Video Article

Identification of Mouse and Human Antibody Repertoires by Next-Generation Sequencing

Lin Sun[1], Naoko Kono[2], Hiroyuki Toh[3], Hanbing Xue[1], Kaori Sano[4,5], Tadaki Suzuki[4], Akira Ainai[4], Yasuko Orba[6], Junya Yamagishi[7,8], Hideki Hasegawa[4,5], Yoshimasa Takahashi[9], Shigeyuki Itamura[2], Kazuo Ohnishi[9,10]

[1]Graduate School of Life and Environmental Sciences, University of Tsukuba

[2]Center for Influenza Virus Research, National Institute of Infectious Diseases

[3]School of Science and Technology, Kwansei Gakuin University

[4]Department of Pathology, National Institute of Infectious Diseases

[5]Division of Infectious Diseases Pathology, Department of Global Infectious Diseases, Tohoku University Graduate School of Medicine

[6]Division of Molecular Pathobiology, Research Center for Zoonosis Control, Hokkaido University

[7]Division of Collaboration and Education, Research Center for Zoonosis Control, Hokkaido University

[8]Global Station for Zoonosis Control, GI-CoRE, Hokkaido University

[9]Department of Immunology, National Institute of Infectious Diseases

[10]Faculty of Life and Environmental Sciences, University of Tsukuba

Correspondence to: Kazuo Ohnishi at ohnishik@nih.go.jp

Abstract

The immense adaptability of antigen recognition by antibodies is the basis of the acquired immune system. Despite our understanding of the molecular mechanisms underlying the production of the vast repertoire of antibodies by the acquired immune systems, it has not yet been possible to arrive at a global view of a complete antibody repertoire. In particular, B cell repertoires have been regarded as a black box because of their astronomical number of antibody clones. However, next-generation sequencing technologies are enabling breakthroughs to increase our understanding of the B cell repertoire. In this report, we describe a simple and efficient method to visualize and analyze whole individual mouse and human antibody repertoires. From the immune organs, representatively from spleen in mice and peripheral blood mononuclear cells in humans, total RNA was prepared, reverse transcribed, and amplified using the 5'-RACE method. Using a universal forward primer and antisense primers for the antibody class-specific constant domains, antibody mRNAs were uniformly amplified in proportions reflecting their frequencies in the antibody populations. The amplicons were sequenced by next-generation sequencing (NGS), yielding more than 10^5 antibody sequences per immunological sample. We describe the protocols for antibody sequence analyses including V(D)J-gene-segment annotation, a bird's-eye view of the antibody repertoire, and our computational methods.

Video Link

The video component of this article can be found at https://www.jove.com/video/58804/

Introduction

The antibody system is one of the fundamentals of the acquired immune system. It is highly potent against invading pathogens due to its vast diversity, fine antigen recognition specificity, and the clonal expansion of antigen-specific B cells. The repertoire of antibody-producing B cells is estimated to be more than 10^{15} in a single individual[1]. This immense diversity is generated with the help of VDJ gene recombination in the immunoglobulin genetic loci[2]. Description of the entire B cell repertoires and their dynamic changes in response to antigen-immunization is therefore challenging, but essential for a complete understanding of the antibody response against invading pathogens.

Because of their astronomical diversity, B cell repertoires have been regarded as a black box; however, the advent of NGS technology has enabled breakthroughs to an enhanced understanding of their complexity[3,4]. Whole antibody repertoires have been successfully analyzed, firstly in zebrafish[5], then mice[6], and humans[6,7]. Although NGS has now become a powerful tool in the study of the adaptive immune response, basic analyses of the commonalities and differences in antibody repertoires among individual animals are lacking.

In mice, it was reported that the IgM repertoires are almost identical between individuals, whereas those of IgG1 and IgG2c are substantially different between individuals[8]. In addition to V-gene usage profile, the observed frequency of VDJ-profile in naive peripheral B cells is highly similar between individuals[8]. The analysis of the amino acid sequences of the VDJ-region also showed the occurrence of the same junctional

sequences in different mice much more frequently than previously thought[8]. These results indicate that the mechanisms for the antibody repertoire formation can be deterministic rather than stochastic[5,8,9]. The process of antibody repertoire development in mice has also been successfully analyzed using NGS to further highlight the potential of NGS to uncover the antibody immune system in detail[10].

In this report, we describe a simple and efficient method to visualize and analyze an antibody repertoire at a global level.

Protocol

All animal experiments were performed according to institutional guidelines and with the approval of the National Institute of Infectious Diseases Animal Care and Use Committee. Sampling of PBMCs from healthy adult volunteers, used as the representative result in this report, was performed with the approval of the Ethics Committee of the National Institute of Infectious Diseases, Tokyo, Japan, and written informed consent was obtained from each participant using an ethics committee-approved form.

1. Primer Design

1. Design a universal forward primer to cDNA to amplify the immunoglobulin mRNA without bias from PCR primers, as used in the 5'-RACE[11,12] and SMART-PCR[13] techniques.
2. For the immunoglobulin VH gene amplification, design the immunoglobulin class-specific sequences in the constant region as reverse primers[8,14] (**Figure 1A**).
   NOTE: Multiplex tag sequences can be added to any of these primers to label the library molecules from different sample sources. Sequences for nested PCR can also be added, according to the manual of the kit used[15].

