Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

Identification of T Cell Receptor Repertoire Signatures in Age, Autoimmune Conditions and Cancer using Next-Generation Sequencing

Doctoral thesis for obtaining the academic degree

Doktor rer. nat.

in the Department of Biology,
Faculty of Mathematics, Informatics and Natural Sciences
at the University of Hamburg

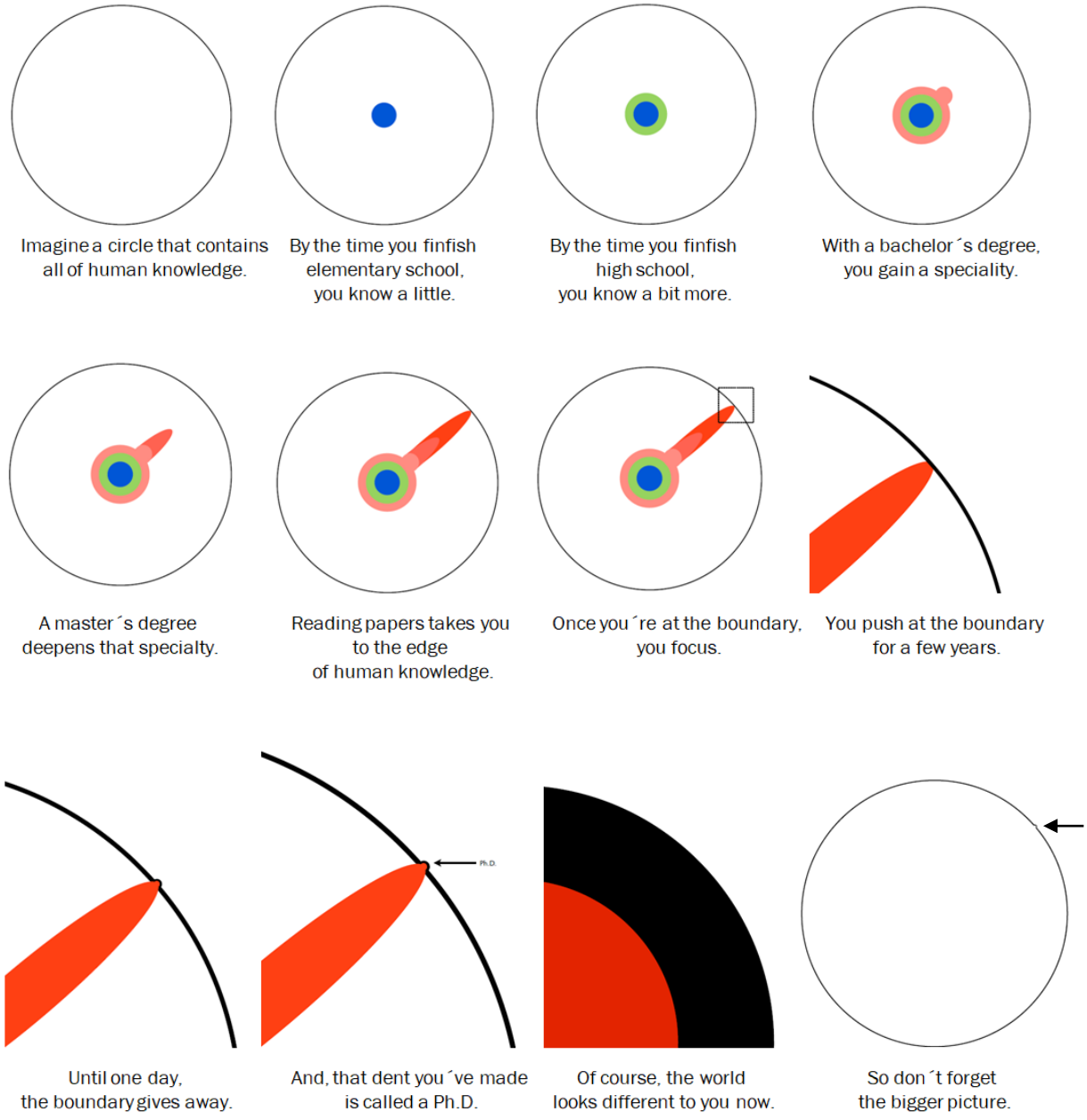
submitted by

Donjetë Simnica, née Statovci

September 2020

Reviewers of the dissertation:
Prof. Dr. med. Mascha Binder
Prof. Dr. rer. nat. Julia Kehr
Disputation: 11.12.2020

The Illustrated guide to a Ph.D.



“Keep pushing.”

Matt Might, a professor in Computer Science at the University of Utah, created The Illustrated Guide to a Ph.D. to explain what a Ph.D. is to new and aspiring graduate students.

ACKNOWLEDGEMENTS

My father used to tell me that receiving your doctorate, does not mean that you are “done”, it is rather the beginning. Because a doctorate, he reasoned, is just the “key” to the lab. This dissertation is the result of my learning curve and of the effort and help of so many people. I am looking forward to continue learning and contributing to push boundaries.

I am grateful to my supervisor Mascha who gave me the opportunity to be part of her lab. She has been a role model to me offering valuable advice yet giving me the space to shape and refine myself as scientist. Thank you for the inspiring discussions, your contagious excitement about science, the learning mindset that you cultivated in the lab and the great laughs even during stressful periods.

Thanks to my great colleagues who make being at work fun and from whom I have learned so much: Anna, Janina, Barbara, Anita and Nuray from the UKE as well as Lisa P., Edith, Marta and Christoph W. from the UKH. I really appreciate your support and help be it with teaching me how to prepare libraries for NGS, properly calculating concentrations of primers or with complex R scripts. Special thanks to Lisa (Master-Lisa) and Christoph (THE-Nature-Christoph) with whom I have gone through some challenging times (like our lab move to Halle) and who have the greatest anecdotes to share which fascinate me every time! Thank you for the inspiring talks - professionally and personally, the great team spirit and for your support during the writing of this thesis!

I owe many thanks to my dearest friends. Your support, patience, love and understanding throughout all my career decisions, which often meant that I lived far away from you, have given me strength and endurance to pass even the most challenging situations.

