTCCCAGTTCCAGCTGAAGCAGTCAGGACCTGCCCTTGTGCCGCCCTCACGGAGCCTGTCCATCACATGCACTGTCTCTGGGTTCTCATTAACCAGCTGTGGTATAAGCTGGCTTCGATGGTCTCCAGGAAACGGAGATAGTGGAATTTTACAACTGTCCCATAGGCCTAGAAAACCCAGTCATCACCCCTGTTCCTCTTCTGCATCCTTACCAATACGTGATCCCCAATCCTTACCCACTATTCATCAAGTATATTATATTTTGCTACTGTTCCCCATTTATACAGGACTTGCCAGATC
L-Part1	Leader	ATGACTGTGCTAGAGCTGCTCCTCTGCCTGGTGGCCTTTCCAAGCT
IGHV8.7	IGHV	CAGGTTACTCTGAAAGAGTCTGGCCCTGGAATATTGCAGCGCTCCCAGACCCTCAGTCTGACTTGTTCTTTCTCTGGGTTTTCACTGAGCACTTCTGGTATGGGTGTAGGCTGGATTCGTCAGCCTTCAGGAAAGGGTCTAGAGTGGCTGGCAGACATTTGGTGGGATGACGATAAGTACTATAACCCATCCCTGAAGAGCCGGCTCACAATCTCCAAGGATACCTCCAAAAACCAGGTATTCCTCAAGAT

		CGCCAGTGTGGACACTGCAGATACTGCCACTTACTACTGTGC TCGAAGAG
IGHV8-8*01	Leader	ATGTCCTGTCC
IGHV8-8*01	Leader	ATGGGCAGGCTTACTTCTTCATTCTTGCTACTGATTGTCCCTG CAT
RSS Pass	RSS	CACAATGAGCAAAAGTTACTGTGAGCTCAAACTAAAACC
IGHV5.1	IGHV	GAGGTGCAGCTGGTGGAGTCTGGGGGAGGCTTAGTGCAGC CTGGAGAGTCCCTGAAACTCTCCTGTGAATCCAATGAATACG AATTCCCTTCCCATGACATGTCTTGGGTCCGCAAGACTCCGG AGAAGAGGCTGGAGTTGGTCGCAGCCATTAATAGTGATGGT GGTAGCACCTACTATCCAGACACCATGGAGAGACGATTCATC ATCTCCAGAGACAATACCAAGAAGACCCTGTACCTGCAAATG AGCAGTCTGAGGTCTGAGGACACAGCCTTGTATTACTGTGCA AGACA
L-Part2	Leader	TACAGTGT
L-Part1	Leader	ATGGACTTCGGGCTCAGCTTGGTTTTCCTTGTCCTTATTTTAA AAA
518355	RSS	CACAGTGTTGCAACCACATCCTGAGAGTGTCAGAAACCC
IGHV1.93	IGHV	CAGGTCCAACTGCAGCAGCCTGGGGCTGAGCTTGTGATGCC TGGGGCTTCAGTGAAGATGTCCTGCAAGGCTTCTGGCTACA CATTCACTGACTACTGGATGCACTGGGTGAAGCAGAGGCCT GGACAAGGCCTTGAGTGGATCGGAGCGATTGATACTTCTGA TAGTTATACTAGCTACAATCAAAAGTTCAAGGGCAAGGCCAC ATTGACTGTAGACGAATCCTCCAGCACAGCCTACATGCAGCT CAGCAGCCTGACATCTGAGGACTCTGCGGTCTATTACTGTGC AAGA
IGHV1-69*01	Leader	GTGTCAACTCC
IGHV1-69*01	Leader	ATGAGATGGAGCTGTATCATCCTCTTCTTGGTAGCAACAGCT ACAG
RSS Pass	RSS	CACAGTGAGGGAAGTCCATTATGAACTTGAACAAAAATT
IGHV2.1	IGHV	CAGGTGCAGCTGAAGCAGTCAGGACCTGGCCTAGTGCAGCC CTCACAGAGCCTGTCCATCACCTGCACAGTCTCTGGTTTCTC ATTAACTAGCTATGGTGTACACTGGGTTCGCCAGTCTCCAGG AAAGGGTCTGGAGTGGCTGGGAGTGATATGGAGTGGTGGAA GCACAGACTATAATGCAGCTTTCATATCCAGACTGAGCATCA GCAAGGACAATTCCAAGAGCCAAGTTTTCTTTAAAATGAACA

		GTCTGCAAGCTAATGACACAGCCATATATTACTGTGCCAGAAA
L-Part2	Leader	GTGTCCTATCC
L-Part1	Leader	ATGGCTGTCTTGGGGCTGCTCTTCTGCCTGGTGACATTCCCAAGCT
IGHV5.2.P	IGHV	CTTTGCTTTGAGGTGTCCAGTGTGAGGTGAAGCTGATAGGGTCAGTGCAGCCTGGAGGGTCCCTGAAACTCTCCTGTGCAGCCTCTGGAGTCACTGTGAGTGACTACTGAATGACCTGGGTCCTTCAGGCTCCAAAGAAGGGGCTGGAGAGGGTGGCAATAATTTTTAATGGTGGAGGTAGCACCTATTATCCAGAAACCATGAAGGGCTGATTCACTATCTACAGAGATGATGCCACAAACACACTTTACCTGAAAATAAACAGTCTGAGGTCTTAGTACACAGCC
L-Part1	Leader	ATGGACTTTGGGCTCAGCTTTGTTTTCCTTGTCCTTATTTTAAAAG
RSS Pass	RSS	CACAATGAGGAAATGTTACTGTGAGCTCAAACTAAAACC
IGHV5.2	IGHV	GAAGTGCAGCTGGTGGAGTCTGGGGGAGGCTTAGTGAAGCCTGGAGGGTCCCTGAAACTCTCCTGTGCAGCCTCTGGATTCACTTTCAGTGACTATTACATGTATTGGGTTCGCCAGACTCCGGAAAAGAGGCTGGAGTGGGTCGCAACCATTAGTGATGGTGGTAGTTACACCTACTATCCAGACAGTGTGAAGGGGCGATTCACCATCTCCAGAGACAATGCCAAGAACAACCTGTACCTGCAAATGAGCAGTCTGAAGTCTGAGGACACAGCCATGTATTACTGTGCAAGAGA
L-Part2	Leader	GTGTCCAGTGT
L-Part1	Leader	ATGAGCACTGAACACGGACCCCTCACCATGAACTTCGGGCTCAGCTTGATTTTCCTTGTCCTTGTTTTAAAAG
IGHV1.51.P	IGHV	CAGGTCCATCTGCAACAATCTGGTCCTGAGATGGTGAGGCTAGGGGCCTCAGTGAAGATATCCTGCAAGGCTTCTGGAAACACATTCACTGACTATGCTATGCACCGGGCAAAACAGTGTCATAGAAAGAGCCTATAGTAGATTGGCTTTATTAATCCTTACAATGATTATACTAGCTACAACCAGAAGTTCAAGGGGAAGGCCACAAAATCTGTAGACAAATCCTCCAGCTCAGCCTATATGCAGCTCAGCAGCCTGACATCTGAGGACTCTGCTGTCTATTACTGTGCAAGACACA
IGHV1S45*01	Leader	GTTTCCACTCC
IGHV1S45*01	Leader	ATAGGTTGTAGCTGTACCATCCTCTTCCTGGTAGCAATAGGTACAG

IGHV5.3.P	IGHV	GTGCCCAGTGTGAGGTGAAGCTAGTGGAGTCTAGGGGAGGCTTAATGCAGCCTGAAAGGTCCGTGATACTCTCCTGTGCTGCCTCTGGATGCACTGTCAGTGACTACTGGTTGGCCTGGGTTTGCCAGGCTCCAAAGAAGGGGCTGGATGGGGGGGGGCATTAATTTTTTCATGGTGGTGGTAGCACCTCCTATGCAGACACCTTGAAGAAGTGGGCTGGACATAACATATTCAGAATCAATATTTAGAGATTCTAATCCTTAAAGACATCACTTTGACCAAGTATATATGCAACATGTTATGGAG
L-Part1	Leader	ATGAACTTTGGGTTCAGCTTGGTTTTCCTTGTCATTATTTGAAAAG
IGHV1.59	IGHV	AAGGTCCAGCTGCAGCAGTCTGGAGCTGAGCTGGTGAAACCCGGGGCATCAGTGAAGCTGTCCTGCAAGGCTTCTGGCTACACCTTCACTGAGTATATTATACACTGGGTAAAGCAGAGGTCTGGACAGGGTCTTGAGTGGATTGGGTGGTTTTACCCTGGAAGTGGTAGTATAAAGTACAATGAGAAATTCAAGGACAAGGCCACATTGACTGCGGACAAATCCTCCAGCACAGTCTATATGGAGCTTAGTAGATTGACATCTGAAGACTCTGCGGTCTATTTCTGTGCAAGACACGAAGA
IGHV1-12*01	Leader	GTGTCCACTCC
IGHV1-62-2*01	Leader	ATGGAATGGTGCTGGGTCTTTCTCTTCCTCCTGTCAGTAACTGCAG
RSS Pass	RSS	CACAATGAGAGAAGTCCATTGTGAGCATTCACAAATACT
IGHV2.2	IGHV	CAGGTGCAGCTGAAGGAGTCAGGACCTGGCCTGGTGGCGCCCTCACAGAGCCTGTCCATCACATGCACTGTCTCAGGGTTCTCATTAACCAGCTATGGTGTAAGCTGGGTTCGCCAGCCTCCAGGAAAGGGTCTGGAGTGGCTGGGAGTAATATGGGGTGACGGGAGCACAAATTATCATTCAGCTCTCATATCCAGACTGAGCATCAGCAAGGATAACTCCAAGAGCCAAGTTTTCTTAAAACTGAACAGTCTGCAAACTGATGACACAGCCACGTACTACTGTGCCAAACC
L-Part2	Leader	GTGTCCTGTCC
L-Part1	Leader	ATGGCTGTCCTGGCACTGCTCCTCTGCCTGGTGACATTCCCAAGCT
IGHV5.4.P	IGHV	GTGTCAAGTGTGAGGTGAAGCTGGTGGAGTCTGTGGGAGGGTTAGAGCAGCCTGGAGGGTCCCTGAAACTCTCCTGTGCAGCCTCTGGAGTCACTGTCAGTGAGTACTGAATGACCTGGGTCCTTCAGGCTCCAAAGAAGGAGCTGGAGAGGGTGGCAATAATTTTTTAATGGTGGAGGTAGCACCTATTATCCAGAAACCATGAAGGGCCAATTCACCATCTGCAGAGATGATGCCAAAAACATACTTTACCTAGAAATAAACACTCTGTGGTCTGAGTA
L-Part1	Leader	ATGGACTTTGGGCTCAGCTTTGTTTTCCTTGTCCTTATTTTAAAAT
485460	RSS	CACATTGTCACAGCCTCAGTTATCAGCTGTACATTAATT

IGHV8.14	IGHV	CAGGTTACTCTGAAAGAGTCTGGCCCTGGGATGTTGCAGCCCTCCCAGACTCTCAGTCTGACTTGTTCTTTCTCTGGGTTTTCAATGAGCACTTTTGGTATGGGTGTAGGCTGGATTCATCAGCCTTCAGGGAAGGGACTGGAGTGGCTGGCAAACATTTGGTGGCATGATAATAAGTACTATAACCCAGCCCTGAAGAGCCGGCTCACAATCTCCAAGGATACCTCCAAAAACCAGGTATTCCTCAAGATCGCCAGTGTGGACACTGCAGATACTGCCACATACTACTGTGCTCGAATAG
IGHV8-8*01	Leader	ATGTCCTGTCC
IGHV8-8*01	Leader	ATGGGCAGGCTTACTTCTTCATTCTTGCTACTGATTGTCCCTGCAT
IGHV8.2.P	IGHV	CAGGTTACTCTGAAAGAGTCTGGCCCTGGGATATTGCAGCCCTCCCAGACCCTCAGTCTGACTTGTTCTTTCTCTGGGTTTTCACTGAGTACTTCTGGTATGGGTGTAGGCTGGATTCGTCAGCCTTCAGGGAAGGGTCTGGAGTGGCTGGCACACATTTGATGGGATGATGTCAAGCGCTATAACCCAGCCCTGAAGAGCCGACTGACAATCTCCAAGGATACCTCCAGCAGCCAGGTATTCCTCAAGATCGCCAGTGTGGACACTGCAGATACTGCCACATACTACTGTGCTCGAATAG
IGHV8-8*01	Leader	ATGTCCTGTCC
IGHV8-8*01 2	Leader	ATGGGCAGGCTTACTTCTTCATTCCTGCTACTGATTGTCCCTGCAT
RSS Pass	RSS	CACAATGAGGAAATGTTACTGTGAGCTCAAACTAAAACC
IGHV5.3	IGHV	GAGGTGCAGCTGGTGGAGTCTGGGGGAGACTTAGTGAAGCCTGGAGGGTCCCTGAAACTCTCCTGTGCAGCCTCTGGATTCACTTTCAGTAGCTATGGCATGTCTTGGGTTCGCCAGACTCCAGACAAGAGGCTGGAGTGGGTCGCAACCATTAGTAGTGGTGGTAGTTACACCTACTATCCAGACAGTGTGAAGGGGCGATTCACCATCTCCAGAGACAATGCCAAGAACACCCTGTACCTGCAAATGAGCAGTCTGAAGTCTGAGGACACAGCCATGTATTACTGTGCAAGACA
L-Part2	Leader	GTGTCCAGTGT
L-Part1	Leader	ATGAACTTCGGGCTCAGCTTGATTTTCCTTGCCCTCATTTTAAAAG
IGHV5.5.P	IGHV	GTGCCCAGTGGGAGGTGAAGCTGGTGGAGTCTAGGGGAGGCTTAGTGTAGCCTGGAAGGTCCGTGATATGCTCATGTGCAGCCTCTGGATTCACTGACAGTGACTAATAGTTGGCCTAGGTTTGCCAAGTTCCAAAGAAGGAGCTGGAATGGGGGGCATTAATTTTTCATGGTGGTGGTAGCACCTCCTATGCAGACACCTTGAAGAAGTGGGTTGGACATAACATATTCAGAATCAATATTTAGAGATTCTAATCCTTGAAGACATCACTTTTGACCAAGTATATATGCAACATGTTACTTAGG

IGHV1.59.P	IGHV	CAGGTCCATCTGCAACAATCTGGTCCTGAGATGGTGAGACTAGGGGCCTCAGTGAAGATGTCCTGCAAGGCTTCTGGAAACACATTCACTGACTATGCTATGCACTGGGTGAAACAGTGTCATGGAAAGAGCCTATAGTAGATTGGCTTTATTAATCCTTACAACGATTATACTAGCTACAACCAGAAGTTCAAGGGGAAGGCCACAAATCTGTAGACAAATCCTCCAGCTCAGCCTATATGCAGCTCAGCAGACTGACATCTGAGGACTCTGCTGTCTATTACTGTGCAAGA
L-Part1	Leader	ATGGACTTTGGGCTCAGCTTGGCTTTACTCGTCCTTATTTTAAAAG
IGHV1S45*01	Leader	GTTTCCACTCC
IGHV1S45*01 3	Leader	ATGGGTTGTAGCTGTACCATCCTCTTCCTGGTAGCAATAGGTACAG
IGHV1.60	IGHV	CAGGTCCAGCTGCAGCAGTCTGGACCTGAGCTGGTGAAGCCTGGGGCTTCAGTGAGGATATCCTGCAAGGCTTCTGGCTACACCTTCACAAGCTACTATATACACTGGGTGAAGCAGAGGCCTGGACAGGGACTTGAGTGGATTGGATGGATTTATCCTGGAAATGTTAATACTAAGTACAATGAGAAGTTCAAGGGCAAGGCCACACTGACTGCAGACAAATCCTCCAGCACAGCCTACATGCAGCTCAGCAGCCTGACCTCTGAGGACTCTGCGGTCTATTTCTGTGCAAGA
IGHV1-77*01 2	Leader	GTGTCCATTGC
IGHV1S56*01 2	Leader	ATGGGATGGAGCCGGATCTTTCTCTTCCTCCTGTCAATAATTGCAG
RSS Fail	RSS	CACTAAGGGAACTCCCATTTGAGCATAAAAAAAAATTTC
IGHV2.2.P	IGHV	GAATCGTCTCCTACTTGCATATAAAAGAGTGAGGACTTGAACTGGTGTAGCCTTCACAGACCCTGCCCCATACCTGCACTGTCTCTGGCTTCTCATTATTCCAGCTACCATGTGCACTGAGTGTGTCAGTCTACAGCAAAGGGTCCACAGTGGATGGGAGCAATGTGTAGTGGGGAAACACAGCATACGGTTCAGCTCTCAACTCCCAACTCGTCATCAATAGGGACACATCTATGGGCAAGTGTTCTTAAAACTGTACAATCTTCAAAAAGGGAAACAGTGATGTACTACTGTGCCAGAGACA
L-Part1	Leader	ATGGCTGTCCTGGTGCTGAACCTCTGCCTGGTGACATTTCCAAGCT
RSS Pass	RSS	CACACTGAGGGAAGTCCATTATGAACTTGAACAAAAATT
IGHV2.3	IGHV	CAGGTGCAGCTGAAGCAGTCAGGACCTGGCCTAGTGCAGCCCTCACAGAGCCTGTCCATCACCTGCACAGTCTCTGGTTTCTCATTAACTAGCTATGGTGTACACTGGGTTCGCCAGCCTCCAGGAAAGGGTCTGGAGTGGCTGGGAGTGATATGGAGTGGTGGAAGCACAGACTATAATGCTGCTTTCATATCCAGACTGAGCATCAGCAAGGACAACTCCAAGAGCCAAGTTTTCTTTAAAATGAACA

		GTCTGCAAGCTGATGACACAGCCATATACTACTGTGCCAGAAA
L-Part2	Leader	GTGTCCTATCC
L-Part1	Leader	ATGGCTGTCCTGGTGCTGCTCTTCTGCCTGGTGACATTCCCAAGCT
187112	RSS	CACAGTGTTGTAACCACATCCTGAGTGTGTCAGAAATCC
IGHV1.94	IGHV	CAGGTTCAGCTGCAGCAGTCTGGAGCTGAGCTGGCGAGGCCCGGGGCTTCAGTGAAGCTGTCCTGCAAGGCTTCTGGCTACACCTTCACTGACTACTATATAAACTGGGTGAAGCAGAGGACTGGACAGGGCCTTGAGTGGATTGGAGAGATTTATCCTGGAAGTGGTAATACTTACTACAATGAGAAGTTCAAGGGCAAGGCCACACTGACTGCAGACAAATCCTCCAGCACAGCCTACATGCAGCTCAGCAGCCTGACATCTGAGGACTCTGCAGTCTATTTCTGTGCAAGA
IGHV1-12*01	Leader	GTGTCCACTCC
IGHV1-77*01	Leader	ATGGAATGGATCTGGATCTTTCTCTTCATCCTCTCAGGAACTGCAG
IGHV5.6.P	IGHV	GTGCCCAGTGTGAGGTGAAGCTGGTGGAGTCTGTGGGAGGCTTAGTGCAGCCTGGAGGGTCGCTGAAACTCTCCTGGGCAGCCTCTGGATTAACTCACTGACAACTGAATGACCTGGGTCCTTCAGGCTCCAAGGAAGGGGCTGGAGAGGGTGGCAATAATTTTTAGTGGTGGAGGTAGCACCTACTATATAGACGCAATGAAGGGCCGATTCACCGTCTCCAGAGATGATAACAAAAACACACTTTACCTGAAAATAAACAGTCTGAGGTCTGAGTACACAGCC
IGHV1.61	IGHV	CAGGTCCAGCTGCAGCAGTCTGGGCCTGAGCTGGTGAGGCCTGGGGTCTCAGTGAAGATTTCCTGCAAGGGTTCCGGCTACACATTCACTGATTATGCTATGCACTGGGTGAAGCAGAGTCATGCAAAGAGTCTAGAGTGGATTGGAGTTATTAGTACTTACTATGGTAATACAAACTACAACCAGAAGTTTAAGGGCAAGGCCACAATGACTGTAGACAAATCCTCCAGCACAGCCTATATGGAACTTGCCAGATTGACATCTGAGGATTCTGCCATCTATTACTGTGCAAGA
IGHV1-67*01	Leader	GTGTGCACTCC
IGHV1-67*01	Leader	ATGGGTTGGAGCTGTATCATCTTCTTTCTGGTAGCAACAGCTACAG
RSS Pass	RSS	CACAGTGAGTGAATGTTACTGTGAGCTCAAACTAAAACC

IGHV5.4	IGHV	GAAGTGAAGCTGGTGGAGTCTGGGGGAGGCTTAGTGAAGCC TGGAGGGTCCCTGAAACTCTCCTGTGCAGCCTCTGGATTCG CTTTCAGTAGCTATGACATGTCTTGGGTTCGCCAGACTCCGG AGAAGAGGCTGGAGTGGGTCGCAACCATTAGTAGTGGTGGT AGTTACACCTACTATCCAGACAGTGTGAAGGGCCGATTCACC ATCTCCAGAGACAATGCCAGGAACACCCTGTACCTGCAAATG AGCAGTCTGAGGTCTGAGGACACGGCCTTGTATTACTGTGC AAGACA
438826	RSS	CACAGTGTTCCAACCACATCCTGAGTGTATTAGAAACCG
IGHV1.95	IGHV	CAGGTCCAACTGCAGCAGTCTGGACCTGAGCTGGTGAAGCC TGGGGCTTCAGTGAAGATATCCTGCAAGGCTTCTGGCTACAC CTTCACTAGCTATGGTATAAACTGGGTAAAGCAGAGGCCTGG ACAGGGCCTAGAGTGGATTGGATATATTTATCCTGGAAGTGG TGGTACTGCCTACAATCAGAAGTTTAAGGGCAAGGCCACACT GACTGCAGACAAGTCCTCCAGCACTGTCTACATGCAGCTCA GCAGCCTGACATCTGAGGACTCTGCGATCTATTTCTGTGCAA GA
L-Part2	Leader	TCCAGTGT
L-Part1	Leader	ATGAACTTTGGGCTGAGCTTGATTTTCCTTGTCCTAATTTTAA AAG
IGHV1-77*01 2	Leader	GTGTCCATTGC
IGHV1-77*01 2	Leader	ATGGGATGGAACTGGATATTTCTCTTCCTCCTGTCAGGAACT GCAG
IGHV5.7.P	IGHV	GTGCCCAGTGGGAGGTAAAGCTGGTGGAGTCTAGGGGAGG CTTAGTGTAGCCTGGAAGGTCCGTGATACGCTCATGTGCAG CCTCTGGATTCACTGACAGTGACTAATAGTTGGCCTAGGTTT GCCAAGTTCCAAAGAAGGAGCTGGAATGGGGGGCATTAATT TTTCATGGTGGTGGTAGCACCTCCTATGCAGACACCTTGAAG AAGTGGGTTGGACATAACATATTCAGAATCAATATTTAGAGAT TCTAATCCTTGAAGACATCACTTTTGACCAAGTATATATGCAA CATGTTACTTAGG
L-Part1	Leader	ATGGACTTTGGGCTCAGCTTGGCTTTCCTTGTCCTTATTTTAA AAG
425033	RSS	CACAGTGTTGTAACCACATCCTGAGTCTGTCAGAAACCC
IGHV1.96	IGHV	CAGGTCCAGCTGAAGCAGTCTGGAGCTGAGCTGGTGAGGCC TGGGGCTTCAGTGAAGCTGTCCTGCAAGACTTCTGGATACAT CTTCACCAGCTACTGGATTCACTGGGTAAAACAGAGGTCTGG ACAGGGCCTTGAGTGGATTGCAAGGATTTATCCTGGAACTG GTAGTACTTACTACAATGAGAAGTTCAAGGGCAAGGCCACAC TGACTGCAGACAAATCCTCCAGCACTGCCTACATGCAGCTCA GCAGCCTGAAATCTGAGGACTCTGCTGTCTATTTCTGTGCAA GA
IGHV1-46*01	Leader	GTGTCCACTGC

IGHV1-76*01	Leader	ATGGGATGGAGCTGGGTCTTTCTCTTCCTCCTGTCAGGAACTGCAG
IGHV2.3.P	IGHV	GAATCGTCTCCTACTTGCATCTGAAAGAGCGAGGACTTGAACTGGTGTAGCCTTCACAGACCCTGCCCCATACCTGCACTGTCTCTGGCTTCTCATTATTCAAGCTACCATGTGCACTGAGTGTGTCAGTCTACAGCAAAGGGTCCACAGTGGATGGGAGCAATGTGTAGTGGGGAAACACAGCATACGGTTCAGCTCTCAACTCCCAACTCGTCATCAATAGGGACACATCTATGGGCAAGTGTTCTTAAAACTGAACAATCTTCAAAAAGGAAAACAGTGATGTACTACTGTGCCAGAGA
RSS Pass	RSS	CACAGTGAGGGAAGTCCATTATGAACCTGAACAAAAATT
IGHV2.4	IGHV	CAGGTGCAGCTGAAGCAGTCAGGACCTGGCCTAGTGCAGCCCTCACAGAGCCTGTCCATAACCTGCACAGTCTCTGGTTTCTCATTAACTAGCTATGGTGTACACTGGGTTCGCCAGTCTCCAGGAAAGGGTCTGGAGTGGCTGGGAGTGATATGGAGAGGTGGAAGCACAGACTACAATGCAGCTTTCATGTCCAGACTGAGCATCACCAAGGACAACTCCAAGAGCCAAGTTTTCTTTAAAATGAACAGTCTGCAAGCTGATGACACTGCCATATACTACTGTGCCAAAAA
L-Part2	Leader	GTGTCCTGTCC
L-Part1	Leader	ATGGCTGTCCTGGTGCTGCTCCTCTGCCTGGTGACATTCCCAA
IGHV5.8.P	IGHV	CTTTGCTTTGAGGTGCCCAGTGTGAGGTGAAGCTGGTAGGGTCAGGGCAGCCTGGAGGGTCCCTGAAACACTCCTGTGCAGCCTCTGTAGTCACTGTGAGTGACTACTGAATGACCTGGGTCCTTCAGGCTCTAAAGAAGGGGCTGGAGAGGGTGGAAATAATTTTTAATGGTGGAGGTAGCACCTATTATCCAGACACCATGAAGGGCTGATTCACCATCTACAGAGATGATGCCAGAAACACACTTTACCTGAAAATAAACAGTCTGAGGTCTGAGTACACAGCC
L-Part1	Leader	ATGGACTTTGGGCTCAGGTTTGTTTTCCTTGTCCTTATTTTAAAAG
IGHV1.62	IGHV	CAGGTCCAACTGCAGCAGCCTGGGGCTGAGCTGGTGAGGCCTGGAGCTTCAGTGAAGCTGTCCTGCAAGGCTTCTGGCTACTCCTTCACCAGCTACTGGATGAACTGGGTGAAGCAGAGGCCTGGACAAGGCCTTGAGTGGATTGGCATGATTCATCCTTCCGATAGTGAAACTAGGTTAAATCAGAAGTTCAAGGACAAGGCCACATTGACTGTAGACAAATCCTCCAGCACAGCCTACATGCAACTCAGCAGCCCGACATCTGAGGACTCTGCGGTCTATTACTGTGCAAGA
IGHV1-12*01	Leader	GTGTCCACTCC
IGHV1-61*01 3	Leader	ATGGGATGGAGCTCTATCATCCTCTTCTTGGTAGCAACAGCTACAG
RSS Pass	RSS	CACAATGAGGAAATGTTACTGTGAGCTCAAACTAAAACC

IGHV5.5	IGHV	GAAGTGAAGCTGGTGGAGTCTGGGGGAGGCTTAGTGCAGC CTGGAGGGTCCCTGAAACTCTCCTGTGCAACCTCTGGATTCA CTTTCAGTGACTATTACATGTATTGGGTTCGCCAGACTCCAG AGAAGAGGCTGGAGTGGGTCGCATACATTAGTAATGGTGGT GGTAGCACCTATTATCCAGACACTGTAAAGGGCCGATTCACC ATCTCCAGAGACAATGCCAAGAACACCCTGTACCTGCAAATG AGCCGTCTGAAGTCTGAGGACACAGCCATGTATTACTGTGCA AGACA
L-Part2	Leader	GTGTCCAGTGT
L-Part1	Leader	ATGAACTTGGGGCTCAGCTTGATTTTCCTTGTCCTTGTTTTAA AAG
Relic.1	IGHV	ATATCACTTTTGACCAAGTATATATGAACCATGTTACTGAGGT TTATGGAGGTTTGAGTATGTTAGGTCCATGGATAGGGAAAAT ATTAGGGGATTTTAGGAGTAACTGAGGCTTGTTGTAGGAAGG ACATCACTGTAGGGGTAGGCTCCGTGTTCCTATCTTCAAGCT CTACCCAGTGCAGAATAGAGCCCTCTTCTGCCGGCAAGTGG AAAGAGTTTCTCTTCCTGGCTGCCTTCAATCCAAGTTGTAGAA TATCAAATCTCCTAACACTATGACTGCGAGCATGCTGACATG CGTCCCACCATAA
IGHV1.5 2.P	IGHV	CAGGTCCATCTGCAACAATCTGGTCCTGAGATGGTGAGGCT AGGGGCCTCAGTGAAGATATCCTGCAAGGCTTCTGGAAACA CATTCACTGACTATGCTATGCACCGGGCAAAACAGTGTCATA GAAAGAGCCTATAGTAGATTGGCTTTATTAATCCTTACAATGA TTATACTAGCTACAACCAGAAGTTCAAGGGGAAGGCCACAAA ATCTGTAGACAAACCCTCCAGCTCAGCCTATATGCAGCTCAG CAGCCTGACATCTGAGGACTCTGCTGTCTATTACTGTGCAAG A
IGHV1S 45*01	Leader	GTTTCCACTCC
IGHV1S 45*01	Leader	ATAGGTTGTAGCTGTACCATCCTCTTCCTGGTAGCAATAGGT ACAG
RSS Pass	RSS	CACAGTGTGGGAAGTCCAATGTGAGCCTGCACAAATACT
IGHV2.5	IGHV	CAGGTGCAGCTGAAGGAGTCAGGACCTGGCCTGGTGGCGC CCTCACAGAGCCTGTCCATCACATGCACCGTCTCAGGGTTCT CATTAACTAGCTATGGTGTACACTGGGTTCGCCAGCCTCCAG GAAAGGGTCTGGAGTGGCTGGTAGTGATATGGAGTGATGGA AGCACAACCTATAATTCAGCTCTCAAATCCAGACTGAGCATC AGCAAGGACAACTCCAAGAGCCAAGTTTTCTTAAAAATGAAC AGTCTCCAAACTGATGACACAGCCATGTACTACTGTGCCAGA AA
L-Part2	Leader	GTGTCCTGTCC
L-Part1	Leader	ATGGCTGTCCTGGGGCTGCTTCTCTGCCTGGTGACTTTCCCA AGCT

IGHV1.63	IGHV	AAGGTCCAGCTGCAGCAGTCTGGAGCTGAGCTGGTGAAACCCGGGGCATCAGTGAAGCTGTCCTGCAAGGCTTCTGGCTACACCTTCACTGAGTATATTATACACTGGGTAAAGCAGAGGTCTGGACAGGGTCTTGAGTGGATTGGGTGGTTTTACCCTGGAAGTGGTAGTATAAAGTACAATGAGAAATTCAAGGACAAGGCCACATTGACTGCGGACAAATCCTCCAGCACAGTCTATATGGAGCTTAGTAGATTGACATCTGAAGACTCTGCGGTCTATTTCTGTGCAAGACACGAAGA
IGHV1-12*01	Leader	GTGTCCACTCC
IGHV1-62-2*01	Leader	ATGGAATGGTGCTGGGTCTTTCTCTTCCTCCTGTCAGTAACTGCAG
375145	RSS	CACAGTGCTGTAACCACATCCAGAGTGTGTCTGAAACCC
IGHV1.97	IGHV	CAGGTCCAGCTGCAGCAGTCTGGAGCTGAGCTGGTAAGGCCTGGGACTTCAGTGAAGGTGTCCTGCAAGGCTTCTGGATACGCCTTCACTAATTACTTGATAGAGTGGGTAAAGCAGAGGCCTGGACAGGGCCTTGAGTGGATTGGAGTGATTAATCCTGGAAGTGGTGGTACTAACTACAATGAGAAGTTCAAGGGCAAGGCAACACTGACTGCAGACAAATCCTCCAGCACTGCCTACATGCAGCTCAGCAGCCTGACATCTGATGACTCTGCGGTTTATTTCTGTGCAAGA
IGHV1-54*01	Leader	GTGTTCACTCC
IGHV1-54*01	Leader	ATGGAATGGAGCGGAGTCTTTATCTTTCTCCTGTCAGTAACTGCAG
IGHV5.9.P	IGHV	GGTGTTCAGGGTGAGGTAGAGCTGGTGAAGTCTAATAAAGGCATAGTAATGCCTAAAAGTCCCTGAAACTCTCATGTTAAACCTCTTGATTCACATTCAGTGATTACTATATAAGTGGGGTCCTCCAGGCTCTAGGGAAGGGTCTTCAGTGGTGACCATACATTAATGGAGATAGTTCCATCAACTGTGCTGAAGCCATGAAAAGCCAATTCATCATCACCAGAGACAATGCCAAGAACAACCTCTACCTGGAAATTTGCAGACTGAGAACACAGCCATGTATCATTGTGCAGGATA
L-Part1	Leader	ATGAAAACAATATATGCCCTGACCCCTGTATCTCTGGCAGAGAAGCCAA
363211	RSS	CACAGTGTTGCAACCACATCCTGAGAGTGTCCGAAAACC
IGHV1.98	IGHV	CAGGTCCAACTCCAGCAGCCTGGGGCTGAACTGGTGAAGCCTGGGGCTTCAGTGAAGTTGTCCTGCAAGGCTTCTGGCTACACCTTCACCAGCTACTGGATGCACTGGGTGAAGCTGAGGCCTGGACAAGGCTTTGAGTGGATTGGAGAGATTAATCCTAGCAATGGTGGTACTAACTACAATGAGAAGTTCAAGAGAAAGGCCACACTGACTGTAGACAAATCCTCCAGCACAGCCTACATGCAACTCAGCAGCCTGACATCTGAGGACTCTGCGGTCTATTACTGTACAATA

IGHV1-12*01	Leader	GTGTCCACTCC
IGHV8.8	IGHV	CAGGTTACTCTGAAAGAGTCTGGCCCTGGGATATTGCAGCCCTCCCAGACCCTCAGTCTGACTTGTTCTTTCTCTGGGTTTTCACTGAGCACTTCTGGTATGGGTGTAGGCTGGATTCGTCAGCCATCAGGGAAGGGTCTGGAGTGGCTGGCACACATTTGGTGGGATGATGTCAAGCGCTATAACCCAGCCCTGAAGAGCCGACTGACTATCTCCAAGGATACCTCCAGCAGCCAGGTATTCCTCAAGATCGCCAGTGTGGACACTGCAGATACTGCCACATACTACTGTGCTCGAATAG
IGHV1S16*01	Leader	ATGGGATGGAGCTATATCATTTTCTTTTTGGTAGCAACAGCTACAG
IGHV8-8*01	Leader	ATGTCCTGTCC
IGHV8-8*01 2	Leader	ATGGGCAGGCTTACTTCTTCATTCCTGCTACTGATTGTCCCTGCAT
RSS Pass	RSS	CACAATGAGGAAATGTTACTGTGAGCTCAAACTAAAACC
IGHV5.6	IGHV	GAAGTGATGCTGGTGGAGTCTGGGGGAGGCTTAGTGAAGCCTGGAGGGTCCCTGAAACTCTCCTGTGCAGCCTCTGGATTCACTTTCAGTAGCTATGCCATGTCTTGGGTTCGCCAGACTCCGGAGAAGAGGCTGGAGTGGGTCGCAACCATTAGTAGTGGTGGTAGTTACACCTACTATCCAGACAGTGTGAAGGGGCGATTCACCATCTCCAGAGACAATGCCAAGAACACCCTGTACCTGCAAATGAGCAGTCTGAGGTCTGAGGACACGGCCATGTATTACTGTGCAAGACA
L-Part2	Leader	TCCAGTGT
L-Part1	Leader	ATGAACTTCGGGCTCAGCTTGATTTTCCTTGTCCTTGTTTTAAAAG
Relic.2	IGHV	ATATCACTTTTGACCAAGTATATATGAACCATGTTACTGAGGTTTATGGAGGTTTGAGTATGTTAGGTCCATGGATAGGGAAAATATTAGGGGATTTTAGGAGTAACTGAGGCTTGTTGTAGGAAGGACATCACTGTAGGGGTAGGCTCTGTGTTCCTATCTTCAAGCTCTACCCAGTGCAGAATAGAGCCCTCTTCTGCCTGCAAGTAGAAAGACTTTCTCTTCCTGGCTGCCTTCAATCCAAGTTGTAGAATATCAAATCTCCTAACACTATGACTGCGAGCATGCTGCCATGCGTCCCACCATAA
RSS Pass	RSS	CACAGTGTGGGAAGTCCAATGTGAGCCTGCACAAATACT
IGHV2.6	IGHV	CAGGTGCAGCTGAAGGAGTCAGGACCTGGCCTGGTGGCGCCCTCACAGAGCCTGTCCATCACATGCACCATCTCAGGGTTCTCATTAACCAGCTATGGTGTACACTGGGTTCGCCAGCCTCCAGGAAAGGGTCTGGAGTGGCTGGTAGTGATATGGAGTGATGGAAGCACAACCTATAATTCAGCTCTCAAATCCAGACTGAGCATCAGCAAGGACAACTCCAAGAGCCAAGTTTTCTTAAAAATGAAC

		AGTCTCCAAACTGATGACACAGCCATGTACTACTGTGCCAGACA
L-Part2	Leader	GTGTCCTGTCC
L-Part1	Leader	ATGGCTGTCCTGGGGCTGCTTCTCTGCCTGGTGACTTTCCCAAGCT
RSS Pass	RSS	CACAGTGAGTGAATGTTACTGTGAGCTCAAACTAAAACC
IGHV5.7	IGHV	GAAGTGAAGCTGGTGGAGTCTGGGGGAGGCTTAGTGAAGCCTGGAGGGTCCCTGAAACTCTCCTGTGCAGCCTCTGGATTCACTTTCAGTAGCTATGGCATGTCTTGGGTTCGCCAGACTCCGGAGAAGAGGCTGGAGTGGGTCGCAACCATTAGTGGTGGTGGTAGTTACACCTACTATCCAGACAGTGTGAAGGGGCGATTCACCATCTCCAGAGACAATGCCAAGAACAACCTGTACCTGCAAATGAGCAGTCTGAGGTCTGAGGACACGGCCTTGTATTACTGTGCAAGACA
L-Part2	Leader	TCCAGTGT
L-Part1	Leader	ATGAACTTCGGGCTCAGCTTGATTTTCCTTGTCCTAATTTTAAAAG
IGHV5.10.P	IGHV	GTGCCCAGTGGGAGGTGAAGCTGGTGAAGTCTAAGGGGAGGCTTAGTGTAGCCTGGAAGGTCCATGATACTCTACTGTGCAGCCTCTGGATTCACTGACAGGGACTAATAGTTTGCCTAGGTTTGCCAGGCTCCAAAGAAGGGGCTGGAATGGGGGGAATTAATTTTTCATGGTGGTGGTAGCACCTCCTATGCAGACACCTTGAAGAAGTGGGTTAGACATAACATATTCAGAATCAATATTTAGAGATTCTAATCCATGAAGACATCACTTTTGACCAAGTATATATGCAACATGTTACTTAGG
L-Part1	Leader	ATGGACTTTGGGCTCAGCTTGGCTTTCCTTGTCCTTATTTTGAAAG
327935	RSS	CACAGTATTATAGCCTGACTCTGAAGTATGTCAGAAAAT
IGHV1.60.P	IGHV	TTCAGTCCTGGTCAGCTTCTTCCTCTGCATTTTTCTGGCAAACTCTGCTCCTGACTCTTTCTCAGGAAAAGGAACTTAGCAGGAAATAGGAGATGGGGAATACAGGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGATTTCACAAAGATCAGAAGTTTGAGGACAGAACCATATTGGATGCAGATATAATGTTCAACACAGCTTGATGGAGCTCAGCAGCCTGTCATTGGAAGGCGCTGATATCTATTAAAGTGTCAGG
IGHV1-9*01	Leader	GTTACCTCTCC
IGHV1-9*01	Leader	CTGCACAATCATTTTTTCTGTTCTCTTACTATGTCTATAATACAAT

RSS Pass	RSS	CACAGTGAGAGAAGTCCATTATGAACCTAAACAAAAATT
IGHV2.4.P	IGHV	CAGGTGCAGCTGAAGCAGTCAGGACCTGGCCTAGTGCAGCCCTCACAGAGCCTGTCCATCACCTGCACATTCTCTGGTTTCTGATTAACCAGCTATGGTGTACACTGGGAGCGCCATTCTCCAGGAAAGGGTCTGGAGTGGCTGGGAGTGATATGGAGTGGTGTACACACAGACTATAATGCAGCTTTCATATCCAGATTGAGCATCAGCAAGGACAACTCCAAGAGCCAAGTTTTCTTTAAAATGAACAGTCTGCAAGCTGATGACACAGCCATATACTACTGTGCCAGAAA
L-Part2	Leader	GTGTCCTGTCC
L-Part1	Leader	GCATGGCTGTCCTGGTGCTACTCCTCTGCCTGCTGACTTTCCCAAGCT
IGHV5.11.P	IGHV	GTGCCCAGTGTGAGGTGAAGCTGGTGGAGTCTGGGGGAGGCTTAGTGCAGCATGGAGGGTCCCTGAAACTCTCCTGGGCAGCCTCTGGATTAACTCACTGACAACTAAATGACCTGTGTCATTCAGGCTCCAAGGAAGGGGCTGGAGAGGGTGGCAATAATTTTTAATGGTGGAGGTAGCACCTACTATATAGACGCAAAGAAGGGCCGATTCACCATCTCCAGAGATGATAACAAAAACACACTTTACCTGAAAATAAACAGTCTGAGGTCTGAGTACACTGCC
L-Part1	Leader	ATGGACTTTGGGCTCAGCTTTGTTTTCCTTGTCCTAATTTTAAAAG
305525	RSS	CACATTGACACAGCTTCAGTTTTCAGCTGTACAGTATTT
IGHV8.15	IGHV	CAGGTTACTCTGAAAGAGTCTGGCCCTGGGATATTGCAGCCCTCCCAGACCCTCAGTCTGACTTGTTCTTTCTCTGGGTTTTCACTGAGCACTTCTGGTATGGGTGTAGGCTGGATTCGTCAGCCTTCAGGAGAGGGTCTAGAGTGGCTGGCAGACATTTGGTGGGATGACAATAAGTACTATAACCCATCCCTGAAGAGCCGGCTCACAATCTCCAAGGATACCTCCAGCAACCAGGTATTCCTCAAGATCACCAGTGTGGACACTGCAGATACTGCCACTTACTACTGTGCTCGAAGA
IGHV8-8*01	Leader	ATGTCCTGTCC
IGHV8-8*01	Leader	ATGGGCAGGCTTACTTCTTCATTCTTGCTACTGATTGTCCCTGCAT
RSS Pass	RSS	CACAATGAGGAAATGTTACTGTGAGCTCAAACTAAAACC
IGHV5.8	IGHV	GAAGTGCAGCTGGTGGAGTCTGGGGGAGGCTTAGTGAAGCCTGGAGGGTCCCTGAAACTCTCCTGTGCAGCCTCTGGATTCACTTTCAGTAGCTATGCCATGTCTTGGGTTCGCCAGACTCCGGAGAAGAGGCTGGAGTGGGTCGCAACCATTAGTAGTGGTGGTAGTTACACCTACTATCCAGACAGTGTGAAGGGTCGATTCACCATCTCCAGAGACAATGCCAAGAACACCCTGTACCTGCAAATGAGCAGTCTGAGGTCTGAGGACACGGCCATGTATTACTGTGCAAGACA

L-Part2	Leader	TCCAGTGT
L-Part1	Leader	ATGAACTTTGTGCTCAGCTTGATTTTCCTTGCCCTCATTTTAAAAG
RSS Pass	RSS	CACACTAAGGATACTCCCATTTGAGCATAAAAAAATTTT
IGHV2.5.P	IGHV	GAATCCTCTCCTACTTGCATCTGAAAGAGTGAGGACTTGGACTGGTGTAGCCTTCACAGACCCTGCCCCATACCTGCACTGTCTCTGGCTTCTCATTATTCCAGCTACTATGTGCACTGAGTGTGTCAGTCTACAGCAAAGGGTCCACAGTGGATGGGAGCAATGTGTAGTGGGGAAGCACAGCATATGGTTCAGCTCTCAAATCCATACTCTTCATCAATAGGGACACATGTATGAGCCAAGTGTTATTAAAACTGTACAATCTTCAAAAAGGGAAACAGTCATATACTACTGTGCCAGAGA
L-Part1	Leader	ATGGCTGTCCTGGTGCTGAACCTCTGCCTGGTGAATTTCTGAGCT
RSS Pass	RSS	CACAGTGAGGGAAGTCCATTATGAACCTAAACAAAAATT
IGHV2.7	IGHV	CAGGTGCAGCTGAAGCAGTCAGGACCTGGCCTAGTGCAGCCCTCACAGAGCCTGTCCATCACCTGCACAGTCTCTGGTTTCTCATTAACTAGCTATGGTGTACACTGGGTTCGCCAGTCTCCAGGAAAGGGTCTGGAGTGGCTGGGAGTGATATGGAGTGGTGGAAGCACAGACTATAATGCAGCTTTCATATCCAGACTGAGCATCAGCAAGGACAACTCCAAGAGCCAAGTTTTCTTTAAAATGAACAGTCTGCAAGCTGATGACACAGCCATATACTACTGTGCCAGAAA
L-Part2	Leader	GTGTCCTATCC
L-Part1	Leader	ATGGCTGTCTTGGGGCTGCTCTTCTGCCTGGTGACATTCCCAAGCT
IGHV5.12.P	IGHV	GTGCCCAGTGTGAGGTGAAGCTGGTGGAGTCTGTGGGAGGCTTAGTACATCATGGAGGGTCCCTGAAACTCTCCTGGGCAGCCTCTGGATTAACTCACTGACAACTGAATGACCTGGGTCCTTCAGGCTCCAAGGAAGGGGCTGGAGAGGGTGGCAATAATTTTTAGTGGTGGAGGTAGCACCTACTATATAGACGCAATGAAGGGCCGATTCACCGTCTCCAGAGATGATAACAAAAACACACTTTACCTGAAAATAAACAGTCTGAGGTCTGAGTACACAGCC
L-Part1	Leader	ATGGACTTTGAGCTCAGCTTTGTTTTCCTTGTCCTAATTTTAAAAG
RSS Pass	RSS	CACAGTGAGTGAATGTTACTGTGAGCTCAAACTAAAACC
IGHV5.9	IGHV	GAAGTGCAGCTGGTGGAGTCTGGGGGAGGCTTAGTGAAGCCTGGAGGGTCCCTGAAACTCTCCTGTGCAGCCTCTGGATTCACTTTCAGTAGCTATGCCATGTCTTGGGTTCGCCAGTCTCCAGAGAAGAGGCTGGAGTGGGTCGCAGAAATTAGTAGTGGTGGTAGTTACACCTACTATCCAGACACTGTGACGGGCCGATTCACCATCTCCAGAGACAATGCCAAGAACACCCTGTACCTGGAAATG

		AGCAGTCTGAGGTCTGAGGACACGGCCATGTATTACTGTGC AAGGGA
L-Part2	Leader	TCCAGTGT
L-Part1	Leader	ATGAACTTCGGGCTCAGCTTGATTTTCCTTGTCCTTATTTTAA AAG
IGHV5.1 3.P	IGHV	GTGTCCAGTGTGAGGTAAAGCTGGAGAAGTCTAAGGGGAGG CTTAGTGCAGCCTGGAAGGTCCATGATACTCTACAGTGCAGC CTCTGGATTCACTGTCAGTGACGAATGGTTTGCCTGGATTGG CCAGGCTCCAAAGAAGGGGCTGTGGTGGGGGGGGGCATAA TTTTTCATGGTTGTGGTAGCCCCTCTTATGCAGACACCTTGAA GAAGTGGGTTGGACCTAAAATATTCAGAATCAATATTTAAAGA TTCTAATCCTTGAAGATGTCACTTTTGACCAAGTATATATGAA ACATGTTACTGA
L-Part1	Leader	ATGGACTTTGTGCTCAGATTGGCTTTCCTTGTCCTTATTTTAA AAG
271257	RSS	CACAATGGAGCAACCACATCCTGAGTGTGTCAGAAACCC
IGHV1.6 1.P	IGHV	CAGGTCCATCTGCAACAATCTGGTCCTGAGATGGTGAGGCT AGGGGCCTCAGTGAAGATGTCCTGCAAGGCTTCTGGAAACA CATTCACTGACTATGCTATGCACTGGGTGAAACAGTGTCATG GAAAGAGCCTATAGTAGATTGGCTTTATTAACCCTTACAATGA TTATACTAGCTACAACCAGAAGTTCAAGGGGAAGGCCACAAA ATCTGTAGACAAATCCTCCAGCTCAGCCTATATGCAGCTCAG CAACCTGACATCTGAGGACTCTGCTGTCTATTACTGTGCAAG A
IGHV1S 45*01	Leader	GTTTCCACTCC
IGHV1S 45*01 3	Leader	ATGGGTTGTAGCTGTACCATCCTCTTCCTGGTAGCAATAGGT ACAG
RSS Pass	RSS	CACAGTGAGGGAAGTCCAGTGTGAGCCTGCACAAATACT
IGHV2.8	IGHV	CAGGTGCAACTGAAGGAGTCAGGACCTGGCCTGGTGGCGC CCTCACAGAGCCTGTCCATTACCTGCACTGTCTCTGGGTTCT CATTAACCAGCTATGATATAAGCTGGATTCGCCAGCCACCAG GAAAGGGTCTGGAGTGGCTTGGAGTAATATGGACTGGTGGA GGCACAAATTATAATTCAGCTTTCATGTCCAGACTGAGCATC AGCAAGGACAACTCCAAGAGCCAAGTTTTCTTAAAAATGAAC AGTCTGCAAACTGATGACACAGCCATATATTACTGTGTAAGA GA
L-Part2	Leader	GTACCCTGTCC

L-Part1	Leader	ATGGCTGTCCTGGCACTACTCCTCTGCCTGGTGGCTTTCCCAAGCT
RSS Fail	RSS	CACAGTGAGTCCATTATGAACCTGAACAAAAATTTCACT
IGHV2.6.P	IGHV	CAGGTGCAGCTGAAGGAGTCAGGACCTGGCCTAGTGCAGCCCTCACAGAGCCTGTCCATCACCTGCACAGTCTCTGGATTCTCATTAACTAGCTATGGTGTACACTGGGTTCGCCAGTCTCCAGGAAAGGGTCTGGAGTGGCTGGGAGTGATATGGAGTGGTGGAAGCACAGACTACAATGCACCTTTCATATCCAGACTGAGCATCAGCAAGGACAACTCCAAGAGCCAAATTTTCTTTAAAATGAACAGTCTGCAAGCTGATGACACTGCCATATACTACTGTGCCAGAAA
L-Part2	Leader	GTGTCCTTTCC
L-Part1	Leader	ATGGCTGTCCTGGTGCTGCTCTTCTGCCTGGTGACATTCCCAAGCT
IGHV5.14.P	IGHV	CTCTGCTTTGAGGTGTCCAGTGTGAGGTGAAGCTGATAGGGTCAGTGCAGCCTGGAGAGTCCCTGAAACTCTCCTGTGCGGCCTCTGGAGGCACTGTCAGTGACTAGTGAAAGACCTGTGTCCTTCAGGCTCCAAAAAAGGGGCTGGAGAGGGTGGCAATAATTTTTAATGGTGGAGGTAGCACCTATTATCCAGAAACCATGAAGGGCTGATTCACCATCTACAGAGATGATAATAAAAACACACTTTACCTGAAAATATACAGTCTGAGTTCTGAGTACACAGCC
L-Part1	Leader	ATGGTCTTGGGGCTCAGCTTTGTTTTCCTTGTACTTATTTTAAAAG
187112	RSS	CACAGTGTTGTAACCACATCCTGAGTGTGTCAGAAATCC
IGHV1.99	IGHV	CAGGTTCAGCTGCAGCAGTCTGACGCTGAGTTGGTGAAACCTGGGGCTTCAGTGAAGATATCCTGCAAGGCTTCTGGCTACACCTTCACTGACCATGCTATTCACTGGGTGAAGCAGAAGCCTGAACAGGGCCTGGAATGGATTGGATATATTTCTCCCGGAAATGGTGATATTAAGTACAATGAGAAGTTCAAGGGCAAGGCCACACTGACTGCAGACAAATCCTCCAGCACTGCCTACATGCAGCTCAACAGCCTGACATCTGAGGATTCTGCAGTGTATTTCTGTAAAAGA
IGHV1-12*01	Leader	GTGTCCACTCC
IGHV1S53*02	Leader	ATGGAATGGAGCTGGGTCTTTCTCTTCTTCCTGTCAGTAACTACAG
RSS Pass	RSS	CACAGTGAGTGAATGTTACTGTGAGCTCAAACTAAAACC
IGHV5.10	IGHV	GACGTGAAGCTCGTGGAGTCTGGGGGAGGCTTAGTGAAGCTTGGAGGGTCCCTGAAACTCTCCTGTGCAGCCTCTGGATTCACTTTCAGTAGCTATTACATGTCTTGGGTTCGCCAGACTCCAGAGAAGAGGCTGGAGTTGGTCGCAGCCATTAATAGTAATGGTGGTAGCACCTACTATCCAGACACTGTGAAGGGCCGATTCACCATCTCCAGAGACAATGCCAAGAACACCCTGTACCTGCAAATGA

		GCAGTCTGAAGTCTGAGGACACAGCCTTGTATTACTGTGCAAGACA
L-Part2	Leader	GTGTCCTGTGT
L-Part1	Leader	ATGAACTTCGGGCTCAGATTGATTTTCCTTGTCCTTGTTTTAAAAG
Relic.3	IGHV	ATATCACTTTTGACCAAGTATATATGAACCATGTTACTGAGGTTTATGGAGGTTTGAGTATGTTAGGTCCATGGATAGGGAAAATATTAGGGGATTTTAGGAGTAACTGAGGCTTGTTGTAGGAAGGACATCACTGTAGGGGTAGGCTCCGTGTTCCTATCTTCAAGCTCTACCCAGTGCAGAATAGAGCCCTCTTCTGCCTGCAAGTGGAAAGAGTTTCTCTTCCTGGCTGCCTTCAATCCAAGTTGTAGAATATCAAATCTCCTAACACTATGACTGCGAGCATGCTGACATGCGTCCCACCATAA
RSS Pass	RSS	CACAGTGTGGGAAGTCCAATGTGAGCCTGCACAAATACT
IGHV2.9	IGHV	CAGGTGCAGCTGAAGGAGTCAGGACCTGACCTGGTGGCGCCCTCACAGAGCCTGTCCATCACATGCACCGTCTCAGGGTTCTCATTAACCAGCTATGGTGTACACTGGGTTCGCCAGCCTCCAGGAAAGGGTCTGGAGTGGCTGGTAGTGATATGGAGTGATGGAAGCACAACCTATAATTCAGCTCTCAAATCCAGACTGAGCATCAGCAAGGACAACTCCAAGAGCCAAGTTTTCTTAAAAATGAACAGTCTCCAAACTGATGACACAGCCATGTACTACTGTGCCAGACA
L-Part2	Leader	GTGTCCTGTCC
L-Part1	Leader	ATGGCTGTCCTGGGGCTGCTTCTCTGCCTGGTGACTTTCCCAAGCT
RSS Pass	RSS	CACAGTGAGTGAATGTTACTGTGAGCTCAAACTAAAACC
IGHV5.11	IGHV	GAAGTGATGCTGGTGGAGTCTGGGGGAGGCTTAGTGAAGCCTGGAGGGTCCCTGAAACTCTCCTGTGCAGCCTCTGGATTCACTTTCAGTAGCTATACCATGTCTTGGGTTCGCCAGACTCCGGAGAAGAGGCTGGAGTGGGTCGCAACCATTAGTAGTGGTGGTGGTAACACCTACTATCCAGACAGTGTGAAGGGTCGATTCACCATCTCCAGAGACAATGCCAAGAACAACCTGTACCTGCAAATGAGCAGTCTGAGGTCTGAGGACACGGCCTTGTATTACTGTGCAAGATA
L-Part2	Leader	TCCAGTGT
L-Part1	Leader	ATGAACTTTGGGCTGAGCTTGATTTTCCTTGTCCTAATTTTAAAAG