Universal forward primer	5'- AAGCAGTGGTATCAACGCAGAGT-3'
Reverse primers for the mouse immunoglobulins (Ref.8)	
IgM_CH1:	5'- CACCAGATTCTTATCAGACAGGGGGCTCTC -3'
IgG1_CH1:	5'- CATCCCAGGGTCACCATGGAGTTAGTTTGG -3'
IgG2c_CH1:	5'- GTACCTCCACACACAGGGGCCAGTGGATAG -3'
IgG3_CH1:	5'-ATGTGTCACTGCAGCCAGGGACCAAGGGA-3'
IgA_CH1:	5'-GAATCAGGCAGCCGATTATCACGGGATCAC-3'
Igκ_CH1:	5'- GCTCACTGGATGGTGGGAAGATGGATACAG -3'
Igλ_CH1:	5'- CTBGAGCTCYTCAGRGGAAGGTGGAAACA -3'
Reverse primers for the human immunoglobulins (Ref.14)	
IgM_CH1:	5'- GGGAATTCTCACAGGAGACG -3'
IgG_CH1:	5'- AAGACCGATGGGCCCTTG -3'
IgD_CH1:	5'- GGGTGTCTGCACCCTGATA -3'
IgA_CH1:	5'- GAAGACCTTGGGGCTGGT -3'
IgE1_CH1:	5'- GAAGACGGATGGGCTCTGT -3'
IgE2_CH1:	5'- TTGCAGCAGCGGGTCAAGGG -3'
Igκ_CH1:	5'- TGCTCATCAGATGGCGGGAAGAT -3'
Igλ_CH1:	5'- AGAGGAGGGCGGGAACAGAGTGA -3'

Table 1: Primer sequences for PCR-amplification of immunoglobulins

2. Nucleic Acid Isolation from Immune Cells and Tissues

NOTE: The procedure given below is for extracting nucleic acids from the mouse spleen. However, it is applicable to other immune tissues and human cells such as lymph nodes or peripheral blood mononuclear cells (PBMCs) (**Figure 1B**).

1. Dissect the tissue, *e.g.*, spleen from an 8-week-old C57BL/6 mouse and pass it through a stainless-steel mesh (200 to 400 μm) with 2 mL of PBS buffer to obtain dispersed cells. Transfer the cell suspension to a 2.0 mL microcentrifuge tube, and centrifuge for 5 min at 600 × g and 4 ˚C. Discard the supernatant.
2. Add 800 μL of ACK lysing buffer (150 mM NH_4Cl, 1 mM $KHCO_3$, 0.1 mM Na_2EDTA, pH 7.2) to the pellet, and incubate on ice for 2 min to lyse red blood cells in the tissue.
3. Wash the tissue cells with 2 mL of PBS 3x, followed by centrifugation for 5 min at 600 × g and 4 ˚C.
4. Add 800 μL of phenol/guanidine isothiocyanate reagent to the pellet, vortex thoroughly, and incubate at about 25 ˚C for 5 min.
5. Add chloroform (200 μL), shake manually for 15 s, and then incubate for 2 min at about 25 ˚C.
6. Separate the phases by centrifugation for 15 min at 12,000 × g and 25 ˚C and transfer the upper aqueous phase to a fresh tube.
7. Add one volume of 70% ethanol, vortex briefly and apply it to the silica spin column.
8. Elute the RNA with 30–100 μL of water.

9. Quantitate initial RNA concentration using a fluorometer (**Table of Materials**).
10. Store the purified RNA at -80 ˚C.

3. cDNA Synthesis and PCR Amplification

NOTE: The method described below is based on the 5'-RACE[11,12] and SMART-PCR techniques[13]. The details and optimization of the reaction are described in the manual of the kit [15]. The starting materials for mouse immunoglobulin are the sample from step 2.10. The starting materials for human immunoglobulin are the sample from human tissues, ex. PBMC, treated as described in steps 2.3 to 2.10.

1. Synthesize the first-strand cDNA from 2 to 10 μg of total RNA template using 5'-RACE CDS primer (oligo-dT-containing) and SMART-PCR oligonucleotide (**Table of Materials**) according to the manufacturer's instructions[15].
    1. For the mouse immunoglobulin, PCR-amplify cDNA with high-fidelity DNA Polymerase using the universal forward primer and immunoglobulin class-specific reverse primers (**Table 1**). Set the thermal cycling conditions as: 94 ˚C for 2 min, then 40 cycles of 94 ˚C for 30 s, 59 ˚C for 30 s, and 72 ˚C for 30 s, followed by a final extension step at 72 ˚C for 5 min.
    NOTE: Typical experiments amplify IgM, IgG1, IgG2c, Igk and Igl immunoglobulin classes to look at the naive, Th1-dependent and Th2-dependent B cells (**Figure 3**).
    2. For the human immunoglobulin, perform the 1st PCR using the universal forward primer and immunoglobulin class-specific reverse primers (**Table 1**) with tag sequences. Include the index sequences for each sample by 2nd PCR using index sequence primers. Use the following PCR conditions and the Taq polymerase: 94 ˚C for 2 min, 21 cycles (1st PCR) or 32 cycles (2nd PCR) at 94 ˚C for 30 s, 59 ˚C for 30 s, 72 ˚C for 30 s.
    NOTE: Typical experiments amplify IgM, IgD, IgG (IgG1, IgG2, IgG3 and IgG4), IgA (IgA1 and IgA2), IgE, Igk and Igl immunoglobulin classes to look at all B cell populations (**Figure 4**).