To my parents who faced great uncertainty by leaving everything familiar behind to provide my brothers and me with a better future in our adopted home Germany: thank you for your trust in our abilities and your unconditional love. To my brothers, their spouses and Elira and Hana: thank you for being part of my life and fulfilling my heart with joy and love.

Thank you Fatos, my dear soul mate, for patiently listing to all the science talk at breakfast; for being excited with me when a paper goes into revision; for helping me to reflect and focus on what really matters. You are a wonderful human being! Your unconditional support, your love and your great sense of humor energize me. I am so lucky to have you in my life!

Finally, I would like to express my gratitude to all the patients who donated their valuable samples without which this work would have been impossible. Although our work might not be of immediate help, I hope that it will contribute prospectively to improve the life of patients.

STATUTORY DECLARATION

This doctoral thesis was conducted at the Department for Internal Medicine II of the University Medical Center Hamburg-Eppendorf as well as the Department for Internal Medicine IV at the University Hospital Halle (Saale) under the supervision of Prof. Dr. med. Mascha Binder in the time between January 2017 and August 2020.

I confirm that I wrote the present dissertation on my own without using any other than the declared sources, references and tools.

……

EIDESSTATTLICHE VERSICHERUNG

Die vorliegende Doktorarbeit entstand zwischen Januar 2017 und August 2020 in der II. Medizinischen Klinik für Onkologie, Hämatologie und Knochenmarktransplantation am Universitätsklinikum Hamburg-Eppendorf und in der Klinik für Innere Medizin IV des Universitätsklinikums Halle (Saale) unter der Leitung von Prof. Dr. med. Mascha Binder.

Hiermit erkläre ich an Eides statt, dass ich die vorliegende Dissertationsschrift selbst verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

Halle (Saale), 04.09.2020 

TABLE OF CONTENTS

ABSTRACT

The immune system is an interactive network composed of cells, tissues and soluble factors which has evolved to protect the host from invading pathogens. T cells are specialized cells of the adaptive arm of the immune system and recognize antigens via their T cell receptor (TCR). The TCR is encoded by the random and irreversible recombination of V(D)J gene segments, a process resulting in a highly diverse repertoire of TCRs. The large diversity of the repertoire offers recognition of, and therefore immunological protection against, a wide range of antigens for the host. The clonal composition of the immune repertoire allows insights into the current state as well as the antigen exposure history of an individual and can be determined by next-generation sequencing (NGS).

Increasing age and diseases like cancer can compromise this fine-tuned network thus leading to chronic inflammation, autoimmunity or insufficient responses towards harmful antigens.

The work presented here discusses age-, cancer and autoimmunity-associated alterations of the TCR repertoire and summarizes fundamental aspects of T cell immunity as well as applications and challenges of adaptive immune receptor repertoire sequencing (AIRR seq).

In the first publication, which arose from this thesis, we profiled more than 300 peripheral blood TCR repertoires from healthy individuals as well as patients with cancer of all age groups. These analyses revealed a continuous age-dependent decline of T cell diversity which was accompanied by a decrease in T cell richness beginning at the fourth decade of life. Moreover, we observed that patients with cancer who received chemotherapy re-established their pre-treatment T cell diversity. Interestingly, the regeneration of the repertoire complexity after chemotherapy was best explained by rebound thymic activity rather than through recovery of T cell counts by peripheral clonal expansion only.

Taken together our data indicate that TCR metrics deteriorate gradually with increasing age, but age-specific repertoire metrics are restored after chemotherapy even in elderly patients with cancer.

In the second publication, we profiled the TCR repertoires of patients with autoimmune cytopenias (AIC) to understand if these patients present disease-specific immunological signatures that could reveal pathophysiological clues and eventually be exploited as blood-based biomarker. We analyzed 25 newly diagnosed patients with primary or secondary (lymphoma-associated) AIC as well as three reference cohorts composed of age- and sex-matched healthy controls, patients with active autoimmune hepatitis (AIH; another primary autoimmune disease) as well as patients with chronic lymphocytic leukemia (CLL, without autoimmune complication). Global TCR repertoire metrics like diversity and clonality as well as VJ gene usage distribution showed uniform characteristics for patients with lymphoma, but no AIC-specific signature. However, clustering of T cells with overlapping antigen specificity using the GLIPH algorithm (grouping lymphocyte interaction by paratope hotspots) revealed a considerable lack of T cell clusters in patients with primary autoimmune disease (AIC as well as AIH) as compared to healthy donors. The functionality of these clusters will have to be elucidated in future studies. Nevertheless, the signature of T cell cluster loss may represent a readily accessible biomarker for autoimmune conditions.

Taken together, AIRR seq has deepened our understanding of the adaptive immune response and has helped thus far to answer pressing questions in fields of immunodiagnostics, vaccines, cancer immunotherapy, and antibody engineering. As technical and computational challenges are being solved, this technology will come to realize its true potential.

ZUSAMMENFASSUNG

Das Immunsystem ist ein interaktives Netzwerk aus Zellen, Geweben und löslichen Faktoren, das sich entwickelt hat, um den Wirt vor Krankheitserregern zu schützen. T-Zellen sind spezialisierte Zellen des adaptiven Arms des Immunsystems und erkennen Antigene über ihren T-Zell-Rezeptor (TCR). Der TCR wird durch die zufällige und irreversible Rekombination von V(D)J-Gensegmenten kodiert; ein Prozess, der zu einem sehr vielfältigen Repertoire an TCRs führt. Die große Vielfalt des Repertoires bietet dem Wirt die Erkennung und damit den immunologischen Schutz gegen eine Vielzahl von Antigenen. Die klonale Zusammensetzung des Immunrepertoires erlaubt Einblicke in den aktuellen Zustand sowie die Antigenexpositionsgeschichte eines Individuums und kann durch Next-Generation-Sequenzierung (NGS) bestimmt werden.

Zunehmendes Alter und Krankheiten wie Krebs können dieses fein abgestimmte Netzwerk beeinträchtigen und so zu chronischer Entzündung, Autoimmunität oder unzureichenden Reaktionen auf schädliche Antigene führen.

