Chapter 11

NGS-Based B-Cell Receptor Repertoire AnalysisRepertoire analyses in the Context of Inborn Errors of Immunity

Pauline A. van Schouwenburg (ORCID)**, Mirjam van der Burg, and Hanna IJspeert**

Abstract

Inborn errors of immunity (IEI) are genetic defects that can affect both the innate and the adaptive immune system. Patients with IEI usually present with recurrent infections, but many also suffer from immune dysregulation, autoimmunity, and malignancies.

Inborn errors of the immune system can cause defects in the development and selection of the B-cell receptor (BCR) repertoire. Patients with IEI can have a defect in one of the key processes of immune repertoire formation like V(D)J recombination, somatic hypermutation (SHM), class switch recombination (CSR), or (pre-)BCR signalling and proliferation. However, also other genetic defects can lead to quantitative and qualitative differences in the immune repertoire.

In this chapter, we will give an overview of protocols that can be used to study the immune repertoire in patients with IEI, provide considerations to take into account before setting up experiments, and discuss analysis of the immune repertoire data using Antigen Receptor Galaxy (ARGalaxy).

Key words Next generation sequencing, Primary immunodeficiency, B-cell receptor repertoire, Inborn errors of immunity

1 Introduction

At this moment, more than 450 monogenetic defects have been reported in patients with inborn errors of immunity (IEI) [1]. The most common forms of IEI are patients with a predominant B-cell disorder leading to primary antibody deficiencies. T-cell disorders also have an effect on the development and function of B cells, because they are required for further differentiation of B cells into memory B cells and plasma cells.

IEI can have a direct or indirect effect on the B-cell receptor (BCR) repertoire. Direct effects are found in patients with genetic defects in genes involved in one of the key processes in the formation or shaping of the B-cell repertoire: V(D)J recombination, somatic hypermutation (SHM), class switch recombination
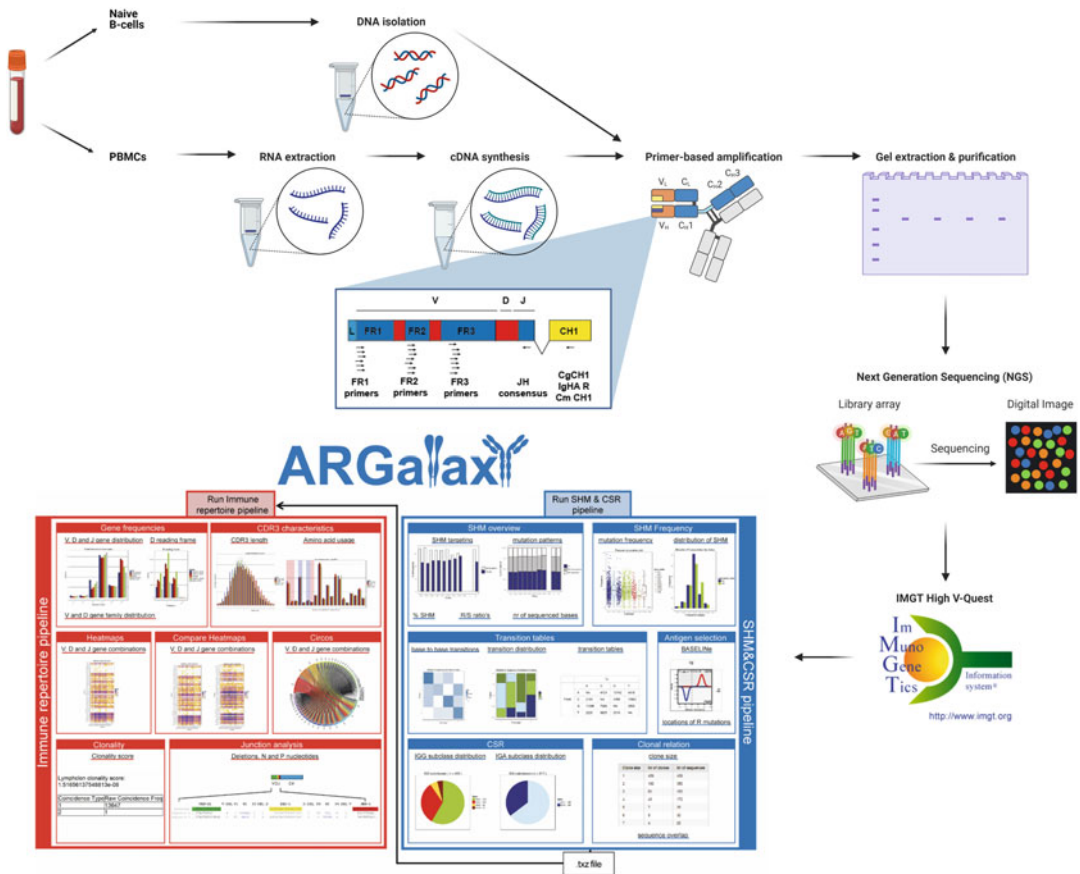
Fig. 1 Schematic overview of the workflow. Summary of the workflow for NGS-based B-cell receptor sequencing using primer-based amplification and analysis using the Antigen Receptor Galaxy (ARGalaxy) pipeline. Created with BioRender.com

(CSR), and (pre-)BCR signalling and proliferation [2–4]. Indirect effects can also be found because recurrent infections and/or auto-immunity can shape the BCR repertoire in IEI patients [5].

The BCR can be studied in several different ways, largely depending on the research question that needs to be answered and on the availability of the material. We will discuss how the BCR repertoire can be studied by amplifying BCR rearrangements from either DNA or cDNA and how to analyze the data using the Antigen Receptor Galaxy (ARGalaxy) analysis tool (Fig. 1). These methods can be applied to every sample, but we will focus on considerations that will affect the setup of the experiments and the data analysis for patients with IEI.