2. Electrophorese the PCR products on an agarose gel and purify 600 to 800 bp fragments using a silica membrane spin-column.
    1. Electrophorese the sample from 3.2.1 or 3.2.2 on 2% agarose gel.
    2. Visualize the DNA bands on UV-transilluminator and excise the gel-slice containing the broad band between 600 to 800 bp.
    3. Add 10 μL of membrane binding solution per 10 mg of gel slice. Mix and incubate at 50–65 °C until the gel slice is completely dissolved.
    4. Transfer the gel solution on silica membrane spin-column. Wash once with washing buffer and elute DNA with 50 μL of nuclease-free water (**Table of Materials**).

3. Quantify the purified amplicons with a fluorometer and pool amplicons from each immunoglobulin class in equal amounts for NGS sequencing.
    NOTE: Typically, 2-10 μg amplicon DNA was recovered for each immunoglobulin class. Mix each sample solution equally in DNA amount to give rise 50 μL solution containing 10-20 ng DNA/μL.
4. Determine the size and concentration of libraries using a micro-capillary based electrophoresis with DNA sizing chip (**Table of Materials**). Store the libraries at - 20 °C.

4. NGS Sequencing of Libraries

1. Generate a SampleSheet.cvs for the sequencing run specifying sample name, index information and instruct to obtain .fastq files only.
2. Thaw the reagent cartridge (**Table of Materials**) and the libraries.
3. Make 0.2 N NaOH and dilute the libraries to obtain the desired molar concentration.
4. Rinse and dry the flow cell. Add 600 μL of diluted and denatured library solution into the well of the reagent cartridge.
5. Start the sequencing run.

5. Quality Control of NGS Data

1. Perform the quality control of FASTQ data using the "FASTX-Toolkit"[16].
    NOTE: A basic example of the parameter settings used is as follows:
    fastq_quality_trimmer -v -t 20 -l 200 -i [InFilename.fastq] -o [InFilename.fastq]
    fastq_quality_filter -v -q 20 -p 80 -i [InFilename.fastq] -o [InFilename.fastq]
    fastx_reverse_complement -v -i [InFilename.fastq] -o [InFilename.fastq]
2. Format the output files to "fasta nucleic acid (.fna)" by the following command:
    fastq_to_fasta -v -n -i [InFilename.fastq] -o [InFilename.fna]

6. Extraction and Analysis of Immunoglobulin Sequences from .fna Data

NOTE: The example programs were implemented in a UNIX environment. Please use them as an example references because performance may depend on the operating system and hardware environment. The authors do not accept any liability for errors or omissions. The programming languages, Perl[17], R[18], and required modules need to be installed according to the instructions on the cited websites. the IgBLAST program need to be installed according to the instructions on the appropriate website[19,20].

1. Download the following examples of in-house programs for repertoire analyses from https://github.com/KzPipeLine/KzPipeLine:
    03_PipeLine_Mouse.zip; A set of example programs for the analyses of mouse antibody sequences.
    05_PipeLine_Human.zip; A set of example programs for the analyses of human antibody sequences.
2. Extract the antibody reads in the sequence data: Extract the immunoglobulin (Ig) sequences of each Ig-class from the data (.fna) by a Perl program that searches the signature sequences in each immunoglobulin constant region (**Table 2**).

1. For the mouse immunoglobulin heavy chain (IgH) genes, extract the reads by the following command:
   $ perl 01_KzMFTIgCmgggaNtdVer3_Kz160607.pl [Input filename] [Output filename (suffix)]
2. For the mouse immunoglobulin light chain (IgL) genes extract the reads by the following command:
   $ perl 01_KzMFTCkltNtdVer1_170810.pl [Input filename] [Output filename (suffix)]
3. For the human immunoglobulin heavy chain (IgH) genes, extract the reads by the following command:
   $ perl 01_KzMfHuIgHCmgadeNtdVer1_Kz180312.pl [Input filename] [Output filename (suffix)]
4. For the human immunoglobulin light chain (IgL) genes, extract the reads by the following command:
   $ perl 01_KzMfHuIgCkltNtd_180316.pl [Input filename] [Output filename (suffix)]

3. Annotate and check the productivity of V(D)J gene recombination:
   NOTE: The method described below utilizes standalone IgBLAST[19] for the annotation of V(D)J gene segments in the sequence. Set the database for the V(D)J genes and the parameter settings for IgBLAST as described[20].
   1. Annotate the mouse immunoglobulin heavy chain (IgH) genes by the following command:
      $ igblastn -germline_db_V $IGDATA/ImtgMouseIghV_NtdDb.txt -germline_db_J $IGDATA/ImtgMouseIghJ_NtdDb.txt -germline_db_D $IGDATA/ImtgMouseIghD_NtdDb.txt -organism mouse -domain_system imgt -query ./$InFile -auxiliary_data $IGDATA/optional_file/mouse_gl.aux -show_translation -outfmt 7 >> ./$OutName
   2. Annotate the mouse immunoglobulin light chain (IgL) genes by the following command:
      $ igblastn -germline_db_V $IGDATA/ImtgMouseIgkV_NtdDb.txt -germline_db_J $IGDATA/ImtgMouseIgkJ_NtdDb.txt -germline_db_D $IGDATA/ImtgMouseIghD_NtdDb.txt -organism mouse -domain_system imgt -query ./$InFile -auxiliary_data $IGDATA/optional_file/mouse_gl.aux -show_translation -outfmt 7 >> ./$OutName
   3. Annotate the human immunoglobulin heavy chain (IgH) genes by the following command:
      $ igblastn -germline_db_V $IGDATA/ImtgHumanIghV_NtdDb.txt -germline_db_J $IGDATA/ImtgHumanIghJ_NtdDb.txt -germline_db_D $IGDATA/ImtgHumanIghD_NtdDb.txt -organism human -domain_system imgt -query ./$InFile -auxiliary_data $IGDATA/optional_file/Human_gl.aux -show_translation -outfmt 7 >> ./$OutName
   4. Annotate the human immunoglobulin light chain (IgL) genes by the following command:
      $ igblastn -germline_db_V $IGDATA/ImtgHumanIgkV_NtdDb.txt -germline_db_J $IGDATA/ImtgHumanIgkJ_NtdDb.txt -germline_db_D $IGDATA/ImtgHumanIghD_NtdDb.txt -organism human -domain_system imgt -query ./$InFile -auxiliary_data $IGDATA/optional_file/human_gl.aux -show_translation -outfmt 7 >> ./$OutName