IGHV5.15.P	IGHV	GTGCCCAGTGGGAGGTAAAGCTGGTGGAGTCTAGGGGAGGCTTAGTGTAGCCTGGAAGGTCCGTGATACGCTCATGTGCAGCCTCTGGATTCACTGACAGTGACTAATAGTTGGCCTAGGTTTGCCAAGTTCCAAAGAAGGAGCTGGAATGGGGGGCATTAATTTTTCATGGTGGTGGTAGCACCTCCTATGCAGACACCTTGAAGAAGTGGGTTGGACATAACATATTCAGAATCAATATTTAGAGATTCTAATCCTTGAAGACATCACTTTTGACCAAGTATATATGCAACATGTTACTTAGG
L-Part1	Leader	ATGGACTTTGGGCTCAGCTTGGCTTTCCTTGTCCTTATTTTAAAAG
222855	RSS	CACAGTGTTGTAACCACATCCTGAGTGTGTCAGAAATCG
IGHV1.100	IGHV	CAGGTTCAGCTGCAGCAGTCTGGACCTGAGCTGGTGAAGCCTGGGGCTTCAGTGAAGATGTCCTGCAAGGCTTCTGGATACACATTCACTGACTATGTTATAAGCTGGGTGAAGCAGAGAACTGGACAGGGCCTTGAGTGGATTGGAGAGATTTATCCTGGAAGTGGTAGTACTTACTACAATGAGAAGTTCAAGGGCAAGGCCACACTGACTGCAGACAAATCCTCCAACACAGCCTACATGCAGCTCAGCAGCCTGACATCTGAGGACTCTGCGGTCTATTTCTGTGCAAGA
IGHV1-12*01	Leader	GTGTCCACTCC
IGHV1-81*01 2	Leader	ATGGAATGGAGGATCTTTCTCTTCATCCTGTCAGGAACTGCAGGTA
IGHV2.7.P	IGHV	GAATCGTCTCCTACTTGCATCTGAAAGAGCGAGGACTTGAACTGGTGTAGCCTTCACAGACCCTGCCCCATACCTGCACTGTCTCTGGCTTCTCATTATTCAAGCTACCATGTGCACTGAGTGTGTCAGTCTACAGCAAAGGGTCCACAGTGGATGGGAGCAATGTGTAGTGGGGAAACACAGCATACGGTTCAGCTCTCAACTCCCAACTCGTCATCAATAGGGACACATCTATGGGCAAGTGTTCTTAAAACTGAACAATCTTCAAAAAGGAAAACAGTGATGTACTACTGTGCCAGAGA
RSS Pass	RSS	CACAGTGAGGGAAGTCCATTATGAACCTGAACAAAAATT
IGHV2.10	IGHV	CAGGTGCAGCTGAAGCAGTCAGGACCTAGCCTAGTGCAGCCCTCACAGAGCCTGTCCATAACCTGCACAGTCTCTGGTTTCTCATTAACTAGCTATGGTGTACACTGGGTTCGCCAGTCTCCAGGAAAGGGTCTGGAGTGGCTGGGAGTGATATGGAGAGGTGGAAGCACAGACTACAATGCAGCTTTCATGTCCAGACTGAGCATCACCAAGGACAACTCCAAGAGCCAAGTTTTCTTTAAAATGAACAGTCTGCAAGCTGATGACACTGCCATATACTACTGTGCCAAAAA
L-Part2	Leader	GTGTCCTGTCC
L-Part1	Leader	ATGGCTGTCCTGGTGCTGCTCCTCTGCCTGGTGACATTCCCAA

IGHV5.16.P	IGHV	CTTTGCTTTGAGGTGCCCAGTGTGAGGTGAAGCTGGTAGGGTCAGGGCAGCCTGGAGGGTCCCTGAAACACTCCTGTGCAGCCTCTGTAGTCACTGTGAGTGACTACTGAATGACCTGGGTCCTTCAGGCTCTAAAGAAGGGGCTGGAGAGGGTGGAAATAATTTTTAATGGTGGAGGTAGCACCTATTATCCAGACACCATGAAGGGCTGATTCACCATCTACAGAGATGATGCCAGAAACACACTTTACCTGAAAATAAACAGTCTGAGGTCTGAGTACACAGCC
L-Part1	Leader	ATGGACTTTGGGCTCAGGTTTGTTTTCCTTGTCCTTATTTTAAAAG
	RSS	CACAGGCAGGTACCAAACCTTGACATTTGTCACTGACAC
RSS Pass	RSS	CACAATGAGGAAATGTTATTGTGAGCTCAAACTAAAACC
IGHV5.12	IGHV	GAAGTGCAGCTGGTGGAGTCTGGGGGAGGCTTAGTGAAGCCTGGAGGGTCCCTGAAACTCTCCTGTGCAGCCTCTGGATTCGCTTTCAGTAGCTATGACATGTCTTGGGTTCGCCAGACTCCGGAGAAGAGGCTGGAGTGGGTCGCATACATTAGTAGTGGTGGTGGTAGCACCTACTATCCAGACACTGTGAAGGGCCGATTCACCATCTCCAGAGACAATGCCAAGAACACCCTGTACCTGCAAATGAGCAGTCTGAAGTCTGAGGACACAGCCATGTATTACTGTGCAAGACA
L-Part2	Leader	GTGTGAAGTGT
L-Part1	Leader	ATGAACTTTGGGCTCAGATTGATTTTCCTTGTCCTTACTTTAAAAG
197154	RSS	CACAGTGTTGTAACCACATCCTGAGTGTGTCAGAAACCC
IGHV1.101	IGHV	CAGATCCAGCTGCAGCAGTCTGGACCTGAGCTGGTGAAGCCTGGGGCTTCAGTGAAGATATCCTGCAAGGCTTCTGGCTACACCTTCACTGACTACTATATAAACTGGGTGAAGCAGAAGCCTGGACAGGGACTTGAGTGGATTGGATGGATTTATCCTGGAAGCGGTAATACTAAGTACAATGAGAAGTTCAAGGGCAAGGCCACATTGACTGTAGACACATCCTCCAGCACAGCCTACATGCAGCTCAGCAGCCTGACATCTGAGGACACTGCTGTCTATTTCTGTGCAAGA
IGHV1-77*01 2	Leader	GTGTCCATTGC
IGHV1-84*02	Leader	ATGGGATGGAGCTGGATCTTTCTTTTCCTCCTGTCAGGAACTGCAG
IGHV5.17.P	IGHV	GTGCCCAATGTGAGGTGAAGCTGGTAGAGACTAGAGGAGGCTTAGGGTAGCCTGGAAGGTCCATGATACTCTCCTTTGCAGCCTCTGGAATCACTCACTGTCAGTGACTACTGGTTGGCTTGGGTTTGCCAGGCTCCAAAGAAGGAGCTGGAATGGGAGGGGGCATTAATTTTTCATGGTGGTGGTAGCACCTCCTATGCAGACACCTTGTAAAAGTGGGTTGGACCTAACATATTCAGAATCAATATTTAGAGATTCTAATCCTTGAAGACATCACTTTTGACCAAGTATATATGCAACATGTT

L-Part1	Leader	ATGGACTTTGGGCTCAGCTTTGTTTTTCTTGTCCTAATTTTAAAAG
187112	RSS	CACAGTGTTGTAACCACATCCTGAGTGTGTCAGAAATCC
IGHV1.102	IGHV	CAGGTTCAGCTGCAGCAGTCTGGAGCTGAACTGGTAAAGCCTGGGGCTTCAGTGAAGTTGTCCTGCAAGGCTTCTGGCTACACCTTCACAAGCTATGATATAAACTGGGTGAGGCAGAGGCCTGAACAGGGACTTGAGTGGATTGGATGGATTTTTCCTGGAGATGGTAGTACTAAGTACAATGAGAAGTTCAAGGGCAAGGCCACACTGACTACAGACAAATCCTCCAGCACAGCCTACATGCAGCTCAGCAGGCTGACATCTGAGGACTCTGCTGTCTATTTCTGTGCAAGA
IGHV1-12*01	Leader	GTGTCCACTCC
IGHV1-46*01	Leader	ATGGGATGGAGCTGGGTCTTTCTCTTCCTCCTGTCAGTAACTGCAG
RSS Pass	RSS	CACAATGAGGAAATGTTACTGTGAGCTCAAACTAAAACC
IGHV5.13	IGHV	GAGGTGCAGCTGGTGGAGTCTGGGGGAGGCTTAGTGCAGCCTGGAGGGTCCCTGAAACTCTCCTGTGCAGCCTCTGGATTCACTTTCAGTAGCTATGGCATGTCTTGGGTTCGCCAGACTCCAGACAAGAGGCTGGAGTTGGTCGCAACCATTAATAGTAATGGTGGTAGCACCTATTATCCAGACAGTGTGAAGGGCCGATTCACCATCTCCAGAGACAATGCCAAGAACACCCTGTACCTGCAAATGAGCAGTCTGAAGTCTGAGGACACAGCCATGTATTACTGTGCAAGAGA
L-Part2	Leader	GTGTCCAGTGT
L-Part1	Leader	ATGAACTTAGGGCTCAGCTTCATTTTCCTTGCCCTTATTTTAAAAG
177616	RSS	CACAGTGCTACAAACACATCCTGAGTGTGTCAGAAACCC
IGHV1.103	IGHV	CAGGTCCAGCTGCAGCAGTCTGGGGCTGAGCTGGTGAGGCCTGGGGCCTCAGTGAAGATTTCCTGCAAGGCTTTTGGCTACACCTTCACAAACCATCATATAAACTGGGTGAAGCAGAGGCCTGGACAGGGCCTGGACTGGATTGGATATATTAATCCTTATAATGATTATACTAGCTACAACCAGAAGTTCAAGGGCAAGGCCACATTGACTGTAGACAAATCCTCCAGCACAGCCTATATGGAGCTTAGCAGCCTGACATCTGAGGACTCTGCAGTCTATTACTGTGCAAGA
IGHV1-67*01	Leader	GTGTGCACTCC
IGHV1S45*01 5	Leader	ATGTGTTGGAGCTGTATCATCCTCTTCCTGTTAGCAACAGCTGCAC

Relic.4	IGHV	GACTTCACTTTGACCAAGTATATATGCAACATGTTACTGAGGT TTGTGTTGGTTTGAGTATGCTTGGTCCATGGAAAGGGAAACT ATTAGGGGATTTCAGGAGTAAGTGTGGCTGGTTATAGGGAG GACATCACTGTAGGGGTAGGCTCTGTGGTCCTATGTTTAAGC TCTACCCAGTGCAGAATAGAGCCCTCTCCTGTCTGCAAGTGG AAAGAGTTTCTCTTCCTGGCTGCCTTCAATTCAAGTTGTAGAG TATCAACTCTTCTAACACTAAGACTGTGAGCATGCTGACATG CTTCCCACCATAG
RSS Pass	RSS	CACAGTGAGGGAAGCCCAGTGTGAGCCTGCACAAATACT
IGHV2.11	IGHV	CAGGTGCAGAATAAGTCAGGACCTGGCCTGGTGGAGCCCTC ACAGAGCCTTTCCATCACATGCACTGTCTATTGGTTCTCGTTA ACCAGCTATGGTGTAAGCTGGGTTCGCCAGCCTCCAGGAAA GGGTCTGAAGTGGCTGGGAGTAATATGGGCTGGTGGAAGCA CAAATTATAATTCAGCTCTCATATCCAGACTGAGCATCAGCAA GGACAACTCCAAGAGCCAAGTTTTCTTAAAAATGAACAGTCT GCAAACTGATGACACAGCCATATACTACTGTGTAAGAGA
L-Part2	Leader	GTGTCCTGTCC
L-Part1	Leader	ATGGCTGACCTGGCTCTGTTCCTCTGCATGCTGACATTCCCA AACT
RSS Fail	RSS	CAACAAACCGGAATCTTCTGCGGCAAAAGCTTTATTGCT
IGHV5.18.P	IGHV	CTTTGAAGTGTGAGGTGAAGTTGTTGGAGTCTGGAGGAGGG TTAGTGCAACCTGGAGGATTCCTGAAACACCCCTGTGCAGC CTCTGGATTCACTGTCAGTGACTACTGGATGACCTGCATCCT TCAGGCTCTAAAGAAGGGGCTGGAGTGGGTGGCATTATTTTT TAATTGTAGAGGAAGTACTTATTGACCAGACATTGTAAAGGA CTGATACACCATCTCCAGAGATGATGCCAAAAACACCCTCTA CCTGAAAATAAGCAGTCTGAGGTCTGAGTATCTTACACGCAT TCGCGACCGGCCA
L-Part1	Leader	ATGGACTTTGGGTTCAGCTTGGTTTTCCTTGTTCTTATTTTAA AAG
IGHV5.19.P	IGHV	GTGTTCAGGGTGAGGTAGAGCTGGTGAAGTCTAGGAAAGGC ATAGTGATGCCTAAAAGTCCCTGAAACTCTCATGTTAAACCTC TGGATTCACATTCAGTGATACTATATAAGTGGGGTCCTCCAC CCTCTAGGGAGGGTCTTCAGTGGTGACCATACATTAATAGAG ACAGCTCCATCAACTGTGCTGAAGCTATGAAAAGCCAACTTA CCATCACCAGAGACAATGCCAAGAACAGCCTCTACCTGGAA GTTTGCAGACTGAGGAATCAGCCATGCATTATTGGGTAGGAT A
L-Part1	Leader	ATGAAAACAATACGTCCTCTGACCCCTGTATCTCTGGCATAG GAACCAA
RSS Pass	RSS	CACAATGAGGAAATGTTACTGTGAGCTCAAACTAAAACC

IGHV5.14	IGHV	GACGTGAAGCTGGTGGAGTCTGGGGGAGGCTTAGTGAAGCCTGGAGGGTCCCTGAAACTCTCCTGTGCAGCCTCTGGATTCACTTTCAGTAGCTATACCATGTCTTGGGTTCGCCAGACTCCGGAGAAGAGGCTGGAGTGGGTCGCAACCATTAGTAGTGGTGGTAGTTACACCTACTATCCAGACAGTGTGAAGGGCCGATTCACCATCTCCAGAGACAATGCCAAGAACACCCTGTACCTGCAAATGAGCAGTCTGAAGTCTGAGGACACAGCCATGTATTACTGTACAAGAGA
L-Part2	Leader	GTGTCCAGTGT
L-Part1	Leader	ATGAACTTCGGGCTCAGATTGATTTTCCTTGTCCTTACTTTAAAAG
IGHV5.20.P	IGHV	GTGCCCAATGTGAGGTGAAGCTGGTGGAGACTAGAGGAGGCTTAGTGTAGCCTGGAAGGTCCATGATACTCTCCTTTGCAGCCTCTGGAATCACTCACTGTCAGGGACTACTGGTTGGCTTGGGTTTGCCAGGCTCCAAAGAAGGAGCTGGAATGGGTGGGGGCATTAATTTTTCATGGTGGTGGTAGCACCTCCTATGCAGACACCTTGTAAAAGTGGGTTGGACCTAACATATTCAGAATCAATATTTAGAGATTCTAATCCTTGAAGACATCACTTTTGACCAAGGATATATGCAACATGTTA
L-Part1	Leader	ATGGACTTTGGGCTCAGCTTTGTTTTTCTTGTCCTAATTTTAAAAG
RSS Pass	RSS	CACAATGAGAAAAAGTTACTATGAGCTCAAACTAAAACC
IGHV5.15	IGHV	GAAGTGAAGCTGGTGGAGTCTGGGGGAGGTTTAGTGCAGCCTGGAGGGTCCCTGAAACTCTCCTGTGCAGCCTCTGGATTCACTTTCAGTAGCTATACCATGTCTTGGGTTCGCCAGACTCCAGAGAAGAGGCTGGAGTGGGTCGCATACATTAGTAATGGTGGTGGTAGCACCTACTATCCAGACACTGTAAAGGGCCGATTCACCATCTCCAGAGACAATGCCAAGAACACCCTGTACCTGCAAATGAGCAGTCTGAAGTCTGAGGACACGGCCATGTATTACTGTGCAAGACA
L-Part2	Leader	GTGTCCTGTGT
L-Part1	Leader	ATGAATTTCGGGCTCAGCTTGATTTTCCTTGTCCTTGTTTTAAAAG
IGHV5.21.P	IGHV	GTGCCCAGTGGGAGGTGAAGCTGGTGGAGTCTAGGGGAGGCTTCGTGTAGCCTGGAAGGTCCGTGATATGCTCATGTGCAGCCTCTGGATTCACTGACAGTGACTAATAGTTGGCCTAGGTTTGCCAGGTTCCAAAGAAGGAGCTGGAATGGGGGGCATTAATTTTTCATGGTGGTGGTAGCACCTCCTATGCAGACACCTTGAAGAAGTGGGTTAGACATAAAATATTCAGAATCAATATTTAGAGATTCTAATCCTTGAAGACATCACTTTTGACCAAGTATATATGCAACATGTTACTTAGG
L-Part1	Leader	ATGGACTTTGGGCTCAGCTTGGCTTTCCTTGTCCTTATTTTAAAAG
RSS Pass	RSS	CACTGTGAGTGAATGTTACTGTGTGCTCAAACTAAAACC

IGHV5.16	IGHV	GAAGTGAAGCTGGTGGAGTCTGGGGGAGGCTTAGTGAAGCCTGGAGGGTCCCTGAAACTCTCCTGTGCAGCCTCTGGATTCACTTTCAGTAGCTATGCCATGTCTTGGGTTCGCCAGACTCCAGAGAAGAGGCTGGAGTGGGTCGCATCCATTAGTAGTGGTGGTAGCACCTACTATCCAGACAGTGTGAAGGGCCGATTCACCATCTCCAGAGATAATGCCAGGAACATCCTGTACCTGCAAATGAGCAGTCTGAGGTCTGAGGACACGGCCATGTATTACTGTGCAAGAGG
L-Part2	Leader	GTGTCCAGTGT
L-Part1	Leader	ATGAACTTCGGGTTCAGCTTGATTTTCCTTGTCCTTGTTTTAAAAG
IGHV5.22.P	IGHV	GTGCCCAGTGTGAGGTGAAGCTGGTGAAGTCTAAGGGGAGGCTTAGTACAGCCTGGAAGGTCCATGATACTCTACTGTGCAGCCTCTGGATTCACTGTCAGAGATGACTGGTTTGCCTGGGTTTGCCAGGCTCCAAAGAAGGGGCTGCAGTGGGGGATGGGAATAATTTTTCATGGTTGTGGTAGCCCCTCTTATGCAGACACCTTGAAGAAGTGGGTTGGACATAACATATTCAGAATCAATATTTAAAGATTCTAATCCTTGAATATGTCACTTTTGACCAAGTATATATGCAAGATGTTTCTG
L-Part1	Leader	ATGGACTTTGTGCTCAGATTGGCTTTCCTTGTCCTTATTTTAAAAG
RSS Pass	RSS	CACAGTGAGGGAAGTCCAATGTGAGCCTGCACAAATACC
IGHV2.12	IGHV	CAGGTGCAGCTGAAGGAGTCAGGACCTGGCCTGGTGGCACCCTCACAGAGCCTGTCCATCACATGCACTGTCTCTGGGTTCTCATTATCCAGATATAGTGTACACTGGGTTCGCCAGCCTCCAGGAAAGGGTCTGGAGTGGCTGGGAATGATATGGGGTGGTGGAAGCACAGACTATAATTCAGCTCTCAAATCCAGACTGAGCATCAGCAAGGACAACTCCAAGAGCCAAGTTTTCTTAAAAATGAACAGTCTGCAAACTGATGACACAGCCATGTACTACTGTGCCAGAAA
L-Part2	Leader	GTGTCCTGTCC
L-Part1	Leader	ATGGCTGTCCTGGGGCTGCTTCTCTGCCTGGTGACGTTCCCAAGCT
RSS Pass	RSS	CACACTAAGGGAACTCCCATTTGAGCATAAAAAAAAATT
IGHV2.8.P	IGHV	AATTGTCTCCTACTTGCATATAAGAGTGAGGACTTGAACTGGTGTAGCCTTCACAGACCCTGGCCCATACCTGCACTGTCTCTGGCTTCTCATTATTCCAGCTACCATGTGCACTGAGTATGTTCGTCTACAGCAAAGGGTCCACAGTGGATGGGAGCAATGTGTAGTGGGGAAGCACTGCATACAGTTCAGCTCTCAACTCCCAACTCGTCATCAATAGGGACACATCTATGGGCAAGTGTTCTTAAAACTGAACAATCTTCAAAAAGGGAAACAGTGATGTACTACTGTGCCAGAGA
L-Part1	Leader	ATGGCTGTCCTGGTGCTGAACCTCTGCCTGGTGACATTTCCAAGCT

RSS Pass	RSS	CACAGTGAGGGAAGTCCATTATGAACTTGAACAAAAATT
IGHV2.13	IGHV	CAGGTGCAGATGAAGCAGTCAGGACCTGGCCTAGTGCAGCCCTCACAGAGCCTGTCCATCACCTGCACAGTCTCTGGTTTCTCATTAACTAGCTATGGTGTACACTGGGTTCGCCAGTCTCCAGGAAAGGGTCTGGAGTGGCTGGGAGTGATATGGAGTGGTGGAAGCACAGACTATAATGCAGCTTTCATATCCAGACTGAGCATCAGCAAGGACAATTCCAAGAGCCAAGTTTTCTTTAAAATGAACAGTCTGCAAGCTGATGACACAGCCATATACTACTGTGTCAGAAA
L-Part2	Leader	GTGTCCTGTCC
L-Part1	Leader	ATGGCTGTCCTGGAGCTGCTCCTCTCCCTGGTGACATTCCCAAGCT
IGHV5.23.P	IGHV	GTGTCCAGGGAGAGGTGAAGGTGGTGGAGTCTGGGGGACGCTTAGTGCAACCTGGAGGGTCGCTGAAACTCTCCTGGGCAGCCTCTGGAATAACTCACTGACAACTGAATGACCTGGGTCCTTCAGGCTCCAAGGAAGCAGCTGGAGAGGGTGGCAATAATTTTTAATGGTGGAGGTAGCACCTACTATACAGAAGCAATGAAGGGCCGATTCACCATCTCCAGAGATGATAACAAAAACACACTTTACCTGAAAATAAACAGTCTGAGGTCTGAGTACACAGCC
L-Part1	Leader	ATGGACTTTGGGCTCAGCTTTGTTTTCCTTGTCCTAATTTTAAAAG
RSS Fail	RSS	CACACTGAGTGAATGTTACTGTGCTGTGAGCTCAAACAA
IGHV5.24.P	IGHV	GAAGTGAAGCTGGTGGCGTCTAGGGGAGGCTTAGTGCAGCTTGGAAGGTCCCTGAAACTCTCCTGTGCAGCCTCTGGATTCACTTTCAGTGACGACTGGATGACCTGGGTTTACCAGGCTCCAAATAAAGGACTGGAGTGGGTGGCATTATTTTTCCTGGTGGTTGTAGCACCTACTCTGCGGACACCTTAAAGGGTAGATTCACCATCTCCAGAGATGATGCCAAGAATAAGGTGTACTTACAAATAACCAGTCTGGGGTCTGAGGAACCAGCCGTGTATTACTGTGAAAGATA
L-Part2	Leader	GTGTCCAGTGT
L-Part1	Leader	ATGGGCTTTGGGCTCAGATTGGTTTTCCTTGTCCTTATTTGAAAAG
IGHV5.25.P	IGHV	GTGTCCAGTGTGAGGTGAAGCTGGTGAAGTCTAAGGGAGGTTTAGTGCAGCCTGGAAGGTCCGTGATACTCTACTGTGCAGCCTCGGATCCACTGTCAGTGACGACTGCTTTGCCTGGGTTTGCCAGGCTCCAAAGAAGGGGCTGAAGGGGGGGCATAATTTTGCATGGTTGTGGTAGCCCCTCTTATGCAGATCCCTTGAAGAAGTGGGTTGGACCTAACATATTCAGAATCTATATTTAGAGTTTCTAATCCTTGAAGATGTCACTTTTGACCAAGTATATATGAAACATGTTACTGAGGTTT
L-Part1	Leader	ATGAGAAACAGAAAAATTGTATTGTTTCTCTATTTTATTTTGTTTT

RSS Pass	RSS	CACAGTGTGGGAAGTCCAATGTGAGCATTCACAAATACT
IGHV2.14	IGHV	CAGGTGCAGCTGAAGGAGTCAGGACCTGGCCTGGTGGCGCCCTCACAGAGCCTGTCCATCACATGCACTGTCTCAGGGTTCTCATTAACCGACTATGGTGTAAGCTGGATTCGCCAGCCTCCAGGAAAGGGTCTGGAGTGGCTGGGAGTAATATGGGGTGGTGGAAGCACATACTATAATTCAGCTCTCAAATCCAGACTGAGCATCAGCAAGGACAACTCCAAGAGCCAAGTTTTCTTAAAAATGAACAGTCTGCAAACTGATGACACAGCCATGTACTACTGTGCCAAACA
L-Part2	Leader	GTGTCCTGTCC
L-Part1	Leader	ATGGCTGTCCTGGGGCTGCTTCTCTGCCTGGTGACGTTCCCAAGCT
RSS Pass	RSS	CACAGTGAGAGAAGTCCAGTGTGAGCATTCACAAATACT
IGHV2.15	IGHV	CAGGTGCAGCTGAAGGAGTCAGGACCTGGCCTCGTTGCGCCCTCACAGAGCCTGTCCATCACATGCACTGTCTCTGGTTTCTCATTAACCAACTCTGGTGTACACTGGGTTCGCCAGTCTCCAGGAAAGGGTCTGGAGTGGCTGGGAGTAATATGGGGTGATGGAAGCACAAATTATAATTCAGCTTTCAAATCCAGACTGAGCATCAGCAAGGACAATTCCAAGAGTCAAGTTTTCTTAAAAATGAACAGTCTGCAAACTGATGACACAGCCAGGTACTACTGTGCCAAACC
L-Part2	Leader	GTGTCCTGTCC
L-Part1	Leader	ATGGCTGTCCTGGAGCTGCTCCTCTCCCTGGTGACATTCCCAAGCT
RSS Pass	RSS	CACAGAGAATGGATGTTGGTGTGGGCTCAGACACAACCA
IGHV5.26.P	IGHV	GTGTTCAGGGTGAGGTAGAGCTGCTGAAATACAAGGAAGGCATAGTGATGCCTAAAAGTCCCTGAAACTCTCATGTTAAACCTCTGGATTCACATTCAGTGATACTATATAAGTGGGATCCTCCAGCCTCTAGGGATGGGTCTTCAGTGGGGAGTATACATTATTAGAGACAGCTCCATCAACTGTGCTGAAGCCATGAAAAGCCAATTCACCATCACCAGAGACAATGCCAAGAACAACCTGTACCTGGATATTTGCAGGCTGAGGACACAGCCATGTTTTCTTGTGCAGGATA
L-Part1	Leader	ATGAAAACAATATATGCTCTGACCCCTGTATCTCTGGCAGAGGGACCAA
RSS Fail	RSS	CAATGAGGAAATGTTATTGTGAGCTCAAAGTAAAACCTG
IGHV5.27.P	IGHV	GAAGTGAAGCTCGTGGAGTCTGGGGGAGGCTTAGTGCAGCCTGGAGGGTCCCTGAAACTCTCCTGTGCAGCCTCTGGATTCACTTTCAGTAGCTATGCCATGTCTTGGATTCGCCAGACTCCGGACAAGAGGCTGGAGTGGGTCGCATCCATTAGTAGTGGTGGTAGTTACACCTACTATCCAGACAGTGTGAAGGGGCGATTCACCATCCCCAGAGACAATACCAAGAACACCCTGTACCTGCAAATGA

		GCAGTCTGAGTTCTAAGGACACAGCCTTGTATTACTGTGCAAGAGA
L-Part2	Leader	GTGTCCAGTGT
L-Part1	Leader	ATGAACTTCGGGTTCAGCTTGATTTTCCTTGTCCTTGTTTTAAAAG
IGHV5.28.P	IGHV	GTGCCCAGTGTGAGGTGAAGCTGGTGAAGTCTAAGGGAGGTTTAGTGCAGCCTGGAAGGTCCGTGATACTCTCCTGTGCACCCTCGGATTCACTGTCAGTGACGACTGGTTTGCCTGGGTTTGCCAGGCTCCAAAGAAGGGGCTGAAGGGGGGAGGCATAATTTTGCATGGTTGTGGTAGCCCCTCTTATGCAGATCCCTTGAAGAAGTGAGTTGGACCTAACATATTCAGAATCAATATTTAGAGTTTCTAATCCTTGAAGATGTCACTTTTGGCCAAGTATATATGAAATATGTTACTGAGGT
L-Part1	Leader	ATGAGAAACAGAAAAATTGTATTGTTTCTCTATTTTATTTTGTTTT
RSS Pass	RSS	CACAGTGAGGGAAGTCCAATGTGAGCCTGCACAAATACC
IGHV2.16	IGHV	CAGGTGCAGCTGAAGGAGTCAGGACCTGGCCTGGTGGCGCCCTCACAGAGCCTGTCCATCACATGCACCGTCTCAGGGTTCTCATTAACCGGCTATGGTGTAAACTGGGTTCGCCAGCCTCCAGGAAAGGGTCTGGAGTGGCTGGGAATGATATGGGGTGATGGAAGCACAGACTATAATTCAGCTCTCAAATCCAGACTGAGCATCAGCAAGGACAACTCCAAGAGCCAAGTTTTCTTAAAAATGAACAGTCTGCAAACTGATGACACAGCCAGGTACTACTGTGCCAGAGA
L-Part2	Leader	GTATCCTTTCC
L-Part1	Leader	ATGGCTGTCCTGGCATTACTCTTCTGCCTGGTAACATTCCCAAGCT
IGHV5.29.P	IGHV	GAGTCCACTGCGCAGTGCAGCTGGGGGAATCTGGTGGACGCTTGCTGTAGCCTCAAGAGTCCTTATAATTCTTCTGTACAGCTTATGGATTCACTTTCAGTGGTTATGTCTTGCATTAGTTCCACAAGGCTCCAGGAAAAGCCCTGGAGTGGGTCTCATCCATCAGTAAATGTCATATTAACATAAACTATGCTAATGCCCTGAAGGACATATTCACCAACATATTGTCTAGAGACTATAACAGAAAAACCGTGTATCTTTTACTGATCAGCCTGAAGGCTGAAGACACAGTCATGTGTTGCTTTCCAAGAGA
RSS Pass	RSS	CACAGTGAAAAAAAGTCCAGTGTGAACCTGAACAAAAAC

IGHV2.9.P	IGHV	GTGTCCTGTCACAAGTGCAGATGAAGGAGTCAGGACCTGACCTTGTGCAGCCATCACAGACTCTGTCTCTCACCTGCACTGTCTCTGGGTTCTCATTAAGTAGCTATGGTGTACATTGGTTTCGCAAGCCTCCGAGAAAGGGATTGGGATTGGAATGGTTGGGAGGAATATGGTCTGGTGGAAGCATATACTATAATCCAGCTCTCAGTTCCCGACTAAGTGTCAGCAGGGACATCTCTAAGAGCCAAGTTTTCTTTAAAATGAGCAGTCTGCAAAGTGAAGACACGGCTGTGTACCACTGTGCAAGATA
RSS Pass	RSS	CACAGTGAGTGAATGTTACTGTGAGGACAAGCACAAACT
IGHV5.17	IGHV	GAGGTGAAGCTGGTGGAGTCTGGGGGAGGCTTAGTGCAGCCTGGAGGGTCCCGGAAACTCTCCTGTGCAGCCTCTGGATTCACTTTCAGTGACTACGGAATGGCGTGGGTTCGACAGGCTCCAGGGAAGGGGCCTGAGTGGGTAGCATTCATTAGTAATTTGGCATATAGTATCTACTATGCAGACACTGTGACGGGCCGATTCACCATCTCTAGAGAGAATGCCAAGAACACCCTGTACCTGGAAATGAGCAGTCTGAGGTCTGAGGACACAGCCATGTACTACTGTGCAAGGGA
L-Part2	Leader	TCCAGTGT
L-Part1	Leader	ATGGACTTCAGGCTCAGCTTACTTATTTTTGTCCTTATTTTAAAAG
RSS Pass	RSS	CACAGTGAGGGAATATTATTGTGAGCTCAAACTAAAACC
IGHV5.18	IGHV	GATGTGCAGCTGGTGGAGTCTGGGGGAGGCTTAGTGCAGCCTGGAGGGTCCCGGAAACTCTCCTGTGCAGCCTCTGGATTCACTTTCAGTAGCTTTGGAATGCACTGGGTTCGTCAGGCTCCAGAGAAGGGGCTGGAGTGGGTCGCATACATTAGTAGTGGCAGTAGTACCATCTACTATGCAGACACAGTGAAGGGCCGATTCACCATCTCCAGAGACAATCCCAAGAACACCCTGTTCCTGCAAATGACCAGTCTAAGGTCTGAGGACACGGCCATGTATTACTGTGCAAGA
L-Part2	Leader	GTGTCCAGTGT
L-Part1	Leader	ATGGACTCCAGGCTCAATTTAGTTTTCCTTGTCCTTATTTTAAAAG
RSS Fail	RSS	CACCATCTGCACCCAGCAGCTAAGGCTGGTCCACAGCCC
IGHV5.30.P	IGHV	GTGTCCAGTGTGAGATGTAGCTGGTGGAGTCTAGGAGAGGCTTAGGACAACCTGGAAGATCCCTGAACTCTCATGAGTAGCCTCTGGATTCACTTTTAGTAACTATGCCATGGCCTGGGTCCAACAAGCTCCAGAGAAGGGGCTGGAGTGGGTTGTGTACATTAGTAGTGGCAGTAGTACCTTCTATTATGCAGACACAGTTAAGGGCCCATTCATCATCTCCAGAGACAATGCCAAGAACACCCTGTTCCAGCAAATGAGCAGTCTAAGGTCTGAGGACAGAGACAAAAAGATAAAAATATATGGC
L-Part1	Leader	ATGTACTTCAGGCACAGCTTGGATTTTTCTTGCCCTTATTTTAAAAG

RSS Fail	RSS	CACTGAGAAAACTTCAGTCTGAGCCTAAACAAACAAATA
IGHV2.10.P	IGHV	TACCCTGTAGAGGTACAACTGAAGTAGTCATGACCTTAACTTCTGCAATCCTCACAGTCTGTCTCTCATCTGCACTGTCTCTGGCTTCTCATTAACCATCTATGGTGTAAACTGGGTCCTCCAGCCACTAGGAAGGGGATTGCAGAGGATACCAGCAATATGGAGAGGTGAAAGCACAGAGCAGAATTCAGATCTCAAATCCTGAATCAGCATCAGTAGGGACACATCCAAGAGTCAACTTTTCTTAAAACTGAACAGAATTCAAACTGAGGACATAGCCATCTATGGCCCTGTCAGAGAAA
RSS Pass	RSS	CACAGTGAGGGAAGTCCAGTGTGAACTTGCACAAAAACC
IGHV2.17	IGHV	CAGGTGCAGCTGAAGGAGTCAGGACCTGGCCTGGTGGCGCCCTCACAGAGCCTGTCCATCACTTGCACTGTCTCTGGGTTTTCATTAACCAGCTATGGTGTACACTGGGTTCGCCAGCCTCCAGGAAAGGGTCTGGAGTGGCTGGGAGTAATATGGGCTGGTGGAAGCACAAATTATAATTCGGCTCTCATGTCCAGACTGAGCATCAGCAAAGACAACTCCAAGAGCCAAGTTTTCTTAAAAATGAACAGTCTGCAAACTGATGACACAGCCATGTACTACTGTGCCAGAGA
L-Part2	Leader	GTGTCCTGTCC
L-Part1	Leader	ATGGCTGTCCTGGTGCTGTTCCTCTGCCTGGTTGCATTTCCAAGCT
RSS Fail	RSS	CACATAATAAGTAAATGTTACTGATAGCTCAGGAAGAAA
IGHV5.31.P	IGHV	TGTTGGAGATATCCAGTGTGAGGTGCAGCTGGTGGGGTCCAGGAGAGTCTTAATGCAGACTAGAAAAACCTTGAAACTCCCCTCTGCAGCACATGGAATCACCTTCAGCAACTGACATGGTACTCCAGGCTCCAGGGAAGGGGCTGGAGTGGGATACATTCATTAATAGTGAGGGTAGTTACATCAACTATGCCAAAGCTATAAAAGAACAATTCACTATCTCCAGAGACAATACCAAGAACACTTTGTACCTGGAAATTAGCAGTTTGAAGTCTGAAAACACAGTCATTTATTACTGTACAACA
L-Part1	Leader	ATGGCTCCAGGCACCAAGAGTGGACTGGCTGGGGTGGGGG
RSS Pass	RSS	CACAGTGAGAGGACGTCATTGTGAGCCCAGACACAAACC
IGHV7.1	IGHV	GAGGTGAAGCTGGTGGAATCTGGAGGAGGCTTGGTACAGCCTGGGGGTTCTCTGAGACTCTCCTGTGCAACTTCTGGGTTCACCTTCAGTGATTTCTACATGGAGTGGGTCCGCCAGCCTCCAGGGAAGAGACTGGAGTGGATTGCTGCAAGTAGAAACAAAGCTAATGATTATACAACAGAGTACAGTGCATCTGTGAAGGGTCGGTTCATCGTCTCCAGAGACACTTCCCAAAGCATCCTCTACCTTCAGATGAATGCCCTGAGAGCTGAGGACACTGCCATTTATTACTGTGCAAGAGATGCA
L-Part2	Leader	GTATCCAGTGT

L-Part1	Leader	ATGAAGTTGTGGTTAAACTGGGTTTTTCTTTTAACACTTTTAC ATG
RSS Fail	RSS	CACACAGTGAGGGTGTGAGCCTTGAAACAAACCTCACTG
IGHV7.1. P	IGHV	GAGGTGAAGCTGGTGGAGTCTGGAGGAGGCTTGGTACAGC CTGGGGCTTCTCTGAGACTCTCCTGTGCATCTTCTGGGTTCA CCTTCACTGATTACTACATGAACTGGGTCCACCAGCCTCCAG GGAAGGCACTTGAGTAGTTGGCTTTGATTAGAAACAAAGCTA ATGGTTACATAACAGAGTACAGTGCATCTATGAAGGGTCGGT TCACCATCTCCAGAGATAATTCCCAAAGCATCCTCTATCTTCA AATGAACACACTGAGCGCACACTGAGGACAGTGCCACTTATT ACTGTGCAAGAGATA
L-Part2	Leader	GTATCCAGTGT
L-Part1	Leader	ATGTTTAACATCATTATCTTCACAGTAACACCTTTAAATG
RSS Pass	RSS	CACAGTGTTGCAACCACATCCTGAGTGTGTCAGAAACCA
IGHV14. 1	IGHV	GAGGTTCAGCTGCAGCAGTCTGGGGCTGAGCTTGTGAGGCC AGGGGCCTTAGTCAAGTTGTCCTGCAAAGCTTCTGGCTTCAA CATTAAAGACTACTATATGCACTGGGTGAAGCAGAGGCCTGA ACAGGGCCTGGAGTGGATTGGATGGATTGATCCTGAGAATG GTAATACTATATATGACCCGAAGTTCCAGGGCAAGGCCAGTA TAACAGCAGACACATCCTCCAACACAGCCTACCTGCAGCTCA GCAGCCTGACATCTGAGGACACTGCCGTCTATTACTGTGCTA GA
L-Part2	Leader	GGGTCAATTCA
L-Part1	Leader	ATGAAATGCAGCTGGGTCATCTTCTTCCTGATGGCAGTGGTT ACAG
RSS Pass	RSS	CACAGTGAGGAAATCTCAGTTTGTACCCAGACATGAACC
IGHV4.1	IGHV	GAGGTGAAGCTTCTCGAGTCTGGAGGTGGCCTGGTGCAGCC TGGAGGATCCCTGAAACTCTCCTGTGCAGCCTCAGGATTCG ATTTTAGTAGATACTGGATGAGTTGGGTCCGGCAGGCTCCAG GGAAAGGGCTAGAATGGATTGGAGAAATTAATCCAGATAGCA GTACGATAAACTATACGCCATCTCTAAAGGATAAATTCATCAT CTCCAGAGACAACGCCAAAAATACGCTGTACCTGCAAATGAG CAAAGTGAGATCTGAGGACACAGCCCTTTATTACTGTGCAAG ACC
L-Part2	Leader	GGGTCCAGTGT
L-Part1	Leader	ATGGATTTTGGGCTGATTTTTTTTATTGTTGCTCTTTTAAAAG
RSS Pass	RSS	CACAGTGTGAGGTCTTTAGTGTGAGCCCAGACATAAACC

IGHV3.1	IGHV	GATGTGCAGCTTCAGGAGTCAGGACCTGACCTGGTGAAACC TTCTCAGTCACTTTCACTCACCTGCACTGTCACTGGCTACTC CATCACCAGTGGTTATAGCTGGCACTGGATCCGGCAGTTTCC AGGAAACAAACTGGAATGGATGGGCTACATACACTACAGTG GTAGCACTAACTACAACCCATCTCTCAAAAGTCGAATCTCTAT CACTCGAGACACATCCAAGAACCAGTTCTTCCTGCAGTTGAA TTCTGTGACTACTGAGGACACAGCCACATATTACTGTGCAAG A
L-Part2	Leader	GTATCCTGTCT
L-Part1	Leader	ATGAGAGTGCTGATTCTTTTGTGCCTGTTCACAGCCTTTCCT G
Relic.5	IGHV	TGTTTTCAGCCATACCATATGAGATACCACTTGTAGAATCTGA AGGCAGCTTGTTATAGCCTGGAGTGTTCATTAAATTCTCTTCT GAAGCCACTGGATCCACCTTCAGTGATTACAGGATTCGCCAA GCCCCAGGGAAGGGGCTAGAGTGGGTCAGCAGATATAAAAT AATACTGAAGTGTGAAAACTAAGGAGAATCTGTGAAGTATAG ATTTGCCATCTACAGGAACAATTCTAAGAGCTTATTGTATCTA CAAATAAACAAATCTAAGAAGTGAGGACAGTGCCATGTGTTA CTGTGTGAGAGA
RSS Pass	RSS	CACAATGAATGGACTTCCCTGTGAAACTAGACATAAACC
IGHV11.1	IGHV	GAAGTGCAGCTGTTGGAGACTGGAGGAGGCTTGGTGCAACC TGGGGGGTCACGGGGACTCTCTTGTGAAGGCTCAGGGTTCA CTTTTAGTGGCTTCTGGATGAGCTGGGTTCGACAGACACCTG GGAAGACCGTGGAGTGGATTGGAGACATTAATTCTGATGGC AGTGCAATAAACTACGCACCATCCATAAAGGATCGATTCACT ATCTTCAGAGACAATGACAAGAGCACCCTGTACCTGCAGATG AGCAATGTGCGATCTGAGGACCCAGCCACGTATTTCTGTATG AGATA
L-Part2	Leader	ATGTCCAGTGT
L-Part1	Leader	ATGGAGTGGGAACTGGGCTTAATTTTCATTTTTGCTCTTTTAA AAG
RSS Pass	RSS	CACAGTGTTGCAACCACATCCTGAGTGTGTCAGAAACCA
IGHV14.1.P	IGHV	TAGGTTAAGCTGCAGCAGTCTGGGGCAGAGCTTGTGAAGCC AGGGGCCTCAGTCAAGTTGTCCTGCAAAGCTTCTGGCTTCAA CATTAAAGACTACTATATGCACTGAGTGAAGCAGAGGCCTGA ACAGGGCCTGGAGTGGATTGGAAGGATTGATCCTGAGGATG GTGAAACTAAATATGCCCCGAAGTTCCAGGGCAAGGCCACT ATAACAGCAGACACATCCTCCAACACAGCCTACCTGCAGCTC AGCAGCCTGACATCTGAGGACACTGCCGTCTATTACTGTGCT AGA
L-Part2	Leader	GGGTCAATTCA
L-Part1	Leader	ATGAAATGCAGCTGGGTCATCTTCTTCCTGATGGCAGTGGTT ACAG

RSS Pass	RSS	CACAGTGAGGAAATCTCAGTTTGTACCCAGACATGAACC
IGHV4.2	IGHV	GAGGTGAAGCTTCTCGAGTCTGGAGGTGGCCTGGTGCAGCC TGGAGGATCCCTGAATCTCTCCTGTGCAGCCTCAGGATTCGA TTTTAGTAGATACTGGATGAGTTGGGCTCGGCAGGCTCCAG GGAAAGGGCAGGAATGGATTGGAGAAATTAATCCAGGAAGC AGTACGATAAACTATACGCCATCTCTAAAGGATAAATTCATCA TCTCCAGAGACAACGCCAAAAATACGCTGTACCTGCAAATGA GCAAAGTGAGATCTGAGGACACAGCCCTTTATTACTGTGCAA GACT
L-Part2	Leader	GGGTCCAGTGT
L-Part1	Leader	ATGGATTTTGGGCTGATTTTTTTTATTGTTGCTCTTTTAAAAG
RSS Pass	RSS	CACAGTGTGAGGTCTTTAGTGTGAGCCCAGACATAAACC
IGHV3.2	IGHV	GATGTGCAGCTTCAGGAGTCGGGACCTGGCCTGGTGAAACC TTCTCAGTCTCTGTCCCTCACCTGCACTGTCACTGGCTACTC AATCACCAGTGATTATGCCTGGAACTGGATCCGGCAGTTTCC AGGAAACAAACTGGAGTGGATGGGCTACATAAGCTACAGTG GTAGCACTAGCTACAACCCATCTCTCAAAAGTCGAATCTCTA TCACTCGAGACACATCCAAGAACCAGTTCTTCCTGCAGTTGA ATTCTGTGACTACTGAGGACACAGCCACATATTACTGTGCAA GA
L-Part2	Leader	GTATCCTGTCT
L-Part1	Leader	ATGAGAGTGCTGATTCTTTTGTGGCTGTTCACAGCCTTTCCT G
Relic.6	IGHV	TGTTTTCAGCCATACCATATGAGATACCACTTGTAGAATCTGA AGGCAGCTTGTTATAGCCTGGAGTGTTCATTAAATTCTCTTCT GAAGCCACTGGATCCACCTTCAGTGATTACTGGATTCGCCAA GCCCCAGGGAAGGGGCTAGAGTGGGTCAGCAGATATAAAAT AATACTGAAGTGTGAAAACTATGCAGAGTCTGTGAAGGATAG ATTTGCCATCTACAGGAACAATTCTAAGAGCTTATTGTATCTA CAAATAAACAAATCTAAGAAGTGAGGACACTGCCATGTGTTA CTGTGTGAGAGA
RSS Pass	RSS	CACAATGAATGGACTTCCCTGTGAAACTAGACATAAACC
IGHV11.2	IGHV	GAAGTGCAGCTGTTGGAGACTGGAGGAGGCTTGGTGCAACC TGGGGGGTCACGGGGACTCTCTTGTGAAGGCTCAGGGTTTA CTTTTAGTGGCTTCTGGATGAGCTGGGTTCGACAGACACCTG GGAAGACCCTGGAGTGGATTGGAGACATTAATTCTGATGGC AGTGCAATAAACTACGCACCATCCATAAAGGATCGATTCACT ATCTTCAGAGACAATGACAAGAGCACCCTGTACCTGCAGATG AGCAATGTGCGATCTGAGGACACAGCCACGTATTTCTGTATG AGATA
L-Part2	Leader	ATGTCCAGTGT

L-Part1	Leader	ATGGAGTGGGAACTGAGCTTAATTTTCATTTTTGCTCTTTTAAAAG
RSS Pass	RSS	CACAGTGTTGCAACCACATCCTGAGCGTGTCAGAAACCA
IGHV14.2	IGHV	GAGGTTCAGCTGCAGCAGTCTGGGGCAGAGCTTGTGAAGCCAGGGGCCTCAGTCAAGTTGTCCTGCACAGCTTCTGGCTTCAACATTAAAGACACCTATATGCACTGGGTGAAGCAGAGGCCTGAACAGGGCCTGGAGTGGATTGGAAGGATTGATCCTGCGAATGGTAATACTAAATATGACCCGAAGTTCCAGGGCAAGGCCACTATAACAGCAGACACATCCTCCAACACAGCCTACCTGCAGCTCAGCAGCCTGACATCTGAGGACACTGCCGTCTATTACTGTGCTAGA
L-Part2	Leader	GGGTCAATTCA
L-Part1	Leader	ATGAAATGCAGCTGGGTTATCTTCTTCCTGATGGCAGTGGTTACAG
Relic.7	IGHV	CTTTCAGCTGGTGAAGTATGTAGGTGTTTTATAGATGTCTGCAGGATCCATGAGACTCACTTGTGAGTCTTCTGCATCCTCATCAAGTAGCTTTGACAGGCACTGATCAAGGATAATCAAGGGAGGGACCTGCAGAGAGACTTGGGGACTACTAATGGCAATAGTGAGTACTAACAGATTCTTTCAATGGCTGACTCGCCATCTGCAAAGATGAAGAACATTGTGTGTCTGCAAGTGAGAAATCTGAGAACTGAGGGCATTGCTGTCCACTACTGTGCAAGACGTCCAG
IGHV8.1.P	IGHV	GTGTATATTATGAGTTGCAGCTTATGGAGTCTTGGGGAAGCTTGTTAAAGCTCCAGGGTTCTGTAAGACATTCTTGTGCAGCCGCTGGATTCACTTTCAGAGAATACTATATCAGATGGGTCCTGTAGATTTTAGAAAAAGATCTTGACTCCTTGATTTAATTACAAACACAGCTGGGGGGTGATTACAGAGTATGCTTCCTATGTGAGAAGGGCACTTCCCATTTCAAGAGATCATACAAAAAAATAAAAATTCATGCCTCCAAATAAACAGATTGGAGACTTCTTTTTTCCCTGAGAAGA
RSS Pass	RSS	CACAGTGTGAAAACCACATCCTGAGGGTGTCAGAAACCA
IGHV9.1	IGHV	CAGATCCAGTTGGTGCAGTCTGGACCTGAGCTGAAGAAGCCTGGAGAGACAGTCAAGATCTCCTGCAAGGCTTCTGGGTATACCTTCACAAACTATGGAATGAACTGGGTGAAGCAGGCTCCAGGAAAGGGTTTAAAGTGGATGGGCTGGATAAACACCTACACTGGAGAGCCAACATATGCTGATGACTTCAAGGGACGGTTTGCCTTCTCTTTGGAAACCTCTGCCAGCACTGCCTATTTGCAGATCAACAACCTCAAAAATGAGGACATGGCTACATATTTCTGTGCAAGA
L-Part2	Leader	GTGCCCAAGCA
L-Part1	Leader	ATGGATTGGCTGTGGAACTTGCTATTCCTGATGGCAGCTGCCCAAA
RSS Pass	RSS	CACCGTGAGGGAAATTCAGTGTGACCCCAGACATAAACT

IGHV12.1.P	IGHV	CAGATTCAGCTTAAGGAATCAGGACCTGCTGTCATCAAGCCA TCATAGTCACTATCACTCACCTGCACAGTATCTGGATTCTCCA TCACTAGTAGTGGTTTTTGCTGGCACTGGATATGCCAGCCCC CAGGAAAGGGGTTAGAGGGGATGGGGCGCATATGTTATGAA GGTTCCATATATTATAGTCCATCCCTCAAAAGTCCCAGCACC ATCTCCAGAGATACATCACTGAATAAATTCTTTATCCAGCTGA GCACTGTACCTGATGAGGACACAGACATGTACTACTATCCCA GGGAAAACCA
L-Part2	Leader	GTATTCTGTCC
L-Part1	Leader	ATGAGACTACTAGGGTTTCTCCTGTGCTTGGCAGCAGCTCTG AAAA
RSS Pass	RSS	CACAGTGTGAAAACCACATCCTGAGGGTGTCAGAAACCA
IGHV9.2	IGHV	CAGATCCAGTTGGTGCAGTCTGGACCTGAGCTGAAGAAGCC TGGAGAGACAGTCAAGATCTCCTGCAAGGCTTCTGGCTATAC CTTCACAAACTATGCAATGCACTGGGTGAAGCAGGCTCCAG GAAAGGGTTTAAAGTGGATGGGCTGGAAATACACCAACACT GGAGAGCCAACATATGGTGATGACTTCAAGGGACGGTTTGC CTTCTCTTTGGAAACCTCTGCCAGCACTGCCTATTTGCAGAT CAACAACCTCAAAAATGAGGACATGGCTACATATTTCTGTGC AAGA
L-Part2	Leader	GTGCCCAAGCA
L-Part1	Leader	ATGGATTGGGTGTGGACCTTGCTATTCCTGATGGCAGCTGCC CAAA
RSS Pass	RSS	CACAGTGAGGGACCTTTGTTTGAGCCCAGACATAAACTT
IGHV12.2.P	IGHV	GTGTTCTGTCCTAGATTCAGCTTAAGGAGTCAGGACCTGCTG TCATCAAGCCATCACAGCCACTGTCTCTCACCTGCACAGTCT CTGGATTCTCTATCACAAGTAGTAGTTATTGAGTTATTGCTGG CACTGGATCCGCCAGCCTTCAGTAAAGGGGTTAGAGTGGAT GGGATGCATATGTTATGAAGTTCAATATACAATAGTCCATCCT TCAAAACCCACAGCACCATCTCCAGAGACACATCTCTGAACA AATTCTTTATCCAGCTGAGCTCTGTGACTGATGAGGACACAG CCATGTACTACTGTTCCAGGGAAAACCA
L-Part1	Leader	ATGAGACTACTAGGGTTACTCCTGTGCTTGGCAGCAGCTCTG AAA
IGHV3.1.P	IGHV	CAGGTGCAGCTTCAGGAGTCAGGTCCTTACAGATGAAGCTT CAGAGTCACTATATTTCACCTACACAATTACTGGTTACTCCAT CACCAATAACTATGCCTGTACTTGGATCCATCAAAGCCATAG AAAGGTCCTGTGGTTGATGAGATACATGAATATGGGAGGTAG TACACAGTACAGACTCTAGAACCAAATATCCATCACTAGAGG CATGTTTAAAAGGCCTCTTCTTAACACTACTGAGCTACATCAC CTCTTAGTGCACAGCCATATATTACTGTATAAGATA
RSS Pass	RSS	CACAGTGTGAAAACCACATCCTGAGGGTGTCAGAAACCA

IGHV9.3	IGHV	CAGATCCAGTTGGTGCAGTCTGGACCTGAGCTGAAGAAGCC TGGAGAGACAGTCAAGATCTCCTGCAAGGCTTCTGGGTATAC CTTCACAAACTATGGAATGAACTGGGTGAAGCAGGCTCCAG GAAAGGGTTTAAAGTGGATGGGCTGGATAAACACCAACACT GGAGAGCCAACATATGCTGAAGAGTTCAAGGGACGGTTTGC CTTCTCTTTGGAAACCTCTGCCAGCACTGCCTATTTGCAGAT CAACAACCTCAAAAATGAGGACACGGCTACATATTTCTGTGC AAGA
L-Part2	Leader	GTATCCAAGCA
L-Part1	Leader	ATGGATTGGCTGTGGAACTTGCTATTCCTGATGGCAGCTGCC CAAA
VH12.a3.101 Regulatory (RSS Pass)	RSS	CACAGTGAGGGAACTTCATTGTGACCCCAGACATAAACT
IGHV12.1	IGHV	GTGTTCTGTCCCAGATTCAGCTTAAGGAGTCTGGACCTGCTG TCATCAAGCCATCACAGTCACTGTCTCTCACCTGCATAGTCT CTGGATTCTCCATCACAAGTAGTAGTTATTGCTGGCACTGGA TCCGCCAGCCCCCAGGAAAGGGGTTAGAGTGGATGGGGCG CATATGTTATGAAGGTTCAATATACTATAGTCCATCCATCAAA AGCCGCAGCACCATCTCCAGAGACACATCTCTGAACAAATTC TTTATCCAGCTGAGCTCTGTGACAAATGAGGACACAGCCATG TACTACTGTTCCAGGGAAAACCA
L-Part1	Leader	ATGAGACTACTAGGGTTTCTCCTGTGCTTGGCAGCAGCTCTA AAAA
IGHV3.2.P	IGHV	CCAAGTGTGCTGTTTCAGAGGCAGCTTCAGGAGTCAGGTCC TGGACAGATGAAGCTTCAGAGACACTATATTTCACCTACTGG TTACTCCATCACCAACAACTATGTCTGTACCTGGATATATCAA AGCCATAGAAAGGTCCTATAGTAGATGAGATACATAAATATG GGAGGAGTGCACAGTACAAACTCAAGAACCAAATATCCATCA CTAGAGGCATGTTCAAAAGGCCTATACTTAAGACAGCTGAAC TACATCACCTCTTAGTACACAGCCATATATTACCGGATATGAG A
RSS Pass	RSS	CACAGTGTGAAAACCACATCCTGAGGGTGTCAAAAACCA
IGHV9.4	IGHV	CAGATCCAGTTGGTGCAGTCTGGACCTGAGCTGAAGAAGCC TGGAGAGACAGTCAAGATCTCCTGCAAGGCTTCTGGGTATAC CTTCACAAACTATGGAATGAACTGGGTGAAGCAGGCTCCAG GAAAGGGTTTAAAGTGGATGGGCTGGATAAACACCTACACTG GAGAGCCAACATATGCTGATGACTTCAAGGGACGGTTTGCCT TCTCTTTGGAAACCTCTGCCAGCACTGCCTATTTGCAGATCA ACAACCTCAAAAATGAGGACACGGCTACATATTTCTGTGCAA GA
L-Part2	Leader	GTGCCCAAGCA