1.1 Selection of the Type of Cell or Tissue

The BCR repertoire can be divided into three classes: the immature BCR repertoire, the naïve BCR repertoire, and the antigen-selected BCR repertoire (Fig. 2). The immature BCR repertoire is derived
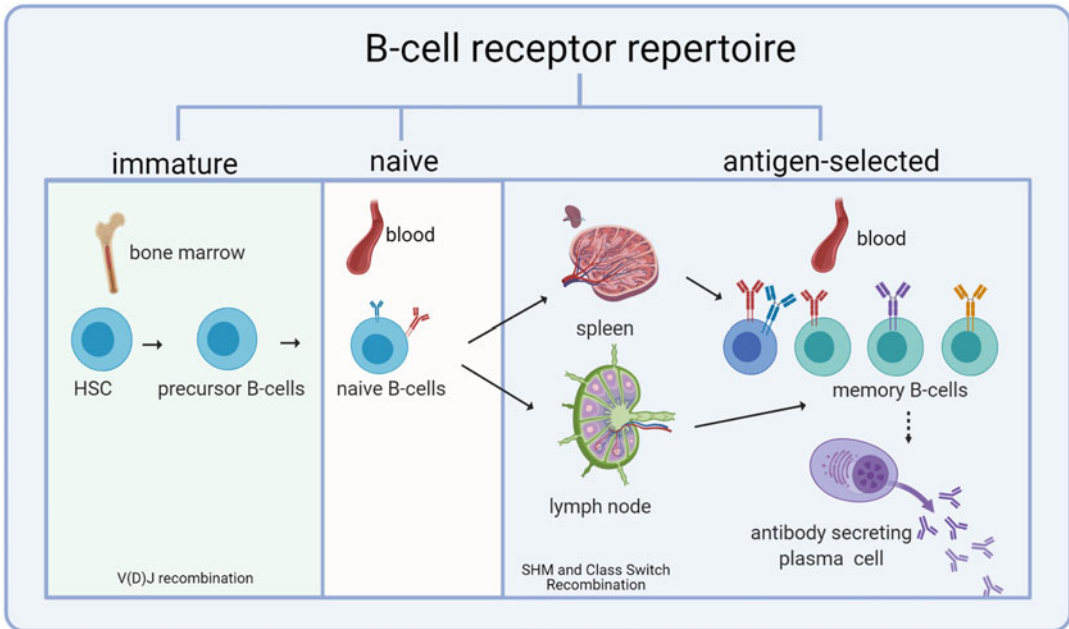
Fig. 2 Overview of the B-cell receptor repertoires. The B-cell receptor (BCR) repertoire can be divided into immature BCR repertoire, naïve BCR repertoire and antigen-selected BCR repertoire. Created with BioRender. com

from precursor B cells that did not undergo selection and/or have not completed their BCR rearrangements. This repertoire is particularly interesting for studying BCR repertoire formation in developing precursor B cells and processes like V(D)J recombination or pre-BCR signalling. Since precursor B-cell development takes place in the bone marrow, the only tissue that can be used to study the immature BCR is bone marrow. The naïve BCR repertoire is derived from naïve B cells that have not been activated. These naïve B cells can be found in peripheral blood. Peripheral blood is the least invasive material to obtain and for most labs easily accessible. However, peripheral blood contains a mixture of B-cell subsets, including naïve, memory, and plasma cells. The antigen-selected repertoire is derived from B cells that have been activated by their antigen. These B cells will differentiate into memory B cells or plasma cells. The antigen-selected B cells can be found in peripheral blood or secondary lymphoid organs, such as spleen or lymph nodes. Because tissues contain a mixture of B-cell subsets, it might be relevant to sort the population of interest before performing immune repertoire analysis of the naïve BCR repertoire or the antigen-selected BCR repertoire.

1.2   DNA Versus RNA   BCR rearrangements can be amplified from either DNA or RNA (cDNA). DNA is more stable than RNA and can be isolated from smaller cell numbers. The advantage of DNA is that is allows to

study unproductive and incomplete (DH-JH) rearrangements, which is not possible with RNA. Furthermore, there is only one DNA copy of a given functional rearrangement per cell, in contrast to RNA where there are many RNA copies per rearrangement per cell. The number of RNA copies is much higher in plasma cells compared to memory B cells. The advantage of RNA is that it allows to only analyze productive rearrangements and to study the constant gene. Furthermore, RNA is also preferred to use unique molecular identifiers (UMI) to identify the single RNA molecules.

1.3 The Number of B Cells that Can Be Studied

The BCR repertoire has been studied for decades by amplifying BCR rearrangements, cloning, and Sanger sequencing. However, since the introduction of next generation sequencing, it is possible to study thousands or even millions of BCR, in a way that is less labor intensive. The challenge of this high-throughput method is to obtain enough B cells to study thousands or millions of BCR rearrangements, especially in patients with a B-cell deficiency. Therefore, in patients with IEI, the starting material is often mononuclear cells obtained from blood or bone marrow. When using mononuclear cells, it is good to determine the frequency of B cells, e.g., using flow cytometry to be able to estimate the number of B-cell rearrangements that can be analyzed.

1.4 Location of the Primers

The IGH locus consist of >100 different variable (V), diversity (D), and joining (J) genes that are recombined to form a BCR. Fortunately, many of the genes have large sequence similarities and can therefore be subdivided in different families, such that primers specific for these gene families can be used in a multiplex PCR to amplify the repertoire. The forward primers can be located in the leader, or the frame work regions (FR) of the VH genes. Preferably, the forward primers should not be located in the complementary determining regions (CDR) regions, because these regions can have a high frequency of somatic hypermutations (SHM) that can decrease the binding efficiency of the primer. The location of the primers is also dependent on the information that is needed from the immune repertoire data. Primers in the leader sequence are least affected by SHM and provide the most accurate information about the hypomorphic alleles, but this results in a long amplicon that might not be suitable for all sequence platforms. In this protocol, we use the 6 IGHV FR1, 7 IGHV FR2, or 7 IGHV FR3 forward primers adapted with the Rd1 adaptor for Illumina sequencing (Fig. 3) (Table 1) [6]. As reverse primer, a single primer in the JH gene is enough to cover all six functional JH genes (Table 1). However, when there is an interest in information about the (sub)-class of the BCR, a primer in the constant (C) gene can be used. These rearrangements can only be amplified using cDNA as starting material. Since the amount of material is often limited in patients
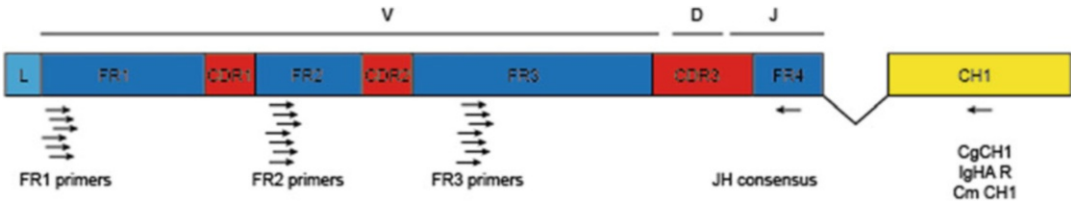
Fig. 3 Overview of IGH locus with primers. The forward primers located in FR1, FR2, or FR2 are indicated. For B-cell receptor rearrangements amplified from DNA the JH consensus can be used. For amplification of the B-cell receptor rearrangements from cDNA, either the JH consensus, the CgCH1, IgHA R, or the Cm CH1 primers can be used

with IEI, using primers in the Cγ or Cα region also allows to select for rearrangements derived from Ig-switched memory B cells without the need of pre-sorting of these cells. Optionally, a reverse primer in the Cμ region can be used. Subsequently, the data can be separated in rearrangements that contain <2% SHM and are likely derived from naïve B cells and rearrangements that have >2% SHM, which are likely derived from memory B cells. The reverse primers should also be adapted by addition of the Rd2 adaptor for Illumina sequencing (Table 1).

1.5   Choosing a Tool to Analyze the Immune Repertoire Data

Next generation sequencing of the BCR repertoire generates thousands of rearrangements and requires bioinformatics tools to analyze. In this last decade, many different analysis tools have been developed. Most tools help to annotate the rearrangements and will aid to visualize the data. The choice of the tool greatly depends on the research question, and it is likely that multiple tools are needed to answer all questions. In this chapter, we will discuss the Antigen Receptor Galaxy (ARGalaxy) tool [7]. This tool is a web-based tool and can be used to analyze many different qualitative measurements. It has two different pipelines, the immune repertoire pipeline which allows the analysis of V, D, and J gene usage, CDR3 characteristics and junction characteristics, and the SHM and CSR pipeline which allows the analysis of SHM, antigen selection, and CSR. Depending on the research question, data can be analyzed with either one or both pipelines.

