



HAL
open science

AIRR Community Guide to Planning and Performing AIRR-Seq Experiments

Anne Eugster, Magnolia Bostick, Nidhi Gupta, Encarnita Mariotti-Ferrandiz,
Gloria Kraus, Wenzhao Meng, Cinque Soto, Johannes Trück, Ulrik Stervbo,
Eline Luning Prak

► **To cite this version:**

Anne Eugster, Magnolia Bostick, Nidhi Gupta, Encarnita Mariotti-Ferrandiz, Gloria Kraus, et al.. AIRR Community Guide to Planning and Performing AIRR-Seq Experiments. Immunogenetics, 2453, Springer US, pp.261-278, 2022, Methods in Molecular Biology, 10.1007/978-1-0716-2115-8_15 . hal-03838024

HAL Id: hal-03838024

<https://hal.sorbonne-universite.fr/hal-03838024>

Submitted on 6 Apr 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Chapter 15

AIRR Community Guide to Planning and Performing AIRR-Seq Experiments

Anne Eugster , Magnolia L. Bostick , Nidhi Gupta ,
Encarnita Mariotti-Ferrandiz , Gloria Kraus, Wenzhao Meng,
Cinque Soto , Johannes Trück , Ulrik Stervbo ,
and Eline T. Luning Prak  and on behalf of the AIRR Community

Abstract

The development of high-throughput sequencing of adaptive immune receptor repertoires (AIRR-seq of IG and TR rearrangements) has provided a new frontier for in-depth analysis of the immune system. The last decade has witnessed an explosion in protocols, experimental methodologies, and computational tools. In this chapter, we discuss the major considerations in planning a successful AIRR-seq experiment together with basic strategies for controlling and evaluating the outcome of the experiment. Members of the AIRR Community have authored several chapters in this edition, which cover step-by-step instructions to successfully conduct, analyze, and share an AIRR-seq project.

Key words AIRR-seq, Immunoglobulin, Antibody, T-cell receptor, Immune repertoire, V(D)J recombination, Next-generation sequencing

1 Introduction

Next-generation sequencing of adaptive immune receptor repertoires (AIRR-seq of immunoglobulin, IG and T-cell receptor, TR rearrangements) has provided a new frontier for in-depth analysis of the immune system. The Adaptive Immune Receptor Repertoire (AIRR) Community was founded with the goal of developing standards for AIRR-seq studies to enable analysis and sharing of AIRR-seq data. In this book, members of the AIRR Community and colleagues have contributed sample methods for immune repertoire profiling studies. These AIRR Community chapters cover experimental (wet lab) and computational (dry lab) methods and encompass all of the many facets of the AIRR Community. While much of our focus in these chapters is on how to adequately control, standardize, annotate, and share data, we found it

impossible to discuss these attributes of AIRR-seq data without also describing the types of data sets that are generated and then integrating those descriptions with data analysis for commonly encountered use cases. In the companion AIRR Community data analysis chapters, information is provided about study design, data analysis, data use, and the AIRR data commons and how data can be reused and shared. In this chapter we describe how to plan and perform AIRR-seq experiments.

2 Planning the Experiment

Understanding the dynamics, selection, and pathology of immune responses has been greatly aided in recent years by next-generation sequencing (NGS)-based approaches to studying the adaptive immune receptor repertoire (AIRR) [1–3]. The AIRR Community is focused on the standardization, sharing, and re-use of these repertoire data [4]. The AIRR is the collection of distinct B-cell and T-cell clones (cells that are derived from a common progenitor cell) that are found in an individual. Each clone is associated with a distinct antigen receptor, which is a B-cell receptor (BCR or IG) or a TR. The DNA sequences that encode IG or TR are very diverse. This diversity is achieved through the recombination of variable (V), diversity (D), and joining (J) gene segments [5, 6]. Moreover, somatic hypermutation (SHM) provides further diversification of IG repertoires through DNA mutation [7, 8]. In addition to facilitating the sampling of diverse and complex immune repertoires, AIRR-seq has opened the door for systematic analysis and comparison of immune responses across different individuals and disease conditions [9–12]. The immune repertoire is dynamic and changes in its composition and diversity with age [13, 14], in different anatomic sites [15] and under diverse conditions such as malignancy, autoimmunity, immunodeficiency, infection, or vaccination [9, 13, 16–21]. In addition to comparing different individuals, AIRR-seq is also a powerful method for studying the evolution of immune responses or tracking specific B- or T-cell populations over time within individuals [22]. For example, clonal expansions can be identified, quantified, and monitored [23]. AIRR-seq studies not only enhance our ability to understand how to diagnose and monitor diseases but also can inform therapeutic approaches [12, 24–31].

When designing a study that leverages AIRR-seq data, there are several considerations including the subjects, sample types, manner in which the samples are processed, timeline and other considerations. The types of samples, their numbers, and budget often drive the types of questions that can be asked and answered using AIRR-seq. Once a suitable question has been defined and appropriate samples have been identified, the next major branch point in the

decision-making process involves the selection of AIRR-seq methods. In this section, we provide a brief overview of the most important considerations when selecting one or more AIRR-seq methods for a research study or clinical evaluation.

2.1 Organisms

This chapter focuses on samples from humans, but of course samples from other vertebrates or synthetic libraries (such as phage display [32]) are possible. If one is planning an experiment with nonhuman or synthetic samples, it is worth considering whether there are established protocols (such as PCR primer sets) and analysis pipelines (to include adequate libraries of validated germline gene sequences for animal species that are not frequently studied) for downstream analysis. With respect to samples derived from humans, there are several considerations [4, 33]. First, are the samples coming from individuals who have been consented for a research study? If not, one should check with the local institutional review board (IRB) or other regulatory body and/or with the investigator who supplies the samples for guidance on whether samples can be studied or if additional regulatory approvals may be required for full analysis and/or sharing of the data. Second, the study design will be impacted by the availability of samples from individuals in different comparison groups or on the availability of samples that are collected over time from the same individuals. Depending on the research question, resources, and time horizon for the project, study participants may be recruited who have a particular disease (in which case the phase of the disease and prior or current therapies may be important). If studying immune responses, longitudinal collections from the same individual at multiple time points and synchronization of those time points across the study cohort may be important to study changes in clonal abundance or, in the case of B cells, the level of SHM within clonal lineages. Demographic characteristics of the individuals in the group under study (including but not limited to age, geographical origin and sex, disease history) and the availability of one or more appropriately matched control groups are additional considerations. For TR-based sequencing, it is also useful to consider the HLA type, as HLA can have a major impact on TRBV gene usage [34]. Finally, if published data are going to be used for comparison, compatibility of the assay platforms and sample types is important.