L-Part1	Leader	ATGGCTTGGGTGTGGACCTTGCTATTCCTGATGGCAGCTGC CCAAA
RSS Pass	RSS	CACAGTGAGGGAAATTCAGTGTGACCCCAGACATAAACA
IGHV12. 3.P	IGHV	GTATTCTGTCCCAGATTCAGCTTAAGGAGTCAGGACCTGCTG TCATCAAGCCATCACAGTCACTGTCTCTCACCTGCACAGTCT CTGAATTCTCCATCACAAGTAGTGGTTTTTGCTGGCACAGGA TATGCCAGCCCCCAGGAAAGGGGTTAGAGTGGATGGGGCG CATATGTTATGAAGGTTCCATATATTATAGTCCATCCCTCAAA AGTCGCAGCACCATCTCCATAGACACATCACTGAATAAAATC TTTATCCAGCTGAGCTCTGTAACTGATGAGGACACAGACATG TACTACTATTCCAGGGAAAACCA
L-Part1	Leader	ATGAGACTCCTAGGGTTTCTCCTGTGCTTGGCAGCAGCGCT GAAAA
RSS Pass	RSS	CATAGTGAGGGAACTTTGTGTGACCCCAGACATAAACTT
IGHV12. 4.P	IGHV	GTGTTCTGTCCAAGATTTAGCTTAAGAAGTCAGGATCTGCTC TCATCAAGCCATCACAGCCACTGTCTCTCACCTGCACAGTCT CTGGATTCTCCATTACAAGTAGTAGTTATTGCTGGCACTGGA TCCGCCAGTCCCCAGGAAAGTGGTTAGAATGGATGGGGCAC ATATGTTATGAAGGTTGACTAAACTATAGTCCATCCCTCAAAA GACGCAGCACCATCTCCAGAGACACATCTCTGAACAAATTCT TTATCCAGCTGAGCTCTGTGACTGATGAGGACACAGCCATGT ACTACTGTTCCAGGGAAAACCA
RSS Pass	RSS	CACAGTGGGAAAACCACATCCTGAGGGTGTCAAAAACCA
IGHV9.5	IGHV	CAGATCCAGTTGGTGCAGTCTGGACCTGAGCTGAAGAAGCC TGGAGAGACAGTCAAGATCTCCTGCAAGGCTTCTGGTTATAC CTTCACAGACTATTCAATGCACTGGGTGAAGCAGGCTCCAG GAAAGGGTTTAAAGTGGATGGGCTGGATAAACACTGAGACT GGTGAGCCAACATATGCAGATGACTTCAAGGGACGGTTTGC CTTCTCTTTGGAAACCTCTGCCAGCACTGCCTATTTGCAGAT CAACAACCTCAAAAATGAGGACACGGCTACATATTTCTGTGC TAGA
L-Part2	Leader	GTATCCAAGCA
L-Part1	Leader	ATGGCTTGGGTGTGGACCTTGCTATTCCTGATGGCAGCTGC CCAAA
RSS Pass	RSS	CACAGTGAGGGTACTTCAGTGTGAGCCTAGACACAAACC
IGHV7.2	IGHV	GAGGTGAAGCTGGTGGAGTCTGGAGGAGGCTTGGTACAGC CTGGGGGTTCTCTGAGACTCTCCTGTGCAACTTCTGGGTTCA CCTTCACTGATTACTACATGAGCTGGGTCCGCCAGCCTCCAG GAAAGGCACTTGAGTGGTTGGGTTTTATTAGAAACAAAGCTA ATGGTTACACAACAGAGTACAGTGCATCTGTGAAGGGTCGGT TCACCATCTCCAGAGATAATTCCCAAAGCATCCTCTATCTTCA

		AATGAACACCCTGAGAGCTGAGGACAGTGCCACTTATTACTG TGCAAGAGATA
L-Part2	Leader	GTATCCAGTGT
L-Part1	Leader	ATGAAGTTGTGGCTGAACTGGATTTTCCTTGTAACACTTTTAA ATG
IGHV15. 1.P	IGHV	TTTCTACACATGCTCACTCCATTGTTTAGCTGCAGCAGTCTGA AGCTGCACTGAGGACTCTTGGAGCTTCAGAATAGGTGCCCTT CAAATGCTGTGATATGGGTAGCTTTCCTTTTTCCTTTAGGAAT TGCATGACATAGAACCCTGAATGTAATGGAAACATAGACCCA AGCATGAGAAGTACACTCTATGGACAGAATTAACAGGGCAGA GTCACAATGGATGCAGACAAAATGTCCAAAACTGCCTACATG TAGGTCAATAATCTGACAGCTGAGGACTCTTCTATCACTGCA GAAGGAAGA
RSS Pass	RSS	CACAGTGTTGCAACCACATCCTGAGAGTGTCAGAAACCA
IGHV14. 3	IGHV	GAGGTTCAGCTGCAGCAGTCTGGGGCAGAGCTTGTGAGGTC AGGGGCCTCAGTCAAGTTGTCCTGCACAGCTTCTGGCTTCAA CATTAAAGACTACTATATGCACTGGGTGAAGCAGAGGCCTGA ACAGGGCCTGGAGTGGATTGGATGGATTGATCCTGAGAATG GTGATACTGAATATGCCCCGAAGTTCCAGGGCAAGGCCACT ATGACTGCAGACACATCCTCCAACACAGCCTACCTGCAGCTC AGCAGCCTGACATCTGAGGACACTGCCGTCTATTACTGTAAT GCA
L-Part2	Leader	GAATCAATTCA
L-Part1	Leader	ATGAAATGCAGCTGGGTCATCTTCTTCCTGATGGCAGTGGTT ATAG
RSS Pass	RSS	CACAGTGTGGAGTCTTCAGTGTGAGCCCAGACATAAACC
IGHV3.3. P	IGHV	GATGTGCAGCTTCTGGAGTCAGGACCTAGCCTGGTGAGACC TTCTCAGACACTCTCCCTTACCTGCACTGTTACTGGCTTCTCC ATCAACAGTGATTGTTACTGGATCTGAATCCGGCAGTTTCCA GGAAACAAACTGGAGTACATCGGGAACACATTCTACAGTGGT ATCACTTACTACAACCCATCTCTCGAAAGTCGAACGTACATAA CGCGTGACACATCTAAGAACCAGTTCTCACTGAAGTTGAGTT CTGTGACTACTGAGGACACAGCCACTTACTACTGTGCGAGA GA
L-Part2	Leader	GTATCCTGTCT
L-Part1	Leader	ATGAGAATGTTGAGTGTTCTGTACCTGTTGTCAGCCCTTCCT G

IGHV10.1.P	IGHV	CCATACCATATATATCATAAAAAGACTAAAAGAGCTAGACAATACTAATTCAAACGCGACAGAAAGACAAAAATAATTTTATAAAAAAATAAAAGATTCATTCAACAAATGCTTAACCCAGAATCAATGTCTCTTATTACCAGAGCAATGAAAACTAAGGTCAGAAAGATATATCACCTCATATGTGGATGTATTCAGGATAAGATACGTGTTCATGGCTGTGGCTATGTTGGCAGTTACACAAGGTTCCAGGAACAGACTGATAAATAAGGACACTACCATGTATTTCTGTGTGAGAGA
IGHV12.5.P	IGHV	TATCATCTATCACATCATCTTAAGTATACTATCATATCATCTACAAATAGTGGTGTTTTTACTTCTTCCTTTCCAACTTATATCCATTTGAACTCTTTCTACTGTCTTATTTTTCTAACTAGAAATTCAAGTACTCCATTGAATAGATAGGGAGAGAGTGAACAGCCTTGTCTAGTTATTGGTAGGTGGTACACAGCCCAAACTCAAGAACCACATCTCCATCATTAAAGGCACATTCAAAGCCCAATTCTTAACACAACTGAACTATATCACCACAAAGGACACAGCCATGCATTACTGCAAAAGA
RSS Pass	RSS	CACAGTGAGGGTACTTCAGTGTGAGCCTAGACACAAACC
IGHV7.3	IGHV	GAGGTGAAGCTGATGGAGTCTGGAGGAGGCTTGGTACAGCCTGGGGCTTCTCTGAGACTCTCCTGTGAAGCTTCTGGATTCACCTTCACTGATTACTACATGAGCTGGGTCCGCCAGCCTCCAGGGAAGTCACCTGAGTGGTTGGCTTTGATTAGAAACAAAGCTAATGGCTATACAACAGAGTATAGTGCATCTGTTAAGGGTCGGTTCACCATCTCCAGAGATAATTCTCAAAACATCCTCTATCTTCAAATGAACACCCTGAGAGCTGAGGCCAGTGCCACTTATTACTGTGCAAAAGATGTA
L-Part2	Leader	GTATCCAGTGT
L-Part1	Leader	ATGAAGTTGTGGCTGAACTGGATTCTACTTGTAGCACTTTTAAATG
IGHV3.4.P	IGHV	AATAACCAAGTGTATTTAGTATTCACCCTACTCAATTATGAAGGCAAAACTCAGACACTGTCCCTCACCTGCACTGTCACTGGGTACTCCATCAGCAGCAATAACTACTGGAAATGGATCAGGAATTCCCTAGAAATAATCTGGAGTGGATGGGGTATATAAGTTATTATGATAGCACTGGCTACAAACCATCACTTCAAAGTCAAATTTCCATCACTAGAGATAATTCCAAAAACCAGCTCTTCCTTAAGATAAACTCTGTGACTATTGAGGACACAGCCACTTATTACTACACAAGAGA
RSS Pass	RSS	CACAGTGTTGAGTCTTCAGTGTGAGCCCAGACAAAAACC
IGHV3.3	IGHV	GATGTACAGCTTCAGGAGTCAGGACCTGGCCTGGTGAAGCCTTCTCAGACAGTGTCCCTCACCTGCACTGTCACTGGCTACTCTATCACTAATGGTAATCACTGGTGGAACTGGATCCGGCAGGTTTCAGGAAACAAACTGGAGTGGATGGGGTACATAAGCTCCAGTGGTAGCACTGACAGCAATCCATCTCTCAAAAGTCAAATCTCCATCACTAGAGACACTTCCAAGAACCAGTTATTCCTGCAGT

		TGAACTCTGTGACTATTGAAGATATAGCCACATATTACTGTGCAAGAGA
L-Part2	Leader	GTATACTGTCT
L-Part1	Leader	ATGAGAGTGTTGATTCTTGTGTACCTGTTGACAGTCCTTCCTG
IGHV10.2.P	IGHV	AGGGGATAATATCTATATCATATATATCATTAAAAAGATGAAAAGAGCTAAATAATACTTATTCAAAAGTGACAAAAAGACAAAAGATGTTTCTATAAACAAAAAGACTAGGTCAACAAAGGCATAGCACAGAATCAATGCCTCTCATTACCACAGCAATGAAAATTAAGGTCAAAAAGAGATATCACCTCATATGTGGATATATTCAGGATAAGATATGTGTTCATGGCTGTTGCTGTGTTGACAGTTACACAAGGTTCCATGAGCAGACTGAGAAATAAGGACACTGCCATGTGTTACTGTTTGAGAGA
RSS Pass	RSS	CACAGTGTGGAATCTTCAGTGTGAGCCCAGATGAAAACC
IGHV3.4	IGHV	GATGTGCAGCTTCAGGAGTCAGGACCTGGTCTGGTGAAACCTTCTCAGACAGTGTCCCTCACCTGCACTGTCACTGGCATCTCCATCACCACTGGAAATTACAGATGGAGCTGGATCCGGCAGTTTCCAGGAAACAAACTGGAGTGGATAGGGTACATATACTACAGTGGTACCATTACCTACAATCCATCTCTCACAAGTCGAACCACCATCACTAGAGACACTTCCAAGAACCAATTCTTCCTGGAAATGAACTCTTTGACTGCTGAAGACACAGCCACATACTACTGTGCACGAGA
L-Part2	Leader	GTATCCTGTCT
L-Part1	Leader	ATGAAAATGTTCACTCTTCTGTACCTGTTGACAGTCGTTCCTG
RSS Pass	RSS	CACAGTGAGGGATAAACTGTGTGAACCCAAACACAAACC
IGHV13.1	IGHV	GAGGTACAGCTGGTAGAGACAGGAGGAGGCTTGGTGCAGCCTGGAAACTCTCTAAAACTTTCCTGTGCCACTTCGGGATACCCTTTTTATGACTACTGGATGGATTGGGTCCGCCACTTTCCAGAAAAGGGGCTGGAGTGGGTTGCTCGAATTGCAACAAAAACTCATAATTATGCAACGTACTATGCAGAGTCTTTGAAAGGCCGATTCATCGTCTCAAGAGATGATTCCAAAAGCAGTGCATACATGCAGATGAACAGCTTAAGAAAGGAAGACACTGCCATTTATTACTGGGCAAGAGAGA
L-Part2	Leader	GTATTCACTGT
L-Part1	Leader	ATGGCATTGGGACTGAGTTGGATTTTCCTTGTTTCCCTTTTAAAGG

IGHV5.32.P	IGHV	TACAAGGTGAGGTGCAGCTGGTAGAATCCGGAGGCAGCTTGATTCAGCTGGGGGGGTGTGTGTCGATTAATCTCTCTTGTGAAGCTTCTGGATTCACCTTCAGTAATTACTGGATTTGACAAGCCTCAAGGAAGGGGTTAGAGTTGTTAGCAAAATCAAGAGGAATTCTATTGTCACAGCATTCTGGTGTTGCCCTGCACACAAAGGAGAGAGATAAAGACCCCACGAAGCCTGATTGTTGAGTTTTTATACAGTTTTCAGAGCAGCACCCATTAGTAACAATGTTATTGGTAGAACAGTGTGACTTT
RSS Pass	RSS	CACAGTGTGGAGTCTTCACTGTGAGCCCAGACATAAACC
IGHV3.5	IGHV	GATGTACAGCTTCAGGAGTCAGGACCTGGCCTCGTGAAACCTTCTCAGTCTCTGTCTCTCACCTGCTCTGTCACTGGCTACTCCATCACCAGTGGTTATTACTGGAACTGGATCCGGCAGTTTCCAGGAAACAAACTGGAATGGATGGGCTACATAAGCTACGACGGTAGCAATAACTACAACCCATCTCTCAAAAATCGAATCTCCATCACTCGTGACACATCTAAGAACCAGTTTTTCCTGAAGTTGAATTCTGTGACTACTGAGGACACAGCTACATATTACTGTGCAAGAGA
L-Part2	Leader	GTATCCTGTCT
L-Part1	Leader	ATGAAAGTGTTGAGTCTGTTGTACCTGTTGACAGCCATTCCTG
RSS Pass	RSS	CACAGTGTGAAAACCACATCCTGAGGGTGTCAGAAACCA
IGHV9.6	IGHV	CAGATCCAGTTGGTGCAGTCTGGACCTGAGCTGAAGAAGCCTGGAGAGACAGTCAGGATCTCCTGCAAGGCTTCTGGGTATACCTTCACAACTGCTGGAATGCAGTGGGTGCAAAAGATGCCAGGAAAGGGTTTGAAGTGGATTGGCTGGATAAACACCCACTCTGGAGTGCCAAAATATGCAGAAGACTTCAAGGGACGGTTTGCCTTCTCTTTGGAAACCTCTGCCAGCACTGCATATTTACAGATAAGCAACCTCAAAAATGAGGACACGGCTACGTATTTCTGTGCGAGA
L-Part2	Leader	GTATCCAAGCA
L-Part1	Leader	ATGGAATGGCTGTGGAACTTGCTATTTCTCATGGCAGCAGCTCAAA
RSS Pass	RSS	CACAAGGTGGTGTCTTCAGTGTGAGCCCAGAAATAAACC
IGHV3.5.P	IGHV	GAGGTACAGCTCCAGGAGTCGGGACCTGACCTGGTGAAAACTTCTCAGTCACTGTCCTTCACCTGCTCTGTCACTGGCTACTCCATCACCAGTGGTTATGACTGGAACTGTATCTGGCAGTTTCCAGGAAACAAACTGTAGTGTATGAGGTACATACACAAGAGTGGTAGCACTAACTACAACCCATCTCTCAAAAGTCAAGTCTCCATCACTGGAGACACTTCCAAGAACCAGTTCTACCTGCAGTTGAATTCTGTGACTACTGAGGACACAGCCATATATTACTGTGCAAATGA
L-Part2	Leader	GTATCCTATCC

L-part1	Leader	ATGAGAGTGTTGAGTCTTGTGTACCTGTTGATAGGTGTTCCTG
IGHV5.33.P	IGHV	GGCTGGTGGAATCTGGAGGCAGCTTGATAGAGCATGAAGGGTACATTCAACTCTTTTGTCAAGCTTCTGGATTCACCCTCAGTGGTTACTGGATGCACAGGATTTGCCAAGCCCCAGGGAAGGGGCCAGAGTGGTTAGCAAATATAAAATTTGATGAGAGTGAAGGGTTTATGGGTTATCCCAGATGATTAAAACCTATGCTATGTACTACAAAAACAGCTGTGAGAATCTAGGCCTAGGTGCATTGTGGGGACATGAGAAATACATCCCAGGGATTAATATACG
RSS Pass	RSS	CACAGTGTGGAGTCTTCAGTGTAGGCCCAGACATAAACC
IGHV3.6	IGHV	GAGGTGCAGCTTCAGGAGTCAGGACCTAGCCTCGTGAAACCTTCTCAGACTCTGTCCCTCACCTGTTCTGTCACTGGCGACTCCATCACCAGTGGTTACTGGAACTGGATCCGGAAATTCCCAGGGAATAAACTTGAGTACATGGGGTACATAAGCTACAGTGGTAGCACTTACTACAATCCATCTCTCAAAAGTCGAATCTCCATCACTCGAGACACATCCAAGAACCAGTACTACCTGCAGTTGAATTCTGTGACTACTGAGGACACAGCCACATATTACTGTGCAAGATA
L-Part2	Leader	GTATCCTGTCA
L-Part1	Leader	ATGATGGTGTTAAGTCTTCTGTACCTGTTGACAGCCCTTCCGG
RSS Pass	RSS	CAGAGTGAGGAACATGAACTGTGAACTCAGACACAAAAC
IGHV6.1.P	IGHV	GAGGTTTAGCTTCTAGAATCTGGAGGTTCTTTGATATAGCCTGGAGGATCCTTTAAACTCTCTTCTAAAGTTGCTGGATTCACCTTCAATGATTACTGGATGTGCTGGTTTTGATAAGGCCCAGGGAATCGGCTAGAGTGGTTTGCAGATAAAAAATATGTGAAAAGTTTAAAAAAACATGCAAAGTCTGTGAAGGGCAGATTTGCCACTACAGAGACAATTCTAAAAACTTATTTTATCTACAAATGAACAGTCTGAGAAATGAGGACATTGCCCCTTATTACTCTATGA
L-Part2	Leader	GCTTAGAGGAT
RSS Fail	RSS	CACATCCTCACTGTGACATAAATC
IGHV1.1.P	IGHV	GTATCCATTTCCAAATCCAACTGCAACAGTCTGGGGCTGACCTGATTAAACCTGGATCCTCAATGAAGGTGTCTTGCAAGGTTCCTGGCTACAGTTTCACTAGCTACTATATGACCTGGGTAAAAGAGAGTCCTGGACAGGGTCTAGAATAGATTAGAGAAACCACCCTAATAGTTGCAGTATAAGTTATGCACAGAAGTTTCAAGGACACATCTCCATAACTAGGGACATATCCTCCAGCATGGCCTACATGGAGCTCAGCAGGCTGACCTCTGAGGACATTTCTGTTTATTGCTGTTTATTGTTATTTATTGTATGT
RSS Pass	RSS	CACACTGTGGGGCCATCAATGAGAACCCATACACAAACC

IGHV13.2	IGHV	CAGGTGCAGCTTGTAGAGACCGGGGGAGGCTTGGTGAGGC CTGGAAATTCTCTGAAACTCTCCTGTGTTACCTCGGGATTCA CTTTCAGTAACTACCGGATGCACTGGCTTCGCCAGCCTCCAG GGAAGAGGCTGGAGTGGATTGCTGTAATTACAGTCAAATCTG ATAATTATGGAGCAAATTATGCAGAGTCTGTGAAAGGCAGAT TCACTATTTCAAGAGATGATTCAAAAAGCAGTGTCTACCTGCA GATGAACAGATTAAGAGAGGAAGACACTGCCACTTATTATTG TAGTAGA
L-Part2	Leader	GTGTCCAGTGT
L-Part1	Leader	ATGGAGTTGGGACTGAGCTGGGTATTTCTTGTGGCTCTTTTG AATG
RSS Pass	RSS	CACAATGAGAAGATTCCAATGTCAACCCACACACAAACC
IGHV12.2	IGHV	CAGATGCAGCTTCAGGAGTCAGGACCTGGCCTGGTGAAACC CTCACAGTCACTCTTCCTCGCCTGCTCTATTACTGGTTTCCC CATCACCAGTGGTTACTACTGGATCTGGATCCGTCAGTCACC TGGGAAACCCCTAGAATGGATGGGGTACATCACTCATAGTG GGGAAACTTTCTACAACCCATCCCTCCAGAGCCCCATCTCCA TTACTAGAGAAACATCCAAGAACCAGTTCTTTCTGCAATTGAA CTCTGTGACCACAGAGGACACAGCCATGTATTACTGTGCAG GAGACAGA
L-Part2	Leader	GTAGCCTGTCT
L-Part1	Leader	ATGAGAGTGCTGGGATTTTTGTGCCTGGTGACAGTCCTTCCT G
IGHV3.6.P	IGHV	TTGTGAAAACTCCAGCCTGTGTCAAGTTGACACACAAAACTA GCCAGTACAGTCACCATCCTTCATCTTCAATGTCACCAGTTC CTCTATTAGCTGTTGTTACTGGTGAGGTTGGATTCATCAGCA ACTACAAAAAGGACTGCAGTGGATGGATAAAACTTATTATATA AGAACCTTCCTGCCATCCAATCCCAAAAAGCCTAATAAATAT GAAAGTAGGTATATAAAAGAATCAGTTTCCTTTGAGGTTGAG CTCTGTGATCATTGAGCAATTGGCTGCACATTACTGTGCAAG AAG
RSS Pass	RSS	CACAGTGAGAAGTCTTCATTGTGAGTCTAGACACAAACT
IGHV6.1	IGHV	GAAGTGAAGCTTGAGGAGTCTGGAGGAGGCTTGGTGCAACC TGGAGGATCCATGAAACTCTCCTGTGTTGCCTCTGGATTCAC TTTCAGTAACTACTGGATGTCCTGGGTCCGCCAGTCTCCAGA GAAGGGGCTTGAGTGGGTTGCTCAAATAAGATTGAAATCTGA TAATTATGCAACACATTATGCGGAGTCTGTGAAAGGGAGGTT CACCATATCAAGAGATGATTCCAAAAGTAGTGTCTACCTGCA AATGAACAACTTAAGGGCTGAAGACACTGGAATTTATTACTG CACAGG
L-Part2	Leader	GGGTCCAGAGT
L-Part1	Leader	ATGGACTTGAGACTGAGCTGTGCTTTTATTATTGTTCTTTTAA AAG

RSS Pass	RSS	CACAGTAAGAAGCCTTCATTGTGAATCTATCCACCAACC
IGHV6.2	IGHV	GATGTGAACCTGGAAGTGTCTGGAGGAGGCTTAGTTAAACCT GGAGGATCCATGCAACGCTCTTGTGTAGACTCTGGATTTACT TTTGTAGATGGCTGGATGGACTGGGTCTGCCAGTCTCCAGA GAAGGGTCTTGAGTGGGTTGCTGAAATTGCAAACAAAGCTAA TAATTACGCAACATATTATCCCGAGTCTGTGAAAGGCAGATT CACCATCTCAAGAGATGATTTCAAAAGTCGTGTCTACCTGCA AAAGAACAGCTTAAGAGCTGAAGATACAGGCATTTATTACTG TACAAGG
L-Part2	Leader	GTGTCCAGAGT
L-Part1	Leader	ATGTACTTGGAAATGAGCTGTGTTTTCATTGTTGTTCTCTTAA AAG
RSS Pass	RSS	CACAGTGAGAAGCCTTCATTGTGAATCTTTCCACAAATC
IGHV6.3	IGHV	GAAGTGAAAATTGAGGAGTCAGGAGGAGGCTTGGTCCAACC TGGAGGATCCATGAAACTCTCTTGTGCAGCCTCTGGATTCAC TTTCAGTGATTACAGGATGGACTGGGTCCACCACTCTACAGA GAATGGGTTGGAGTGGGTTGCTGAAATTAGAAACAAAGCTAG TAATTATGCAACATATTATGCGGAGTCTGTGAAAGGGAGGTT CACCATCTCAAGAGATGATTCCAAAAGTAGTGTCTACCTGCA AATGAACAGCTTAAGAGCTGAAGATACTGGCATTTATTACTGT AAAAGG
L-Part2	Leader	GTGTCCAGAGT
L-Part1	Leader	ATGGACTTGGAACTGAGCTGTGTTTCATTGTTGTCCTCTTAAA AG
RSS Pass	RSS	CACAGTGAGAAGCCTTTATTATGAACCTATCCACAAACC
IGHV6.4	IGHV	GAAGTGAAGCTTGAGGAGTCTGGAGGAGGCTTGGTGCAACC TGGAGGATCCATGAAACTCTCCTGTGTTGCCTCTGGATTCAC TTTCAGTAACTACTGGATGAACTGGGTCCGCCAGTCTCCAGA GAAGGGGCTTGAGTGGGTTGCTGAAATTAGATTGAAATCTAA TAATTATGCAACACATTATGCGGAGTCTGTGAAAGGGAGGTT CACCATCTCAAGAGATGATTCCAAAAGTAGTGTCTACCTGCA AATGAACAACTTAAGAGCTGAAGACACTGGCATTTATTACTGT ACCAGG
L-Part2	Leader	GTGTCCAGAGT
L-Part1	Leader	ATGTACTTGGGACTGAACTGTGTATTCATAGTTTTTCTCTTAA AAG
RSS Pass	RSS	CACAGTGAGAAGTCTTCATTGTGATCCTAGACACAAACC

IGHV6.5	IGHV	GAGGTGAAGCTGGATGAGACTGGAGGAGGCTTGGTGCAACC TGGGAGGCCCATGAAACTCTCCTGTGTTGCCTCTGGATTCAC TTTTAGTGACTACTGGATGAACTGGGTCCGCCAGTCTCCAGA GAAAGGACTGGAGTGGGTAGCACAAATTAGAAACAAACCTTA TAATTATGAAACATATTATTCAGATTCTGTGAAAGGCAGATTC ACCATCTCAAGAGATGATTCCAAAAGTAGTGTCTACCTGCAA ATGAACAACTTAAGAGCTGAAGACATGGGTATCTATTACTGT ACATGG
L-Part2	Leader	GTGTCCAGTGT
L-Part1	Leader	ATGTACTTGGGACTGAGCTGTGTATTCATTGTTTTTCTCTTAA AAG
RSS Pass	RSS	CACAGTGGTGCAACCGTGACCCACAGCTGTGCAATATTT
IGHV8.1	IGHV	CAGGTTACTCTGAAAGAGTCTGGCCCTGGGATATTGCAGCC ATCACAGACGCTTAGCCTGGCCTGTACTTTCTCTGGGATTTC ACTGAGTACTTCTGGTATGGGTTTGAGCTGGCTTCGTAAGCC CTCAGGGAAGGCTTTAGAGTGGCTGGCAAGCATTTGGAATA ATGATAACTACTACAACCCATCTTTGAAGAGCCGGCTCACAA TCTCCAAGGAGACCTCCAACAACCAAGTATTCCTTAAACTCA CCAGTGTGGACACTGCAGATTCTGCCACATACTACTGTGCTT GGAGAGAG
L-Part2	Leader	ATGTCCTATCT
L-Part1	Leader	ATGGGCAGACTTACATTCTCATTCCTGCTGCTGCTGCCTGTC CCTGCAT
RSS Pass	RSS	CACAGTGTTGTGACCACATCCTGAGTATGTCAGAAAACT
IGHV1.1	IGHV	CAGGTTCAGCTCCAGCAGTCTGGGGCTGAGCTGGCAAGACC TGGGGCTTCAGTGAAGTTGTCCTGCAAGGCTTCTGGCTACAC CTTTACTAGCTACTGGATGCAGTGGGTAAAACAGAGGCCTG GACAGGGTCTGGAATGGATTGGGGCTATTTATCCTGGAGAT GGTGATACTAGGTACACTCAGAAGTTCAAGGGCAAGGCCAC ATTGACTGCAGATAAATCCTCCAGCACAGCCTACATGCAACT CAGCAGCTTGGCATCTGAGGACTCTGCGGTCTATTACTGTGC AAGA
L-Part2	Leader	GTGTCTACTCA
L-Part1	Leader	ATGGAATGTAACTGGATACTTCCTTTTATTCTGTCAGTAACTT CAG
RSS Pass	RSS	CACAGTGTTGTGACCACATCCTGAGTGTGTCAGAAAACT
IGHV1.2	IGHV	GAGGTTCAGCTCCAGCAGTCTGGGACTGTGCTGGCAAGGCC TGGGGCTTCCGTGAAGATGTCCTGCAAGGCTTCTGGCTACA GCTTTACCAGCTACTGGATGCACTGGGTAAAACAGAGGCCT GGACAGGGTCTAGAATGGATTGGTGCTATTTATCCTGGAAAT AGTGATACTAGCTACAACCAGAAGTTCAAGGGCAAGGCCAAA CTGACTGCAGTCACATCCGCCAGCACTGCCTACATGGAGCT

		CAGCAGCCTGACAAATGAGGACTCTGCGGTCTATTACTGTACAAGA
L-Part2	Leader	GGGTCTACTCA
L-Part1	Leader	ATGGAATGTAACTGGATACTTCCTTTTATTCTGTCGGTAATTTCAG
RSS Pass	RSS	CACAGTGTGGAATCTTCAGTGTGAGCCTAGACACAAACC
IGHV10.1	IGHV	GAGGTGCAGCTTGTTGAGTCTGGTGGAGGATTGGTGCAGCCTAAAGGATCATTGAAACTCTCATGTGCCGCCTCTGGTTTCACCTTCAATACCTATGCCATGCACTGGGTCTGCCAGGCTCCAGGAAAGGGTTTGGAATGGGTTGCTCGCATAAGAAGTAAAAGTAATAATTATGCAACATATTATGCCGATTCAGTGAAAGACAGATTCACCATCTCCAGAGATGATTCACAAAGCATGCTCTATCTGCAAATGAACAACCTGAAAACTGAGGACACAGCCATGTATTACTGTGTGAGAGA
L-Part2	Leader	GTGTGCATTGT
L-Part1	Leader	ATGGTGTTGGGGCTGAAGTGGGTTTTCTTTGTTGTTTTTTATCAAG
RSS Pass	RSS	CACAGTGGTGCAACCACATCCCGACTGTGTCAGAAACCC
IGHV1.3	IGHV	CAGGTCCAGCTTCAGCAGTCTGGGGCTGAACTGGCAAAACCTGGGGCCTCAGTGAAGATGTCCTGCAAGGCTTCTGGCTACACCTTTACTAGCTACTGGATGCACTGGGTAAAACAGAGGCCTGGACAGGGTCTGGAATGGATTGGATACATTAATCCTAGCACTGGTTATACTGAGTACAATCAGAAGTTCAAGGACAAGGCCACATTGACTGCAGACAAATCCTCCAGCACAGCCTACATGCAACTGAGCAGCCTGACATCTGAGGACTCTGCAGTCTATTACTGTGCAAGA
L-Part2	Leader	GTGTCCACTCC
L-Part1	Leader	ATGGAAAGGCACTGGATCTTTCTCTTCCTGTTTTCAGTAACTGCAG
IGHV10.3.P	IGHV	GGTACAGCCTGTTGGGTCTGGTGAAGGATTGGTGCAGCCTAAAGGGTCACTGAAACACTTTTATTCAGCCTCTGGATTCATCTTCAATACATATACCATGGAATGGGTCTGCCAGGCTCCAAGAAACGGTCTGGAATGGGTTGCATGCATAAGAACTAAAAGTAATAATTATGCCACATATACTGATTCAGTGAAAGACAGATTCGCCATCTCCAGAGATGATTCTCAAAGCATGGTCAACCTGTAAATGAACAATCTGAAAACTGAGGACATAGATATGTATTAATTACAAGAGA
RSS Pass	RSS	CACTGTGTTAAAACCACATACTGATTGTGTCAAAAACCC

IGHV1.4	IGHV	CAGGATCAGCTACATCAGTCTGGAGCTGAGCTGCAGCAACC TGGGACATCAGTGAAGATGCCCTGCAAGGCTACTGGCTACA CCTTCACTAAGTATCGAATGCGTTGGGTGAGGCAGAAGCTTG GACAGGGCCTGGAATGGATTGCATCTGTTGATCCTGGAAATA GTAATACTGAATACAATCAGAAGTTCAAAGGCAAGACCTCAC TAACTGAATACAAATCCTCCAGCACAGCCTACATAGAGCTTA GCAGCCTGACCTCTGAGGACTCTGTGGTCTATTACTGTACAA GA
L-Part2	Leader	GTGTCCACTCA
L-Part1	Leader	ATGGAAGAGAGTGGTATCTTTCTCTTCCTCTTGTCAGTAACTT
RSS Pass	RSS	CACAGTGTAACAGCTCATATCTGGAGCATGTCAAAAAGT
IGHV15.1	IGHV	CAGGTTCACCTACAACAGTCTGGTTCTGAACTGAGGAGTCCT GGGTCTTCAGTAAAGCTTTCATGCAAGGATTTTGATTCAGAA GTCTTCCCTATTGCTTATATGAGTTGGGTTAGGCAGAAGCCT GGGCATGGATTTGAATGGATTGGAGACATACTCCCAAGTATT GGTAGAACAATCTATGGAGAGAAGTTTGAGGACAAAGCCACA CTGGATGCAGACACAGTGTCCAACACAGCCTACTTGGAGCT CAACAGTCTGACATCTGAGGACTCTGCTATCTACTACTGTGC AAGG
L-Part2	Leader	GTATCCAATCC
L-Part1	Leader	ATGGACTGGATTTGGATCATGCTCCATCTGCTGGCAGCAGCT ACAG
RSS Pass	RSS	CACAGTGTTGTAACCACATCCTGAGTGTGTCAGAAACTC
IGHV1.5	IGHV	CAGGTTCAGCTGCAGCAGTCTGGAGCTGAGCTGATGAAGCC TGGGGCCTCAGTGAAGATATCCTGCAAGGCTACTGGCTACA CATTCAGTAGCTACTGGATAGAGTGGGTAAAGCAGAGGCCT GGACATGGCCTTGAGTGGATTGGAGAGATTTTACCTGGAAGT GGTAGTACTAACTACAATGAGAAGTTCAAGGGCAAGGCCACA TTCACTGCAGATACATCCTCCAACACAGCCTACATGCAACTC AGCAGCCTGACATCTGAGGACTCTGCCGTCTATTACTGTGCA AGA
L-Part2	Leader	GTGTCCACTCC
L-Part1	Leader	ATGGAATGGACCTGGGTCTTTCTCTTCCTCCTGTCAGTAACT GCAG
IGHV1.2.P	IGHV	CCGGCCCAGCTGCAGCATTCTGGGGCTGACATGGAGGAGC CTGGTCCTCAGTGAAGTTTTCCTGCATGGCTTCTGGCTATAC CTTCACTGACCACTCTATTCATGTGGTAAAACAATGAAACAGA GGCCTGGATAGGGACTGGAGTGGATTGAATGGGTTGGACCT ACATGTGGTGGTACTGTATATGCTAGGAAGTTTCAAGGCAAA GCCACACTGACTGTAGACAAATCAGCCATCACAACCTACATG CAGCTCAGTAGCCTGACATCTGAAAACTCTGCAGTCTATTAC TGTGCCATGCAAGGA

RSS Pass	RSS	CACAGTGCTACAACCACATCCTGAGTGTGTCAGAAACCA
IGHV1.3.P	IGHV	CAGATTCAGCTGCAGCAGTCAGGAGCTGAGCTGGCGAGTCCTGGGGCATCAGTGACACTGTCCTGCAAGGCTTCTGGCTACACATTTACTGACCATATTATGAATTGGGTAAAAAAGAGGCCTGGACAGGGCCTTGAGTGGATTGGAAGGATTTATCCAGTAAGTGGTGAAACTAACTACAATCAAAAGTTCATGGGCAAGGCCACATTCTCTGTAGACCGGTCCTCCAGCACAGTGTACATGGTGTTGAACAGCCTGACATCTGAGGACCCTGCTGTCTATTACTGTGGAAGG
L-Part2	Leader	GTGCCTCCTCC
L-Part1	Leader	ATGGGATGGAGCCAGATCACCCTCTTTCTGGTGGCAGCAATTACAT
RSS Pass	RSS	CACAGTGTTGTGACCACATCCTGCATGTGTCAGAAACCC
IGHV1.6	IGHV	CAGGCTTATCTACAGCAGTCTGGGGCTGAGCTGGTGAGGTCTGGGGCCTCAGTGAAGATGTCCTGCAAGGCTTCTGGCTACACATTTACCAGTTACAATATGCACTGGGTAAAGCAGACACCTGGACAGGGCCTGGAATGGATTGGATATATTTATCCTGGAAATGGTGGTACTAACTACAATCAGAAGTTCAAGGGCAAGGCCACATTGACTGCAGACACATCCTCCAGCACAGCCTACATGCAGATCAGCAGCCTGACATCTGAAGACTCTGCGGTCTATTTCTGTGCAAGA
L-Part2	Leader	GTGTCCACTCC
L-Part1	Leader	ATGGGATTCAGCAGGATCTTTCTCTTCCTCCTGTCAATAACTACAG
RSS Pass	RSS	CACAGAGTTGAAACCACATCCTGAGTGTGTCAGAAACCC
IGHV1.4.P	IGHV	CAGGTCCAACTTCAGCAGTCTGGACCGGAGCTGGTAATACCTGGGGCTTCAGTGAAGTTGTCCTGCAAGGCTTCTGGCTACAATTTTAATGACTATGAAAATCAATGGGTGAAGCAGAGTCTGAAGCAGGCACTGGAATGGATTGGAGCTATTCATCCTGAAAATGGTGGTATTACCTACAATCAGAAGTTCAAAGGCAAGGCCACATTTACTGTAGACACATCCTCCAACACAGCCTACATGCAACTCAGAAGCCTGACATCTGAGGACACTGCTGACTATTATTGTGAGAGA
L-Part2	Leader	GCATCCTCTCC
L-Part1	Leader	ATGGAATGAAGCTGCATCTTTTCTTCCTCCTGTCAATAACTATAG
RSS Pass	RSS	CACAGTGTTGTAACCACATCCCGAATGTGTCATAAACCC

IGHV1.7	IGHV	GAGGTCCAGCTGCAGCAGTCTGGACCTGAGCTGGTAAAGCC TGGGGCTTCAGTGAAGATGTCCTGCAAGGCTTCTGGATACA CATTCACTAGCTATGTTATGCACTGGGTGAAGCAGAAGCCTG GGCAGGGCCTTGAGTGGATTGGATATATTAATCCTTACAATG ATGGTACTAAGTACAATGAGAAGTTCAAAGGCAAGGCCACAC TGACTTCAGACAAATCCTCCAGCACAGCCTACATGGAGCTCA GCAGCCTGACCTCTGAGGACTCTGCGGTCTATTACTGTGCAA GA
L-Part2	Leader	GTGTCCACTCT
L-Part1	Leader	ATGGAATGGAGTTGGATATTTCTCTTTCTCCTGTCAGGAACT GCAG
RSS Pass	RSS	CACAGTGTTGAAACCACATCCTGACTGTGTCAGAAACCC
IGHV1.8	IGHV	CAGGTTCAACTGCAGCAGTCTGGGGCTGAGCTGGTGAGGCC TGGGGCTTCAGTGAAGCTGTCCTGCAAGGCTTTGGGCTACA CATTTACTGACTATGAAATGCACTGGGTGAAGCAGACACCTG TGCATGGCCTGGAATGGATTGGAGCTATTCATCCAGGAAGT GGTGGTACTGCCTACAATCAGAAGTTCAAGGGCAAGGCCAC ACTGACTGCAGACAAATCCTCCAGCACAGCCTACATGGAGCT CAGCAGCCTGACATCTGAGGACTCTGCTGTCTATTACTGTAC AAGA
L-Part2	Leader	GTGTCCAATCC
L-Part1	Leader	ATGGAATGGAGCTGGGTCTTTCTCTTCCTCCTGTCAGTAACT GCAG
RSS Pass	RSS	CACAGTGCTGCAACCACCTCCTGAGTGTGTCAGAAACCC
IGHV1.9	IGHV	GAGGTCCAGCTGCAGCAGTCTGGACCTAAGGTAGTGAATGC TGGGGCTTCAGTGAAGCTGTCCTGCAAGTCTTCTGGTTACTC ATTCAGTAGATACAAAATGGAATGTGTGAAACAGAGCCATGT AAAGAGCCTTGAGTGGATTGAACATATTAATCTTTTCAATGGT ATTACTAACTACAATGGAAACTTTAAAAGCAAGGCCACATTGA CTGTAGACATATCCTCTAGCACAGCCTATATGGAACTTAGCA GATTGACATCTGAAGACTCAGAGGTATATTACTGTGCAAGA
L-Part2	Leader	GTGTCCACTCT
L-Part1	Leader	ATGGAATGGAGTTGGGTCTTTATTTTATTCCTGTCAGTAACTA CAG
RSS Fail	RSS	CAGTGTTGTAACCATATTCTTAGAGTGTCAGAAATCCTG
IGHV1.5.P	IGHV	GAAGTCCACCTGCAGCAGTCTCTACCTAAGGTAGTGAAGGC TGGGCCTTCAGTGAAGATATCCTGCAAGGCTTCTGGTTACTC ATTCACTGGCTACTACATGCACTGGGTGAAGCAGAGCCATG GAAAGATCCTTCAAAGGATTGAATATGTTAATCCTTATAATGG TGGTACTGGCTACAATGAAAAGTTCAAGGACAAGGCCACATT GACTGCAGACAAATCCTTCAGCACAGCCTATATGCAATTCAG CAGCCTGACATCTGAGGACTCTCTGGTCTATTACTGTGCAAG ATA

L-Part2	Leader	GTGTCAACTCT
L-Part1	Leader	ATGGAGCTGGAACTTTCTCTTCCTCCTTTCAGGAACTACAG
RSS Pass	RSS	CACAGTGTTGTAACCACAGCCTGAGTGTGTCAGAAAACT
IGHV1.6.P	IGHV	TGAGGTCCAGCTGCAGCAGTCTGGACCTGAGCTGGTGAGCC TGGGGCTTCAGTGAAGATATCCTGCAAGGCTTCTGGTTATTC ATTCAGTGACTACTACATGGAATGGGTGAAGCAGAGCCATAG AAAGAGTCTTGAATGTATTGGAGAAATTAATCCTTACAATGGT GGTACTACCTACAACCAGAAGTTCAAGGGCAAGGCCACATT GACTGTAGACACATCTTCCAGCACAGCGTACATGGAGCTCC GCAGCCTGACATCTGAGGACTCTTCGGTCTATTACTGTGCAA GA
L-Part2	Leader	GTGTCCACTC
L-Part1	Leader	ATGTCCTCTCCACAGTACCTTAACACACTGACTCTAACCATAA AATTGAGCTGGATCTTTCTCTTCCTCCTGTCAGGAACTGCAG
RSS Pass	RSS	CACAGTGCTACAAACACTTCCTGAGTGTGTCAGAAATCC
IGHV1.10	IGHV	GAGGTCCAGCTGCAACAGTCTGGACCTGAGCTGGTGAAGCC TGGGGCTTCAGTGAAGATATCCTGCAAGACTTCTGGATACAC ATTCACTGAATACACCATGCACTGGGTGAAGCAGAGCCATG GAAAGAGCCTTGAGTGGATTGGAGGTATTAATCCTAACAATG GTGGTACTAGCTACAACCAGAAGTTCAAGGGCAAGGCCACA TTGACTGTAGACAAGTCCTCCAGCACAGCCTACATGGAGCTC CGCAGCCTGACATCTGAGGATTCTGCAGTCTATTACTGTGCA AGA
L-Part2	Leader	GTGTCCTCTCT
L-Part1	Leader	ATGGGATGGAGCTGGATCTTTCTCTTTCTCCTGTCAGGAACT GCAG
RSS Pass	RSS	CACAGTGTTATAGCCACATCCTTTGTGTGACAGAAAACC
IGHV1.7.P	IGHV	TAAGGTCAAGCTGCAGCAGTCTGGACCTGAGCTGGTGAGCC TGGGGCTTCAGTGAAGATATCCTGCAAGACTTCTAGTTACTC ATTCACTGGCTACTACATGCACTGGGTGAAGCAGAGCCATTG AAAGAGCCTTGAGTGGATTGGACTTATTATTCCTTACAATGGT GATACTGGCTACAACCAGAAGTTCAAGGAAAAGGCCACATTG ACTGTAGACTAGTCCTTCAGCACAGCCTACATGGAGCTCCGC AGCCTGAAATCTTAGGACTCTGTGGTCTATTACTGTGCAAGA
L-Part2	Leader	GCATCCACTC
L-Part1	Leader	ACGGAGCTGGCTCTTTCTCTTCCTCCTGTCAGTAAAT
RSS Pass	RSS	CACAGTGTTGTAACCACATCCTGGAGTGTGTCAGAAAAC

IGHV1.11	IGHV	GAGATCCAGCTGCAGCAGTCTGGACCTGAGCTGATGAAGCCTGGGGCTTCAGTGAAGATATCCTGCAAGGCTTCTGGTTACTCATTCACTAGCTACTACATGCACTGGGTGAAGCAGAGCCATGGAAAGAGCCTTGAGTGGATTGGATATATTGATCCTTTCAATGGTGGTACTAGCTACAACCAGAAATTCAAGGGCAAGGCCACATTGACTGTAGACAAATCTTCCAGCACAGCCTACATGCATCTCAGCAGCCTGACATCTGAGGACTCTGCAGTCTATTACTGTGCAAGA
L-Part2	Leader	GTGTCCACTCT
L-Part1	Leader	ATGGGATGGAGCTGGATCTTTCTCTTCCTTCTGTCAGGAACTGCAG
RSS Fail	RSS	CAGTGCTGTAACCACATCGTGAGTGTGTCAGTAATCCTG
IGHV1.8.P	IGHV	GAGGTCCACCTGAAGCAGTCTGGACCTAAGGTAGTGAGGGCTGGGGCTTCAGTGAAGATATCCTGCAAGGCTTCTGGTTACTCATTCACTGGCTACTACAGGCACTGGGTGAAGTCCTGTAAAAAGCCTTCAAAGGATTGGATATATTATTCCTTACAGTGGTGGTACTGGCTACAATGAAAAGTTCAAGGGGAAGGCCACATTGACTGCAGACAAATCCTTCAGCACAGCCTATATGCAATTCAGCAGCCTGACATCTGAGGACTCTGTGGACTCTTACTGTGCGAGATA
L-Part2	Leader	GTGTCCATTCT
L-Part1	Leader	ATGGAATGGAACTGGTTGAAACTTACTCTTTCTCCTGTCAGGAACTACAG
RSS Pass	RSS	CACAGTGTTGTAACCACAGCCTGAGTGTGTCAGAAAATG
IGHV1.12	IGHV	GAGGTCCAACTGCAGCAGTCTGGACCTGAGCTGGTGAAGCCTGGGGCTTCAGTGAAGATATCCTGCAAGGCTTCTGGTTACTCATTCACTGGCTACTACATGCACTGGGTGAAGCAAAGTCCTGAAAATAGTCTTGAGTGGATTGGAGAGATTAATCCTAGCACTGGGGGTACTAGCTACAACCAGAAGTTCAAGGGCAAGGCCACATTAACTGTAGATAAATCCTCCAGCACAGCCTACATGCAGCTCAAGAGCCTGACATCTGAAGAGTCTGCAGTCTATTACTGTACAAGA
L-Part2	Leader	GTGTCCACTCT
L-Part1	Leader	ATGGGATGGAGCTGGACCTTTATTTTAATCCTGTCAGTAACTACAG
RSS Pass	RSS	CACGGTGCTACAAACACATCCTGTGTGTGTCAGAAACCC
IGHV1.9.P	IGHV	GAGGTCCAACTGCAGCAGTCTGGACCTGAGCTGGTGAAGCCTGGGGCTTCAGTGAAGATATCCTGCAAGGCTTCTGGTTACTCATTCACTGTCTACTACATGCACTGGGTGAAGCAGAGCCATGGAAAGAGCCTTGAGTGGACTGGATAAGTTTATCCTTACAATGGTGTTTCTAGCTACAACCAGAAATTCAAGGGCAAGGCCACATTGACTGTAGACAAGTCCTCTATAGCACAGCCTACATGGAGCTCCACAGCCTGACATCTGAGGACTCTGCAGTCTATTACTGTGCAAGA

L-Part2	Leader	GTGTCCACTCT
L-Part1	Leader	ATGGAATGGAGGTGGATCTTTCTCTTCCTCCTTTCAGCAACTACAG
RSS Pass	RSS	CAAAGTGCTACAAATACATCCTGAGAGTTTCAGAAATCC
IGHV1.10.P	IGHV	GAGGTCCAGCTGCAGCAGTCTTTACCTGACCTGGTGAAGCCTGGGGCTTCAGTGAAGATATCCTGCAAGGCTTCTGGTTACTCTTTCACTGGCTACTACATGCACTGGGTGAAGCAGAGCCATGGAAATAGCCTTGAGTGGATTTGATATATTGATCCTTATGATGGTGTTACTAGCTACAATAGGAAGTTCAAGGGAAAGGCCACATTGACTGTAGACAAGTCCTCCAGCACAGCCTACATGGAGCTATGAAACCTGACATCTGAGGAAACTGCAGTCTATTACTGTGCAAGA
L-Part2	Leader	GTGTCCACTCT
L-Part1	Leader	ATGGAATAGAGCTGCATCTTTCTCTTCCTTCTGTCAGTAACTGCAG
RSS Pass	RSS	CACAGTGCTACAAACACATCCTGAGTGTGTCAGAAAACT
IGHV1.13	IGHV	GAGGTCCAGCTGCAACAGTCTGGACCTGAGCTGGTGAAGCCTGGGGCTTCAGTGAAGATGTCCTGTAAGGCTTCTGGATACACATTCACTGACTACTACATGGACTGGGTGAAGCAGAGCCATGGAGAAAGCTTTGAGTGGATTGGACGTGTTAATCCTTACAATGGTGGTACTAGCTACAACCAGAAGTTCAAGGGCAAGGCCACATTGACTGTTGACAAGTCCTCCAGCACAGCCTACATGGAGCTCAACAGCCTGACATCTGAGGACTCTGCGGTCTATTACTGTGCAAGA
L-Part2	Leader	GTGTCCACTCT
L-Part1	Leader	ATGGGATGGAGCTGGATCTTTCTCTTCCTCTTGTCAGGAACTGCAG
	RSS	CAGACTGATTTCCAGAGAGGTTGTACAAGCCTGAAATCC
	RSS	CACAATGAGATGAACAAACCTATGGATAATAGGAGAAGA
	RSS	CACAGGGCTGAAATCAACAAAATAGAAACACAAAAAACT
	RSS	CACATTCCTTGCAAGAGGAGAGCTTGCCTGCAGAGAGTT
RSS Pass	RSS	CACAGTGTTGCAACCACATCCTGAGTGTGTCAGAAATCC
IGHV1.11.P	IGHV	GAGGTCCAGCTGCAGCAGTCTGGACCTGAGCTGGTGAATCCAGTGGCTTCAGTGAAGATATCCTGCAAGGCTTCTGGTTACTCATTCACTGGCTACTACATACACTGGGTGAAGCAGGGTCCTAGAAAGAGCCTTGAGTGGATTGGATATATTAGTTGTTACAATCGTGCTACTAGCTACAACCAGAAGTTCAAAGGCAAGGCCATATTTACTGTAGACAAATCCTCCAGCACAGCCTACATGCAATTCAACAGCCTGACATCTGAGGACTCTGCTGTCTATTACTGTGCAAGA
L-Part2	Leader	GTGTCCACTCT

L-Part1	Leader	ATGGAGCTGAATCTTTCTCTTCCTCCTGTCAGGAACTGCAG
RSS Fail	RSS	CAAGATACAGTGTTGTAACCACATTCTTAGTGTGTTAGA
IGHV1.12.P	IGHV	GAGGTCCACATGCAGCAGTCTGGACCTAAGGTAGTGAAGGCTGGACTTTCAGTGAAGATATCCTGCAAGGCTTCTGGGTACTCATTCACTGACTACTACACTCACTGGGTGAAACAGAGTCCTTTAAAGAGCCTTCAAAAGTTTGAATATATTAATCCTTATAATTGTGCTACTGGCTACAATGAAAAGTTCAAGGACAAGGCCACATTGACTGTAGACAAATCCTTCAGCACAGCCTATATGCATCTCAGCAGCCTGACATCTGAGGACTCTGAGGTCTATTACTGTG
L-Part2	Leader	GTGTCCACTCT
L-Part1	Leader	AGGGAATGGAGCTGGAACTTTCTCTTCCTCTTGTCAGGAACTACAG
	RSS	CAGAGCGAGGGGTCCATGACTCCAGGGTGATAGAATGGA
RSS Pass	RSS	CACAGTGTTGTAACCACAGCCTGAGTGTGTCAGAAAACT
IGHV1.13.P	IGHV	GAGGTCCAGCTGAAGCAGTCTTGACCTGAGCTGGTGAGCCTGGGGCTTCAGTGAAGATATCCTGCAAGGCTTCTGGTTACTCATTCACTGGCTACATTATGAACTGGGTGAAGCAGAGCCATGGAAAGAGCCTTGAGTGGATTGGAGAAATTAATCCTTACAAAGGTGGTACTAGCTACAACCAGAAGTTCAAGGGCAAGGCCACATTGACTGTAGACACATCCTCCAGCACAGCGTACATGGAGCTCAAGAGCCTGACATCTGAGGACTCTGCAGTCTATTACTGTGCAAGA
L-Part2	Leader	GTGTCCACTCT
L-Part1	Leader	ATGGAGCTGGATCTTTCTCTTCCTCCTGTCAGGAACTGAAG
	RSS	CACATTGTCTGAAAGATCTCAGGAGTCTAAGTTTATATG
RSS Pass	RSS	CACAGTGCTACAAACACATCCTGAGTGTGTCAGAAAACT
IGHV1.14	IGHV	GAGGTCCAGCTGCAACAGTCTGGACCTGAGCTGGTGAAGCCTGGGGCTTCAGTGAAGATATCCTGCAAGGCTTCTGGTTACTCATTCACTGGCTACTACATGCACTGGGTGAAGCAAAGCCATGTAAAGAGCCTTGAGTGGATTGGACGTATTAATCCTTACAATGGTGCTACTAGCTACAACCAGAATTTCAAGGACAAGGCCAGCTTGACTGTAGATAAGTCCTCCAGCACAGCCTACATGGAGCTCCACAGCCTGACATCTGAGGACTCTGCAGTCTATTACTGTGCAAGA
L-Part2	Leader	GTGTCCTCTCT
L-Part1	Leader	ATGGGATGGAGCTGGATCTTTCTCTTTCTCCTGTCAGGAACTGCAG
	RSS	CATAGTGTATAATTCAAGTCATGGAACCACACATAAATG
	RSS	CACAGTCTGAGACATTAGTCACCATCTCATTATAAAGGA