Die hier vorgestellte Arbeit diskutiert alters-, krebs- und autoimmunitätsassoziierte Veränderungen des TCR-Repertoires und fasst grundlegende Aspekte der T-Zell-Immunität sowie Anwendungen und Herausforderungen der Sequenzierung des adaptiven Immunrezeptor-Repertoires (AIRR seq) zusammen.

In der ersten Publikation, die aus dieser Arbeit hervorging, sequenzierten wir mehr als 300 TCR-Repertoires aus peripherem Blut von gesunden Personen sowie von Patienten mit Krebs aller Altersgruppen. Diese Analysen zeigten einen kontinuierlichen altersabhängigen Rückgang der T-Zell-Diversität, der ab dem vierten Lebensjahrzehnt mit einer Abnahme der T-Zell-Abundanz einherging. Darüber hinaus beobachteten wir, dass Krebspatienten, die eine Chemotherapie erhielten, ihre Vortherapie-T-Zell-Diversität wiederherstellten. Interessanterweise ließ sich die Regeneration der Repertoirekomplexität nach der Chemotherapie am besten durch die Produktion neuer T Zellen aus dem Thymus erklären

und weniger durch die Wiederherstellung der T-Zellzahl nur durch periphere klonale Expansion.

Zusammengenommen deuten unsere Daten darauf hin, dass sich die TCR-Metriken mit zunehmendem Alter allmählich verschlechtern, aber die altersspezifischen Repertoire-Metriken selbst bei älteren Patienten mit Krebs nach der Chemotherapie wiederhergestellt werden können.

In der zweiten Publikation analysierten wir die TCR-Repertoires von Patienten mit autoimmunen Zytopenien (AIC), um zu verstehen, ob diese Patienten krankheitsspezifische immunologische Signaturen aufweisen, die pathophysiologische Hinweise aufzeigen und schließlich als blutbasierter Biomarker genutzt werden könnten. Wir analysierten 25 neu diagnostizierte Patienten mit primärer oder sekundärer (Lymphom-assoziierter) AIC sowie drei Referenzkohorten, die sich aus gesunden Kontrollen (angepasst für Alter und Geschlecht), Patienten mit aktiver Autoimmunhepatitis (AIH; eine weitere primäre Autoimmunerkrankung) sowie Patienten mit chronischer lymphatischer Leukämie (CLL, ohne Autoimmunkomplikation) zusammensetzten. Globale TCR-Repertoire-Metriken wie Diversität und Klonalität sowie die Verteilung der VJ-Genverwendung zeigten einheitliche Merkmale für Patienten mit Lymphom, aber keine AIC-spezifische Signatur. Die Gruppierung von T-Zellen mit überlappender Antigenspezifität unter Verwendung des GLIPH-Algorithmus (Gruppierung der Lymphozyten-Interaktion durch Paratope-Hotspots) zeigte jedoch einen erheblichen Mangel an T-Zell-Clustern bei Patienten mit primärer Autoimmunerkrankung (AIC sowie AIH) im Vergleich zu gesunden Spendern. Die Funktionalität dieser Cluster muss in zukünftigen Studien geklärt werden. Dennoch könnte die Signatur des T-Zell-Cluster-Verlusts einen leicht zugänglichen Biomarker für Autoimmunerkrankungen darstellen.

Zusammengenommen hat AIRR seq unser Verständnis der adaptiven Immunantwort vertieft und bisher dazu beigetragen, drängende Fragen in den Bereichen Immundiagnostik, Impfstoffe, Krebsimmuntherapie und Antikörperentwicklung zu beantworten. Technischer Fortschritt sowie optimierte Berechnungsprozesse werden dazu beitragen, dass diese Technologie ihr volles Potenzial entfaltet.

1 INTRODUCTION

The immune system is a very complex but well-orchestrated interdependent network of different tissues, cells and soluble factors set out to protect the body from internal or external threats. Increasing age and diseases, like cancer, can perturb fine-tuned immune responses thereby compromising immunological key functions. On the one hand, this can result in an overly aggressive immune response towards harmless (self-) antigens leading to uncontrolled chronic inflammation and autoimmunity or, on the other hand, in lack of response towards harmful antigens. Roughly, two arms of the immune response are distinguished, the innate and adaptive immunity.

In contrast to the innate immune response which confers immediate protection against a wide variety of pathogens, the adaptive immune responses are the ones that lead to lasting immunity and confer protection from disease if reinfection with the same pathogen occurs. The two major cell types of the adaptive immune system are B and T cells, both characterized by the expression of highly specialized cell surface receptors mediating antigen recognition. While B cells eventually mature to antibody-producing plasma cells, T cells play major roles as initiators, regulators and effectors of adaptive immunity and also modulate innate immune responses. The unique characterizing feature of a single T cell is its T cell receptor (TCR) generated during development through complex and diverse genomic rearrangements. While the majority of TCRs are assembled as heterodimers consisting of an alpha (α) and beta (β) polypeptide chain (encoded by the TRA and TRB genes, respectively), a small subpopulation of functionally divergent T cells encompasses TCRs assembled by the alternative gamma (γ) and delta (δ) chains (encoded by TRG and TRD genes, respectively). TCRs require the presentation of their cognate antigen by major histocompatibility complex (MHC) I or MHC II molecules on the surface of either target cells

or so-called antigen-presenting cells (APCs) for efficient initiation of an immune response. The functional state of T cells is characterized by specialized membrane-bound co-receptor molecules, most importantly cluster of differentiation (CD) 4 and CD8, in combination with their secretion of signaling molecules mediating pro- and anti-inflammatory information, the cytokines. Based on these molecular profiles, T cells can be roughly classified as $CD8^+$ cytotoxic T cells, $CD4^+$ helper T cells, and regulatory T cells. The entirety of all T cell receptor specificities of an individual is termed TCR repertoire.

Since the work presented here focuses on the T cell arm of the adaptive immune response and aims to elucidate its functional role in increasing age, cancer and autoimmunity, the following chapters will shortly introduce fundamental aspects of T cell development and function.

