The past, present, and future of immune repertoire biology – the rise of next-generation repertoire analysis

Adrien Six [1,2,3,4,5] *[†], Maria Encarnita Mariotti-Ferrandiz [1,2,3,5], Wahiba Chaara [1,2,3,4,5], Susana Magadan [6], Hang-Phuong Pham [1,2], Marie-Paule Lefranc [7], Thierry Mora [8], Véronique Thomas-Vaslin [1,2,3,5], Aleksandra M. Walczak [9] and Pierre Boudinot [6][†]

[1] UPMC University Paris 06, UMR 7211, Immunology-Immunopathology-Immunotherapy (I3), Paris, France
[2] CNRS, UMR 7211, Immunology-Immunopathology-Immunotherapy (I3), Paris, France
[3] INSERM, UMR_S 959, Immunology-Immunopathology-Immunotherapy (I3), Paris, France
[4] AP-HP, Hôpital Pitié-Salpêtrière, CIC-BTi Biotherapy, Paris, France
[5] AP-HP, Hôpital Pitié-Salpêtrière, Département Hospitalo-Universitaire (DHU), Inflammation-Immunopathology-Biotherapy (i2B), Paris, France
[6] Institut National de la Recherche Agronomique, Unité de Virologie et Immunologie Moléculaires, Jouy-en-Josas, France
[7] IMGT®, The International ImMunoGeneTics Information System®, Institut de Génétique Humaine, UPR CNRS 1142, Université Montpellier 2, Montpellier, France
[8] Laboratoire de Physique Statistique, UMR8550, CNRS and Ecole Normale Supérieure, Paris, France
[9] Laboratoire de Physique Théorique, UMR8549, CNRS and Ecole Normale Supérieure, Paris, France

T and B cell repertoires are collections of lymphocytes, each characterized by its antigen-specific receptor. We review here classical technologies and analysis strategies developed to assess immunoglobulin (IG) and T cell receptor (TR) repertoire diversity, and describe recent advances in the field. First, we describe the broad range of available methodological tools developed in the past decades, each of which answering different questions and showing complementarity for progressive identification of the level of repertoire alterations: global overview of the diversity by flow cytometry, IG repertoire descriptions at the protein level for the identification of IG reactivities, IG/TR CDR3 spectratyping strategies, and related molecular quantification or dynamics of T/B cell differentiation. Additionally, we introduce the recent technological advances in molecular biology tools allowing deeper analysis of IG/TR diversity by next-generation sequencing (NGS), offering systematic and comprehensive sequencing of IG/TR transcripts in a short amount of time. NGS provides several angles of analysis such as clonotype frequency, CDR3 diversity, CDR3 sequence analysis, V allele identification with a quantitative dimension, therefore requiring high-throughput analysis tools development. In this line, we discuss the recent efforts made for nomenclature standardization and ontology development. We then present the variety of available statistical analysis and modeling approaches developed with regards to the various levels of diversity analysis, and reveal the increasing sophistication of those modeling approaches. To conclude, we provide some examples of recent mathematical modeling strategies and perspectives that illustrate the active rise of a "next-generation" of repertoire analysis.

Keywords: diversity analysis, immune receptors, next-generation sequencing, modeling, statistics, gene nomenclature, B cell repertoire, T cell repertoire

INTRODUCTION

T and B cell repertoires are collections of lymphocytes, each characterized by its antigen-specific receptor. The resources available to generate the potential repertoires are described by the genomic T cell receptor (TR) and immunoglobulin (IG) loci. TR and IG are produced by random somatic rearrangements of V, D, and J genes during lymphocyte differentiation. The product of the V-(D)-J joining, called the complementarity determining region 3 (CDR3) and corresponding to the signature of the rearrangement, binds the antigen and is responsible for the specificity of the recognition. During their differentiation, lymphocytes are subjected to selective processes, which lead to deletion of most auto-reactive cells, selection, export, and expansion, of mature T and B cells to the periphery. Primary IG and

TR repertoires are therefore shaped to generate the available peripheral or mucosal repertoires. In addition, several different functional T and B cells subsets have been identified, with differential dynamics and antigen-specific patterns. These available repertoires are dramatically modified during antigen-driven responses especially in the inflammatory context of pathogen infections, autoimmune syndromes, and cancer to shape actual repertoires. When considering the importance of efficient adaptive immune responses to get rid of infections naturally or to avoid auto-reactive damages, but also for therapeutic purposes such as vaccination or cell therapy, one realizes the relevance of understanding how lymphocyte repertoires are selected during differentiation, from ontogeny to aging, and upon antigenic challenge. However, immune repertoires of expressed antigen receptors are

built by an integrated system of genomic recombination and controlled expression, and follow complex time-space developmental patterns. Thus, an efficient repertoire analysis requires both (1) methods that sample and describe the diversity of receptors at different levels for an acceptable cost and from a little amount of material and (2) analysis strategies that reconstitute the best multidimensional picture of the immune diversity from the partial information provided by the repertoire description as reviewed in Ref. (1). In the following sections, we summarize technologies developed over the past decades to describe lymphocyte repertoires and we present the growing number of analysis tools, evolving from basic to sophisticated statistics and modeling strategies with regards to the level of complexity of the data produced.

METHODS DEVELOPED TO DESCRIBE THE IG AND TR REPERTOIRES

B and T lymphocyte repertoires can be studied from different lymphoid tissues and at various biological levels, such as cell membrane or secreted proteins, transcripts or genes, according to the techniques used. Fluorescence microscopy or flow cytometry techniques allow to track and sort particular cell phenotypes and to quantify the expressed repertoire at the single-cell level with V subgroup-specific monoclonal antibodies. Alternatively, the IG or TR diversity may be also analyzed using proteomics methods from either the serum (for IG) or dedicated cell extracts. Finally, molecular biology techniques assess the repertoire at the genomic DNA or transcriptional levels, qualitatively and/or quantitatively.

ANALYSIS OF IG AND TR REPERTOIRES AT THE PROTEIN LEVEL
Flow cytometry single-cell repertoire analysis
The frequency of lymphocytes expressing a given IG or TR can be determined using flow cytometry when specific monoclonal antibodies are available. This technique allows for the combined analysis of the antigen receptor and of other cell surface markers. Currently, using flow cytometry, up to 13 parameters can be routinely studied at once, reaching 20 parameters with the last generation flow cytometers and 70–100 parameters with mass cytometry (2). Seminal studies in mice using specific anti-TRBV antibodies have led to the characterization of the central tolerance selection processes that occur in thymus (3–5). Later on, a comprehensive description of the human TRBV repertoire was setup (6), when monoclonal antibodies became available for most of the TRBV subgroups. Repertoire analysis with flow cytometry provides a qualitative and quantitative analyses of the variable region, often done on heterogeneous cell populations, in order to decipher, for example, selection events related to aging, perturbations, and treatments (7). However, this technology is naturally limited by the availability of specific monoclonal antibodies, and does not address more detailed issues such as junction diversity. Furthermore, polymorphism of the IG or TR genes (8, 9) may constitute a serious limitation for a systematic survey using these approaches.

Proteomic repertoire analysis for serum immunoglobulins
Recent developments of proteomics tools now offer sensitivity levels applicable to IG repertoire analysis. Such a description at the protein level takes into account all post transcriptional and translational modifications.

PANAMA-blot technology.   A semi-quantitative immunoblot, called the PANAMA-blot technique (10), allows for the identification of the antibody reactivities present in collection of sera (or cell culture supernatant) against a given source of antigens (10–12). Briefly, a selected source of antigens is subjected to preparative SDS-PAGE, transferred onto nitrocellulose membranes, then incubated with the serum to be tested allowing for the revelation of the bound antibodies using an appropriate secondary antibody coupled to alkaline phosphatase. Computer-assisted analysis of the densitometric profiles allows for the rescaling and the quantitative comparison of patterns of antibody reactivity from individuals in different groups. A large amount of data is generated when testing a range of sera against various sources of antigens. Statistical analyses are included in the PANAMA-Blot approach (as described further). This global analysis helped to reveal that the IgM repertoire in mice is selected by internal ligands and independent of external antigens (13).

This method can also lead to identify IG reactivity patterns specific for a type of pathology or clinical status and has been applied to both fundamental and clinical analysis. In particular, it was used to analyze human self-reactive antibody repertoires and their potential role for down-modulating autoimmune processes (14–16).

Antigen micro-array chips.   More recently, antigen micro-array-based technology coupled to a complex two-way clustering bioinformatics analysis was developed to evaluate the serum repertoire antibodies from diabetes-prone individuals and revealed their predictive or diagnostic value. In brief, a range of antigens (proteins, peptides, nucleotides, phospholipids...) were plated onto glass plates and incubated with sera from individuals (human diabetes patients or mice in an experimental model of diabetes). The intensity of reactivity of the serum IG for each peptide was determined and scored against the control reactivity. Clustering analysis was then implemented to determine a potential antigen signature that significantly sorts out diabetes from non-diabetes individuals. In this way, it was found that the patterns of IgG antibodies expressed early in male NOD mice can mark susceptibility or resistance to diabetes induced later and that it is different than the pattern characteristic of healthy or diabetic mice after disease induction (17). Similarly, this clustering approach was applied in humans to successfully separate human subjects that are already diabetic from healthy people (18).

REPERTOIRE ANALYSIS AT THE GENOMIC DNA LEVEL
Other strategies that cover IG or TR repertoire analyses have been developed at the genomic DNA level. Firstly, CDR3 spectratyping studies (detailed in the following section) have been carried out at the DNA level mostly to address issues related to B or T cell development (19, 20). More recently, an original multiplex genomic PCR assay coupled to real-time PCR analysis was developed to provide a comprehensive description of the mouse T cell receptor alpha (TRA) repertoire during development (21). Although these approaches can be applied to all IG isotypes and TR, they have not been used as much as transcript CDR3 spectratyping due to sensitivity and heterozygosity issues.

Immunoglobulin or T cell receptor repertoires can also be assessed by following the diversity of rearrangement deletion circles. Since they are produced by the V-(D)-J recombination machinery when the joint signal is formed and diluted in daughter cells, they give a good representation of recently generated T or B cells. This technique has been particularly useful for describing the restoration of T cell diversity following highly active antiretroviral therapy in HIV-infected patients (22) and has been used to model thymic export (23, 24) as well as to demonstrate continued contribution of the thymus to repertoire diversity, even in older individuals (25). It also reveals that thymic output is genetically determined, and related to the extent of proliferation of T cells at DN4 stage in mice (26). However, their analysis does not provide much insight into the level of diversity since the signal joint does not vary for a given combination of genes. Therefore, the interest of such analyses is reached when combined with CDR3 spectratyping analyses to know whether a repertoire perturbation is rather attributable to newly produced T cells or peripheral T cell proliferation.

V-(D)-J JUNCTION ANALYSIS OF IG AND TR TRANSCRIPT REPERTOIRES

Original molecular-based strategies for analyzing repertoire diversity relied on cloning and hybridization of molecular probes specific for IGHV gene subgroups first by RNA colony blot assay (27). This led to the observation that IGHV gene usage is characteristic of mouse strain and is a process of random genetic combination by equiprobable expression of IGHV genes (28). The study of selection processes revealed that the IGHV region-dependent selection determines clonal persistence of B cells (29) and that selection with age leads to biased IGHV gene expression (30).

In situ hybridization on single-cells revealed that during mouse ontogeny and early development of B cells in bone marrow, there is a non-random position-dependent IGHV gene expression, favoring D-proximal IGHV gene subgroup usage (31). Thereafter, sequencing of PCR-amplified cDNA collections were obtained from samples of interest. Although fastidious, these early studies have been useful in defining the basis of human IG and TR repertoires in terms of overall distribution, CDR3-length distribution, and V-(D)-J use (32–35), sometimes leading to the identification of new IG or TR genes. Later, more practical techniques have been developed for large-scale analysis of lymphocyte repertoires, such as quantitative PCR, micro-array, and junction length spectratyping, as described below.

Quantitative RT-PCR for repertoire analysis

In parallel to qualitative CDR3 spectratyping techniques (see section below), quantitative PCR strategies were developed (36). Coupling the two techniques for all V domain-C region combinations provides a complete qualitative and quantitative picture of the repertoire (37–39) described by up to 2,000 measurements per IG isotype or TR for one sample. With the development of real-time quantitative PCR, this approach opened the possibility for a more precise evaluation of repertoire diversity (39–41). Complementary tools have been also developed in order to allow normalization of spectratype analysis such as studies by Liu et al. (42) and Mugnaini et al. (43).

Matsutani et al. (44) developed another method to quantify the expression of the human TRAV and TRBV repertoires based on hybridization with gene specific primers coated plates. The cDNA from PBMC extracted RNA are ligated to a universal adaptor which allows for a global amplification of all TRAV or TRBV cDNAs. The PCR products are then transferred onto microplates coated with oligonucleotides specific for each TRAV or TRBV regions, and the amount of hybridized material is quantified. This technique was used to analyze the TR repertoire diversity of transplanted patients (45) and adapted to the study of mouse TRAV and TRBV repertoires (46). VanderBorght et al. also developed a semi-quantitative PCR-ELISA-based method for the human TRAV and TRBV repertoire analysis (38). The combined usage of digoxigenin (DIG)-coupled nucleotides and DIG-coupled reverse TRAC or TRBC primers allowed for a quantitative measurement of the amount of amplified DNA by a sandwich ELISA.

Du et al. (47) later setup a megaplex PCR strategy to characterize the antigen-specific TRBV repertoire from sorted IFNγ-producing cells after *Mycobacterium* infection. The clonotypic TRBV PCR products were used for Taqman probes design to quantify the expression of the corresponding clonotypes from ATLAS-amplified SMART cDNAs.

Direct measurement of lymphocyte diversity using micro-arrays

Another technology, similar to the one just discussed, has been developed by the group of Cascalho et al. which allows for a direct measurement of the entire population of lymphocyte-receptors. This is accomplished by hybridization of lymphocyte-receptor specific cRNA of a lymphocyte population of interest to random oligonucleotides on a gene chip; the number of sites undergoing hybridization corresponds to the level of diversity. This method was validated and calibrated using control samples of random oligonucleotides of known diversity ($1, 10^3, 10^6, 10^9$) (48, 49) and successfully demonstrated that central and peripheral diversification of T lymphocytes is dependent on the diversity of the circulating IG repertoire (49, 50). Similarly, a highly sensitive micro-array-based method has been proposed to monitor TR repertoire at the single-cell level (51).