RSS Pass	RSS	CACAGTGAGAAGTCTTCATTGTGAGTCTAGACACAAACT
IGHV6.6	IGHV	GAAGTGAAGCTTGAGGAGTCTGGAGGAGGCTTGGTGCAACC TGGAGGATCCATGAAACTCTCCTGTGTAGCCTCTGGATTTAC TTTCAGTAGCTACTGGATGTCTTGGGTCCGCCAGTCTCCAGA GAAGGGGCTTGAGTGGGTTGCTGAAATTAGATTGAAATCTGA TAATTATGCAACACATTATGCGGAGTCTGTGAAAGGGAAGTT CACCATCTCAAGAGATGATTCCAAAAGTCGTCTCTACCTGCA AATGAACAGCTTAAGAGCTGAAGACACTGGAATTTATTACTG TACAGA
L-Part2	Leader	GGGTCCAGAGT
L-Part1	Leader	ATGGACTTGAGACTGAGCTGCGCTTTTATTATTGTTCTTTTAA AAG
RSS	RSS	CACAGTAAGAAGCCTTCATTGTGAATCTATCCACCAACC
IGHV6.7	IGHV	GATGTGAACCTGGAAGTGTCTGGAGGAGGCTTAGTTAAACCT GGAGGATCCATGCAACTCTCTTGTGTAGCCTCTGGATTTACT TTTGTAGATGGCTGGATGGACTGGGTCCGCCAGTCTCCAGA GAAGGGTCTTGAGTGGGTTGCTGAAATTGCAAACAAAGCTAA TAATTACGCAACATATTATCCCGAGTCTGTGAAAGGCAGATT CACCATCTCAAGAGATGATTTCAAAAGTCGTGTCTACCTGCA AAAGAACAGCTTAAGAGCTGAAGATACAGGCATTTATTACTG TACAAGG
L-Part2	Leader	GTGTCCAGAGT
L-Part1	Leader	ATGTACTTGGAACTGAGCTGTGTTTTCATTGTTGTTTTCTTAA AAG
RSS	RSS	CACAGTGAGAAGCCTTCATTGTGAATCTATCCACAAACC
IGHV6.8	IGHV	GAAGTGAAAATTGAGGAGTCAAGAGGAGGCTTCATCCAACCT GGAGGATCCATGAAATTCTCTTGTGCAGCCTCTGGATTCACT TTCAGTGATTACAGGATGGACTGGGTCCACCACTCTCCAGAG AATGGGTTGGACTGGGTTGCTGAAATTAGAAACAAAGCTAGT AATTATGCAACATATTATGTGGAGTCTGTGAAAGGGAGGTTC ACCATCTCAAGAGATGATTCCAAAAGTAGTGTCTACCTGCAA ATGAACAGCTTAAGAGCTGAAGATACTGGCATTTATTACTGTA CAAGG
L-Part2	Leader	GTGTCCAGAGT
L-Part1	Leader	ATGTACTTGGAAATGAGCTGTATTTTCATTGTTGTCCTCTTAA AAG
RSS Pass	RSS	CACAGTGAGAAGGCTTTATTATGAACCTATCCACAAAC

IGHV6.9	IGHV	GAAGTGAAGCTTGAGGAGTCTGGAGGAGGCTTGGTGCAACC TGGAGGATCCATGAAACTCTCTTGTGCTGCCTCTGGATTCAC TTTTAGTGACGCCTGGATGGACTGGGTCCGCCAGTCTCCAG AGAAGGGGCTTGAGTGGGTTGCTGAAATTAGAAGCAAAGCT AATAATCATGCAACATACTATGCTGAGTCTGTGAAAGGGAGG TTCACCATCTCAAGAGATGATTCCAAAAGTAGTGTCTACCTG CAAATGAACAGCTTAAGAGCTGAAGACACTGGCATTTATTAC TGTACCAGG
L-Part2	Leader	GTGTCCAGAGT
L-Part1	Leader	ATGTACTTGGGACTGAACTATGTATTCATAGTTTTTCTCTTAA ATG
IGHV6.1 0	IGHV	GAGGAGAAGCTGGATGAGTCTGGAGGAGGCTTGGTGCAACC TGGGAGGTCCATGAAACTCTCCTGTGTTGCCTTTGGATTCAC TTTTACTAACTCCTGGATGAACTGGTTCTGCCAGTCTCCAGA GAAAGGACTGGAGTGGGTAGCACAAATTAAAAGCAAACCTTA TAATTATGAAACATATTATTCAGATTCTGTGAAAGGCAGATTG ACCATCTCAAGAGATGATTCCAAAAGTAGTGTATACCTGCAA ATGAACAACTTAAGAGCTGAAGACACGGGCATCTATTACTGT ACATGG
L-Part2	Leader	GTGTCCAGTGT
L-Part1	Leader	ATGTACTTGGGACTGAGCTGTGTATTCATTGTTTTTCTCTTAA AAG
RSS Pass	RSS	CACAGTGGTGCAACCGTGACCCACAGCTGTGCAATATTT
IGHV8.2	IGHV	CAGGTTACTCTGAAAGTGTCTGGCCCTGGGATATTGCAGCCA TCACAGACTCTCGGCCTGGCCTGTACTTTCTCTGGGATTTCA CTGAGTACTTCTGGTATGGGTTTGAGCTGGCTTCGTAAGCCC TCAGGGAAGGCTTTAGAGTGGCTGGCAAGCATTTGGAATAAT GATAACTACTACAACCCATCTTTGAAGAGCCGGCTCACAATC TCCAAGGAGACCTCCAACAACCAAGTATTCCTTAAACTCACC AGTGTGGACACTGCAGATTCTACCACATACTACTGTGCTTGG AGAGAG
L-Part2	Leader	ATGTCCTATCT
L-Part1	Leader	ATGGGCAGACTTACATTCTCATTCCTGCTGCTGCTGCCTGTC CCTGCAT
IGHV1.1 4.P	IGHV	AGCAGTCTGGGCCTGAGCTGGCAAGGCCTTGGGCTTCAGTG AAGATATCCTGCCAGGCTTTCTACACCTTTTCCAGAAGGGTG TACTTTGCCATTAGGGATACCAACTACTGGATGCAGTGGGTA AAACAGAGGCCTGGACAGGGTCTGGAATGGATCGGGGCTAT TTATCCTGGAAATGGTGATACTAGTTACAATCAGAAGTTCAAG GGCAAGGCCACATTGACTGCAGACAAATCCTCCAGCACAGC CTACATGCAACTCAGCAGCCTGACATCTGAGGACTCTGCGG TCTATTACTGTGCATGA
RSS Pass	RSS	CACAGTGTGGAATCTTCAGTGTTAGCCTTGACACAAACC

IGHV10.2	IGHV	GAGGTGCAGCTTGTTGAGTCTGGTGGAGGATTGGTGCAGCCTAAAGGGTCATTGAAACTCTCATGTGCAGCCTCTGGATTCACCTTCAATACCTACGCCATGAACTGGGTCCGCCAGGCTCCAGGAAAGGGTTTGGAATGGGTTGCTCGCATAAGAAGTAAAAGTAATAATTATGCAACATATTATGCCGATTCAGTGAAAGACAGGTTCACCATCTCCAGAGATGATTCACAAAGCATGCTCTATCTGCAAATGAACAACTTGAAAACTGAGGACACAGCCATGTATTACTGTGTGAGACA
L-Part2	Leader	GTGTGCATTGT
L-Part1	Leader	ATGCTGTTGGGGCTGAAGTGGGTTTTCTTTGTTGTTTTTTATCAAG
RSS Pass	RSS	CACAGTGGTGCAACCACATCCCGACTGTGTCACAAACCC
IGHV1.15	IGHV	CAGGTCCAGCTGCAGCAGTCTGGGGCTGAACTGGCAAGACCTGGGGCCTCAGTGAAGATGTCCTGCAAGGCTTCTGGCTACACCTTTACTAGCTACACGATGCACTGGGTAAAACAGAGGCCTGGACAGGGTCTGGAATGGATTGGATACATTAATCCTAGCAGTGGTTATACTAATTACAATCAGAAGTTCAAGGACAAGGCCACATTGACTGCAGACAAATCCTCCAGCACAGCCTACATGCAACTGAGCAGCCTGACATCTGAGGACTCTGCAGTCTATTACTGTGCAAGA
L-Part2	Leader	GTGTCCACTCC
L-Part1	Leader	ATGGAAAGGCACTGGATCTTTCTACTCCTGTTGTCAGTAACTGCAG
IGHV10.4.P	IGHV	GCTTACAGGTGTGCATTGTGAGGTACAGCTTGTTGGGTCTGGTGAAGGATTGGTGCAGCATAAAGGGTCACTTGTATTCAGCCTCTGGATTCATCTTCAATACATACACCTTGGAGTGGGTGTGTCAGACTCCAAGAAAGGGTCTGGAATGGGTTGCATGCATAAGAACAAAAAGTAATAATTATGCCACATATACTGATTCAGTGAAAGACAGATTCACCATCTCCAGAGATGATTCTCAAAGCATGGTCAACCTGCAAATGAACAATTTGAAAACTGAGGACATAGACCTGTATTAATTACAAGAGA
RSS Pass	RSS	CACAGTGTTGTGACCACATCCTGAGTGTGTCAGAAACTT
IGHV1.15.P	IGHV	GAGGTTCAGCTCCAGCAGTCTGGGACTGTGCTGGCAAGGCCTGGGGCTTCAGTGAAGATGTCCTGCAAGGCTTCTGGCTACACCTTTACCAGCTACTGGATGCACTGGGTAAAACAGAGGCCTGGACAGGGTCTGGAATGGATTGGCGTATTTATCCTGGAAATAGTGATACTAGCTACAACCAGAAGTTCAAGGGCAAGGCCAAACTGACTGCAGTCACATCCACCAGCACTGCCTACATGGAGCTCACAGCCTGACAAATGAGGACTCTGCGGTCTATTACTGTACAAGA
L-Part2	Leader	GGGTCTACTCA
L-Part1	Leader	ATGGAATGTAACTGGATACTTCCTTTTATTCTGTCGGTACTTCAG

IGHV10.5.P	IGHV	GAGTGCAGCTTGTTGAGACTGTGGAGGATTGTGCAGCCTAAAGGGTCATTGAAACTCTCATGTGCAGCCTCTGGATTCACCTTCATACCAATGCCATGAACTGGGTCCGCCAGGCTCCAGGAAAGGTTGGAATGGGTTGCTCGCATAAGAAGTAAAAGTAATAATTATGCAACATATTATGCCGATTCAGTGAAAGACAGGTTCACCATCTCCAGAGATGATTCACAAAGCATGCTCTATCTGCAAATGAACAACTTGAAAAACTGAGACACAGCCATGTATTACTGTGAGAA
IGHV1.16.P	IGHV	CAGGATCCAGCTGCAGCAGTCTGGAACTGAGTTTGGAAGGCCTGGAGTCTCAGAGAAGTAGTCCTGCAAGTCTTCTGTTACAAATTCACCAACTACTAGATTCACCGGATAGGGCAGAATCGTGGAAATAGCTAGGCTAGATTGGATATATTTATCGTGGAATTGGCCATCCTGGCTATGCTCACAACTTCAAGGACAAGACCACACTGACTTCAGACAAATCCTCTAGCACAGCCTACATGAATTTCACATGCATTACTCCACCAATCTTCTTGGGTATCTCCCCATGACGTAACTC
RSS Pass	RSS	CACAGTGGTGCAACCACATCCCGACTGTGTCAGAAACC
IGHV1.16	IGHV	CAGGTCCAGCTGCAGCAGTCTGCAGCTGAACTGGCAAGACCTGGGGCCTCAGTGAAGATGTCCTGCAAGGCTTCTGGCTACACCTTTACTAGCTACACGATGCACTGGGTAAAACAGAGGCCTGGACAGGGTCTGGAATGGATTGGATACATTAATCCTAGCAGTGGATATACTGAGTACAATCAGAAGTTCAAGGACAAGACCACATTGACTGCAGACAAATCCTCCAGCACAGCCTACATGCAACTGAGCAGCCTGACATCTGAGGACTCTGCGGTCTATTACTGTGCAAGA
L-Part2	Leader	GTGTCCACTCC
L-Part1	Leader	AAGGCACTGGATCTTCTCTTCCTGTTGTCAGTAACTGCAGGTAAGG
IGHV10.6.P	IGHV	GAGGTACAGCTTGTTGGGTCTGGTGAAGGATTGGTGCAGCCTAAAGGTCACTGAACACTTGTATTTCAGCCTCTGGATTCATCTTCATACAATACCATGGAATGGGTCTGCCAGGCTCCAAGAAGGGTCTGAATGGGTTGCATGCATTAAGAACTAAAAGTAATAATTATGCACATATACTGATTCAGTTGAAAGACAGATTGCCCATCTCCAGAGATGATTCTCAAAGCATGATCAACCTGTAAATGAACAATCTGAAAACTGAGGACATAGACATGTATTAATTACAAGAGA
IGHV15.2.P	IGHV	CAGGTTCACCTACAACAGTCTGGTTTGAACTGAGGAGTCCTGGGTCTTCAGTAAAGCTGTCATGCAAGGATTTTGATTCAGAAGTCTTCCCTATTGCTTATATGAGTTGGGTAGGCAGAAGCCTGGACATGGATTTGAATGGATTGGAGACATACTCCAAGTATTGTAGAAGAATCTATGGAGTGAAGTTTGAGGACAAAGCCACACTGGATGCAGACACAGTGTCCAACACAGCCTACTTGGAGCTCAACAGTCTGACATCTGAGGACTCTGCTATCTACTACTGTGCAAGG
RSS Pass	RSS	CACAGTGTTATAACCACTTCCTGAGTGTGTTAGAAACTC

IGHV1.17	IGHV	CAGGGTCAGATGCAGCAGTCTGGAGCTGAGCTGGTGAAGCCTGGGGCTTCAGTGAAGCTGTCCTGCAAGACTTCTGGCTTCACCTTCAGCAGTAGCTATATAAGTTGGTTGAAGCAAAAGCCTGGACAGAGTCTTGAGTGGATTGCATGGATTTATGCTGGAACTGGTGGTACTAGCTATAATCAGAAGTTCACAGGCAAGGCCCAACTGACTGTAGACACATCCTCCAGCACAGCCTACATGCAATTCAGCAGCCTGACAACTGAGGACTCTGCCATCTATTACTGTGCAAGACAC
L-Part2	Leader	GTGTCTATGCC
L-Part1	Leader	ATGGAATGGAACTGGGTCGTTCTCTTCCTCCTGTCATTAACTGCAG
IGHV1.18	IGHV	CAGATCCAGCTGCAACAGTCAGGAGCTGAGCTGGCGAGTCCTGGGGCATCAGTGACACTGTCCTGCAAGGCTTCTGGCTACACATTTACTGACCATATTATGAATTGGGTAAAAAAGAGGCCTGGACAGGGCCTTGAGTGGATTGGAAGGATTTATCCAGTAAGTTGTGAAACTAACTACAATCAAAAGTTCATGGGCAAGGCCACATTCTCTGTAGACCGGTCCTCCAGCACAGTGTACATGGTGTTGAACAGCCTGACATCTGAGGACCCTGCTGTCTATTACTGTGGAAGG
L-Part2	Leader	GTGCCTACTCC
L-Part1	Leader	ATGGGATGGAGCCAGATCACCCTCTTTCTGGTGGCAGCAATTACAT
IGHV1.17.P	IGHV	CTCAGTGAAGATGTCCTGCAAGGCTTCTGGCTACACATTTACCAGTTACAATATGCACTGGGTAAAGCAGACACCTGGACAGGGCCTGGAATGGATTGGAGCTATTTATCCAGGAAATGGTGATACTTCCTACAATCAGAAGTTCAAGGGCAAGGCCACACTGACTGTAGACACATCCTCCAGCACAGCGTACATGCAGATCAGCAGCCTGACATCTGAAGACTCTGCGGAGAAGACTCTGCGGTCTATTTCTGTGCAAGACACAGTGTTGTGACCA
RSS Pass	RSS	CACAGTGTTGTAACCACATCCCGAATGTGTCATTAAAC
IGHV1.18.P	IGHV	GAGGTCCAGCTGCAGCAGTCTGGACCTGAGCTGGTGAAGCCTGGGGCTTCAGTGAAGATGTCCTGCAAGGCTTCTGGATACACATTCACTGACTATGTTATGCACTGGGTGAAGCAGAAGCCTGGGCAGGGCCTTGAGTGGATTGGATATATTTATCCTTACAATGATGGTACTGAGTACACTGAGAAGTTCAAAGGCAAGGCCACACTGACTTTAGACAAATCCTCCAGCACAGCCTACATGGATCTCAGCAGCCTGACCTCTGAGGACTCTACGGTCTATTACTGTGCAAGA
L-Part2	Leader	GTGTCCACTCT
L-Part1	Leader	ATGGAGCTGGATATTTCTCTTCCTCCTGTCAGGAACTGCAG
RSS Pass	RSS	CACAGTGTTGAAACCACATCCTGACTGTATCAGAAACC

IGHV1.19	IGHV	CAGGTTCAACTGCAGCAGTCTGGGGCTGAGCTGGTGAGGCCTGGGGCTTCAGTGACGCTGTCCTGCAAGGCTTCGGGCTACACATTTACTGACTATGAAATGCACTGGGTGAAGCAGACACCTGTGCATGGCCTGGAATGGATTGGAGCTATTGATCCTGAAACTGGTGGTACTGCCTACAATCAGAAGTTCAAGGGCAAGGCCACACTGACTGCAGACAAATCCTCCAGCACAGCCTACATGGAGCTCCGCAGCCTGACATCTGAGGACTCTGCCGTCTATTACTGTACAAGA
L-Part2	Leader	GTGTCCAATCC
L-Part1	Leader	ATGGAATGGAGCTGGGTCTTTCTCTTCCTCCTGTCAGTAATTGCAG
IGHV1.19.P	IGHV	GAGGTCCAGCTGCAGCAGTCTGGACCTAAGGTAGTGAATGCTGGGGCTTCAGTGAAGCTGTCCTGCAAGTCTTCTGGTTACTCATTCAGTAGATACAAAACGGAATATGTGAAGCAGAGCCATGTAAAGAGCCTTGAGTGGATTTAACATATTAATCTTTTCAATGGTGTTACTAACTACAATGGAAAGTTCAAAAGCAAGGCCACATTGACTGTAGACATATCCTCTAGCACAGCCTATATGGAGCTTAGCAGATTGACATCTGAAGACTCAGAGGTATATTACTGTGCAAGA
IGHV1.20.P	IGHV	GAGGTCTACCTGCAGCAGTCTGTACCTAAGGTAGTGAAGGCTGGGCCTTCAGTGAAGATATCCTGCAAGGCTTCTGGTTACTCATTCACTGGCTGCTACATGCACTGAGTGAAGCAGAGTCCTTTAAAGAGACTTCAAAGGATTGAATATATTAATCTTTATAATGGTGGTACTGGCTACAATGAAAAGTTCAAGGTCAAGGCAACATTGACTGCAGACAAATCCTTCAGCACAGCCTATATGCACCTCAGCGGCCTGACATCTGAGGACTCGGCCTTATATTACCGTGCAAGA
IGHV1.20	IGHV	GAAGTCCAGCTGCAGCAGTCAGGACCTGAGCTGGTGAAGCCTGGCGCTTCAGTGAAGATATCCTGCAAGGCTTCTGATTACTCATTCACTGGCTACATCATGAACTGGGTGAAGCAGAGCCATGGAAAGAGCCTTGAATGGATTGGAGAAATTAATCCTTACAATGGTGGTACTAGCTACAACCAGAAGTTCAAGGGCAAGGCCACATTGACTGTAGACACATCCTCCAGCACAGCGTACATGGAGCTCCACAGCCTGACATCTGAGGACTCTTTGGTCTATTACTGTGCAAGA
L-Part2	Leader	GTGTCCACTCT
L-Part1	Leader	ATAAAATGGAGCTGGATCTCTCTCTTCCTCCTGTCAGGAACTGCAG
RSS Pass	RSS	CACAGTGCTACAAACACATCCTGAGTGTGTCAGAAACC
IGHV1.21	IGHV	GAGGTCCAGCTTCAGCAGTCAGGACCTGAGCTGGTGAAACCTGGGGCCTCAGTGAAGATATCCTGCAAGGCTTCTGGATACACATTCACTGACTACAACATGCACTGGGTGAAGCAGAGCCATGGAAAGAGCCTTGAGTGGATTGGATATATTTATCCTTACAATGGTGGTACTGGCTACAACCAGAAGTTCAAGAGCAAGGCCACATTGACTGTAGACAATTCCTCCAGCACAGCCTACATGGAGCTCCGCAGCCTGACATCTGAGGACTCTGCAGTCTATTACTGTGCAAGA

L-Part2	Leader	GCGTCCACTCT
L-Part1	Leader	ATGGGATGGAGCTGGATCTTTCTCTTCCTCCTGTCAGGAACTGCAG
IGHV1.21.P	IGHV	AAGGTCACTGAAGCAGTCTGGACCTGAGCTCGTGAGCCTGGGGCTTCAGTGAAGATATCCTGCAAGGCTTCTGGTTACTCATTCACTAGCTACTAAATGCACTGGGTGAAGCAGAGTCATGGAAAGAGCCTTGATTGGATTGGACTTATTATTCCTTACAATGGTGATACTGGCTACAACCAGAAGTTCAAAGACAAGGCCACATTGACTGTAGACTAGTCCTTCAGCACAGCCTACATGGAGCTCCGCAGCCTGAAATCTTAGGACTCTGTGGTCTATTACTGTGCAAGA
RSS Pass	RSS	CACAGTGTTGTAACCACATCCTGAGTGTGTCAGAAAAC
IGHV1.22	IGHV	GAGATCCAGCTGCAGCAGTCTGGACCTGAGCTGGTGAAGCCTGGGGCTTCAGTGAAGGTATCCTGCAAGGCTTCTGGTTACTCATTCACTGACTACAACATGTACTGGGTGAAGCAGAGCCATGGAAAGAGCCTTGAGTGGATTGGATATATTGATCCTTACAATGGTGGTACTAGCTACAACCAGAAGTTCAAGGGCAAGGCCACATTGACTGTTGACAAGTCCTCCAGCACAGCCTTCATGCATCTCAACAGCCTGACATCTGAGGACTCTGCAGTCTATTACTGTGCAAGA
L-Part2	Leader	GTGTCCACTCT
L-Part1	Leader	ATGGAATGGAGCTGGATCTTTCTCTTCCTCCTGTCAGGAACTACAG
IGHV1.22.P	IGHV	CCACACAAGTGTCCACTCTGAGGTCCACCTGTAGCAGTCTGGACCTGAGCTGGTGAAGCCTGGGGATTCAGTGAAGATATCCTGCAAGGCTTCTGGTTACTCATTCACTGGCTACTACATGCACTGAGTGAAGTCCTGTAAAAAGCCTTAAAAGGATTGGATATATTAATCCAGTGGTGGTACTGGCTACACTGAAAAGTTCAAGGGCAAGGCCACATTGACTGCAGACAAATCCTTCAGCACAGCCTATATGTACTTCAGCAGCCTGACATCTGAGGACTCTGTGGACTCTTACTGTGCAAGA
RSS	RSS	CACAGTGTTGTAACCACATCCTGAATGTGTCAGAAACCT
IGHV1.23	IGHV	GAGGTCCAGTTGCAGCAGTCTGGACCTGAGCTGGTGAAGACTGGGGCTTCAGTGAAGATATCCTGCAAGGCTTCTGGTTACTCATTCACTGACTACTACATGCACTGGGTGAAGCAAAGTCCTGAAAAGAGTTTTGAGTGGATTGGAGAGATTAATCCTAGCACTGGTGGTACTAGCTACAACCAGAAGTTCAAGGGCAAGGCCACATTGACTGTAGACAAATCCTCCAGCACAGCCTACATGCAGCTCAAGAGCCTGACATCTGAGGACTCTGCAGTCTATTACTGTGCAAGA
L-Part2	Leader	GTGTCCACTCT
L-Part1	Leader	TGGAGCTGGATCTTTATTTTAATCCTGTCAGTAACTACAGGTAAGG

IGHV1.24	IGHV	GAGGTTCAGCTGCAGCAGTCTGGACCTGAGCTGGTGAAGCCTGGGGCTTCAGTGAAGATATCCTGCAAGGCTTCTGGTTACTCATTCACAGGCTACTACATGCACTGGGTGAAGCAGAGCCATGGAAATAGCCTCGATTGGATTGGATATATTTATCCTTACAATGGTGTTTCTAGCTACAACCAGAAATTCAAGGGCAAGGCCACATTGACTGTAGACAAGTCCTCTAGCACAGCCTACATGGAGCTCCGCAGCCTGACATCTGAGGACTCTGCAGTCTATTACTGTGAAAGA
L-Part2	Leader	GTGTCCACTCT
L-Part1	Leader	ATGGAATGGATCTGGATCTTTCTCTTCCTCCTGTCAGCAACTACAG
IGHV1.23.P	IGHV	GGTCCAGCTGCAGCAGTCTGGAACTGAGCTGGTGAAGCCTGGGGCTTCAGTGAAGATATCCCGCAAGACTTCTGGTAAATCTTTCACTGGCTACTACATGCGCTGGGTGAAGCAGAGCCATGGAAATAGCCTTGAGTGGATTTGATATATTGATCCTTACGATGGTGTTACTAGCTACAATAAGAAGTTCAAGAGAAAGGCCACATTGACTGTAGAAAAGTCCTCCAGCACAGCCTACATGGAGCTCTGAAACCTGACATCTGAGGAAACTGAAGTCTATTACTGTGTAAGA
RSS Pass	RSS	CACAGTTCTACAAACATGTCCTGAGTGTGTCAGAAAACT
IGHV1.25	IGHV	GAGGTCCAGCTGCAACAATCTGGACCTGAGCTGGTGAAGCCTGGGGCTTCAGTGAAGATGTCCTGTAAGGCTTCTGGATACACATTCACTGACTACTACATGAAGTGGGTGAAGCAGAGTCATGGAAAGAGCCTTGAGTGGATTGGAGATATTAATCCTAACAATGGTGGTACTAGCTACAACCAGAAGTTCAAGGGCAAGGCCACATTGACTGTAGACAAATCCTCCAGCACAGCCTACATGCAGCTCAACAGCCTGACATCTGAGGACTCTGCAGTCTATTACTGTGCAAGA
L-Part2	Leader	GTGTCCTCTCT
L-Part1	Leader	ATGGGATGGAGCTGGATCTTTCTCTTCCTCTTGTCAGGAACTGCAG
IGHV1.24.P	IGHV	GAGGTTCAGCTGCAACAGTCTGGACCTGAGCTGGTGAAGCCTGGGCTTCAGTGAAGATGTCCTGCAAGCCTTCTGGTTACTCATTTACTGACTACTTTATGAACTGGGTGAAGCAGAGCCATGGAAAGAGCCTTGACTGGATTGGACGTATTAATCCTTACAATGGTGATACTTTCTACAACCAGAAGTTCAAGGGCAAGGCCACATTGACTGTAGACAAATCCTCTAGCACAGCCCACATGGAGCTCCGGAGCCTGACATCTGAGGACTCTGCAGTCTATTATTGTGCAAGA
RSS Pass	RSS	CACAGTGTTGCAACCACATCCTGAGTGTGTCAGAAATC

IGHV1.26	IGHV	GAGGTCCAGCTGCAGCAGTCTGGACCTGAGCTAGTGAAGAC TGGGGCTTCAGTGAAGATATCCTGCAAGGCTTCTGGTTACTC ATTCACTGGTTACTACATGCACTGGGTCAAGCAGAGCCATGG AAAGAGCCTTGAGTGGATTGGATATATTAGTTGTTACAATGG TGCTACTAGCTACAACCAGAAGTTCAAGGGCAAGGCCACATT TACTGTAGACACATCCTCCAGCACAGCCTACATGCAGTTCAA CAGCCTGACATCTGAAGACTCTGCGGTCTATTACTGTGCAAG A
L-Part2	Leader	GTGTCCACTCT
L-Part1	Leader	ATGGGATGGATCTGGATCTTTCTCTTCCTCCTGTCAGGAACT GCAG
IGHV1.27	IGHV	GAGGTCCACCTGCAGCAGTCTCTACCTAAGGTAGTGAAGGC TGGGCCTTCAGTGAAGATATCCTGCAAGGCTTCTGGTTACTC ATTCACTGGCTACTACATGCACTGGGTGAAGCAGAGCCATG GAAAGATCCTTCAAAGGATTGAATATATTAATCCTTATAATGG TGGTACTAGCTACAATGAAAAGTTCAAGGACAAGGCCACATT GACTGCAGATAAATCCTTCAGCACAGCCTACATGCACCTCAG CGGCCTGACATCTGAGGACTCTCTGGTCTATTACTGTGCAAG A
L-Part2	Leader	GTGTCCACTCT
L-Part1	Leader	AGGGAATGGAGCTGGAACTTTCTCTTCCTCCTTTCAGGAACT ACAG
RSS	RSS	CACAGTGTTGTAACCACAGCCTGAGTGTGTCAGAAAACT
IGHV1.28	IGHV	GAGGTCCAGCTGCAGCAGTCTGGACCTGAGCTGGTGAAGCC TGGGCCTTCAGTGAAGATATCCTGCAAGGCTTCTGGTTACTC ATTCACTGGCTACTACATGCACTGGGTGAAGCAGAGCCATG GAAAGAGCCTTGAGTGGATTGGAGAAATTAATCCTTACAATG GTGGTACTAGCTACAACCAGAAGTTCAAGGGCAAGGCCACA TTGACTGTAGACACATCCTCCAGCACAGCGTACATGGAGCTC CACAGCCTGACATCTGAGGACTCTTTGGTCTATTACTGTGCA AGA
L-Part2	Leader	GTGTCCACTCT
L-Part1	Leader	ATAAAATGGAGCTGGATCTTTCTCTTCCTCCTGTCAGGAACT GCAC
IGHV1.25.P	IGHV	GAGGTCCAGCTGCAACAGTCTGGACCTGAGCTGGTGAAGCC TGGGGCTTCAGTGAAGATATCCTGCAAGGCTTCTTGATACAC ATTCACTGACTACAACATGCACTGGGTGAAGCAGAGCCATG GAAAGAGCCTTGAGTGGATTGGAGGTATTAATCCTAACAATG GTGCTACTAGCTACAACCAGAAGTTCAAGGGCAAGGCCACTT TGACTGTAGACAAGTCCTCCAGCACAGCCTACATGGAGCTC CGCAGCCTGACATCTGAGGATTCTGAAGTCTATTACTGTGCA AGA
RSS	RSS	CACAGTGCTACAAACACATCCTGAGTGTGTCAGAAACCT

IGHV1.29	IGHV	GAGGTCCAGCTGCAACAGTTTGGAGCTGAGCTGGTGAAGCCTGGGGCTTCAGTGAAGATATCCTGCAAGGCTTCTGGCTACACATTCACTGACTACAACATGGACTGGGTGAAGCAGAGCCATGGAAAGAGCCTTGAGTGGATTGGAGATATTAATCCTAACTATGATAGTACTAGCTACAACCAGAAGTTCAAGGGAAAGGCCACATTGACTGTAGACAAGTCCTCCAGCACAGCCTACATGGAGCTCCGCAGCCTGACATCTGAGGACACTGCAGTCTATTACTGTGCAAGA
L-Part2	Leader	GTGTCCTCTCT
L-Part1	Leader	ATGGAATGGAGCTGGATCTTTCTCTTTCTCCTGTCAGGAACTGCAG
RSS Pass	RSS	CACAGTGCTACAAACACATCCTGAGTGTGTCAGAAACC
IGHV1.30	IGHV	GAGGTCCTGCTGCAACAGTCTGGACCTGAGCTGGTGAAGCCTGGGGCTTCAGTGAAGATACCCTGCAAGGCTTCTGGATACACATTCACTGACTACAACATGGACTGGGTGAAGCAGAGCCATGGAAAGAGCCTTGAGTGGATTGGAGATATTAATCCTAACAATGGTGGTACTATCTACAACCAGAAGTTCAAGGGCAAGGCCACATTGACTGTAGACAAGTCCTCCAGCACAGCCTACATGGAGCTCCGCAGCCTGACATCTGAGGACACTGCAGTCTATTACTGTGCAAGA
L-Part2	Leader	GTGTCCTCTCT
L-Part1	Leader	ATGGGATGGAGCTGGATCTTTCTCTTCTTCCTGTCAGGAACTGCAG
IGHV1.26.P	IGHV	AAGGTCACTGAAGCAGTCTGGACCTGAGCTCGTGAGCCTGGGGCTTCAGTGAAGATATCCTGCAAGGTTCTGGTTACTCATTCACTGCTACTAAATGCACTGGGTGAAGCAGAGCATGGAAAGAGCCTTGATTGGATTGGACTTATTATTCCTTACAATGGTGATACTGGCTACAACCAGAAGTTCAAAGACAAGGCCACATTGACTGTAGACTAGTCCTTCAGCACAGCCTACATGGAGCTCCGCAGCCTGAAATCTTAGGACTCTGTGGTCTATTACTGTGCAAGA
RSS Pass	RSS	CACAGTGTTGTAACCACATCCTGAGTGTGTCAGAAAACC
IGHV1.31	IGHV	GAGATCCAGCTGCAGCAGTCTGGACCTGAGCTGGTGAAGCCTGGGGCTTCAGTGAAGGTATCCTGCAAGGCTTCTGGTTATCATTCACTGACTACAACATGTACTGGGTGAAGCAGAGCCATGGAAAGAGCCTTGAGTGGATTGGATATATTGATCCTTACAATGGTGGTACTAGCTACAACCAGAAGTTCAAGGGCAAGGCCACATTGACTGTTGACAAGTCCTCCAGCACAGCCTTCATGCATCTCAACAGCCTGACATCTGAGGACTCTGCAGTCTATTACTGTGCAAGA
L-Part2	Leader	GTGTCCACTCT
L-Part1	Leader	ATGGAATGGAGTGGATCTTTCTCTTCCTCCTGTCAGGAACTACAG

IGHV1.27.P	IGHV	CCACACAAGTGTCCACTCTGAGGTCCACCTGTAGCAGTCTG GACCTGAGCTGGTGAAGCCTGGGGATTCAGTGAAGATATCC TGCAAGGCTTCTGGTTACTCATTCACTGGCTACTACATGCAC TGAGTGAAGTCCTGTAAAAAGCCTTAAAAGGATTGGATATATT AATCCAGTGGTGGTACTGGCTACACTGAAAAGTTCAAGGGCA AGGCCACATTGACTGCAGACAAATCCTTCAGCACAGCCTATA TGTACTTCAGCAGCCTGACATCTGAGGACTCTGTGGACTCTT ACTGTGCAAGA
IGHV1.28.P	IGHV	GAGGTCCAGTTGCAGCAGTCTGGACCTGAGCTGGTGAAGAC TGGGGCTTCAGTGAAGATATCCTGCAAGGCTTCTGGTTACTC ATTCACTGGCTACTACATGCACTGGGTGAAGCAAAGTCCTGA AAAGAGTTTTGAGTGAATTGGAGAGATCAATCCTAGCACTGG TGGTACTAGCTACAACCAGAAGTTCAAGGGCAAGGCCACATT GACTGTAGACAAATCCTCCAGCACAGCCTACATGCAGCTCAA GAGCCTGACATCTGAAGACTGCAGTCTATTACTGTGCAAGA
IGHV1.29.P	IGHV	AAGGTCAAGCTGCAGCAGTCTGGACCTGAGCTGGTGAGCCT GGGGCTTCAGTGAAGATATCCTGCAAGGCTTCTCATTACTCA TTCACTGGCTACTACATGTACTGGGTGAAGCAGAGCCATGGA AATAGCCTTGAGTAGATTGGACTTATTATTCCTTACAATGGTG ATACTGGCTACAACCAGAAGTTTAAGGGCAAGGCCACATTGA CTGTAGACTAGTCCTTCAGCACAGCCTACATGGAGCTCCACA GCCTGAAATCTTAGGACTCTGTGGTCTATTACTGTACAAGA
RSS Pass	RSS	CACATTGTTGTAACCACATCCTGAGTGTGTCAAAAACCT
IGHV1.32	IGHV	GAGGTCCAGCTTCAGCAGTCTGGGCCTGAGCTGGGGAAGC CTGGGGCTTCAGTGAAGATATCCTGCAAGGCTTCTGGTTACT CATTCACTGGCTACAACATGTACTGGGTGAAGCAGAGCCATA GGAAAAGCCTTGAGTGGATTGGATATATTGATCCTTACAATG GTGGTACTAGCTACAACCAGAAGTCCAAGGGCAAGGCCACA TTGACTGTAGACAAATCTTCCAGCACAGCCTACATGCATCTC AACAGCCTGACATCTGAGGACTCTGCAATCTATTACTGTGCA AGA
L-Part2	Leader	GTGTCCACTCT
L-Part1	Leader	ATGGAATGGAGCTGGATCTTTCTCTTCCTCCTGTCAGGAACT GCAG
IGHV1.30.P	IGHV	GGTCCAGCTGTAGCAGTCTGGACCTGAGCTGGTGAAGCCTG GCTCTTCAGTGAAGATATCCTGCAAGGCTTCTGGTTACTCAT TCACTGGCTACTACATACACTGGGTGAAGTCCTGTAAAAAGC CTTCAAAGGATTGGATATATTAATCCAGTGGTGGTACTGGCT ACAATGAAAAGTTCAAGGACAAGGCCACATTGACTGCAAACA AATCCTTCAGCACAGCCTATATGCAATTCAGCAGCCTGACAT CTGAGGACTCTGTGGACTCTTACTGTGCAAGA
RSS Pass	RSS	CACGGTGCTACAAACACATCCTGAGTGTGTCAGAAACCC

IGHV1.33	IGHV	GAGGTTCAGCTGCAGCAGTCTAGACCTGAGCTGGTGAAGCCTGGGGCTTCAGTGAAGATATCCTGCAAGGCTTCTGGTTACTCATTCACTGGCTACTACATGCACTGGGTGAAGCAGAGCCATGGAAAGAGCCTTGAGTGGATTGGATATATTTATCCTTACAATGGTTTTTCTAGCTACAACCAGAAGTTCAAGGGCAAGGCCACATTGACTGTAGACAAGTCCTCTAGCACAGCCTACATGGAGCTCCGCAGCCTGACATCTGAGGACTCTGCAGTCTATTACTGTGCAAGA
L-Part2	Leader	GGTTCCACTCT
L-Part1	Leader	ATGGAATGGAGCTGGATCTTTCTTTTCCTCCTGTCAGCAACTACAG
IGHV1.31.P	IGHV	CCAGCTGCAGCAGTCTGGACCTGAACTGGTGAAGCCTGGGGCTTCAGTGAAGATACCTGCAAGACTTCTGGTTAATCTTTCACTGGCTACTACATGCACTGGGTGAAGCAGAGCCATGGAAATAGCCTTGAGTGGATTTGATATATTGATCCTTACGATGGTGTTACTAGCTACAATAAGAAGTTCAAGGGAAAGGCCACATTGACTGTAGACAAGTTCTCCAGCACAGCCTACATGGAGCTCTAAAACCTGACATCTGAGAAAACTGCGCTCTATTACTGTGCAAGA
RSS Pass	RSS	CACAGTGCTACAAACACATCCTTAGTGTGTCAGAAAAC
IGHV1.34	IGHV	GAGGTCCAGCTGCAACAGTCTGGACCTGAGCTGGTGAAGCCTGGGGCTTCAGTGAAGATGTCCTGCAAGGCTTCTGGATACACCTTCACTGACTACTACATGAAGTGGGTGAAGCAGAGCCATGGAAAGAGCCTTGAGTGGATTGGAGATATTAATCCTAACAATGGTGATACTTTCTACAACCAGAAGTTCAAGGGCAAGGCCACATTGACTGTAGACAAATCCTCCAGCACAGCCTACATGCAGCTCAACAGCCTGACATCTGAGGACTCTGCAGTCTATTACTGTGCAAGA
L-Part2	Leader	GTGTCCTCTCT
L-Part1	Leader	ATGGGATGGAGCTGGATCTTTCTCTTTCTCTTGTCAGGAACTGGAG
RSS Pass	RSS	CACTGTACTACTAACACATCCTTAGTGTGTCAGAAAAC
IGHV1.35	IGHV	GAGGTTCAGCTGCAGCAGTCTGGACCTGAGCTGGTGAAGCCTGGGGCTTCAGTGAAGATATCCTGCAAGGCTTCTGGTTACTCATTTACTGGCTACTTTATGAACTGGGTGATGCAGAGCCATGGAAAGAGCCTTGAGTGGATTGGACGTATTAATCCTTACAATGGTGATACTTTCTACAACCAGAAGTTCAAGGGCAAGGCCACATTGACTGTAGACAAATCCTCTAGCACAGCCCACATGGAGCTCCGGAGCCTGGCATCTGAGGACTCTGCAGTCTATTATTGTGCAAGA
L-Part2	Leader	GTGTGTTCTCT
L-Part1	Leader	ATGGGATGGAGCTGGATCTTTCTCTTTCTCCTGTCAGTAACTGCAG

IGHV1.32.P	IGHV	GAGGTCCAGCTGCAACAGTCTGGACCTGAGTTGGTGAAGCCTGGGCCTTCAGTGAAGATGTCCTGCAAGGCTTCTGGATTCACATTCACTGACTAATACATGCACTGGGTGAAGCAGAGCCATGGAAAGAGCCTTGAGTGGATTGGACATGTTTATCCTTACAATGGTGGTACTAGCTACAACCAGAAATTCAAGGGCAAGGCCACATTGACTGTCGACAATACCTCCAGCACAGCCTACATGGAGCTCGGCAGCCTGACTTCTGAGGACTCTGTGGTCTATTACTCTGCAAGA
IGHV1.33.P	IGHV	GAGGTCCAGCTGAAGCAGCCTGGCACTGTGGTGGTGAAACCTGGGGCTTCAGTTAAGATATCCTGCCAGGCTTCTGGTTACTCATTTACTGGCTACTACATGCACTGGGTGAAGCAGAGCCATGAAAAGAGCCTTTAGTGGATTGGACTTATTATTCCTTACAATGGTGATACTAGCAACAACCAGAAGTTCAAGGGCAAAGCCACATTGACTGTAGACAAGTCCTCCAGCACAGCCAACATGGAGCTCTGCAGCCTGACATCTGAGGACTCTACGGTCTATTACCGTGCAAGA
RSS Pass	RSS	CACAGTGCTACAAACACATCCTGAGTGTGTCAGAAAACT
IGHV1.36	IGHV	GAGGTCCAGCTGCAACAGTCTGGACCTGAGCTGGTGAAGCCTGGGGCTTCAGTGAAGATATCCTGCAAGGCTTCTGGCTACACATTCACTGACTACTACATGAACTGGGTGAAGCAGAGCCATGGAAAGAGCCTTGAGTGGATTGGACTTGTTAATCCTAACAATGGTGGTACTAGCTACAACCAGAAGTTCAAGGGCAAGGCCACATTGACTGTAGACAAGTCCTCCAGCACAGCCTACATGGAGCTCCGCAGCCTGACATCTGAGGACTCTGCGGTCTATTACTGTGCAAGA
L-Part2	Leader	GTGTCCACTCT
L-Part1	Leader	TGGGGTGGAGTCTGGATCTTTTTCTTCCTCCTGTCAGGAACTGCAG
RSS Pass	RSS	CACTGTACTACAAACACATCCTTAGTGTGTCAGAAAAC
IGHV1.37	IGHV	GAGGTTCAGCTGCAGCAGTCTGGACCTGAGCTGGTGAAGCCTGGGGCTTCAGTGAAGATATCCTGCAAGGCTTCTGGTTACTCATTTACTGGCTACTTTATGAACTGGGTGAAGCAGAGCCATGGAAAGAGCCTTGAGTGGATTGGACGTATTAATCCTTACAATGGTGATACTTTCTACAACCAGAAGTTCAAGGGCAAGGCCACATTGACTGTAGACAAATCCTCTAGCACAGCCCACATGGAGCTCCTGAGCCTGACATCTGAGGACTCTGCAGTCTATTATTGTGGAAGA
L-Part2	Leader	GTGTGTTCTCT
L-Part1	Leader	ATGGGATGGAGCTGTATCTTTCTCTTTCTCCTGTCAGTAACTGTAG

IGHV1.34.P	IGHV	GAGGTCCAGCTGCAGCAGTTTGGACCTGAGCTGGTGAAGCCTGGGGCTTCAGTGAAGATATCCTGCAAGGCTTCTGGTTACCCATTCACTGGTTACTACATGCACTGGGTGAAGCAGAGCCATGGAAAGAGCCTTGAGTAGATTAGACTTATTATTCCTTACGATGGTGATACTTTCTACAACCAGAAGTTCAAGGGCAAGGCCACATTGACTGTAGACTAGTCCTTCAGCACAACCTACATGGAGCTCCGCAGCCTGAAATCTGAGGACTCTGCGGTCTATTACTCTACAAGA
RSS Pass	RSS	CACAGTGTTGTAACCACATCCTGAGTGTGTCAGAAAAC
IGHV1.38	IGHV	GAGATCCAGCTGCAGCAGACTGGACCTGAGCTGGTGAAGCCTGGGGCTTCAGTGAAGATATCCTGCAAGGCTTCTGGTTATTCATTCACTGACTACATCATGCTCTGGGTGAAGCAGAGCCATGGAAAGAGCCTTGAGTGGATTGGAAATATTAATCCTTACTATGGTAGTACTAGCTACAATCTGAAGTTCAAGGGCAAGGCCACATTGACTGTAGACAAATCTTCCAGCACAGCCTACATGCAGCTCAACAGTCTGACATCTGAGGACTCTGCAGTCTATTACTGTGCAAGA
L-Part2	Leader	GTGCCCACTCT
L-Part1	Leader	ATGGGAGGGATCTGGATCTTTCTCTTCCTCCTGTCAGGAACTGCAG
IGHV1.35.P	IGHV	GAGGTCCAGATGCAGCAGTCTGGATCTAAGGTAGTGAATGCTGGGGCTTCAGTGAAGATATCCTGCAAGGCTTCTGGTTACTCATTCACTGGCTACTACATGCACTGGGTGAAGCAGAGCCATGGAAAGAGCCTTGAGTCGATTGGAGATATTAATCTTTCCAATGGTGGTACTAACTACAATGGAAAGTTCAAAAGCAAGGCCACATTGACTGTAGATATATCCTCTAGCACAGCCTATATGGAGCTAGCATATTGACATCTGAGGTCTCTGCAGTCTCTCACTGTGCAATA
RSS Pass	RSS	CACAGTGTTGTAACCACATCCTGAGTGTGTCAGAAAACC
IGHV1.39	IGHV	GAGGTCCAGCTGCAGCAGTCTGGACCTGAGCTGGAGAAGCCTGGCGCTTCAGTGAAGATATCCTGCAAGGCTTCTGGTTACTCATTCACTGGCTACAACATGAACTGGGTGAAGCAGAGCAATGGAAAGAGCCTTGAGTGGATTGGAAATATTGATCCTTACTATGGTGGTACTAGCTACAACCAGAAGTTCAAGGGCAAGGCCACATTGACTGTAGACAAATCCTCCAGCACAGCCTACATGCAGCTCAAGAGCCTGACATCTGAGGACTCTGCAGTCTATTACTGTGCAAGA
L-Part2	Leader	GTGTCCACTCT
L-Part1	Leader	ATGGGATGGACCTGGATCTTTATTTTAATCCTGTCAGTAACTACAG
IGHV1.36.P	IGHV	GAGGTCCAGCTGCAACAGTCTGGACCTGAGCTGGTGAAGCCTGGAGCTTCAATGAAGATATCCTGCAAGGCTACTCATTCACTGGCTACACCATGAACTGGGTGAAGCAGAGCCATGGAAAGAACCTTGAGTGGATTGGACTTATTAATCCTTACAATGGTGGTACTAGCTACAACCAGAAGTTCAAGGGCAAGGCCACATTAACTGTA

		GACAAGTCATCCAGCACAGCCTACATGGAGCTCCTCAGTCT GACATCTGAGGACTCTGCAGTCTATTACTGTGCAAGA
RSS Pass	RSS	CACGATGCTACAAATACACCCAGAGTGAGTCAGAAACC
IGHV1.40	IGHV	GAGGTCCAGCTGCAGCAGTCTGGACCTGACCTGGTGAAGCC TGGGGCTTCAGTGAAGATATCCTGCAAGGCTTCTGGTTACTC ATTCACTGGCTACTACATGCACTGGGTGAAGCAGAGCCATG GAAAGAGCCTTGAGTGGATTGGACGTGTTAATCCTAACAATG GTGGTACTAGCTACAACCAGAAGTTCAAGGGCAAGGCCATAT TAACTGTAGACAAGTCATCCAGCACAGCCTACATGGAGCTCC GCAGCCTGACATCTGAGGACTCTGCGGTCTATTACTGTGCAA GA
L-Part2	Leader	GTGTCCACTCT
L-Part1	Leader	ATGGAATGGAGCTGGGTCTTTCTCTTCCTCCTGTCAGTAACT ACAG
IGHV1.37.P	IGHV	CAGGTCCAGCTGCAGCAGTCTGGGGCTGAAATGGTGAGGCC TGGGGCCTCAGTGAAAATGTCCTGCAAGGCTTTTGGCTACAC ATTCACTGACTATGGCATGCATTGGGTGAAGCATAGTCACGG AAGGCATCCTTGAGAGGATTGGAAATGTTAATACTTACTATG ACAATACTGGCTACAATGAAAAGTTCAAGGGCAAGTCCAAAT TGACTGTAGACAAATCCTCCAGCACAGCCTATGTGGAGTTTA GCAGAATGACATCTGAGGATTCTGTAGTCTATTACTGTGAAA
IGHV1.38.P	IGHV	CAGGTTCATATGCAGCAGTCTAGGAGAAGAGGGGATGAAGC CTGTGTTCTCAGTGAAGATGTCCTGTAAGGGTTCTGGCTACA GCTTTACCAACTACTATATGCCCTGGGTAAAATAGTGGACTG GACACAGCCTTGAGTGGATTAGATGGATTCATCCTGGAAATG GTGATACTTACTACAATCAAAAGTTCAAGGGAAAGGCAACAC TGACCAAGTACAAATTCTCCAGCACAGCCTACTTACATCACA ACAGGCTGACATCTGAGCACCCAGTAGTTTATAAATGTGCGA GA
IGHV1.39.P	IGHV	CAGGTCCAGCTGCAGCAGTCTGCAGCTGAGCTGGTGAAGCC TGGGAGTCCAGTGAAGCTGTCCTGCAAAGCTTCTGGCTACA CCGTCAATGACAACTATATGGAGCAGGTAAAGCAGAGGCCT GGACAGAGCATGGAATGGATTGGATAGATTCATTTTGTATAT GGTGGTACTTAATACAATGAAAAGTTCTAGGGCAAGTCCACA TTAACTGTAGAAAAATCCTCCAACACAGCCTACATGGAACTC CACAGCTCAACATCTGAGGTCTCTGCGGTTTATTTCTGTGCA CGG
RSS Pass	RSS	CACAGTGTTGTAACCACATCCTGTGTGTGTCAGAAAAC

IGHV1.41	IGHV	CAGGTTCAGCTGCAGCAGTCTGGGGCTGAGCTGGTGAAGCCTGGGGCCTCAGTGAAGATGTCCTGCAAGGCTTTTGGCTACACCTTCACTACCTATCCAATAGAGTGGATGAAGCAGAATCATGGGAAGAGCCTAGAGTGGATTGGAAATTTTCATCCTTACAATGATGATACTAAGTACAATGAAAAATTCAAGGGCAAGGCCAAATTGACTGTAGAAAAATCCTCTAGCACAGTCTACTTGGAGCTCAGCCGATTAACATCTGATGACTCTGCTGTTTATTACTGTGCAAGG
L-Part2	Leader	GTGTCCACTCC
L-Part1	Leader	ATGGCGTGGATCTCTATCATCCTCTTCCTAGTGGCAACAGCTATAG
RSS Pass	RSS	CACATTGACACAGAGTCAGTTTTCAGCTGTACAATAGTT
IGHV8.3	IGHV	CAGATTACTCTGAAAGAGTCTGGCCCTGGGATAGTGCAGCCATCCCAGCCCTTCAGACTTACTTGCACTTTCTCTGGGTTTTCACTGAGCACTTCTGGTATAGGTGTAACCTGGATTCGTCAGCCCTCAGGGAAAGGTCTGGAGTGGCTGGCAACGATTTGGTGGGATGATGATAACCGCTACAACCCATCTCTAAAGAGCAGGCTCACAGTCTCCAAAGACACCTCCAACAACCAAGCATTCCTGAATATCATCACTGTGGAAACTGCAGATACTGCCATATACTACTGTGCTCAGAGTG
L-Part2	Leader	ATGTTATGTCC
L-Part1	Leader	ATGGGCAGGCTTATGTCTTCATTCCTGCTTCTGGTTGCCCCTTTAT
RSS Pass	RSS	CACAGTGTTGTAACCACATCCTGAGTGTGTCAGAAACCC
IGHV1.42	IGHV	CAGGTCCAGTTGCAGCAGTCTGGAGCTGAGCTGGTAAGGCCTGGGACTTCAGTGAAGATATCCTGCAAGGCTTCTGGCTACACCTTCACTAACTACTGGCTAGGTTGGGTAAAGCAGAGGCCTGGACATGGACTTGAGTGGATTGGAGATATTTACCCTGGAGGTGGTTATACTAACTACAATGAGAAGTTCAAGGGCAAGGCCACACTGACTGCAGACACATCCTCCAGCACTGCCTACATGCAGCTCAGTAGCCTGACATCTGAGGACTCTGCTGTCTATTTCTGTGCAAGA
L-Part2	Leader	ATGTCCACTCC
L-Part1	Leader	ATGGAATGGAGCGGGGTCTTTATCTTTCTCTTGTCAGTAACTGCAG
RSS Pass	RSS	CACAGTGTTGCAACCACATCCTGAGAGTGTCAGAAACCC
IGHV1.43	IGHV	CAGGTCCAACTGCAGCAGCCTGGGGCTGAGCTGGTGAGGCCTGGGGCTTCAGTGAAGCTGTCCTGCAAGGCTTCTGGCTACACCTTCACCAGCTACTGGATAAACTGGGTGAAGCAGAGGCCTGGACAAGGCCTTGAGTGGATCGGAAATATTTATCCTTCTGATAGTTATACTAACTACAATCAAAAGTTCAAGGACAAGGCCACATTGACTGTAGACAAATCCTCCAGCACAGCCTACATGCAGCT

		CAGCAGCCCGACATCTGAGGACTCTGCGGTCTATTACTGTACAAGA
L-Part2	Leader	GTGTCCACTCC
L-Part1	Leader	ATGGGATGGAGCTGTATCATCCTCTTCTTGGTAGCAACAGCTACAG
RSS Pass	RSS	CACAGTGTTGCAACCACATCCTGAGAGTGTCAGAAACCC
IGHV1.44	IGHV	CAGGTCCAACTGCAGCAACCTGGGTCTGAGCTGGTGAGGCCTGGAGCTTCAGTGAAGCTGTCCTGCAAGGCTTCTGGCTACACATTCACCAGCTACTGGATGCACTGGGTGAAGCAGAGGCATGGACAAGGCCTTGAGTGGATTGGAAATATTTATCCTGGTAGTGGTAGTACTAACTACGATGAGAAGTTCAAGAGCAAGGGCACACTGACTGTAGACACATCCTCCAGCACAGCCTACATGCACCTCAGCAGCCTGACATCTGAGGACTCTGCGGTCTATTACTGTACAAGA
L-Part2	Leader	GTGTCCACTCC
L-Part1	Leader	ATGGGATGGAGCTGTATCATCCTCTTCTTGGTAGCAACAGCTACAG

Supplemental Table 3. Annotated BALB/cByJ IGHD sequences, heptamers, and nonamers.