2   Materials

2.1   Amplification (VH-Cg or VH-Ca from cDNA or VH-JH from DNA)

1. cDNA or 50 ng/μl DNA.
2. PCR cycler.
3. PCR tubes.
4. AmpliTaqGold (Thermo Fisher Scientific) with 10× Buffer Gold.
5. 25 mM $MgCl_2$.

Table 1
Overview of primers sequences

Name primer	Rd1 or Rd2 adaptor illumina	Template-specific sequence	Primer 5′ - 3′
VH1-FR1 Rd1	ACACTCTTTCCCTACACGACGCTCTTCCGATCT	GGCCTCAGTGAAGGTCTCCTGCAAG	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGGCCTCAGTGAAGGTCTCCTGCAAG
VH2-FR1 Rd1	ACACTCTTTCCCTACACGACGCTCTTCCGATCT	GTCTGGTCCTACGCTGGTGAAACCC	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGTCTGGTCCTACGCTGGTGAAACCC
VH3-FR1 Rd1	ACACTCTTTCCCTACACGACGCTCTTCCGATCT	CTGGGGGGTCCCTGAGACTCTCCTG	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCTGGGGGGTCCCTGAGACTCTCCTG
VH4-FR1 Rd1	ACACTCTTTCCCTACACGACGCTCTTCCGATCT	CTTCGGAGACCCTGTCCCTCACCTG	ACACTCTTTCCCTACACGAGCGCTCTTCCGATCTCTTCGGAGACCCTGTCCCTCACCTG
VH5-FR1 Rd1	ACACTCTTTCCCTACACGACGCTCTTCCGATCT	CGGGGAGTCTCTGAAGATCTCCTGT	ACACTCTTTCCCTACACGAGCGCTCTTCCGATCTCGGGGGAGTCTCTGAAGATCTCCTGT
VH6-FR1 Rd1	ACACTCTTTCCCTACACGACGCTCTTCCGATCT	TCGCAGACCCTCTCACTCACCTGTG	ACACTCTTTCCCTACACGAGCGCTCTTCCGATCTTCGCAGACCCTCTCACTCACCTGTG
VH1-FR2 Rd1	ACACTCTTTCCCTACACGACGCTCTTCCGATCT	CTGGGTGCGACAGGCCCCTGGACAA	ACACTCTTTCCCTACACGAGCGCTCTTCCGATCTCTGGGGTGCGACAGGCCCCTGGACAA
VH2-FR2 Rd1	ACACTCTTTCCCTACACGACGCTCTTCCGATCT	TGGATCCGTCAGCCCCCAGGGAAGG	ACACTCTTTCCCTACACGAGCGCTCTTCCGATCTTGGATCCGTCAGCCCCCAGGGAAGG
VH3-FR2 Rd1	ACACTCTTTCCCTACACGACGCTCTTCCGATCT	GGTCCGCCAGGCTCCAGGGAA	ACACTCTTTCCCTACACGAGCGCTCTTCCGATCTGGTCCGCCAGGCTCCAGGGAA

VH4-FR2 Rd1	ACACTCTTTCCCTACACGACGC TCTTCCGATCT	TGGATCCGGCCAGCCC CCAGGGAAGG	ACACTCTTTCCCTACA CGACGGCTCTTCCGATCTTGGA TCCGCCAGCCCCCAGGGAAGG
VH5-FR2 Rd1	ACACTCTTTCCCTACACGACGC TCTTCCGATCT	GGGTGCGCCAGA TGCCCGGGAAAGG	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGGG TGCGCCAGATGCCCGGGAAAGG
VH6-FR2 Rd1	ACACTCTTTCCCTACACGACGC TCTTCCGATCT	TGGATCAGGCAGTCCCCA TCGAGAG	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTGGA TCAGGCAGTCCCCATCGAGAG
VH7-FR2 Rd1	ACACTCTTTCCCTACACGACGC TCTTCCGATCT	TTGGGTGCGACAGGCCCC TGGACAA	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTTGGG TGCGACAGGCCCCTGGACAA
VH1-FR2 Rd1	ACACTCTTTCCCTACACGACGC TCTTCCGATCT	TGGAGCTGAGCAGCCTGAGA TCTGA	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTGGAGC TGAGCAGCCTGAGATCTGA
VH2-FR2 Rd1	ACACTCTTTCCCTACACGACGC TCTTCCGATCT	CAATGACCAACATGGACCCTG TGGA	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCAA TGACCAACATGGACCCTGTGGA
VH3-FR2 Rd1	ACACTCTTTCCCTACACGACGC TCTTCCGATCT	TCTGCAAATGAACAGCC TGAGAGCC	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTC TGCAAATGAACAGCCTGAGAGCC
VH4-FR2 Rd1	ACACTCTTTCCCTACACGACGC TCTTCCGATCT	GAGCTCTG TGACCGCCGCCGCGGCACG	ACACTCTTTCCCTACACGAGGCTCTTCCGATCTGAGCTC TGTGACCGCCGCCGCGGCACG
VH5-FR2 Rd1	ACACTCTTTCCCTACACGACGC TCTTCCGATCT	CAGCACCGCCTACCTGCAG TGGAGC	ACACTCTTTCCCTACACGACGCTCTTCCGATC TCAGCACCGCCTACCTGCAGTGGAGC
VH6-FR2 Rd1	ACACTCTTTCCCTACACGACGC TCTTCCGATCT	GTTCTCCCTGCAGCTGAACTC TGTG	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGTTC TCCCTGCAGCTGAACTCTGTG

(continued)

Table 1
(continued)

Name primer	Rd1 or Rd2 adaptor illumina	Template-specific sequence	Primer 5′ - 3′
VH7-FR2 Rd1	ACACTCTTTCCCTACACGACGC TCTTCCGATCT	CAGCACGGCATATCTGCAGA TCAG	ACACTCTTTCCCTACACGACGCTCTTCCGATC TCAGCACGGCATATCTGCAGATCAG
JH cons Rd2	TCGGGAGTTAATGCAACGATCG TCGAAATTCGC	CTTACCTGAGGAGACGG TGACC	TCGGGAGTTAATGCAACGATCGTCGAAATTCGCCTTACC TGAGGAGACGGTGACC
CgCH1 Rd2	TCGGGAGTTAATGCAACGATCG TCGAAATTCGC	GGAAGGTGTGCACGCCGC TGGTC	TCGGGAGTTAATGCAACGATCGTCGAAA TTCGCGGAAGGTGTGCACGCCGCTGGTC
IgHA R Rd2	TCGGGAGTTAATGCAACGATCG TCGAAATTCGC	CTTTCGCTCCAGGTCACAC TGAG	TCGGGAGTTAATGCAACGATCGTCGAAATTCGCC TTTCGCTCCAGGTCACACTGAG
Cm CH1 Rd2	TCGGGAGTTAATGCAACGATCG TCGAAATTCGC	GGGAATTCTCACAGGAGACGA	TCGGGAGTTAATGCAACGATCGTCGAAATTCGCGGGGGAA TTCTCACAGGAGACGA

6. dNTP solution; prepare a mix with 20 mM of each nucleotide.