4. Visualize the global feature of an antibody repertoire.
   1. Visualize the mouse IgH repertoire by the following command:
      $. 00a1_3DView_MoIgH_Kz180406.sh
      NOTE: The input file is filename.fna (sequence data), preferably the output file from 6.2.1. This file needs to be placed in a lower directory (folder) named "filename". In line 50 of the shell script, assign a "filename" for Para_4.
   2. Visualize the human IgH repertoire by the following command:
      $. 00a1_3DView_HuIgH_Kz180411.sh
      NOTE: The input file is filename.fna sequence data, preferably the output file of 6.2.3. This file needs to be placed in the lower directory (folder) that name is "filename". In line 46 of the shell script, assign a "filename" for Para_4.
   3. Visualize the mouse IgL repertoire by the following command:
      $. 00_2DViewS_MoIgL_Kz180406.sh
      NOTE: With this pipeline, Igk and Igl are processed concomitantly. The input file is filename.fna (sequence data), preferably the output file from 6.2.2. This file needs to be placed in the lower directory (folder) named "filename". In line 53 of the shell script, assign a "filename" for Para_4. The output file's name, ending with "_IgKlCount.txtDim2Rpm.txt" gives the coordinates for a two-dimensional bar graph (**Figure 3**, IgL).
   4. Visualize the human IgL repertoire by the following command:
      $. 00_2DView_HuIgL_Kz180319.sh
      NOTE: With this pipeline, Igk and Igl are processed concomitantly. The input file is filename.fna sequence data, preferably the output file of 6.2.4. This file needs to be placed in a lower directory (folder) named "filename". In line 53 of the shell script, assign "filename" for Para_4. The output file name ending with "_IgKlCount.txtDim2Rpm.txt" gives the coordinates for a two-dimensional bar graph (**Figure 4**, IgL).

Representative Results

Antibody repertoires of mouse

A perspective of a murine antibody repertoire as a whole can be obtained from cells or tissues such as the spleen, bone marrow, lymph node, or blood. **Figure 3** shows representative results of IgM, IgG1, IgG2c, and immunoglobulin light chain (IgL) repertoires from a naïve mouse spleen. The summary of the read numbers is shown in **Table 3**. For example, 166,175/475,144 reads contained IgM-specific signature sequence (**Table 2**) and 133,371/166,175 reads were VDJ-productive inferred by IgBLAST[19].

Figure 3 shows a repertoire profile of VDJ-rearrangement by 3D-VDJ-plot, in which the size of each ball represents the relative number of reads; in other words, the number of antibody mRNAs in whole B cells. The 3-D mesh consists of 110 IGHV, 12 IGHD, and 4 IGHJ, which are aligned to reflect their order on the chromosome. In addition, the genes ambiguously assigned by IgBLAST were collected separately in the last position for each IGHV, IGHD and IGHJ line, giving rise to 7,215 nodes in the cuboid.

Also, shown in **Figure 3** is a 2D-VJ-plot showing the profile of VJ-rearrangement in the IgL repertoire. The length of each bar on this plot represents the relative number of reads. The x-axis represents 101 IGLVκ and 3 IGLVλ genes, and the y-axis represents 4 IGLJκ and 3 IGLJλ genes. The unannotated V- and J-genes are represented on the right borderline.

The complementarity-determining region 3 (CDR3) sequences of these productive reads, which give rise to the majority of antigen-binding specificity, are given in IgBLAST outputs. The CDR3 sequences can be analyzed statistically, including biological or technical replicates, as described previously[8,10].

Human antibody repertoires

A perspective of a human antibody repertoire as a whole can be analyzed from various tissues including peripheral blood mononuclear cells (PBMCs) or pathological tissues. **Figure 4** shows representative results of IgM, total IgG (IgG1, IgG2, IgG3, and IgG4), total IgA (IgA1 and IgA2), IgD, IgE and IgL repertoires from normal PBMCs. A summary of the read numbers is shown in **Table 3**. For example, 90,238/1,582,754 reads contained IgM-specific signature sequence and 67,896/90,238 reads were VDJ-productive.

The repertoire profile of VDJ rearrangement is shown on a 3D-VDJ-plot in which the size of each ball represents the relative number of reads; in other words, the number of antibody mRNAs from whole PBMCs (**Figure 4**). The 3-D mesh consists of 56 IGHV, 27 IGHD, and 6 IGHJ, aligned in the order they appear on the chromosome. In addition, genes ambiguously assigned by IgBLAST are represented separately in the last position for each IGHV, IGHD and IGHJ line, giving rise to 11,172 nodes in the cuboid.