Name	Type	Length	Sequence
	Nonamer	9	ACAAAAACC
	Heptamer	7	CACGGTG
IGHD4-1*01 (DQ52)	D-Gene	11	CTAACTGGGAC
	Heptamer	7	CACTGTG
	Nonamer	9	GTCAAAACC
	Nonamer	9	ACCAAAACT
	Heptamer	7	CACTGTA
IGHD3-1 (DST4) (3-2*01)	D-Gene	16	AGACAGCTCGGGCTAC
	Heptamer	7	CACTGTG
	Nonamer	9	TCAAAATCC
	Nonamer	9	CTCAAATCC
	Heptamer	7	CACATTG
IGHD5	D-Gene	10	GAATACCTAC
	Heptamer	7	CACGGTG
	Nonamer	9	ACAGAATCC
	Nonamer	9	ACAAAAACC
	Heptamer	7	CACAGTG
IGHD2-10*01(DSP2.7) (2-10*01)	D-Gene	17	CCTACTATGGTAACTAC
	Heptamer	7	CACAGTA
	Nonamer	9	ACAAAAATC
	Nonamer	9	CTCAAATTC
	Heptamer	7	CACATTG
IGHD5	D-Gene	10	GAATACCTAC

	Heptamer	7	CACGGTG
	Nonamer	9	ACAGAATCC
	Nonamer	9	ACAAAAACC
	Heptamer	7	CACAGTG
IGHD2-8 (DSP2.8) (2-10*02)	D-Gene	17	CCTAGTATGGTAACTAC
	Heptamer	7	CACAGTA
	Nonamer	9	ACAAAAATC
	Nonamer	9	ACAAGAAAG
	Heptamer	7	CACAATG
IGHD6	D-Gene	29	AGGCAGCTAGCCTCTGCAGTGCCACAACC
	Heptamer	7	GAATTCA
	Nonamer	9	TTATAGAGG
	Nonamer	9	CTCAAATTC
	Heptamer	7	CACATTG
IGHD5	D-Gene	10	GAATACCTAC
	Heptamer	7	CACGGTG
	Nonamer	9	ACAGAATCC
	Nonamer	9	ACAAAAACC
	Heptamer	7	CACAGTG
IGHD2-4*01(DSP2.2b)	D-Gene	17	TCTACTATGATTACGAC
	Heptamer	7	CACAGTA
	Nonamer	9	ACAAAAATC
	Nonamer	9	ACAAGAAAG
	Heptamer	7	CACAATG
IGHD6	D-Gene	29	AGGCAGCTAGCCTCTGCAGTGCCACAACC
	Heptamer	7	GAATTCA
	Nonamer	9	TTATAGAGG
	Nonamer	9	CTCAAATTC

	Heptamer	7	CACATTG
IGHD5-6	D-Gene	10	GAATACCTAC
	Heptamer	7	CACGGTG
	Nonamer	9	ACAGAATCC
	Nonamer	9	ACAAAAACC
	Heptamer	7	CACAGTG
IGHD2-14*01(DSP2.11)	D-Gene	17	CCTACTATAGGTACGAC
	Heptamer	7	CACAGTA
	Nonamer	9	ACAAAAATC
	Nonamer	9	CTCAAATTC
	Heptamer	7	CACATTG
IGHD5-6	D-Gene	10	GAATACCTAC
	Heptamer	7	CACGGTG
	Nonamer	9	ACAGAATCC
	Nonamer	9	ACAAAAACC
	Heptamer	7	CACAGTG
IGHD2-4*01(DSP2.2a)	D-Gene	17	TCTACTATGATTACGAC
	Heptamer	7	CACAGTA
	Nonamer	9	ACAAAAATC
	Nonamer	9	ACAAGAAAG
	Heptamer	7	CACAATG
IGHD6	D-Gene	29	AGGCAGCTAGCCTCTGCAGTGCCACAACC
	Heptamer	7	GAATTCA
	Nonamer	9	TTATAGAGG
	Nonamer	9	CTCAAATTC
	Heptamer	7	CACATTG
IGHD5	D-Gene	10	GAATACCTAC
	Heptamer	7	AACGGTG

	Nonamer	9	ACAGAATCC
	Nonamer	9	ACAAAAACC
	Heptamer	7	CACAGTG
IGHD2-8 (DSP2.5)	D-Gene	17	TCTACTATGGTAACTAC
	Heptamer	7	CACAGTA
	Nonamer	9	ACAAAAATC
	Nonamer	9	ACAAAAACA
	Heptamer	7	CACAGTG
IGHD1-2*01 (DFL16.2)	D-Gene	17	TTCATTACTACGGCTAC
	Heptamer	7	CACAGTA
	Nonamer	9	ACAAAAAGC
	Nonamer	9	ACCAAACTG
	Heptamer	7	CACTGAA
IGHD3-3*01(DST4.3)	D-Gene	17	GGGACAGCTAGGGCTGT
	Heptamer	7	CACTGTG
	Nonamer	9	TCAAAATCC
	Nonamer	9	CTCAAATCC
	Heptamer	7	TACAGTG
IGHD5-1*01 (DMB3)	D-Gene	10	GAGTACCTAC
	Heptamer	7	TACAGTG
	Nonamer	9	ATAGAATCC
	Nonamer	9	ACAAAAACC
	Heptamer	7	CACAGTG
IGHD2-7*01/2-2*01 (DSP2.3)	D-Gene	17	TCTACTATGGTTACGAC
	Heptamer	7	CACAGTA
	Nonamer	9	ACAAAAATC
	Nonamer	9	ACAAGAAAG
	Heptamer	7	CACAATG

IGHD6	D-Gene	29	AGGCAGCTAGTCTCTGCAGTGCCAC AACC
	Heptamer	7	GAATTCA
	Nonamer	9	TTATAGAGG
	Nonamer	9	CTCAAATCC
	Heptamer	7	TACAGTG
IGHD5-1*01 (DMB2)	D-Gene	10	GAGTACCTAC
	Heptamer	7	CACAGTG
	Nonamer	9	ATAGAATCC
	Nonamer	9	ACAAAAACC
	Heptamer	7	CACAGTG
IGHD2-3*01 (DSP2.9)	D-Gene	17	TCTATGATGGTTACTAC
	Heptamer	7	CACAGTA
	Nonamer	9	ACAAAAATC
	Nonamer	9	ACAAGAAAG
	Heptamer	7	CACAATG
IGHD6	D-Gene	29	AGACAGCTAGCCTCTGCAGTGCCAC AACC
	Heptamer	7	GAATTCA
	Nonamer	9	TTATAGAGG
	Nonamer	9	GCAAAAACC
	Heptamer	7	CACAGTG
IGHD1-1*01 (DFL16.1)	D-Gene	23	TTTATTACTACGGTAGTAGCTAC
	Heptamer	7	CACAGTA
	Nonamer	9	ACAAAAAGC
	Nonamer	9	CAAAACTGC
	Heptamer	9	CACTGTAAG
IGHD3-1*01 (DST4.2)	D-Gene	17	GGCACAGCTCGGGCTAC
	Heptamer	7	CACTGTG

	Nonamer	9	TCAAAATCC
	Nonamer	9	CATGGAAGA
	Heptamer	7	CACAGTG
IGHD5 (DMB1)	D-Gene	10	GAGTACCTAC
	Heptamer	7	CACAGTG
	Nonamer	9	ACTGAATTC
	Nonamer	9	GCAAAAGTT
	Heptamer	7	CACAATG
IGHD1-3*01 (DFL 16.3)	D-Gene	23	TATATAACTAAAGTGGTAGCTCA
	Heptamer	7	CACAGTA
	Nonamer	9	TCCAAAAAC
	Nonamer	9	ACTGGGATA
	Heptamer	7	AAAAGAC
IGHD5	D-Gene	10	GAATACCTAC
	Heptamer	7	CTCTGTT
	Nonamer	9	CACATATAA

Supplemental Table 4. Annotated BALB/cByJ IGHJ sequences, heptamers, and nonamers.

Name	Type	Length	Sequence
IGHJ4*01	J-Gene	54	ATTACTATGCTATGGACTACTGGGGTCAAGGAACCTCAGTCACCGTCTCCTCAG
	Heptamer	7	CACAATA
	Nonamer	9	CAAAAACCC
IGHJ3*01	J-Gene	48	CCTGGTTTGCTTACTGGGGCCAAGGGACTCTGGTCACTGTCTCTGCAG
	Heptamer	7	CACATTG
	Nonamer	9	ACAATAAAT
IGHJ2*01	J-Gene	48	ACTACTTTGACTACTGGGGCCAAGGCACCACTCTCACAGTCTCCTCAG
	Heptamer	7	CACACTA
	Nonamer	9	ACAAAAACC
IGHJ1*01	J-Gene	53	CTACTGGTACTTCGATGTCTGGGGCGCAGGGACCACGGTCACCGTCTCCTCAG
	Heptamer	7	CACAGTC
	Nonamer	9	CTAAAACTC

Supplemental Table 5. Annotated BALB/cByJ IGHC sequences.

Name	Type	Length	Sequence
IGHA Exon 4 of 4	C-Gene	197	AACGTCAAGAGCCACTTTCCTATGTGCTACTGGAC CAGTCACAAGACATCCTGGAGGAAGAGGCCCCAG GTGCCAGCCTGTGGCCCACCACTGTGACCTTCCT CACCCTCTTCCTACTGAGCTTGTTCTACAGCACAG CACTCACTGTTACAACTGTTCGAGGCCCGTTTGGC AGCAAAGAGGTCCCCCAGTACTGA
IGHA Exon 3 of 4	C-Gene	333	TGAACACCTTCCCACCCCAGGTCCACCTGCTACCG CCGCCGTCGGAGGAGCTGGCCCTGAATGAGCTCT TGTCCCTGACATGCCTGGTGCGAGCTTTCAACCCT AAAGAAGTGCTGGTGCGATGGCTGCATGGAAATG AGGAGCTGTCCCCAGAAAGCTACCTAGTGTTTGAG CCCCTAAAGGAGCCAGGCGAGGGAGCCACCACCT ACCTGGTGACAAGCGTGTTGCGTGTATCAGCTGAA ACCTGGAAACAGGGTGACCAGTACTCCTGCATGG TGGGCCACGAGGCCTTGCCCATGAACTTCACCCA GAAGACCATCGACCGTCTGTCGG
IGHA Exon 2 of 4	C-Gene	336	GTCCTACTCCTCCTCCTCCTATTACTATTCCTTCCT GCCAGCCCAGCCTGTCACTGCAGCGGCCAGCTCT TGAGGACCTGCTCCTGGGTTCAGATGCCAGCATC ACATGTACTCTGAATGGCCTGAGAAATCCTGAGGG AGCTGTCTTCACCTGGGAGCCCTCCACTGGGAAG GATGCAGTGCAGAAGAAAGCTGTGCAGAATTCCT GCGGCTGCTACAGTGTGTCCAGCGTCCTGCCTGG CTGTGCTGAGCGCTGGAACAGTGGCGCATCATTC AAGTGCACAGTTACCCATCCTGAGTCTGGCACCTT AACTGGCACAATTGCCAAAGTCACAG
IGHA Exon 1 of 4	C-Gene	303	AGTCTGCGAGAAATCCCACCATCTACCCACTGACA CTCCCACCAGCTCTGTCAAGTGACCCAGTGATAAT CGGCTGCCTGATTCACGATTACTTCCCTTCCGGCA CGATGAATGTGACCTGGGGAAAGAGTGGGAAGGA TATAACCACCGTAAACTTCCCACCTGCCCTGGCCT CTGGGGGACGGTACACCATGAGCAGCCAGTTGAC CCTGCCAGCTGTCGAGTGCCCAGAAGGAGAATCC GTGAAATGTTCCGTGCAACATGACTCTAACCCCGT CCAAGAATTGGATGTGAATTGCTCTG
IGHE Exon 6 of 6	C-Gene	84	GTGAAGTGGGTCTTCTCCACACCGATGCAGGATA CACCCCAGACCTTCCAAGACTATGCCAACATCCTC CAGACCAGGGCATAG

IGHE Exon 5 of 6	C-Gene	134	AGCTAGACCTCCAGGACCTATGTATTGAAGAGGTG GAGGGCGAGGAGCTGGAAGAGCTGTGGACCAGT ATTTGTGTCTTCATCACCCTGTTCCTGCTCAGTGTG AGCTATGGGGCCACTGTCACCGTCCTCAAG
IGHE Exon 3 of 6 (CH4)	C-Gene	327	GCCAGCGCTCAGCCCCCGAGGTATATGTGTTCCC ACCACCAGAGGAGGAGAGCGAGGACAAACGCACA CTCACCTGTTTGATCCAGAACTTCTTCCCTGAGGA TATCTCTGTGCAGTGGCTGGGGGATGGCAAACTG ATCTCAAACAGCCAGCACAGTACCACAACACCCCT GAAATCCAATGGCTCCAATCAAGGCTTCTTCATCTT CAGTCGCCTAGAGGTCGCCAAGACACTCTGGACA CAGAGAAAACAGTTCACCTGCCAAGTGATCCATGA GGCACTTCAGAAACCCAGGAAACTGGAGAAAACA ATATCCACAAGCCTTG
IGHE Exon 3 of 6 (CH3)	C-Gene	321	ATCATGAGCCACGGGGTGTGATTACCTACCTGATC CCACCCAGCCCCCTGGACCTGTATCAAAACGGTG CTCCCAAGCTTACCTGTCTGGTGGTGGACCTGGAA AGCGAGAAGAATGTCAATGTGACGTGGAACCAAG AGAAGAAGACTTCAGTCTCAGCATCCCAGTGGTAC ACTAAGCACCACAATAACGCCACAACTAGTATCAC CTCCATCCTGCCTGTAGTTGCCAAGGACTGGATTG AAGGCTACGGCTATCAGTGCATAGTGGACCACCC TGATTTTCCCAAGCCCATTGTGCGTTCCATCACCA AGACCCCAG
IGHE Exon 2 of 6 (CH2)	C-Gene	321	TTCGACCTGTCAACATCACTGAGCCCACCTTGGAG CTACTCCATTCATCCTGCGACCCCAATGCATTCCA CTCCACCATCCAGCTGTACTGCTTCATTTATGGCC ACATCCTAAATGATGTCTCTGTCAGCTGGCTAATG GACGATCGGGAGATAACTGATACACTTGCACAAAC TGTTCTAATCAAGGAGGAAGGCAAACTAGCCTCTA CCTGCAGTAAACTCAACATCACTGAGCAGCAATGG ATGTCTGAAAGCACCTTCACCTGCAAGGTCACCTC CCAAGGCGTAGACTATTTGGCCCACACTCGGAGA TGCCCAG
IGHE Exon 1 of 6 (CH1)	C-Gene	273	CCTCTATCAGGAACCCTCAGCTCTACCCCTTGAAG CCCTGTAAAGGCACTGCTTCCATGACCCTGGGCT GCCTGGTAAAGGACTACTTCCCTGGTCCTGTGACT GTGACCTGGTATTCAGACTCCCTGAACATGAGCAC TGTGAACTTCCCTGCCCTCGGTTCTGAACTCAAGG TCACCACCAGCCAAGTGACCAGCTGGGGCAAGTC AGCCAAGAACTTCACATGCCACGTGACACATCCTC CATCATTCAACGAAAGTAGGACTATCCTAG

IGHG2A M2	C-Gene	81	GTAAAGTGGATCTTCTCCTCTGTGGTGGAGCTGAAGCAGACGATCTCCCCTGACTACAGAAACATGATTGGGCAGGGAGCC
IGHG2A M1	C-Gene	131	GGCTAGACCTGGATGATGTCTGTGCTGAGGCCCAGGACGGGGAGCTGGACGGCCTCTGGACGACCATCACCATCTTCATCAGCCTCTTCCTGCTCAGCGTGTGCTACAGCGCCTCTGTCACACTCTTCAAG
IGHG2B Exon 4 of 6 (CH3)	C-Gene	320	GGTCAGTAAGAGCTCCACAGGTATATGTCTTGCCTCCACCAGAAGAAGAGATGACTAAGAAACAGGTCACTCTGACCTGCATGGTCACAGACTTCATGCCTGAAGACATTTACGTGGAGTGGACCAACAACGGGAAAACAGAGCTAAACTACAAGAACACTGAACCAGTCCTGGACTCTGATGGTTCTTACTTCATGTACAGCAAGCTGAGAGTGGAAAAGAAGAACTGGGTGGAAAGAAATAGCTACTCCTGTTCAGTGGTCCACGAGGGTCTGCACAATCACCACACGACTAAGAGCTTCTCCCGGACTCCGGGTAAA
IGHG2B Exon 3 of 6 (CH2)	C-Gene	330	CACCTAACCTCTTGGGTGGACCATCCGTCTTCATCTTCCCTCCAAAGATCAAGGATGTACTCATGATCTCCCTGAGCCCCATAGTCACATGTGTGGTGGTGGATGTGAGCGAGGATGACCCAGATGTCCAGATCAGCTGGTTTGTGAACAACGTGGAAGTACACACAGCTCAGACACAAACCCATAGAGAGGATTACAACAGTACTCTCCGGGTGGTCAGTGCCCTCCCCATCCAGCACCAGGACTGGATGAGTGGCAAGGAGTTCAAATGCAAGGTCAACAACAAAGACCTCCCAGCGCCCATCGAGAGAACCATCTCAAAACCCAAAG
IGHG2B Exon 2 of 6 (H)	C-Gene	48	AGCCCAGAGGGCCCACAATCAAGCCCTGTCCTCCATGCAAATGCCCAG
IGHG2A Exon 1 of 6 (CH1)	C-Gene	291	CCAAAACAACAGCCCCATCGGTCTATCCACTGGCCCCTGTGTGTGGAGATACAACTGGCTCCTCGGTGACTCTAGGATGCCTGGTCAAGGGTTATTTCCCTGAGCCAGTGACCTTGACCTGGAACTCTGGATCCCTGTCCAGTGGTGTGCACACCTTCCCAGCTGTCCTGCAGTCTGACCTCTACACCCTCAGCAGCTCAGTGACTGTAACCTCGAGCACCTGGCCCAGCCAGTCCATCACCTGCAATGTGGCCCACCCGGCAAGCAGCACCAAGGTGGACAAGAAAATTG

IGHG2B Exon 6 of 6	C-Gene	1433	GTAAAGTGGATCTTCTCCTCAGTGGTGGAGCTGAA GCAGAAGATCTCCCCTGACTACAGAAACATGATTG GGCAGGGAGCCTAGGCCACTTCCTCTGGGATCAG AAGAGCTTCCTAAGTCCTGCAGAAGCCCATCCATC CTACTGTGCAGCCTAACAGGGAGGCCACACTCTA GCCCTATGACTCTCTGATCCAGAACTCTCCATGGT CTCTCCTTTGGAGGACCATGTGCAGAGCAGGCTTT ACCCCAGACCTAAACACTTCCACAGCAGTTGCCAA ATAACTAACTGATCAGCGCCAGAATAACCATCTCC ATCCAAACCAGCACTAGGGGATCTGAGATCTCAGA ATGTTTGCCTAAGATGAACTGGAAATCCAGGTTTC CTGAGTTCAGCTGCATTGACATCAGCCCTGGATTT GGCCTGGACCACACATTTTCATCTGAAATGGATCT TCCCAGAAAGGGCAATGTTTTCAGATACTTGAATG GAGACATCTGCTCAGAAGGCCAGGGTCACAATAG CCCAAAGGTCACTCTGGTATCCTTTGTACCAGCTA TACTTGGACAGCTTCCTTCCTGGTACTTCTGTGAA CCCTGGGCTGAGGAAAGGACCTGTGATCCTCAAA GTGGTGGGATGAGCCCTCTGAAGAAGATGCAGTC TTCTCTTCCTCCTACCCACTCAAGATGCTGGGTTC TCAGATACAGAATCCAGTCCACAGACACTGTTGTC CACAGAACTTCAAGACAAGGCCACTGCCACCTCCA GAGGCAGAAGGACTGCTTTGTGTAGCCTCACAAA GCTCTAGCAGGGACAAGTGTATACACCACTGAGTA CTGGCTAGGTAATCAGGCAGGAAAAGTTCCAGGT GAGGGCTCTGATGAGAAGATTCCTGTTGGGCTCA ATAAAATGGGGGTTGGGGAGAGGAGCAATGTGGG AGAGATGTGAGCACATAGTTAAAAGAAAACCATGG TGCCACAGGTCTGCTCTGAAGGCCTGGTGTCCAG ACTGGACCACTCACAGCAGAAAACATCAGTTGCAA TGGACCTCTAGGTGGACAGATCTAGAATACCAGG GTGATTTTAGCTAATGAAAAGTAGCCCTGTATCCTT TCCTCTTCTACCTGCCTAGGACAATAAGAAGCCAG AGTCTCCAGACAGAAAGAAAGAGCACGCACTTGCT TTTGCGTTAATGGGTCTGATAGGAAATGTTTAGAG AACTTGGACACTTAAGAGCTGTGGATTTTTCTAGA GGTCATGAATGTCTTAAAGTGTTCTCTCTCCCTCTC TCTCTCTCTCTCAGTGTCTCTTTGTGTGTATGTGTG TATGTGTGTTTGTGTGCATGTGTGCGTGCACATGC GGGCGCGCACAGGTGCACACTCATGCATGCTGTT CTTTTGGAATTTGAAACAATAAAAGATTCTTTTAATA TTGTG
IGHG2B	C-Gene	131	GGCTAGACCTGGATGATATCTGTGCTGAGGCCAA GGATGGGGAGCTGGACGGCCTCTGGACGACCATC

Exon 5 of 6			ACCATCTTCATCAGCCTCTTCCTGCTCAGCGTGTG CTACAGCGCCTCTGTCACACTCTTCAAG
IGHG2B Exon 4 of 6	C-Gene	315	GGCTAGTCAGAGCTCCACAAGTATACATCTTGCCG CCACCAGCAGAGCAGTTGTCCAGGAAAGATGTCA GTCTCACTTGCCTGGTCGTGGGCTTCAACCCTGGA GACATCAGTGTGGAGTGGACCAGCAATGGGCATA CAGAGGAGAACTACAAGGACACCGCACCAGTCCT GGACTCTGACGGTTCTTACTTCATATATAGCAAGC TCAATATGAAAACAAGCAAGTGGGAGAAAACAGAT TCCTTCTCATGCAACGTGAGACACGAGGGTCTGAA AAATTACTACCTGAAGAAGACCATCTCCCGGTCTC CGG
IGHG2B Exon 3 of 6	C-Gene	330	CTCCTAACCTCGAGGGTGGACCATCCGTCTTCATC TTCCCTCCAAATATCAAGGATGTACTCATGATCTCC CTGACACCCAAGGTCACGTGTGTGGTGGTGGATG TGAGCGAGGATGACCCAGACGTCCAGATCAGCTG GTTTGTGAACAACGTGGAAGTACACACAGCTCAGA CACAAACCCATAGAGAGGATTACAACAGTACTATC CGGGTGGTCAGCACCCTCCCCATCCAGCACCAGG ACTGGATGAGTGGCAAGGAGTTCAAATGCAAGGT CAACAACAAAGACCTCCCATCACCCATCGAGAGAA CCATCTCAAAAATTAAAG
IGHG2B Exon 2 of 6	C-Gene	66	AGCCCAGCGGGCCCATTTCAACAATCAACCCCTGT CCTCCATGCAAGGAGTGTCACAAATGCCCAG
IGHG2B Exon 1 of 6	C-Gene	291	CCAAAACAACACCCCCATCAGTCTATCCACTGGCC CCTGGGTGTGGAGATACAACTGGTTCCTCCGTGA CTCTGGGATGCCTGGTCAAGGGCTACTTCCCTGA GTCAGTGACTGTGACTTGGAACTCTGGATCCCTGT CCAGCAGTGTGCACACCTTCCCAGCTCTCCTGCA GTCTGGACTCTACACTATGAGCAGCTCAGTGACTG TCCCCTCCAGCACCTGGCCAAGTCAGACCGTCAC CTGCAGCGTTGCTCACCCAGCCAGCAGCACCACG GTGGACAAAAAACTTG

IGHG1 Exon 6 of 6	C-Gene	1396	GTAAAGTGGATCTTCTCCTCGGTGGTGGAGCTGAA GCAGACACTGGTTCCTGAATACAAGAACATGATTG GGCAAGCGCCCTAGGCCACCTCTTGTAATGGCAG GGGATTTCCCAGGCCCCAAAGGACCCTGTCCAAT ATGCCAAGCAGCACAACTGAGATCACACTGTCTGC TCATCTCGCTTTCCTCCGACCCCGAGACTCAGCTA CTCTCAAATTTTCCCTCTCTGAAGGACCATGTGGA CATTACATTGCTCCAGGCCACAGCCACCAGGACCT AAAACACCATCACAGCAGCACCAAAGACACTGGAT AGACCCACAAGAGCAATAGCTTCCTCAACAGTATA TCCAAACTGTTGGGACAAACGAGCAATCACTGAAG AAGTGACAAGTTCCCACAATGTCAGTGTCCAGCTG AGAAGGGGCAAAAAGTGGTACCAGCCCTGTCCAC ACCACCTTCTAATTCACAGGAATCCGTGATAGAAG AGGCAGGTTGTAGATCCGAAAGATGAGACAGATTT TATCAACTCCAGAAAGAGCTGGGCCCAACTGAATC TAACTGAATTATTCTAGCGACCTTGGCATTGCCAT GACCTGCCATGACCTTCCTCCTTAGCACTTCGATG AACCCTGGGATATGGAAAATGCCTGTGTTTCTCAG GGTTTGGGAAGAACCATCCATGTTGGGATTCTTGT GTAGATCCTCCTCCTGGTCACAGATGCAATACACT GGATTTTCAGGCAAAGGAGCAAATTCACAGACAAC TCTGGCCCTACAGTCCTCAGACCTAGACACCACCA TCTCCTTGGAATTATCAAATCTAACACCCGGCACA CAACAAAGAAGGACTGGGACTTTGAGGCCTTTGTG TAGCCCTAGAGGGGGCAGAGGCCACTGAGCAGG GATTGGGTGATCAGCAAGGACCTCCTGGAGAGGG ACCTGAGGAGCAGGTTCCAATTGGGCCAAAGAAA GAAGAAGAACAATAGAGGTGAAGGATGCTGGAAA GAGCCATGGTACAGCAGTCTTGTCCTTCAGACATG ACTCTTACAGCCCAGGACTCTTACAGTAGCTAGCT GGAGCAGAAGTCCAAGGGATTACCATGCCCTAGG GCCACAGGCTACTGGAGGGTGGAGTGAGTCTACT ACACAGGTCCAATGCCTGTTTCTCCATCGTTTCTC AGCCAATGAGAAATCAGAGTCTCCAAACAGGAAGA AAAAGGAAGGCGGAAATGAAAGGTGAGCACCTGC CTTCCCGTGACTGGCAGAAAGATCTCCACGGACT CAAGGCTTTGACTTCAAACTTCCCGTAGATCCCTG ATGTCTTTAAGGACTCTGTCTGTTGGTTTTGTTTGT TTGTTTTGTACTTTGTCAATAAAACATTTTTCAAATA TTCTA
IGHG1 Exon 5 of 6	C-Gene	131	GGCTGCAACTGGACGAGACCTGTGCTGAGGCCCA GGACGGGGAGCTGGACGGGCTCTGGACGACCAT CACCATCTTCATCAGCCTCTTCCTGCTCAGCGTGT GCTACAGCGCTGCTGTCACACTCTTCAAG

IGHG1 Exon 4 of 6	C-Gene	315	GCAGACCGAAGGCTCCACAGGTGTACACCATTCC ACCTCCCAAGGAGCAGATGGCCAAGGATAAAGTC AGTCTGACCTGCATGATAACAGACTTCTTCCCTGA AGACATTACTGTGGAGTGGCAGTGGAATGGGCAG CCAGCGGAGAACTACAAGAACACTCAGCCCATCAT GGACACAGATGGCTCTTACTTCGTCTACAGCAAGC TCAATGTGCAGAAGAGCAACTGGGAGGCAGGAAA TACTTTCACCTGCTCTGTGTTACATGAGGGCCTGC ACAACCACCATACTGAGAAGAGCCTCTCCCACTCT CCTG
IGHG1 Exon 3 of 6	C-Gene	321	TCCCAGAAGTATCATCTGTCTTCATCTTCCCCCCAA AGCCCAAGGATGTGCTCACCATTACTCTGACTCCT AAGGTCACGTGTGTTGTGGTAGACATCAGCAAGG ATGATCCCGAGGTCCAGTTCAGCTGGTTTGTAGAT GATGTGGAGGTGCACACAGCTCAGACGCAACCCC GGGAGGAGCAGTTCAACAGCACTTTCCGCTCAGT CAGTGAACTTCCCATCATGCACCAGGACTGGCTCA ATGGCAAGGAGTTCAAATGCAGGGTCAACAGTGC AGCTTTCCCTGCCCCCATCGAGAAAACCATCTCCA AAACCAAAG
IGHG1 Exon 2 of 6	C-Gene	39	TGCCCAGGGATTGTGGTTGTAAGCCTTGCATATGT ACAG
IGHG1 Exon 1 of 6	C-Gene	291	CCAAAACGACACCCCCATCTGTCTATCCACTGGCC CCTGGATCTGCTGCCCAAACTAACTCCATGGTGAC CCTGGGATGCCTGGTCAAGGGCTATTTCCCTGAG CCAGTGACAGTGACCTGGAACTCTGGATCCCTGTC CAGCGGTGTGCACACCTTCCCAGCTGTCCTGCAG TCTGACCTCTACACTCTGAGCAGCTCAGTGACTGT CCCCTCCAGCACCTGGCCCAGCGAGACCGTCACC TGCAACGTTGCCCACCCGGCCAGCAGCACCAAGG TGGACAAGAAAATTG
IGHG3 Exon 6 of 6 (M2)	C-Gene	84	GTGAAGTGGATCTTCTCCTCAGTGGTGCAGGTGAA GCAGACGGCCATCCCTGACTACAGGAACATGATT GGACAAGGTGCCTAG
IGHG3 Exon 5 of 6 (M1)	C-Gene	131	AGCTGGAACTGAATGAGACCTGTGCTGAGGCCCA GGATGGGGAGCTGGACGGGCTCTGGACGACCAT CACCATCTTCATCAGCCTCTTCCTGCTCAGCGTGT GCTACAGCGCCTCTGTCACCCTCTTCAAG
IGHG3 CHS	C-Gene	6	GGTAAA

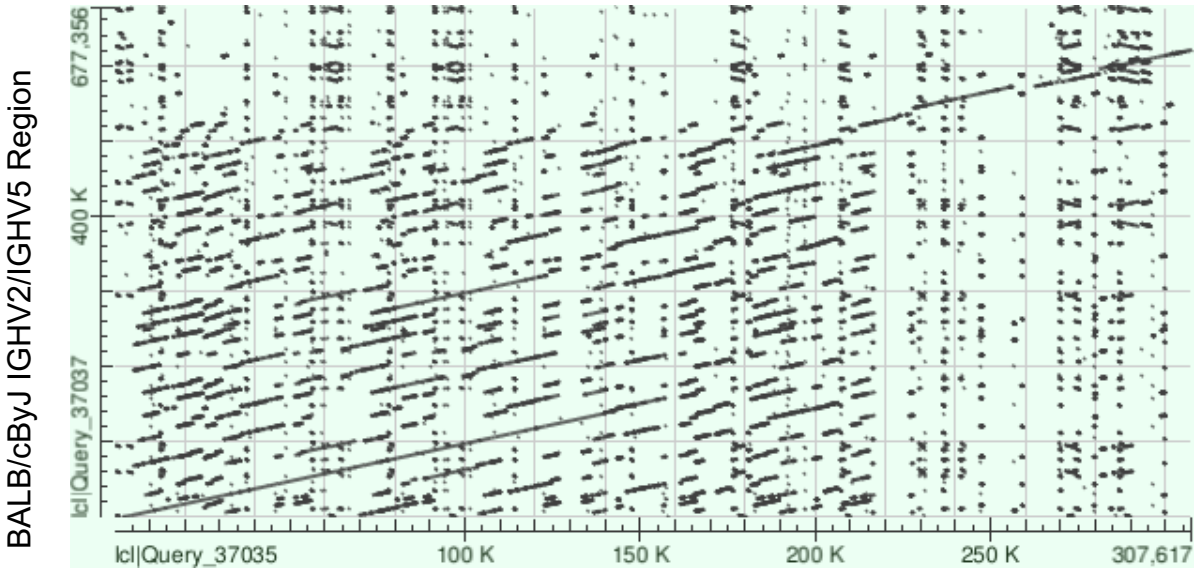
IGHG3 Exon 4 of 6 (CH3)	C-Gene	314	GAAGAGCCCAGACACCTCAAGTATACACCATACCC CCACCTCGTGAACAAATGTCCAAGAAGAAGGTTAG TCTGACCTGCCTGGTCACCAACTTCTTCTCTGAAG CCATCAGTGTGGAGTGGGAAAGGAACGGAGAACT GGAGCAGGATTACAAGAACACTCCACCCATCCTG GACTCAGATGGGACCTACTTCCTCTACAGCAAGCT CACTGTGGATACAGACAGTTGGTTGCAAGGAGAAA TTTTTACCTGCTCCGTGGTGCATGAGGCTCTCCAT AACCACCACACACAGAAGAACCTGTCTCGCTCCCC T
IGHG3 Exon 3 of 6 (CH2)	C-Gene	330	CTGGTAACATCTTGGGTGGACCATCCGTCTTCATC TTCCCCCCAAAGCCCAAGGATGCACTCATGATCTC CCTAACCCCCAAGGTTACGTGTGTGGTGGTGGAT GTGAGCGAGGATGACCCAGATGTCCATGTCAGCT GGTTTGTGGACAACAAAGAAGTACACACAGCCTG GACACAGCCCCGTGAAGCTCAGTACAACAGTACC TTCCGAGTGGTCAGTGCCCTCCCCATCCAGCACC AGGACTGGATGAGGGGCAAGGAGTTCAAATGCAA GGTCAACAACAAAGCCCTCCCAGCCCCCATCGAG AGAACCATCTCAAAACCCAAAG
IGHG3 Exon 2 of 6 (H)	C-Gene	48	AGCCTAGAATACCCAAGCCCAGTACCCCCCCAGG TTCTTCATGCCCAC
IGHG3 Exon 1 of 6 (CH1)	C-Gene	291	CTACAACAACAGCCCCATCTGTCTATCCCTTGGTC CCTGGCTGCAGTGACACATCTGGATCCTCGGTGA CACTGGGATGCCTTGTCAAAGGCTACTTCCCTGAG CCGGTAACTGTAAAATGGAACTATGGAGCCCTGTC CAGCGGTGTGCGCACAGTCTCATCTGTCCTGCAG TCTGGGTTCTATTCCCTCAGCAGCTTGGTGACTGT ACCCTCCAGCACCTGGCCCAGCCAGACTGTCATC TGCAACGTAGCCCACCCAGCCAGCAAGACTGAGT TGATCAAGAGAATCG

129 C psi gamma 0 gene	C-Gene	1722	TGAGGGGCAAGGAATTCAAATGCAAGGTCAAAAGT GGACCCTTCCCAGTTCTCATTGAGAAAACAATCTG TAAAGTTGAAGGTGGGGTCCGGGAGTGGGAGGAG GTAGATGGTTTTGAGACTGTGATGGGCAGAGTGTT GAGCTTCGCAAATTCTTTCCTTTGGTAGCAACTGC TATGGCTTCTCTCTCCAACAGGGCATGCCCAGGTG CCTTGGGAATACACCACAGGCCTCCATCCAGGTA GCAGATTAACAAGAACAAAGTCAGTCTCACCTGCA TGGTCATGCTCTTCTATCCTGGAGACATTGATGTA GAGTGCTATAGTCATGGAGTTCTTTTCTCCATACAA ATAAGCCACCCAATACTTCCCTGGGACCCAATGAA AGTGTCTTGGTTCTTTCTGAGGTTACACCAATGTG GCTAAGTCAATTGCTGTGGAGTCTTTAGTGTCCAA GTCCATACCATCAGACTCTGAGCATTTATATATATC TGGCCAGGCCTGAGGTCATCCCTCAACAGGGGAG GAGGTTTGGTTCAAGGGGCCTTCTCTCATCTTTGC TTACAATCAGCTGAAGAGTTATGCTGAACTGGGAC ATCTCCAAATAATCTCCTAAGGAAAGGAACACAGA GAACATATAAATCCTGACTCTTTCCATCTGTCCCTT CTTCCTGACCAAGGGTGTCTCTGTGACACTGCAGA GCCATCTCTAGGCTATGTGTACAATAAGTCATTAA GGCCAACACACTCAGTGCAGCCTGATTTGCCTGC CCTTGACTCCACAAGGAATTTGGATGCCTATGTGT GTGGATATTGCCTATTTATGGGACACTAATTGTGC CCCAAAGGGACTGAGGCCAGTGTTCTGTCTACCTC AGAGGAGGGCCTACAAGATACACACCTGATCCATT AGACCAGTGTACTTAGTATCACTCAGCCAATATGC AATCTGTACACAGATGTGTAATACATGTACTTGCCT AATATACTAGGGTCCTTACAGATGCATCTTGCATAA CACCCACTCCCTGTTGGGTATCAGAATATCCCTAA CCACCACTTCAATGTTCATATTTGGGTCATCAAAGC TAAACACAGACCTCACCTTCTGAGGGCACCTGCAC CCATTTCCCTATACACACAGTTTTTGAGGATCAATA TGTCTCTGAAGGCCTGTCCTTCTAATCTACACAGC TGATATTTCCTATAGTTTGTGGCCTGGCCTTCTTGC TGGTGTGAACCTTCAGTGACATCAAGATATTTCTA GGCACAACTGTGGCCCTCTGGGCTAGGTGACCGA AGGACATGTAGAGGGTCAGGCAGACTTCTGAGCC ACTTCCTATAGGAGAACAGGTCTTTACTTAGATCA GGTAGACCCAGAACTGAGAGGTAGAGCAGGCATG GAGGACTCCTGGATAATCTAGGTCATACCATTAGG TTCAATATGTCTATTGAACCCAGGGACTAAGCCAG AACAGTTAGAGAAGAGGGCTATTGGCCTAGACAG GGCAAGAAGCAGAAAGCTGGCCCCAGGATGGGTC TGCCTGACCAGCTCAGCTGGTTCTTTCTCTTGCAA CCCCTGATGTCCCACCTTGCCACAAAAGAAATTAC

			CCCCAGCATCTCCTTGTTGGAAAAGACACTGACTG CTCCTCTCTGTGTAGAGCTGGAATTGGATGGTACC TGGGGGGAGCTGGACGAACTCTGGATGACCATCA CAATCTTCA
IGHD Exon 5 (M2)	C-Gene	6	GTGAAG
IGHD Exon 4 (M1)	C-Gene	158	GCATAGTCAACACCATCCAACACTCGTGTATCATG GATGAGCAAAGTGACAGCTACATGGACTTAGAGG AGGAGAACGGCCTGTGGCCCACAATGTGCACCTT CGTGGCCCTCTTCCTGCTCACACTGCTCTACAGTG GCTTCGTCACCTTCATCAAG
IGHD Exon 3 (CH3)	C-Gene	321	GGGCCATGGCACCCAGCAACCTCACTGTGAACAT CCTGACCACATCCACCCATCCTGAGATGTCATCTT GGCTCCTGTGTGAAGTATCTGGCTTCTTCCCGGAA AATATCCACCTCATGTGGCTGAGTGTCCACAGTAA AATGAAGTCTACAAACTTTGTCACTGCAAACCCCA CCGCCCAGCCTGGGGGCACATTCCAGACCTGGAG TGTCCTGAGACTACCAGTCGCTCTGAGCTCATCAC TTGACACTTACACATGTGTGGTGGAACATGAGGCC TCAAAGACAAAGCTTAATGCCAGCAAGAGCCTAGC AATTAGTG

IGHD Exon 2 (H)	C-Gene	105	AGTCATGGGATTCCCAGTCCTCTAAGAGAGTCACT CCAACTCTCCAAGCAAAGAATCACTCCACAGAAGC CACCAAAGCTATTACCACCAAAAAGGACATAGAAG
IGHD Exon 1 (CH1)	C-Gene	282	GTGATAAAAAGGAACCTGACATGTTCCTCCTCTCA GAGTGCAAAGCCCCAGAGGAAAATGAAAAGATAAA CCTGGGCTGTTTAGTAATTGGAAGTCAGCCACTGA AAATCAGCTGGGAGCCAAAGAAGTCAAGTATAGTT GAACATGTCTTCCCCTCTGAAATGAGAAATGGCAA TTATACAATGGTCCTCCAGGTCACTGTGCTGGCCT CAGAACTGAACCTCAACCACACTTGCACCATAAAT AAACCCAAAAGGAAAGAAAAACCTTTCAAGTTTCC TG
IGHM Exon 6 of 6 (M2)	C-Gene	6	GTGAAA
IGHM Exon 5 of 6 (M1)	C-Gene	116	AGGGGGAGGTGAATGCTGAGGAGGAAGGCTTTGA GAACCTGTGGACCACTGCCTCCACCTTCATCGTCC TCTTCCTCCTGAGCCTCTTCTACAGCACCACCGTC ACCCTGTTCAAG
IGHM CHS	C-Gene	60	GGTAAACCCACACTGTACAATGTCTCCCTGATCAT GTCTGACACAGGCGGCACCTGCTAT
IGHM Exon 4 of 6 (CH4)	C-Gene	332	AGGTGCACAAACATCCACCTGCTGTGTACCTGCTG CCACCAGCTCGTGAGCAACTGAACCTGAGGGAGT CAGCCACAGTCACCTGCCTGGTGAAGGGCTTCTC TCCTGCAGACATCAGTGTGCAGTGGCTTCAGAGA GGGCAACTCTTGCCCCAAGAGAAGTATGTGACCA GTGCCCCGATGCCAGAGCCTGGGGCCCCAGGCTT CTACTTTACCCACAGCATCCTGACTGTGACAGAGG AGGAATGGAACTCCGGAGAGACCTATACCTGTGTT GTAGGCCACGAGGCCCTGCCACACCTGGTGACCG AGAGGACCGTGGACAAGTCCACT
IGHM Exon 3 of 6 (CH3)	C-Gene	318	GTCCCTCCACAGACATCCTAACCTTCACCATCCCC CCCTCCTTTGCCGACATCTTCCTCAGCAAGTCCGC TAACCTGACCTGTCTGGTCTCAAACCTGGCAACCT ATGAAACCCTGAATATCTCCTGGGCTTCTCAAAGT GGTGAACCACTGGAAACCAAAATTAAAATCATGGA AAGCCATCCCAATGGCACCTTCAGTGCTAAGGGT GTGGCTAGTGTTTGTGTGGAAGACTGGAATAACAG GAAGGAATTTGTGTGTACTGTGACTCACAGGGATC TGCCTTCACCACAGAAGAAATTCATCTCAAAACCC AATG

IGHM Exon 2 of 6 (CH2)	C-Gene	339	CTGTCGCAGAGATGAACCCCAATGTAAATGTGTTC GTCCCACCACGGGATGGCTTCTCTGGCCCTGCAC CACGCAAGTCTAAACTCATCTGCGAGGCCACGAA CTTCACTCCAAAACCGATCACAGTATCCTGGCTAA AGGATGGGAAGCTCGTGGAATCTGGCTTCACCAC AGATCCGGTGACCATCGAGAACAAAGGATCCACA CCCCAAACCTACAAGGTCATAAGCACACTTACCAT CTCTGAAATCGACTGGCTGAACCTGAATGTGTACA CCTGCCGTGTGGATCACAGGGGTCTCACCTTCTTG AAGAACGTGTCCTCCACATGTGCTGCCA
IGHM Exon 1 of 6 (CH1)	C-Gene	315	AGAGTCAGTCCTTCCCAAATGTCTTCCCCCTCGTC TCCTGCGAGAGCCCCCTGTCTGATAAGAATCTGGT GGCCATGGGCTGCCTGGCCCGGGACTTCCTGCCC AGCACCATTTCCTTCACCTGGAACTACCAGAACAA CACTGAAGTCATCCAGGGTATCAGAACCTTCCCAA CACTGAGGACAGGGGGCAAGTACCTAGCCACCTC GCAGGTGTTGCTGTCTCCCAAGAGCATCCTTGAAG GTTCAGATGAATACCTGGTATGCAAAATCCACTAC GGAGGCAAAAACAGAGATCTGCATGTGCCCATTC CAG



Supplemental Figure 2. Dot plot comparison showing expansion of BALB/cByJ-IGH assembly IGHV2 and IGHV5 region relative to C57BL/6 IGHV2 and IGHV5 region.

Supplemental Table 6. Annotated NOD/ShiLtJ assembly IGKV sequences, RSS, and leader sequences.

Contig	Name	Type	Sequence	Start	End
NOD.IGK.SuperContig.B	144118	RSS	CACAGTGATTCAAGCCATGACATAAACC	35370	35397
NOD.IGK.SuperContig.B	IGKV12.1	V-Gene	GACATCCAGATGACTCAGTCTCCAGCCTCCCTATCTGTATCTGTGGGAGAAACTGTCACCATCACATGTCGAGCAAGTGAGAATATTTACAGTAATTTAGCATGGTATCAGCAGAAACAGGGAAAATCTCCTCAGCTCCTGGTCTATGCTGCAACAAACTTAGCAGATGGTGTTCCATCAAGGTTCAGTGGCAGTGGATCAGGCACACAGTATTCCCTCAAGATCAACAGCCTGCAGTCTGAAGATTTTGGGAGTTATTACTGTCAACATTTTTGGGGTACTCCTCC	35398	35684
NOD.IGK.SuperContig.B	L-Part2	Leader	ATGCCAGATGT	35685	35695
NOD.IGK.SuperContig.B	L-Part1	Leader	ATGAGTGTGCCCACTCAGGTCCTGGGGTTGCTGCTGCTGTGGCTTACAG	35819	35867
NOD.IGK.SuperContig.B	100090	RSS	CACAGTGATGCAGACCATAGCAAAAATC	46415	46442
NOD.IGK.SuperContig.B	IGKV5.1	V-Gene	GATATTGTGCTAACTCAGTCTCCAGCCACCCTGTCTGTGACTCCAGGAGATAGAGTCAGTCTTTCCTGCAGGGCCAGCCAAAGTATTAGCAACTACCTACACTGGTATCAACAAAAATCACATGAGTCTCCAAGGCTTCTCATCAAGTAT	46443	46729

			GCTTCCCAGTCCATCT CTGGGATCCCCTCCAG GTTCAGTGGCAGTGGA TCAGGGACAGATTTCA CTCTCAGTATCAACAGT GTGGAGACTGAAGATT TTGGAATGTATTTCTGT CAACAGAGTAACAGCT GGCCTCA		
NOD.IGK.SuperContig.B	L-Part2	Leader	CCTCCAGAGGT	46730	46740
NOD.IGK.SuperContig.B	L-Part1	Leader	ATGGTTTTCACACCTCA GATTCTTGGACTTATGC TTTTCTGGATTTCAG	46951	46999
NOD.IGK.SuperContig.B	144118	RSS	CACAGTGATTCAAGCC ATGACATAAACC	91319	91346
NOD.IGK.SuperContig.B	IGKV12.2	V-Gene	GACATCCAGATGACTC AGTCTCCAGCCTCCCT ATCTGCATCTGTGGGA GAAACTGTCACCATCA CATGTCGAGCAAGTGA GAATATTTACAGTTATT TAGCATGGTATCAGCA GAAACAGGGAAAATCT CCTCAGCTCCTGGTCT ATAATGCAAAAACCTTA GCAGAAGGTGTGCCAT CAAGGTTCAGTGGCAG TGGATCAGGCACACAG TTTTCTCTGAAGATCAA CAGCCTGCAGCCTGAA GATTTTGGGAGTTATTA CTGTCAACATCATTATG GTACTCCTCC	91347	91633
NOD.IGK.SuperContig.B	L-Part2	Leader	GTGCCAGATGT	91634	91644
NOD.IGK.SuperContig.B	L-Part1	Leader	ATGAGTGTGCCCACTC AGGTCCTGGGGTTGCT GCTGCTGTGGCTTACA G	91768	91816
NOD.IGK.SupercontigA	L-Part1	Leader	ATGAGGTGCCTAGCTG AGTTCCTGGGGCTGCT	95997	96045

			TGTGCTCTGGATCCCT G		
NOD.IGK. Supercont igA	L-Part2	Leader	GAGCCATTGGG	96265	96275
NOD.IGK. Supercont igA	IGKV2.1	V-Gene	GATATTGTGATGACTCA GGCTGCACCCTCTGTA CCTGTCACTCCTGGAG AGTCAGTATCCATCTCC TGCAGGTCTAGTACGA GTCTCCTGCACAGTAG TGGCAAGCATAGGTTG TATTGGTTCCTACAGAG GCCAGGCCAGTCTCCT CAGCTCCTGATATATTA TATGTCCAACCTTGCCT CAGGAGTCCCAGACAG GTTCAGTGGCAGTGGG TCAGGAACTGATTTCAC ACTGAGAATCAGTAGA GTGGAGGCTGAGGATT TTGGTGTTTATTATTGT ATGCAAAGTCTAGAATA TCCTCT	96276	96577
NOD.IGK. Supercont igA	96605	RSS	CACAGTGATACAGCCC TGAACAAAAACC	96578	96605
NOD.IGK. SuperCon tig.B	100090	RSS	CACAGTGATGCAGACC ATAGCAAAAATC	100090	100117
NOD.IGK. SuperCon tig.B	IGKV5.2	V-Gene	GATATTGTGCTAACTCA GTCTCCAGCCACCCTG TCTGTGACTCCAGGAG ATAGCGTCAGTCTTTCC TGCAGGGCCAGCCAAA GTATTAGCAACAACCTA CACTGGTATCAACAAAA ATCACATGAGTCTCCAA GGCTTCTCATCAAGTAT GCTTCCCAGTCCATCT CTGGGATCCCCTCCAA GTTCAGTGGCAGTGGA TCAGGGACAGATTTCA CTCTCAGTATCAACAGT GTGGAGACTGAAGATT TTGGAATGTATTTCTGT	100118	100404

			CAACAGAGTAACAGCTGGCCTCA		
NOD.IGK.SuperContig.B	L-Part2	Leader	CCTCCAGAGGT	100405	100415
NOD.IGK.SuperContig.B	L-Part1	Leader	ATGGTTTTCACACCTCAGATTCTTGGACTTATGCTTTTCTGGATTTCAG	100626	100674
NOD.IGK.SupercontigA	L-Part1	Leader	ATGAGGTGCCTAGCTGAGTTCCTGGGGCTGCTTGTGCTCTGGATCCCTG	109647	109695
NOD.IGK.SupercontigA	L-Part2	Leader	GAGCCATTGGG	109915	109925
NOD.IGK.SupercontigA	IGKV2.2	V-Gene	GATATTGTGATGACTCAGGCTGCACCCTCTGTACCTGTCACTCCTGGAGAGTCAGTATCCATCTCCTGCAGGTCTAGTAAGAGTCTCCTGCATAGTAATGGCAACACTTACTTGTATTGGTTCCTACAGAGGCCTGGCCAGTCTCCTCAGCTCCTGATATATCGGATGTCCAACCTTGCCTCAGGAGTCCCAGACAGGTTCAGTGGCAGTGGGTCAGGAACTGCTTTCACACTGAGAATCAGTAGAGTCGAGGCTGAGGATGTGGGTGTTTATTACTGTATGCAACATCTAGAGTATCCTTT	109926	110227
NOD.IGK.SupercontigA	110255	RSS	CACAGTGATACAGCCCTGAACAAAATCC	110228	110255
NOD.IGK.SuperContig.B	117248	RSS	CACAGTGATTCAGTCCATGATATAAGTA	117248	117275

NOD.IGK.SuperContig.B	IGKV12.3	V-Gene	GACATTCAGATGACTCAGTCTCCAGCCTCCCTATCTGCATCGCTGGGAGAAAGTGTCACCATCACATGTCAAGCAAGTGAGAATATTTACAAGTATTTAACATGGTATCAGCAAAAACCAGGGACATCTCTCCATCTCCTCATTTATGGTGCAACCAGCTTGGCCGGTTCAGTGGCATTGGATCTGGCACACGGTATTCTGTAAAGATTAACAGTTTGCAGCCTGAAGATGTTGCAACTTATTACTGTCAAATTGCTTTAAATACCCCTTC	117276	117546
NOD.IGK.SuperContig.B	L-Part2	Leader	ATGCCAGATGT	117547	117557
NOD.IGK.SuperContig.B	L-Part1	Leader	AGGAGTGTGCCCACTCAGCTCCTGGGGTTGCTGCTGCTGTGGCTTACAG	117681	117729
NOD.IGK.SupercontigA	L-Part2	Leader	AATCCAGGGCA	119254	119264
NOD.IGK.SupercontigA	IGKV17.1.P	V-Gene	GAAACAACTGTGACCCAGTCTCCAGCATCCATGTCCATGGCTTCATGAGAAAAAGTCCCTATCAGTTGTATCACCAGCACCAATATTGATGATGATATGAACTGGTACCCGCAGAAACCAGGAGAAGCTCCTAAGCTCCTTATTTCAGAAGGCAATACTCTTCGTCCTGGAGTCCCATCCTGGTTCTCCAGCAGTGGCTATGGCACAGATTTTGTATTTACAGTTGAAAACATGCTCTCAGGAGATGCTGCAGATTATTACTGTCAGTGAAGTGATAAATTGCCTCT	119265	119551

NOD.IGK.SupercontigA	119579	RSS	CACGGTGCTATATCCTCTTACAGACACC	119552	119579
NOD.IGK.SupercontigA	L-Part1	Leader	ATGAGGGCCCCTGCCTTGTTTCTTGGCATCTTGTTGCTCTGGTTTCCAG	136396	136444
NOD.IGK.SupercontigA	L-Part2	Leader	GTGCCAGATGT	136559	136569
NOD.IGK.SupercontigA	IGKV9.1	V-Gene	GACATCCAGATGACCCAGTCTCCATCCTCCATGTCTGCCTCTCTGGGAGAAAGAGTCAGTCTCACTTGCCGGGCCAGTCAGGGCATTAACGGTAATTTACACTGGTTTCAGCAGAAGTCAGGTGGAACTCTTAAACGCCTGATCTACTCCACGTCCAATTTAGATTCTGGTGTTCCATCAAGGTTCAGTGGCAGCGGGTCTGGGTCAGATTATTCTCTCACCATCAGCAGCCTGGAGTCTGAAGATTTTGCAATCTATTACTATCTACAGTATGATGAACATCCTCC	136570	136856
NOD.IGK.SupercontigA	136884	RSS	CACAGTGATACAAGTTATAACATAAGCC	136857	136884
NOD.IGK.SuperContig.B	144118	RSS	CACAGTGATTCAAGCCATGACATAAACC	144118	144145
NOD.IGK.SuperContig.B	IGKV12.4	V-Gene	GACATCCAGATGACTCAGTCTCCAGCCTCCCTATCTGCATCTGTGGGAGAAACTGTCACCATCACATGTCGAGCAAGTGAGAATATTTACAGTAATTTAGCATGGTATCAGCAGAAACAGGGAAAATCTCCTCAGCTCCTGGTCTATGCTGCAACAAATTTAGCAGATGGTGTGCCATCAAGGTTCAGTGGCAGTGGATCAGGCACACAG	144146	144432

			TTTTCTCTGAAGATCAACAGCCTGCAGCCTGAAGATTTTGGGAGTTATTACTGTCAACATTTTTATGGTACTCCTCC		
NOD.IGK.SuperContig.B	L-Part2	Leader	GTGCCAGATGT	144433	144443
NOD.IGK.SuperContig.B	L-Part1	Leader	ATGAGTGTGCCCACTCAGGTCCTGGGGTTGCTGCTGCTGTGGCTTACAG	144567	144615
NOD.IGK.SuperContig.B	100090	RSS	CACAGTGATGCAGACCATAGCAAAAATC	155001	155028
NOD.IGK.SuperContig.B	IGKV5.3	V-Gene	GATATTGTGCTAACTCAGTCTCCAGCCACCCTGTCTGTGACTCCAGGAGATAGAGTCAGTCTTTCCTGCAGGGCCAGCCAAAGTATTAGCAACTACCTACACTGGTATCAACAAAAATCACATGAGTCTCCAAGGCTTCTCATCAAGTATGCTTCCCAGTCCATCTCTGGGATCCCCTCCAGGTTCAGTGGCAGTGGATCAGGGACAGATTTCACTCTCAGTATCAACAGTGTGGAGACTGAAGATTTTGGAATGTATTTCTGTCAACAGAGTAACAGCTGGCCTCA	155029	155315
NOD.IGK.SuperContig.B	L-part2	Leader	CCTCCAGAGGT	155316	155326
NOD.IGK.SuperContig.B	L-Part1	Leader	ATGGTTTTCACACCTCAGATTCTTGGACTTATGCTTTTCTGGATTTCAG	155537	155585
NOD.IGK.SupercontigA	L-Part1	Leader	ATGATGAGTCCTGCCCAGTTCCTGTTTCTGCTAGTGCTCTCGATTCAGG	173213	173261