2.2 Samples and Processing

Studies on humans are often limited by sample availability. The most commonly used sample is peripheral blood, which serves as starting material for a range of different sample types including whole blood (drawn into a tube with an anticoagulant such as EDTA), peripheral blood mononuclear cells (PBMCs, which are typically isolated by centrifugation over a Ficoll gradient), or plasma (the liquid portion of anticoagulated whole blood, which is typically prepared by centrifugation and stored in aliquots frozen for

isolation of cell-free DNA). Samples from other body fluids such as cerebrospinal fluid or bronchoalveolar lavage may also provide important insights if sampled in certain disease states. Tissue samples can be obtained from fine-needle aspirations (where sample quantities may be very limited, particularly if the same samples are being used for both clinical and research purposes) or from biopsies, where larger amounts of tissue can be sampled. In the case of the bone marrow, the aspirate is typically used for the evaluation of clonally expanded populations. In some cases, it is possible to obtain multiple tissues (surveillance biopsies for transplant rejection or bone marrow samples) as well as peripheral blood from the same individual over time. Finally, different tissues can be accessed from the same individual in organ donors or living individuals, as has been described for studies of human tissue-based immunity [35] and in certain disease states, such as type 1 diabetes, lupus, or rheumatoid arthritis [36–43]. From most of these samples, either total cells or isolated cell subsets (obtained after cell sorting using flow cytometry or magnetic bead-based methods) can be analyzed. The sample size and purity of the cell population of interest are important to consider when designing the experiment and interpreting the results.

How samples are processed is a critical consideration for the design of AIRR-seq experiments. Bulk sequencing methods can use samples that are formalin-fixed, lysed, or non-viably cryopreserved. Fixation significantly reduces the quality of the input nucleic acid and may require larger amounts of input DNA or RNA as well as protocols that use shorter amplicons (such as primers that are positioned in FR3 instead of FR1). The longer a sample sits in a fixative or is stored as a formalin-fixed paraffin-embedded (FFPE) tissue section, the poorer the template quality becomes. If it is possible to obtain snap frozen tissues that are not fixed, this is preferable. For certain cell types, such as diffuse large B-cell lymphoma, using tissue sections may provide a higher yield of cells of interest than single-cell suspensions [44]. For single-cell-based methods, viable cells are essential and typically consist of either freshly isolated cells or cryopreserved cells. In the case of cryopreserved cells, one needs to consider whether the method of initial sample preparation has influenced the recovery or phenotype of the cell population of interest.

Cell sorting or enrichment with magnetic beads can be used to selectively recover larger numbers of cells of interest, as, for example, with antigen-specific T cells identified by multimer staining, but these methods can also result in significant loss of sample. Sorting time should be kept to a minimum for plate-based single-cell methods, as cell viability decreases rapidly in the plate; ideally, the time from the addition of a life/dead staining solution to the end of the sort should not exceed 30 min. If longer sorting times are necessary, as is often the case for rare cells, cells can be sorted

into PCR strips instead. For droplet sequencing-based single-cell methods, batches of 1000–20,000 cells are usually collected in PCR tubes that need to be coated to ensure complete recovery of the cells for further processing.

2.3 Bulk vs. Single-Cell Sequencing

There are two complementary approaches to analyze the AIRR by sequencing that are usually driven by the number of cells available and the research question. On the one hand, bulk AIRR-seq methods allow systematic and global analysis of TR and IG repertoires from as few as 1000 cells to hundreds of thousands of cells or more. Bulk methods provide information about the TR (usually alpha + beta) or IG (heavy + light) rearrangements, although the pairing information is lost during the cell lysis step. On the other hand, single-cell AIRR-seq offers the possibility to reconstruct paired chain information for each TR or IG. However, most single-cell methods use lower cell input numbers (usually <20,000 cells, due to constraints in costs associated with kits and sequencing). Hence single-cell approaches, when used on bulk populations, generally tend to be focused on specific cell subsets or antigen-enriched cells to ensure sufficient sampling of the population of interest. In some cases, for example, when multiple samples with different amounts of cell inputs are available from the same individual, it may be preferable to use a tiered approach. For example, one might rely on bulk sequencing to get a view of the overall clonal landscape and then leverage single-cell sequencing to gain detailed insights into the association of specific clones (with paired chain information) and cell phenotypes (either through flow cytometry or by single-cell RNA-seq). The single-cell approach is discussed in detail in the AIRR Community chapter (Chapter 20)

2.4 Template Amplification from DNA vs. RNA

Bulk AIRR-seq can be performed on libraries that have been generated from either genomic DNA (gDNA) or RNA. gDNA-based methods are exclusively based on multiplex PCR approaches, where primers targeting the different V genes (or leader regions) and J genes are combined in the same reaction. Advantages of DNA-based sequencing are the stability of the template and its parsimonious nature (one template per cell), which allows for studies in which large numbers of cells are studied at modest cost. Disadvantages include the potential for primer bias, as PCR primers are usually positioned in the V gene and J gene (due to constraints on sequence length) and the potential loss of amplification in heavily mutated IG sequences. The bulk DNA approach is discussed in the AIRR Community chapter (Chapter 18).

Messenger RNA-based methods can be based on multiplex PCR (with either V and J primer combinations or V and constant region (C) primer combinations), or they can use rapid amplification of cDNA Ends (RACE)-PCR. Advantages of RNA-based sequencing are (1) more “shots on goal” with RNA than DNA

(with individual B/T cells harboring multiple RNA copies vs. only a single DNA copy), allowing for higher yield of amplicons when there are low cell numbers; (2) reduced PCR bias with primers that are in the constant region, (3) the incorporation of unique molecular identifiers (UMI) at the cDNA synthesis step (allowing for the generation of high-fidelity consensus sequences); and (4) the ability to generate data on the constant region usage for isotyping. Disadvantages of RNA-based sequencing methods include greater cost associated with the higher sequencing depths that are required (particularly if UMIs are used) and biases introduced by differences in transcript abundance in different cell types (if mixed rather than sorted populations are used for input). In the AIRR Community chapter (Chapter 19), we focus on the mRNA-based approach to AIRR-seq.

2.5 Commercial Kit vs. Homebrew Bulk Methods

Several commercial kits are now available to generate AIRR-seq data. Currently available commercial kits include gDNA-based methods (e.g., Adaptive Biotechnologies, iRepertoire) as well as mRNA-based methods (e.g., Illumina, Takara Bio, iRepertoire, MiLaboratory). Advantages of commercial-grade AIRR-seq assays are that kit reagents are produced following standards and rigorous quality controls such as qualifying primers, controlling for contamination, and verifying yield and amplification standards. Some vendors obtain certification in meeting rigorous quality standards in their laboratories that manufacture reagents, such as those set forth by the International Organization for Standardization (e.g., ISO 9001). In addition, service providers such as Adaptive Biotechnologies and iRepertoire offer large data sets for comparison and a series of user-friendly data analysis tools. Some disadvantages of commercial methods are that kits are expensive and sometimes these assays are not easily adapted to specific experimental needs. On the other hand, with homebrew assays, there is considerable variation in assay linearity and reproducibility (e.g., *see* ref. 45), and it can take months or even years to set up robust, well-validated assays that are then also not easy to adjust. The use of commercially available kits for in-house experiments can be a compromise to ensure reliability of the reagents and protocol customization.