CDR3 spectratyping techniques

Immunoscope technology. Among various techniques used to analyze the T or B cell repertoires, Immunoscope, also known as CDR3 spectratyping (52, 53) consists in the analysis of the CDR3-length usage so that antigen-specific receptor repertoires can be described by thousands of measurements. In the case of naive murine repertoires, T cell populations are polyclonal and analysis typically yields eight-peak regular bell-shaped CDR3 displays (wrongly assumed to be Gaussian), each peak corresponding to a given CDR3-length. When an immune response occurs, this regular polyclonal display can be perturbed: one can see one or several prominent peaks that correspond to the oligoclonal or clonal expansion of lymphocytes. A complete description of this technique and its applications to clinical studies has been published elsewhere (54).

In the original Immunoscope publication, Cochet et al. (55) analyzed the T cell repertoire after the immunization of mice with the pigeon cytochrome c. They provided the first description of

an *ex vivo* follow-up of a primary T cell specific response in a mouse model. Their second paper analyzed the average CDR3-lengths as a function of TRBV-TRBJ combinations. In particular, the authors found a correlation between TRBV CDR1 and major histocompatibility (MH) haplotype (52). This group later published a large amount of original studies in various models such as lymphocyte development (40, 56–63), kinetics of antigen-specific responses (64–67), viral infection (68, 69), autoimmunity (70, 71), tumor-associated disease (72), and analysis of allogeneic T cell response and tolerance after transplantation (73). Notably, the combination of CDR3 spectratyping with flow cytometry-based IG or TR V frequency analysis provides a more comprehensive assessment, such as in Pilch et al. (74). For example, such an approach revealed the constriction of repertoire diversity through age-related clonal CD8 expansion (75). Similarly, a combination of CDR3 spectratyping, flow cytometry, and TR deletion circle analysis has allowed to define age-dependent incidence on thymic renewal in patients (76) or to evaluate the effects of caloric restriction in monkeys to preserve repertoire diversity (77). CDR3-length spectratyping was also used in other models, such as rainbow trout, to analyze TRB repertoire and its modifications induced by viral infection (78–80). While no tool such as monoclonal antibodies to T cell marker(s) was available in this model, this approach demonstrated that fish could mount specific T cell responses against virus, which could be found in all individuals (public clonotypes) or not (private clonotypes). Similar strategies, developed by other groups (81) and following the same approach in parallel, analyzed the IG repertoire in *Xenopus* at different stages of development, describing a more restricted IG junction diversity in the tadpole compared to the adult.

Gorski et al. (82) developed their own CDR3 spectratyping technique to analyze the complexity and stability of circulating αβT cell repertoires in patients following bone marrow transplantation as compared to normal adults. They showed that repertoire complexity of bone marrow recipients correlates with their state of immune function; in particular, individuals suffering from recurrent infections associated with T cell impairment exhibited contractions and gaps in repertoire diversity. The detailed procedure for this technique has been published in Maslanka et al. (83). A variation of this technique has been reported later by Lue et al. (84), relying on a compact glass cassette, a simpler device than the usual automated plate DNA sequencers.

Alternative technologies. Alternative CDR3 spectratyping techniques have been described such as single-strand conformation polymorphism (85–87) and heteroduplex analysis (88–91). These methods differ from the CDR3 spectratyping/Immunoscope technique mostly in the way PCR products are analyzed by performing non-denaturing polyacrylamide electrophoresis. The main advantage of these techniques is a more direct assessment of clonal expansion since PCR products migrate according to their conformation properties; therefore, presence of a predominant peak is strongly indicative of clonality when a smear migration pattern indicates polyclonality. However, these techniques have been less widely used probably because of the difficulty to make clear correlations between the expanded peaks across samples.

Another original alternative technique has been described by Bouffard et al. (92), analyzing products obtained after *in vitro* translation of PCR-amplified TR-specific products by isoelectric focusing. With this technique, clonality can also directly be assessed by looking at the obtained migration profile.

IG/TR REARRANGEMENT SEQUENCING: FROM CLONING-BASED- TO NEXT-GENERATION-SEQUENCING

In order to get a better description of IG/TR diversity at the nucleotide sequence level, thus providing fine-tuned description of the actual diversity, Sanger sequencing approaches relying on bacterial cloning of rearrangements were performed in physiological conditions globally (60, 93–99) or partially to characterize particular expansions identified by other technologies such as CDR3 spectratyping (40, 59, 100–102), flow cytometry (103). They were also used in pathological/infectious conditions (104–107) sometimes leading to antigen-specific T cell TR identification and quantification through the combination of antigen-specific T cell stimulation and cytometry-based cell sorting, anchor-PCR, and bacterial cloning-based sequencing (108).

These studies pioneered the description of the repertoire and provided fruitful information regarding the extent and modification of the diversity. However, besides being time and cost-extensive, such approaches have allowed for the analysis of 10^2–10^3 sequences, far under the estimated diversity reaching 10^6–10^7 unique clonotypes in mice and humans (40, 59, 109).

In the last decade, DNA sequencing technologies have made tremendous progresses (110) with the development of so called next-generation sequencers, already reaching four generations (111). Those instruments are designed to sequence mixtures of up to millions of DNA molecules simultaneously, instead of individual clones separately. Second generation sequencers became affordable in the last 5 years and have been used for immune repertoire analysis, starting with the seminal work of Weinstein et al. (112) where the IG repertoire of Zebrafish has been described by large-scale sequencing. Consequently, exploratory works by other groups provided an overview of the complex sequence landscape of immune repertoires in humans (113–118). More recent work aimed at addressing fundamental questions such as lineage cells commitment (119–122), generation of the diversity processes (123–125), and diversity sharing between individuals (126, 127). Finally, the power of this technology has been validated in the clinic as well (128, 129).

As seen above for other technologies, combinations of approaches have been applied to NGS. Notably, deep sequencing has been used in combination with CDR3-length spectratyping by some groups to study human (130) or rainbow trout IG (131) repertoire modifications after vaccination against bacteria or viruses. In the latter, pyrosequencing performed for relevant VH/Cμ or VH/Cτ junctions identified the clonal structure of responses, and showed, for example, that public responses are made of different clones identified by (1) distinct V-(D)-J junctions encoding the same protein sequence or (2) distinct V-(D)-J sequences differing by one or two conservative amino acid changes (131) as described for public response in mammals (132, 133). These studies showed that NGS and traditional spectratyping techniques lead to remarkably similar CDR3 distributions.

Several NGS have been developed in the past years using different sequencing technologies characterized with different speed, deepness and read length. Metzker thoroughly reviewed their principles and properties (134). Among them, three platforms, all offering benchtop sequencers with reduced cost and setup, fit with immune repertoire analysis in terms of read length and deepness. The 454/Roche platform uses pyrosequencing technology (135), which combines single nucleotide addition (SNA) with chemoluminescent detection on templates that are clonally amplified by emulsion PCR and loaded on a picotiter plate. Pyrosequencing currently has a 500 bp (GS Junior) to 700 bp (GS FLX) sequencing capacity with a respective deepness of 150,000–3,000,000 reads per run (134). The Illumina/Solexa platform technology is based on cyclic reversible termination (CRT) sequencing (an adaptation of Sanger sequencing) performed on templates clonally amplified on solid-phase bridge PCR. Protected fluorescent nucleotides are added, imaged, delabeled, and deprotected cyclically (134), providing a deeper sequencing (from 15 to 6 billion reads per run for the MiSeq to the HiSeq2500/2000) of shorter reads (100–250 bp for the very recent MiSeq) with the possibility to perform pair-end sequencing (two-side sequencing) to increase the read length after aligning the generated complementary sequences. A more recent platform, Ion Torrent/Life Technologies using an imaging free detection system may open a new era in terms of deepness (one billion reads per run) of 200 bp reads (136) in a very short time and on a benchtop sequencer. Importantly, depending on the technology, errors due to the PCR-based sample preparation and the sequencing are of major concern. Bolotin et al. (137) evaluated this issue on TR repertoire analysis of the same donor performed on the three platforms described previously; algorithms for error correction have been developed. Indeed, PCR- and sequencing-related errors represent the major concern for immune repertoire diversity analysis as they may generate artificial diversity. Illumina and 454 appear to be the most robust technologies, with Illumina having the highest throughput and 454 generating the longest reads. The currently available Ion Torrent platform, although very promising, has been shown to display the highest rate of errors in TR (137) and bacterial DNA (138) sequencing. However, such error corrections must be used with caution since they may inadvertently underestimate repertoire diversity by removing rare sequences.

With the power of such approach for genomics and transcriptomics studies in general, constant improvements are achieved to increase the sequencing deepness and read length as well as to reduce the cost, therefore offering multitude of biological explorations (139). NGS now permits a comprehensive and quantitative view of IG and TR diversity by combining and improving the sensitivity of classical approaches with accurate and large-scale sequencing. NGS has the power to identify IG or TR specific for given antigens (in combination with antigen-specific assays) and to define more complex signatures (i.e., TR sets) related to disease and/or treatment from heterogeneous T and B cell populations. Still, most of the deep sequencing efforts have been limited to only one chain of the receptor at the repertoire level (usually the β chain for TR and the heavy chain for IG). Indeed, current high-throughput approaches do not allow one to assign which combination of chains (TRA and TRB, or IGH and IGK

or IGL) belong to which cell (140). A recent development by DeKosky et al. proposed a reasonably high-throughput technology to assess massively paired IG VH and VL from bulk population (141). In parallel, Turchaninova et al. (142) have proposed a similar approach for the paired analysis of the TRA and TRB chains. The parallel development of high-throughput microfluidic-based single-cell sorting will certainly push forward new developments in the field (143).

However, despite the technological advance, studies so far have mainly reported CDR3 counting and identification of major expansions. The complexity of immune repertoires is still a matter that such approach cannot completely overcome, due to the paucity of powerful analytical methods. Besides data management tools, studies are now starting to extract most of the benefit from such approach to model the immune repertoire diversity and dynamics (144), an approach that may help in understanding the interplay between cells and repertoire shaping. Accurate and powerful statistical analyses are required to manage such amount of information. Current state will be reviewed in the following sections.

POTENTIAL AND GENOMIC REPERTOIRES: A QUESTION OF ONTOLOGY AND ORTHOLOGY

Immune repertoires *sensu stricto* are expressed by lymphocyte clones, each carrying a single receptor for the antigen. Such receptors comprise IG and TR in jawed vertebrates (8, 9) and VLR in Agnathans (145). The sequences of these receptors are available in databases such as GenBank or EMBL, which are difficult to use for transversal studies due to inconsistent annotation. The IMGT® information system (see below) has largely solved this problem setting standardized gene nomenclatures, ontologies and a universal numbering of the IG/TR V and C domains, thus giving a common access to standardized data from genome, proteome, genetics, two-dimensional, and three-dimensional structures (146). The accuracy and the consistency of the IMGT® data are based on IMGT-ONTOLOGY, the first, and so far, unique ontology for immunogenetics and immunoinformatics (147).

With the development of high-throughput sequencing, large numbers of new sequences of antigen receptor genes have become available, which can be classified into different categories: genomic sequences of IG or TR (in germline configuration in genome assemblies) or fragments of IG/TR transcripts, containing the CDR3 or not. Also, these datasets can be produced from species newly sequenced, as well as from new haplotypes of well-described species.

The annotation of such sequences remains an open question. Manual annotation is not applicable, and no good automated approach has been validated yet. A relevant annotation of these massive datasets will require the integration of genomic and expression data with existing standardized description charts, as offered by IMGT®. A standardized annotation is an important issue since it facilitates the re-utilization of datasets and comparison of analyses. Thus, the description of IG and TR polymorphisms, the integration of repertoire studies with structural features of antigen-specific domains, and even the usage of new genes in genetic engineering rely on a common standard for nomenclature, numbering, and annotation (147).

To take advantage of the current standards that have been established from classical sequencing data during the last 25 years, new, fast, reliable, and human-supervised annotation methods will have to be developed, integrating directly high-throughput sequence information from the increasing number of deep sequencing platforms and technologies, at different genetic levels (genome, transcriptome, clonotype repertoires). Along this line, IMGT/HighV-QUEST offers online tools to the scientific community for the analysis of long IG and TR sequences from NGS (148).

Special attention can be paid to the orthology/paralogy relationships between similar antigen receptor genes from different species. These characteristics are essential to understand the dynamics of IG and TR loci. In fact, with many important lymphocyte subsets characterized by canonical/invariant antigen receptors, such relationships are critical to transfer functional knowledge between models. Importantly, the phylogenetic analyses required to reconstitute the evolution of antigen receptor genes are based on multiple alignments, the quality of which is highly dependent on common numbering and precise annotation of sequences.

As far as immune repertoires are characterized by the diversity of receptors specifically binding antigen/pathogen motifs to initiate a defense response, they might not be limited to lymphocyte diversifying receptors, e.g., IG, TR, and VLR. The particularity of these systems is a somatic diversification combined to a clonal structure of the repertoire, each lymphocyte clone expressing the product of a recombination/hypermutation and/or conversion process. However, many other arrays of diverse receptors binding or sensing pathogens have been discovered in metazoans, in invertebrates as well as in vertebrates.

In some cases, their diversity is really "innate," i.e., encoded in the genome as multiple genes produced by duplications. Fish NLR, finTRIMs, and NITR, primate KIR, chicken CHIR, or TLR in sea urchin, constitute good examples of such situations. While these repertoires may appear as relatively limited, polymorphism within populations, and differential expression of receptors per cell upon stimulation represent complex issues, which fall well into "traditional" repertoire approaches.

In other cases, receptors are subject to diversification processes much faster than gene duplication, which does not comply with a clonal selection pattern. The best examples are probably the DSCAM in arthropods, which hugely diversify by alternative splicing of exons encoding half-IgSF domains (149, 150), and the FREP lectins in mollusks, of which sequences are highly variable at the population level, and even between parents and offspring produced by auto-fecundation (151).