The profile of VJ-rearrangement in the IgL repertoire is depicted in a 2D-VJ-plot in which the length of each bar represents the relative number of reads (**Figure 4**). The x-axis represents 41 IGLVκ and 32 IGLVλ genes, and the y-axis represents 5 IGLJκ and 5 IGLJλ genes. The un-annotated V- and J-genes are represented on the right borderline.

The human CDR3 sequences are given in IgBLAST outputs and can be analyzed statistically as described previously[8,10].

Immunoglobulin class	Sense	Antisense
Mouse immunoglobulin heavy chains (C57BL/6)		
IgM	AGTCAGTCCTTCCCAAATGTC	GACATTTGGGAAGGACTGACT
IgG1	AAAACGACACCCCCATCTGTC	GACAGATGGGGGTGTCGTTTT
(IgG1 variant)	AAAACAACACCCCCATCAGTC	GACTGATGGGGGTGTTGTTTT
IgG2c	AAAACAACAGCCCCATCGGTC	GACCGATGGGGCTGTTGTTTT
IgG3	GTGATCCCGTGATAATCGGCT	AGCCGATTATCACGGGATCAC
IgA	TCCCTTGGTCCCTGGCTGCAG	TCCCTTGGTCCCTGGCTGCAG
Mouse immunoglobulin light chains (C57BL/6)		
Igκ	CTGTATCCATCTTCCCACCATCCAGTGAGC	GCTCACTGGATGGTGGGAAGATGGATACAG
Igλ1	TGTTTCCACCTTCCTCTGAAGAGCTCGAG	CTCGAGCTCTTCAGAGGAAGGTGGAAACA
Igλ2	TGTTTCCACCTTCCTCTGAGGAGCTCAAG	CTTGAGCTCCTCAGAGGAAGGTGGAAACA
Igλ3	TGTTTCCACCTTCCCCTGAGGAGCTCCAG	CTGGAGCTCCTCAGGGGAAGGTGGAAACA
Igλ4	TGTTCCCACCTTCCTCTGAAGAGCTCAAG	CTTGAGCTCTTCAGAGGAAGGTGGGAACA
Human immunoglobulin heavy chains		
IgM	GGGAGTGCATCCGCCCCAAC	GTTGGGGCGGATGCACTCCC
IgG	GCTTCCACCAAGGGCCCATC	GATGGGCCCTTGGTGGAAGC
IgA	GCATCCCCGACCAGCCCCAA	GACCGATGGGGCTGTTGTTTT
IgD	GCACCCACCAAGGCTCCGGA	TCCGGAGCCTTGGTGGGTGC
IgE	GCCTCCACACAGAGCCCATC	GATGGGCTCTGTGTGGAGGC
Human immunoglobulin light chains		
Igκ	ACTGTGGCTGCACCATCTGC	GCAGATGGTGCAGCCACAGT
Igλ1,2,6	GTCACTCTGTTCCCGCCCTC	GAGGGCGGGAACAGAGTGAC
Igλ3,7	GTCACTCTGTTCCCACCCTC	GAGGGTGGGAACAGAGTGAC

Table 2: Summary of the immunoglobulin signature sequences

Mouse IgH	Total reads	IgM	IgG1	IgG2c		
Input	475,144					
IgC-containing		166,175	229,671	36,628		
VDJ-productive		133,371	196,583	31,446		
Mouse IgL	Total reads	IgKappa	IgLambda			
Input	527,668					
IgC-containing		178,948	21,446			
VJ-productive		160,924	16,988			
Human IgH	Total reads	IgM	IgG	IgA	IgD	IgE
Input	1,582,754					
IgC-containing		90,238	5,298	94,061	75,549	2,932
VDJ-productive		67,896	2,775	78,203	56,495	3
Human IgL	Total reads	IgKappa	IgLambda			
Input	1,582,754					
IgC-containing		120,316	64,148			
VJ-productive		97,169	52,324			

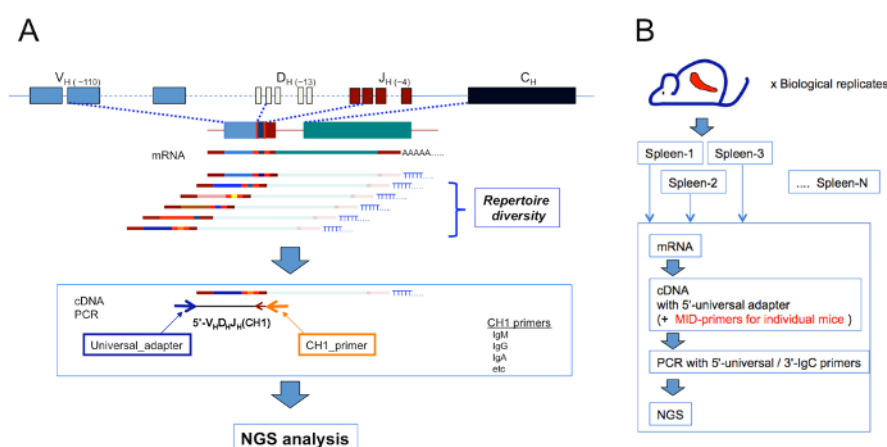**Table 3: Summary of the read numbers in the experiments**