7. Bovine serum albumin (BSA) (20 mg/ml).

8. Nuclease-free PCR water.

9. 10 μM (10 pmol/μl) primer: pipet every primer separately.

10. Ethidium bromide (Sigma).

11. Agarose.

12. Tris-Borate-EDTA (TBE) buffer.

13. Loading dye for DNA gels.

14. 100 bp DNA ladder.

15. Gel extraction kit (Qiagen).

16. Scalpels: use 1 scalpel per PCR reaction.

2.2 Nested PCR

1. PCR cycler.

2. PCR tubes.

3. KAPA HiFi Hotstart Ready mix (Roche).

4. TruSeq Custom Amplicon Index Kit (Illumina).

2.3 Merging, Trimming, and Alignment of Reads and Data Analysis

1. https://argalaxy.researchlumc.nl/.

2. PEAR (https://cme.h-its.org/exelixis/web/software/pear/) [8].

3. Cutadapt (http://code.google.com/p/cutadapt) [9].

4. FASTQ to FASTA converter (http://usegalaxy.org/u/dan/p/fastq) [10].

5. IMGT High-V-Quest (http://www.imgt.org/HighV-QUEST/home.action) [11].

6. Immune repertoire tool of ARGalaxy (https://argalaxy.researchlumc.nl/).

7. SHM and CSR tool of ARGalaxy (https://argalaxy.researchlumc.nl/).

3 Methods

3.1 Amplification of VH-Cg, VH-Ca, or VH-Cμ from cDNA

1. Prepare PCR master mix consisting of 28.3 μl water, 5 μl 10× Buffer Gold, 3 μl MgCl$_2$, 0.5 μl dNTPs, 1 μl BSA, and 0.2 μl Taq Gold (*see* **Note 1**).

2. Transfer PCR master mix into PCR reaction tubes (38 μl into each well).

3. Deposit 1 μl of each primer into the corresponding well (*see* **Notes 2**, **3**, and **4**).

4. Add 5 μl cDNA to the corresponding well, and carefully add the lid of the PCR tubes (*see* **Notes 4** and **5**).

5. Run PCR at 95 °C for 7 min; 25–35 cycles at 94 °C for 30 s, 57 °C for 30 s, 72 °C 1 min; 72 °C for 10 min (*see* **Note 6**).

6. Load 50 μl PCR product with 10 μl loading dye onto a 1% agarose gel in TBE buffer containing ethidium bromide and run for 1 h at 180 V.

7. Visualize DNA band under ultraviolet (UV) light (*see* **Note 7**), and cut the PCR band of approximately 500 bp from gel using a scalpel (*see* **Note 8**).

8. Purify the PCR product from gel using the gel extraction kit. Follow the instructions in the manual and eluate with 20 μl elution buffer.

9. Continue with Subheading 3.3, Nested PCR.

3.2 Amplification of VH-JH from DNA

1. Prepare PCR master mix consisting of 31.3 μl water, 5 μl 10× Buffer Gold, 3 μl $MgCl_2$, 0.5 μl dNTPs, 1 μl BSA, and 0.2 μl Taq Gold (*see* **Note 1**).

2. Transfer PCR master mix into PCR reaction tubes (41 μl into each well).

3. Deposit 1 μl of each primer (6 5′ primers and 1 Cg or Ca primer per well) into the corresponding well (*see* **Notes 2** and **4**).

4. Add 2 μl 50 ng/μl DNA to the corresponding well, and carefully add the lid of the PCR tubes (*see* **Notes 4** and **9**).

5. Run PCR at 95 °C for 7 min; 25–35 cycles at 94 °C for 30 s, 57 °C for 30 s, 72 °C 1 min; 72 °C for 10 min (*see* **Note 6**).

6. Load 50 μl PCR product with 10 μl loading dye onto a 1% agarose gel in TBE buffer containing ethidium bromide and run for 1 h at 180 V.

7. Visualize DNA band under ultraviolet (UV) light (*see* **Note 7**), and cut the PCR band of approximately 500 bp from gel using a scalpel (*see* **Note 8**).

8. Purify the PCR product from gel using the gel extraction kit. Follow the instructions in the manual and eluate with 20 μl elution buffer.

9. Continue with Subheading 3.3, Nested PCR.

3.3 Nested PCR and Pooling

1. Add 12.5 μl KAPA HiFi Hotstart Ready mix, 2 μl TruSeq Custom Amplicon Index forward primer, 2 μl TruSeq Custom Amplicon reverse primer, and 8.5 μl purified PCR product from Subheading 3.1 or Subheading 3.2 to a PCR reaction tube.

2. Run PCR at 95 °C for 5 min; 10 cycli at 98 °C for 20 s, 66 °C for 30 s, 72 °C 30 s; 72 °C for 1 min.

3. Measure the concentration of the PCR products (*see* **Note 10**).

4. Mix the PCR product at an equimolar concentration of 50 mM.

5. Purify the pool of PCR products (*see* **Note 11**).

6. The PCR pool can be sequenced using the Illumina platform.

3.4 Merging, Trimming, and Alignment of Reads Using Galaxy

1. Sequencing with the Illumina platform results R1 and R2 reads that need to be merged before they can be aligned to a reference database. This merging can be done with PEAR, which is a pair-end read merger [8], which can be found on "pre-processing" at https://argalaxy.researchlumc.nl/.

2. After the reads are merged, the Illumina Rd1 and Rd2 primer adapters have to be removed from the reads as well as the forward primers. This can be done with the Cutadapt tool [9] (*see* **Note 12**). which can be found on "pre-processing" at https://argalaxy.researchlumc.nl/.

3. Before the reads can be aligned using IMGT/HighV-Quest, the FASTQ files have to be adapted to the FASTA file format. This can be done with the FASTQ to FASTA converter [10], which can be found on "pre-processing" at https://argalaxy.researchlumc.nl/.

4. For alignment and annotation of the BCR rearrangements, the international ImMunoGeneTics system IMGT/HighV-Quest can be used (http://www.imgt.org/HighV-QUEST/analysis.action) [11]. This tool will produce a compressed .txz file that contains 12 text files with alignment information.

3.5 Data Analysis Using the Immune Repertoire Pipeline in Antigen Receptor Galaxy (ARGalaxy) (See Note 13)

1. Open ARGalaxy from https://argalaxy.researchlumc.nl/ [7].

2. Upload the compressed .txz files using: get data → upload file (*see* **Note 12**). The file will appear on the right site of your screen under "History."

3. Select under "Tools" on the left site of the screen "ARGalaxy," and click on the "Immune repertoire pipeline."

4. Select the .txz file you would like to analyze (*see* **Notes 14** and **15**).

5. Enter a name in the "ID" field (*see* **Note 16**).

6. Select the definition of the clonotype (*see* **Note 17**).

7. Select the order in which the V, D, and J genes have to appear in the graphs. The default setting is on alphabetical order and not in the order they appear on the IGH locus.