The number of such "innate" repertoires which are not expressed by clonally selected lymphocytes will likely increase with deep sequencing of new genomes/transcriptomes, as illustrated by a recent report from mussel (152). A good example of the importance of a proper structural description of key domains of receptors is provided by the extensive analysis of LRR motifs in studies on TLR evolution (153, 154). Further insights into the functions of such diverse proteins will be provided by the characterization of their expressed (available) repertoire, at different levels such as single-cells, cell populations, and animal populations.

Such analyses will require precise identification of genes and sequences as well as mutations, and a standardized approach of nomenclature and structural description will be as useful as it is for the vertebrate IG and TR sequences. Importantly, these receptors are made of a small number of structural units, such as IgSF domain or LRR domains, which suggests that standardized system(s) for sequence annotation could be developed following IMGT standards (155).

STATISTICAL ANALYSIS AND MODELING OF IMMUNE REPERTOIRE DATA

STATISTICAL REPERTOIRE ANALYSIS

The description of the repertoire modifications using flow cytometry or Immunoscope provided clear-cut and detailed insight into the clonal expansion processes during the responses against a defined antigen (64, 66). However, it is difficult to identify the relevant alterations of the repertoires in more complex situations such as pathogen infections or variable genetic backgrounds. For example, it appeared impossible to identify all significant modifications of TRB Immunoscope profiles during cerebral malaria by direct ocular comparison (107). Different methods were therefore developed to extract from IG and TR repertoire descriptions the relevant information, to encode it as numerical tables and to analyze them with statistical models.

CDR3 spectratype perturbation indices

Since the initial description of the CDR3 spectratyping technique, different scoring indices were developed or derived from the literature: "relative index of stimulation" (RIS) (55), "overall complexity score" (156), Reperturb (157), "complexity scoring system" (158), COPOM (159), Oligoscore (160), TcLandscape (161), "spectratype diversity scoring system" (162), Morisita-Horn index and Jaccard index (95–97), "absolute perturbance value" (163). A comparative review of such scoring strategies was published by Miqueu et al. (164).

In particular, the perturbation index Reperturb was developed by Gorochov et al. to perform TR repertoire analysis in HIV patient during progression to AIDS and under antiretroviral therapy. They could show drastic restrictions in the CD8[+] T cell repertoire at all stages of natural progression that persisted during the first 6 months of treatment. In contrast, CD4[+] T cell repertoire perturbations correlated with progression to AIDS with a return to a diversified repertoire in good responders to treatment (157).

Soulillou et al. refined this approach by combining the qualitative information obtained with usual CDR3 spectratyping with quantitative information of TRBV usage obtained by real-time quantitative PCR. They devised a four-dimension representation that represents TRBV subgroups, CDR3-length and percentage of TRBV use on three axis chart in addition to a color-coded representation of the CDR3 profile perturbation. Using this original approach, they were able to show that graft rejection is associated with a vigorous polyclonal accumulation of TRBV mRNA among graft-infiltrating T lymphocytes, whereas in tolerated grafts T cell repertoire is strongly altered (161, 165). Their study puts the emphasis on the importance of not only qualitative but also quantitative analysis of lymphocyte repertoires.

Platforms for repertoire data management and statistical analysis

Several platforms have been developed and rely mostly on CDR3 spectratyping and sequencing data, with recent developments to manage and analyze NGS data.

The ISEApeaks strategy and software were developed in order to satisfy the needs for efficient automated electrophoresis data retrieval and management (160, 166). ISEApeaks extracts peak area and length data generated by software used to determine fragment intensity and size. CDR3 spectratype raw data, consisting of peak areas and nucleotide lengths for each V-(D)-J-C combination, is extracted, smoothed, managed, and analyzed. The repertoires of different samples are gathered in a peak database and CDR3 spectratypes can be analyzed by different perturbation indices and multivariate statistical methods implemented in ISEApeaks. We have applied our ISEApeaks strategy in several studies. In an experimental model of cerebral malaria, we established a correlation between the quality of TR repertoire alterations and the clinical status of infected mice, whether they developed cerebral malaria or not (107). We contributed to the characterization of the membrane-associated *Leishmania* antigens (MLA) that stimulates a large fraction of naive CD4 lymphocytes. Repertoire analyses showed that MLA-induced T cell expansions used TR with various TRBV rearrangements and CDR3 lengths, a feature closer to that of polyclonal activators than of a classic antigen (167). We also revealed repertoire age-related perturbations in mice (7). ISEApeaks functions for statistical analysis was successfully applied to analyze the TR repertoire in fish as shown by our detailed analysis of the TRB repertoire of rainbow trout IELs, performed in both naive and virus-infected animals. Rainbow trout IEL TRBV transcripts were highly diverse and polyclonal in adult naive individuals, in sharp contrast with the restricted diversity of IEL oligoclonal repertoires described in birds and mammals (102). More recently, our study of the CD8$^+$ and CD8$^-$ αβ T cell repertoire suggests different regulatory patterns of those T cell patterns in fish and in mammals (168). ISEApeaks was also used to implement a new statistically based strategy for quantification of repertoire diversity (159).

Kepler et al. described another original statistical approach for CDR3 spectratype analysis, using complex procedures for testing hypotheses regarding differences in antigen receptor distribution and variable repertoire diversity in different treatment groups. This approach is based on the derivation of probability distributions directly from spectratype data instead of using *ad hoc* measures of spectratype differences (169). A software (called SpA) implementing this method has been developed and made available online (170). This approach has been used in a longitudinal analysis of TRBV repertoire during acute GvHD after stem cell transplantation (171).

Another group (163) reported the development of a new software platform, REPERTOIRE, which allows handling of CDR3 spectratyping data. This software implements a perturbation index based upon an expected normal Gaussian distribution of CDR3 length profiles.

Owing to the complexity and diversity of the immune system, immunogenetics represents one of the greatest challenges for data interpretation: a large biological expertise, a considerable effort of standardization, and the elaboration of an efficient system

for the management of the related knowledge were required. To answer that challenge, IMGT®, the international ImMunoGeneTics information system®(http://www.imgt.org), was created in 1989 by one of the authors (146). Overtime, it developed standards that, since 1995, have been endorsed by the World Health Organization-International Union of Immunological Societies (WHO-IUIS) Nomenclature Committee and by the WHO-International Nonproprietary Names (INN) (172–175). IMGT® comprises seven databases (sequence, gene, and structure databases), 17 online tools and more than 15,000 pages of web resources. Among the databases, IMGT/LIGM-DB, the database for nucleotide sequences (170,685 sequences from 335 species as of July 2013) and IMGT/GENE-DB, the gene database (3,081 genes and 4,687 alleles) are of great interest for repertoire analysis. Freely available since 1997, IMGT/V-QUEST is an integrated system for the standardized analysis of collections of IG and TR rearranged nucleotide sequences (176, 177). A high-throughput version, IMGT/HighV-QUEST (148), has been released in 2010 for the analysis of long IG and TR sequences from NGS using the 454 Life Sciences technology. In the same line, other analysis tools are becoming available showing the renewed interest for repertoire analyses and modeling consecutive to NGS technology developments (178–181).

Altogether, these efforts highlight the relevance of developing more efficient and powerful technologies for the evaluation of repertoire diversity. Notably, two successful French biotech companies (TcLand, Nantes; ImmunID, Grenoble) were created in the field of repertoire analysis, using different technologies. In collaboration with ImmunID, we have proposed a novel strategy for statistical modeling of T lymphocyte repertoire data obtained in humans and humanized mice. With this model, we revealed that half of the human TRB repertoire, in terms of proportion of TRBV-TRBJ combinations, is genetically determined, the other half occurring stochastically (182). In addition, the biotechnology company "Adaptive" and the "Repertoire 10K (R10K) Project" have been recently founded by researchers respectively from the Fred Hutchinson Cancer Research Center (Seattle and Washington) and the HudsonAlpha Institute (Huntsville). Both have developed platforms (immunoSEQ®, iRepertoire®) providing researchers with a global analysis of the T or B cell receptor sequence repertoires (183). However, despite the power of this technology, studies are still limited by the ability to process the complexity of the information provided. Specific software developments for the automatic treatment and annotation of IG and TR sequences and the statistical modeling of repertoire diversity can still be improved.

Multivariate analysis

As mentioned above, the PANAMA-Blot technique also includes statistical analysis of the data. Multi-parametric analysis was introduced to compare the global reactivity of antibodies of different individuals in different groups with a given antigenic extract. This analysis has been successfully implemented to identify reactivity patterns specific for a given pathology or clinical status (10–12, 14, 15, 184). Similarly, multi-parametric analysis was also applied to TRBV spectratype analysis in an experimental cerebral malaria model (107).

Hierarchical clustering or classification algorithms have become very popular with the growing of micro-array-based transcriptome analysis. Although still uncommon for immune repertoire analysis, such approaches have been employed to categorize large sets of repertoire data without *a priori* (17, 102, 107).

Diversity indices

The concept of immune repertoire has been devised to describe the diversity of cells involved in the immune system of an individual (1). As described above, different scoring systems were developed to assess this diversity, some are heuristics but others have been borrowed from theoretical ecology and evolution. As reviewed by Magurran (185), the Shannon entropy, introduced by Claude Shannon in 1948 for the information theory, is the most used because it not only integrates the number of different species but also the relative proportion of each of these species. In 1961, Alfred Rényi generalized this entropy to a family of functions, like Species Richness, Simpson, Quadratic, and Berger–Parker indices, for quantifying the diversity, the uncertainty or randomness of a system. Most of these indices are implemented in the free software application Estimates (http://purl.oclc.org/estimates) (186). Altogether, these diversity indices constitute a collection of tools with their own sensitivity to the variety and the relative abundances of the species that are perfectly suitable for assessing immune repertoire diversity. Indeed, the very famous index of variability proposed by Kabat and Wu (187) corresponds to the ratio of Species Richness and Berger–Parker indices. In 1990, Jores et al. showed that the resolving power of this Wu-Kabat variability coefficient can be enhanced by increasing the weight on the frequency distribution of the amino acids in the formula (188). This approach inspired Stewart et al. (189) to use the Shannon entropy to demonstrate that TR amino acid composition is significantly more diverse than that of IG. In the same way, CDR3 spectratyping data can be analyzed using the relative abundance of each peak within CDR3 length global distribution. By doing so, we adjusted the original Shannon entropy, making it reaching its maximum for a Gaussian distribution, to compare the CDR3 length diversity of splenic IgM, IgD, and IgT in infected Teleost Fish (131). Recently, the Gini index, used in ecology or economics to measure the equality of distributions, was applied to individual TR clones and compared naive and memory repertoires (190). The development of deep sequencing techniques ignited a renewed interest in IG/TR repertoire. Indeed, several studies used high-throughput analysis to describe TR repertoire of key T cell subsets in human peripheral blood (115, 126, 191). This approach assessing the repertoire diversity from the relative abundance of each species in the global distribution can be decomposed hierarchically into components attributable, respectively, to variations in TRBV-TRBJ combinations and in CDR3-length (113, 117). However, most of these studies have been limited to the counting of the observed unique clonotypes. Beside the species richness, ecology-derived indices have also been applied to assess and compare immune repertoire diversity. Föhse et al. (119) used the Morisita-Horn similarity index to compare regulatory T cell repertoires between several lymphoid organs. In addition, Simpson diversity index, associated with Shannon entropy, was used to monitor TR repertoire

diversity of HIV-specific CD8 T cells during antiretroviral therapy (192) but also to quantify TR repertoire recovery in the blood after allogeneic hematopoietic stem cell transplantation (128). In the same manner, Koning et al. (193) used Shannon's and Simpson's indices to show the role for the peptide component of the peptide-MH1 complex on the molecular frontline of CD8$^+$ T cell–mediated immune surveillance, by comparing the repertoire diversity of CD8$^+$ T cell populations directed against a variety of epitopes. In parallel, using Simpson's index as a metric allowed Johnson et al. (194) to model mathematically the naïve CD4 T cell repertoire contraction with age leading them to conclude that diversity plummet observed around the age of 70 could be correlated to cell-intrinsic mutations affecting cell division rate or death.

MODELING STRATEGIES

Modeling approaches have a strong tradition in immunology, usually at the boundary with other disciplines such as physics (195). Before deep sequencing data was available, general design principles were proposed as desirable features of immune repertoires, with implications for the observed repertoire diversity and dynamics (196–198). Many efforts have involved the modeling of immune cell dynamics and the effects of antigens on repertoire diversity, using differential equations descriptions of the population dynamics (199–201). Recognition in the immune system is often studied both theoretically and experimentally by probing the dynamics of cells with a specific type of receptor with respect to infections (202). Alternatively one can look at the response of a small set of chosen receptors to a specific pathogenic challenge, or careful biochemical investigation of particular receptor/antigen pairs (203, 204). Much work has been devoted to systems-biology approaches to signal processing in immune cells, as reviewed in Germain et al. (205) and Emonet and Altan-Bonnet (206). Here we focus on approaches inspired by recent advances in sequencing technologies (112, 113, 115, 116, 125, 191, 207, 208) that have opened the way for data-driven modeling of the immune repertoires and interactions between receptors and antigen.

A common modeling approach for describing receptors at the amino acid level is to choose a relevant interaction parameter (e.g., chemical affinity or hydrophobicity) and assign it a simplified digit-string representation (209). These methods are extensions of the string model, which describes both receptor and epitopes as strings of length L, with values chosen from natural numbers, and quantify their interaction by the match between the two strings (197, 210, 211). Such quantitative, physically inspired descriptions of immune receptors, despite the arbitrary choice of interaction coordinates, have proven a valuable first step in statistically describing recognition in T cells (195, 212–215). Recently, lower hydrophilicity of regulatory vs. conventional T cells was suggested from CDR3 sequencing (216).

High-throughput sequencing of immune receptors raises specific challenges compared to traditional genomic sequencing. It is harder to distinguish sequencing errors from new polymorphisms, since no corresponding pre-existing sequence exists. One of the most interesting regions when studying diversity is the CDR3 with its many insertions and deletions added to the germline sequence. These regions are often hard to align to the genomic templates, or

with each other (217). Therefore, extra care is needed when generating and analyzing sequence data. Not all sequencing technologies are equally good for all purposes (218): while 454 sequencing gives longer reads than Illumina it is known to have a greater probability of frameshift errors. In addition, primer-dependent PCR amplification biases require that raw sequence counts be normalized using control experiments (112) in order to accurately report clone sizes, as demonstrated by spike-in experiments (219). In TR repertoire studies, this is circumvented by using 5′RACE which provides an unbiased amplification of fully rearranged sequences, as recently demonstrated for TRB V-(D)-J transcripts (191).