Figure 1: Schematic representation of sequencing strategy for analyzing antibody repertoires in individual mice. (A) Total RNA from the immune cells or tissues was reverse-transcribed and PCR-amplified using the universal forward primer and immunoglobulin class-specific reverse primers. The amplicons from each immunoglobulin class were pooled and rendered for next-generation sequencing.**(B)** The biological replicates such as spleens from C57BL/6 mice were treated as follows: total RNAs were purified from spleen samples, and cDNAs were amplified by 5'-RACE using the universal primer and antibody class-specific primer. They were then rendered for next-generation sequencing with labeling primers for individual mice. Parts of the figure are adapted from[8] with permission. Please click here to view a larger version of this figure.
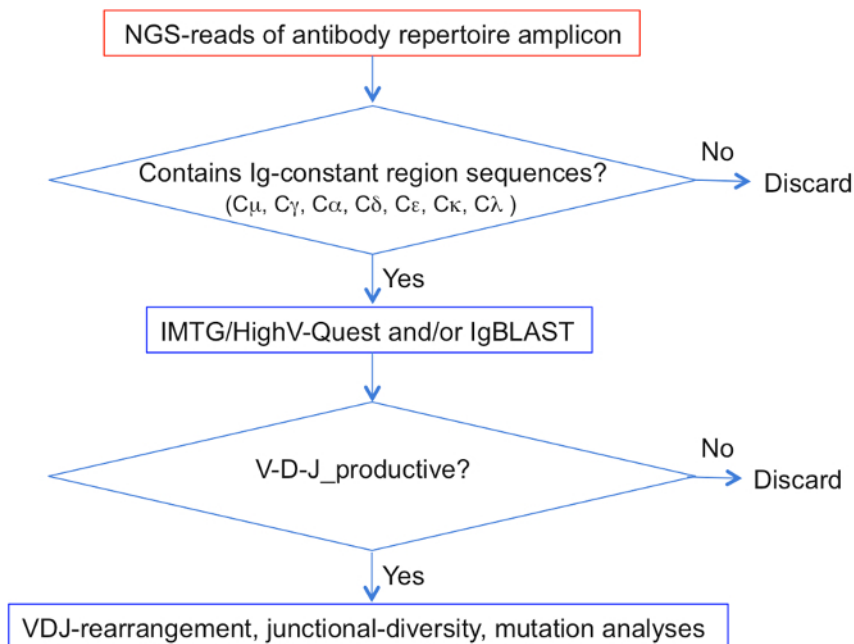
Figure 2: Schematic of data-processing flowchart for analyzing antibody repertoires in individual mice. Amplicon reads obtained after next-generation sequencing were processed as follows: (1) read sequences were checked for the presence of antibody class-specific signature sequences; (2) sequences were examined for the V, D, and J gene fragments using IMGT/HighV-Quest and/or IgBLAST; (3) the sequences containing a productive VDJ junction were collected; and (4) these sequences were used for the analysis of overall repertoire features, CDR3, etc. Please click here to view a larger version of this figure.
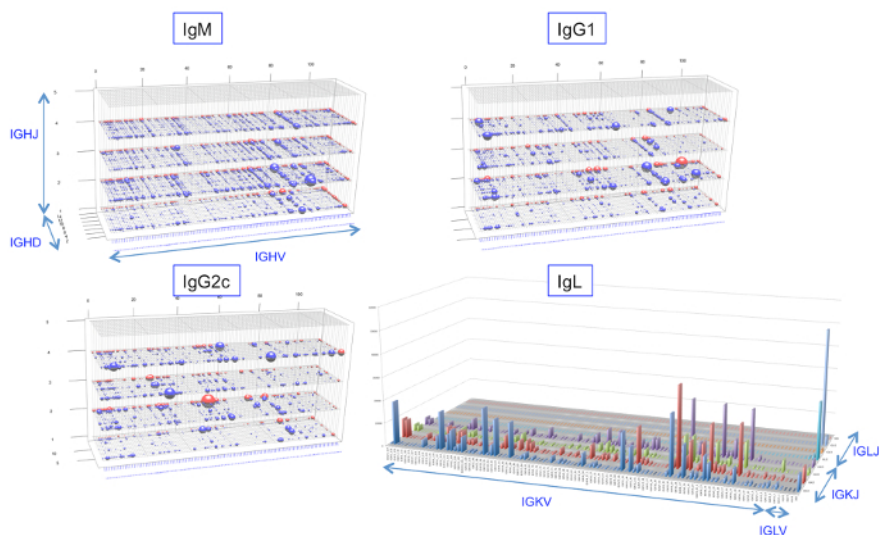