2.6 Single Cell: Index Sorting and Bead-Based Emulsion Approaches

Single-cell AIRR-seq (scAIRR-seq), as any other single-cell sequencing technology, relies on partitioning each cell. In early protocols, cells were index sorted into plates, and multiplex PCR was used to amplify both chains of immune receptors of a cell concomitantly [46, 47]. The emergence of single-cell RNA-seq (scRNA-seq) has provided another tool for AIRR-seq. Many protocols to recover and sequence mRNA from single cells have been developed and differ in their approaches for cell capture, cDNA synthesis (full-length or tag-based) and amplification (only PCR or PCR following reverse transcription), and library preparation steps

[48]. Probably the most frequently used current commercial protocol for sequencing small cell numbers leverages the scSMARTer technology. With this approach, paired IG/TR information became accessible by combining full-length scRNA-seq amplification approaches with the development of the de novo assembly-based bioinformatics tools (TraCer, scTCR Seq, TRAPes, VDJ Puzzle) [49–52]. Unfortunately, these approaches remain computationally intensive, relatively costly, and are constrained with respect to cell throughput. More recently, bead-based emulsion methods have been developed for higher-throughput single-cell sequencing, allowing access to repertoires of tens of thousands of cells [53]. The formation of droplets in an oil-water emulsion using microfluidics allows single-cell encapsulation, barcoding, and the production of cDNA from each cell and culminates in parallel sequencing of the transcriptomes of thousands of cells [54]. These approaches have been adapted to sequence both TR or IG chains in parallel [55] and are available commercially, via the 10× Genomics platform (Chromium 10×), thereby allowing the processing of samples of 5×10^2 to 1.5×10^4 cells. In addition to paired immune receptor data, it is also possible to obtain scRNA-seq data. Similar approaches are also commercially available including the BD Rhapsody VDJ CDR3 protocol, which relies on cell compartmentation by microwells and allows processing of 1×10^3 to 4×10^4 cells, and the Takara Bio ICELL8 Single-Cell System, which can process $\sim 1 \times 10^3$ cells. Recent progress on the throughput of single-cell sorting has been described with CelliGO, which combines cell encapsulation in droplets through microfluidics [56], but sequencing costs are still limiting the widespread adoption of these approaches.

2.7 Cost

Finally, cost may influence the choice of a particular protocol. There are many factors that contribute to the cost of AIRR-seq data generation. For example, the number of samples, the cost of sequencing, the sequencing depth, and the number of cells analyzed per sample are all important considerations. Furthermore, the choice between service providers, commercial kits, and “home-brew” methods will influence costs. In general, gDNA analysis is the most cost-effective method, because it usually requires the lowest-sequencing depth with the largest representation of cells per sample, whereas single-cell analysis is on the opposite end of the spectrum, with bulk cDNA sequencing in the middle [45].

2.8 Overview of Companion AIRR Community Method Chapters

The correct choice of method for a given experimental question is crucial and has to be carefully evaluated. The companion AIRR Community method chapters concern (1) “Bulk gDNA Sequencing of Antibody Heavy-Chain Gene Rearrangements for Detection and Analysis of B-Cell Clone Distribution” (Chapter 18), (2) “Bulk Sequencing from mRNA with UMI for Evaluation of B-Cell

Table 1
Overview of highlighted use cases in associated chapters

Highlighted use case	Major steps
Bulk gDNA Sequencing of Antibody Heavy Chain Gene Rearrangements for Detection and Analysis of B-Cell Clone Distribution; a method by the AIRR Community	
Analysis of the clonal landscape in different samples from the same individual	Bulk gDNA FR1 + JH primers (BIOMED2 adapted) 2 × 300 bp reads Multiple replicates per sample
Bulk Sequencing from mRNA with UMI for Evaluation of B-Cell Isotype and Clonal Evolution; a method by the AIRR Community	
Evaluation of an antibody response to viral infection with clonal evolution	Antigen-enriched cells Bulk RNA (SMARTer Kit) UMI at cDNA synthesis step Amplification of isotypes
Single-Cell Analysis and Tracking of Antigen-Specific T Cells: Integrating Paired-Chain AIRR-Seq and Transcriptome Sequencing; a method by the AIRR Community	
<i>Part A</i> Single-cell-paired TCR chain and/or mRNA sequencing of memory and/or whole CD8+ T cells from COVID-19 patients	Bulk memory CD8+ T cells Feature barcode and sample hashtagging possible Template amplification and size fractionation Paired TCR chain data Single-cell RNA-seq data, feature barcode data
<i>Part B</i> Sequencing of activated and/or SARS-CoV-2 antigen-specific CD8+ T-cell populations to map to whole repertoires	T-cells (sorted from PBMCs) binding to antigen or with activation marker expression Index sort possible Single-cell mRNA sequencing (SMART-Seq) Paired TCR chain data Single-cell RNA-seq data
Quality Control: Chain Pairing Precision and Monitoring of Cross-Sample Contamination; a method by the AIRR Community	
Chain pairing reference data and estimation of within sample AIRR-seq reproducibility	Isolation of memory B cells or CD8+ T cells Establishment of replicate stimulation cultures prior to single-cell sequencing

Isotype and Clonal Evolution” (Chapter 19), (3) “Single-Cell Analysis and Tracking of Antigen-Specific T Cells: Integrating Paired-Chain AIRR-Seq and Transcriptome Sequencing” (Chapter 20), and (4) “Quality Control: Chain Pairing Precision and Monitoring of Cross-Sample Contamination” (Chapter 21). These chapters illustrate four basic workflows for AIRR-seq, with a focus on IG for bulk sequencing, TR for single-cell sequencing, and IG and TR replicate analyses for quality control. The four methods are summarized in Table 1 and are discussed further below.