Despite sequencing issues, statistical algorithms are often able to extract information from the data. Many studies of diversity focus on the V, D, and J gene usage of each rearranged sequence. Algorithms and tools have been developed to rapidly identify the V, D, and J genes for massive numbers of sequences (148, 178, 181). In many cases however, the assignment of a D gene to each sequence read is unreliable if the D region is too short owing to extensive trimming. Mora et al. (217) learned from data and analyzed statistical models of the D gene flanked by its junctions. These models are based on the principle of maximum entropy and make minimal assumptions about the mechanisms of diversity – they only rely on the observed frequencies of amino acid pairs along the sequence. These models were used to describe global features of the sequence ensemble, such as the probability distribution following Zipf's law (220) – the observation that the probability of sequences is inversely proportional to their frequency-rank, or the observation of peaks of frequency in sequence landscape as possible signatures of past pathogenic challenges. Recently, the estimation of repertoire diversity and clonal size distribution were analyzed by Poisson abundance models (221) and simple bivariate-Poisson-lognormal (BPLN) parametric model for fitting and analyzing TR repertoire data was proposed (222). Similarly, network analysis of IG repertoire from Weinstein et al. study revealed the possibility to identify subgroups of individuals on the basis of IG network similarity (223).

The task of characterizing the CDR3 at the nucleotide level is made difficult by the fact that a deterministic assignment of the V-(D)-J recombination process is impossible, because any given sequence can be generated by many possible recombination processes. A previous study proposed a probabilistic model of nucleotide trimming of rearranged TR genes derived from a benchmark data set of TRA and TRG V-(D)-J junctions obtained by comparison to the germline genes in the IMGT® tools (224). Recently a statistical method based on the expectation-maximization algorithm was proposed to circumvent this issue and to extract the statistical properties of junctional diversity accurately from data (124). Applying it to human non-productive DNA sequences gave insight into a universal generation mechanism, reproducible from individual to individual. It was shown that each sequence could potentially be generated by the equivalent of ~30 equally likely ways by convergent recombination. This method showed that the potential diversity of the recombination machinery was equivalent to $\sim10^{14}$ *equally likely* sequences (and a practically infinite total number of possible sequences), much more than the estimated 10^{12} T cells that a single human body can hold. The frequencies of the V, D, and J genes is non-uniform, even at the level of recombination, suggesting underlying physical mechanisms at work. Ndifon et al. (125) proposed a polymer model that accounts for the likelihood of connecting given genomic fragments, giving insight into the mechanistic process.

One of the ultimate goals of deep repertoire sequencing is to find signatures of the repertoire's response to its antigenic environment. A combination of clustering methods and tree reconstruction techniques have been developed (225, 226) to identify lineages in B cells and study the response to pathogenic challenges. Statistical methods have been devised to detect and quantify the extent of antigen-driven selection acting on B cells, by analyzing the patterns of hypermutations in a Bayesian framework, with applications to deep sequencing data (227, 228).

A lot remains to be done in terms of both data-driven and small-scale models of repertoire-antigen interactions. Ultimately, a close collaboration and development of experimental techniques and models can shed light on how selection at different stages shapes the repertoire, how affinity maturation changes the diversity and the link between sequence diversity and function.

FUTURE PROSPECTS OF BIOMATHEMATICAL ANALYSIS OF REPERTOIRE DATA

One of the current challenging issues in antigen-specific repertoire analysis is the development of relevant statistical analysis strategies. Biologists are usually keen on parametric tests, such as ANOVA, *t*-test, Fischer's test, among others. However, such statistical methods assume that the inherent probability distribution of the observed variable follows a normal distribution. Rock et al. (229) described that the distribution of the TR diversity is far from following this distribution, thus they proposed the use of non-parametric tests. Nevertheless, different groups are dealing with this issue in order to determine the relevant way to analyze repertoire diversity data and to propose new biostatistics strategies, including principal component analysis, discriminant analysis, hierarchical clustering, specific statistics (164, 169).

In fact, the traditional use of statistics in biology aims at the falsification of a defined hypothesis, i.e., at validating significant differences between defined situations. The recent development of "systems immunology" reverses this point of view and establishes a new usage of multi-parametric statistical approaches to represent the biological data by projections and "landscapes" in the N-dimensional space of considered parameters (230). Thus, the traditional description of separate repertoires for distinct cell subsets defined from a few markers is being replaced by overlapping clouds of data, setting the limits of the different classification groups (tissue of origin, infection contexts, combination of marker expression, repertoire expression. . .). Moreover, repertoire diversity technologies can now be combined to complementary approaches to decipher the complexity of lymphocyte populations, such as microwell array cell culture and high-resolution imaging (231), mass cytometry (232, 233), cellular barcoding (234), intravital imaging (235, 236), single-cell gene expression (237). In addition, high-throughput repertoire descriptions will enrich mathematical and computer models of lymphocyte repertoire diversity and dynamics such as those proposed by Mehr (238), Ciupe et al. (239), or Stirk et al. (240).

As advocated by others, the concepts developed by systems biology, such as the signatures emerging from clustering and the modularity regulating gene networks, will probably need to be adapted to the constraints of immunology data (241). However, this is probably through this kind of representation that global analysis of immune repertoires will have to be addressed (242).

The upcoming challenge is now to merge data produced through the different technological approaches available to achieve full integration of these data and make them available for interactive meta-analysis. This necessitates more than the simple juxtaposition of annotated raw data but rather requires (1) the codification and standardization of this multi-level data and (2) the integration of complexity science into immunology. Along this line, recent developments of multi-parametric flow cytometry naturally led to systematic clustering and multivariate statistical analysis approaches for searching functional signatures (2, 232, 233, 243–245).

ACKNOWLEDGMENTS

REFERENCES

1. Boudinot P, Marriotti-Ferrandiz ME, Du Pasquier L, Benmansour A, Cazenave PA, Six A. New perspectives for large-scale repertoire analysis of immune receptors. *Mol Immunol* (2008) **45**:2437–45. doi:10.1016/j.molimm.2007.12.018

2. Bendall SC, Nolan GP, Roederer M, Chattopadhyay PK. A deep profiler's guide to cytometry. *Trends Immunol* (2012) **33**:323–32. doi:10.1016/j.it.2012.02.010

3. MacDonald HR, Pedrazzini T, Schneider R, Louis JA, Zinkernagel RM, Hengartner H. Intrathymic elimination of Mlsᵃ-reactive (Vβ6⁺) cells during neonatal tolerance induction to Mlsᵃ-encoded antigens. *J Exp Med* (1988) **167**:2005–10. doi:10.1084/jem.167.6.2005

4. MacDonald HR, Schneider R, Lees RK, Howe RC, Acha-Orbea H, Festenstein H, et al. T-cell receptor Vβ use predicts reactivity tolerance to Mlsᵃ-encoded antigens. *Nature* (1988) **332**:40–5. doi:10.1038/332040a0

5. Salaun J, Bandeira A, Khazaal I, Burlen-Defranoux O, Thomas-Vaslin V, Coltey M, et al. Transplantation tolerance is unrelated to superantigen-dependent deletion and anergy. *Proc Natl Acad Sci U S A* (1992) **89**:10420–4. doi:10.1073/pnas.89.21.10420

6. Faint JM, Pilling D, Akbar AN, Kitas GD, Bacon PA, Salmon M. Quantitative flow cytometry for the analysis of T cell receptor Vβ chain expression. *J Immunol Methods* (1999) **225**:53–60. doi:10.1016/S0022-1759(99)00027-7

7. Thomas-Vaslin V, Six A, Pham HP, Dansokho C, Chaara W, Gouritin B, et al. Immunodepression & Immunosuppression during aging. In: Portela MB editor. *Immunosuppression*. Rijeka: InTech open access publisher (2012). p. 125–463.

8. Lefranc MP, Lefranc G. *The Immunoglobulin FactsBook*. London: Academic Press (2001).

9. Lefranc MP, Lefranc G. *The T Cell Receptor FactsBook*. London: Academic Press (2001).

10. Nobrega A, Haury M, Grandien A, Malanchere E, Sundblad A, Coutinho A. Global analysis of antibody repertoires. II. Evidence for specificity, self-selection and the immunological "homunculus" of antibodies in normal serum. *Eur J Immunol* (1993) **23**:2851–9. doi:10.1002/eji.1830231119

11. Haury M, Grandien A, Sundblad A, Coutinho A, Nobrega A. Global analysis of antibody repertoires. 1. An immunoblot method for the quantitative screening of a large number of reactivities. *Scand J Immunol* (1994) **39**:79–87.

12. Fesel C, Coutinho A. Serum IgM repertoire reactions to MBP/CFA immunization reflect the individual status of EAE susceptibility. *J Autoimmun* (2000) **14**:319–24. doi:10.1006/jaut.2000.0373

13. Haury M, Sundblad A, Grandien A, Barreau C, Coutinho A, Nobrega A. The repertoire of serum IgM in normal mice is largely independent of external antigenic contact. *Eur J Immunol* (1997) **27**:1557–63. doi:10.1002/eji.1830270635

14. Stahl D, Lacroix-Desmazes S, Heudes D, Mouthon L, Kaveri SV, Kazatchkine MD. Altered control of self-reactive IgG by autologous IgM in patients with warm autoimmune hemolytic anemia. *Blood* (2000) **95**:328–35.

15. Stahl D, Lacroix-Desmazes S, Mouthon L, Kaveri SV, Kazatchkine MD. Analysis of human self-reactive antibody repertoires by quantitative immunoblotting. *J Immunol Methods* (2000) **240**:1–14. doi:10.1016/S0022-1759(00)00185-X

16. Costa N, Pires AE, Gabriel AM, Goulart LF, Pereira C, Leal B, et al. Broadened T-cell repertoire diversity in ivIg-treated SLE patients is also related to the individual status of regulatory T-cells. *J Clin Immunol* (2013) **33**:349–60. doi:10.1007/s10875-012-9816-7

17. Quintana FJ, Hagedorn PH, Elizur G, Merbl Y, Domany E, Cohen IR. Functional immunomics: microarray analysis of IgG autoantibody repertoires predicts the future response of mice to induced diabetes. *Proc Natl Acad Sci U S A* (2004) **101**:14615–21. doi:10.1073/pnas.0404848101

18. Quintana FJ, Getz G, Hed G, Domany E, Cohen IR. Cluster analysis of human autoantibody reactivities in health and in type 1 diabetes mellitus: a bioinformatic approach to immune complexity. *J Autoimmun* (2003) **21**:65–75. doi:10.1016/S0896-8411(03)00064-7

19. Delassus S, Darche S, Kourilsky P, Cumano A. Ontogeny of the heavy chain immunoglobulin repertoire in fetal liver and bone marrow. *J Immunol* (1998) **160**:3274–80.

20. Yassai M, Gorski J. Thymocyte maturation: selection for in-frame TCRα-chain rearrangement is followed by selection for shorter TCRβ-chain complementarity-determining region 3. *J Immunol* (2000) **165**:3706–12.

21. Pasqual N, Gallagher M, Aude-Garcia C, Loiodice M, Thuderoz F, Demongeot J, et al. Quantitative and qualitative changes in V-Jα rearrangements during mouse thymocytes differentiation: implication for a limited T cell receptor α chain repertoire. *J Exp Med* (2002) **196**:1163–73. doi:10.1084/jem.20021074

22. Douek DC, McFarland RD, Keiser PH, Gage EA, Massey JM, Haynes BF, et al. Changes in thymic function with age and during the treatment of HIV infection. *Nature* (1998) **396**:690–5. doi:10.1038/25374

23. Ribeiro RM, Perelson AS. Determining thymic output quantitatively: using models to interpret experimental T-cell receptor excision circle (TREC) data. *Immunol Rev* (2007) **216**:21–34.

24. Bains I, Thiebaut R, Yates AJ, Callard R. Quantifying thymic export: combining models of naive T cell proliferation and TCR excision circle dynamics gives an explicit measure of thymic output. *J Immunol* (2009) **183**:4329–36. doi:10.4049/jimmunol.0900743

25. Poulin JF, Viswanathan MN, Harris JM, Komanduri KV, Wieder E, Ringuette N, et al. Direct evidence for thymic function in adult humans. *J Exp Med* (1999) **190**:479–86. doi:10.1084/jem.190.4.479

26. Dulude G, Cheynier R, Gauchat D, Abdallah A, Kettaf N, Sékaly RP, et al. The magnitude of thymic output is genetically determined through controlled intrathymic precursor T cell proliferation. *J Immunol* (2008) **181**:7818–24.

27. Wu GE, Paige CJ. VH gene family utilization in colonies derived from B and pre-B cells detected by the RNA colony blot assay. *EMBO J* (1986) **5**:3475–81.