NOD.IGK.SupercontigA	L-Part2	Leader	AAACCAATGGT	173664	173674
NOD.IGK.SupercontigA	IGKV1.1	V-Gene	GATGTTGTGATGACCCAGACTCCACTGTCTTTGTCGGTTACCATTGGACAACCAGCCTCCATCTCTTGCAAGTCAAGTCAGAGCCTCTTACATAGTGATGGAAAGACATATTTGAATTGGTTATTACAGAGGCCAGGCCAGTCTCCAAAGCTCCTAATCTATCTGGTGTCTAAACTGGAATCTGGCATCCCTGACAGGTTCAGTGGCAGTGGATCAGGGACAGATTTCACACTGAAAATCAGCAGAGTGGAGGTTGAGGATTTGGGAGTTTATTACTGCTTGCAACATACACATTTTCCTCA	173675	173976
NOD.IGK.SupercontigA	174004	RSS	CACAGTGATTCAGACCTGAACAAAAACT	173977	174004
NOD.IGK.SuperContig.B	144118	RSS	CACAGTGATTCAAGCCATGACATAAACC	186344	186371
NOD.IGK.SuperContig.B	IGKV12.5	V-Gene	GATATCCAGATGACTCAGTCTCCAGCCTCCCTATCTGCATCTGTGGGAGAAACTGTCACCATCACATGTCGAGCAAGTGGGAATATTCACAATTATTTAGCATGGTATCAGCAGAAACAGGGAAAATCTCCTCAGCTCCTGGTCTATAATGCAAAAACCTTAGCGGAAGGTGTGCCATCAAGGTTCAGTGGCAGTGGATCAGGAACACAATATTCTCTCAAGATCAACAGCCTGCAGCCTGAGGATTTTGGGAGTTATTACTGTCATCATTATTATAGTACTCCTCC	186372	186658

NOD.IGK.SuperContig.B	L-Part2	Leader	ATGCCAGATGT	186659	186669
NOD.IGK.SuperContig.B	L-Part1	Leader	ATGAGTGTGCCCACTCAGGTCCTGGCATTGCTGCTGCTGTGGCTTACAG	186791	186839
NOD.IGK.SuperContig.B	144118	RSS	CACAGTGATTCAAGCCATGACATAAACC	207597	207624
NOD.IGK.SuperContig.B	IGKV12.6	V-Gene	GACATCCAGATGACTCAGTCTCCAGCTTCCCTGTCTGCATCTGTGGGAGAAACTGTCACCATCACATGTCGAGCAAGTGAGAATATTGACAGTTATTTAGCATGGTATCAGCAGAAACAGGGAAAATCTCCTCAGCTCCTGGTCTATGCTGCAACACTCTTAGCAGATGGTGTGCCATCAAGGTTCAGTGGCAGTGGATCAGGCACACAGTATTCTCTCAAGATCAACAGCCTGCAGTCTGAAGATGTTGCGAGATATTACTGTCAACATTATTATAGTACTCCTCC	207625	207911
NOD.IGK.SuperContig.B	L-Part2	Leader	ATGCCAGATGT	207912	207922
NOD.IGK.SuperContig.B	L-Part1	Leader	ATGAGTGTGCCCACTCAGCTCCTGGGGTTGCTGCTGCTGTGGCTTACAG	208045	208093
NOD.IGK.SupercontigA	230551	RSS	CACAGTGCTATGTCCTCTTACAGAAACC	230524	230551
NOD.IGK.SupercontigA	IGKV17.2.P	V-Gene	GAAATAACTGTGACCCAGTCTCCAGCATCCCTATCCATGGCTACAGGAGAAAAAGTCACTATCAGACGCATAACACTGATATTGATGATGAAATGAACTAGTACCAGCAGAAGCCAGGGGAACCTCCTAAG	230552	230834

			CTCCTTATTTCAGAAGG CAATACTCTCCATCCTG GAGTCCCATCCCAGTT CTCCAGCAGTGGCTAT GGCACAGATTTTGTTTT TACAATTGAAAACATGC TCTCAGAAGATGTTGCA GATTACTACTGTTTGCA AAGTGGTAACATGCCT CT		
NOD.IGK. SupercontigA	L-Part2	Leader	ATTCTAGGGCA	230835	230845
NOD.IGK. SupercontigA	L-Part1	Leader	ATGACCATGCTCTCACT AGTTCTTCTCCTCAGCT TTCTTCTCCTCTGTGTC ACTG	230994	231048
NOD.IGK. SuperContig.B	231694	RSS	CACAATGATGCAGACC ATAGCAAAAACC	231694	231721
NOD.IGK. SuperContig.B	IGKV5.4	V-Gene	GACATTGTGATGACTCA GTCTCCAGCCACCCTG TCTGTGACTCCAGGAG ATAGGGTCTCTCTTTCC TGCAGGGCCAGTCAGA GTATTAGCGACTACTTA CACTGGTATCAACAAAA ATCACATGAGTCTCCAA GGCTTCTCATCAAGTAT GCTTCCCAATCCATCTC TGGGATCCCCTCCAGG TTCAGTGGCAGTGGAT CAGGGTCAGATTTCAC TCTCAGTATCAACAGTG TGGAACCTGAAGATGT TGGAGTGTATTACTGTC AAAATGGTCACAGCTTT CCTCC	231722	232008
NOD.IGK. SuperContig.B	L-Part2	Leader	CCTCCAGATGT	232009	232019
NOD.IGK. SuperContig.B	L-Part1	Leader	ATGGTGCCCACTTCTC AGCTCCTTGGATTTTTG CTTTTCTGGACTTCAG	232227	232275

NOD.IGK.Supercontig A	L-Part1	Leader	ATGATGAGTCCTGCCCAGTTCCTGTTTCTGCTAGTGCTCTGGATTCAGA	234669	234717
NOD.IGK.Supercontig A	L-Part2	Leader	AAACCAACGGT	235104	235114
NOD.IGK.Supercontig A	IGKV1.2	V-Gene	GATGTTGTGATGACCCAGACTCCACTCACTTTGTCTGTTACCATTGGACAGCCAGCTTCCATTTCTTGCAAGTCAAGTCAGAGCCTCTTATATAGTGATGGAAAAACCTATTTGAATTGGTTATTACAGAGTCCAGGCCAGTCTCCAAAGCTCCTAATCTATCTGGTGTCTAAACTGGAATCTGGAGTCCCTGACAGATTCAGTGGCAGTGGATCAGGGACAGATTTTACACTGAAAATCAGCAGAGTGGAGGCTGAGGATTTGGGAGTTTATTACTGCGTGCAAGGTACACATTTCCCTCA	235115	235416
NOD.IGK.Supercontig A	235444	RSS	CACAGTGATACAGACTCTATCAAAAACT	235417	235444
NOD.IGK.Supercontig A	L-Part1	Leader	ATGATGAGTCCTGCCCAGTTCCTGTTTCTGCTAGTGCTCTCGATTCAGG	257329	257377
NOD.IGK.Supercontig A	L-Part2	Leader	AAACCAACGGT	257778	257788
NOD.IGK.Supercontig A	IGKV1.3	V-Gene	GATGTTGTGATGACTCAGACCCCACTCACTTTGTCGGTTACCATTGGACAACCAGCCTCCATCTCTTGCAAATCAAGTCAGAGCCTCTTACATAGTAATGGAAAGACATATTTGAATTGGTTATTACAGAGGCCAGGCCAGTCTCCAAAGCTCCTAATCTATCTGGTGTCTAAACTGGAATCTGGAGTCCCTGACAGGT	257789	258090

			TCAGTGGCAGTGGATC AGGGACAGATTTCACA CTGAAAATCAGCAGAG TGGAGGCTGAGGATTT GGGAGTTTATTACTGCT TGCAAGCTACACATTTT CCTCA		
NOD.IGK. SupercontigA	174004	RSS	CACAGTGATTCAGACC TGAACAAAAACT	258091	258118
NOD.IGK. SuperContig.B	144118	RSS	CACAGTGATTCAAGCC ATGACATAAACC	278139	278166
NOD.IGK. SuperContig.B	IGKV12.7	V-Gene	GACATCCAGATGACTC AGTCTCCAGCCTCCCT GGCTGCATCTGTGGGA GAAACCATCACCATCA CATGTCAAGCAAGTGA GAACATTTACTTCAGTT TAGCATGGTATCAGCA GAAGCAAGGGAAATCT CCTCAGCTCCTGATCTA TAATGCAAACAGCTTG GAAGATGGTGTCCCAT CGAGGTTCAGTGGCAG TGGATCTGGGACACAG TATTCTATGAAGATCAA CAGCATGCAGCCTGAA GATACTGCAACTTATTT CTGTAAACAGGCTTATG ACTTTCCTCC	278167	278453
NOD.IGK. SuperContig.B	L-Part2	Leader	ACGCAGGATGT	278454	278464
NOD.IGK. SuperContig.B	L-Part1	Leader	ATGAGTGTGCCCACTC AGCTCCTGGGGTTGCT GCTGCTGTGGCTTACA G	278581	278629
NOD.IGK. SuperContig.B	294842	RSS	CACAGTGATGCAGACC ATAGCAAAAACC	294842	294869

NOD.IGK.SuperContig.B	IGKV5.5	V-Gene	GACATCCTGATGACCCAGTCTCCAGCCACCCTGTCTGTGACTCCAGGAGAAACAGTCAGTCTTTCCTGTAGGGCCAGCCAGAATACTTACAAGAACCTACACTGGTATCAACAGAAATCACATGGGACTCCAAAGCTTCTCATCAAGTATGCATCTGATCCCATCTCTGGGATCCCCTCCAGGTTCACTGGCAGTGGATCAGGGACAGATTACACTCTCAGTATCAACAGTGTGAAGCCCGAAGATGAAGGAATATATTACTGTCTTCAAGGTTACAGCATGCCTCC	294870	295156
NOD.IGK.SuperContig.B	L-Part2	Leader	CCTCCACAGGT	295157	295167
NOD.IGK.SuperContig.B	L-Part1	Leader	ATGGTATTCACACCTCATATCCTTGGACTTCTGCTTTTCTGGATTTCAG	295371	295419
NOD.IGK.SupercontigA	230551	RSS	CACAGTGCTATGTCCTCTTACAGAAACC	311257	311284
NOD.IGK.SupercontigA	IGKV17.1	V-Gene	GAAATAACTGTGACCCAGTCTCCAGCATCCCTATCCATGGCTACAGGAGAAAAAGTCACTATCAGACGCATAACACTGATATTGATGATGAAATGCACTAGTACCAGCAGAAGCCAGGGGAACCTCCTAAGCTCCTTATTTCAGAAGGCAATACTCTCCATCCTGGAGTCCCATCCCAGTTCTCCAGCAGTGGCTATGGCACAGATTTTGTTTTTACAATTGAAAACATGCTCTCAGAAGATGTTGCAGATTATTACTGTTTGCAAAGTGGTAACATGCCTCT	311285	311567

NOD.IGK.SupercontigA	L-Part2	Leader	ATTCTAGGGCA	311568	311578
NOD.IGK.SupercontigA	L-Part1	Leader	ATGACCATGCTCTCACTAGTTCTTCTCCTCAGCTTTCTTCTCCTCTGTGTCACTG	311727	311781
NOD.IGK.SupercontigA	L-Part1	Leader	ATGATGAGTCCTGCCCAGTTCCTGTTTCTGTTAGTGCTCTGGATTCAGG	315618	315666
NOD.IGK.SupercontigA	L-Part2	Leader	AAACCAACGGT	316053	316063
NOD.IGK.SupercontigA	IGKV1.4	V-Gene	GATGTTGTGATGACCCAGACTCCACTCACTTTGTCAGTTACCATTGGACAACCAGCCTCTATCTCTTGCAAGTCAAGTCAGAGCCTCTTATATAGTAATGGAAAAACCTATTTGAACTGGTTATTACAGAGGCCAGGCCAGTCTCCAAAGCGCCTAATCTATCTGGTGTCTAAACTGGACTCTGGAGTCCCTGACAGGTTCACTGGCAGTGGATCAGGAACAGATTTTACACTGAAAATCAGCAGAGTGGAGGCTGAGGATTTGGGAGTTTATTACTGCGTGCAAGGTACACATTTTCCTCA	316064	316365
NOD.IGK.SupercontigA	235444	RSS	CACAGTGATACAGACTCTATCAAAAACT	316366	316393
NOD.IGK.SuperContig.B	326735	RSS	CACAGTGATTCAACATGTCACAAAAACC	326735	326762
NOD.IGK.SuperContig.B	IGKV18.1	V-Gene	ACTGGAGAAACAACACAGTCTCCAGCTTCTCTGAGTTTTTCTCTTGGTGAAACAGCAACACTGTCATGCAGGTCCAGTGAGAGTGTTGGCAGCTACTTAGCCTGGTACCAGCAGAAAGCAGAGCAAGTTCC	326763	327049

			CCGGCTCCTTATCCATA GTGCCTCCACTAGGGC CGGTGGTGTCCCAGTC CGATTCAGTGGCACTG GGTCTGGGACAGACTT CACTCTCACCATCAGC AGTCTAGAACCTGAAG ATGCTGCAGTTTACTAC TGTCAACCTTTCAAAAG TTGGTCATA		
NOD.IGK.SuperContig.B	L-Part2	Leader	TCTCAGCTGTC	327050	327060
NOD.IGK.SuperContig.B	L-Part1	Leader	ATGGAAACTCCAGCTT CATTTCTTTGCCTCCTT TTACTGTGGACCACGG	327223	327271
NOD.IGK.SuperContig.B	346032	RSS	CACAGTGATATAGACA CTTAAAAAAATT	346032	346059
NOD.IGK.SuperContig.B	IGKV1.12	V-Gene	GACATTGTGATGACCC AGACTCCACTCACTTTA TCAGCTACCATTGGAC AATCAGCCTCCATCTCT TGCAGGTCAAGTCAGA GTCTCTTACATAGTAAT GGAAACACATACTTGAA TTGGTTTCTACAGAGG CCAGGCCAATCTCCAC AGCTTCTGATTTATGGG GTGTTTGAACGGGAAT CTGGGGTTCCTGACAG GTTCAGTGGCAGTGGG TCAGGAACAGATTTCAC ACTCAAGATCAGCAGA GTGGAGGCTGAGGATT TGGGAGTTTATTACTGC ATGCAAGCTACCTATGA ACCTCC	346060	346361
NOD.IGK.SuperContig.B	L-Part2	Leader	TCAGTGGT	346362	346369
NOD.IGK.SuperContig.B	L-Part1	Leader	ATGATGAGTCCTGTCC ACTCCATATTTATATTG TTGCTTTGGATTGTGG	346736	346784

NOD.IGK.Supercontig A	L-Part1	Leader	ATGATGAGTCCTGTCC AGTTCCTGTTTCTGTTA ATGCTCTGGATTCAGG	353241	353289
NOD.IGK.Supercontig A	L-Part2	Leader	AAACCAATGGT	353642	353652
NOD.IGK.Supercontig A	IGKV1.5	V-Gene	GATGTTGTGATGACCC AGACTCCACTGTCTTTG TCGGTTACCATTGGAC AACCAGCCTCTATCTCT TGCAAGTCAAGTCAGA GCCTCTTATATAGTGAT GGAAAGACATATTTGAA TTGGTTACAACAGAGG CCAGGCCAGTCTCCAA AGCGCCTAATGTATCA GGTGTCCAAACTGGAC CCTGGCATCCCTGACA GGTTCAGTGGCAGTGG ATCAGAGACAGATTTTA CACTTAAAATCAGCAGA GTGGAGGCTGAGGATT TGGGAGTTTATTACTGC TTGCAAGGTACATATTA TCCTCA	353653	353954
NOD.IGK.Supercontig A	353982	RSS	CACAGTGATACAGACC CTAACAAAAACT	353955	353982
NOD.IGK.Supercontig A	IGKV1.6	V-Gene	GATGTTGTGATGACCC AGCCTCCACTGTCTTTG TCGGTTACCATTGGAC AACCAGCCTCTATCTCT TGCAAGTCAAGTCAGA GCCTCTTACATAGTAAT GGAAAGACATATTTGAA CTGGTTATTACAGAGG CCAGGCCAGTCTCCAA AGCGCCTATTCTATCTG GTATCCAAACTGGACC CTGGCATCCCTGACAG TTTCAGTGGCAGTGGA TCAGAGACAGATTTTAC ACTGAAAATCAGCAGA GTGGAGGCTGAAGATT TGGGAGTTTATTACTGC	359586	359887

			GTGCAAGGTACACATTTTCCTCT		
NOD.IGK.SupercontigA	L-Part2	Leader	AAACCAACGGT	359888	359898
NOD.IGK.SupercontigA	L-Part1	Leader	ATGATGAGTCCTGTCCAGTTCCTGTTTCTGCTAATGCTCTGGATTCAGGGTAAGG	360268	360322
NOD.IGK.SuperContig.B	374170	RSS	CACAGTGCTTTAGCCTTCTACACAAACC	374170	374197
NOD.IGK.SuperContig.B	IGKV8.1	V-Gene	GACATTTTGATGACTCAGTCTCCATCCTCCCTGACTGTGTCAGCAGGAGAGAAGGTCACTATGAGCTGCAAGTCCAGTCAGAGTCTTTTAGCTAGTGCCAACCAAAATAACTACTTGGCCTGGCACCAGCAGAAACCAGGACGATCTCCTAAAATGCTGATAATTTGGGCATCCACTAGGGTATCTGGAGTCCCTGATCGCTTCATAGGCAGTGGATCTGGGACGGATTTCACTCTGACCATCAACAGTGTGCAGGCTGAAGATCTGGCTGTTTATTACTGTCAGCAGTCCTACAGCGCTCCTAC	374198	374502
NOD.IGK.SuperContig.B	L-Part2	Leader	GTACCTGTGGA	374503	374513
NOD.IGK.SuperContig.B	L-Part1	Leader	ATGGATTCACAGGCCCAGGTCCTCATGTTGCTGCTGCTATCGGTATCTG	374716	374764

NOD.IGK.SupercontigA	L-Part1	Leader	ATGGACATGAGGGCTCCTGCTCAGTTTCTTGGGATCTTGTTGCTCTGGTTCCCAG	384641	384695
NOD.IGK.SupercontigA	L-Part2	Leader	CCAGATGT	384812	384819
NOD.IGK.SupercontigA	IGKV14.1	V-Gene	GAAATCCAGATGACCCAGTCTCCATCCTCTATGTCTGCATCTCTGGGAGACAGAATAACCATCACTTGCCAGGCAACTCAAGACATTGTTAAGAATTTAAACTGGTATCAGCAGAAACCAGGGAAACCCCCTTCATTCCTGATCTATTATGCAACTGAACTGGCAGAAGGGGTCCCATCAAGGTTCAGTGGCAGTGGGTCTGGGTCAGACTATTCTCTGACAATCAGCAACCTGGAGTCTGAAGATTTTGCAGACTATTACTGTCTACAGTTTTATGAGTTTCCTCC	384820	385106
NOD.IGK.SupercontigA	385134	RSS	CACAGTGATACAAGTCATAACATAAACC	385107	385134
NOD.IGK.SuperContig.B	394023	RSS	CACAGTGCTTCAGCCTCCTACACAAACC	394023	394050
NOD.IGK.SuperContig.B	IGKV7.1	V-Gene	GACATTGTGATGACTCAGTCTCCAACTTTCCTTGCTGTGACAGCAAGTAAGAAGGTCACCATTAGTTGCACGGCCAGTGAGAGCCTTTATTCAAGCAAACACAAGGTGAACTACTTAGCTTGGTACCAGAAGAAACCAGAGCAATCTCCTAAACTGCTGATTTATGGGGCATCCAACCGATTCACTGGGGTCCCTGATCGCTTCACAGGCAGTGGATCTGGGACAGATTTCACTCTGACCATCAGC	394051	394355

			AGTGTGCAGGCTGAAG ACCTCACACATTATTAC TGTGCACAGTTTTACAG CTATCCTCT		
NOD.IGK. SuperCon tig.B	L-Part2	Leader	CCTGTGCA	394356	394363
NOD.IGK. SuperCon tig.B	L-Part1	Leader	ATGGAGTTTCAGACCC AGGTACTCATGTCCCT GCTGCTCTGCGTGTCT G	394570	394618
NOD.IGK. SuperCon tig.B	394023	RSS	CACAGTGCTTCAGCCT CCTACACAAACC	410653	410680
NOD.IGK. SuperCon tig.B	IGKV6.1	V-Gene	AGTATTGTGATGACCCA GACTCCCAAATTCCTG CCTGTATCAGCAGGAG ACAGGGTTACCATGAC CTGCAAGGCCAGTCAG AGTGTGGGTAATAATGT AGCCTGGTACCAACAG AAGCCAGGACAGTCTC CTAAACTGCTGATATAC TATGCATCCAATCGCTA CACTGGAGTCCCTGAT CGCTTCACTGGCAGTG GATCTGGGACAGATTT CACTTTCACCATCAGCA GTGTGCAGGTTGAAGA CCTGGCAGTTTATTTCT GTCAGCAGCATTATAG CTCTCCTCC	410681	410967
NOD.IGK. SuperCon tig.B	L-Part2	Leader	GTGCTCATGGG	410968	410978
NOD.IGK. SuperCon tig.B	L-part1	Leader	ATGAAGTCACAGACCC AGGTCTTCATATTTCTA CTGCTCTGTGTGTCTG	411194	411242
NOD.IGK. SupercontigA	L-Part1	Leader	ATGGACATGAGGGCTC CTGCTCAGGTTTTTGG	432584	432638

			CTTCTTGTTGCTCTGGTTTCCAG		
NOD.IGK.SupercontigA	L-Part2	Leader	CCAGATGT	432763	432770
NOD.IGK.SupercontigA	IGKV9.2	V-Gene	GACATCCAGATGACCCAGTCTCCATCCTCCTTATCTGCCTCTCTGGGAGAAAGAGTCAGTCTCACTTGCCGGGCAAGTCAGGACATTCATGGTTATTTAAACTTGTTTCAGCAGAAACCAGGTGGAACTATTAAACACCTGATCTATGAAACATCCAATTTAGATTCTGGTGTCCCAAAAAGGTTCAGTGGCAGTAGGTCTGGGTCAGATTATTCTCTCATTATCAGCAGCCTTGAGTCTGAAGATTTTGCAGACTATTACTGTCTACAATATGCTAGTTATCCTCC	432771	433057
NOD.IGK.SupercontigA	433085	RSS	CACAGTGATACAAATCACAACATAAACC	433058	433085
NOD.IGK.SuperContig.B	394023	RSS	CACAGTGCTTCAGCCTCCTACACAAACC	435517	435544
NOD.IGK.SuperContig.B	IGKV8.1.P	V-Gene	GACATTGTGATGTCACAGTCTCCATCCGCCCTAGCTGTGTCAGTTGGAGAGAAGGTCACTATGAGCTGCAAGTCCAGTCAGAGCCTTTTAATAGTAGAATCAAAGAACTACTTGGCTGGTACCAGCAGAAACCAGGGCAGTCTCCTAAACTGTTAATCTACTGGGCATCCACTAGGGAATCTGGGGTCCCTGACCGCTTCACAGGCAGTAGATCAGGGACAGATTTCACTCTCACCATCAGCAGTGTGCAGGCTGAAGACCTGGCCGTTTATTACTG	435545	435842

			CAAGCAATCTTATAATC TTCC		
NOD.IGK. SuperCon tig.B	L-Part2	Leader	GTACCTGTGGG	435843	435853
NOD.IGK. SuperCon tig.B	L-Part1	Leader	ATGGATTCACAGGCCC AGGTTCTTATTTGCTGC TGCTATGGGTATCTG	436050	436097
NOD.IGK. Supercont igA	L-Part1	Leader	ATGGACATGAGGGCCC CTGCTCAGTTTCTTGG GATCTTGTTGCTCTGGT TTCCAG	440197	440251
NOD.IGK. Supercont igA	L-Part2	Leader	GTGCCAGATGT	440365	440375
NOD.IGK. Supercont igA	IGKV9.3	V-Gene	GACATCTGGATGACTC AGTCTCCATCGTCCAT GTTTGGCTCTCTGGGA GACAGAGTCAGTCTCA CTTGCCAGGCTAGTCA GAGCATTAGAGTTTATT TAAGCTGGTATCAGCA GAAACAGGTTGGAACT ATTAAACTCCTGATCTA CTCTACATCCAAATTAG ATTCTGGTGTCCCATCA AGGTTCAGTGGCAGTG GGTCTTGGTCAGATTAT TCTCTCACCATCAACAG CCTAGAGTCTGAAGAT GTAGCAATTTATTACTG TCTACAACATGCTAGTT CTCCTCC	440376	440662
NOD.IGK. Supercont igA	440690	RSS	CACAGTGATTCAAATCA TAACATAAACC	440663	440690
NOD.IGK. Supercont igA	L-Part1	Leader	ATGACCATGCTCTCACT AGCTCCTCTCCTCAGC	453848	453902

			CTTCTTCTCCTCTGTGT CTCTG		
NOD.IGK. Supercont igA	L-Part2	Leader	ATTCTAGGGCA	454051	454061
NOD.IGK. Supercont igA	IGKV17.2	V-Gene	GAAACAACTGTGACCC AGTCTCCAGCATCCCT GTCCGTGGCTACAGGA GAAAAAGTCACTATCAG ATGCATAACCAACACTG ATATTGATGATGATATG AACTGGTACCAGCAGA AGCCAGGGGAACCTCC TAAGCTCCTTATTTCAG AAGGCAATACTCTTCGT CCTGGAGTCCCATCCC GATTCTCCAGCAGTGG CTATGGCACAGATTTTG TTTTTACAATTGAAAAC ACGCTCTCAGAAGATG TTGCAGATTACTACTGT TTGCAAAGTGATAACAT GCCTCT	454062	454348
NOD.IGK. Supercont igA	230551	RSS	CACAGTGCTATGTCCT CTTACAGAAACC	454349	454376
NOD.IGK. SuperCon tig.B	394023	RSS	CACAGTGCTTCAGCCT CCTACACAAACC	457033	457060
NOD.IGK. SuperCon tig.B	IGKV8.2	V-Gene	GACATTGTGATGACAC AGTCTCCATCCTCCCT GAGTGTGTCAGCAGGA GATAAGGTCACTATGA GCTGCAAGTCCAGTCA GAGTCTGTTAAACAGTA GAAACCAAAAGAACTA CTTGGCCTGGTACCAG CAGAAACCATGGCAGC CTCCTAAACTGCTGATC TACGGGGCATCCACTA GGGAATCTGGGGTCCC TGATCGCTTCACAGGC AGTGGATCTGGAACAG ATTTCACTCTCACCATC AGCAGTGTGCAGGCTG AAGACCTGGCAGTTTAT	457061	457365

			TACTGTCAGAATGATTATAGTTATCCTCC		
NOD.IGK.SuperContig.B	L-Part2	Leader	GTACCTGTGGG	457366	457376
NOD.IGK.SuperContig.B	L-Part1	Leader	ATGGAATCACAGACTCAGGTCCTCATGTCCCTGCTGCTCTGGGTATCTG	457581	457629
NOD.IGK.SupercontigA	L-Part1	Leader	ATGGACATGAGGGCCCCTGCTCAGTTTTTTGGGATCTTGTTGCTCTGGTTTCCAG	485437	485491
NOD.IGK.SupercontigA	L-Part2	Leader	GTATCAGATGT	485605	485615
NOD.IGK.SupercontigA	IGKV14.2	V-Gene	GACATCAAGATGACCCAGTCTCCATCCTCCATGTATGCATCGCTGGGAGAGAGAGTCACTATCACTTGCAAGGCGAGTCAGGACATTAAAAGCTATTTAAGCTGGTACCAGCAGAAACCATGGAAATCTCCTAAGACCCTGATCTATTATGCAACAAGCTTGGCAGATGGGGTCCCATCAAGATTCAGTGGCAGTGGATCTGGGCAAGATTATTCTCTAACCATCAGCAGCCTGGAGTCTGACGATACAGCAACTTATTACTGTCTACAGCATGGTGAGAGCCCTCC	485616	485902
NOD.IGK.SupercontigA	485930	RSS	CACGTTGATACAAGTAATAACATAAACC	485903	485930

NOD.IGK.SuperContig.B	L-part1	Leader	ATGGAATCACAGACCCATGTCCTCATGTTCTTGCTGCTTTGGGTATCTG	492015	492063
NOD.IGK.SuperContig.B	L-Part2	Leader	ATACCTGTGGG	492262	492272
NOD.IGK.SuperContig.B	IGKV8.3	V-Gene	GACATTGTGATGATCCAGTCTCCATCCTCCCTGGCTGTGACAGCAGGAGAGAAGGTCACTATGAGATGCAAGTCCAGTCAGAGTCTTTTGTGGAGTGTAAACCAAAAGAACTACTTGTCCTGGTACCAGCAGAAACAAGGGCAGCCTCCTAAACTGCTTATCTATGGGGCATCCATTAGAGAATCTTGGGTCCCTGATCGATTCACAGGAAGTGGATCTGGGACAGACTTCACTCTCACCATTAGCAATGTGCATGCTGAAGACCTAGCAGTTTATTACTGTCAGCACAATCATGGCAGCTTTCTCCCCC	492273	492582
NOD.IGK.SuperContig.B	492583	RSS	CACAGAGCTTCAGCTGCCTACACAAACC	492583	492610
NOD.IGK.SuperContig.B	L-Part1	Leader	ATGGGGTCAGCTACTCTCCTGCTGTGGGTGCTGATGCCCTGAGTACCAG	499211	499259
NOD.IGK.SuperContig.B	L-Part2	Leader	GCTCTACTGGG	499502	499512
NOD.IGK.SuperContig.B	IGKV3.1.P	V-Gene	GACATTGTGCTGACCCAGACTTCAGCTTGTTTTGTTGTGTCTCTGGGACAAAGAGACAACATATCCTGCAAGTCCTGTGAGAGCGTCACTGACACACTGTGCAACAACAGTATGCAGTGGTATCAGCAGAACCCAGAACAGCTTCCCATAGTCCTGATTTATGAAGCATACAGTGTAGA	499513	499814

			GTCTGGTGTTTATGTCA GATTCTGTGGTAGCTG GTGTGAGACAGATTTC AAACTCACAGTTTATCC TGTAGAGGACAGCAAT GCTGCAAGCTATTATTG TCAGCAGAGTAAGGAA TTTTTTCA		
NOD.IGK. SuperCon tig.B	499815	RSS	CACAGTTTTCTATGTCT AAGCAAGAAAT	499815	499842
NOD.IGK. Supercont igA	L-Part1	Leader	ATGAACATGAGGGCCC CTGCTGAGTTCCTTGG GTTCCTGTTGCTCTGGT TTTTAG	509292	509346
NOD.IGK. Supercont igA	L-Part2	Leader	GTGCCAGATGT	509474	509484
NOD.IGK. Supercont igA	IGKV11.1	V-Gene	GATGTCCAGATGATTCA GTCTCCATCCTCCCTGT CTGCATCTTTGGGAGA CATAGTCACCATGACTT GCCAGGCAAGTCAGGG CACTAGCATTAATTTAA ACTGGTTTCAGCAAAAA CCAGGGAAAGCTCCTA AGCTCCTGATCTATGGT GCAAGCAACTTGGAAG ATGGGGTCCCATCAAG GTTCAGTGGCAGTAGA TATGGGACAGATTTCAC TCTCACCATCAGCAGC CTGGAGGATGAAGATA TGGCAACTTATTTCTGT CTACAGCATAGTTATCT CCCTCC	509485	509771
NOD.IGK. Supercont igA	385134	RSS	CACAGTGATACAAGTC ATAACATAAACC	509772	509799
NOD.IGK. SuperCon tig.B	L-Part1	Leader	ATGGAGTCACAGATTC AGGTCTTTGTATTCGTG TTTCTCTGGTTGTCTG	511351	511399
NOD.IGK. SuperCon tig.B	L-Part2	Leader	GTGT	511578	511581

NOD.IGK.SuperContig.B	IGKV6.1.P	V-Gene	TGCATTGTGATGACCCAGTCTCACAAATTCATGTCCACATCAGTAGGAGACAGGGTCAGCATCACCTGCAAGGCCAGTCAGGATGTGAGTACTGCTGTAGCCTGGTATCAACAAAAACCAGGGCAATCTCCTAAACTACTGATTTACTGGGCATCCACCCGGCACACTGGAGTCCCTGATCGCTTCACAGGCAGTGGATCTGGGACAGATTTCACTCTTACCATCAGCAGTGTGCAGGCTGAAGACCTGGCACTTTATTACTGTCAGCAACATTATAGCACTCCTCC	511582	511868
NOD.IGK.SuperContig.B	511869	RSS	CACAGTGCTTCAGCCTTCTACATAAACC	511869	511896
NOD.IGK.SuperContig.B	511869	RSS	CACAGTGCTTCAGCCTTCTACATAAACC	512741	512768
NOD.IGK.SuperContig.B	IGKV8.4	V-Gene	GACATTGTGATGACACAGTCTCCATCCTCCCTGGCTATGTCAGTAGGACAGAAGGTCACTATGAGCTGCAAGTCCAGTCAGAGCCTTTTAAGTAGTAGCAATCAAAAGAACTATTTGGCCTGGTACCAGCAGAAACCAGGACAGTCTCCTAAACTTCTGGTATACTTTGCATCCACTAGGGAATCTGGGGTCCCTGATCGCTTCATAGGCAGTGGATCTGGGACAGATTTCACTCTTACCATCAGCAGTGTGCAGGCTGAAGACCTGGCACTTTATTACTGTCAGCAACATTATAGCACTCCTCC	512769	513073
NOD.IGK.SuperContig.B	L-Part2	Leader	GTGCCTGTGCA	513074	513084

NOD.IGK.SuperContig.B	L-Part1	Leader	ATGGAATCACAGACCCGGGTCCTCATGTTTCTTCTGCTCTGGGTATCTG	513284	513332
NOD.IGK.SuperContig.B	394023	RSS	CACAGTGCTTCAGCCTCCTACACAAACC	536689	536716
NOD.IGK.SuperContig.B	IGKV8.2.P	V-Gene	GAAATTGTGTTGACACAGTCTATACCATCCCTGACTGTGTCAGCAGGAGAGAGGGTCACTATCAGCTGCAAATCCAATCAGAATCTTTTATGGAGTGGAAACCAACGGTACTGTTTGGTCTGGCACCAGTGGAAACCAGGGCAAACTCCTACACCGTTGATCACCTGGACATCTGATAGGATACTCTGGAGTCCCTGATCGTTTCATAGGCAGTGGATCTGTGACAGATTTCACTCTGACCATCAGCAGTGTGCAGGATGAAGATGTGGCAGTTTATTTCTGTCAGCAGCATTTACACATTCCTCC	536717	537022
NOD.IGK.SuperContig.B	L-Part2	Leader	GTGCTTGTGGG	537023	537033
NOD.IGK.SuperContig.B	L-Part1	Leader	ATGGTATCAGAGACCCATGTCCTCATATTTTTGCTGCTATGGGTGTCTG	537236	537284
NOD.IGK.SupercontigA	545070	RSS	CACAGTGATACAAATCATAACATAAACC	545043	545070
NOD.IGK.SupercontigA	IGKV9.4	V-Gene	GACATCCAGATGACCCAGTCTCCATCCTCCTTATCTGCCTCTCTGGGAGAAAGAGTCAGTCTCACTTGTCGGGCAAGTCAGGAAATTAGTGGTTACTTAAGCTGGCTTCAGCAGAAACCAGATGGAACTATTAAACGCCTGATCTACAGCACATCCACTTTAGATTCTGGTGTCCCAAAAAGGTTCAGTGGCAGTAG	545071	545357

			GTCTGGGTCAGATTATT CTCTCACCATCAGCAG CCTTGAGTCTGAAGATT TTGCAGACTATTACTGT CTACAATATGCTAGTTC TCCTCC		
NOD.IGK. SupercontigA	L-Part2	Leader	GTGCCAGATGT	545358	545368
NOD.IGK. SupercontigA	L-Part1	Leader	ATGGACATGAGGGTTC CTGCTCACGTTTTTGGC TTCTTGTTGTTCTGGTT TCCAG	545490	545544
NOD.IGK. SuperContig.B	555769	RSS	CACACTACACTACACTA ATGGTAAATAT	555769	555796
NOD.IGK. SuperContig.B	IGKV6.2.P	V-Gene	GACATTGTGATGACCC AGTTCAAAAAATTTATG TCCACTTCAGTAGGAG AGAGGGTCAGCATCAC TTGCAAGGCCAGTCAG AATGTGGGCACTGCTG TAGCCTGGTATCAACA GAAACCAGGGCAGTCT CCTAAACTACTGATTTA CTGGGCATCCAACTGG TACACTGGAGTCCCTG ATCGCTTCACAGGCAG TGGATCTGGAGAAAAT ATGATGTTTGAATTGTT AACAACATAATTGATCT AGGCTTACTATTCTTAG GCCATATCTGCCTGTG CTTTCTTCTCATTTACTT CCCAATGAAGAGACTG TGATACCTCATTC	555797	556120
NOD.IGK. SuperContig.B	L-Part2	Leader	GTGTTGATGGA	556121	556131
NOD.IGK. SuperContig.B	L-Part1	Leader	ATGCTCAGATTCTGTGT ATTGAACATAACAAATT TACCTTTTTCTATGTG	556411	556460

NOD.IGK.SuperContig.B	556650	RSS	CACAGTGCTTCAGCCTCCTACATAAACC	556650	556677
NOD.IGK.SuperContig.B	IGKV6.3.P	V-Gene	GACATTGTGATGACCCAGTCTCAAAAATTCATGTCCACATCAGTAGGAGACAGGGTCAGCGTCACCTGCAAGGCCAGTCAGTATGTGGGTACTTATGTAGCCTGGTATCAACAGAAACCAGGACAATCTCCTAAAGCACTGATTTACTCGGCATCCACCCGGTCACACTGGAGTCCCTGATCGCTTCACAGGCAGTGGATCTGGGACAGATTTTCACTCTCACCATTAGCAATGTGCAGTCTGAAGACTTGGCAGAGTATTTCTGTGAGCAATACAGCAGCTCTCCTCT	556678	556965
NOD.IGK.SuperContig.B	L-Part2	Leader	GTGTTGATGGA	556966	556976
NOD.IGK.SuperContig.B	L-part1	Leader	ATGGAGACACATTCTCAGGTCTTTGTATACATGTTGCTGTGGTTGTCTG	557155	557203
NOD.IGK.SupercontigA	545070	RSS	CACAGTGATACAAATCATAACATAAACC	563622	563649
NOD.IGK.SupercontigA	IGKV9.5	V-Gene	GACATCCAGATGATTCAGTCTCCATCGTCCATGTTTGGCTCTCTGGGAGACAGAGTCAGTCTCTCTTGCCGGGCTAGTCAGGGCATTAGAGGTAATTTAGACTGGTATCAGCAGAAACCAGGTGGAACTATTAAACTCCTGATCTACTCCACATCCAATTTAAATTCTGGTGTCCCATCAAGGTTCAGTGGCAGTGGGTCTGGGTCAGATTATTCTCTCACCATCAGCAGCCTAGAGTCTGAAGATTTTGCAGACTATTACTGTC	563650	563936

			TACAGCGTAATGCGTATCCTCT		
NOD.IGK.SupercontigA	L-Part2	Leader	GTGCCAGATGT	563937	563947
NOD.IGK.SupercontigA	L-Part1	Leader	ATGGACATCAGGGCTCCTGCTCAGTTTCTTGGCATCTTGTTGCTCTGGTTTCCAG	564058	564112
NOD.IGK.SuperContig.B	579577	RSS	CACAGTGCTTCACCTCCTACACAGACCT	579577	579604
NOD.IGK.SuperContig.B	IGKV8.3.P	V-Gene	GACATTGTAATGTCTCAGTCTCCATCCTCCCCGACTATGTCAGCAGGAAAGAAGGTTACTATGAGCTGCAAGTCCAGCCAGAGCCTTTTAGCTAGTGGCAACCAAAAGAACTACTTGGCCTGGTACCAGCAAAAACCAGGATAGTCTCCTAACCTGGTGATCTACTATGCATCCACTAGAGTACCTGATCTGTGGTACCTGATTGCTTTATAGGCAATGGATTTAGGACAGATTTCACTCTGACCATCAGAAGTATGCTGGCTGAAGACCTGACAGTTTATTACTGTCAACAACATTTTAGCTTTCCTAC	579605	579914
NOD.IGK.SuperContig.B	L-Part2	Leader	GTACCTGTGGG	579915	579925
NOD.IGK.SuperContig.B	L-Part1	Leader	ATGGATTCACAGGCCCAGGTCCTCATGTTGCTGCTGCTATGGGTATCTG	580125	580173

NOD.IGK.SuperContig.B	394023	RSS	CACAGTGCTTCAGCCTCCTACACAAACC	590801	590828
NOD.IGK.SuperContig.B	IGKV8.5	V-Gene	GACATTGTGATGTCACAGTCTCCATCCGCCCTAGCTGTGTCAGTTGGAGAGAAGGTCACTATGAGCTGCAAGTCCAGTCAGAGCCTTTTATATAGTAGCAATCAAAAGAACTACTTGGCCTGGTACCAGCAGAAACCAGGGCAGTCTCCTAAACTGTTAATCTACTGGGCATCCACTAGGGAATCTGGGGTCCCTGACCGCTTCACAGGCAGTAGATCAGGGACAGATTTCACTCTCACCATCAGCAGTGTGCAGGCTGAAGACCTGGCCGTTTATTACTGCAAGCAATCTTATAATCTTCC	590829	591130
NOD.IGK.SuperContig.B	L-Part2	Leader	GTACCTGTGGG	591131	591141
NOD.IGK.SuperContig.B	L-Part1	Leader	ATGGATTCACAGGCCCAGGTTCTTATGTTGCTGCTGCTATGGGTATCTG	591338	591386
NOD.IGK.SuperContig.B	612516	RSS	CACAGTGCTTCAGCCTCCTACATAAACA	612516	612543
NOD.IGK.SuperContig.B	IGKV6.2	V-Gene	AACATTGTAATGACCCAATCTCCCAAATCCATGTCCATGTCAGTAGGAGAGAGGGTCACCTTGAGCTGCAAGGCCAGTGAGAATGTGGGTACTTATGTATCCTGGTATCAACAGAAACCAGAGCAGTCTCCTAAACTGCTGATATACGGGGCATCCAACCGGTACACTGGGGTCCCCGATCGCTTCACAGGCAGTGGATCTGCAACAGATTTCACTCTGACCATCAGCAGTGTGCAGGCTGAAGA	612544	612830

			CCTTGCAGATTATTACT GTGGACAGAGTTACAG CTATCCTCC		
NOD.IGK. SuperCon tig.B	L-Part2	Leader	GTGCTGATGGG	612831	612841
NOD.IGK. SuperCon tig.B	L-Part1	Leader	ATGGAATCACAGACTCT GGTCTTCATATCCATAC TGCTCTGGTTATATG	613125	613173
NOD.IGK. SuperCon tig.B	394023	RSS	CACAGTGCTTCAGCCT CCTACACAAACC	617609	617636
NOD.IGK. SuperCon tig.B	IGKV8.6	V-Gene	GACATTGTGATGACAC AGTCTCCATCCTCCCT GACTGTGACAGCAGGA GAGAAGGTCACTATGA GCTGCAAGTCCAGTCA GAGTCTGTTAAACAGT GGAAATCAAAAGAACTA CTTGACCTGGTACCAG CAGAAACCAGGGCAGC CTCCTAAACTGTTGATC TACTGGGCATCCACTA GGGAATCTGGGGTCCC TGATCGCTTCACAGGC AGTGGATCTGGAACAG ATTTCACTCTCACCATC AGCAGTGTGCAGGCTG AAGACCTGGCAGTTTAT TACTGTCAGAATGATTA TAGTTATCCTCC	617637	617941
NOD.IGK. SuperCon tig.B	L-Part2	Leader	GTACCTGTGGG	617942	617952
NOD.IGK. SuperCon tig.B	L-Part1	Leader	ATGGAATCACAGACTC AGGTCCTCATGTCCGT GCTGTTCTGGGTATCT G	618161	618209
NOD.IGK. SuperCon tig.B	L-Part1	Leader	ATGGAATCACAGACCC ATGTCCTCATGTTCTTG CTGCTTTGGGTATCTG	632596	632644

NOD.IGK.SuperContig.B	L-Part2	Leader	ATAGCTGTGGG	632842	632852
NOD.IGK.SuperContig.B	IGKV8.7	V-Gene	GACATTGTGATGACCCAGTCTCCATCCTCCCTGGCTGTGACAGCAGGAGAGAAGGTCACTATGAGATGCAAGTCCAGTCAGAGTCTTTTGTGGAGTGTAAACCAAAAGAACTACTTATCCTGGTACCAGCAGAAACAAGGGCAGCCTCCTAAACTGCTTATCTATGGGGCATCCATTAGAGAATCTTGGGTCCCTGATCGATTCACAGGAAGTGGATCTGGGACAGACTTCACTCTCACCATTAGCAATGTGCATGCTGAAGACCTAGCAGTTTATTACTGTCAGCACAATCATGGCAGCTTTCTCCCCC	632853	633162
NOD.IGK.SuperContig.B	492583	RSS	CACAGAGCTTCAGCTGCCTACACAAACC	633163	633190
NOD.IGK.SupercontigA	L-Part1	Leader	ATGAAGTTGCCTGTTAGGTTGTTGGTGCTGTTGTTCTGGATTCCTG	633183	633228
NOD.IGK.SupercontigA	L-Part2	Leader	CTTCCAGCAGT	633601	633611
NOD.IGK.SupercontigA	IGKV1.7	V-Gene	GATGCTGTGATGACCCAAACTCCACTCTCCCTGCCTGTCAGTCTTGGAGATCAAGCCTCCATCTCTTGCAGGTCTAGTCAGAGCCTTGAAAACAGTAATGGAAACACCTATTTGAACTGGTACCTCCAGAAACCAGGCCAGTCTCCACAGCTCCTGATCTACAGGGTTTCCAACCGATTTTCTGGGGTCCTAGACAGGTTCAGTGGCAGTGGTTCAGGGACAGATTTCA	633612	633913

			CACTCAAGATCAGCAG AGTGGAGGCTGAGGAT TTGGGAGTTTATTTCTG CCTCCAAGTTACACATG TCCCTCC		
NOD.IGK. SupercontigA	633941	RSS	CACAGTGATGCAGACC CTAACAAAAACA	633914	633941
NOD.IGK. SuperContig.B	L-Part1	Leader	ATGGAGTCACAGATTC AGGTCTTTGTATTCGTG TTTCTCTGGTTGTCTG	648186	648234
NOD.IGK. SuperContig.B	L-Part2	Leader	GTGTTGAAGGA	648413	648423
NOD.IGK. SuperContig.B	IGKV6.3	V-Gene	GACATTGTGATGACCC AGTCTCACAAATTCATG TCCACATCAGTAGGAG ACAGGGTCAGCATCAC CTGCAAGGCCAGTCAG GATGTGGGTACTGCTG TAGCCTGGTATCAACA GAAACCAGGGCAATCT CCTAAACTACTGATTTA CTGGGCATCCACCCGG CACACTGGAGTCCCTG ATCCCTTCACAGGCAG TGGATCTGGGACAGAT TTCACTCTCACCATCAG CAGTGTGCAGGCTGAA GACCTGGCAGTTTATTA CTGTCAGCAAGATTATA GCACTCCTCC	648424	648710
NOD.IGK. SuperContig.B	394023	RSS	CACAGTGCTTCAGCCT CCTACACAAACC	648711	648738
NOD.IGK. SuperContig.B	394023	RSS	CACAGTGCTTCAGCCT CCTACACAAACC	670669	670696

NOD.IGK.SuperContig.B	IGKV8.8	V-Gene	GAAATTGTGTTGACACAGTCTATACCATCCCTGACTGTGTCAGCAGGAGAGAGGGTCACTATCAGCTGCAAATCCAATCAGAATCTTTTATGGAGTGGAAACCAACGGTACTGTTTGGTCTGGCACCAGTGGAAACCAGGGCAAACTCCTACACCGTTGATCACCTGGACATCTGATAGGTACTCTGGAGTCCCTGATCGTTTCATAGGCAGTGGATCTGTGACAGATTTCACTCTGACCATCAGCAGTGTGCAGGATGAAGATGTGGCAGTTTATTTCTGTCAGCAGCATTTACACATTCCTC	670698	671001
NOD.IGK.SuperContig.B	L-Part1	Leader	GTGCTTGTGGG	671002	671012
NOD.IGK.SuperContig.B	L-Part1	Leader	ATGGTATCAGAGACCCATGTCCTCATATTTTTGCTGCTATGGGTGTCTG	671215	671263
NOD.IGK.SupercontigA	L-Part1	Leader	ATGGACATGAGGGCTCCTGCACAGGTTTTTGGCTTCTTGTTGCTCTTGTTTCCAG	682342	682396
NOD.IGK.SupercontigA	L-Part2	Leader	GTGCCAGATGT	682518	682528
NOD.IGK.SupercontigA	IGKV9.6	V-Gene	GACATCCAGATGACCCAGTCTCCATCCTCCTTATCTGCCTCTCTGGGAGAAAGAATCAGTCTCACTTGCCGGGCAAGTCAGGACATTTATGGTAGCTTAAACTGGTTTCAGCAGAAACCAGATGGAACTATTAAACTCCTGATCTACGGCACATCCAGTTTAGATTCTGGTGTCCCCAAAAGGTTCAGTGGCAGTAGGTCTGGGTCAGATTATTCTCTCACCATCAGCAG	682529	682815

			CCTTGAGTCTGAAGATTTTGCAGACTATTACTGTCTACAATATGCTAGTTCTCCTCC		
NOD.IGK.SupercontigA	682843	RSS	CACAGTGATAAAAATCATAACATAAACC	682816	682843
NOD.IGK.SupercontigA	L-Part1	Leader	ATGGACATGCGGACCCCTGCTCAGTTTCTTGGGATCTTGTTGCTCTGGTTTCCAG	688696	688750
NOD.IGK.SupercontigA	L-Part2	Leader	GTGCCAGATGT	688864	688874
NOD.IGK.SupercontigA	IGKV9.1.P	V-Gene	GACATCCGGATGACTCAGTCTCCATCGTCTATGTTTGCCTCTCTGGGAGACAGAGTCAGTCTCACTTGAAGGGGTAGTCAGAGCATTAGAGTTTATTTAAGCTGGTATCAGCAGAAACCAGGTGGAACTATTAAACTCCTGATCTACTCCACATCCATATTAGATTCTGGTGTCGCATCAAGGTTCAGGGGCAGTGGGACTGGGTCAGATTATTCTCTCACCATCAACAGCCTAGAGTCTGAAGATGTGGCAATTTATTACTCTCTACAATATGCTAGTTCTCCTCC	688875	689161
NOD.IGK.SupercontigA	433085	RSS	CACAGTGATACAAATCACAACATAAACC	689162	689189
NOD.IGK.SuperContig.B	556650	RSS	CACAGTGCTTCAGCCTCCTACATAAACC	690388	690415

NOD.IGK.SuperContig.B	IGKV6.4.P	V-Gene	GACATTGTGATGACCCAGTCTCAAAATTCATGTCCACATCAGTAGGAGACAGGGTCAGCGTCACCTGCAAGGCCAGTCAGAATGTGGGTACTATGTAGCCTGGTATCAACAGAAACCAGGCAATCTCCTAAAGCACTGATTTACTGGCATCCACCGGTCACTGGAGTCCCTGATCGCTTCACAGGCAGTGGATCTGGGACAGATTTCACTCTCACCATTAGCAATGTGCAGTCTGAAGACTTGGCAGATATTTCTGTGAGCAATACAGCAGCTCTCCTCT	690416	690695
NOD.IGK.SuperContig.B	L-Part2	Leader	GTGTTGATGGA	690696	690706
NOD.IGK.SuperContig.B	L-Part1	Leader	ATGAAGACACATTCTCAGGTCTTTGTATACATGTTGCTGTGGTTGTCTG	690881	690929
NOD.IGK.SuperContig.B	579577	RSS	CACAGTGCTTCACCTCCTACACAGACCT	713137	713164
NOD.IGK.SuperContig.B	IGKV8.4.P	V-Gene	GACATTGTAATGTCTCAGTCTCCATCCTCCCCGACTATGTCAGCAGGAAAGAAGGTTACTATGAGCTGCAAGTCCAGCCAGAGCCTTTTAGCTAGTGGCAACCAAAAGAACTACTTGGCCTGGTACCAGCAAAAACCAGGATAGTCTCCTAACCTGGTGATCTACTATGCATCCACTAGAGTACCTGATCTGTGGTACCTGATTGCTTTATAGGCAATGGATTTAGGACAGATTTCACTCTGACCATCAGAAGTATGCTGGCTGAAGACCTGACAGTTTATTACTGTCAACAACATTTTAGCTTTCCTAC	713165	713474

NOD.IGK.SuperContig.B	L-Part2	Leader	GTACCTGTGGG	713475	713485
NOD.IGK.SuperContig.B	L-Part1	Leader	ATGGATTCACAGGCCCAGGTCCTCATGTTGCTGCTGCTATGGGTATCTG	713685	713733
NOD.IGK.SuperContig.B	394023	RSS	CACAGTGCTTCAGCCTCCTACACAAACC	724350	724377
NOD.IGK.SuperContig.B	IGKV8.9	V-Gene	GACATTGTGATGTCACAGTCTCCATCCGCCCTAGCTGTGTCAGTTGGAGAGAAGGTCACTATGAGCTGCAAGTCCAGTCAGAGCCTTTTATATAGTAGCAATCAAAAGAACTACTTGGCCTGGTACCAGCAGAAACCAGGGCAGTCTCCTAAACTGTTAATCTACTGGGCATCCACTAGGGAATCTGGGGTCCCTGACCGCTTCACAGGCAGTAGATCAGGGACAGATTTCACTCTCACCATCAGCAGTGTGCAGGCTGAAGACCTGGCCGTTTATTACTGCAAGCAATCTTATAATCTTCC	724378	724679
NOD.IGK.SuperContig.B	L-Part2	Leader	GTACCTGTGGG	724680	724690
NOD.IGK.SuperContig.B	L-Part1	Leader	ATGGATTCACAGGCCCAGGTTCTTATGTTGCTGCTGCTATGGGTATCTG	724887	724935
NOD.IGK.SupercontigA	L-Part1	Leader	ATGAATTTGCCTGGTAGGCTGTTGGTGTTGTTGTTCTGGATTTCAG	729421	729466
NOD.IGK.SupercontigA	L-Part2	Leader	CTTCCAGAGGT	729855	729865

NOD.IGK.SupercontigA	IGKV1.1.P	V-Gene	GGTGTTGTGGTGACACAAACTCCAGTCTCCCTGACTGTCAGCCTTGGAGATCAAGCCTCTATCTCTTGCAGGTCTAGTCAAAGCCTTGTACACAAAAAATGGAAATACCTATTTGAAATGGTTCGTGCAGAAGCCTGGCCAGTCTCCACAGCTCCTCATCTATGAGATTTCCAGCCGATTTTCTGGGGTCCCAGACAGGTTCAGCGGCAGTGGTCCAAGGACAGATTTCACACTTAAGATCAACAGAGTGGAGCTTGAAGAACTGGGCGTTTATTACTGCTTCCAAGGTACACATTTACCTCC	729866	730168
NOD.IGK.SupercontigA	1071654	RSS	CACAGTGATGCAGACCCTAACAAAAATG	730169	730196
NOD.IGK.SuperContig.B	612516	RSS	CACAGTGCTTCAGCCTCCTACATAAACA	746046	746073
NOD.IGK.SuperContig.B	IGKV6.4	V-Gene	AACATTGTAATGACCCAATCTCCCAAATCCATGTCCATGTCAGTAGGAGAGAGGGTCACCTTGAGCTGCAAGGCCAGTGAGAATGTGGGTACTTATGTATCCTGGTATCAACAGAAACCAGAGCAGTCTCCTAAACTGCTGATATACGGGGCATCCAACCGGTACACTGGGGTCCCCGATCGCTTCACAGGCAGTGGATCTGCAACAGATTTCACTCTGACCATCAGCAGTGTGCAGGCTGAAGACCTTGCAGATTATTACTGTGGACAGAGTTACAGCTATCCTCC	746074	746360
NOD.IGK.SuperContig.B	L-Part2	Leader	GTGCTGATGGG	746361	746371