1.1. T CELL DEVELOPMENT

Most constituents of the cellular immune system derive from bone marrow resident hematopoietic stem cells (HSCs) and develop within specialized lymphopoietic environments to fully committed lymphocytes through the successive ordered acquisition of lineage defining properties, especially rearrangements of antigen-receptor genes and expression of developmental stage specific signaling and surface molecules. T cells arise from common lymphoid progenitor cells (CLPs), descendants of the HSCs, that, mediated by chemotactic agents, migrate to the thymus (thymus homing) where they are instructed by Notch1-dependent signaling to become a T cell [1]. These thymic progenitor cells committed to the T lineage are termed thymocytes. Lymphoid progenitor cells enter the thymus through blood vessels in the corticomedullary junction (CMJ) which separates two anatomically and functionally discrete regions within the thymus: the central medulla and the peripheral cortex enclosed by the subcapsular zone. Thymocytes constantly migrate through these distinct microenvironments thereby interacting with stromal cells, thymic

epithelial cells (TECs) and other thymic resident hematopoietic cells, like dendritic cells (DCs) [2] (Figure 1). This tightly regulated process i.e. through spatially restricted cues, is essential for the generation of a functional yet self-tolerant T cell repertoire.

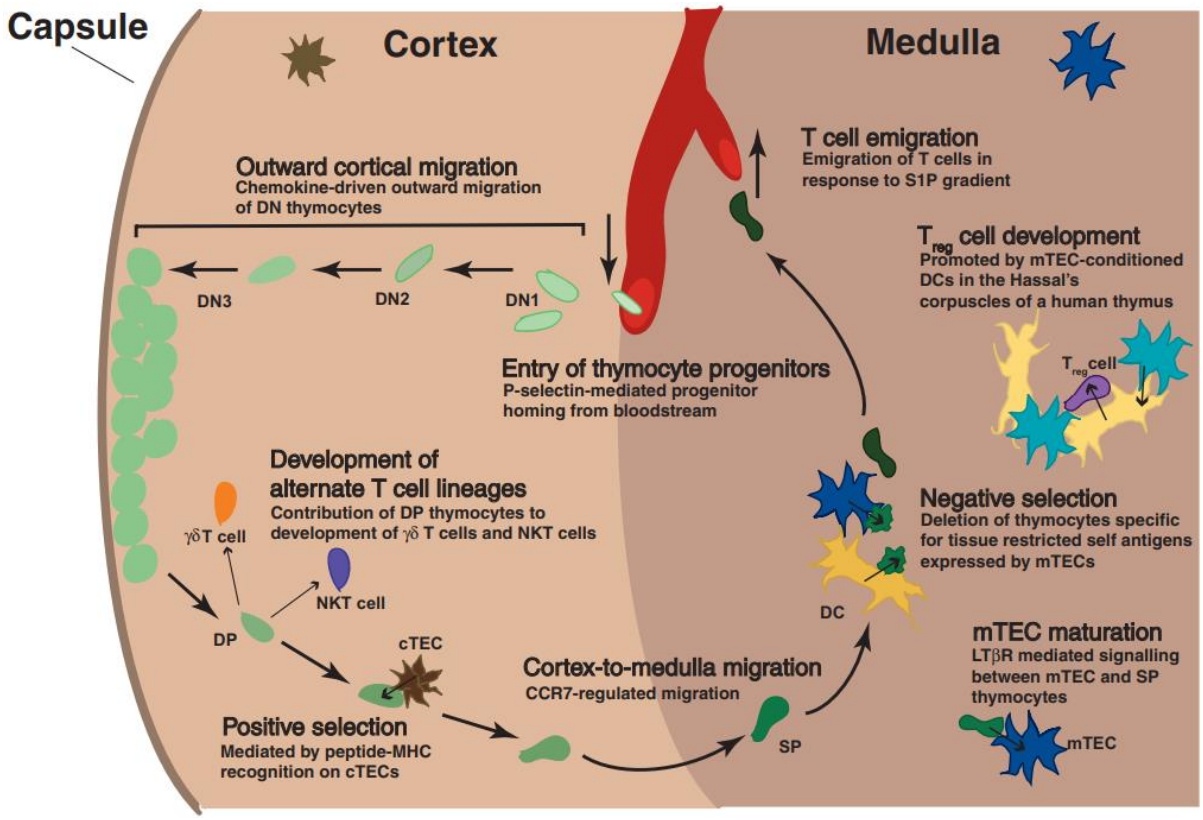


Figure 1. T cell development in the human thymus. *DN, double-negative; DP, double positive; cTEC, cortical thymus epithelial cell; mTEC, medullary thymus epithelial cell; NKT, natural killer T cell; SP, single positive.* Reprinted with permission from Springer Nature Ltd.: Nature Immunology [3], copyright 2006.

In the early phases of differentiation, thymocytes do not express the T cell characteristic co-receptors CD4 or CD8, a developmental stage described as double-negative (DN) [4]. Three major substages of DN cells are distinguished, termed DN1, DN2 and DN3 [5]. DN1 cells ($CD44^{+}CD25^{-}CD117^{hi}$) enter the thymus through the vasculature at the CMJ and migrate to the midcortex where they differentiate to DN2 cells ($CD44^{+}CD25^{+}CD117^{hi}$). Through the interaction with TECs, signaling molecules like Notch1, interleukin 7 receptor (IL-7R) and C-X-C chemokine receptor type 4 (CXCR4) are activated and promote survival, differentiation and T-lineage commitment of DN1 and DN2 cells [6, 7]. *RAG* gene expression is significantly upregulated in DN2 cells located in the midcortex [8] and the