In Chapter 18, we illustrate, using a homebrew method with primer sequences adapted for NGS from the BIOMED2 immunoglobulin heavy-chain (IGH) PCR assays [57], how to evaluate the clonal landscape, including clone size distributions, clonal lineage analysis, and tracking of clones in different samples from the same individual. This method uses multiplex PCR and can be scaled to very high cell inputs as described [15]. The method shown uses long reads that are adequate for robust IGHV gene alignment and SHM evaluation but can also be performed with shorter reads, depending upon the sample type and DNA quality. In Chapter 19, IGH rearrangements are amplified from bulk RNA with UMIs incorporated at the cDNA synthesis step for the generation of high-fidelity consensus sequences using a commercial kit from Takara Bio. This method can be used for low to moderate throughput analysis of antigen-enriched cell populations, for evaluation of SHM, selection, and isotype usage. In Chapter 20, two different but parallel workflows are used to analyze single cells, both for paired TR transcripts as well as for their transcriptome, using two commercial kits, one from Takara Bio and one from 10× Genomics. Single-cell technologies can use a multiplex or RACE-based amplification and can generate long high-quality reads that can be mapped to individual cells but can also be based on AIRR target enrichment. One kit allows for the analysis of small numbers of antigen-enriched, index-sorted cells, useful in the case the cells of interest are present at very low frequencies in the overall sample, while the other kit allows for the analysis of larger cell numbers, providing insights into the overall T-cell repertoire as well as into other immune cell populations, if desired. The combination of paired-chain information and RNA-seq data can provide insights into the nature of the different T-cell populations that are found among expanded clones in various disease settings. Furthermore, through clonal overlap analysis, the data from the antigen-enriched cells can be integrated with the larger data set to further characterize the populations with respect to antigen-binding. In Chapter 21, two workflows are presented. The first is for the isolation of CD27+ memory B cells and their expansion in replicate cultures *in vitro*, using a cell line that expresses CD40L and a cocktail of cytokines. The second workflow is for the isolation of CD8+ T cells and their expansion using CD3/CD28 and IL-2 stimulation. The generation of these expanded cell cultures provides a larger input of more readily resampled cells that can be used as reference libraries for IG- or TR-paired chain combinations, respectively, as well as providing diverse libraries for the evaluation of within-sample reproducibility.

3 Interpreting the Results

3.1 Overview

Immune repertoire profiling experiments are affected by numerous pre-analytical, experimental, and post-analytical variables. Pre-analytical variables include the quality, quantity, and purity of the target cell population(s) in the sample. Experimental variables include the quality and length of the template for amplification, contamination at the level of the sample or PCR, hybrid PCR products, and PCR jackpots. The sequencing run can be affected by the concentration of the library, which can influence the clustering density; there can be cross-clustering in the flow cell, poor quality or short reads, and issues with controlling for sequencing depth (reads per template). Many technical problems with experiments can be evaluated during data analysis (please *see* the companion AIRR Community commentaries on “TR and IG Gene Annotation” (Chapter 16) and “Repertoire Analysis” (Chapter 17), so here we will limit our comments to basic strategies for controlling and evaluating the adequacy of the experiment on the wet bench side.

3.2 General QC Considerations and Controls

For sample and amplification QC, spectroscopy, agarose gel electrophoresis, or capillary electrophoresis can be used for the evaluation of nucleic acid purity and size distribution. Standardized samples that are put through the same workflow can be used to compare the entire AIRR-seq procedure in one assay run to another run, to help identify and control for batch effects. Bead purification and/or further gel purification can be performed to remove primer dimers, which can swamp sequencing runs and reduce the fraction of informative reads. Capillary electropherograms (e.g., Bioanalyzer) can be used to evaluate library quality, while KAPA quantitation and real-time PCR can be performed to quantify the library. For the sequencing run, the clustering density is important (as described in the individual protocol chapters). Another helpful metric is the fraction of reads that have quality scores of 30 or higher (projected sequencing error rates below 1 per 1000 nucleotides).

3.3 Clonal Recovery

The quality and type of sample have significant effects on the efficiency of amplification and clonal yield. FFPE tissue samples yield ~10-fold fewer clones than the same tissue snap frozen without fixation. Furthermore, the longer a tissue sits in FFPE, the poorer the sample quality becomes. For FFPE samples, using larger amounts of input DNA or RNA into the initial amplification can improve clonal recovery, as can the use of primers that target shorter amplicons (e.g., primers that flank the CDR3 sequence such as FR3 and JH [58]). Another reason for low numbers of clones is if the initial amplification uses primers that do not capture

a high enough fraction of the rearrangements in the sample. With RNA as the starting material, there is bias toward recovering more templates from cells that are activated. Plasma cells, for example, can produce ~100 times as much IG RNA as naive B cells [59]. Primers that amplify DNA are not subject to this problem, but can have other issues, such as the potential for nonuniform amplification of different templates. To correct for PCR bias, some assays use internal calibrators [60, 61]. Amplification of IG rearrangements has an additional challenge if these are highly somatically hypermutated. One hint that this may be occurring is if there is an elevated frequency of nonproductive rearrangements (from a bulk gDNA amplification). Alternative approaches in this situation are to amplify templates that are less prone to SHM such as the leader region in the VH genes or focus on RNA-based sequencing with primers that extend from the constant region [15]. Another approach is to amplify alternative loci (such as light chains, which have about half the level of SHM of heavy chains [62], RS (recombining sequence also known as kappa deleting element) rearrangements [63], or DJ rearrangements [58]).

3.4 PCR Cycle Number

For RNA-based protocols, the gene expression of each IG/TR chains can vary significantly from one cell to another. Therefore, it is challenging to predict how many cycles of PCR will amplify sufficient material for downstream sequencing without overamplification such that there are significant off-target PCR products. One approach is to focus on sorted cell populations to control for the effects of different transcript levels. In addition, one can amplify each chain of interest (e.g., IgH, IgK, IgL, etc.) separately, with different library index combinations for each chain. This can allow for separate optimization of cycling conditions for each chain, as discussed in Chapter 19. It is also possible that the suggested number of cycles will not generate enough material for downstream sequencing. If there is insufficient material for sequencing, we recommend increasing the number of cycles. Conversely, if the library yield is too high, the number of cycles in the library PCR amplification (e.g., PCR2 in Chapter 19) can be decreased.

3.5 Sensitivity

The sensitivity of an AIRR-seq experiment can be determined by titrating spike-ins, such as mixing cells with a known gene rearrangement into a diverse sample at different ratios, as described by Barennes and colleagues [45]. The linearity of the titration also reveals the range of clone concentrations where the method is quantitative or semiquantitative. The threshold of detection of the assay depends upon the biological question being asked, but if rare clonotypes need to be detected (as is the case for detection of minimal residual disease), then it is important to power the analysis on clone sizes. This can be accomplished experimentally by running multiple biological replicates (independent PCR amplifications) on the same sample and determining the fraction of rearrangements

that can be repeatedly sampled in two, three, four, or more replicates, as described previously [15, 64]. Using within-sample clonal overlap as a maximal estimate, one can then evaluate (with greater rigor) the expected overlap between one sample and a different sample [15]. If sensitivity falls below the level required, there are several potential reasons for this including poor-quality sample, too few cells (of the relevant type) in the sample, too small a sample, or a clone size that is too small to be detected. The depth of sequencing can also influence the detection of clones, particularly if one uses rigorous cutoffs for clone size or requires a minimum number of UMIs per clone.