NOD.IGK.SuperContig.B	L-Part1	Leader	ATGGAATCACAGACTCTGGTCTTCATATCCATACTGCTCTGGTTATATG	746654	746702
NOD.IGK.SupercontigA	L-Part1	Leader	ATGAAGTTGCCTGTTAGGCTGTTGGTGCTGATGTTCTGGATTCCTG	748683	748728
NOD.IGK.SupercontigA	L-Part2	Leader	CTTCCAGCAGT	749115	749125
NOD.IGK.SupercontigA	IGKV1.8	V-Gene	GATGTTGTGATGACCCAAACTCCACTCTCCCTGCCTGTCAGTCTTGGAGATCAAGCCTCCATCTCTTGCAGATCTAGTCAGAGCATTGTACATAGTAATGGAAACACCTATTTAGAATGGTACCTGCAGAAGCCAGGCCAGTCTCCAAAGCTCCTGATCTACAAAGTTTCCAACCGATTTTCTGGGGTCCCAGACAGGTTCAGTGGCAGTGGATCAGGGACAGATTTCACACTCAAGATCAGCAGAGTGGAGGCTGAGGATCTGGGAGTTTATTACTGCTTTCAAGGTTCACATGTTCCTCA	749126	749427
NOD.IGK.SupercontigA	749455	RSS	CACAGTGATACAGACCCTAACAAAAATA	749428	749455
NOD.IGK.SuperContig.B	394023	RSS	CACAGTGCTTCAGCCTCCTACACAAACC	751129	751156
NOD.IGK.SuperContig.B	IGKV8.10	V-Gene	GACATTGTGATGACACAGTCTCCATCCTCCCTGACTGTGACAGCAGGAGAGAAGGTCACTATGAGCTGCAAGTCCAGTCAGAGTCTGTTAAACAGTGGAAATCAAAAGAACTACTTGACCTGGTACCAGCAGAAACCAGGGCAGCCTCCTAAACTGTTGATCTACTGGGCATCCACTAGGGAATCTGGGGTCCC	751157	751461

			TGATCGCTTCACAGGC AGTGGATCTGGAACAG ATTTCACTCTCACCATC AGCAGTGTGCAGGCTG AAGACCTGGCAGTTTAT TACTGTCAGAATGATTA TAGTTATCCTCC		
NOD.IGK.SuperContig.B	L-Part2	Leader	GTACCTGTGGG	751462	751472
NOD.IGK.SuperContig.B	L-Part1	Leader	ATGGAATCACAGACTC AGGTCCTCATGTCCGT GCTGTTCTGGGTATCT G	751681	751729
NOD.IGK.SupercontigA	L-Part1	Leader	ATGAGGTTCTCTGCTCA GCTTCTGGGGCTGCTT GTGCTCTGGATCCCTG	771577	771625
NOD.IGK.SupercontigA	L-Part2	Leader	GATCCACTGCA	771991	772001
NOD.IGK.SupercontigA	IGKV2.3	V-Gene	GATATTGTGATGACGC AGGCTGCATTCTCCAAT CCAGTCACTCTTGGAA CGTCAGCTTCCATCTC CTGCAGGTCTAGTAAG AGTCTCCTACATAGTGA TGGCATCACTTATTTGT ATTGGTATCTGCAGAG GCCAGGCCAGTCTCCT CAGCTCCTGATATATCG GATGTCCAACCTTGCC TCAGGAGTCCCAGACA GGTTCAGTGGCAGTGG GTCAGGAACTGATTTCA CACTGAGAATCAGCAG AGTGGAGGCTGAGGAT GTGGGTGTTTATTACTG TGCTCAAATGCTAGAAT TCCCTCC	772002	772303
NOD.IGK.SupercontigA	772331	RSS	CACACTGATACAGCCC TGAACAAAAACA	772304	772331

NOD.IGK.Supercontig A	L-Part1	Leader	ATGAGGTGCTCTCTTCAGTTCTTGGGGATGCTTATGTTCTGGATCTCTG	810273	810321
NOD.IGK.Supercontig A	L-Part2	Leader	GAGTCAGTGGG	810687	810697
NOD.IGK.Supercontig A	IGKV2.4	V-Gene	GATATTGTGATAACCCAGGATGAACTCTCCAATCCTGTCACTTCTGGAGAATCAGTTTCCATCTCCTGCAGGTCTAGTAAGAGTCTCCTATATAAGGATGGGAAGACATACTTGAATTGGTTTCTGCAGAGGCCAGGACAGTCTCCTCAGCTCCTGGTCTATTGGATGTCCACCCGTGCATCAGGAGTCTCAGACCGGTTTAGTGGCAGTGGGTCAGGAACAGATTTCACACTGGAAATCAGTAGAGTGAAGGCTGAGGATGTCGGTGTGTATTACTGTCAACAAGTTGTAGAGTATCCTC	810698	810998
NOD.IGK.Supercontig A	811026	RSS	CACAGTGCTAGAGCCCTGAACAAAAACC	810999	811026
NOD.IGK.Supercontig A	L-Part1	Leader	ATGGACATGAGGACCCCTGCTCAGTTTCTTGGGATCTTGTTGCTCTGGTTTCCAG	825327	825381
NOD.IGK.Supercontig A	L-Part2	Leader	GTATCAACTGT	825495	825505
NOD.IGK.Supercontig A	IGKV14.3	V-Gene	GACATCAAGATGACCCAGTCTCCATCTTCCATGTATGCATCTCTAGGAGAGAGAGTCACTATCACTTGCAAGGCGAGTCAGGACATTAATAGCTATTTACGCTGGTACCAGCAGAAACCAGGAAAATCTCCTAAGACCCTGATCTATGGTGCAAACAGCTTGGTAGATGGGGTCCCATCAA	825506	825792

			GGTTCAGTGGCAGTGG ATCTGGGCAAGATTATT CTCTCACCATCAGCAG TCTGGAGTATGAAGATA TGGGAATTTATTATTGT CTACAGTATGATGAGTT TCCTCC		
NOD.IGK. Supercont igA	385134	RSS	CACAGTGATACAAGTC ATAACATAAACC	825793	825820
NOD.IGK. Supercont igA	L-Part1	Leader	ATGAAGTTGCCTGTTAG GTTGTTGGTGCTGTTGT TCTGGATTCCTG	839595	839640
NOD.IGK. Supercont igA	L-Part2	Leader	CTTCCAGCAGT	840027	840037
NOD.IGK. Supercont igA	IGKV1.9	V-Gene	GATGTTGTGATGACCC AAACTCCACTCTCCCTG CCTGTCAGTCTTGGAG ATCAAGCTTCCATCTCT TGCAGATCTAGTCAGA GCCTTGTACACAGCAA TGGAAACACCTATTTAT ATTGGTACCTGCAGAA GCCAGGCCAGTCTCCA AAGCTCCTGATCTACA GGGTTTCCAACCGATTT TCTGGGGTCCCAGACA GGTTCAGTGGCAGTGG ATCAGGGACAGATTTC ACACTCAAGATCAGCA GAGTGGAGGCTGAGGA TCTGGGAGTTTATTTCT GCTTTCAAGGTACACAT GTTCCTCA	840038	840339
NOD.IGK. Supercont igA	749455	RSS	CACAGTGATACAGACC CTAACAAAAATA	840340	840367
NOD.IGK. Supercont igA	L-Part1	Leader	ATGAGGTTCTCTGCTCA GCTTCTGGGGCTGCTT GTGCTCTGGATCCCTG	852076	852124
NOD.IGK. Supercont igA	L-Part2	Leader	GATCCACTGCA	852492	852502

NOD.IGK.Supercontig A	IGKV2.5	V-Gene	GATATTGTGATGACGCAGGCTGCATTCTCCAATCCAGTCACTCTTGGAACATCAGCTTCCATCTCTTGCAGTTCTAGTAAGAGTCTCCTACATAGTAATGGCATCACTTATTTGTATTGGTATCTGCAGAGGCCAGGCCAGTCTCCTCAGCTCCTGATATATCGGATGTCCAACCTTGCCTCAGGAGTCCCAGACAGGTTCAGTGGCAGTGGGTCAGGAACTGATTTCACACTGAGAATCAGCAGAGTGGAGGCTGAGGATGTGGGTGTTTATTACTGTGCTCAAATGCTAGAACGCCCTCC	852503	852804
NOD.IGK.Supercontig A	772331	RSS	CACACTGATACAGCCCTGAACAAAAACA	852805	852832
NOD.IGK.Supercontig A	IGKV1.2.P	V-Gene	TCTAGTCAGAGCCTTGTACACAGTAATGGAAATTCCTATTTGGAATGGCACCTGCAGAAGCCAGGCCAGTCTCTACAACTCCTGATCTATGAGGTTTCCAAACGACATTCTGGGGTTCCAGACAGGTTCAGTGGCAGTGGATCAGGGACAGATTTCACACTTAAGATCAGCAGAGTATAGCCTGAGGATTTGGGAGTTTATTACTGCTTCCAAGGTACACATTTACCTCA	861827	862056
NOD.IGK.Supercontig A	862084	RSS	CACAGTGATGCAGACCTTAACAAAAATG	862057	862084
NOD.IGK.Supercontig A	IGKV1.3.P	V-Gene	TCTAGTCAGAGTCTTGTACACAGTAATGGAAATTCCTATTTGGATTGGCACCTGCAGAAGCCAGACCAGTCTCTACAACTCCTGATCTATGAGGTTTCCAAACGAAATTCTGGGGTT	867564	867793

			CCAGACAGGTTCAGTG GCAGTGGATCAGGAAC AGATTTCACACTTAAGA TCAGCAGAGTAGAGCC TGAGGATTTGGGAGTT TATTACTGCTTCCAACG TACACATTTACCTCC		
NOD.IGK. Supercont igA	862084	RSS	CACAGTGATGCAGACC TTAACAAAAATG	867794	867821
NOD.IGK. Supercont igA	L-Part1	Leader	ATGAGAACCCTGCTCC TTTCCTTGGGCTCCTGT TATTCTGTTTTCTAG	903462	903509
NOD.IGK. Supercont igA	L-Part2	Leader	GTGTCAGATGT	903631	903641
NOD.IGK. Supercont igA	IGKV11.1. P	V-Gene	GATGTTCAAATGACCCA GTCTCCATCCTCCCTGT CTGCATCTTTGGGAGA AAGAGTCTCCCTGACC TGCCAGGCAAGTCAGA GCATTAGCAATTATTTA AACTGGTATCAGCAAA CACTAGGGAAAGCTGC TAGGCTCTTGATCTATG GTGCAAGCAAATTGGA AGATGGGGTCCCTTCA AGGTTCAGTGGAACTG GATATGGGACAGATTT CACTTTCACCATCAGCA GCCAGGAGGAAGAAGA TGTGGCAACTTATTTCT GTCTACAGCATAGGTAT CTCCCTCC	903642	903928
NOD.IGK. Supercont igA	903956	RSS	CACAATGATACAAGTAA TAACATAAACC	903929	903956
NOD.IGK. Supercont igA	L-Part1	Leader	ATGAGTTACTCTCTTCA GCTCCTGAGGATGCTT GTGCTCTGGATTCCTG	910642	910690
NOD.IGK. Supercont igA	L-Part2	Leader	GAATCAGTAAA	911063	911073

NOD.IGK.Supercontig A	IGKV2.1.P	V-Gene	GTTATTGTGATGACACAGGGTGCACTGTCCAATCCTTTCACTTCTGTAGAGTCAGCTTCCATCTCTTGCAGGTCTAGTAAGAGTCTCCTACTTAAGGATGGGAAGGCATACTTGGATTGGTTTCTGCAAAGGCCCGGACACCCTACTTGGCTCCTGATCCATTTGATGCCAACCTGTGCCTCAGGAGTCTCAGACAAGTTTAGTGACAGTGGGTCAGGAACAGATTTCACACTGAAAAATCAGTAAAGTGGAGGTTGAGGATGTGGGTGTGCATTACTGTCAGCAAGTTGTAGAGTATATTC	911074	911375
NOD.IGK.Supercontig A	911403	RSS	CACAGTGATACAGCTCTGAACAAAATCC	911376	911403
NOD.IGK.Supercontig A	L-Part1	Leader	ATGAGGTTCCAGGTTCAGGTTCTGGGGCTCCTTCTGCTCTGGATACCAG	960427	960475
NOD.IGK.Supercontig A	L-Part2	Leader	GTGCCCAGTGT	960599	960609
NOD.IGK.Supercontig A	IGKV16.1	V-Gene	GATGTCCAGATAACCCAGTCTCCATCTTATCTTGCTGCATCTCCTGGAGAAACCATTACTATTAATTGCAGGGCAAGTAAGAGCATTAGCAAATATTTAGCCTGGTATCAAGAGAAACCTGGGAAAACTAATAAGCTTCTTATCTACTCTGGATCCACTTTGCAATCTGGAATTCCATCAAGGTTCAGTGGCAGTGGATCTGGTACAGATTTCACTCTCACCATCAGTAGCCTGGAGCCTGAAGATTTTGCAATGTATTACTGT	960610	960897

			CAACAGCATAATGAATCCCCGTAC		
NOD.IGK.Supercontig A	960925	RSS	CACAGTGATACAAGTTATAACAAAAACC	960898	960925
NOD.IGK.Supercontig A	L-Part1	Leader	ATGAAACTCCTTGCTGAGCTCCTGGGGCTGCTGCTGTTCTGCTTTTTAG	971068	971116
NOD.IGK.Supercontig A	L-Part2	Leader	GTGTGAGATGT	971228	971238
NOD.IGK.Supercontig A	IGKV15.1	V-Gene	GACATCCAGATGAACCAGTCTCCATCCAGTCTGTCTGCATCCCTCGGAGACACAATTACCATCACTTGCCGTGCCAGTCAGAACATTAATATTTGGTTAAGCTGGTACCAGCAGAAACCAGGAAATATTCCTAAACTATTGATCTATAAGGCTTCCAACTTGCACACAGGCGTCCCATCAAGGTTTAGTGGCAGTGGATCTGGAACAGATTTCACATTAACCATCAGCAGTCTGCAGCCTGAAGACATTGCCACTTACTACTGTCTACAGGGTCAAAGTTATCCTCT	971239	971525
NOD.IGK.Supercontig A	971553	RSS	CACAGTGATATAAGTCATAACACAAACC	971526	971553
NOD.IGK.Supercontig A	994491	RSS	CACAGTGATACAAGTCATAACACAAACA	994464	994491

NOD.IGK.Supercontig A	IGKV15.1.P	V-Gene	GATACCTAGATGAACCAGTCTCCATGCACTCTATCTGCATCAATTAGAGAATGAATAGTTATCAATTATCATACCAGTGAGAAAGTTAATACTTGGTTATCCTGTAATCAGTAGAAACTAGGGAATTATCCTAAACTACTGATCTATAACACATCCAACTTGCATACTGGGGTCCCATCAAGGTTAAGTGGCAGTAGATCTGGGACAGATTACTCTCTCTTCATCAGCAGCCTGCAGCTTGAAGATATTGCCACTTACTACTGTGTACAGTCTAGCAGTCTTTCTCT	994492	994778
NOD.IGK.Supercontig A	L-Part2	Leader	GTACGACATGT	994779	994789
NOD.IGK.Supercontig A	L-Part1	Leader	ATGGCTCTCGTTCAGCTCCTAGGGCTGCTGATGCTCTGGCTCCAAG	1003150	1003195
NOD.IGK.Supercontig A	L-Part2	Leader	GCATGAGCTGT	1003304	1003314
NOD.IGK.Supercontig A	IGKV20.1	V-Gene	AATATCCAGGTGATCCAGTCACCATTTCTGTCTGCATCTGTGGGAGAGAGGGTCACAATCAGCTGCAAGACACATCAGCATATTAACAGTTCCATAGCCTGGTACCAGCAAAAAGTTGGAAAAGCTCCCAAACTCCTGATAAGAGATGCAAGTTTTTCTCTAACAGACACCCCATCAAGGTTCACTGGGAATGGATTTGGCACAGATTTCACACTCAGCATCAGCAGTATGCAGCCTGAAGATGGTGCCACATACTTCTGCCAGCAGCATTTTAACTATTAC	1003315	1003596

NOD.IGK.Supercontig A	1003624	RSS	CACAGTGATATAAGTCATAACATAACCC	1003597	1003624
NOD.IGK.Supercontig A	IGKV15.2.P	V-Gene	AATACCCAGATGAACCAGCCTCCATCCACTCTATCTGCATCTAGCGGAGAACAAGTAATTATCAATTGTTGAGCCAGTGAGAACATTAATAGTTGGTCTTCCTGGCACCAGCAGAAACCAGGGAATGCTCCTCAAATATTGATCTATAAGGCATCCACCTTGCGTACTTGGGTCCCATCAAGGTTCAGTGGCGTTGATCCGGGGCAGATTACTCTCTCATCAGCAGCAGCCTGCAGCCTGAAGACATTGCTACTTACTATTGTGTACAGACTAAGAATTTTCCCCT	1009961	1010246
NOD.IGK.Supercontig A	L-Part2	Leader	GTGTAAGATGT	1010247	1010257
NOD.IGK.Supercontig A	L-Part1	Leader	ATGGACATGAGGGTCCTTGCTCAGTTTCTTGCATTCTTGTTGCTTTGGTTTCCAG	1048262	1048316
NOD.IGK.Supercontig A	L-Part2	Leader	CAAGATGT	1048433	1048440
NOD.IGK.Supercontig A	IGKV14.4	V-Gene	GACATCCTGATGACCCAATCTCCATCCTCCATGTCTGTATCTCTGGGAGACACAGTCAGCATCACTTGCCATGCAAGTCAGGGCATTAGCAGTAATATAGGGTGGTTGCAGCAGAAACCAGGGAAATCATTTAAGGGCCTGATCTATCATGCAACCAACTTGGAAGATGGAGTTCCATCAAGGTTCAGTGGCAGTGGATCTGGAGCAGATTATTCTCTCACCATCAGCAGCCTGGAATCTGAAG	1048441	1048727

			ATTTTGCAGACTATTAC TGTGTACAGTATGCTCA GTTTCCTCC		
NOD.IGK. Supercont igA	1048755	RSS	CACAGTGATTCAAGTCA TAACATAAACC	104872 8	1048755
NOD.IGK. Supercont igA	L-Part1	Leader	ATGAAGCTGCCTGTTCT GCTGGTGGTGCTGCTA TTGTTCATGAGTCCAG	107086 2	1070910
NOD.IGK. Supercont igA	L-Part2	Leader	CAAGCAGT	107131 7	1071324
NOD.IGK. Supercont igA	IGKV1.10	V-Gene	GATGTTGTTCTGACCCA AACTCCACTCTCTCTGC CTGTCAATATTGGAGAT CAAGCCTCTATCTCTTG CAAGTCTACTAAGAGC CTTCTGAATAGTGATGG ATTCACTTATTTGGGCT GGTACCTGCAGAAGCC AGGCCAGTCTCCACAG CTCCTAATATATTTGGT TTCTAATCGATTTTCTG GAGTTCCAGACAGGTT CAGTGGTAGTGGGTCA GGGACAGATTTCACCC TCAAGATCAGCAGAGT GGAGGCTGAGGATTTG GGAGTTTATTATTGCTT CCAGAGTAACTATCTTC CTCT	107132 5	1071626
NOD.IGK. Supercont igA	1071654	RSS	CACAGTGATGCAGACC CTAACAAAAATG	107162 7	1071654
NOD.IGK. Supercont igA	L-Part1	Leader	ATGAACATGCTCACTCA GCTCCTGGGATTACTG GTGCTCTGGTTTGCAG	110035 5	1100403
NOD.IGK. Supercont igA	L-Part2	Leader	GTGGTAAATGT	110053 0	1100540

NOD.IGK.Supercontig A	IGKV12.1.P	V-Gene	GACATTCAGATGACCCAGTCTCCTGCCTCCTAGTCTGCATCTCTGGGAGAAAGTGTCACCATCACATGCCTGGCAAGTCAGACCATTGATACATGGTTAGCATGGTATCAGCAGAAACCAGGGAAATCTCCTCAGCTCCTGATTTATGCTGCAACCAGCTTGGCAGATGGGGTCCCATCAAGGTTCAGTGGTAGTGGATCTGGCACAAAGTTTTCTTTCAAGATCAGCAGCCTACAGGCTGAAGATTTTGCAAGTTATTACTGTCAACAACATTACAGTACTCCTCT	1100541	1100827
NOD.IGK.Supercontig A	1100855	RSS	CACAGTGATTCAAGCCATAACATAAACC	1100828	1100855
NOD.IGK.Supercontig A	1151733	RSS	CACAATGATATAAGTCATAACATAAACC	1151706	1151733
NOD.IGK.Supercontig A	IGKV10.1	V-Gene	GATATCCAGATGACACAGACTACATCCTCCCTGTCTGCCTCTCTGGGAGACAGAGTCACCATCAGTTGCAGGGCAAGTCAGGACATTAGCAATTATTTAAACTGGTATCAGCAGAAACCAGATGGAACTGTTAAACTCCTGATCTACTACACATCAAGATTACACTCAGGAGTCCCATCAAGGTTCAGTGGCAGTGGGTCTGGGACAGATTATTCTCTCACTATTAGCAACCTGGAACAAGAAGATATTGCCACTTACTTTTGCCAACAGGATAGTAAGCATCCTCC	1151734	1152020
NOD.IGK.Supercontig A	L-Part1	Leader	GTACCAGATGT	1152021	1152031

NOD.IGK.Supercontig A	L-Part1	Leader	ATGATGTCCTCTGCTCAGTTCCTTGGTCTCCTGTTGCTCTGTTTTCAAG	1152153	1152201
NOD.IGK.Supercontig A	1151733	RSS	CACAATGATATAAGTCATAACATAAACC	1192031	1192058
NOD.IGK.Supercontig A	IGKV10.2	V-Gene	GATATCCAGATGACACAGACTACATCCTCCCTGTCTGCCTCTCTGGGAGACAGAGTCACCATCAGTTGCAGGGCAAGTCAGGATATTAGCAATTATTTAAACTGGTATCAGCAGAAACCAGATGGAACTGTTAAACTCCTGATCTACTACACATCAAGATTACACTCAGGAGTCCCATCAAGGTTCAGTGGCAGTGGGTCTGGGACAGATTATTCTCTCACCATCAGCAACCTGGAACCTGAAGATATTGCCACTTACTATTGTCAGCAGTATAGTAAGCTTCCTCC	1192059	1192345
NOD.IGK.Supercontig A	L-Part2	Leader	GTACCAGATGT	1192346	1192356
NOD.IGK.Supercontig A	L-Part1	Leader	ATGATGTCCTCTGCTCAGTTCCTTGGTCTCCTGTTGCTCTGTTTTCAAG	1192481	1192529
NOD.IGK.Supercontig A	1236495	RSS	CACAGTGATACAAATCATAACAAAAACC	1236468	1236495
NOD.IGK.Supercontig A	IGKV19.1	V-Gene	GACATCCAGATGACACAGTCTCCATCCTCACTGTCTGCATCTCTGGGAGGCAAAGTCACCATCACTTGCAAGGCAAGCCAAGACATTAACAAGAATATAGCTTGGTACCAACACAAGCCTGGAAAAGGTCCTAGGCTGCTCATATGGTACACATCTACATTACAGCCAGGCATCCCATCAAGGTTCAGTGGAAGTGGGTCTGGGAGAGATT	1236496	1236782

			ATTCCTTCAGCATCAGC AACCTGGAGCCTGAAG ATATTGCAACTTATTAC TGTCTACAGTATGATAA TCTTCCACC		
NOD.IGK. Supercont igA	L-Part2	Leader	CTCAGTGT	123678 3	1236790
NOD.IGK. Supercont igA	L-Part1	Leader	ATGAGACCCTCCATTCA GTTCCTGGGGCTCTTG TTGTTCTGGCTTCATG	123691 5	1236963
NOD.IGK. Supercont igA	1254300	RSS	CACAGTGCTACAGACT GGAACAAAATGC	125427 3	1254300
NOD.IGK. Supercont igA	IGKV4.1	V-Gene	GAAATGGTTCTCACCC AGTCTCCAGTATCCATA ACTGCATCTCGAGGGG AGAAGATCACCATCAC CTGCCGTGCCAGCTCA AGTATAAGTTCCAATTA CTTACACTGGTACCAG CAGAAGCCAGGATCCT CCCCTAAACTTTTGATT TATAGGACATCCATCCT GGCATCTGGAGTCCTA GACAACTTCAGTGGCA GTGGGTCTGAGAGCTC TTACACTCTGACAATCA GCTGCATGCAGGACGA TGTTGCTGCCACTTACT ACTGTCAGCAGGGGAG TAGTAGCCCACCA	125430 1	1254591
NOD.IGK. Supercont igA	L-Part2	Leader	TGTCAAAGAGA	125459 2	1254602
NOD.IGK. Supercont igA	L-Part1	Leader	ATGGATATGTGGGTGC AGATTTTCAGCTTACTG CTAATCTGTGTCACAG	125480 8	1254856
NOD.IGK. Supercont igA	1276225	RSS	CACCATGCTACAGACT AGAACAAGAACT	127619 8	1276225

NOD.IGK.Supercontig A	IGKV4.2	V-Gene	GAAATTGTGCTTACCCAGTCTCCAACCACCATGGCTGCATCTCCTGGGGAGAAGGTCACCATCACCTGCAGTGCCAGCTCAAGTATAAGTTCCAATTACTTGCACTGGTATCAGCAGAAGCCAGGATTCCCTCCTAAACTCTTGATATATAGGACATCCAATCTGGCTTCTGGAGTCCCAGCTCGCTTCAGTGGCAGTGGGTCTGGGACCTCTTACTCTCTCACAATTGGCACCATGGAGGCTGAAGGTGCTGCCACTTATTACTGCCAGCATGGTAGTAGTTTACTACGCA	1276226	1276517
NOD.IGK.Supercontig A	L-Part2	Leader	TGTCTAATGGA	1276518	1276528
NOD.IGK.Supercontig A	L-Part1	Leader	ATGGATTTTCAGATGCAGATTATCAGCTTGCTGCTAATCAGTGTCACAG	1276708	1276756
NOD.IGK.Supercontig A	1300219	RSS	CACAGTGATTCAAGCCATGACATAAATC	1300192	1300219
NOD.IGK.Supercontig A	IGKV12.2.P	V-Gene	GACATCCAGATGACTCAGTCTCCAGCTTCACTGTCTGCATCTGTGGGAGAAACTGTCACCATCACATGTGGAGCAAGTGAGAATATTTACGGTGCTTTAAATTGGTATCAGCAGAAACAGGGAAAATCTCCTCAGCTCCTGATCTATGGTGCAACCAACTTGGCAGATGGAATTCATCGAGGTTCAGTGGCAGTGGATCTGGTAGACAGTATTCTCTCAAGATCAGTAGCCTGCATCCTGACGATGTTGCAACGTATTACTGTCAAAATGTGTTAAGTACCCCTCC	1300220	1300505

NOD.IGK.Supercontig A	L-Part2	Leader	TCAGATGT	1300506	1300513
NOD.IGK.Supercontig A	L-Part1	Leader	ATGGGTGTACCCACTCAGCTCCTGGCGTTGCTGCTGCTGTGGCTTACCG	1300641	1300689
NOD.IGK.Supercontig A	1320029	RSS	CACAGTGATTTATACCCCAACAAAAACT	1320002	1320029
NOD.IGK.Supercontig A	IGKV1.11	V-Gene	GATGTTGTGGTGACTCAAACTCCACTCTCCCTGCCTGTCAGCTTTGGAGATCAAGTTTCTATCTCTTGCAGGTCTAGTCAGAGTCTTGCAAACAGTTATGGGAACACCTATTTGTCTTGGTACCTGCACAAGCCTGGCCAGTCTCCACAGCTCCTCATCTATGGGATTTCCAACAGATTTTCCGGGGTGCCAGACAGGTTCAGTGGCAGTGGTTCAGGGACAGATTTCACACTCAAGATCAGCACAATAAAGCCTGAGGACTTGGGAATGTATTACTGCTTACAAGGTACACATCAGCCTCC	1320030	1320331
NOD.IGK.Supercontig A	L-Part2	Leader	CTTCCAGAGGT	1320332	1320342
NOD.IGK.Supercontig A	L-Part1	Leader	ATGAATTTGCCTGTTCATCTCTTGGTGCTTCTGTTGTTCTGGATTCCTG	1320746	1320794
NOD.IGK.Supercontig A	1335504	RSS	CACAGTGATATAAGTCATAATATAAACC	1335477	1335504
NOD.IGK.Supercontig A	IGKV14.5	V-Gene	GACATCCAAATGACCCAGTCCTCATCATTTCTCTCTGCATCTCTAGGAGATCATCTTACAATCAACTGCAGGGCCAGTAAGGATATTAACAAGTATTTTGCTTGGGTTCAACAGAAGCCAGGGAAGGCTCC	1335505	1335791

			AAGGATGTTGATTCATT TTGCTTCCACCTTGCTA CCTGGGGTTCCAGAAA AGTTCAGTGGGAGTGG ATCTGGGACAGATTTTT CTCTCACTATCAGAAAC ATAGAGTCTGAAGATAT TGCAATGTATTACTGTC TACAGTATTCTGAGCAT CCTCC		
NOD.IGK.Supercontig A	L-Part2	Leader	GTGTATTGTGT	1335792	1335802
NOD.IGK.Supercontig A	L-Part1	Leader	ATGAGAATTCGTGGTC AGCTCTTGGGGCTACT GGTGCTCTGGATCACA G	1335927	1335975
NOD.IGK.Supercontig A	1365663	RSS	CACAGTGATACAGACT GGAACAAAAACT	1365636	1365663
NOD.IGK.Supercontig A	IGKV4.3	V-Gene	GAAATTGTGCTCACTCA GTCTCCAGCCATCACA GCTGCATCTCTGGGGC AAAAGGTCACCATCAC CTGCAGTGCCAGCTCA AGTGTAAGTTACATGCA CTGGTACCAGCAGAAG TCAGGCACCTCCCCCA AACCATGGATTTATGAA ATATCCAAACTGGCTTC TGGAGTCCCAGCTCGC TTCAGTGGCAGTGGGT CTGGGACCTCTTACTCT CTCACAATCAGCAGCA TGGAGGCTGAAGATGC TGCCATTTATTACTGCC AGCAGTGGAATTATCC ACTTA	1365664	1365946
NOD.IGK.Supercontig A	L-Part2	Leader	CCAGAGGA	1365947	1365954
NOD.IGK.Supercontig A	L-Part1	Leader	ATGGACTTTCGGGTGC AGATTTTCAGCTTCCTG CTAATCAGTGTCACAG	1366143	1366191

NOD.IGK.Supercontig A	L-Part1	Leader	ATGAAGTTTCCTTTTCAACTTCTGCTCTTCCTGCTGTTCGGAATCCCAG	1390080	1390128
NOD.IGK.Supercontig A	L-Part2	Leader	GCATGATATGT	1390248	1390258
NOD.IGK.Supercontig A	IGKV13.1	V-Gene	GACATCCAGATGACACAATCTTCATCCTCCTTGTCTGTATCTCTAGGAGACAGAGTCACCATTACTTGCAAGGCAAGTGAGCACATTAATAGTTGGTTAGCCTGGTATCAGCAAAAACCAGGAAATGCTCCTAGGCTCTTAATATCTGGTGCAACCAGTTTGGAAACTGGGGTTCCTTCAAGATTCAGTGGCAGTGCATCTGGAAAGGATTACACTCTCAGCATTACTAGTCTTCAGACTGAAGATGTTGCTACTTATTACTGTCAACAGTATTGGGGTACTCCTCC	1390259	1390545
NOD.IGK.Supercontig A	1390573	RSS	CACAGTGATACAAGCCATGACATAAACC	1390546	1390573
NOD.IGK.Supercontig A	1410559	RSS	CACAGTGATACAGACTAGAACAAAAACT	1410532	1410559
NOD.IGK.Supercontig A	IGKV4.1.P	V-Gene	GGAATTGTGCTCACCCAATCTCCAACAACCATGACTGCATTTCCAGGGGAAAATGTCACCATCACCTGCAGTGCCAGCTCAAGTATAAATTACATTCACTGGTACCAGCAGAAGTCAGGAACTACCCCCAAACAATGAATTTATAAGACATCCAACCTGCCTTCTGGAGTCCCAGCTCTCTTCAGTGGCAGTGGGTCTGGGACCTCTTACTCTCTCACAATCAGCAGTGTGGAGGCTGAAGATGCTGCCACTTATTACTGCCA	1410560	1410845

			GCAGTGGAGTGGTTAC TAACCCA		
NOD.IGK. Supercont igA	L-Part2	Leader	TATCCAGTGGA	141084 6	1410856
NOD.IGK. Supercont igA	L-Part1	Leader	GTGCAGATTTTCAGGTT CCTGCTAATCAGTGTCA CAG	141115 7	1411193
NOD.IGK. Supercont igA	L-Part1	Leader	ATGATGTCTCCCTCTCA GCTTCTGCTTATCTTAC TATTCTGATTCTAAG	142233 8	1422386
NOD.IGK. Supercont igA	L-Part2	Leader	GCCTAATATGT	142250 7	1422517
NOD.IGK. Supercont igA	IGKV12.2. P	V-Gene	GACATCCAAATGGCAT AATCTTCATCCTCCTTG TCTTCATCTCTAGGAGA CAGAGTCACCATTACTT GCAGGCCAGATGAAGG TATTAATGATTGGTGAT CCTGGTATCAGCAGAA ACCAGGGAATGTTCCT AAACTCCTAATATACCA TTGTACCAGTGTTGAAT CTGGAGTTTCTTCAAG GTTCAGTGGCAGTGAA TATGGGAAAGATTTTAC TCTTGCTGTTAGCAATC TGCAGCATGAAAATATT GCTACTTATTACTGTCA ACACTATTTAGTATACC TAC	142251 8	1422803
NOD.IGK. Supercont igA	1422831	RSS	CACAATGAAACAAGCC ATGACAAAAACT	142280 4	1422831
NOD.IGK. Supercont igA	1434717	RSS	CACAGTGATACAGACT GGAACAAAAACC	143469 0	1434717

NOD.IGK.Supercontig A	IGKV4.4	V-Gene	GAAAATGTGCTGACCCAGTCTCCAGCAATCATGGCTGCATCTCCAGGGGAGAAGGTCACCATGACCTGCAGTGCCAGCTCAAGTGTAAGTTCTGGTAACTTTCACTGGTACCAGCAGAAGCCAGGCACTTCTCCCAAACTCTGGATTTATAGGACATCCAACCTGGCTTCTGGAGTCCCCGCTCGCTTCAGTGGCAGTGGGTCTGGGACCTCTTACTCTCTTACAATCAGCAGCATGGAGGCCGAAGATGCTGCCACTTATTACTGCCAGCAGTGGAGTGGTTACCCACCCA	1434718	1435009
NOD.IGK.Supercontig A	L-Part2	Leader	TCTTAATGTCCAGAGGA	1435010	1435026
NOD.IGK.Supercontig A	L-Part1	Leader	ATGGATTTTCAGGTGCAGATTTTCAGCTTCCTGTTAATCAGTGTCTCAG	1435207	1435255
NOD.IGK.Supercontig A	1455190	RSS	CACAGTGATACAGACAAGAACAAAAACC	1455163	1455190
NOD.IGK.Supercontig A	IGKV4.5	V-Gene	CAAATTGTTCTCACCCAGTCTCCAGCAATCATGTCTGCATCTCTAGGGGAGGAGATCACCCTAACCTGCAGTGCCAGCTCGAGTGTAAGTTACATGCACTGGTTCCAGCAGAAGTCAGGCACTTCTCCCAAACTCTTGATTTATAGCACATCCAACCTGGCTTCTGGAGTCCCTTCTCGCTTCAGTGGCAGTGGGTCTGGGACCTTTTATTCTCTCACAATCAGCAGTGTGGAGGCTGAAGATGCTGCCGATTATTACTGCCATCAGTGGAGTAGTTATCCA	1455191	1455472

NOD.IGK.Supercontig A	L-Part2	Leader	TCATAATGTCCAGAGGA	1455473	1455489
NOD.IGK.Supercontig A	L-Part1	Leader	ATGGATTTTCAGGTGCAGATTTTCAGCTTCCTGCTAATCAGTGCCTCAG	1455671	1455719
NOD.IGK.Supercontig A	1434717	RSS	CACAGTGATACAGACTGGAACAAAAACC	1470631	1470658
NOD.IGK.Supercontig A	IGKV4.6	V-Gene	CAAATTGTTCTCACCCAGTCTCCAGCAATCATGTCTGCATCTCCAGGGCAGAAAGTCACCATAACCTGCAGTGCCAGCTCAAGTGTAAATTACATGCACTGGTACCAGCAGAAGCTAGGATCCTCCCCCAAACTCTGGATTTATGACACATCCAAACTGGCTCCTGGAGTCCCTGCTCGCTTCAGTGGCAGTGGGTCTGGGACCTCTTACTCTCTCACAATCAGCAGCATGGAGGCTGAAGATGCTGCCTCTTATTTCTGCCATCAGTGGAGTAGTTACCCACCCA	1470659	1470944
NOD.IGK.Supercontig A	L-Part2	Leader	TCATAATGTCCAGAGGA	1470945	1470961
NOD.IGK.Supercontig A	L-part1	Leader	ATGGATTTTCAGGTGCAGATTTTCAGCTTCCTGCTAATCAGTGCCTCAG	1471139	1471187
NOD.IGK.Supercontig A	1487909	RSS	CACAGTGCTACAGACTAGAACAAAAACC	1487882	1487909
NOD.IGK.Supercontig A	IGKV4.7	V-Gene	CAAATTGTTCTCACCCAGTCTCCAGCAATCATGTCTGCATCTCCTGGGGAACGGGTCACCATGACCTGCAGTGCCAGCTCAAGTGTAAGTTCCAGCTACTTGTACTGGTACCAGCAGAAGCCAGGATCCTCCCCAAACTATGGATTTATAGCACATCCAACCT	1487910	1488199

			GGCTTCTGGAGTCCCT GCTCGCTTCAGTGGCA GTGGGTCTGGGACCTC TTATTCTCTCACAATCA GCAGCATGGAGGCTGA AGATGCTGCCACTTATT ACTGCCAGCAGTACAG TGGTTACCATCCA		
NOD.IGK. Supercont igA	L-Part2	Leader	TCATAATGTCCAGAGG A	148820 0	1488216
NOD.IGK. Supercont igA	L-Part1	Leader	ATGGATTTTCAGGTGCA GATTTTCAGCTTCCTGC TAATCAGTGCCTCAG	148839 1	1488439
NOD.IGK. Supercont igA	L-Part1	Leader	ATGATGTCTCCCTCTCA ACATCTGCTGATCTTAC TATTCTGATTCCTAG	150774 9	1507797
NOD.IGK. Supercont igA	L-Part2	Leader	GCCTGATATGT	150791 8	1507928
NOD.IGK. Supercont igA	IGKV13.2. P	V-Gene	GACATCCAGATGACAC AATCTCCATCCTCCTTG TCTGTATCTCTAGGAGA CAGAGTTACCATTTCTT GCAGGCCAGATGTGAG TATTAATGATTGGTGAC CCTGATTTCAGTAGAAA CCAGGGAATGCTCATA AACACCTAATATACCAT TCTACCAGTGTGGAAT CTGGAGTTCCTTCAAG GTTCAGTGGCAGTGAA TTTGGGAAAGATTTTAC TCTGTTAACAAACTGAA GTGTGAAAATATTGCTA CTTATTACTGTCAACAG TATTTTTTATAACTAC	150792 9	1508210
NOD.IGK. Supercont igA	1508238	RSS	CACAATGAAACAAGTCA TGACAAAAATT	150821 1	1508238
NOD.IGK. Supercont igA	1530322	RSS	CACAGTGATACAGACT AGAACAAAAACC	153029 5	1530322

NOD.IGK.Supercontig A	IGKV4.2.P	V-Gene	AAAAATGTGCTGACCCAGTCTCCGGCAATCATGGCTACATCTCCAGGGGAGAAGGTCACCATGACCTGCAGTGCTAGCTCAAATGTAAGTTCTGGTAAGTTGCACTGGTAAAAGCAGTAGTCAGGCACTTCTCCCAAACTCTGGATTTATAGCACATTCAACCTGGCTTCTGGAGTTCCAGCTCGCTTCAGTGGCAGTGGGTCTGGGACCTCTTACTCTCTCACAATCAGCAGCATGGAGGCTGAAGATGATGCCAATTATTACTGCCAGAAGTGGAGTGGTTACCCACCCA	1530323	1530614
NOD.IGK.Supercontig A	L-Part2	Leader	TCATAATATCCAGAGGG	1530615	1530631
NOD.IGK.Supercontig A	L-Part1	Leader	ATGGATTTTCAGGTGCAGATTTTTCAAATTCCTGTTAATCAGTGTCTCAG	1530815	1530864
NOD.IGK.Supercontig A	1568979	RSS	CACAATGATACAGACTGGAACAAAAACC	1568952	1568979
NOD.IGK.Supercontig A	IGKV4.8	V-Gene	CAGATTGTTCTCACCCAGTCTCCAGCAATCATGTCTGCATCTCCAGGGGAGAAGGTCACCATGACCTGCAGGGCCAGCTCAAGTGTAAGTTCCAGTTACTTGCACTGGTACCAGCAGAAGCCAGGATCTTCCCCCAAACTCTGGATTTATAGCACATCCAACCTGGCTTCAGGAGTCCCAGCTCGCTTCAGTGGCAGTGGGTCTGGGACCTCTTACTCTCTCACAATCAGCAGTGTGGAGGCTGAGGATGCTGCCACTTATTACTGCCAGCAGTATGATAGTTCCCCATCCA	1568980	1569271

NOD.IGK.Supercontig A	L-Part2	Leader	TCATAATGTCCAGAGGA	1569272	1569288
NOD.IGK.Supercontig A	L-Part1	Leader	ATGGATTCTCAAGTGCAGATTTTCAGCTTCCTTCTAATCAGTGCCTTAG	1569464	1569512
NOD.IGK.Supercontig A	1601602	RSS	CACAGTGGCACAAGCAATGACATAAACC	1601575	1601602
NOD.IGK.Supercontig A	IGKV12.3.P	V-Gene	GACATCCAGATGCCTCAGTGTCCAGCCACCCCTTTCTGAATCTCTGGGAGAAAGTGTCACCATCACATGTCAAGCAAGTGAGAATATTGACAATTATTTATCATGGTATCAGCAAAAACCAAGGAAATCTCCTCAGCCCCTGATCAATTATACAACCAGCTTGGCAGATGGGGTTCCATCAAGGTCTAGTGGCAGTGGATCAGGCACACAGTTTTCTCTCAAGATCAACAACTTGCAAACAGAAGATGTTGCAAGTTACTATTGTCAACATCATTATTGTACTCCTCC	1601603	1601890
NOD.IGK.Supercontig A	L-Part2	Leader	ATGCCAGATGT	1601891	1601901
NOD.IGK.Supercontig A	L-Part1	Leader	ATGAGAACTTCTTCACAAAAGATCACACCCTGTGCTGGAGTCAG	1602095	1602138
NOD.IGK.Supercontig A	1611629	RSS	CACAGTAGTACAAGCAATGACATAAACC	1611602	1611629
NOD.IGK.Supercontig A	IGKV12.4.P	V-Gene	GACATCCAGATAACTCAGTCTCCAGTCACCCATTTCTGCATCTCTGGGAGAAAGTGTCACCATCACATGTCAAGCAAGTGAGAATATTGACAATTATTTATCATGGTATCAGCAAAAACCAAGGAAATCTCCTCAGCCCCTGATCAATTATACAACCAGCTTGGCAG	1611630	1611917

			ATGGGGTTCCAAAAAG GTCTAGTGGCAGTGGA TCAGGCACACAGTTTTC TCTCAAGATCAACAGC CTACAACCAGAAGATG TTGCAAGTCATTACTGT CAACATCATTATAGTAC TCCTCC		
NOD.IGK.Supercontig A	L-Part2	Leader	TCTCAGATGACAGATGT	1611918	1611934
NOD.IGK.Supercontig A	1619546	RSS	CACAGGGATACAGACA AGAACAAGAACC	1619519	1619546
NOD.IGK.Supercontig A	IGKV4.3.P	V-Gene	GAAATTGTGCTCATTAA GTCTTCAACAATCATGG CTGCATCTCCAGGGGA GAAGGTCACTATCACC TGCAGTGTCAGCTCAA GTGTAAGTTCCAGCTA CTAGCACTGGTACCAG CAGAAGTCAGGAGCCT CCCCCAAACTCTGGAT TTATGGCTCATTCAACC TGGCTTCTGAAGTTCCA GCTCACTTCAGTGGTA GTGGGTCTGGGACCTC TTGCTCTCTCACAATCA GCAGCATGGAAGCTGA AGGTGCTGCCACTTATT ACTGCCAGCAGGTTAC TTGTTCCCTACCCA	1619547	1619838
NOD.IGK.Supercontig A	L-Part2	Leader	TCATAGGGTCCAGTGG A	1619839	1619855
NOD.IGK.Supercontig A	L-Part1	Leader	ATGAATTTTCATGTGCA GATTTTCAGCTTCATGC TTATCAGTGTCACAG	1620031	1620079
NOD.IGK.Supercontig A	L-part2	Leader	GCCTGATATGT	1640107	1640117
NOD.IGK.Supercontig A	1654732	RSS	CACAGTGATACAGACT GGAACAAAAAAC	1654705	1654732

NOD.IGK.Supercontig A	IGKV4.9	V-Gene	GAAAATGTTCTCACCCAGTCTCCAGCAATCATGTCTGCATCTCCAGGGGAGAAGGTCACCATGACCTTCAGTGCCAGCTCAAGTGTAAGTTACATGCACTGGTACCAGCAGAAGTCAGGCACCTCCCCAAAACTCTGGATTTATGACACATCCAAACTGGCTTCTGGAGTCCCAGGTATCTTCAGTGGCAGTGGGTCTGGAAACTCTTACTCTCTCACGATCAGCAGCATGGAGGCTGAAGATGTTGCCACTTATTACTGCTTTCAGGGGAGTGGGTACCCACTCA	1654733	1655018
NOD.IGK.Supercontig A	L-Part2	Leader	TCATAATGTCCAGAGGA	1655019	1655035
NOD.IGK.Supercontig A	L-Part1	Leader	ATGGATTTTCAGGTGCAGATTTTCAGCTTCCTGCTAATCAGTGCCTCAG	1655213	1655261
NOD.IGK.Supercontig A	1434717	RSS	CACAGTGATACAGACTGGAACAAAAACC	1678470	1678497
NOD.IGK.Supercontig A	IGKV4.4.P	V-Gene	CAAATTGTTCTCACCCAGTCTCCAGCAATCATGTCTGCATCTCCAGGGGAGAAGGTCACCATGACCTGCAGTGCCAGCTCAAGTGTAAGTTACATGTACTGGTACCAGCAGAAGCCATGATCCTCCCCCAGACTCTGGATTTATGACACATCCAACCTGGCTTCTGGAGTCCCTGCTCGCTTCAGTGGCAGTGTGTCTGGGACGTCTTATTCTCTCACAATCAGCAGCATGGAGGCTGAAGATGCTGCCACTTATTACTGCCAGCAGTGGAGTAGTAACCAACCCA	1678498	1678783

NOD.IGK.Supercontig A	L-Part2	Leader	TCATACTGTCCAGAGGA	1678784	1678800
NOD.IGK.Supercontig A	L-Part1	Leader	ATGGATTTTCAAGTGCAGATTTTCAGCTTCCTGCTAATCAGTGCCTCAG	1678976	1679024
NOD.IGK.Supercontig A	1434717	RSS	CACAGTGATACAGACTGGAACAAAAACC	1694362	1694389
NOD.IGK.Supercontig A	IGKV4.10	V-Gene	CAAATTGTTCTCACCCAGTCTCCAGCAATCATGTCTGCCTCTCCAGGGGAGAAGGTCACCATGACCTGCAGTGCCAGCTCAAGTGTAAGTTCCAGGTACTTGCACTGGTACCAGCAGAAGTCAGGAGCCTCCCCCAAACTCTGGATTTATGGCACATCCAACCTGGCTTCTGGAGTCCCTGCTCGCTTCAGTGGCAGTGGGTCTGGGACCTCTTACTCTCTCACAATCAGCAGCGTGGAGGCTGAAGATGCTGCCACTTATTACTGCCAGCAGTATCATAGTGACCCACCCA	1694390	1694681
NOD.IGK.Supercontig A	L-Part2	Leader	TCATAATGACCAGAGGA	1694682	1694698
NOD.IGK.Supercontig A	L-Part1	Leader	ATGGATTTTCAAGTGCAGATTTTCAGCTTCTTGCTGATCAGTGCCTCAG	1694885	1694933
NOD.IGK.Supercontig A	1714166	RSS	CACAGTGGTACAAGCAATGACATAAATC	1714139	1714166
NOD.IGK.Supercontig A	IGKV12.5.P	V-Gene	GACATCCAGATGACTCAGTCTCCAGCCTCCCCTATATGCATCTCTGGGAGAAAGTGTCACCATCACATGTCAAGCAAGTGAGAATATTGACAATTATTTATCATGGTATCAGCAAAAACCAAGGAAATCTCCTCAGCCCCTGATCAATTATACAACCAGCTTGG	1714167	1714447

			CAGATGGGGTTCCATC AAGGTCTAGTGGCAGT GGATCAGGCACACAGT TTTCTCTCAAGATTAAC AACCTGCAAACAGAAG AAGTTATTACTGTCAAC ATCATTATAGTACTCCT CC		
NOD.IGK. Supercont igA	L-part2	Leader	ATGCCAGATGT	171444 8	1714458
NOD.IGK. Supercont igA	L-Part1	Leader	ATGAGTTTACCCACTCA ACTCCAGGGGTTGCTG CTGCTGTGGCTTACAG	171458 2	1714630
NOD.IGK. Supercont igA	1728210	RSS	CACAGGGATACAGACT AGAGCAAGAACC	172818 3	1728210
NOD.IGK. Supercont igA	IGKV4.11	V-Gene	GAAATTGTGCTCATTCA GTCTTCAACAATCATGG CTGCATCTCCAGGGGA AAAGGTCACCATGACC TGCAGTGCGAGCTCAA GTGTAAGTTCCAGCTA CTTGCACTTGTACCAG CAGAAGTCAGGAGCCT TCCCCAAATTCTGGATT TATGGCATATCCAACCT GGCTTCTGGAGTCCCT GTTCGCTTCAATGGCA GTGGGTCTGGGACCTC TTACTCTCTCACAATCA GCAGCATGGAGGCTGA AGATGCTGCCTCTTATT ACTGCCAGCAGGTTAC TAGTTCCCTACCCA	172821 1	1728502
NOD.IGK. Supercont igA	L-Part2	Leader	TCATAGGGTCCAGTGG A	172850 3	1728519
NOD.IGK. Supercont igA	L-Part1	Leader	ATGAATTTTCATGTGCA GATTTTCAGCTTCATGC TTATCAGTGTCACAG	172869 5	1728743
NOD.IGK. Supercont igA	L-Part2	Leader	GCTTTATATGT	175693 1	1756941

NOD.IGK.Supercontig A	IGKV13.3.P	V-Gene	GACATCCAGATGACAC AATCTCCATCCTCCTTG TCTGTATCTCTAGGAGA CAGAGTCACCATTTCTT GCAGGCCAGATGAGGG TATTAATGATTGGTGAG CCTGATTTCAGTAGAAA CCAGGGAATGTTCCTA AACTCCTAATATACCAT TCTATAAGTGTGAAATC TGGAGTTCCTTCAAGG TTCAGTGGCAGTGAATT TGGGAAAGATTTTACTC TTACTGTTAACAAACTG CAGTGTGAAAATATTGC TACTTATTACTGACAAC AGTATTTTTGTATAACT AC	1756942	1757228
NOD.IGK.Supercontig A	1422831	RSS	CACAATGAAACAAGCC ATGACAAAAACT	1757229	1757256
NOD.IGK.Supercontig A	1365663	RSS	CACAGTGATACAGACT GGAACAAAAACT	1778002	1778029
NOD.IGK.Supercontig A	IGKV4.12	V-Gene	CAAATTGTTCTCACCCA GTCTCCAGCAATCATGT CTGCATCTCCAGGGGA GAAGGTCACCATGACC TGCAGTGCCAGCTCAA GTGTAAGTTACATGTAC TGGTACCAGCAGAAGC CAGGATCCTCCCCCAG ACTCTGGATTTATGACA CATCCAACCTGGCTTCT GGAGTCCCCGCTCGCT TCAGTGGCAGTAGGTC TGGGACCTCTTATTCTC TCACAATCAGCAGCAT GGAGGCTGAAGATGCT GCCACTTATTACTGCCA TCAGCGGAGTAGTTAC CCA	1778030	1778311
NOD.IGK.Supercontig A	L-Part2	Leader	TCATACTGTCCAGAGG A	1778312	1778328

NOD.IGK.Supercontig A	L-Part1	Leader	ATGGATTTTCAAGTGCAGATTTTCAGCTTCCTGCTAATCAGTGCCTCAG	1778505	1778553
NOD.IGK.Supercontig A	1530322	RSS	CACAGTGATACAGACTAGAACAAAAACC	1791915	1791942
NOD.IGK.Supercontig A	IGKV4.13	V-Gene	CAAATTGTTCTCACCCAGTCTCCAGCAATCATGTCTGCATCTCCAGGGGAGAAGGTCACCATGACATGCAGTGCCAGCTCAAGTGTAAGTTACATGCACTGGTACCAGCAGAAGTCAGGCACCTCCCCCAAACCATGGATTTATGAAATATCCAAACTGGCTTCTGGAGTCCCAGCTCGCTTCAGTGGCAGTGGGTCTGGGACCTCTTATTCTCTCACAATCTGCAGCATGGAGGCTGAAGATGCTGCCACTTATTACTGCTATCAGTGGAGTAGTTACCCA	1791943	1792224
NOD.IGK.Supercontig A	L-Part2	Leader	TCATACTGTCCAGAGGA	1792225	1792241
NOD.IGK.Supercontig A	L-Part1	Leader	ATGGATTTTCAAGTGCAGATTTTCAGCTTCCTGCTAATCAGTGCCTCAG	1792417	1792465
NOD.IGK.Supercontig A	1530322	RSS	CACAGTGATACAGACTAGAACAAAAACC	1809528	1809555
NOD.IGK.Supercontig A	IGKV4.14	V-Gene	CAAATTGTTCTCACCCAGTCTCCAGCAATCATGTCTGCATCTCCAGGGGAGAAGGTCACCATGACATGCAGTGCCAGCTCAAGTGTAAGTTACATGCACTGGTACCAGCAGAAGTCAGGCACCTCCCCCAAAAGATGGATTTATGACACATCCAAACTGGATTCTGGAGTCCCTGCTCGCTTCAGTGGCAGTGGGTCTGGGACCTCTTATTCTC	1809556	1809837

			TCACAATCAGCAGCAT GGAGGCTGAAGATGCT GCCACTTATTACTGCTA TCAGTGGAGTAGTTAC CCA		
NOD.IGK. Supercont igA	L-Part2	Leader	TCATAATGTCCAGAGG A	180983 8	1809854
NOD.IGK. Supercont igA	L-Part1	Leader	ATGGATTTTCAAGTGCA GATTTTCAGCTTCCTGA TAATCAGTGCCTCAG	181003 2	1810080
NOD.IGK. Supercont igA	1410559	RSS	CACAGTGATACAGACT AGAACAAAAACT	182709 9	1827126
NOD.IGK. Supercont igA	IGKV4.15	V-Gene	GAAATTGTGCTCACCC AGTCTCCAGCACTCAT GGCTGCATCTCCAGGG GAGAAGGTCACCATCA CCTGCAGTGTCAGCTC AAGTATAAGTTCCAGCA ACTTACACTGGTACCA GCAGAAGTCAGGAACC TCCCCCAAACCCTGGA TTTATGGCACATCCAAC CTTGCTTCTGGAGTCC CTGTTCGCTTCAGTGG CAGTGGATCTGGGACC TCTTATTCTCTCACAAT CAGCAGCATGGAGGCT GAAGATGCTGCCACTT ATTACTGTCAACAGTGG AGTAGTTACCCACCCA	182712 7	1827418
NOD.IGK. Supercont igA	L-Part2	Leader	TCATATTGTCCAGTGGA	182741 9	1827435
NOD.IGK. Supercont igA	L-Part1	Leader	ATGGATTTTCATGTGCA GATTTTCAGCTTCATGC TAATCAGTGTCACAG	182760 9	1827657
NOD.IGK. Supercont igA	1434717	RSS	CACAGTGATACAGACT GGAACAAAAACC	185670 9	1856736