8. Select "IGH" at the "Locus" field.

9. Choose if you want to visualize the unproductive rearrangements in the graphs (*see* **Note 18**).

Table 2
Example of the overview table of the Immune repertoire pipeline in ARGalaxy showing the number and percentage of (unique) productive and unproductive sequences per donor and per replicate. The definition of unique sequences is based on the clonal type definition filter setting chosen

Donor/replicate	All	Productive	Unique productive	Unproductive	Unique unproductive
TEST	42,488	32,207 (76%)	14,905 (35%)	10,010 (24%)	6067 (14%)
TEST_1	18,911	15,036 (80%)	6452 (34%)	3772 (20%)	2164 (11%)
TEST_2	11,880	9390 (79%)	4852 (41%)	2394 (20%)	1583 (13%)
TEST_3	11,697	7781 (67%)	3601 (31%)	3844 (33%)	2320 (20%)

10. Select if you want to identify overlapping sequences between different replicates within one donor (*see* **Note 19**).

11. Press execute. A new item will be displayed in your history and turn green when the tool is ready with processing the data.

12. Click on the "eye" symbol to open the table that shows an overview of the rearrangements, including the percentage of productive, productive unique, unproductive, and unproductive unique (*see* Table 2 for an example).

13. Press on "Click here for the results" to open the page with the different analysis tabs.

14. The tab "Gene frequencies" shows the percentage of V, D, an J gene usage (*see* **Note 20**). The frequency of the different V, D, and J genes vary slightly between individuals and also between different primer sets that are used to amplify the BCR rearrangements. However, there are some important parameters that can give an indication of changes in the BCR repertoire (*see* Table 3). These changes can be specific for patients with IEI, but are also present between the naïve and antigen-selected BCR repertoire in healthy individuals. For example, the frequency of BCR with the IGHV4–34 and IGHJ6 genes are relatively high in the naïve BCR repertoire, but are significantly lower in antigen-selected B cells. In contrast to IGHJ4 which is less frequently used than IGHJ6 in the naïve BCR repertoire, it is the most frequently used IGHJ gene in the antigen-selected repertoire in healthy individuals (Fig. 4a) [15].

15. The tab "CDR3 characteristics" contains plots that show the distribution of the CDR3 length and the frequency of the different amino acids used in the CDR3. The median CDR3 length is longer in naïve B cells compared to memory B cells, which is likely caused by selection against long CDR3 lengths, because they are more likely to be autoreactive (Fig. 4b) [15].

Table 3
Overview of V, D, and J genes that can be affected in the B-cell receptor repertoire

Gene	Remark
IGHV4-34	B cells expressing this IGHV gene are almost all intrinsically autoreactive [12, 13]. The frequency of VH4-34 is high in naive B cells, but very low in memory B cells. An increased frequency has been observed in several IEI patients
IGHJ4 and IGHJ6	JH4 and JH6 are the most frequently used JH genes in healthy donors; however the distribution might differ between naive and memory B cells. The length of the JH6 gene is significantly longer than the other JH genes. Since memory B cells have shorter CDR3 length, the frequency of JH6 is lower in memory B cells compared to naive B cells in the same healthy donor [4]
IGHD7–27	IGHD7-27 is the smallest D gene and is located immediately adjacent to the IGHJ locus. High frequencies of IGHD7-27 have been observed in fetal B-cell receptor rearrangements [14]

16. In the tab "Heatmaps," the frequency of the different combinations of V-J, V-D, and D-J genes are visualized in heatmaps.

17. In the tab "Compare heatmaps," the heatmaps between different donors can be compared.

18. In the tab "Circos," the frequency of the different combination of V-J, V-D, and D-J genes are visualized using circus plots [16].

19. When the option is chosen to determine the number of sequences that share the same clonal type between replicates or to determine the clonality of the donor, the tab "Shared Clonal Types" or "Clonality" is shown. These tabs include a table with information about the number of BCR rearrangement that is present in multiple replicates of the same donor. When three replicates are present and the option "determine the clonality of the donor" is chosen, the clonality score based on the publication by Boyd et al. is given [17]. In patients with IEI, the diversity of the repertoire is often reduced (Fig. 4c) [2, 7].

20. The tab "Junction analysis" contains a table with the median or mean number of deletions, palindromic (P) nucleotides, or non-templated (N) nucleotides in the productive and unproductive rearrangements. Genetic defects in the non-homologous end joining (NHEJ) pathway have been shown to affect the number of deletions, N-nucleotides and P-nucleotides (*see* Table 4).

21. In the "Download" tab, all data used to create the tables and graphs can be downloaded.
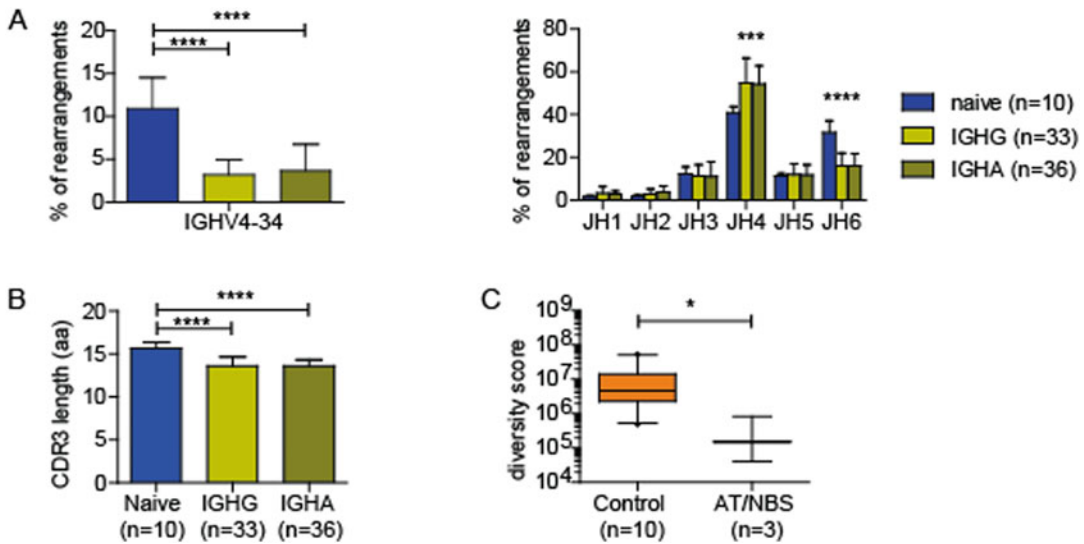
Fig. 4 Examples of analyses with the immune repertoire pipeline. Naïve B cells have a higher frequency of IGHV4–34 and IGHJ6 compared to antigen-selected switched B cells (IGHG, and IGHA) (**a**). The CDR3 length is shorter in antigen-selected switched B cells (IGHG and IGHA) compared to naïve B cells (**b**). Patients with ataxia telangiectasia (AT) or Nijmegen breakage syndrome have a reduced diversity of the naïve BCR repertoire (**c**). The number of samples analyzed is indicated per group. *P*-values <0.001 are indicated by *** and *P*-values <0.0001 are indicated by ****

Table 4
Overview junction characteristics of IEI patients with defects in the non-homologous end joining pathway

Gene	Protein name	Deletions	N-nucleotides	P-nucleotides	References
DCLRE1C	ARTEMIS	↓	↓	↑↑	[18]
PRKDC	DNA-PKcs	↓	↓	↑	[19]
NHEJ1	XLF/Cernunnos	Normal	↓↓↓	Normal	[4]
LIG4	DNA ligase 4	↑↑↑	↓	Normal	[20]
XRCC4	XRCC4	Normal	↓	Normal	[21]

3.6 Data Analysis Using the SHM and CSR Tool in ARGalaxy (See Note 13)

1. Open ARGalaxy from https://argalaxy.researchlumc.nl/ [7].

2. Upload the compressed .txz files using: get data → upload file (*see* **Note 14**). The file will appear on the right site of the screen under "History."