first rearrangements of the TCRγ and δ encoding genes occur here (but not TCRβ yet) [9, 10]. Early DN3 cells ($CD44^{-}CD25^{+}CD117^{lo}$) develop by migrating further outward in the cortex toward the capsule. Here, the TRB locus is rearranged, representing the irreversible commitment to the T-cell lineage (β-selection) [11]. Moreover, DN3 cells are known for their pronounced proliferation potential most likely induced and supported by continuous Notch1 and IL-7 signaling [12, 13]. The development to late DN3 cells and subsequently pre-double positive (DP) cells ($CD4^{lo}CD8^{lo}CD44^{lo}CD25^{-}$) takes place in the subcapsular area of the thymus. The events occurring in this critical transition phase are the completion of TRB rearrangement and expression of a pre-TCR. Only cells with an in-frame TRB rearrangement receive survival signals and increase proliferation drastically [14-18]. Moreover, CD4 and CD8 are upregulated here [19] and TRA gene rearrangements are initiated [20]. Upon completion of this developmental stage the DP cells ($CD4^{hi}CD8^{hi}$) reverse their journey inwards the medulla [21]. One major change that occurs during this stage is the transient loss of proliferative capacity [16, 22] . Moreover, DP cells are the first T cell progenitors to no longer depend on Notch signaling which is solely mediated through ligands on the cell surfaces of stromal cells [23]. At this stage, positive selection and CD4/CD8 lineage divergence of thymocytes occurs through interaction with MHC-expressing non-hematopoietic stromal cells of the thymic cortex [24]. Different models of lineage choice are proposed: (1) random repression of CD4 or CD8 and only cells that still recognize self-MHC survive, (2) instructive repression of CD4 on cells with MHC I-restricted TCR or of CD8 on cells with MHC II-restricted TCR [25] or (3) as suggested by more recent data, DP cells terminate CD8 expression and upregulate IL-7R by default. If the TCR signal is not disrupted by CD8 loss, the CD4 lineage is produced. However, if CD8 loss interrupts the positive selection signal, IL-7R signaling mediated by stroma-derived IL-7 leads to the silencing of CD4 followed by re-induction of CD8 expression [26]. Thus, positive selection and lineage divergence is controlled by thymic stromal cells in a direct

(TCR/MHC) and indirect (IL-7/IL-7R) manner. The development into single-positive (SP) cells coincides with the migration of these cells to the outer medulla of the thymus. Here, the abundant dendritic cells (DCs) govern the selection of self-tolerant TCRs (negative selection) by presenting a wide array of self-peptides (thymic and non-thymic tissue specific) on MHC molecules. [24, 27-29]. Most self-antigens, however, are synthesized by medullary thymic epithelial cells (mTECs) under the control of transcription factor autoimmune regulator (Aire) and "supplied" to DCs via cross-presentation [30-33]. The interaction between DCs and SP T cells at this stage is recurrent in order to increase effectiveness of self-tolerance, a process termed central tolerance [34, 35]. Cells successfully undergoing negative selection receive survival signals through stromal IL-7 which itself leads to re-expression of B-cell lymphoma 2 (Bcl-2) [36-38]. Bcl-2 together with immune-associated nucleotide-binding protein (IAN) 4 and 5 promote the survival of SP T cells [39]. Moreover, at this stage a subset of cells is selected to become regulatory T cells (T_{reg}) which in addition to IL-7 require thymic stromal lymphopoietin (TSLP) for their survival and differentiation [40-43]. Post-selection SP T cells are functionally not mature and require further interactions with medullary stroma until final maturation. This interaction involves CD69 mediated signaling which extends the time in this thymic zone and allows for further maturation as well as differentiation of potentially self-reactive SP T cells to T_{regs} [41]. Proper differentiation of mTECs, on the other hand, also depends on signals from maturing thymocytes (e.g. through the lymphotoxin-β receptor (LTβR) pathway) [44]. This highlights the essential crosstalk between thymocytes and TECs for the proper organization and function of the thymus. Finally, functionally mature T cells downregulate CD69 and exit the thymus through the vasculature at the CMJ.

1.1.1 V(D)J RECOMBINATION AND T CELL RECEPTOR ASSEMBLY

The T cell receptor (TCR) is composed of two different polypeptide chains, both of which contribute to the antigen binding domain (Figure 2A right). This heterodimer is encoded by rearranged variable (V), diversity (D) joining (J) and constant (C) gene segments of either the TRA and TRB loci (αβ T cells) or TRG and TRD loci (γδ T cells), respectively. The so-called V(D)J rearrangement occurs in an ordered fashion whereby D to J recombination takes place before the V gene segment is joined to the recombined DJ segments [45] (Figure 2A left).

Each T lymphocyte expresses only one receptor specificity. Since the gene rearrangement is irreversible, all progeny of a particular lymphocyte inherits that receptor specificity thereby forming a clone of cells with identical antigen receptors (Figure 2B).

The TCR repertoire of a human is estimated to have 10^8 to 10^{10} different specificities at any given time [46, 47]. This almost infinite diversity is reached during the assembly of the antigen receptor, through (1) combinatorial diversity (random combination of different gene segments) (2) junctional diversity (random addition of non-templated (N) nucleotides and palindromic (P) sequences or deletion of nucleotides at the junctions of gene segments) and (3) as mentioned before by the combination of two different variable chains (αβ or γδ) which are encoded from two different sets of gene segments.