28. Schulze DH, Kelsoe G. Genotypic analysis of B cell colonies by in situ hybridization. Stoichiometric expression of three VH families in adult C57BL/6 and BALB/c mice. *J Exp Med* (1987) **166**:163–72. doi:10.1084/jem.166.1.163

29. Thomas-Vaslin V, Andrade L, Freitas A, Coutinho A. Clonal persistence of B lymphocytes in normal mice is determined by variable region-dependent selection. *Eur J Immunol* (1991) **21**:2239–46. doi:10.1002/eji.1830210935

30. Andrade L, Huetz F, Poncet P, Thomas-Vaslin V, Goodhardt M, Coutinho A. Biased VH gene expression in murine CD5 B cells results from age-dependent cellular selection. *Eur J Immunol* (1991) **21**:2017–23. doi:10.1002/eji.1830210908

31. Freitas AA, Lembezat MP, Coutinho A. Expression of antibody V-regions is genetically and developmentally controlled and modulated by the B lymphocyte environment. *Int Immunol* (1989) **1**:342–54. doi:10.1093/intimm/1.4.342

32. Rosenberg WMC, Moss PAH, Bell JI. Variation in human T cell receptor Vβ and Jβ repertoire: analysis using anchored polymerase reaction. *Eur J Immunol* (1992) **22**:541–9. doi:10.1002/eji.1830220237

33. Moss PAH, Rosenberg WMC, Zintzaras E, Bell JI. Characterization of the human T cell receptor α-chain repertoire and demonstration of α genetic influence on the Vα usage. *Eur J Immunol* (1993) **23**:1155–9. doi:10.1002/eji.1830230526

34. Moss PAH, Bell JI. Sequence analysis of the human αβ T-cell receptor CDR3 region. *Immunogenetics* (1995) **42**:10–8. doi:10.1007/BF00164982

35. Moss PA, Bell JI. Comparative sequence analysis of the human T cell receptor TCRA and TCRB CDR3 regions. *Hum Immunol* (1996) **48**:32–8. doi:10.1016/0198-8859(96)00084-5

36. Pannetier C, Delassus S, Darche S, Saucier C, Kourilsky P. Quantitative titration of nucleic acids by enzymatic amplification reactions run to saturation. *Nucleic Acids Res* (1993) **21**:577–83. doi:10.1093/nar/21.3.577

37. Manfras BJ, Rudert WA, Trucco M, Boehm O. Analysis of the αβ T-cell receptor repertoire by competitive and quantitative family-specific PCR with exogenous standards and high resolution fluorescence based CDR3 size imaging. *J Immunol Methods* (1997) **210**:235–49. doi:10.1016/S0022-1759(97)00197-X

38. VanderBorght A, Van der Aa A, Geusens P, Vandevyver C, Raus J, Stinissen P. Identification of overrepresented T cell receptor genes in blood and tissue biopsies by PCR-ELISA. *J Immunol Methods* (1999) **223**:47–61. doi:10.1016/S0022-1759(98)00201-4

39. Lim A, Baron V, Ferradini L, Bonneville M, Kourilsky P, Pannetier C. Combination of MHC-peptide multimer-based T cell sorting with the Immunoscope permits sensitive ex vivo quantitation and follow-up of human CD8+ T cell immune responses. *J Immunol Methods* (2002) **261**:177–94. doi:10.1016/S0022-1759(02)00004-2

40. Casrouge A, Beaudoing E, Dalle S, Pannetier C, Kanellopoulos J, Kourilsky P. Size estimate of the αβ TCR repertoire of naive mouse splenocytes. *J Immunol* (2000) **164**:5782–7.

41. Gallard A, Foucras G, Coureau C, Guery JC. Tracking T cell clonotypes in complex T lymphocyte populations by real-time quantitative PCR using fluorogenic complementarity-determining region-3-specific probes. *J Immunol Methods* (2002) **270**:269–80. doi:10.1016/S0022-1759(02)00336-8

42. Liu DB, Callahan JP, Dau PC. Intrafamily fragment analysis of the T cell receptor β chain CDR3 region. *J Immunol Methods* (1995) **187**:139–50. doi:10.1016/0022-1759(95)00178-D

43. Mugnaini EN, Egeland T, Syversen AM, Spurkland A, Brinchmann JE. Molecular analysis of the complementarity determining region 3 of the human T cell receptor β chain. Establishment of a reference panel of CDR3 lengths from phytohaemagglutinin activated lymphocytes. *J Immunol Methods* (1999) **223**:207–16. doi:10.1016/S0022-1759(99)00004-6

44. Matsutani T, Yoshioka T, Tsuruta Y, Iwagami S, Suzuki R. Analysis of TCRAV and TCRBV repertoires in healthy individuals by microplate hybridization assay. *Hum Immunol* (1997) **56**:57–69. doi:10.1016/S0198-8859(97)00102-X

45. Matsutani T, Yoshioka T, Tsuruta Y, Iwagami S, Toyosaki-Maeda T, Horiuchi T, et al. Restricted usage of T-cell receptor α-chain variable region (TCRAV) and T-cell receptor β-chain variable region (TCRBV) repertoires after human allogeneic haematopoietic transplantation. *Br J Haematol* (2000) **109**:759–69. doi:10.1046/j.1365-2141.2000.02080.x

46. Yoshida R, Yoshioka T, Yamane S, Matsutani T, Toyosaki-Maeda T, Tsuruta Y, et al. A new method for quantitative analysis of the mouse T-cell receptor V region repertoires: comparison of repertoires among strains. *Immunogenetics* (2000) **52**:35–45. doi:10.1007/s002510000248

47. Du G, Qiu L, Shen L, Sehgal P, Shen Y, Huang D, et al. Combined megaplex TCR isolation and SMART-based real-time quantitation methods for quantitating antigen-specific T cell clones in mycobacterial infection. *J Immunol Methods* (2006) **308**:19–35. doi:10.1016/j.jim.2005.09.009

48. Ogle BM, Cascalho M, Joao C, Taylor W, West LJ, Platt JL. Direct measurement of lymphocyte receptor diversity. *Nucleic Acids Res* (2003) **31**:e139. doi:10.1093/nar/gng139

49. Joao C, Ogle BM, Gay-Rabinstein C, Platt JL, Cascalho M. B cell-dependent TCR diversification. *J Immunol* (2004) **172**:4709–16.

50. Joao C. Immunoglobulin is a highly diverse self-molecule that improves cellular diversity and function during immune reconstitution. *Med Hypotheses* (2007) **68**:158–61. doi:10.1016/j.mehy.2006.05.062

51. Bonarius HP, Baas F, Remmerswaal EB, van Lier RA, ten Berge I, Tak PP, et al. Monitoring the T-cell receptor repertoire at single-clone resolution. *PLoS One* (2006) **1**:e55. doi:10.1371/journal.pone.0000055

52. Pannetier C, Cochet M, Darche S, Casrouge A, Zöller M, Kourilsky P. The size of the CDR3 hypervariable regions of the murine T-cell receptor β chains vary as a function of the recombined germ-line segments. *Proc Natl Acad Sci U S A* (1993) **90**:4319–23. doi:10.1073/pnas.90.9.4319

53. Pannetier C, Even J, Kourilsky P. T-cell repertoire diversity and clonal expansions in normal and clinical samples. *Immunol Today* (1995) **16**:176–81. doi:10.1016/0167-5699(95)80117-0

54. Pannetier C, Levraud JP, Lim A, Even J, Kourilsky P. The immunoscope approach for the analysis of T-cell repertoires. In: Oksenberg J editor. *The Human Antigen T Cell Receptor. Selected Protocols and Applications.* Georgetown, TX: Landes RG (1997). p. 287–325.

55. Cochet M, Pannetier C, Darche S, Leclerc C, Kourilsky P. Molecular detection and *in vivo* analysis of the specific T cell response to a protein antigen. *Eur J Immunol* (1992) **22**:2639–47. doi:10.1002/eji.1830221025

56. Regnault A, Cumano A, Vassalli P, Guy-Grand D, Kourilsky P. Oligoclonal repertoire of the CD8αα and the CD8αβ TCR-α/β murine intestinal intraepithelial T lymphocytes: evidence for the random emergence of T cells. *J Exp Med* (1994) **180**:1345–58. doi:10.1084/jem.180.4.1345

57. Regnault A, Levraud JP, Lim A, Six A, Moreau C, Cumano A, et al. The expansion and selection of T cell receptor αβ intestinal intraepithelial T cell clones. *Eur J Immunol* (1996) **26**:914–21. doi:10.1002/eji.1830260429

58. Ema H, Cumano A, Kourilsky P. TCRβ repertoire development in the mouse embryo. *J Immunol* (1997) **159**:4227–32.

59. Arstila TP, Casrouge A, Baron V, Even J, Kanellopoulos J, Kourilsky P. A direct estimate of the human αβ T cell receptor diversity. *Science* (1999) **286**:958–61. doi:10.1126/science.286.5441.958

60. Bousso P, Lemaître F, Laouini D, Kanellopoulos J, Kourilsky P. The peripheral CD8 T cell repertoire is largely independent of the presence of intestinal flora. *Int Immunol* (2000) **12**:425–30. doi:10.1093/intimm/12.4.425

61. Arstila TP, Even J. Size of the αβ TCR repertoire. *Med Sci (Paris)* (2000) **16**:1257–60. doi:10.4267/10608/1566

62. Cabaniols JP, Fazilleau N, Casrouge A, Kourilsky P, Kanellopoulos JM. Most α/β T cell receptor diversity is due to terminal deoxynucleotidyl transferase. *J Exp Med* (2001) **194**:1385–90. doi:10.1084/jem.194.9.1385

63. Fazilleau N, Cabaniols JP, Lemaitre F, Motta I, Kourilsky P, Kanellopoulos JM. Valpha and Vbeta public repertoires are highly conserved in terminal deoxynucleotidyl transferase-deficient mice. *J Immunol* (2005) **174**:345–55.

64. Cibotti R, Cabaniols JP, Pannetier C, Delarbre C, Vergnon I, Kanellopoulos JM, et al. Public and private Vβ T cell receptor repertoires against hen egg white lysozyme (HEL) in nontransgenic versus HEL transgenic mice. *J Exp Med* (1994) **180**:861–72. doi:10.1084/jem.180.3.861

65. Gapin L, Fukui Y, Kanellopoulos J, Sano T, Casrouge A, Malier V, et al. Quantitative analysis of the T cell repertoire selected by a single peptide-major histocompatibility complex. *J Exp Med* (1998) **187**:1871–83. doi:10.1084/jem.187.11.1871

66. Bouneaud C, Kourilsky P, Bousso P. Impact of negative selection on the T cell repertoire reactive to a self-peptide: a large fraction of T cell clones escapes clonal deletion. *Immunity* (2000) **13**:829–40. doi:10.1016/S1074-7613(00)00080-7

67. Fukui Y, Oono T, Cabaniols JP, Nakao K, Hirokawa K, Inayoshi A, et al. Diversity of T cell repertoire shaped by a single peptide ligand is critically affected by its amino acid residue at a T cell receptor contact. *Proc Natl Acad Sci U S A* (2000) **97**:13760–5. doi:10.1073/pnas.250470797

68. Musette P, Bureau JF, Gachelin G, Kourilsky P, Brahic M. T lymphocyte repertoire in Theiler's virus encephalomyelitis: the nonspecific infiltration of the central nervous system of infected SJL/J mice is associated with a selective local T cell expansion. *Eur J Immunol* (1995) **25**:1589–93. doi:10.1002/eji.1830250618

69. Sourdive DJD, Murali-Krishna K, Altman JD, Zajac AJ, Whitmire JK, Pannetier C, et al. Conserved T cell receptor repertoire in primary and memory CD8 T cell responses to an acute viral infection. *J Exp Med* (1998) **188**:71–82. doi:10.1084/jem.188.1.71

70. Musette P, Bequet D, Delarbre C, Gachelin G, Kourilsky P, Dormont D. Expansion of a recurrent Vβ5.3[+] T-cell population in newly diagnosed and untreated HLA-DR2 multiple sclerosis patients. *Proc Natl Acad Sci U S A* (1996) **93**:12461–6. doi:10.1073/pnas.93.22.12461

71. Fazilleau N, Delarasse C, Sweenie CH, Anderton SM, Fillatreau S, Lemonnier FA, et al. Persistence of autoreactive myelin oligodendrocyte glycoprotein (MOG)-specific T cell repertoires in MOG-expressing mice. *Eur J Immunol* (2006) **36**:533–43. doi:10.1002/eji.200535021

72. Musette P, Bachelez H, Flageul B, Delarbre C, Kourilsky P, Dubertret L, et al. Immune-mediated destruction of melanocytes in halo nevi is associated with

the local expansion of a limited number of T cell clones. *J Immunol* (1999) **162**:1789–94.

73. Douillard P, Pannetier C, Josien R, Menoret S, Kourilsky P, Soulillou JP, et al. Donor-specific blood transfusion-induced tolerance in adult rats with a dominant TCR-Vβ rearrangement in heart allografts. *J Immunol* (1996) **157**:1250–60.

74. Pilch H, Höhn H, Freitag K, Neukirch C, Necker A, Haddad P, et al. Improved assessment of T-cell receptor (TCR) VB repertoire in clinical specimens: combination of TCR-CDR3 spectratyping with flow cytometry-based TCR VB frequency analysis. *Clin Diagn Lab Immunol* (2002) **9**:257–66.

75. Messaoudi I, LeMaoult J, Guevara-Patino JA, Metzner BM, Nikolich-Zugich J. Age-related CD8 T cell clonal expansions construct CD8 cell repertoire and have the potential to impair immune defense. *J Exp Med* (2004) **200**:1347–58. doi:10.1084/jem.20040437

76. Hakim FT, Memon SA, Cepeda R, Jones EC, Chow CK, Kasten-Sportes C, et al. Age-dependent incidence, time course, and consequences of thymic renewal in adults. *J Clin Invest* (2005) **115**:930–9. doi:10.1172/JCI200522492

77. Messaoudi I, Warner J, Fischer M, Park B, Hill B, Mattison J, et al. Delay of T cell senescence by caloric restriction in aged long-lived nonhuman primates. *Proc Natl Acad Sci U S A* (2006) **103**:19448–53. doi:10.1073/pnas.0606661103

78. Boudinot P, Boubekeur S, Benmansour A. Rhabdovirus infection induces public and private T cell responses in teleost fish. *J Immunol* (2001) **167**:6202–9.

79. Boudinot P, Boubekeur S, Benmansour A. Primary structure and complementarity-determining region (CDR) 3 spectratyping of rainbow trout TCRβ transcripts identify ten Vβ families with Vβ6 displaying unusual CDR2 and differently spliced forms. *J Immunol* (2002) **169**:6244–52.

80. Boudinot P, Bernard D, Boubekeur S, Thoulouze MI, Bremont M, Benmansour A. The glycoprotein of a fish rhabdovirus profiles the virus-specific T-cell repertoire in rainbow trout. *J Gen Virol* (2004) **85**:3099–108. doi:10.1099/vir.0.80135-0

81. Desravines S, Hsu E. Measuring CDR3 length variability in individuals during ontogeny. *J Immunol Methods* (1994) **168**:219–25. doi:10.1016/0022-1759(94)90058-2

82. Gorski J, Yassai M, Zhu X, Kissella B, Keever C, Flomenberg N. Circulating T cell repertoire complexity in normal individuals and bone marrow recipients analyzed by CDR3 spectratyping. *J Immunol* (1994) **152**:5109–19.