Figure 3: Global data visualization for mouse antibody repertoires. The overall repertoire profiles of each antibody class were visualized by 3D-VDJ-plot. The x-axis represents 110 IGHV genes ordered as on the chromosome. The y- and z-axis represents 12 IGHD and 4 IGHJ genes, respectively. The volume of spheres on each node represents the number of reads. Red spheres: un-annotated V, D, and J genes. The IgL read distributions are shown on a 2D-VJ-plot in which the length of each bar represents the relative number of reads. The x-axis represents 101 x IGLVκ and 3 x IGLVλ genes, and the y-axis represents 4 x IGLJκ and 3 x IGLJλ genes. The un-annotated V and J genes are represented on the right borderline. Please click here to view a larger version of this figure.
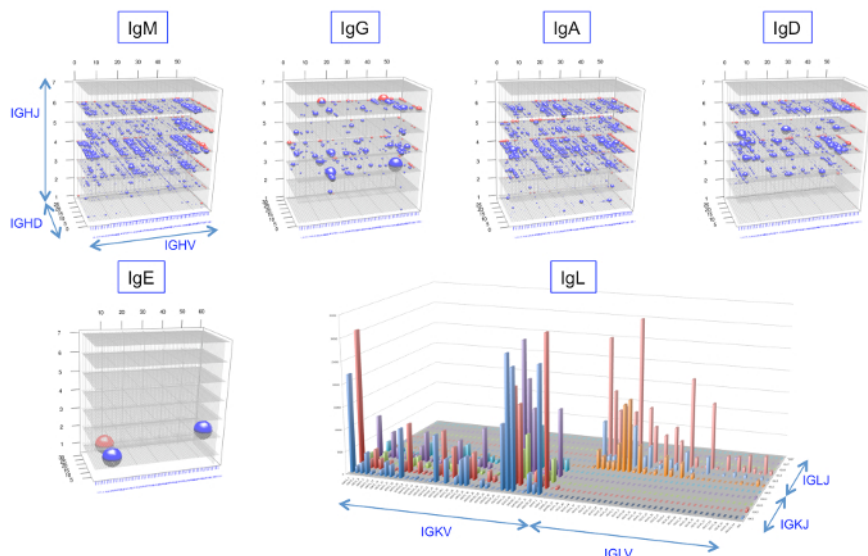
Figure 4: Global data visualization for human antibody repertoires. The overall repertoire profiles of each antibody class were visualized by 3D-VDJ-plot. The x-axis represents 56 IGHV genes ordered as on the chromosome. The y- and z-axis represents 27 IGHD and 6 IGHJ genes, respectively. The volume of spheres on each node represents the number of reads. Red spheres: un-annotated V, D, and J genes. The IgL reads are arrayed on the 2D-VJ-plot in which the length of each bar represents the relative number of the reads. The x-axis represents 41 x IGLVκ and 32 x IGLVλ genes, and the y-axis represents 5 x IGLJκ and 5 x IGLJλ genes. The un-annotated V- and J-genes are represented on the right borderline. Please click here to view a larger version of this figure.

Discussion

The method described here utilizes NGS for antibody RNA amplified using the 5'-RACE method. In contrast to methods that use degenerate 5'-V_H gene primers, mRNAs of each antibody class are amplified evenly using universal forward primers. In addition, the use of antisense primers specific for the constant-region 1 (CH1) of the antibody gene enables repertoire profiling of specific immunoglobulin classes. This is very beneficial for dissecting the class-specific antibody response, as well as for comparing naive and immunized repertoires[8,9].

A most likely pitfall of the method is a paucity of amplified immunoglobulin messages. The depth of antibody repertoire obtained by this protocol substantially depends on the PCR amplification described in steps 3.1 and 3.2. If the repertoire depth is not properly obtained, changing the ratios of template cDNA and primers in steps 3.2.1 or 3.2.2 is strongly recommended.

Generally, approximately 20% of the antibody reads produced by NGS are ambiguous sequences[21]. Even with established "correction methods", 5-10% remain ambiguous[3]. We, therefore, analyzed the sequence and filtered raw reads containing signature sequences corresponding to immunoglobulin constant regions (CμH1, Cγ1H1, Cγ2cH1, etc.). Hence the analysis of somatic hyper-mutations needs the careful examinations.

One of the limitations of this method is that immunoglobulin heavy and light chain pair is unable to be inferred. Hence the repertoire view obtained by this method is not holistic. However, it is possible to approximate the top-ranking pairs by statistical analysis of the data[10]. Also, a novel method to sequence the immunoglobulin pairs was reported recently[3,4].

The immunoglobulin sequences in the output .fna data were extracted based on the presence of immunoglobulin gene signature sequences. The V, D, and J gene segments were then annotated and the productivity of V(D)J rearrangements were assessed. The complementarity-determining region 3 (CDR3) sequences were also annotated. These systematic examinations of immunoglobulin sequences in .fna data were usefully provided by the IMGT/HighV-QUEST server[22,23,24]. However, building an automated processing pipeline has the merit to analyze the big experimental data. The pipeline customized for each purpose is possible to set up by using the standalone IgBLAST protocol[19]. This approach needs basic programming literacy but is very useful for detailed analyses of the immunoglobulin system. The pipelines described are the examples of the customized protocol (**Figure 2**).

The number of antibody reads is proportional to the amount of antibody RNAs in the sample, reflecting the antibody constituents of the antibody system at given time points[5,8,25]. The method described here gives a bird's eye view of the V(D)J constitution of an antibody repertoire using R programs[8,18,26].

The global view of IgM antibody repertoires of individual naive mice revealed a highly conserved VDJ-profile as compared to those of IgG1 or IgG2c[8]. It was reported that VDJ combinations of immature zebrafish are highly stereotyped[9]. In contrast, human VDJ combinations are reported to be highly skewed[6]. The highly conserved deterministic VDJ-profiles in naive B cells are probably generated either by skewed VDJ-rearrangements or negative selection with auto-antigens presented in the body. For example, IGHV11-2 is expressed preferentially in the fetal IgM repertoire[27] and this predominance is attributed to the autoreactivity of IGHV11-2 against senescent erythrocytes[27]. Interestingly, IGHV11-2 was also the most common major repertoire in our previously published analysis of naive IgM[8].