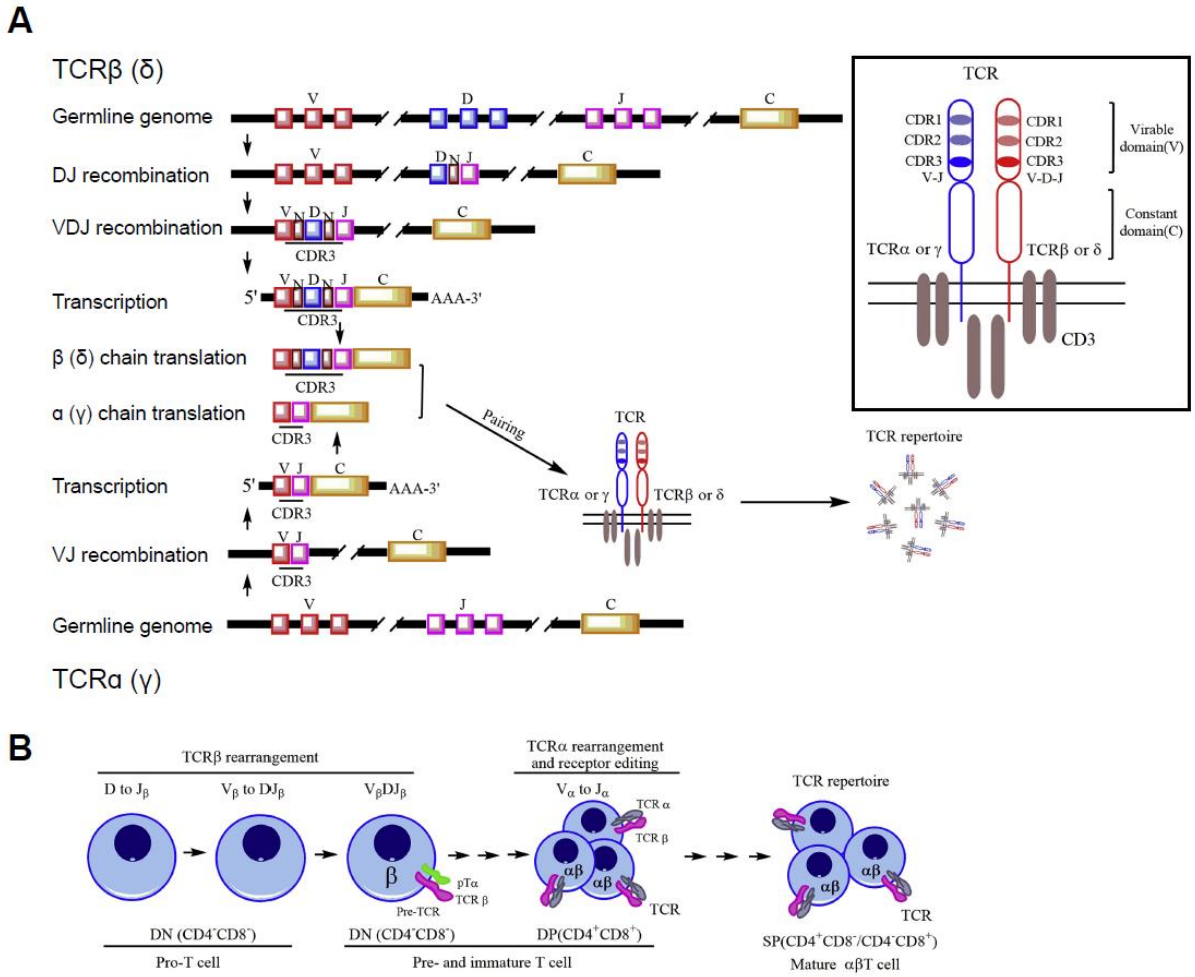


Figure 2. Generation of the TCR repertoire. A. (left) Sequential recombination of T cell receptor (TCR) V(D)J genes for β(δ) chain and α(γ) chain. (right) Structure of TCR. The TCR is a heterodimer of αβ chains (or γδ chains). It consists of a constant domain and a variable domain, which is encoded by the rearranged V(D)J genes and consists of three hypervariable complementarity determining-regions (CDRs). **B.** TCR assembly during T cell development in the thymus. Reprinted with permission from Elsevier B.V.: Cancer Letters (adapted from [48]), copyright 2018.

Downstream to each V, adjacent to each D and upstream of each J gene segment a recombination signal sequence (RSS) is found (Figure 3A). It consists of a conserved heptamer and nonamer sequence which are separated by a 12 or 23 nucleotide spacer (12RSS or 23RSS) [49]. The "12/23 rule" stipulates that efficient recombination can only occur between RSS with different spacer length. However, the "beyond 12/23 rule" defines additional restrictions, like direct Vβ-to-Jβ recombination, that are put in place. V(D)J

recombination requires RSSs and based on their orientation the reaction proceeds by inversion or deletion [50] (Figure 3B).