3. Select under "Tools" on the left site of the screen "ARGalaxy" and click on the "SHM and CSR pipeline."

4. Select the .txz file to be analyzed.

5. Select which regions of the BCR rearrangements should be included in the analysis (*see* **Note 21**).

6. Select if only the productive, only the unproductive, or both productive and unproductive sequences should be analyzed.

7. Select if the sequences should be filtered by "remove unique" or "keep unique." The "remove unique filter" removes all sequences that occur only once and the duplicates (based on the nucleotides sequence of the "analyzed region" and the C gene or the sequences that have the same V, J, and amino acid sequence of the CDR3 region). When choosing "remove unique," an additional filter appears that allows to choose the minimal number of duplicates that have to be in a group in order to keep one of the sequences (based on the nucleotides sequence of the "analyzed region" and the C gene or the sequences that have the same V, J, and amino acid sequence of the CDR3 region) . The "keep unique" filter removes all duplicate sequences based on the nucleotides sequence of the "analyzed" region and the C gene.

8. Select if duplicates should be removed based on V, CDR3, and C region (different options possible).

9. The class/subclass filter should only be applied when part of the C region is present. The SHM and CSR pipeline identifies human Cμ, Cα, Cγ, and Cε constant genes by dividing the reference sequences for the subclasses (NG_001019) in eight nucleotide chunks, which overlap by four nucleotides. These overlapping chunks are then individually aligned in the right order to each input sequence. This alignment is used to calculate the chunk hit percentage and the nt hit percentage. The chunk hit percentage is the percentage of the chunks that is aligned. The Nt hit percentage is the percentage of chunks covering the subclass-specific nucleotide match with the different subclasses. The most stringent filter for the subclass is 70% "nt hit percentage" which means that five out of seven subclass-specific nucleotides for Cα or six out of eight subclass specific nucleotides of Cγ should match with the specific subclass. The option ">19% class" can be chosen when only the class (Cα/Cγ/Cμ/Cε) of the sequences is of interest and the length of the sequence is not long enough to assign the subclasses. With the location of the primers used in this protocol, assignment of subclass is not possible and the class can be assigned with the >19% filter.

10. Select if a new IGMT archive output is needed in the history that contains only the sequences based on the filtering options used before (*see* **Note 13**).

11. Select if the generation of new IMGT archives and the analysis of Change-O/Baseline need to be skipped to decrease the time ARGalaxy needs to run the pipeline.

12. Press execute. A new item will be displayed in the history and turns green when the tool is ready with processing the data.

13. Click on the "eye" symbol to open the table that shows the number of rearrangement after each filtering step.

14. Press on "Click here for the results" to open the page with the different analysis tabs.

15. The "SHM overview" tab gives a table with detailed information on the SHM including frequency of SHM, the transversion and transition mutations, replacement and silent mutations, etc. Furthermore, it also contains graphs visualizing the percentage of mutations in AID and pol eta motives, the relative mutation patterns, and the absolute mutation patterns. The frequency of SHM increased during childhood (Fig. 5a) [15], but can be affected in patients with IEI (Fig. 5b). This can be caused by genetic defects in one of the genes involved in the SHM process [22], but can also be the consequence of recurrent infections or immune dysregulation.

16. The "SHM frequency" tab contains graphs that visualize the frequency of SHM per (sub)class.

17. The "transition table" tab contains tables, heatmaps, and bar graph that visualize the SHM per base. This information provides a lot of information about the SHM process and can be used to study the SHM pathway. In patients with genetic defects in genes involved in the DNA repair pathways (*UNG*, *MSH2*, *MSH6*, *PMS2*) crucial for the induction of SHM, the frequency as well as the pattern of SHM is affected (Fig. 5c) [22].

18. The "antigen selection" tab contains bar plots showing the frequency of replacement mutations per amino acid. These graphs can be used to study in which region or amino acids positions replacement mutations are most/least frequent. Furthermore, in this tab, also the plots showing the score for antigen selection based on the BASELINe method are given [23].

19. The "CSR" tab contains circle plots that indicate the subclass distribution of the IGHA or IGHG rearrangements. In patients with IEI, the subclass distribution is often affected (Fig. 5d). This can be caused by defects in CSR, e.g., in patients with ataxia telangiectasia (AT) [2], but is also observed in patients with common variable immunodeficiency [24].

20. The "clonal relation" tab gives a table which indicates the number of clones and the number of sequences within a clone (the definition of the clone is based on the filter settings used) based on the Change-O method [25] (*see* **Note 22**).