83. Maslanka K, Piatek T, Gorski J, Yassai M. Molecular analysis of T cell repertoires – Spectratypes generated by multiplex polymerase chain reaction and evaluated by radioactivity or fluorescence. *Hum Immunol* (1995) **44**:28–34.

84. Lue C, Mitani Y, Crew MD, George JF, Fink LM, Schichman SA. An automated method for the analysis of T-cell receptor repertoires: rapid RT-PCR fragment length analysis of the T-cell receptor β chain complementarity-determining region 3. *Am J Clin Pathol* (1999) **111**:683–90.

85. Yamamoto K, Masuko-Hongo K, Tanaka A, Kurokawa M, Hoeger T, Nishioka K, et al. Establishment and application of a novel T cell clonality analysis using single-strand conformation polymorphism of T cell receptor messenger signals. *Hum Immunol* (1996) **48**:23–31. doi:10.1016/0198-8859(96)00080-8

86. Shiokawa S, Nishimura J, Ohshima K, Uike N, Yamamoto K. Establishment of a novel B cell clonality analysis using single-strand conformation polymorphism of immunoglobulin light chain messenger signals. *Am J Pathol* (1998) **153**:1393–400. doi:10.1016/S0002-9440(10)65726-4

87. Raaphorst FM, Gokmen E, Teale JM. Analysis of clonal diversity in mouse immunoglobulin heavy chain genes selected for size of the antigen combining site. *Immunol Invest* (1998) **27**:355–65. doi:10.3109/08820139809022709

88. Sottini A, Quiròs Roldan E, Albertini A, Primi D, Imberti L. Assessment of T-cell receptor beta-chain diversity by heteroduplex analysis. *Hum Immunol* (1996) **48**:12–22. doi:10.1016/0198-8859(96)00087-0

89. Wack A, Montagna D, Dellabona P, Casorati G. An improved PCR-heteroduplex method permits high-sensitivity detection of clonal expansions in complex T cell populations. *J Immunol Methods* (1996) **196**:181–92. doi:10.1016/0022-1759(96)00114-7

90. Shen DF, Doukhan L, Kalam S, Delwart E. High-resolution analysis of T-cell receptor β-chain repertoires using DNA heteroduplex tracking: generally stable, clonal CD8+ expansions in all healthy young adults. *J Immunol Methods* (1998) **215**:113–21. doi:10.1016/S0022-1759(98)00066-0

91. Wedderburn LR, Maini MK, Patel A, Beverley PCL, Woo P. Molecular fingerprinting reveals non-overlapping T cell oligoclonality between an inflamed site and peripheral blood. *Int Immunol* (1999) **11**:535–43. doi:10.1093/intimm/11.4.535

92. Bouffard P, Gagnon C, Cloutier D, MacLean SJ, Souleimani A, Nallainathan D, et al. Analysis of T cell receptor β chain expression by isoelectric focusing following gene amplification and *in vitro* translation. *J Immunol Methods* (1995) **187**:9–21. doi:10.1016/0022-1759(95)00161-3

93. Sant'Angelo DB, Lucas B, Waterbury PG, Cohen B, Brabb T, Goverman J, et al. A molecular map of T cell development. *Immunity* (1998) **9**:179–86. doi:10.1016/S1074-7613(00)80600-7

94. Correia-Neves M, Waltzinger C, Mathis D, Benoist C. The shaping of the T cell repertoire. *Immunity* (2001) **14**:21–32. doi:10.1016/S1074-7613(01)00086-3

95. Hsieh CS, Liang Y, Tyznik AJ, Self SG, Liggitt D, Rudensky AY. Recognition of the peripheral self by naturally arising CD25+ CD4+ T cell receptors. *Immunity* (2004) **21**:267–77. doi:10.1016/j.immuni.2004.07.009

96. Hsieh CS, Zheng Y, Liang Y, Fontenot JD, Rudensky AY. An intersection between the self-reactive regulatory and nonregulatory T cell receptor repertoires. *Nat Immunol* (2006) **7**:401–10. doi:10.1038/ni1318

97. Pacholczyk R, Ignatowicz H, Kraj P, Ignatowicz L. Origin and T cell receptor diversity of Foxp3+CD4+CD25+ T cells. *Immunity* (2006) **25**:249–59. doi:10.1016/j.immuni.2006.05.016

98. Pacholczyk R, Kern J, Singh N, Iwashima M, Kraj P, Ignatowicz L. Nonself-antigens are the cognate specificities of Foxp3+ regulatory T cells. *Immunity* (2007) **27**:493–504. doi:10.1016/j.immuni.2007.07.019

99. Wong J, Obst R, Correia-Neves M, Losyev G, Mathis D, Benoist C. Adaptation of TCR repertoires to self-peptides in regulatory and nonregulatory CD4+ T cells. *J Immunol* (2007) **178**:7032–41.

100. Kang JA, Mohindru M, Kang BS, Park SH, Kim BS. Clonal expansion of infiltrating T cells in the spinal cords of SJL/J mice infected with Theiler's virus. *J Immunol* (2000) **165**:583–90.

101. Apostolou I, Cumano A, Gachelin G, Kourilsky P. Evidence for two subgroups of CD4−CD8− NKT cells with distinct TCRαβ repertoires and differential distribution in lymphoid tissues. *J Immunol* (2000) **165**:2481–90.

102. Bernard D, Six A, Rigottier-Gois L, Messiaen S, Chilmonczyk S, Quillet E, et al. Phenotypic and functional similarity of gut intraepithelial and systemic T cells in a teleost fish. *J Immunol* (2006) **176**:3942–9.

103. Mancini S, Candéias SM, Fehling HJ, von Boehmer H, Jouvin-Marche E, Marche PN. TCR α-chain repertoire in pTα-deficient mice is diverse and developmentally regulated: implications for pre-TCR functions and TCRA gene rearrangement. *J Immunol* (1999) **163**:6053–9.

104. Halapi E, Werner A, Wahlström J, Österborg A, Jeddi-Tehrani M, Yi Q, et al. T cell repertoire in patients with multiple myeloma and monoclonal gammopathy of undetermined significance: clonal CD8+ T cell expansions are found preferentially in patients with a low tumor burden. *Eur J Immunol* (1997) **27**:2245–52. doi:10.1002/eji.1830270919

105. Brawand P, Cerottini JC, MacDonald HR. Hierarchal utilization of different T-cell receptor Vβ gene segments in the CD8+-T-cell response to an immunodominant Moloney leukemia virus-encoded epitope *in vivo*. *J Virol* (1999) **73**:9161–9.

106. Matsuzaki G, Takada H, Nomoto K. *Escherichia coli* infection induces only fetal thymus-derived γδ T cells at the infected site. *Eur J Immunol* (1999) **29**:3877–86. doi:10.1002/(SICI)1521-4141(199912)29:12<3877::AID-IMMU3877>3.3.CO;2-3

107. Collette A, Bagot S, Ferrandiz ME, Cazenave PA, Six A, Pied S. A profound alteration of blood TCRB repertoire allows prediction of cerebral malaria. *J Immunol* (2004) **173**:4568–75.

108. Douek DC, Betts MR, Brenchley JM, Hill BJ, Ambrozak DR, Ngai KL, et al. A novel approach to the analysis of specificity, clonality, and frequency of HIV-specific T cell responses reveals a potential mechanism for control of viral escape. *J Immunol* (2002) **168**:3099–104.

109. Lim A, Lemercier B, Wertz X, Pottier SL, Huetz F, Kourilsky P. Many human peripheral VH5-expressing IgM+ B cells display a unique heavy-chain rearrangement. *Int Immunol* (2008) **20**:105–16. doi:10.1093/intimm/dxm125

110. Voelkerding KV, Dames SA, Durtschi JD. Next-generation sequencing: from basic research to diagnostics. *Clin Chem* (2009) **55**:641–58. doi:10.1373/clinchem.2008.112789

111. McGinn S, Gut IG. DNA sequencing – spanning the generations. *Nat Biotechnol* (2013) **30**:366–72. doi:10.1016/j.nbt.2012.11.012

112. Weinstein JA, Jiang N, White RA III, Fisher DS, Quake SR. High-throughput sequencing of the zebrafish antibody repertoire. *Science* (2009) **324**:807–10. doi:10.1126/science.1170020

113. Robins HS, Campregher PV, Srivastava SK, Wacher A, Turtle CJ, Kahsai O, et al. Comprehensive assessment of T-cell receptor β-chain diversity in αβ T cells. *Blood* (2009) **114**:4099–107. doi:10.1182/blood-2009-04-217604

114. Freeman JD, Warren RL, Webb JR, Nelson BH, Holt RA. Profiling the T-cell receptor beta-chain repertoire by massively parallel sequencing. *Genome Res* (2009) **19**:1817–24. doi:10.1101/gr.092924.109

115. Wang C, Sanders CM, Yang Q, Schroeder HW, Wang E, Babrzadeh F, et al. High throughput sequencing reveals a complex pattern of dynamic interrelationships among human T cell subsets. *Proc Natl Acad Sci U S A* (2010) **107**:1518–23. doi:10.1073/pnas.0913939107

116. Warren RL, Freeman JD, Zeng T, Choe G, Munro S, Moore R, et al. Exhaustive T-cell repertoire sequencing of human peripheral blood samples reveals signatures of antigen selection and a directly measured repertoire size of at least 1 million clonotypes. *Genome Res* (2011) **21**:790–7. doi:10.1101/gr.115428.110

117. Venturi V, Quigley MF, Greenaway HY, Ng PC, Ende ZS, McIntosh T, et al. A mechanism for TCR sharing between T cell subsets and individuals revealed by pyrosequencing. *J Immunol* (2011) **186**:4285–94. doi:10.4049/jimmunol.1003898

118. Wu D, Sherwood A, Fromm JR, Winter SS, Dunsmore KP, Loh ML, et al. High-throughput sequencing detects minimal residual disease in acute T lymphoblastic leukemia. *Sci Transl Med* (2012) **4**:134ra63. doi:10.1126/scitranslmed.3003656

119. Föhse L, Suffner J, Suhre K, Wahl B, Lindner C, Lee CW, et al. High TCR diversity ensures optimal function and homeostasis of Foxp3+ regulatory T cells. *Eur J Immunol* (2011) **41**:3101–13. doi:10.1002/eji.201141986

120. Sherwood AM, Desmarais C, Livingston RJ, Andriesen J, Haussler M, Carlson CS, et al. Deep sequencing of the human TCRγ and TCRβ repertoires suggests that TCRβ rearranges after αβ and γδ T cell commitment. *Sci Transl Med* (2011) **3**:90ra61. doi:10.1126/scitranslmed.3002536

121. Cebula A, Seweryn M, Rempala GA, Pabla SS, McIndoe RA, Denning TL, et al. Thymus-derived regulatory T cells contribute to tolerance to commensal microbiota. *Nature* (2013) **497**:258–62. doi:10.1038/nature12079

122. Bashford-Rogers RJM, Palser AL, Huntly BJ, Rance R, Vassiliou GS, Follows GA, et al. Network properties derived from deep sequencing of human B-cell receptor repertoires delineate B-cell populations. *Genome Res* (2013) **23**:1874–84. doi:10.1101/gr.154815.113

123. Srivastava SK, Robins HS. Palindromic nucleotide analysis in human T cell receptor rearrangements. *PLoS One* (2012) **7**:e52250. doi:10.1371/journal.pone.0052250

124. Murugan A, Mora T, Walczak AM, Callan CG. Statistical inference of the generation probability of T-cell receptors from sequence repertoires. *Proc Natl Acad Sci U S A* (2012) **109**:16161–6. doi:10.1073/pnas.1212755109

125. Ndifon W, Gal H, Shifrut E, Aharoni R, Yissachar N, Waysbort N, et al. Chromatin conformation governs T-cell receptor Jβ gene segment usage. *Proc Natl Acad Sci U S A* (2012) **109**:15865–70. doi:10.1073/pnas.1203916109

126. Robins HS, Srivastava SK, Campregher PV, Turtle CJ, Andriesen J, Riddell SR, et al. Overlap and effective size of the human CD8+ T cell receptor repertoire. *Sci Transl Med* (2010) **2**:47ra64. doi:10.1126/scitranslmed.3001442

127. Prabakaran P, Chen W, Singarayan MG, Stewart CC, Streaker E, Feng Y, et al. Expressed antibody repertoires in human cord blood cells: 454 sequencing and IMGT/HighV-QUEST analysis of germline gene usage, junctional diversity, and somatic mutations. *Immunogenetics* (2012) **64**:337–50. doi:10.1007/s00251-011-0595-8

128. van Heijst JW, Ceberio I, Lipuma LB, Samilo DW, Wasilewski GD, Gonzales AM, et al. Quantitative assessment of T cell repertoire recovery after hematopoietic stem cell transplantation. *Nat Med* (2013) **19**:372–7. doi:10.1038/nm.3100

129. Meier J, Roberts C, Avent K, Hazlett A, Berrie J, Payne K, et al. Fractal organization of the human T cell repertoire in health and after stem cell transplantation. *Biol Blood Marrow Transplant* (2013) **19**:366–77. doi:10.1016/j.bbmt.2012.12.004

130. Ademokun A, Wu Y-C, Martin V, Mitra R, Sack U, Baxendale H, et al. Vaccination-induced changes in human B-cell repertoire and pneumococcal IgM and IgA antibody at different ages. *Aging Cell* (2011) **10**:922–30. doi:10.1111/j.1474-9726.2011.00732.x

131. Castro R, Jouneau L, Pham HP, Bouchez O, Giudicelli V, Lefranc MP, et al. Teleost fish mount complex clonal IgM and IgT responses in spleen upon systemic viral infection. *PLoS Pathog* (2013) **9**:e1003098. doi:10.1371/journal.ppat.1003098

132. Bousso P, Casrouge A, Altman JD, Haury M, Kanellopoulos J, Abastado JP, et al. Individual variations in the murine T cell response to a specific peptide reflect variability in naive repertoires. *Immunity* (1998) **9**:169–78. doi:10.1016/S1074-7613(00)80599-3

133. Lin MY, Welsh RM. Stability and diversity of T cell receptor repertoire usage during lymphocytic choriomeningitis virus infection of mice. *J Exp Med* (1998) **188**:1993–2005. doi:10.1084/jem.188.11.1993

134. Metzker ML. Sequencing technologies – the next generation. *Nat Rev Genet* (2010) **11**:31–46. doi:10.1038/nrg2626

135. Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, et al. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* (2005) **437**:376–80.