3.6 Amplification Bias

As discussed in the amplification section in Chapter 18, DNA-based amplification methods can exhibit bias in the form of preferential amplification of certain genes over others. RNA-based amplification methods can be biased by transcript abundance, which is higher in certain cell types than others. To evaluate an AIRR-seq experiment for amplification bias, one can use an alternative method, such as flow cytometry with antibodies against known TCR V β chains, as a basis for comparison, as described in [45]. In single-cell experiments, one can quantify the recovery of receptors in different cell subsets using RNA-seq profiles to assign cells to different subsets. In addition, spike-in controls and cell mixtures with defined rearrangements can be used during protocol development to quantify bias. Primers with conserved sequence tags can also be used to evaluate bias, as described by Reddy and colleagues [61]. Bias can also occur during the sequencing step. For example, a higher depth of sequencing can result in greater coverage and the detection of smaller clones. However, in samples with few clones, a higher-depth sequencing can also create more sequencing errors which, depending on the bioinformatic pipeline, can result in skewed clone size or SHM profiles. If samples from different sequencing runs are being compared, it is important to consider potential batch effects due to differences in depth of sequencing, clustering density, and sequence quality. To minimize problems associated with batch effects, it is useful to include samples that are being compared to each other in the same run, whenever this is possible. One way to potentially control for (or at least recognize) batch effects is to include an external reference sample (such as pooled spleen or PBMCs) in each run.

3.7 Contamination

During data analysis, one can check for contamination by computing clonal overlap between different samples in the same experiment. Samples from the same individual will exhibit numerous overlapping clones, depending upon the level of sampling, whereas samples between individuals have far fewer overlapping clones. Overlapping clones or identical CDR3 sequences between different individuals cause concern for contamination if they have identical nucleotide sequences and if there are multiple shared sequences

(which is nearly impossible to achieve by chance, particularly for IG sequences, [65]). Spurious clonal overlap between different individuals can arise through mixing of samples prior to nucleic acid amplification, by erroneous assignment of sample barcodes, by PCR contamination, by cross-clustering of samples in the same flow cell, or some combination of these difficulties. Sample mixing can occur during flow cytometry if the instrument is not rigorously flushed between samples. Samples that are assigned the wrong barcode will associate with the “wrong” individual, or if samples come from different species, processing with the wrong pipeline (including the wrong database for reference germline genes) will result in sequences that have very low levels of sequence homology to the (incorrect) germline genes. If this occurs, an IgBLAST [66] search with a few sequences will quickly resolve to which species the genes correspond. With PCR contamination, one may see spurious amplification in the negative control samples (such as water or fibroblast DNA). PCR contamination can also often result in high-copy sequences that are shared by multiple subjects in the same experiment. In contrast, with cross-clustering, there is often a very-high-copy sequence and then a low number of copies of that same sequence in an unrelated individual. There are several process controls that can reduce the risk of contamination. First, there should be physically separate areas for pre- and post-PCR workstations. Second, primers with different barcodes can be used for diagnostic samples (where high-copy clones might be present) vs. MRD samples. Unique dual indices can be used to control for sequencing barcode crosstalk [67]. Third, when in doubt and if more samples are available, repeat the experiment to confirm the results.

3.8 Spurious Amplification Products

Sometimes one obtains unexpected sequences due to technical artifacts. Large clonal expansions can appear with PCR jackpot. In the case of gDNA, independent PCR amplifications of the same sample are sampling different gene rearrangements. If the same expanded clone is present in both biological replicates, it is far more likely to be due to a *bona fide* expansion instead of a PCR jackpot. Another artifact is a hybrid PCR product. With hybrid PCR products, templates with partial sequence homology can cross-amplify [68]. Hybrid products will tend to share sequences at either the 5' or 3' end and then exhibit a sharp boundary where the templates crossed over into the other sequence. One way to distinguish hybrid products from gene conversion events or biological variants in V gene sequences or potential convergence (with sharing of CDR3 sequences) is to amplify sequences with TRBV or IGHV gene specific primers and see if the same products can be recreated. In addition, using protocols with fewer PCR cycle numbers can sometimes be helpful in reducing spurious amplification products.

3.9 Data Reporting

The AIRR Community has published a series of data and experimental metadata sharing standards called MiAIRR [33]. The MiAIRR data standards guide the publication, curation, and sharing of AIRR-seq data and metadata and consist of six high-level data sets for study and subject, sample collection, sample processing and sequencing, raw sequences, processing of sequence data, and processed AIRR sequences. All current data fields in the MiAIRR standard can be accessed here: https://docs.airr-community.org/en/stable/miairr/data_elements.html.

More details on how to annotate and report AIRR-seq data and metadata are provided in the AIRR Community companion method chapter “Data sharing and re-use” (Chapter 23).

4 Conclusion

In this chapter, we have given an overview of the considerations needed to plan and execute a successful AIRR-seq experiment. We have also broadly discussed basic strategies for controlling and evaluating the adequacy of the experiment. Each topic touched upon in this chapter is explored in depth in the corresponding AIRR Community companion chapters.

Acknowledgments

This work is supported by NIH research grants awarded to E.L.P. (AI144288, AI106697, P30-AI0450080, P30-CA016520). U.S. was supported by grants from Mercator Stiftung, Germany; German Research Foundation, Germany (DFG, grant 397650460); BMBF e:KID, Germany (01ZX1612A); and BMBF NoChro, Germany (FKZ 13GW0338B). A.E. and G.K. are supported by grants from the Deutsche Forschungsgemeinschaft (BO 3429/3-1 and BO 3429/4-1) and the BMBF (RESET-AID). E.M.F. contributions were funded by iMAP (ANR-16-RHUS-0001), Transimmunom LabEX (ANR-11-IDEX-0004-02), TriPoD ERC Research Advanced Grant (Fp7-IdEAS-ErC-322856), AIR-MI (ANR-18-ECVD-0001), iReceptorPlus (H2020 Research and Innovation Programme 825821), and SirocCo (ANR-21-CO12-0005-01) grants. J.T. was supported by the Swiss National Science Foundation (Ambizione-SCORE: PZ00P3_161147 and PZ00P3_183777).

The authors thank Andrew Farmer for constructive criticism of the manuscript.