21. In the "Download" tab, all data used to create the tables and graphs can be downloaded.
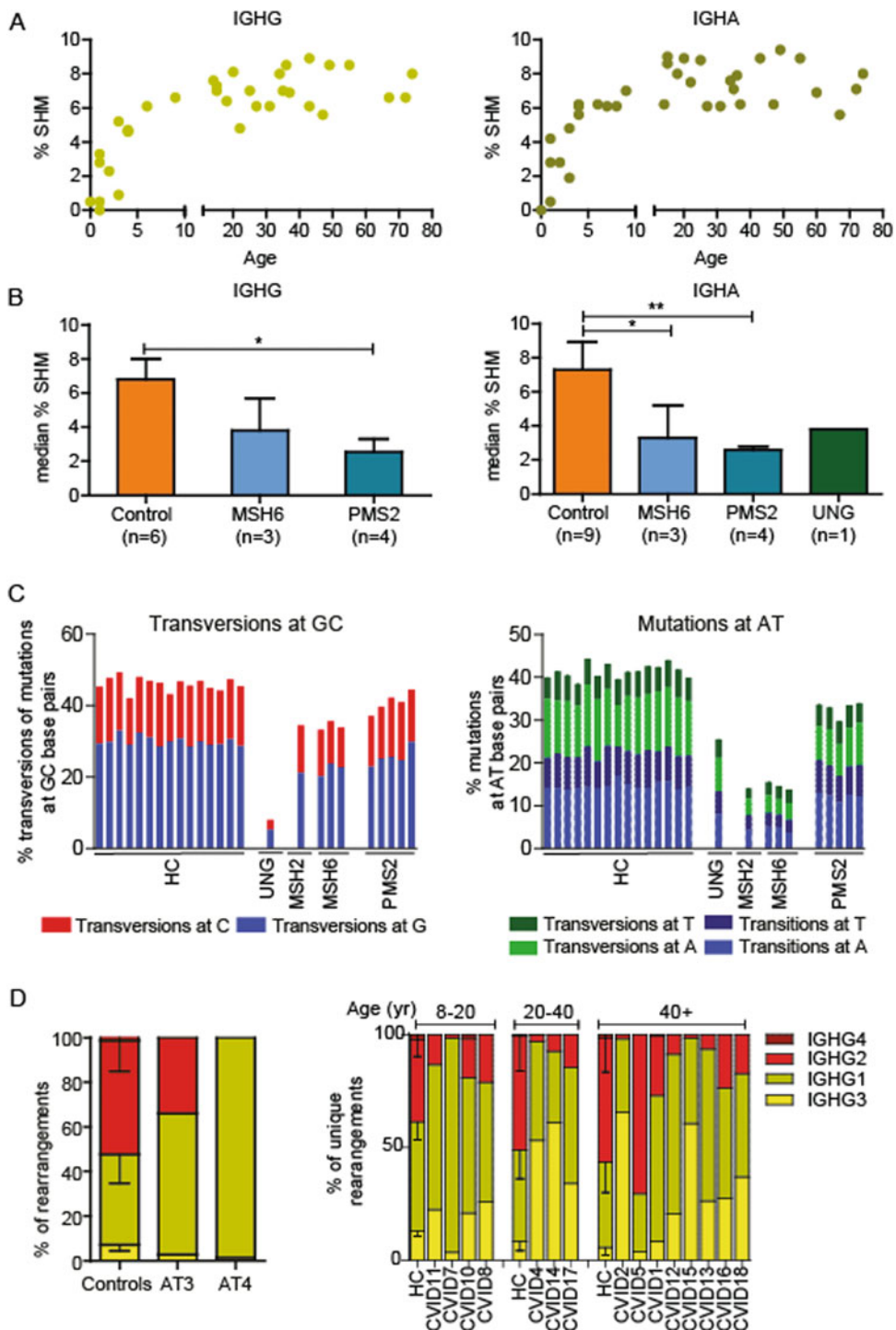
Fig. 5 Examples of analyses with the SHM and CSR pipeline. The median frequency of somatic hypermutations (SHM) increases during childhood in both IGHG and IGHA antigen-selected switched B cells (**a**). The median frequency of SHM is reduced in patients with MSH6, PMS2, or UNG deficiency (**b**). Patients with defects MSH2 and MH6 have a strong reduction in mutation at A and T base pairs compared to controls. Patients with UNG deficiency have a strong reduction in transversion mutations at G and C base pairs (**c**). Patients with ataxia telangiectasia (AT) and common variable immunodeficiency (CVID) have reduced frequency of IGHG2 and IGHG4 switched B cells (**d**). *$P < 0.05$ and **$P < 0$

4   Notes

1. Prepare a master mix for all reactions. Allow a surplus of 10%.

2. When using IGH VH FR1 primers, add six forward primers or seven primers when using IGH VH FR2 or IGH VH FR3.

3. When using reverse primers in the constant region, prepare a separate reaction to amplify the VH-Cα, VH-Cγ, or VH-Cμ rearrangements with the IgHA R Rd2, CgCH1 Rd2, or Cm CH1 Rd2 primers, respectively.

4. In these steps, prevention of cross-contaminations between samples is essential.

5. The amount of cDNA that needs to be added is dependent on the amount of B cells in the samples, and the number of B-cell rearrangement to be analyzed. Furthermore, it has to be taken into account that plasma cells typically have a 1000 times higher copy number of the RNA copies of the B-cell rearrangement compared to other B cells. If less cDNA is needed, nuclease-free PCR water can be added to reach a total volume of 5 μl.

6. The number of PCR cycli should preferably be low enough to remain in the linear amplification stage of the PCR which will reduce amplification bias. However, the number of cycli should be high enough to be able to visualize the PCR product on the agarose gel. The lower the number of B cells in the sample, the higher the number of PCR cycli that should be used.

7. Keep the exposure to UV as short as possible since UV can damage the PCR products.

8. Use a new scalpel for every PCR product to avoid contamination.

9. The amount of DNA that needs to be added is dependent on the amount of B cells in the sample, and the number of B-cell rearrangements to be analyzed. If less volume is needed to add the accurate amount of DNA, nuclease-free PCR water can be added to reach the total volume of 2 μl. The amount of DNA per cells is estimated to be 6 pg. So for DNA isolated from only B cells can be divided by 6 pg. 100 ng DNA corresponds to approximately 16,667 B cells. Since every B cells can have one unproductive and 1 productive rearrangement, 100 ng B-cell DNA can result in maximally 33,334 unique B-cell rearrangements.

10. The amount of PCR product should be measures with a sensitive method for low quantities of double-stranded DNA, e.g., Qubit™ dsDNA BCR Assay Kit.

11. Purification of the PCR library pool can be done with AMPure XP beads (Beckman Coulter) according to the manufacturer's instructions.

12. Removing the adaptor sequences from the reads improves the alignment of the BCR rearrangements. Removal of primer sequences located in the V(D)J region is essential to prevent mismatching of degenerate primers to be classified as SHM.

13. Dependent on your research question, analyze the data using the immune repertoire pipeline, the SHM and CSR pipeline, or both. When interested in using both using the same filtering, one can start with the analysis of the SHM and CSR pipeline, and then select "yes" in the filter "Output new IMGT archives per class into the history." This will provide a new data set in the history with the filtered data (split per class if class is assigned) which then can be analyzed using the immune repertoire pipeline.

14. Select "imgt_archive" by "Type (set all)."

15. In the "Immune repertoire pipeline," multiple .txz files can be analyzed simultaneously. These can be replicates from the same donor or can be derived from multiple donors.

16. Spaces and special characters (except "_") cannot be used in the ID field. Leaving this field empty will result in an error.

17. The data likely contains multiple reads which are identical or nearly identical. These reads can be derived from unique B cells with the same IGH rearrangements, but these can also be technical duplicates. When the BCR rearrangements are amplified from a low number of B-cells and/or many PCR cycles had to be used to obtain a PCR product, the presence of reads with the same clonotype is more likely caused by technical duplicates. Importantly, IGH rearrangements with the same CDR3 sequence at the amino acid level can be derived from unique B cells with a different IGH rearrangement at the nucleotide level. This filter will only include one sequence with the same clonotype definition in the analysis.

18. Unproductive rearrangements are rearrangements that are out-of-frame or contain a stop codon. When IGH rearrangements were amplified from DNA, a large fraction of the rearrangements are non-productive, while in case of amplification of IGH rearrangements from RNA, only a very small fraction of the IGH rearrangements will be unproductive, since unproductive rearrangements are mostly not transcribed.

19. Option 1 "Do not determine overlap (only 1 replicate present)" should be used if only one replicate is analyzed per donor or if there is no interest to determine the presence of overlapping sequences. Option 2 "Determine the number of

sequences that share the same clonal type between the replicate" should be used if the overlap between at least two replicates within the same donor should be determined. Option 3 "Determine the clonality of the donor (minimal 3 replicates) can be used to determine the number of overlapping sequences between at least three replicates within one donor and provides the clonality score described by Boyd et al. [17].