136. Rothberg JM, Hinz W, Rearick TM, Schultz J, Mileski W, Davey M, et al. An integrated semiconductor device enabling non-optical genome sequencing. *Nature* (2011) **475**:348–52. doi:10.1038/nature10242

137. Bolotin DA, Mamedov IZ, Britanova OV, Zvyagin IV, Shagin D, Ustyugova SV, et al. Next generation sequencing for TCR repertoire profiling: platform-specific features and correction algorithms. *Eur J Immunol* (2012) **42**:3073–83. doi:10.1002/eji.201242517

138. Bragg LM, Stone G, Butler MK, Hugenholtz P, Tyson GW. Shining a light on dark sequencing: characterising errors in Ion Torrent PGM data. *PLoS Comput Biol* (2013) **9**:e1003031. doi:10.1371/journal.pcbi.1003031

139. Shendure J, Aiden EL. The expanding scope of DNA sequencing. *Nat Biotechnol* (2012) **30**:1084–94. doi:10.1038/nbt.2421

140. Dash P, McClaren JL, Oguin TH III, Rothwell W, Todd B, Morris MY, et al. Paired analysis of TCRα and TCRβ chains at the single-cell level in mice. *J Clin Invest* (2011) **121**:288–95. doi:10.1172/JCI44752

141. DeKosky BJ, Ippolito GC, Deschner RP, Lavinder JJ, Wine Y, Rawlings BM, et al. High-throughput sequencing of the paired human immunoglobulin heavy and light chain repertoire. *Nat Biotechnol* (2013) **31**:166–9. doi:10.1038/nbt.2492

142. Turchaninova MA, Britanova OV, Bolotin DA, Shugay M, Putintseva EV, Staroverov DB, et al. Pairing of T-cell receptor chains via emulsion PCR. *Eur J Immunol* (2013) **43**:2507–15. doi:10.1002/eji.201343453

143. Plessy C, Desbois L, Fujii T, Carninci P. Population transcriptomics with single-cell resolution: a new field made possible by microfluidics: a technology for high throughput transcript counting and data-driven definition of cell types. *Bioessays* (2013) **35**:131–40. doi:10.1002/bies.201200093

144. Mehr R, Sternberg-Simon M, Michaeli M, Pickman Y. Models and methods for analysis of lymphocyte repertoire generation, development, selection and evolution. *Immunol Lett* (2012) **148**:11–22. doi:10.1016/j.imlet.2012.08.002

145. Pancer Z, Amemiya CT, Ehrhardt GR, Ceitlin J, Gartland GL, Cooper MD. Somatic diversification of variable lymphocyte receptors in the agnathan sea lamprey. *Nature* (2004) **430**:174–80. doi:10.1038/nature02740

146. Lefranc MP, Giudicelli V, Ginestoux C, Jabado-Michaloud J, Folch G, Bellahcene F, et al. IMGT, the international ImMunoGeneTics information system. *Nucleic Acids Res* (2009) **37**:D1006–12. doi:10.1093/nar/gkn838

147. Giudicelli V, Lefranc MP. IMGT-ONTOLOGY 2012. *Front Genet* (2012) **3**:79. doi:10.3389/fgene.2012.00079

148. Alamyar E, Giudicelli V, Li S, Duroux P, Lefranc MP. IMGT/HighV-QUEST: the IMGT(R) web portal for immunoglobulin (IG) or antibody and T cell receptor (TR) analysis from NGS high throughput and deep sequencing. *Immunome Res* (2012) **8**:26. doi:10.1007/978-1-61779-842-9_32

149. Watson FL, Puttmann-Holgado R, Thomas F, Lamar DL, Hughes M, Kondo M, et al. Extensive diversity of Ig-superfamily proteins in the immune system of insects. *Science* (2005) **309**:1874–8. doi:10.1126/science.1116887

150. Du Pasquier L. Insects diversify one molecule to serve two systems. *Science* (2005) **309**:1826–7. doi:10.1126/science.1118828

151. Zhang SM, Adema CM, Kepler TB, Loker ES. Diversification of Ig superfamily genes in an invertebrate. *Science* (2004) **305**:251–4. doi:10.1126/science.1088069

152. Philipp EER, Kraemer L, Melzner F, Poustka AJ, Thieme S, Findeisen U, et al. Massively parallel RNA sequencing identifies a complex immune gene repertoire in the Lophotrochozoan *Mytilus edulis*. *PLoS One* (2012) **7**:e33091. doi:10.1371/journal.pone.0033091

153. Matsushima N, Tanaka T, Enkhbayar P, Mikami T, Taga M, Yamada K, et al. Comparative sequence analysis of leucine-rich repeats (LRRs) within vertebrate toll-like receptors. *BMC Genomics* (2007) **8**:124. doi:10.1186/1471-2164-8-124

154. Matsushima N, Miyashita H, Mikami T, Kuroki Y. A nested leucine rich repeat (LRR) domain: the precursor of LRRs is a ten or eleven residue motif. *BMC Microbiol* (2010) **10**:235. doi:10.1186/1471-2180-10-235

155. Kaas Q, Ehrenmann F, Lefranc MP. IG, TR and IgSF, MHC and MhcSF: what do we learn from the IMGT Colliers de Perles? *Brief Funct Genomic Proteomic* (2007) **6**:253–64. doi:10.1093/bfgp/elm032

156. Bomberger C, Singh-Jairam M, Rodey G, Guerriero A, Yeager AM, Fleming WH, et al. Lymphoid reconstitution after autologous PBSC transplantation with FACS-sorted CD34+ hematopoietic progenitors. *Blood* (1998) **91**:2588–600.

157. Gorochov G, Neumann AU, Kereveur A, Parizot C, Li TS, Katlama C, et al. Perturbation of CD4+ and CD8+ T-cell repertoires during progression to AIDS and regulation of the CD4+ repertoire during antiviral therapy. *Nat Med* (1998) **4**:215–21. doi:10.1038/nm0298-215

158. Wu CJ, Chillemi A, Alyea EP, Orsini E, Neuberg D, Soiffer RJ, et al. Reconstitution of T-cell receptor repertoire diversity following T-cell depleted allogeneic bone marrow transplantation is related to hematopoietic chimerism. *Blood* (2000) **95**:352–9.

159. Hori S, Collette A, Demengeot J, Stewart J. A new statistical method for quantitative analyses: application to the precise quantification of T cell receptor repertoires. *J Immunol Methods* (2002) **268**:159–70. doi:10.1016/S0022-1759(02)00187-4

160. Collette A, Six A. ISEApeaks: an excel platform for GeneScan and Immunoscope data retrieval, management and analysis. *Bioinformatics* (2002) **18**:329–30. doi:10.1093/bioinformatics/18.2.329

161. Guillet M, Brouard S, Gagne K, Sebille F, Cuturi MC, Delsuc MA, et al. Different qualitative and quantitative regulation of Vβ TCR transcripts during early acute allograft rejection and tolerance induction. *J Immunol* (2002) **168**:5088–95.

162. Peggs KS, Verfuerth S, D'Sa S, Yong K, Mackinnon S. Assessing diversity: immune reconstitution and T-cell receptor BV spectratype analysis following stem cell transplantation. *Br J Haematol* (2003) **120**:154–65. doi:10.1046/j.1365-2141.2003.04036.x

163. Long SA, Khalili J, Ashe J, Berenson R, Ferrand C, Bonyhadi M. Standardized analysis for the quantification of Vbeta CDR3 T-cell receptor diversity. *J Immunol Methods* (2006) **317**:100–13. doi:10.1016/j.jim.2006.09.015

164. Miqueu P, Guillet M, Degauque N, Dore JC, Soulillou JP, Brouard S. Statistical analysis of CDR3 length distributions for the assessment of T and B cell repertoire biases. *Mol Immunol* (2007) **44**:1057–64. doi:10.1016/j.molimm.2006.06.026

165. Guillet M, Sebille F, Soulillou JP. TCR usage in naive and committed alloreactive cells: implications for the understanding of TCR biases in transplantation. *Curr Opin Immunol* (2001) **13**:566–71. doi:10.1016/S0952-7915(00)00260-0

166. Collette A, Cazenave PA, Pied S, Six A. New methods and software tools for high throughput CDR3 spectratyping. Application to T lymphocyte repertoire modifications during experimental malaria. *J Immunol Methods* (2003) **278**:105–16. doi:10.1016/S0022-1759(03)00225-4

167. Sassi A, Largueche-Darwaz B, Collette A, Six A, Laouini D, Cazenave PA, et al. Mechanisms of the natural reactivity of lymphocytes from noninfected individuals to membrane-associated *Leishmania infantum* antigens. *J Immunol* (2005) **174**:3598–607.

168. Castro R, Takizawa F, Chaara W, Lunazzi A, Dang TH, Koellner B, et al. Contrasted TCRβ diversity of CD8+ and CD8− T cells in rainbow trout. *PLoS One* (2013) **8**:e60175. doi:10.1371/journal.pone.0060175

169. Kepler TB, He M, Tomfohr JK, Devlin BH, Sarzotti M, Markert ML. Statistical analysis of antigen receptor spectratype data. *Bioinformatics* (2005) **21**:3394–400. doi:10.1093/bioinformatics/bti539

170. He M, Tomfohr JK, Devlin BH, Sarzotti M, Markert ML, Kepler TB. SpA: web-accessible spectratype analysis: data management, statistical analysis and visualization. *Bioinformatics* (2005) **21**:3697–9. doi:10.1093/bioinformatics/bti600

171. Liu C, He M, Rooney B, Kepler TB, Chao NJ. Longitudinal analysis of T-cell receptor variable beta chain repertoire in patients with acute graft-versus-host disease after allogeneic stem cell transplantation. *Biol Blood Marrow Transplant* (2006) **12**:335–45. doi:10.1016/j.bbmt.2005.09.019

172. Lefranc MP. From IMGT-ONTOLOGY CLASSIFICATION Axiom to IMGT standardized gene and allele nomenclature: for immunoglobulins (IG) and T cell receptors (TR). *Cold Spring Harb Protoc* (2011) **2011**:627–32. doi:10.1101/pdb.ip84

173. Lefranc MP. From IMGT-ONTOLOGY DESCRIPTION axiom to IMGT standardized labels: for immunoglobulin (IG) and T cell receptor (TR) sequences and structures. *Cold Spring Harb Protoc* (2011) **2011**:614–26. doi:10.1101/pdb.ip84

174. Lefranc MP. From IMGT-ONTOLOGY IDENTIFICATION axiom to IMGT standardized keywords: for immunoglobulins (IG), T cell receptors (TR), and conventional genes. *Cold Spring Harb Protoc* (2011) **2011**:604–13. doi:10.1101/pdb.ip84

175. Lefranc MP. IMGT unique numbering for the variable (V), constant (C), and groove (G) domains of IG, TR, MH, IgSF, and MhSF. *Cold Spring Harb Protoc* (2011) **2011**:633–42. doi:10.1101/pdb.ip85

176. Giudicelli V, Chaume D, Lefranc MP. IMGT/V-QUEST, an integrated software program for immunoglobulin and T cell receptor V-J and V-D-J rearrangement analysis. *Nucleic Acids Res* (2004) **32**:W435–40. doi:10.1093/nar/gkh412

177. Brochet X, Lefranc MP, Giudicelli V. IMGT/V-QUEST: the highly customized and integrated system for IG and TR standardized V-J and V-D-J sequence analysis. *Nucleic Acids Res* (2008) **36**:W503–8. doi:10.1093/nar/gkn316

178. Gaëta BA, Malming HR, Jackson KJL, Bain ME, Wilson P, Collins AM. iHMMune-align: hidden Markov model-based alignment and identification of germline genes in rearranged immunoglobulin gene sequences. *Bioinformatics* (2007) **23**:1580–7. doi:10.1093/bioinformatics/btm147

179. Rogosch T, Kerzel S, Hoi KH, Zhang Z, Maier RF, Ippolito GC, et al. Immunoglobulin analysis tool: a novel tool for the analysis of human and mouse heavy and light chain transcripts. *Front Immunol* (2012) **3**:176. doi:10.3389/fimmu.2012.00176

180. Ye J, Ma N, Madden TL, Ostell JM. IgBLAST: an immunoglobulin variable domain sequence analysis tool. *Nucleic Acids Res* (2013) **41**:W34–40. doi:10.1093/nar/gkt382

181. Thomas N, Heather J, Ndifon W, Shawe-Taylor J, Chain B. Decombinator: a tool for fast, efficient gene assignment in T-cell receptor sequences using a finite state machine. *Bioinformatics* (2013) **29**:542–50. doi:10.1093/bioinformatics/btt004

182. Pham HP, Manuel M, Petit N, Klatzmann D, Cohen-Kaminsky S, Six A, et al. Half of the T-cell repertoire combinatorial diversity is genetically determined in humans and humanized mice. *Eur J Immunol* (2012) **42**:760–70. doi:10.1002/eji.201141798

183. Eisenstein M. Personalized, sequencing-based immune profiling spurs startups. *Nat Biotechnol* (2013) **31**:184–6. doi:10.1038/nbt0313-184b

184. Stahl D, Lacroix-Desmazes S, Barreau C, Sibrowski W, Kazatchkine MD, Kaveri SV. Altered antibody repertoires of plasma IgM and IgG toward nonself antigens in patients with warm autoimmune hemolytic anemia. *Hum Immunol* (2001) **62**:348–61. doi:10.1016/S0198-8859(01)00225-7

185. Magurran AE. *Measuring Biological Diversity*. Oxford: Wiley-Blackwell (2004).

186. Colwell RK. *EstimateS: Statistical Estimation of Species Richness and Shared Species from Samples. [Version 9]. User's Guide and application* (2013). Available from: http://purl.oclc.org/estimates

187. Wu TT, Kabat EA. An analysis of the sequence of the variable regions of Bence Jones proteins and myeloma light chains and their implications for antibody complementarity. *J Exp Med* (1970) **132**:211–50. doi:10.1084/jem.132.2.211