E.L.P. is the former Chair of the Adaptive Immune Receptor Repertoire Community, receives research funding from Roche Diagnostics and Janssen Pharmaceuticals for projects unrelated to

the methods presented in this chapter, and is consulting or an advisor for Roche Diagnostics, Enpicom, the Antibody Society, IEDB, and the American Autoimmune Related Diseases Association. J.T. is consulting or an advisor for Enpicom and Merck, Sharp & Dohme (MSD).

References

- Boudinot P, Mariotti-Ferrandiz ME, Pasquier LD, Benmansour A, Cazenave PA, Six A (2008) New perspectives for large-scale repertoire analysis of immune receptors. *Mol Immunol* 45(9):2437–2445. <https://doi.org/10.1016/j.molimm.2007.12.018>
- Robins HS, Campregher PV, Srivastava SK, Wacher A, Turtle CJ, Kahsai O et al (2009) Comprehensive assessment of T-cell receptor beta-chain diversity in alphabeta T cells. *Blood* 114(19):4099–4107. <https://doi.org/10.1182/blood-2009-04-217604>
- Georgiou G, Ippolito GC, Beausang J, Busse CE, Wardemann H, Quake SR (2014) The promise and challenge of high-throughput sequencing of the antibody repertoire. *Nat Biotechnol* 32(2):158–168. <https://doi.org/10.1038/nbt.2782>
- Breden F, Luning Prak ET, Peters B, Rubelt F, Schramm CA, Busse CE et al (2017) Reproducibility and reuse of adaptive immune receptor repertoire data. *Front Immunol* 8:1418. <https://doi.org/10.3389/fimmu.2017.01418>
- Tonegawa S (1983) Somatic generation of antibody diversity. *Nature* 302(5909):575–581. <https://doi.org/10.1038/302575a0>
- Sakano H, Kurosawa Y, Weigert M, Tonegawa S (1981) Identification and nucleotide sequence of a diversity DNA segment (D) of immunoglobulin heavy-chain genes. *Nature* 290(5807):562–565. <https://doi.org/10.1038/290562a0>
- Weigert MG, Cesari IM, Yonkovich SJ, Cohn M (1970) Variability in the lambda light chain sequences of mouse antibody. *Nature* 228(5276):1045–1047. <https://doi.org/10.1038/2281045a0>
- Papavasiliou FN, Schatz DG (2002) Somatic hypermutation of immunoglobulin genes: merging mechanisms for genetic diversity. *Cell* 109(Suppl):S35–S44. [https://doi.org/10.1016/s0092-8674\(02\)00706-7](https://doi.org/10.1016/s0092-8674(02)00706-7)
- Benichou J, Ben-Hamo R, Louzoun Y, Efroni S (2012) Rep-Seq: uncovering the immunological repertoire through next-generation sequencing. *Immunology* 135(3):183–191. <https://doi.org/10.1111/j.1365-2567.2011.03527.x>
- Six A, Mariotti-Ferrandiz ME, Chaara W, Magadan S, Pham HP, Lefranc MP et al (2013) The past, present, and future of immune repertoire biology—the rise of next-generation repertoire analysis. *Front Immunol* 4:413. <https://doi.org/10.3389/fimmu.2013.00413>
- Langerak AW, Bruggemann M, Davi F, Darzentas N, van Dongen JJM, Gonzalez D et al (2017) High-throughput immunogenetics for clinical and research applications in immunohematology: potential and challenges. *J Immunol* 198(10):3765–3774. <https://doi.org/10.4049/jimmunol.1602050>
- Arnaout RA, Luning Prak ET, Schwab N, Rubelt F, the AIRR Community (2021) The future of blood testing is the immunome. *Front Immunol*. <https://doi.org/10.3389/fimmu.2021.626793>
- Ghraichy M, Galson JD, Kelly DF, Truck J (2018) B-cell receptor repertoire sequencing in patients with primary immunodeficiency: a review. *Immunology* 153(2):145–160. <https://doi.org/10.1111/imm.12865>
- Gibson KL, Wu YC, Barnett Y, Duggan O, Vaughan R, Kondeatis E et al (2009) B-cell diversity decreases in old age and is correlated with poor health status. *Aging Cell* 8(1):18–25. <https://doi.org/10.1111/j.1474-9726.2008.00443.x>
- Meng W, Zhang B, Schwartz GW, Rosenfeld AM, Ren D, Thome JJC et al (2017) An atlas of B-cell clonal distribution in the human body. *Nat Biotechnol* 35(9):879–884. <https://doi.org/10.1038/nbt.3942>
- Kirsch IR, Watanabe R, O'Malley JT, Williamson DW, Scott LL, Elco CP et al (2015) TCR sequencing facilitates diagnosis and identifies mature T cells as the cell of origin in CTCL. *Sci Transl Med* 7(308):308ra158. <https://doi.org/10.1126/scitranslmed.aaa9122>
- Dziubianau M, Hecht J, Kuchenbecker L, Sattler A, Stervbo U, Rodelsperger C et al (2013) TCR repertoire analysis by next generation sequencing allows complex differential diagnosis of T cell-related pathology. *Am J*