NOD.IGK.Supercontig A	IGKV4.16	V-Gene	GAAATTGTGCTCACCCAGTCTCCAGCACTCATGGCTGCATCTCCAGGGGAGAAGGTCACCATGACCTGCAGTGCCAGCTCAAGTGTAGGTTCCAGTTACTTGCACTGGTACCAGCAGAAGTCAGGAGCCTCCCCCAAACTCTGGATTTACGGCACATCCAACCTGGCTTCTGGAGTCCCTGCTTGCTTCAGTGGCAGTGGGTCTGGGACCTCTTACTCTCTCACAATCAGCAGCGTGGAGGCTGAAGATGATGCAACTTATTACTGCCAGCAGGGGTGGGATTACCCACCCA	1856737	1857028
NOD.IGK.Supercontig A	L-Part2	Leader	TCGTAATGTCCAGAGGA	1857029	1857045
NOD.IGK.Supercontig A	L-Part1	Leader	ATGGATTTACAGGTGCAGATTATCAGCTTCCTACTAATCAGTGCCTCAG	1857221	1857269

Supplemental Table 7. BALB/cByJ genes present in BALB/cByJ-IGH congenic line.

mm9 Chromosome 12 Start Position	mm9 Chromosome 12 End Position	Gene Symbol	Gene Description
112278008	112351597	Rcor1	REST corepressor 1
112404758	112505359	Traf3	TNF receptor-associated factor 3 isoform a
112509321	112514637	Amn	protein amnionless precursor
112522972	112532199	AK039023	Mus musculus adult male diencephalon cDNA, RIKEN full-length enriched library, clone:9330195H18 product:unclassifiable, full insert sequence.
112531182	112615929	Cdc42bpb	serine/threonine-protein kinase MRCK beta
112531182	112533527	mKIAA1124	Mus musculus mRNA for mKIAA1124 protein.
112645019	112650288	A230065H16Rik	hypothetical protein LOC380787
112655640	112669394	1200009I06Rik	SEC6-like protein C14orf73 homolog
112680871	112693229	Tnfaip2	tumor necrosis factor alpha-induced protein 2
112688615	112737833	AK080484	Mus musculus 7 days neonate cerebellum cDNA, RIKEN full-length enriched library, clone:A730046I10 product:tumor necrosis factor, alpha-induced protein 2, full insert sequence.
112722819	112724034	Gm266	hypothetical protein LOC212539
112776311	112784964	Eif5	eukaryotic translation initiation factor 5

112779156	112779277	Snora28	Mus musculus small nucleolar RNA, H/ACA box 28 (Snora28), small nucleolar RNA.
112810531	112812613	AK012841	Mus musculus 10, 11 days embryo whole body cDNA, RIKEN full-length enriched library, clone:2810029C07 product:unclassifiable, full insert sequence.
112812720	112894438	Mark3	MAP/microtubule affinity-regulating kinase 3
112907565	112910549	Ckb	creatine kinase B-type
112916315	112922113	Trmt61a	tRNA (adenine-N(1)-)-methyltransferase catalytic
112947703	112951467	Bag5	BAG family molecular chaperone regulator 5
112951479	112972435	2810002N01Rik	apoptogenic 1 isoform 2
112997059	113044986	Klc1	kinesin light chain 1 isoform 1D
113041403	113052052	Xrcc3	DNA repair protein XRCC3
113052480	113066287	Zfyve21	zinc finger FYVE domain-containing protein 21
113066668	113146266	Ppp1r13b	apoptosis-stimulating of p53 protein 1
113180201	113182693	5033406O09Rik	Mus musculus 11 days pregnant adult female ovary and uterus cDNA, RIKEN full-length enriched library, clone:5033406O09 product:unclassifiable, full insert sequence.
113199586	113205188	2010107E04Rik	6.8 kDa mitochondrial proteolipid
113209769	113307065	Tdrd9	putative ATP-dependent RNA helicase TDRD9
113217533	113219066	BC048943	hypothetical protein LOC217874
113344893	113365784	Aspg	60 kDa lysophospholipase

113347448	113347533	Mir3073	Mus musculus microRNA 3073 (Mir3073), microRNA.
113369090	113369166	Mir203	Mus musculus microRNA 203 (Mir203), microRNA.
113384418	113419958	Kif26a	kinesin-like protein KIF26A
113629736	113644023	AK054515	Mus musculus 2 days pregnant adult female ovary cDNA, RIKEN full-length enriched library, clone:E330021B06 product:unclassifiable, full insert sequence.
113649233	113661409	A730018C14Rik	Mus musculus 7 days neonate cerebellum cDNA, RIKEN full-length enriched library, clone:A730018C14 product:hypothetical protein, full insert sequence.
113727658	113738138	A530016L24Rik	transmembrane protein C14orf180 homolog
113738394	113749371	Tmem179	transmembrane protein 179
113826994	113853768	Inf2	inverted formin-2
113858257	113879566	Adssl1	adenylosuccinate synthetase isozyme 1
113883038	113887363	Siva1	apoptosis regulatory protein Siva isoform 1
113892031	113912487	Akt1	RAC-alpha serine/threonine-protein kinase
113917050	113920958	Zbtb42	zinc finger and BTB domain-containing protein
113960384	113984802	AW555464	hypothetical protein LOC217882
113998865	114007197	Pld4	phospholipase D4
114010404	114013851	Ahnak2	SubName: Full=AHNAK nucleoprotein 2; SubName: Full=Putative uncharacterized protein;

114023208	114040868	Ahnak2	SubName: Full=Putative uncharacterized protein; Flags: Fragment;
114047185	114054456	BC022687	hypothetical protein LOC217887
114058445	114067600	Cdca4	cell division cycle-associated protein 4
114089086	114099127	Gpr132	probable G-protein coupled receptor 132
114146800	114167706	Jag2	protein jagged-2 precursor
114172943	114180329	Nudt14	uridine diphosphate glucose pyrophosphatase
114198072	114238832	Brf1	transcription factor IIIB 90 kDa subunit
114214773	114217151	Btbd6	BTB/POZ domain-containing protein 6 isoform 1
114252718	114312612	Pacs2	phosphofurin acidic cluster sorting protein 2
114312712	114327125	Tex22	testis expressed gene 22
114336488	114375417	Mta1	metastasis-associated protein MTA1
114378446	114383717	Crip2	cysteine-rich protein 2
114390222	114392090	Crip1	cysteine-rich protein 1
114394631	114403669	4930427A07Rik	hypothetical protein LOC104732
114424113	114427733	Tmem121	transmembrane protein 121
114493109	114666777	Igh-A (1g2)	RecName: Full=Ig alpha chain C region;
114498399	117248164	abParts	Parts of antibodies, mostly variable regions.
114509390	114511358	J00476	RecName: Full=Ig epsilon chain C region;
114525497	114667344	Igh	RecName: Full=Ig gamma-2A chain C region secreted form; AltName: Full=B allele;

114563456	114577593	Ighg1	RecName: Full=Ig gamma-1 chain C region, membrane-bound form;
114594433	114604628	AI324046	SubName: Full=MFLJ00385 protein; Flags: Fragment;
114727775	114730046	Adam6b	a disintegrin and metalloproteinase domain
114782118	114784625	Adam6a	a disintegrin and metalloprotease domain 6
115693838	115694294	Igh	Mus musculus immunoglobulin heavy chain complex, mRNA (cDNA clone MGC:68276 IMAGE:3489841), complete cds.
115802787	115803099	U39293	Mus musculus clone H246 monoclonal autoantibody heavy chain variable region mRNA, partial cds.
116960143	116960166	AB345949	AB345949
117285934	117299062	Zfp386	zinc finger protein 386 (Kruppel-like) isoform
117316195	117384734	Vipr2	vasoactive intestinal polypeptide receptor 2
117341264	117377121	AK080151	Mus musculus mRNA for Bmh1 transcript.
117445522	117501498	Wdr60	WD repeat-containing protein 60
117519694	117611571	Esyt2	extended synaptotagmin-2
117640419	117643634	D430020J02Rik	Mus musculus 13 days embryo lung cDNA, RIKEN full-length enriched library, clone:D430020J02 product:similar to ENVELOPE PROTEIN (FRAGMENT) [Friend spleen focus-forming virus], full insert sequence.
117643874	117702004	Ncapg2	condensin-2 complex subunit G2

117724192	118516640	Ptprn2	receptor-type tyrosine-protein phosphatase N2
117980520	117995690	AK076914	Mus musculus adult male testis cDNA, RIKEN full-length enriched library, clone:4930550J05 product:unclassifiable, full insert sequence.
118382995	118389742	Gm10421	Mus musculus adult male hypothalamus cDNA, RIKEN full-length enriched library, clone:A230083A21 product:hypothetical protein, full insert sequence.
118489289	118489358	Mir153	Mus musculus microRNA 153 (Mir153), microRNA.
118521065	118575495	AK039060	Mus musculus adult male hypothalamus cDNA, RIKEN full-length enriched library, clone:A230091C17 product:unclassifiable, full insert sequence.
118754951	118995451	Rapgef5	rap guanine nucleotide exchange factor 5
119082333	119117179	Cdca7l	cell division cycle-associated 7-like protein
119116454	119437516	Dnahc11	dynein, axonemal, heavy chain 11
119473405	119539913	Sp4	transcription factor Sp4 isoform 1
120084801	120091051	Sp8	transcription factor Sp8
120106296	120204894	Abcb5	ATP-binding cassette, sub-family B (MDR/TAP),
120401275	120476749	Itgb8	integrin beta-8
120681882	120705405	Macc1	metastasis associated in colon cancer 1

Supplemental Table 8. NOD/ShiLtJ genes present in NOD/ShiLtJ-IGK congenic line.

mm9 Chromosome 6 Start Position	mm9 Chromosome 6 End Position	Gene Symbol	Gene Description
38237393	38249259	Zc3hav1l	zinc finger CCCH-type antiviral protein 1-like
38249034	38253791	AK137583	Mus musculus adult male bone cDNA, RIKEN full-length enriched library, clone:9830147F01 product:unclassifiable, full insert sequence.
38260496	38304603	Zc3hav1	zinc finger CCCH-type antiviral protein 1
38331523	38377647	Ttc26	tetratricopeptide repeat protein 26
38383037	38383783	BC099561	Mus musculus adult male testis cDNA, RIKEN full-length enriched library, clone:1700021L22 product:unclassifiable, full insert sequence.
38383924	38462763	Ubn2	ubinuclein-2
38484860	38489449	1110001J03Rik	formation of mitochondrial complexes 1 homolog
38501443	38559470	Luc7l2	putative RNA-binding protein Luc7-like 2 isoform
38502518	38504037	AK015545	Mus musculus adult male testis cDNA, RIKEN full-length enriched library, clone:4930471C18 product:unclassifiable, full insert sequence.
38576659	38587239	Klrg2	killer cell lectin-like receptor subfamily G

38613068	38630864	Clec2l	C-type lectin domain family 2 member L
38647838	38826189	Hipk2	homeodomain-interacting protein kinase 2 isoform
38668833	38692699	AK076827	Mus musculus adult male testis cDNA, RIKEN full-length enriched library, clone:4930417C18 product:unclassifiable, full insert sequence.
38868984	39034578	Tbxas1	thromboxane-A synthase
39036410	39068348	Parp12	poly [ADP-ribose] polymerase 12
39068471	39098311	AK016417	Mus musculus adult male testis cDNA, RIKEN full-length enriched library, clone:4930599N23 product:hypothetical protein, full insert sequence.
39086618	39156772	Jhdm1d	lysine-specific demethylase 7
39284769	39327706	Slc37a3	sugar phosphate exchanger 3
39331426	39340378	Rab19	ras-related protein Rab-19
39347819	39370368	Mkrn1	E3 ubiquitin-protein ligase makorin-1
39370376	39371293	AnMKRN1	Mus musculus AnMKRN1 mRNA for hypothetical protein, complete cds.
39412376	39507833	Dennd2a	DENN domain-containing protein 2A
39477520	39494484	AK047163	Mus musculus 10 days neonate cerebellum cDNA, RIKEN full-length enriched library, clone:B930030E13 product:unclassifiable, full insert sequence.
39523874	39538768	Adck2	aarF domain containing kinase 2

39542581	39549470	Ndufb2	NADH dehydrogenase [ubiquinone] 1 beta
39553236	39675462	Braf	serine/threonine-protein kinase B-raf
39751806	39760935	Mrps33	28S ribosomal protein S33, mitochondrial isoform
40060251	40198352	Gm5567	hypothetical protein LOC434008
40275476	40346761	Agk	acylglycerol kinase, mitochondrial precursor
40357496	40386132	E330009J07Rik	protein LCHN
40392861	40416814	Wee2	wee1-like protein kinase 2
40421413	40428586	Ssbp1	single-stranded DNA-binding protein,
40441236	40442238	Tas2r137	taste receptor type 2 member 3
40443590	40444484	Tas2r108	taste receptor type 2 member 4
40464822	40469507	Prss37	probable inactive serine protease 37 precursor
40494034	40494976	Olfr461	olfactory receptor 461
40521386	40522331	Olfr460	olfactory receptor 460
40524896	40535804	Clec5a	C-type lectin domain family 5 member A isoform
40562313	40563309	Tas2r138	taste receptor type 2 member 38
40578829	40719122	Mgam	maltase-glucoamylase
40828792	40837493	Moxd2	DBH-like monooxygenase protein 2 precursor
40841227	40841886	AF479018	RecName: Full=T-cell receptor beta chain V region E1; Flags: Precursor;
40845260	40850386	BC048599	putative trypsin-X3 precursor
40870458	40890556	1700074P13Rik	trypsin X5

40914770	40918426	1810009J06Rik	RIKEN cDNA 1810009J06
40945820	40949478	Gm2663	trypsinogen 4
40980266	40985508	2210010C04Rik	trypsinogen 7
40997338	40997994	TCRB	Mus musculus T-cell receptor beta chain (TCRB) mRNA, partial cds.
40998296	40998823	TCRB	Mus musculus T-cell receptor beta chain (TCRB) mRNA, partial cds.
41009392	41009878	TCR BV10S1A2Dbeta1Jbeta2.1	Mus musculus T cell receptor beta chain (TCR BV10S1A2Dbeta1Jbeta2.1) mRNA, partial cds.
41012342	41013101	TCRB5-2	RecName: Full=T-cell receptor beta chain V region A20.2.25; Flags: Precursor;
41063548	41064066	TRB	Mus musculus clone TCRB12-1-1lung T-cell receptor beta chain mRNA, partial cds.
41066008	41066461	TCRB13-1-1	Mus musculus clone 14S11 T cell receptor beta V-D-J region mRNA, partial cds.
41068862	41071846	TCRBVbeta5.1/Jbeta1.5	SubName: Full=Beta-chain; Flags: Precursor; Fragment;
41080145	41080583	TCRB13-3-3	RecName: Full=T-cell receptor beta chain V region C5; Flags: Precursor; Fragment;
41085165	41085609	Tcrb	Mus musculus T cell receptor V13D2J2.5 beta chain mRNA, partial cds.
41089320	41483908	TCRB	Mus musculus clone 114908252830411G T-cell receptor beta chain (TCRB) mRNA, partial cds.
41091186	41091661	BV12S11J2S3	Mus musculus T-cell receptor beta chain VJ

			region (BV12S11J2S3) mRNA, partial cds.
41093745	41493500	Z12226	M.musculus rearranged T-cell receptor beta chain Vbeta8 repertoire (VDJ).
41101172	41102227	TCR-beta chain	RecName: Full=T-cell receptor beta chain V region CTL-F3; Flags: Precursor;
41113060	41113551	TCRB	Mus musculus T cell receptor Vb9 chain mRNA, partial cds.
41128744	41497103	TCR-beta chain	RecName: Full=T-cell receptor beta-1 chain C region;
41138277	41138976	TRB	Mus musculus clone MR3.3 T-cell receptor beta chain mRNA, partial cds.
41168089	41168557	U07657	Mouse T-cell beta-chain receptor variable region V-beta-17.
41174067	41493405	TCRB	Mus musculus clone 115255110649604F T-cell receptor beta chain (TCRB) mRNA, partial cds.
41177694	41488591	H4	RecName: Full=T-cell receptor beta-1 chain C region;
41221423	41221873	TRB	RecName: Full=T-cell receptor beta chain V region PHDS203; Flags: Precursor; Fragment;
41231394	41231984	TCRB	Mus musculus T-cell receptor beta chain (TCRB) mRNA, partial cds.
41252270	41255532	Try4	trypsin 4
41261230	41264709	Try5	trypsin 5
41304103	41307943	Try10	trypsin 10
41323757	41327612	Prss3	mesotrypsin

41342354	41347255	Gm5771	trypsinogen 12
41365305	41369592	Gm5409	Try10-like trypsinogen
41392212	41396096	Gm10334	mesotrypsin-like
41408928	41413785	Prss1	protease, serine, 1
41471774	41475078	Prss2	anionic trypsin-2 precursor
41483110	41483210	TCRB	Mus musculus T-cell receptor beta chain mRNA, partial sequence.
41483862	41487582	H4	Mus musculus T cell receptor beta chain, minor H4 antigen graft infiltrating CDR3 region B6.19 u2, mRNA, partial cds.
41492087	41493270	TCR-beta chain	Mus musculus TCR-beta chain mRNA for T cell receptor beta chain, complete cds.
41507696	41508364	Gm16809	SubName: Full=V-beta 14 segment; Flags: Precursor; Fragment;
41555480	41570506	Ephb6	ephrin type-B receptor 6
41570617	41586404	Trpv6	transient receptor potential cation channel
41602768	41630722	Trpv5	transient receptor potential cation channel
41634429	41635716	1700034O15Rik	hypothetical protein LOC76606 precursor
41636328	41654324	Kel	kell blood group glycoprotein homolog
41721351	41722296	Olfr459	olfactory receptor 459
41797547	41802061	Pip	prolactin-inducible protein homolog precursor
41810337	41814321	Sval2	seminal vesicle antigen-like 2
41901626	41906097	Sval1	seminal vesicle antigen-like 1

41918138	41923089	Sval3	seminal vesicle antigen-like 3
41988392	41992850	Sva	seminal vesicle antigen
42090934	42091894	Tas2r139	taste receptor type 2 member 39
42165326	42166286	Tas2r144	taste receptor type 2 member 40
42195933	42200440	Gstk1	glutathione S-transferase kappa 1
42211968	42214554	Tmem139	transmembrane protein 139
42215037	42232495	Casp2	caspase-2
42236683	42264655	Clcn1	chloride channel protein 1
42265310	42274639	Fam131b	hypothetical protein LOC76156 isoform a
42278711	42300861	AK081414	Mus musculus 16 days embryo head cDNA, RIKEN full-length enriched library, clone:C130015E19 product:unclassifiable, full insert sequence.
42299826	42308394	Zyx	zyxin
42308485	42323267	Epha1	ephrin type-A receptor 1 precursor
42320669	42330554	AK008560	Mus musculus adult male small intestine cDNA, RIKEN full-length enriched library, clone:2010310C07 product:unclassifiable, full insert sequence.
42350236	42351118	Tas2r143	taste receptor type 2 member 143
42355527	42356493	Tas2r135	taste receptor type 2 member 135
42384533	42385460	Tas2r126	taste receptor type 2 member 41
42410074	42411016	Olfr458	olfactory receptor 458
42421233	42422175	Olfr457	olfactory receptor 457

42436161	42437213	Olfr456	olfactory receptor 456
42488065	42489019	Olfr455	olfactory receptor 455, pseudogene 1
42537211	42547371	Fam115e	experimental autoimmune prostatitis antigen 2
42573041	42595040	Fam115c	hypothetical protein LOC232748
42622545	42643058	Fam115a	hypothetical protein LOC77574
42694037	42694991	Olfr453	olfactory receptor 453
42712052	42713006	Olfr38	olfactory receptor 38
42740039	42740993	Olfr452	olfactory receptor 452
42767471	42768404	Olfr450	olfactory receptor 450
42787881	42788817	Olfr449	olfactory receptor 449
42846451	42847384	Olfr448	olfactory receptor 448
42861523	42862456	Olfr447	olfactory receptor 447
42877231	42878158	Olfr446	olfactory receptor 446
42905498	42906431	Olfr444	olfactory receptor 444
43017306	43045581	Olfr446	Mus musculus olfactory receptor 446, mRNA (cDNA clone IMAGE:30251255), complete cds.
43031513	43031707	IgK	Mus musculus 375.3 monoclonal anti-DNA IgM kappa-chain variable region mRNA, partial cds.
43031714	43031818	IgK	Mus musculus 375.3 monoclonal anti-DNA IgM kappa-chain variable region mRNA, partial cds.
43065742	43066675	Olfr441	olfactory receptor 441
43103305	43104238	Olfr237-ps1	olfactory receptor 438
43112452	43122315	AK006308	Mus musculus adult male testis cDNA, RIKEN full-length enriched library, clone:1700024N05

			product:unclassifiable, full insert sequence.
43117058	43117991	Olfr437	olfactory receptor 437
43123986	43124919	Olfr13	olfactory receptor 13
43151644	43152586	Olfr435	olfactory receptor 435
43166913	43167879	Olfr434	olfactory receptor 434
43185608	43186574	Olfr47	olfactory receptor 47
43215642	43239319	Arhgef5	Rho guanine nucleotide exchange factor (GEF) 5
43253672	43259553	Nobox	homeobox protein NOBOX
43295005	43616174	Tpk1	thiamin pyrophosphokinase 1
45010059	47251370	Cntnap2	contactin-associated protein-like 2 isoform a
47092053	47095304	AK018905	Mus musculus adult male testis cDNA, RIKEN full-length enriched library, clone:1700083J10 product:unclassifiable, full insert sequence.
47110742	47114232	AK018905	Mus musculus adult male testis cDNA, RIKEN full-length enriched library, clone:1700083J10 product:unclassifiable, full insert sequence.
47381782	47382132	AK211659	Mus musculus cDNA, clone:Y2G0122H09, strand:plus, reference:ENSEMBL:Mouse-Transcript-ENST:ENSMUST00000023813, based on BLAT search.
47404322	47476138	Cul1	cullin-1
47480272	47545029	Ezh2	histone-lysine N-methyltransferase EZH2 isoform

47604919	47702924	Rn4.5s	Mus musculus 0 day neonate thymus cDNA, RIKEN full-length enriched library, clone:A430093I18 product:4.5S RNA, full insert sequence.
47746139	47763511	Pdia4	protein disulfide-isomerase A4
47753574	47753651	Mir704	Mus musculus microRNA 704 (Mir704), microRNA.
47769264	47780504	Zfp786	zinc finger protein 786
47785659	47818256	Zfp398	zinc finger protein 398 isoform 1
47827553	47858483	Zfp282	zinc finger protein 282
47870566	47882636	Zfp212	Zinc finger protein 212 isoform 1
47893173	47898145	Zfp783	SubName: Full=Putative uncharacterized protein ENSMUSP00000110236;
47903388	47915298	Zfp956	hypothetical protein LOC101197
47974186	47998113	Zfp777	zinc finger protein 777
48012393	48036592	Zfp746	zinc finger protein 746
48080904	48345407	AK158926	Mus musculus 7 days neonate cerebellum cDNA, RIKEN full-length enriched library, clone:A730086M21 product:unclassifiable, full insert sequence.
48345584	48369854	Krba1	protein KRBA1
48363334	48369854	Krba1	protein KRBA1
48386611	48395824	Zfp467	zinc finger protein 467 isoform a
48414787	48451249	Sspo	SCO-spondin
48454337	48484831	Zfp862	Mus musculus 10 days neonate skin cDNA, RIKEN full-length enriched library, clone:4732460K03

			product:unclassifiable, full insert sequence.
48487567	48491799	Atp6v0e2	V-type proton ATPase subunit e 2
48504797	48520721	Lrrc61	leucine-rich repeat-containing protein 61
48519696	48522669	Rarres2	retinoic acid receptor responder protein 2
48539443	48540583	Gm5111	hypothetical protein LOC330305
48543881	48549081	Repin1	replication initiator 1
48563176	48573226	Zfp775	zinc finger protein 775
48578165	48583687	AI854703	Mus musculus adult male hypothalamus cDNA, RIKEN full-length enriched library, clone:A230078L20 product:weakly similar to ZNF74=KRUPPEL-TYPE ZINC FINGER (22 KRUPPEL-RELATED ZINC FINGER PROTEIN) (FRAGMENT) [Homo sapiens], full insert sequence.
48597359	48610874	Gimap8	GTPase IMAP family member 8
48626133	48628703	Gimap9	GTPase, IMAP family member 9
48634576	48642061	Gimap4	GTPase IMAP family member 4 isoform a
48651581	48658243	Gimap6	GTPase IMAP family member 6
48668619	48674635	Gimap7	GTPase, IMAP family member 7
48689045	48693794	Gimap1	GTPase IMAP family member 1
48696195	48704199	Gimap5	GTPase IMAP family member 5

48714462	48720850	Gimap3	GTPase IMAP family member 3
48783810	48791373	Tmem176b	transmembrane protein 176B
48791654	48795512	Tmem176a	transmembrane protein 176A
48810327	48816082	Gm7932	SubName: Full=Putative uncharacterized protein;
48845253	48859186	Abp1	amiloride-sensitive amine oxidase
48879894	48883686	1600015I10Rik	diamine oxidase-like protein 1
48925141	48928744	Doxl2	diamine oxidase-like protein 2
48936859	48941723	Svs1	seminal vesicle-secreted protein I precursor
48986516	49008181	Gpnmb	transmembrane glycoprotein NMB precursor
49023793	49034716	2410003K15Rik	hypothetical protein LOC75593
49035216	49164953	Igf2bp3	insulin-like growth factor 2 mRNA-binding
49193919	49214051	Tra2a	transformer-2 protein homolog alpha
49269351	49291580	Ccdc126	coiled-coil domain-containing protein 126
49317737	49339903	D330028D13Rik	hypothetical protein LOC231946 isoform 1
49345602	49419501	Stk31	serine/threonine-protein kinase 31
49772727	49779504	Npy	neuropeptide Y precursor
50060239	50148597	Mpp6	MAGUK p55 subfamily member 6 isoform a
50157401	50211916	Dfna5	non-syndromic hearing impairment protein 5

50243325	50332836	Osbpl3	oxysterol-binding protein-related protein 3
50512561	50516473	Cycs	cytochrome c, somatic
50516641	50544864	5430402O13Rik	Mus musculus 6 days neonate head cDNA, RIKEN full-length enriched library, clone:5430402O13 product:unclassifiable, full insert sequence.
50523302	50546589	4921507P07Rik	hypothetical protein LOC70821
50600869	50604392	Npvf	FMRFamide-related peptides precursor
50726113	50764893	AK043004	Mus musculus 7 days neonate cerebellum cDNA, RIKEN full-length enriched library, clone:A730047D12 product:unclassifiable, full insert sequence.
50798985	50828078	AK145307	Mus musculus mammary gland RCB-0527 Jyg-MC(B) cDNA, RIKEN full-length enriched library, clone:G930045G22 product:hypothetical protein, full insert sequence.
51219810	51219909	Mir148a	Mus musculus microRNA 148a (Mir148a), microRNA.
51350964	51352218	AK002748	Mus musculus adult male kidney cDNA, RIKEN full-length enriched library, clone:0610033M10 product:unclassifiable, full insert sequence.
51382668	51408767	Nfe2l3	nuclear factor erythroid 2-related factor 3
51410433	51419893	Hnrnpa2b1	heterogeneous nuclear ribonucleoproteins A2/B1
51420614	51433703	Cbx3	chromobox protein homolog 3

51473901	51540669	Snx10	sorting nexin-10
51809163	51962548	Skap2	src kinase-associated phosphoprotein 2
52105365	52108316	Hoxa1	homeobox protein Hox-A1
52112509	52114830	Hoxa2	homeobox protein Hox-A2
52115672	52119450	5730446D14Rik	SubName: Full=5730446D14Rik protein; SubName: Full=MCG121171;
52119060	52163066	Hoxa3	homeobox protein Hox-A3
52122879	52124051	AK142386	Mus musculus 13 days embryo lung cDNA, RIKEN full-length enriched library, clone:D430032P18 product:unclassifiable, full insert sequence.
52125215	52143019	AK039882	Mus musculus 0 day neonate thymus cDNA, RIKEN full-length enriched library, clone:A430025B13 product:unclassifiable, full insert sequence.
52139685	52141702	Hoxa4	homeobox protein Hox-A4
52151122	52163596	2700086A05Rik	Mus musculus 11 days embryo whole body cDNA, RIKEN full-length enriched library, clone:2700086A05 product:unclassifiable, full insert sequence.
52151752	52154586	Hoxa5	homeobox protein Hox-A5
52156363	52158623	Hoxa6	homeobox protein Hox-A6
52165622	52168572	Hoxa7	homeobox protein Hox-A7
52174052	52177369	Hoxa9	homeobox protein Hox-A9
52180079	52180164	Mir196b	Mus musculus adult female vagina cDNA, RIKEN full-length enriched library, clone:9930038K12 product:unclassifiable, full insert sequence.

52181195	52184938	Hoxa10	homeobox protein Hox-A10 isoform a
52192104	52195766	Hoxa11	homeobox protein Hox-A11
52195983	52199768	Hoxa11as	Mus musculus 8 days embryo whole body cDNA, RIKEN full-length enriched library, clone:5730529O17 product:homeo box A11, opposite strand transcript, full insert sequence.
52195983	52199768	Hoxa11as	Mus musculus 8 days embryo whole body cDNA, RIKEN full-length enriched library, clone:5730529O17 product:homeo box A11, opposite strand transcript, full insert sequence.
52208851	52210874	Hoxa13	homeobox protein Hox-A13
52212777	52217597	AK033508	Mus musculus adult male colon cDNA, RIKEN full-length enriched library, clone:9030414G15 product:hypothetical protein, full insert sequence.
52258382	52264826	AK031498	Mus musculus 8 days embryo whole body cDNA, RIKEN full-length enriched library, clone:5730457N03 product:unclassifiable, full insert sequence.
52263491	52268372	Evx1	homeobox even-skipped homolog protein 1
52496223	52590294	Hibadh	3-hydroxyisobutyrate dehydrogenase,
52663722	52716773	Tax1bp1	tax1-binding protein 1 homolog
52718061	53018618	Jazf1	juxtaposed with another zinc finger protein 1

53171795	53176593	AK043311	Mus musculus 7 days neonate cerebellum cDNA, RIKEN full-length enriched library, clone:A730083G08 product:unclassifiable, full insert sequence.
53523367	53645826	Creb5	cyclic AMP-responsive element-binding protein 5
53765461	53770819	Tril	TLR4 interactor with leucine rich repeats
53768736	53818175	AK039090	Mus musculus adult male hypothalamus cDNA, RIKEN full-length enriched library, clone:A230094J03 product:unclassifiable, full insert sequence.
53823374	53928656	Cpvl	probable serine carboxypeptidase CPVL precursor
53928684	53968024	AK076597	Mus musculus adult male testis cDNA, RIKEN full-length enriched library, clone:4930590G02 product:unclassifiable, full insert sequence.
53989925	54251806	Chn2	beta-chimaerin isoform 2
54096880	54100933	AK020588	Mus musculus adult male urinary bladder cDNA, RIKEN full-length enriched library, clone:9530036M11 product:unclassifiable, full insert sequence.
54195735	54229070	AK080009	Mus musculus adult male aorta and vein cDNA, RIKEN full-length enriched library, clone:A530046H22 product:unclassifiable, full insert sequence.
54219674	54380215	9130019P16Rik	Mus musculus adult male cecum cDNA, RIKEN full-length enriched library, clone:9130019P16

			product:hypothetical protein, full insert sequence.
54277005	54280194	Prr15	proline-rich protein 15
54402876	54453762	Wipf3	WAS/WASL-interacting protein family member 3
54458809	54516376	Scrn1	secernin-1
54501512	54504443	AK029418	Mus musculus 0 day neonate head cDNA, RIKEN full-length enriched library, clone:4833432I19 product:unclassifiable, full insert sequence.
54527598	54543122	Fkbp14	peptidyl-prolyl cis-trans isomerase FKBP14
54545104	54595816	Plekha8	pleckstrin homology domain-containing family A
54631765	54650400	2410066E13Rik	hypothetical protein LOC68235
54766909	54840218	Znrf2	E3 ubiquitin-protein ligase ZNRF2
54873935	54921655	Nod1	nucleotide-binding oligomerization
54935088	54942861	Ggct	gamma-glutamylcyclotransferase
54987994	55029498	Gars	glycyl-tRNA synthetase
55040041	55082966	Crhr2	corticotropin-releasing factor receptor 2
55120620	55124984	Inmt	indolethylamine N-methyltransferase
55153376	55270216	Fam188b	hypothetical protein LOC330323 isoform 1
55286292	55298549	Aqp1	aquaporin-1
55326288	55338524	Ghrhr	growth hormone-releasing hormone receptor
55401973	55451449	Adcyap1r1	pituitary adenylate cyclase-activating

55627812	55631257	Neurod6	neurogenic differentiation factor 6
55634564	55644568	AK038589	Mus musculus adult male hypothalamus cDNA, RIKEN full-length enriched library, clone:A230048G09 product:unclassifiable, full insert sequence.
55787011	55928592	Ccdc129	coiled-coil domain-containing protein 129
55967508	55982682	Gsbs	G-substrate
56019797	56312386	Pde1c	calcium/calmodulin-dependent 3',5'-cyclic
56319698	56321936	AK076685	Mus musculus adult male testis cDNA, RIKEN full-length enriched library, clone:4930414C11 product:unclassifiable, full insert sequence.
56651056	56654693	Lsm5	U6 snRNA-associated Sm-like protein LSm5
56664898	56711905	Avl9	late secretory pathway protein AVL9 homolog
56727518	56747807	Kbtbd2	kelch repeat and BTB (POZ) domain containing 2
56782052	56829354	Fkbp9	FK506 binding protein 9
56832394	56873926	Nt5c3	cytosolic 5'-nucleotidase 3
56906506	56907400	Vmn1r4	vomeronasal 1 receptor 4
56935335	56936283	Vmn1r5	vomeronasal 1 receptor, C19
56952348	56953260	Vmn1r6	vomeronasal 1 receptor 6
56974331	56975267	Vmn1r7	vomeronasal 1 receptor 7
56985906	56987119	Vmn1r8	vomeronasal 1 receptor, C32
57020888	57021939	Vmn1r9	vomeronasal 1 receptor, C30
57063418	57064354	Vmn1r10	vomeronasal 1 receptor, C1

57087346	57088246	Vmn1r11	vomeronasal 1 receptor, C3
57108913	57109837	Vmn1r12	vomeronasal 1 receptor 12
57159851	57160754	Vmn1r13	vomeronasal 1 receptor, C5
57183432	57184344	Vmn1r14	vomeronasal 1 receptor, C7
57208142	57209042	Vmn1r15	vomeronasal 1 receptor, C6
57272717	57273629	Vmn1r16	vomeronasal 1 receptor, C29
57310460	57311372	Vmn1r17	vomeronasal 1 receptor 17
57339661	57340561	Vmn1r18	vomeronasal 1 receptor 18
57354457	57355384	Vmn1r19	vomeronasal 1 receptor 19
57381684	57382596	Vmn1r20	vomeronasal 1 receptor 20
57456495	57485420	Ppm1k	protein phosphatase 1K, mitochondrial precursor
57530985	57615130	Herc6	hect domain and RLD 5
57634732	57642072	Pigy	protein preY, mitochondrial precursor
57652448	57689443	Lancl2	lanC-like protein 2
57702257	57775119	Vopp1	EGFR-coamplified and overexpressed protein
57793557	57794451	Vmn1r21	vomeronasal 1 receptor 21
57850075	57850984	Vmn1r22	vomeronasal 1 receptor 22
57875876	57876785	Vmn1r23	vomeronasal 1 receptor 23
57905634	57906525	Vmn1r24	vomeronasal 1 receptor, C18
57928387	57929296	Vmn1r25	vomeronasal 1 receptor, C8
57958176	57959196	Vmn1r26	vomeronasal 1 receptor 26
58051720	58051773	DQ701874	DQ701874
58165099	58166011	Vmn1r27	vomeronasal 1 receptor 27
58215167	58216076	Vmn1r28	vomeronasal 1 receptor 28
58257290	58258202	Vmn1r29	vomeronasal 1 receptor 29
58384930	58385839	Vmn1r30	vomeronasal 1 receptor 30
58421960	58422872	Vmn1r31	vomeronasal 1 receptor 31

58534520	58536554	AK015257	Mus musculus adult male testis cDNA, RIKEN full-length enriched library, clone:4930430M16 product:unclassifiable, full insert sequence.
58546665	58642445	Abcg2	ATP-binding cassette sub-family G member 2
58783693	58870390	Herc3	hect domain and RLD 3
58855226	58857120	Nap1l5	nucleosome assembly protein 1-like 5
58883529	58974496	Fam13a	family with sequence similarity 13, member A1
59158863	59162027	Tigd2	tigger transposable element-derived protein 2
59302454	59376284	Gprin3	G protein-regulated inducer of neurite outgrowth
60346342	60353701	A530053G22Rik	Mus musculus adult male aorta and vein cDNA, RIKEN full-length enriched library, clone:A530001F19 product:unclassifiable, full insert sequence.
60681566	60778990	Snca	alpha-synuclein
60894310	60939372	Mmrn1	multimerin-1 isoform a
61125597	61130264	AK158619	Mus musculus 10 days neonate cortex cDNA, RIKEN full-length enriched library, clone:A830012A11 product:unclassifiable, full insert sequence.
61130318	62332857	Fam190a	hypothetical protein LOC232035
63206850	64616273	Grid2	glutamate receptor delta-2 subunit precursor
63274665	63278435	AK029641	Mus musculus adult male testis cDNA, RIKEN full-length enriched library, clone:4930419F23

			product:unclassifiable, full insert sequence.
64679139	64681229	Atoh1	protein atonal homolog 1
64992660	65066043	Smarcad1	SWI/SNF-related matrix-associated
65067286	65094724	Hpgds	hematopoietic prostaglandin D synthase
65331287	65408144	C130060K24Rik	hypothetical protein LOC243407
65540391	65584034	Tnip3	RIKEN cDNA 9030611K07
65621604	65656924	A930038C07Rik	fibronectin type-III domain-containing protein
65728955	65886371	Prdm5	PR domain zinc finger protein 5
65902564	65904008	4930544G11Rik	ras homolog gene family, member A
66485461	66490985	Mad2l1	mitotic spindle assembly checkpoint protein
66502176	66509702	Vmn1r32	vomeronasal 1 receptor 32
66561644	66562562	Vmn1r33	vomeronasal 1 receptor 33
66586816	66587746	Vmn1r34	vomeronasal 1 receptor 34
66628787	66629678	Vmn1r35	vomeronasal 1 receptor 35
66665965	66666883	Vmn1r36	vomeronasal 1 receptor 36
66681385	66682294	Vmn1r37	vomeronasal 1 receptor 37
66726215	66727124	Vmn1r38	vomeronasal 1 receptor 38
66754408	66755326	Vmn1r39	vomeronasal 1 receptor 39
66844637	66846515	AK087755	Mus musculus adult male testis cDNA, RIKEN full-length enriched library, clone:4930597O21 product:unclassifiable, full insert sequence.
66846390	66971355	Gng12	guanine nucleotide-binding protein
66985089	66987401	Gadd45a	growth arrest and DNA damage-inducible protein

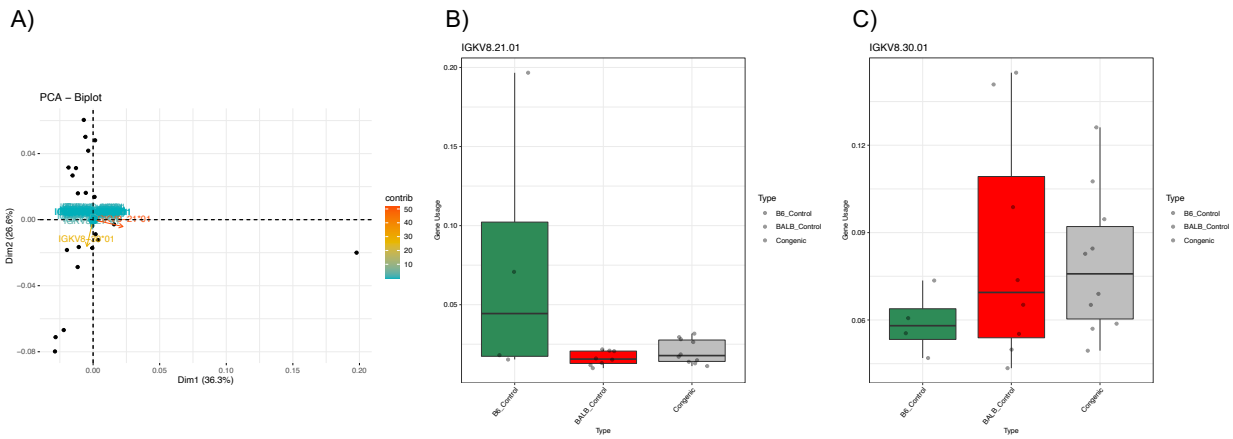
66986592	67030646	AK054076	Mus musculus 0 day neonate head cDNA, RIKEN full-length enriched library, clone:4833443G20 product:growth arrest and DNA-damage-inducible 45 alpha, full insert sequence.
67065666	67118420	AK039826	Mus musculus 0 day neonate thymus cDNA, RIKEN full-length enriched library, clone:A430010J10 product:unclassifiable, full insert sequence.
67216972	67239296	Serbp1	plasminogen activator inhibitor 1 RNA-binding
67242011	67326131	Il12rb2	interleukin-12 receptor subunit beta-2
67372925	67441849	Il23r	interleukin-23 receptor
67484052	67485816	Tacstd2	tumor-associated calcium signal transducer 2
67505629	70676960	abParts	Parts of antibodies, mostly variable regions.
67515227	67515693	4930515G16Rik	Mus musculus adult male testis cDNA, RIKEN full-length enriched library, clone:4930515G16 product:similar to 1810047J07RIK PROTEIN [Mus musculus], full insert sequence.
67892067	70408032	abParts	Parts of antibodies, mostly variable regions.
68169933	68170698	M19910	Mouse Ig rearranged kappa-chain mRNA, clone AN05K.
69276676	69277511	Rprl1	Mus musculus ribonuclease P RNA-like 1 (Rprl1), RNase P RNA.
70715713	70742169	Rpia	ribose-5-phosphate isomerase

70794520	70855234	Eif2ak3	eukaryotic translation initiation factor 2-alpha
70863082	70868916	1700011F03Rik	protein TSC21
70906599	70911060	Foxi3	forkhead box protein I3
71078391	71094370	Thnsl2	threonine synthase-like 2
71149881	71155017	Fabp1	fatty acid-binding protein, liver
71163933	71212275	Smyd1	SET and MYND domain-containing protein 1 isoform
71222012	71235313	Krcc1	lysine-rich coiled-coil protein 1
71243758	71252712	Gm1070	SubName: Full=Gm1070 protein; SubName: Full=Putative uncharacterized protein;
71272805	71287445	Cd8b1	T-cell surface glycoprotein CD8 beta chain
71307142	71310823	AK042210	Mus musculus 3 days neonate thymus cDNA, RIKEN full-length enriched library, clone:A630071F18 product:unclassifiable, full insert sequence.
71323420	71329165	Cd8a	T-cell surface glycoprotein CD8 alpha chain
71338627	71390631	Rmnd5a	protein RMD5 homolog A
71443887	71460875	Rnf103	E3 ubiquitin-protein ligase RNF103
71493847	71531568	Vps24	charged multivesicular body protein 3
71538965	71582680	Kdm3a	lysine-specific demethylase 3A
71657853	71760696	Reep1	receptor expression-enhancing protein 1
71762990	71773778	Mrpl35	39S ribosomal protein L35, mitochondrial
71781324	71825260	Immt	mitochondrial inner membrane protein

71830631	71858756	Ptcd3	pentatricopeptide repeat-containing protein 3,
71859046	71930261	Polr1a	DNA-directed RNA polymerase I subunit RPA1
72047606	72104564	St3gal5	lactosylceramide alpha-2,3-sialyltransferase
72070206	72072940	AK016225	Mus musculus adult male testis cDNA, RIKEN full-length enriched library, clone:4930565C19 product:unclassifiable, full insert sequence.
72112842	72117701	AK039589	Mus musculus adult male spinal cord cDNA, RIKEN full-length enriched library, clone:A330068C15 product:unclassifiable, full insert sequence.
72156171	72185571	Atoh8	protein atonal homolog 8
72253241	72264364	Sftpb	pulmonary surfactant-associated protein B
72268669	72295169	Usp39	U4/U6.U5 tri-snRNP-associated protein 2
72297022	72298452	BC054353	Mus musculus cDNA clone IMAGE:3962411, partial cds.
72297310	72303154	0610030E20Rik	hypothetical protein LOC68364
72305476	72309756	Tmem150a	transmembrane protein 150A precursor
72309707	72312375	Rnf181	E3 ubiquitin-protein ligase RNF181
72318042	72330462	Vamp5	vesicle-associated membrane protein 5
72335214	72340661	Vamp8	vesicle-associated membrane protein 8
72364326	72380701	Ggcx	vitamin K-dependent gamma-carboxylase

72382792	72389552	Mat2a	S-adenosylmethionine synthase isoform type-2
72463645	72470621	Sh2d6	SubName: Full=Putative uncharacterized protein Sh2d6;
72494432	72512974	Capg	macrophage-capping protein isoform 1
72515920	72548336	Elmod3	ELMO domain-containing protein 3
72548621	72557482	Retsat	all-trans-retinol 13,14-reductase precursor
72558414	72566994	Tgoln2	trans-Golgi network integral membrane protein 2
72576373	72738950	Tcf7l1	transcription factor 7-like 1 isoform 1
72586246	72587711	AK131839	Mus musculus 10 day old male pancreas cDNA, RIKEN full-length enriched library, clone:1810035C01 product:hypothetical protein, full insert sequence.
72649207	72655084	AK053994	Mus musculus 2 days pregnant adult female oviduct cDNA, RIKEN full-length enriched library, clone:E230011F03 product:unclassifiable, full insert sequence.
72791107	72849973	Kcmf1	E3 ubiquitin-protein ligase KCMF1
72907340	72908477	Tmsb10	thymosin beta-10
72967600	73171625	Dnahc6	axonemal dynein heavy chain
73198498	73226901	Suclg1	succinyl-CoA ligase [GDP-forming] subunit alpha,
73386958	73421015	AK006896	Mus musculus adult male testis cDNA, RIKEN full-length enriched library, clone:1700065L07

			product:unclassifiable, full insert sequence.
73418575	73419661	4931417E11Rik	hypothetical protein LOC66740
75246831	75246892	AK219219	Mus musculus cDNA, clone:Y2G0101J20, strand:minus, reference:ENSEMBL:Mouse-Transcript-ENST:ENSMUST00000015800, based on BLAT search.
76831630	77929661	Ctnna2	catenin alpha-2 isoform 2
77192710	77195511	Lrrtm1	leucine-rich repeat transmembrane neuronal



Supplemental Figure 3. Examination of potential IGKV outlier genes.

A) PCA biplot showing potential outlier genes. B and C) Gene usage plots of potential outlier genes indicated by PCA biplot.

CURRICULUM VITAE

Justin T. Kos

Department of Biochemistry and Molecular Genetics

University of Louisville School of Medicine

580 S. Preston Street, Louisville KY 40202

(314) 566-7690

justin.kos@louisville.edu

Education

2016 — present	**University of Louisville SOM**	Louisville, KY
	PhD, Biochemistry and Molecular Genetics	
	GPA: 3.85/4.0	
	Qualifying Exam Passed: 5/2018, PhD expected spring 2022	
2013 — 2015	**University of Missouri — St. Louis**	St. Louis, MO
	M.S., Cellular and Molecular Biology	
	GPA: 3.6/4.0	
2008 — 2012	**Saint Louis University**	St. Louis, MO
	B.S., Biology	

GPA: 3.5/4.0

Research Experience

2016 — present **University of Louisville SOM** Louisville, KY

Doctoral student; Dissertation Advisor: Corey T. Watson

PhD Dissertation: *Genomic Tools and Models for Investigating the Role of Germline Diversity in Mouse Antibody Repertoire Development*

i) Antibody light chain AIRR-seq of biomedically relevant mouse strains

ii) Developed new immunoglobulin genomic resources to study the effect of germline variation on the antibody repertoire

iii) Developed congenic mouse models to control immunoglobulin genetic variation

2014 — 2015 **University of Missouri — St. Louis** St. Louis, MO

Graduate Research; Advisor: Adam B. Smith, PhD

Project: *Systematic Assessment of Threats Affecting the Rare Plants of the United States*

i) Compiled, analyzed, and organized data on over 2000 threatened and endangered plant species, which revealed recreation was the most prominent threat

ii) Led primary literature analysis comparing the prevalence of threats to the distribution of research effort

2012 **Saint Louis University** St. Louis, MO

Undergraduate Research; Advisor: Blythe Janowiak, PhD

Project: *Microbial Diversity On A Heavily Fertilized Urban College Campus*

i) Conducted cultivation independent and dependent analyses utilizing Illumina 16S RNA sequencing and a variety of growth conditions, respectively

2010 — 2011 **Saint Louis University** St. Louis, MO

Undergraduate Research; Advisor: Robert Aldridge, PhD

Project: *Sexual Kidneys in the Red-Spotted Newt*

i) Performed histology and staining on kidney nephrons to elucidate structural changes during seasonal activity

Teaching Experience

2017 Teaching Assistant, Advanced Biochemistry, University of Louisville SOM

Led weekly review sessions on lecture material and graded exams.

2010 - 2012 Teaching Assistant, General Chemistry Lab, Saint Louis University

Conducted weekly discussion sessions with students on how to integrate lecture material into the laboratory, graded weekly lab reports, and led weekly preparation meetings with team for laboratory coordinator.

Publications

Jackson, K. J., **Kos, J. T.**, Lees, W., Gibson, W. S., Smith, M. L., Peres, A., Yaari, G., Corcoran, M., Busse, C. E., Ohlin, M., Watson, C. T. & Collins, A. M. (2022). A BALB/c IGHV Reference Set, defined by haplotype analysis of long-read VDJ-C sequences from F1 (BALB/c / C57BL/6) mice. BioRxiv, 2022.02.28.482396. https://doi.org/10.1101/2022.02.28.482396

Lee, J. H., Toy, L., **Kos, J. T.**, Safonova, Y., Schief, W. R., Havenar-Daughton, C., Watson, C. T., & Crotty, S. (2021). Vaccine genetics of IGHV1-2 VRC01-class broadly neutralizing antibody precursor naïve human B cells. *Npj Vaccines*, *6*(1), 113. https://doi.org/10.1038/s41541-021-00376-7

Zhou, M., Dascani, P., Ding, C., **Kos, J. T.**, Tieri, D., Lin, X., Caster, D., Powell, D., Wen, C., Watson, C. T., & Yan, J. (2021). Integrin CD11b Negatively

Regulates B Cell Receptor Signaling to Shape Humoral Response during Immunization and Autoimmunity. The Journal of Immunology, 207(7), 1785–1797. https://doi.org/10.4049/jimmunol.2100070

Waide, M. L., Polidoro, R., Powell, W. L., Denny, J. E., **Kos, J.**, Tieri, D. A., Watson, C. T., & Schmidt, N. W. (2020). Gut Microbiota Composition Modulates the Magnitude and Quality of Germinal Centers during Plasmodium Infections. Cell Reports, 33(11), 108503. https://doi.org/10.1016/j.celrep.2020.108503

Gadala-Maria, D., Gidoni, M., Marquez, S., Heiden, J. A. V., **Kos, J. T.**, Watson, C. T., O'Connor, K. C., Yaari, G., & Kleinstein, S. H. (2019). Identification of Subject-Specific Immunoglobulin Alleles From Expressed Repertoire Sequencing Data. Frontiers in Immunology, 10, 129. https://doi.org/10.3389/fimmu.2019.00129

Watson, C. T., **Kos, J. T.**, Gibson, W. S., Newman, L., Deikus, G., Busse, C. E., Smith, M. L., Jackson, K. J., & Collins, A. M. (2019). A comparison of immunoglobulin IGHV, IGHD and IGHJ genes in wild-derived and classical inbred mouse strains. Immunology and Cell Biology. https://doi.org/10.1111/imcb.12288

Hernández-Yáñez, H., **Kos, J. T.**, Bast, M. D., Griggs, J. L., Hage, P. A., Killian, A., Loza, M. I., Whitmore, M. B., & Smith, A. B. (2016). A systematic assessment of threats affecting the rare plants of the United States. Biological Conservation, 203, 260–267. https://doi.org/10.1016/j.biocon.2016.10.009

Research Conferences and Poster Presentations

JT Kos, OL Rodriguez, RG Gregg, CT Watson. Genomic Tools and Models for Investigating the Role of Germline Diversity in Mouse Antibody Repertoire Development, Research Louis, University of Louisville, October 2021

JT Kos, CT Watson. Genomic Tools and Models for Investigating the Role of Germline Diversity in Mouse Antibody Repertoire Development. AAI 2021, Virtual, May 2021

K Gilbert, **JT Kos**, WS Gibson, D Tieri, CE Busse, KJL Jackson, AM Collins, CT Watson. Examining Antibody Genomic Variation in Inbred Mouse Strains. Research Louisville, University of Louisville, October 2019

JT Kos, WS Gibson, KJL Jackson, L Newman, G Deikus, CE Busse, ML Smith, AM Collins, CT Watson. A Comparison of Immunoglobulin IGHV, IGHD, and IGHJ Genes in Wild-Derived and Classical Inbred Mouse Strains. Department of Biochemistry & Molecular Genetics Retreat, University of Louisville, August 2019

JT Kos, WS Gibson, KJL Jackson, L Newman, G Deikus, CE Busse, ML Smith, AM Collins, CT Watson. Antibody Repertoire Sequencing Uncovers Similarities and Differences Between Immunoglobulin Heavy Chain Variable Gene Loci of Inbred Mouse Strains. Immunogenomics Conference, Huntsville AL, October 2018

JT Kos, AM Collins, KJL Jackson, ML Smith, RG Gregg, CT Watson. Developing Genomic Tools and Models for Investigating the Role of Epi/Genetic Diversity In Mouse Antibody Repertoire Development. Department of Biochemistry & Molecular Genetics Retreat, University of Louisville, August 2017

Awards and Honors

2021	1st Place, Basic Doctoral Sciences, Research Louisville poster competition, University of Louisville SOM
2010 — 2012	Dean's list, Saint Louis University
2011	Saint Louis University Bright Ideas Grant Recipient
2010 — 2011	Bill Norman Community Service Award for Collegiate Honors Preparatory Program, Saint Louis University Leadership and Service Awards
2008 — 2012	Alpha Epsilon Delta National Honor Society
2008 — 2012	Saint Louis University Provost Scholarship

Service

2017 — 2020 Student Representative, University of Louisville SOM Graduate Council, Louisville KY

Engaged with University Directors of Graduate Studies to review faculty appointments, curriculum, and student concerns.

2020 Science Fair Judge, St. Francis of Assisi Catholic School, Louisville KY

2019 — 2020 Student President, Department of Biochemistry & Molecular Genetics, University of Louisville SOM, Louisville KY

Led recruitment events for prospective students and worked with Graduate Executive Committee to implement curriculum changes.

2010 — 2012 President, Collegiate Honors Preparatory Program, Saint Louis University, St. Louis MO

Led, planned, and started ACT test prep course for inner-city St. Louis youth. Created individualized lessons, 10-week curriculum, and individual short and long-term goals for students.

Professional Memberships and Other Experience

2021 — current	American Association of Immunologists
2021	Peer reviewer for Frontiers in Immunology
2016 — current	AIRR Community, Germline Database Working Group