The method described here is useful for deciphering antigen-responsive antibody repertoires by inclusively analyzing the antibody-repertoire space generated in individual bodies, avoiding inadvertent omission of key antibody repertoires[8,10]. This method also allows the examination of detailed antibody network dynamism, which would facilitate accelerated discovery of protective antibodies against newly emerging pathogens.

Disclosures

The authors have no conflicts of interest to disclose.

Acknowledgments

References

1. Schroeder, H. W., Jr. Similarity and divergence in the development and expression of the mouse and human antibody repertoires. *Developmental & Comparative Immunology.* **30** (1-2), 119-135 (2006).
2. Tonegawa, S. Somatic generation of antibody diversity. *Nature.* **302** (5909), 575-581 (1983).
3. Georgiou, G. *et al.* The promise and challenge of high-throughput sequencing of the antibody repertoire. *Nature Biotechnology.* **32** (2), 158-168 (2014).
4. Lees, W. D., & Shepherd, A. J. Studying Antibody Repertoires with Next-Generation Sequencing. *Methods in Molecular Biology.* **1526,** 257-270 (2017).
5. Weinstein, J. A., Jiang, N., White, R. A., 3rd, Fisher, D. S., & Quake, S. R. High-throughput sequencing of the zebrafish antibody repertoire. *Science.* **324** (5928), 807-810 (2009).
6. Arnaout, R. *et al.* High-resolution description of antibody heavy-chain repertoires in humans. *PLoS One.* **6** (8), e22365 (2011).
7. Boyd, S. D. *et al.* Individual variation in the germline Ig gene repertoire inferred from variable region gene rearrangements. *Journal of Immunology.* **184** (12), 6986-6992 (2010).
8. Kono, N. *et al.* Deciphering antigen-responding antibody repertoires by using next-generation sequencing and confirming them through antibody-gene synthesis. *Biochemical and Biophysical Research Communications.* **487** (2), 300-306 (2017).
9. Jiang, N. *et al.* Determinism and stochasticity during maturation of the zebrafish antibody repertoire. *Proceedings of the National Academy of Sciences of the United States of America.* **108** (13), 5348-5353 (2011).
10. Sun, L. *et al.* Distorted antibody repertoire developed in the absence of pre-B cell receptor formation. *Biochemical and Biophysical Research Communications.* **495** (1), 1411-1417 (2018).
11. Olivarius, S., Plessy, C., & Carninci, P. High-throughput verification of transcriptional starting sites by Deep-RACE. *Biotechniques.* **46** (2), 130-132 (2009).
12. Yeku, O., & Frohman, M. A. Rapid amplification of cDNA ends (RACE). *Methods in Molecular Biology.* **703,** 107-122 (2011).
13. Zhu, Y. Y., Machleder, E. M., Chenchik, A., Li, R., & Siebert, P. D. Reverse transcriptase template switching: a SMART approach for full-length cDNA library construction. *Biotechniques.* **30** (4), 892-897 (2001).
14. Vollmers, C., Sit, R. V., Weinstein, J. A., Dekker, C. L., & Quake, S. R. Genetic measurement of memory B-cell recall using antibody repertoire sequencing. *Proceedings of the National Academy of Sciences of the United States of America.* **110** (33), 13463-13468 (2013).
15. *SMARTer RACE 5'/3' Kit User Manual (634858, 634859).* (2018).
16. *FASTX-Toolkit.* http://hannonlab.cshl.edu/fastx_toolkit/ (2018).
17. *Perl.* <https://perldoc.perl.org/> (2018).
18. *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. (2016).
19. Ye, J., Ma, N., Madden, T. L., & Ostell, J. M. IgBLAST: an immunoglobulin variable domain sequence analysis tool. *Nucleic Acids Research.* **41** (Web Server issue), W34-40 (2013).
20. *IgBLAST.*, <https://www.ncbi.nlm.nih.gov/igblast/faq.html - standalone> (2018).
21. Prabakaran, P., Streaker, E., Chen, W., & Dimitrov, D. S. 454 antibody sequencing - error characterization and correction. *BMC Research Notes.* **4,** 404 (2011).
22. Lefranc, M. P. *et al.* IMGT, the international ImMunoGeneTics information system. *Nucleic Acids Research.* **37** (Database issue), D1006-1012 (2009).
23. Alamyar, E., Duroux, P., Lefranc, M. P., & Giudicelli, V. IMGT((R)) tools for the nucleotide analysis of immunoglobulin (IG) and T cell receptor (TR) V-(D)-J repertoires, polymorphisms, and IG mutations: IMGT/V-QUEST and IMGT/HighV-QUEST for NGS. *Methods in Molecular Biology.* **882**, 569-604 (2012).
24. *IMGT/HighV-QUEST.*, <http://www.imgt.org/HighV-QUEST/login.action> (2018).
25. Glanville, J. *et al.* Precise determination of the diversity of a combinatorial antibody library gives insight into the human immunoglobulin repertoire. *Proceedings of the National Academy of Sciences of the United States of America.* **106** (48), 20216-20221 (2009).
26. *rgl: 3D Visualization Using OpenGL.* R package version 0.95.1247 (2015).
27. Hardy, R. R., Wei, C. J., & Hayakawa, K. Selection during development of VH11+ B cells: a model for natural autoantibody-producing CD5+ B cells. *Immunological Reviews.* **197**, 60-74 (2004).