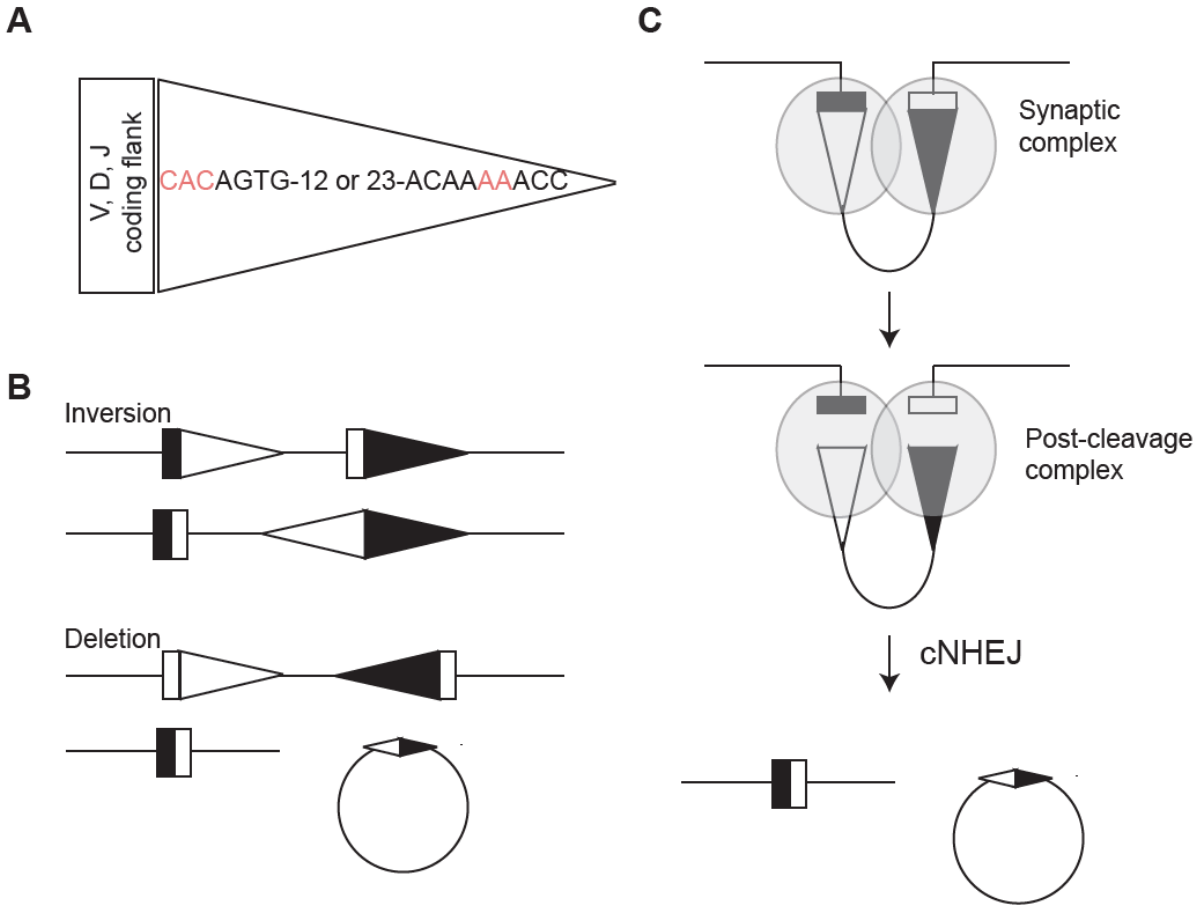


Figure 3. Schematic overview of V(D)J recombination mechanism. A. Recombination signal sequence (RSS) with conserved heptamer and nonamer. Red letters represent the most conserved positions. Sequences from [51]. **B.** V(D)J recombination reaction orientation and products. **C.** V(D)J recombination process with synaptic and post-cleavage complex resulting in the coding and signal joint. *cNHEJ, classical non-homologous end-joining.* Adapted from [45].

The V(D)J recombination process can be divided into two steps: the cleavage and the joining phase [49]. During the cleavage phase RAG proteins together with high mobility group box protein (HMGB) 1 mediate the combination of a 12RSS and a 23RSS thereby forming a synaptic complex [52, 53] (Figure 3C top). In this complex the RSSs are nicked, and the cleavage of both RSSs is catalyzed by RAG proteins to form a pair of blunt signal

ends and a pair of hairpin coding ends [54]. In a post-cleavage complex (Figure 3C middle), RAG proteins retain the four DNA ends and direct them to the classical non-homologous end joining (cNHEJ) pathway preventing access to the low fidelity, translocation-prone alternative NHEJ (aNHEJ) [55]. The transition to the joining phase happens within this complex, during which the DNA ends are re-organized, processed, and repaired. The hairpin structures are resolved by structure-specific endonucleases followed by joining of the signal and coding ends (Figure 3C bottom) through cNHEJ involving proteins like Ku70, Ku80, XRCC4, DNA Ligase IV and the Cernunnos/XLF protein [56, 57]. Typically, two types of joined products are generated: the coding joint formed by combining two coding segments together, which involve junctional nucleotide additions or deletions. This process is mainly mediated through TdT, polμ and polλ (members of the Pol X family of polymerases) [58]. The other type of joined product is the signal joint formed by the ligation of the two blunt signal ends.

Two main T cell classes are distinguished: αβ and γδ T cells. The choice of the lineage fate is believed to happen at the DN3/pre-DP stage where thymocytes rearrange almost simultaneously three out of four TCR loci: TRB, TRG and TRD [59]. Proliferation arrest occurs until a functional TCR is expressed. If TRB is rearranged in frame, the cells express the TCRβ combined with a germline-encoded pre-TCRα (pTα) chain leading to a burst in proliferation (β-selection), silencing of TRG and initiation of TRA rearrangement. The rearrangement of TRA results in the excision of the TRD locus, which is located within the TRA locus. The upregulation of CD4 and CD8 in these cells determines their migration into the center of thymus as DP cells to finish maturation.

Productive rearrangement of TRG and TRD, on the other hand, leads to the expression of a γδTCR and subsequent increase in proliferation (γδ-selection). The majority of these cells

avoid progression through the DP phase and egress to the periphery with a DN phenotype (only rarely as $CD4^+$ or $CD8^+$ SP cell) comprising 0.5 – 5% of all T lymphocytes [60].

More current models suggest that lineage fate is controlled by cell intrinsic signaling and TCR signal strength, rather than TCR type *per se*, with stronger TCR signaling leading to the development of γδ and weaker signaling favoring the αβ linage [61]. Moreover, it was shown that γδ T cells can be redirected to the αβ linage by impairing their signaling capacity [62]. DN T cells which prematurely express an αβTCR have been described and exhibit high similarity to γδ T cells (phenotype and rapid effector responses) [59]. It is hypothesized that these DN T cells (with γδ or αβ TCR) leave the thymus through vasculature in the subcapsular zone before their re-migration towards the medulla is induced [63]. Therefore, lineage definition currently is based on the progression through the DP phase (mainly αβ) or lack thereof (mainly γδ).

In the following sections, T cell(s) or TCR will, for simplicity, solely refer to αβ T cells, unless otherwise specified, since the work presented here focuses on this subtype/population.

1.1.2 T CELL ACTIVATION

Functional T cells that have undergone maturation in the thymus are found in the periphery and especially in lymphoid organs and tissues. Upon encounter of their cognate antigen, these naïve T cells proliferate and differentiate into effector T cells mediating the adaptive immune response. However, productive activation of naïve T cells requires the convergence of signals arising from TCR engagement as well as co-stimulatory signals which ultimately induce proliferation and differentiation of the cell.

In addition to the variable αβ chains, the TCR complex consists of CD3 protein dimers including γε, δε and ζζ, responsible for intracellular signal transmission [64].

APCs present antigens to T cells via MHC class I or II surface molecules whereby either $CD8^+$ or $CD4^+$ T cells are triggered, respectively. MHCs are highly polymorphic

transmembrane glycoproteins that are encoded in the cluster of genes known as MHC. The human MHCs are also called HLA (human leukocyte antigen). The extracellular part of the protein has a cleft in which peptides are bound and thereby presented to the TCR. In order to bind to the peptide-MHC (pMHC) complex, a TCR has to be specific for a unique combination of peptide and MHC molecule, a phenomenon which is called MHC restriction. In other words, a specific peptide presented on MHC class II molecule can only be bound with high avidity by a $CD4^+$ T cell with a TCR with suitable αβ chains. Subsequently, the TCR signal is transmitted via cytosolic domains of the CD3 proteins, referred to as immunoreceptor tyrosine-based activation motifs (ITAMs). ITAMs consist of two Tyr-X-X-Leu/Ile motifs separated by six to eight amino acids [65]. Antigen recognition by the αβ chains leads to the phosphorylation of ITAMs at the tyrosine residues by Src-family protein tyrosine kinases (PTKs) and in turn to the activation of downstream signaling events including Erk/MAP kinase cascade, protein kinase C theta (PKCθ), and phosphoinositide-dependent kinase (PDK) 1 mediated pathways [66, 67].