- Transplant 13(11):2842–2854. <https://doi.org/10.1111/ajt.12431>
18. Hou D, Chen C, Seely EJ, Chen S, Song Y (2016) High-throughput sequencing-based immune repertoire study during infectious disease. *Front Immunol* 7:336. <https://doi.org/10.3389/fimmu.2016.00336>
 19. Galson JD, Kelly DF, Truck J (2015) Identification of antigen-specific B-cell receptor sequences from the total B-cell repertoire. *Crit Rev Immunol* 35(6):463–478. <https://doi.org/10.1615/CritRevImmunol.2016016462>
 20. Tipton CM, Fucile CF, Darce J, Chida A, Ichikawa T, Gregoretti I et al (2015) Diversity, cellular origin and autoreactivity of antibody-secreting cell population expansions in acute systemic lupus erythematosus. *Nat Immunol* 16(7):755–765. <https://doi.org/10.1038/ni.3175>
 21. Turner JS, Zhou JQ, Han J, Schmitz AJ, Rizk AA, Alsoussi WB et al (2020) Human germinal centres engage memory and naive B cells after influenza vaccination. *Nature* 586(7827):127–132. <https://doi.org/10.1038/s41586-020-2711-0>
 22. Davis CW, Jackson KJL, McElroy AK, Halfmann P, Huang J, Chennareddy C et al (2019) Longitudinal analysis of the human B cell response to ebola virus infection. *Cell* 177(6):1566–1582.e1517. <https://doi.org/10.1016/j.cell.2019.04.036>
 23. Miho E, Yermanos A, Weber CR, Berger CT, Reddy ST, Greiff V (2018) Computational strategies for dissecting the high-dimensional complexity of adaptive immune repertoires. *Front Immunol* 9:224. <https://doi.org/10.3389/fimmu.2018.00224>
 24. Robinson WH (2015) Sequencing the functional antibody repertoire—diagnostic and therapeutic discovery. *Nat Rev Rheumatol* 11(3):171–182. <https://doi.org/10.1038/nrrheum.2014.220>
 25. Fink K (2019) Can we improve vaccine efficacy by targeting T and B cell repertoire convergence? *Front Immunol* 10:110. <https://doi.org/10.3389/fimmu.2019.00110>
 26. Jiang N, Schonnesen AA, Ma KY (2019) Ushering in integrated T cell repertoire profiling in cancer. *Trends Cancer* 5(2):85–94. <https://doi.org/10.1016/j.trecan.2018.11.005>
 27. Jacobsen LM, Posgai A, Seay HR, Haller MJ, Brusko TM (2017) T cell receptor profiling in type 1 diabetes. *Curr Diab Rep* 17(11):118. <https://doi.org/10.1007/s11892-017-0946-4>
 28. Theil A, Wilhelm C, Kuhn M, Petzold A, Tuve S, Oelschlagel U et al (2017) T cell receptor repertoires after adoptive transfer of expanded allogeneic regulatory T cells. *Clin Exp Immunol* 187(2):316–324. <https://doi.org/10.1111/cei.12887>
 29. Sellner L, Bruggemann M, Schlitt M, Knecht H, Herrmann D, Reigl T et al (2017) GvL effects in T-prolymphocytic leukemia: evidence from MRD kinetics and TCR repertoire analyses. *Bone Marrow Transplant* 52(4):544–551. <https://doi.org/10.1038/bmt.2016.305>
 30. Stervbo U, Nienen M, Hecht J, Viebahn R, Amann K, Westhoff TH et al (2020) Differential diagnosis of interstitial allograft rejection and BKV nephropathy by T-cell receptor sequencing. *Transplantation* 104(4):e107–e108. <https://doi.org/10.1097/TP.0000000000003054>
 31. Babel N, Stervbo U, Reinke P, Volk HD (2019) The identity card of T cells—clinical utility of T-cell receptor repertoire analysis in transplantation. *Transplantation* 103(8):1544–1555. <https://doi.org/10.1097/TP.0000000000002776>
 32. Glanville J, Zhai W, Berka J, Telman D, Huerta G, Mehta GR et al (2009) Precise determination of the diversity of a combinatorial antibody library gives insight into the human immunoglobulin repertoire. *Proc Natl Acad Sci U S A* 106(48):20216–20221. <https://doi.org/10.1073/pnas.0909775106>
 33. Rubelt F, Busse CE, Bukhari SAC, Burckert JP, Mariotti-Ferrandiz E, Cowell LG et al (2017) Adaptive Immune Receptor Repertoire Community recommendations for sharing immune-repertoire sequencing data. *Nat Immunol* 18(12):1274–1278. <https://doi.org/10.1038/ni.3873>
 34. DeWitt WS 3rd, Smith A, Schoch G, Hansen JA, Matsen FA, Bradley P (2018) Human T cell receptor occurrence patterns encode immune history, genetic background, and receptor specificity. *eLife* 7. <https://doi.org/10.7554/eLife.38358>
 35. Szabo PA, Miron M, Farber DL (2019) Location, location, location: tissue resident memory T cells in mice and humans. *Sci Immunol* 4(34). <https://doi.org/10.1126/sciimmunol.aas9673>
 36. Kaestner KH, Powers AC, Naji A, Consortium H, Atkinson MA (2019) NIH initiative to improve understanding of the pancreas, islet, and autoimmunity in type 1 diabetes: the human pancreas analysis program (HPAP). *Diabetes* 68 (7):1394–1402. <https://doi.org/10.2337/db19-0058>
 37. Zhang J, Hu M, Wang B, Gao J, Wang L, Li L et al (2018) Comprehensive assessment of