188. Jores R, Alzari PM, Meo T. Resolution of hypervariable regions in T-cell receptor β chains by a modified Wu-Kabat index of amino acid diversity. *Proc Natl Acad Sci U S A* (1990) **87**:9138–42. doi:10.1073/pnas.87.23.9138

189. Stewart JJ, Lee CY, Ibrahim S, Watts P, Shlomchik M, Weigert M, et al. A Shannon entropy analysis of immunoglobulin and T cell receptor. *Mol Immunol* (1997) **34**:1067–82. doi:10.1016/S0161-5890(97)00130-2

190. Thomas PG, Handel A, Doherty PC, La Gruta NL. Ecological analysis of antigen-specific CTL repertoires defines the relationship between naïve and immune T-cell populations. *Proc Natl Acad Sci U S A* (2013) **110**:1839–44. doi:10.1073/pnas.1222149110

191. Li S, Lefranc MP, Miles J, Alamyar E, Giudicelli V, Duroux P, et al. IMGT/HighV-QUEST paradigm for T cell receptor IMGT clonotype diversity and next generation repertoire immunoprofiling. *Nat Commun* (2013) **4**:2333. doi:10.1038/ncomms3333

192. Conrad JA, Ramalingam RK, Duncan CB, Smith RM, Wei J, Barnett L, et al. Antiretroviral therapy reduces the magnitude and T cell receptor repertoire

diversity of HIV-specific T cell responses without changing T cell clonotype dominance. *J Virol* (2012) **86**:4213–21. doi:10.1128/JVI.06000-11

193. Koning D, Costa AI, Hoof I, Miles JJ, Nanlohy NM, Ladell K, et al. CD8+ TCR repertoire formation is guided primarily by the peptide component of the antigenic complex. *J Immunol* (2013) **190**:931–9. doi:10.4049/jimmunol.1202466

194. Johnson PLF, Yates AJ, Goronzy JJ, Antia R. Peripheral selection rather than thymic involution explains sudden contraction in naive CD4 T-cell diversity with age. *Proc Natl Acad Sci U S A* (2012) **109**:21432–7. doi:10.1073/pnas.1209283110

195. Perelson AS, Weisbuch G. Immunology for physicist. *Rev Mod Phys* (1997) **69**:1219–67. doi:10.1103/RevModPhys.69.1219

196. Perelson AS, Oster GF. Theoretical studies of clonal selection: minimal antibody repertoire size and reliability of self-non-self discrimination. *J Theor Biol* (1979) **81**:645–70. doi:10.1016/0022-5193(79)90275-3

197. Percus JK, Percus OE, Perelson AS. Predicting the size of the T-cell receptor and antibody combining region from consideration of efficient self-nonself discrimination. *Proc Natl Acad Sci U S A* (1993) **90**:1691–5. doi:10.1073/pnas.90.5.1691

198. Bergstrom CT, Antia R. How do adaptive immune systems control pathogens while avoiding autoimmunity? *Trends Ecol Evol* (2006) **21**:22–8. doi:10.1016/j.tree.2005.11.008

199. Perelson AS. Modelling viral and immune system dynamics. *Nat Rev Immunol* (2002) **2**:28–36. doi:10.1038/nri700

200. Antia R, Ganusov VV, Ahmed R. The role of models in understanding CD8+ T-cell memory. *Nat Rev Immunol* (2005) **5**:101–11. doi:10.1038/nri1550

201. Thomas-Vaslin V, Six A, Bellier B, Klatzmann D. Lymphocytes dynamics repertoires, modeling. In: Dubitzky W, Wolkenhauer O, Cho K-H, Yokota H editors. *Encyclopedia of Systems Biology*. Heidelberg: Springer Verlag (2013). p. 1149–52. doi:10.1007/978-1-4419-9863-7_96

202. De Boer RJ, Homann D, Perelson AS. Different dynamics of CD4+ and CD8+ T cell responses during and after acute lymphocytic choriomeningitis virus infection. *J Immunol* (2003) **171**:3928–35.

203. Verkoczy LK, Martensson AS, Nemazee D. The scope of receptor editing and its association with autoimmunity. *Curr Opin Immunol* (2004) **16**:808–14. doi:10.1016/j.coi.2004.09.017

204. Wucherpfennig KW, Allen PM, Celada F, Cohen IR, De BR, Garcia KC, et al. Polyspecificity of T cell and B cell receptor recognition. *Semin Immunol* (2007) **19**:216–24. doi:10.1016/j.smim.2007.02.012

205. Germain RN, Meier-Schellersheim M, Nita-Lazar A, Fraser IDC. Systems biology in immunology: a computational modeling perspective. *Annu Rev Immunol* (2011) **29**:527–85. doi:10.1146/annurev-immunol-030409-101317

206. Emonet T, Altan-Bonnet G. Systems immunology: a primer for biophysicists. In: Egelman E editor. *Comprehensive Biophysics*. New York: Academic Press (2012). p. 389–413.

207. Quigley MF, Greenaway HY, Venturi V, Lindsay R, Quinn KM, Seder RA, et al. Convergent recombination shapes the clonotypic landscape of the naïve T-cell repertoire. *Proc Natl Acad Sci U S A* (2010) **107**:19414–9. doi:10.1073/pnas.1010586107

208. Martins VC, Ruggiero E, Schlenner SM, Madan V, Schmidt M, Fink PJ, et al. Thymus-autonomous T cell development in the absence of progenitor import. *J Exp Med* (2012) **209**:1409–17. doi:10.1084/jem.20120846

209. Farmer JD, Packard NH, Perelson AS. The immune system, adaptation and machine learning. *Physica D* (1986) **22**:187–204. doi:10.1016/0167-2789(86)90240-X

210. De Boer RJ, Perelson AS. Size and connectivity as emergent properties of a developing immune network. *J Theor Biol* (1991) **149**:381–424. doi:10.1016/S0022-5193(05)80313-3

211. Celada F, Seiden PE. A computer model of cellular interactions in the immune system. *Immunol Today* (1992) **13**:56–62. doi:10.1016/0167-5699(92)90135-T

212. Goldstein B, Faeder JR, Hlavacek WS. Mathematical and computational models of immune-receptor signalling. *Nat Rev Immunol* (2004) **4**:445–56. doi:10.1038/nri1374

213. Chao A, Chazdon RL, Colwell RK, Shen TJ. A new statistical approach for assessing similarity of species composition with incidence and abundance data. *Ecol Lett* (2005) **8**:148–59. doi:10.1111/j.1461-0248.2004.00707.x

214. Kosmrlj A, Jha AK, Huseby ES, Kardar M, Chakraborty AK. How the thymus designs antigen-specific and self-tolerant T cell receptor sequences. *Proc Natl Acad Sci U S A* (2008) **105**:16671–6. doi:10.1073/pnas.0808081105

215. Kosmrlj A, Chakraborty AK, Kardar M, Shakhnovich EI. Thymic selection of T-cell receptors as an extreme value problem. *Phys Rev Lett* (2009) **103**:068103. doi:10.1103/PhysRevLett.103.068103

216. Verhagen J, Genolet R, Britton GJ, Stevenson BJ, Sabatos-Peyton CA, Dyson J, et al. CTLA-4 controls the thymic development of both conventional and regulatory T cells through modulation of the TCR repertoire. *Proc Natl Acad Sci U S A* (2013) **110**:E221–30. doi:10.1073/pnas.1208573110

217. Mora T, Walczak AM, Bialek W, Callan CG. Maximum entropy models for antibody diversity. *Proc Natl Acad Sci U S A* (2010) **107**:5405–10. doi:10.1073/pnas.1001705107

218. Baum PD, Venturi V, Price DA. Wrestling with the repertoire: the promise and perils of next generation sequencing for antigen receptors. *Eur J Immunol* (2012) **42**:2834–9. doi:10.1002/eji.201242999

219. Robins H, Desmarais C, Matthis J, Livingston R, Andriesen J, Reijonen H, et al. Ultra-sensitive detection of rare T cell clones. *J Immunol Methods* (2012) **375**:14–9. doi:10.1016/j.jim.2011.09.001

220. Zipf GK. *Human Behaviour and the Principle of Least Effort: An Introduction to Human Ecology*. Cambridge, MA: Addison-Wesley (1949).

221. Sepulveda N, Paulino CD, Carneiro J. Estimation of T-cell repertoire diversity and clonal size distribution by Poisson abundance models. *J Immunol Methods* (2009) **353**:124–37. doi:10.1016/j.jim.2009.11.009

222. Rempala GA, Seweryn M, Ignatowicz L. Model for comparative analysis of antigen receptor repertoires. *J Theor Biol* (2011) **269**:1–15. doi:10.1016/j.jtbi.2010.10.001

223. Ben-Hamo R, Efroni S. The whole-organism heavy chain B cell repertoire from Zebrafish self-organizes into distinct network features. *BMC Syst Biol* (2011) **5**:27. doi:10.1186/1752-0509-5-27

224. Bleakley K, Lefranc MP, Biau G. Recovering probabilities for nucleotide trimming processes for T cell receptor TRA and TRG V-J junctions analyzed with IMGT tools. *BMC Bioinformatics* (2008) **9**:408. doi:10.1186/1471-2105-9-408

225. Kleinstein SH, Louzoun Y, Shlomchik MJ. Estimating hypermutation rates from clonal tree data. *J Immunol* (2003) **171**:4639–49.

226. Anderson SM, Khalil A, Uduman M, Hershberg U, Louzoun Y, Haberman AM, et al. Taking advantage: high-affinity B cells in the germinal center have lower death rates, but similar rates of division, compared to low-affinity cells. *J Immunol* (2009) **183**:7314–25. doi:10.4049/jimmunol.0902452

227. Uduman M, Yaari G, Hershberg U, Stern JA, Shlomchik MJ, Kleinstein SH. Detecting selection in immunoglobulin sequences. *Nucleic Acids Res* (2011) **39**:W499–504. doi:10.1093/nar/gkr413

228. Yaari G, Uduman M, Kleinstein SH. Quantifying selection in high-throughput Immunoglobulin sequencing data sets. *Nucleic Acids Res* (2012) **40**:e134. doi:10.1093/nar/gks457

229. Rock EP, Sibbald PR, Davis MM, Chien Y. CDR3 length in antigen-specific immune receptors. *J Exp Med* (1994) **179**:323–8. doi:10.1084/jem.179.1.323

230. Hyatt G, Melamed R, Park R, Seguritan R, Laplace C, Poirot L, et al. Gene expression microarrays: glimpses of the immunological genome. *Nat Immunol* (2006) **7**:686–91. doi:10.1038/ni0706-686

231. Han Q, Bagheri N, Bradshaw EM, Hafler DA, Lauffenburger DA, Love JC. Polyfunctional responses by human T cells result from sequential release of cytokines. *Proc Natl Acad Sci U S A* (2012) **109**:1607–12. doi:10.1073/pnas.1117194109

232. Bendall SC, Simonds EF, Qiu P, Amir E, Krutzik PO, Finck R, et al. Single-cell mass cytometry of differential immune and drug responses across a human hematopoietic continuum. *Science* (2011) **332**:687–96. doi:10.1126/science.1198704

233. Newell EW, Sigal N, Bendall SC, Nolan GP, Davis MM. Cytometry by time-of-flight shows combinatorial cytokine expression and virus-specific cell niches within a continuum of CD8+ T cell phenotypes. *Immunity* (2012) **36**:142–52. doi:10.1016/j.immuni.2012.01.002

234. Schepers K, Swart E, van Heijst JW, Gerlach C, Castrucci M, Sie D, et al. Dissecting T cell lineage relationships by cellular barcoding. *J Exp Med* (2008) **205**:2309–18. doi:10.1084/jem.20072462

235. Sumen C, Mempel TR, Mazo IB, von Andrian UH. Intravital microscopy: visualizing immunity in context. *Immunity* (2004) **21**:315–29. doi:10.1016/j.immuni.2004.08.006

236. Marangoni F, Murooka TT, Manzo T, Kim EY, Carrizosa E, Elpek NM, et al. The transcription factor NFAT exhibits signal memory during serial T cell interactions with antigen-presenting cells. *Immunity* (2013) **38**:237–49. doi:10.1016/j.immuni.2012.09.012

237. Flatz L, Roychoudhuri R, Honda M, Filali-Mouhim A, Goulet JP, Kettaf N, et al. Single-cell gene-expression profiling reveals qualitatively distinct CD8 T cells elicited by different gene-based vaccines. *Proc Natl Acad Sci U S A* (2011) **108**:5724–9. doi:10.1073/pnas.1013084108

238. Mehr R. Modeling and analysis of the meta-population dynamics of lymphocyte repertoires. *J Comput Appl Math* (2005) **184**:223–41. doi:10.1016/j.cam.2004.07.033

239. Ciupe SM, Devlin BH, Markert ML, Kepler TB. The dynamics of T-cell receptor repertoire diversity following thymus transplantation for DiGeorge anomaly. *PLoS Comput Biol* (2009) **5**:e1000396. doi:10.1371/journal.pcbi.1000396

240. Stirk ER, Molina-Paris C, van den Berg HA. Stochastic niche structure and diversity maintenance in the T cell repertoire. *J Theor Biol* (2008) **255**:237–49. doi:10.1016/j.jtbi.2008.07.017

241. Benoist C, Germain RN, Mathis D. A plaidoyer for 'systems immunology.' *Immunol Rev* (2006) **210**:229–34. doi:10.1111/j.0105-2896.2006.00374.x

242. Cohen IR. Autoantibody repertoires, natural biomarkers, and system controllers. *Trends Immunol* (2013). doi:10.1016/j.it.2013.05.003

243. Petrausch U, Haley D, Miller W, Floyd K, Urba WJ, Walker E. Polychromatic flow cytometry: a rapid method for the reduction and analysis of complex multiparameter data. *Cytometry* (2006) **69A**:1162–73. doi:10.1002/cyto.a.20342

244. Hofmann M, Zerwes HG. Identification of organ-specific T cell populations by analysis of multiparameter flow cytometry data using DNA-chip analysis software. *Cytometry A* (2006) **69**:533–40.

245. Lugli E, Pinti M, Troiano L, Nasi M, Patsekin V, Robinson JP, et al. Subject classification obtained by cluster analysis and principal component analysis applied to flow cytometric data. *Cytometry A* (2007) **71A**:334–44. doi:10.1002/cyto.a.20387