20. The rearrangements used for making the graphs are filtered based on the settings "Clonal type definition" and "Remove the unproductive sequences from graphs." When choosing to filter the data based on clonal type and remove the unproductive sequences from the graph, only the total number of unique productive sequences are included in the graphs.

21. The regions that are/can be included in the analysis depend on the forward primer being used. When using primers in the leader sequence, the complete BCR rearrangement can be used. However, when using primers in the FR regions, these regions have to be excluded because the primers sequences can cause false-positive SHM.

22. To calculate clonal relation, Change-O is used [26]. Transcripts are considered clonally related, if they have maximally three nucleotides difference in their CDR3 sequence and the same first V gene (as assigned by IMGT). Change-O settings used are the nucleotide hamming distance substitution model with a complete distance of maximally three. For clonal assignment, the first genes were used, and the distances were not normalized. In case of asymmetric distances, the minimal distance was used.

References

1. Tangye SG, Al-Herz W, Bousfiha A, Cunningham-Rundles C, Franco JL, Holland SM et al (2021) The ever-increasing array of novel inborn errors of immunity: an interim update by the IUIS committee. J Clin Immunol 41:666–679

2. Driessen GJ, IJspeert H, Weemaes CM, Haraldsson A, Trip M, Warris A et al (2013) Antibody deficiency in patients with ataxia telangiectasia is caused by disturbed B- and T-cell homeostasis and reduced immune repertoire diversity. J Allergy Clin Immunol 131: 1367–1375

3. Kolhatkar NS, Brahmandam A, Thouvenel CD, Becker-Herman S, Jacobs HM, Schwartz MA et al (2015) Altered BCR and TLR signals promote enhanced positive selection of autoreactive transitional B cells in Wiskott-Aldrich syndrome. J Exp Med 212:1663–1677

4. IJspeert H, Rozmus J, Schwarz K, Warren RL, van Zessen D, Holt RA et al (2016) XLF deficiency results in reduced N-nucleotide addition during V(D)J recombination. Blood 128: 650–659

5. Galson JD, Clutterbuck EA, Truck J, Ramasamy MN, Munz M, Fowler A et al (2015) BCR repertoire sequencing: different patterns of B-cell activation after two Meningococcal vaccines. Immunol Cell Biol 93:885–895

6. van Dongen JJ, Langerak AW, Bruggemann M, Evans PA, Hummel M, Lavender FL et al (2003) Design and standardization of PCR primers and protocols for detection of clonal immunoglobulin and T-cell receptor gene recombinations in suspect lymphoproliferations: report of the BIOMED-2 Concerted Action BMH4-CT98-3936. Leukemia 17: 2257–2317

7. IJspeert H, Van Schouwenburg P, Van Zessen D, Pico-Knijnenburg I, Stubbs AP, Van der Burg M (2017) Antigen receptor galaxy: a user-friendly web-based tool for analysis and visualization of T and B cell receptor repertoire data. J Immunol 198:4156–4165

8. Zhang J, Kobert K, Flouri T, Stamatakis A (2013) PEAR: a fast and accurate Illumina Paired-End reAd mergeR. Bioinformatics 30: 614–620

9. Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal [S.l.], 17(1):10–12. ISSN 2226-6089. https://doi.org/10.14806/ej.17.1.200

10. Blankenberg D, Gordon A, Von Kuster G, Coraor N, Taylor J, Nekrutenko A, Galaxy T (2010) Manipulation of FASTQ data with Galaxy. Bioinformatics 26:1783–1785

11. Alamyar E, Duroux P, Lefranc MP, Giudicelli V (2012) 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 Mol Biol 882:569–604

12. Klonowski KD, Primiano LL, Monestier M (1999) Atypical VH-D-JH rearrangements in newborn autoimmune MRL mice. J Immunol 162:1566–1572

13. Wardemann H, Yurasov S, Schaefer A, Young JW, Meffre E, Nussenzweig MC (2003) Predominant autoantibody production by early human B cell precursors. Science 301: 1374–1377

14. Shiokawa S, Mortari F, Lima JO, Nunez C, Bertrand FE 3rd, Kirkham PM et al (1999) IgM heavy chain complementarity-determining region 3 diversity is constrained by genetic and somatic mechanisms until two months after birth. J Immunol 162: 6060–6070

15. IJspeert H, van Schouwenburg PA, van Zessen D, Pico-Knijnenburg I, Driessen GJ, Stubbs AP et al (2016) Evaluation of the antigen-experienced B-cell receptor repertoire in healthy children and adults. Front Immunol 7:410

16. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D et al (2009) Circos: an information aesthetic for comparative genomics. Genome Res 19:1639–1645

17. Boyd SD, Marshall EL, Merker JD, Maniar JM, Zhang LN, Sahaf B et al (2009) Measurement and clinical monitoring of human lymphocyte clonality by massively parallel VDJ pyrosequencing. Sci Transl Med 1:12–23

18. van der Burg M, Verkaik NS, den Dekker AT, Barendregt BH, Pico-Knijnenburg I, Tezcan I et al (2007) Defective Artemis nuclease is characterized by coding joints with microhomology in long palindromic-nucleotide stretches. Eur J Immunol 37:3522–3528

19. van der Burg M, IJspeert H, Verkaik NS, Turul T, Wiegant WW, Morotomi-Yano K et al (2009) A DNA-PKcs mutation in a radiosensitive T-B- SCID patient inhibits Artemis activation and nonhomologous end-joining. J Clin Invest 119:91–98

20. van der Burg M, van Veelen LR, Verkaik NS, Wiegant WW, Hartwig NG, Barendregt BH et al (2006) A new type of radiosensitive T-B-NK+ severe combined immunodeficiency caused by a LIG4 mutation. J Clin Invest 116:137–145

21. Murray JE, van der Burg M, IJspeert H, Carroll P, Wu Q, Ochi T et al (2015) Mutations in the NHEJ component XRCC4 cause primordial Dwarfism. Am J Hum Genet 96: 412–424

22. IJspeert H, van Schouwenburg PA, Pico-Knijnenburg I, Loeffen J, Brugieres L, Driessen GJ et al (2019) Repertoire sequencing of B cells elucidates the role of UNG and mismatch repair proteins in somatic hypermutation in humans. Front Immunol 10:1913

23. Yaari G, Vander Heiden JA, Uduman M, Gadala-Maria D, Gupta N, Stern JN et al (2013) Models of somatic hypermutation targeting and substitution based on synonymous mutations from high-throughput immunoglobulin sequencing data. Front Immunol 4: 358

24. van Schouwenburg PA, IJspeert H, Pico-Knijnenburg I, Dalm V, van Hagen PM, van Zessen D et al (2018) Identification of CVID patients with defects in immune repertoire formation or specification. Front Immunol 9: 2545

25. Yaari G, Uduman M, Kleinstein SH (2012) Quantifying selection in high-throughput immunoglobulin sequencing data sets. Nucleic Acids Res 40:e134

26. Gupta NT, Vander Heiden JA, Uduman M, Gadala-Maria D, Yaari G, Kleinstein SH (2015) Change-O: a toolkit for analyzing large-scale B cell immunoglobulin repertoire sequencing data. Bioinformatics 31 (20):3356–3358