- T-cell repertoire following autologous hematopoietic stem cell transplantation for treatment of type 1 diabetes using high-throughput sequencing. *Pediatr Diabetes* 19(7):1229–1237. <https://doi.org/10.1111/pedi.12728>
38. Seay HR, Yusko E, Rothweiler SJ, Zhang L, Posgai AL, Campbell-Thompson M et al (2016) Tissue distribution and clonal diversity of the T and B cell repertoire in type 1 diabetes. *JCI Insight* 1(20):e88242. <https://doi.org/10.1172/jci.insight.88242>
 39. Li H, Adamopoulos IE, Moulton VR, Stillman IE, Herbert Z, Moon JJ et al (2020) Systemic lupus erythematosus favors the generation of IL-17 producing double negative T cells. *Nat Commun* 11(1):2859. <https://doi.org/10.1038/s41467-020-16636-4>
 40. Liu X, Zhang W, Zhao M, Fu L, Liu L, Wu J et al (2019) T cell receptor beta repertoires as novel diagnostic markers for systemic lupus erythematosus and rheumatoid arthritis. *Ann Rheum Dis* 78(8):1070–1078. <https://doi.org/10.1136/annrheumdis-2019-215442>
 41. Thapa DR, Tonikian R, Sun C, Liu M, Dearth A, Petri M et al (2015) Longitudinal analysis of peripheral blood T cell receptor diversity in patients with systemic lupus erythematosus by next-generation sequencing. *Arthritis Res Ther* 17:132. <https://doi.org/10.1186/s13075-015-0655-9>
 42. Sui W, Hou X, Zou G, Che W, Yang M, Zheng C et al (2015) Composition and variation analysis of the TCR beta-chain CDR3 repertoire in systemic lupus erythematosus using high-throughput sequencing. *Mol Immunol* 67(2Pt B):455–464. <https://doi.org/10.1016/j.molimm.2015.07.012>
 43. Bashford-Rogers RJM, Bergamaschi L, McKinney EF, Pombal DC, Mescia F, Lee JC et al (2019) Analysis of the B cell receptor repertoire in six immune-mediated diseases. *Nature* 574(7776):122–126. <https://doi.org/10.1038/s41586-019-1595-3>
 44. Bertram HC, Check IJ, Milano MA (2001) Immunophenotyping large B-cell lymphomas. Flow cytometric pitfalls and pathologic correlation. *Am J Clin Pathol* 116(2):191–203. <https://doi.org/10.1309/BA3U-RMTU-D7UJ-M8DR>
 45. Barennes P, Quiniou V, Shugay M, Egorov ES, Davydov AN, Chudakov DM et al (2021) Benchmarking of T cell receptor repertoire profiling methods reveals large systematic biases. *Nat Biotechnol* 39(2):236–245. <https://doi.org/10.1038/s41587-020-0656-3>
 46. Kim SM, Bhonsle L, Besgen P, Nickel J, Backes A, Held K et al (2012) Analysis of the paired TCR alpha- and beta-chains of single human T cells. *PLoS One* 7(5):e37338. <https://doi.org/10.1371/journal.pone.0037338>
 47. Eugster A, Lindner A, Catani M, Heninger AK, Dahl A, Klemroth S et al (2015) High diversity in the TCR repertoire of GAD65 autoantigen-specific human CD4+ T cells. *J Immunol* 194(6):2531–2538. <https://doi.org/10.4049/jimmunol.1403031>
 48. Ziegenhain C, Vieth B, Parekh S, Reinus B, Guillaumet-Adkins A, Smets M et al (2017) Comparative analysis of single-cell RNA sequencing methods. *Mol Cell* 65(4):631–643.e634. <https://doi.org/10.1016/j.molcel.2017.01.023>
 49. Afik S, Yates KB, Bi K, Darko S, Godec J, Gerdemann U et al (2017) Targeted reconstruction of T cell receptor sequence from single cell RNA-seq links CDR3 length to T cell differentiation state. *Nucleic Acids Res* 45(16):e148. <https://doi.org/10.1093/nar/gkx615>
 50. Stubbington MJT, Lonnberg T, Proserpio V, Clare S, Speak AO, Dougan G et al (2016) T cell fate and clonality inference from single-cell transcriptomes. *Nat Methods* 13(4):329–332. <https://doi.org/10.1038/nmeth.3800>
 51. Redmond D, Poran A, Elemento O (2016) Single-cell TCRseq: paired recovery of entire T-cell alpha and beta chain transcripts in T-cell receptors from single-cell RNAseq. *Genome Med* 8(1):80. <https://doi.org/10.1186/s13073-016-0335-7>
 52. Eltahlia AA, Rizzetto S, Pirozyan MR, Betz-Stablein BD, Venturi V, Kedzierska K et al (2016) Linking the T cell receptor to the single cell transcriptome in antigen-specific human T cells. *Immunol Cell Biol* 94(6):604–611. <https://doi.org/10.1038/icb.2016.16>
 53. DeKosky BJ, Kojima T, Rodin A, Charab W, Ippolito GC, Ellington AD et al (2015) In-depth determination and analysis of the human paired heavy- and light-chain antibody repertoire. *Nat Med* 21(1):86–91. <https://doi.org/10.1038/nm.3743>
 54. Klein AM, Macosko E (2017) InDrops and Drop-seq technologies for single-cell sequencing. *Lab Chip* 17(15):2540–2541. <https://doi.org/10.1039/c7lc90070h>
 55. Zemmour D, Zilionis R, Kiner E, Klein AM, Mathis D, Benoist C (2018) Single-cell gene expression reveals a landscape of regulatory T cell phenotypes shaped by the TCR. *Nat Immunol* 19(3):291–301. <https://doi.org/10.1038/s41590-018-0051-0>

56. Gerard A, Woolfe A, Mottet G, Reichen M, Castrillon C, Menrath V et al (2020) High-throughput single-cell activity-based screening and sequencing of antibodies using droplet microfluidics. *Nat Biotechnol* 38(6):715–721. <https://doi.org/10.1038/s41587-020-0466-7>
57. 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(12):2257–2317. <https://doi.org/10.1038/sj.leu.2403202>
58. Scheijen B, Meijers RWJ, Rijntjes J, van der Klift MY, Mobs M, Steinhilber J et al (2019) Next-generation sequencing of immunoglobulin gene rearrangements for clonality assessment: a technical feasibility study by EuroClonality-NGS. *Leukemia* 33(9):2227–2240. <https://doi.org/10.1038/s41375-019-0508-7>
59. Perry RP, Kelley DE, Coleclough C, Kearney JF (1981) Organization and expression of immunoglobulin genes in fetal liver hybridomas. *Proc Natl Acad Sci U S A* 78(1):247–251. <https://doi.org/10.1073/pnas.78.1.247>
60. Carlson CS, Emerson RO, Sherwood AM, Desmarais C, Chung MW, Parsons JM et al (2013) Using synthetic templates to design an unbiased multiplex PCR assay. *Nat Commun* 4:2680. <https://doi.org/10.1038/ncomms3680>
61. Friedensohn S, Lindner JM, Cornacchione V, Iazeolla M, Miho E, Zingg A et al (2018) Synthetic standards combined with error and bias correction improve the accuracy and quantitative resolution of antibody repertoire sequencing in human naive and memory B cells. *Front Immunol* 9:1401. <https://doi.org/10.3389/fimmu.2018.01401>
62. Brard F, Shannon M, Luning Prak ET, Litwin S, Weigert M (1999) Somatic mutation and light chain rearrangement generate autoimmunity in anti-single-stranded DNA transgenic MRL/lpr mice. *J Exp Med* 190(5):691–704. <https://doi.org/10.1084/jem.190.5.691>
63. Hieter PA, Korsmeyer SJ, Waldmann TA, Leder P (1981) Human immunoglobulin kappa light-chain genes are deleted or rearranged in lambda-producing B cells. *Nature* 290(5805):368–372. <https://doi.org/10.1038/290368a0>
64. Rosenfeld AM, Meng W, Chen DY, Zhang B, Granot T, Farber DL et al (2018) Computational evaluation of B-cell clone sizes in bulk populations. *Front Immunol* 9:1472. <https://doi.org/10.3389/fimmu.2018.01472>
65. Japp AS, Meng W, Rosenfeld AM, Perry DJ, Thirawatananond P, Bacher RL et al (2021) TCR(+)/BCR(+) dual-expressing cells and their associated public BCR clonotype are not enriched in type 1 diabetes. *Cell* 184(3):827–839.e814. <https://doi.org/10.1016/j.cell.2020.11.035>
66. Ye J, Ma N, Madden TL, Ostell JM (2013) IgBLAST: an immunoglobulin variable domain sequence analysis tool. *Nucleic Acids Res* 41(Web server issue):W34–W40. <https://doi.org/10.1093/nar/gkt382>
67. MacConaill LE, Burns RT, Nag A, Coleman HA, Slevin MK, Giorda K et al (2018) Unique, dual-indexed sequencing adapters with UMIs effectively eliminate index cross-talk and significantly improve sensitivity of massively parallel sequencing. *BMC Genomics* 19(1):30. <https://doi.org/10.1186/s12864-017-4428-5>
68. Shuldiner AR, Nirula A, Roth J (1989) Hybrid DNA artifact from PCR of closely related target sequences. *Nucleic Acids Res* 17(11):4409. <https://doi.org/10.1093/nar/17.11.4409>

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