As mentioned above, in addition to the TCR-pMHC signal, a co-stimulatory signal via the T cell surface molecule CD28 is crucial for productive T cell activation. CD28 interacts with its ligands CD80/CD86 (B7-1/B7-2) on the surface of APCs and the absence of this co-stimulatory signal results in T cell anergy, a state of unresponsiveness often seen in chronic inflammation and cancer.

1.2 NEXT-GENERATION SEQUENCING

Next-generation sequencing (NGS) relies on the parallel sequencing of billions of DNA bases. As opposed to Sanger sequencing [68] which is considered the “first-generation” with a low throughput of approximately one million bases per day, sequencers of the next generation are able to produce up to four trillion bases in one day (e.g. Illumina NovaSeq). This makes it now possible to sequence the whole human genome in approx. 2 days - an endeavor that took over 10 years to be accomplished during the Human Genome Project (HGP) on numerous Sanger sequencing machines [69, 70]. The introduction of massive parallel sequencing led to a drastic drop in sequencing costs. While the sequencing costs of one human genome accounted for approx. 100 Mio US Dollar at the end of the HGP in 2001, the costs have dropped to 1000 US Dollar per human genome today [71]. Widely used NGS technologies nowadays are Illumina sequencing, Ion torrent and Pacific biosciences technology [72]. These technologies differ in the strategies that are used to prepare the sequence libraries, to detect the signal and ultimately to read the DNA sequence. While on Illumina and Ion torrent machines a clonal amplification of the initial template is required to increase the signal-to-noise ratio, the SMRT systems (Pacific biosciences) are able to detect single molecule template extensions and therefore do not need any pre-amplification.

Since the work presented here is based on Illumina sequencing, the following paragraphs focus on this technology.

During library preparation, specific adapters are added to each DNA fragment of interest, for example through enzymatic ligation or PCR amplification of the region of interest with primer-overhang. Moreover, the tagging of samples with 6-8 nucleotides long molecular barcodes allows multiplexing of numerous samples on one sequencing run. The barcoded and adapter-tagged library is de-natured using NaOH (Sodium hydroxide) and loaded onto the flow cell, which is covered with a lawn of oligos that are complementary to the adapters

on the DNA fragments. The single-stranded template DNA hybridizes to the oligos on the flow cell. An initial copy from the template molecule is synthesized, using the flow-cell oligo as a primer. Following the wash-off step of the initial library molecules, the immobilized copy of the template is amplified by bridge amplification to generate a cluster of identical DNA sequences. This step yields approx. 1000 copies of the initial molecule. Before sequencing starts, one of the oligos is cleaved off from the flow cell leaving identical stranded sequences in one cluster. During four-color sequencing the fluorescently labeled dNTPs are incorporated one at a time and the emitted light serves to identify the corresponding base. Moreover, the fluorescent group serves as reversible terminator to prevent multiple extension events.

Once the imaging step is completed the fluorescent group is cleaved off, the 3´hydroxyl group is regenerated, and the template strands are ready for the next dNTP incorporation. This repetitive process of base addition, imaging and removal of the fluorescent group is termed sequencing by synthesis. Successful sequencing depends on library titration. Defining the optimal number of molecules for cluster formation is essential to maximize data output per sequencing run while at the same time avoiding over-clustering (which means the camera cannot distinguish between different clusters). Another potentially confounding factor is the length of molecules to be sequenced, because short molecules yield clusters that are denser with smaller areas, hence generating a more intense signal. Therefore, libraries with a wide range of fragment size will lead to clusters widely varying in area size and signal strength thus impairing the number of reads passing filter [72]. Another phenomenon known as dephasing has an impact on sequencing quality. The longest reads that can be generated are 300 bases long (on the Illumina MiSeq platform). This limitation is owed to errors in chemistry that add up over the increasing number of cycles. Thereby, a small portion of molecules within a cluster lags behind during extension (phasing) or advance a base (pre-phasing) which leads to an accumulation of errors and

decrease the signal-to-noise ratio thereby decreasing the quality towards the ends of reads [73].

1.2.1 T CELL RECEPTOR BETA (TRB) SEQUENCING

The assembled V(D)J sequences of TRA and TRB encode for the variable, antigen-recognizing domain of the α:β heterodimer, respectively. These regions present a very high variability which originates from the process of V(D)J recombination. The variability within this region is lower in the four so-called framework regions (FR) that surround the more hypervariable complementarity determining regions (CDR). Each chain of the TCR consists of three CDRs and the highest sequence variability is found in each CDR3, respectively. The CDR3 spans the junction of V, D and J gene segments which due to the non-templated (N/P) insertion and deletion of nucleotides during the recombination process can produce an almost unlimited diversity of amino acid sequences. The likelihood that an identical CDR3 is rearranged by random V(D)J recombination in two different T cell clones is virtually zero. If rearrangement of the TRB locus results in a non-productive gene (through frameshift or premature stop), a second recombination is attempted at the other chromosome. In contrast, the TRA loci are recombined simultaneously on both chromosomes and several recombination attempts are undertaken using more distal V and J gene segments to produce a productive gene [20]. Potentially, this would give rise to four different TCRs by combining two different α and β chains, respectively. Through allelic exclusion it is ensured that only one receptor combination is expressed on the surface of each cell. However, this process is leaky: approximately 7% of T cells have two productively rearranged beta chains [74, 75] and 1% express both of them on the surface [76-78]. For the alpha chain this number is even higher with 7 to 30% of T cells expressing two different functional alpha chains [79, 80].

Since a given CDR3 sequence is unique (not present in two clones) as well as the majority of clones express only one functional TCRβ CDR3, the sequence of this region can be used as a barcode to distinguish one T cell clone from another.

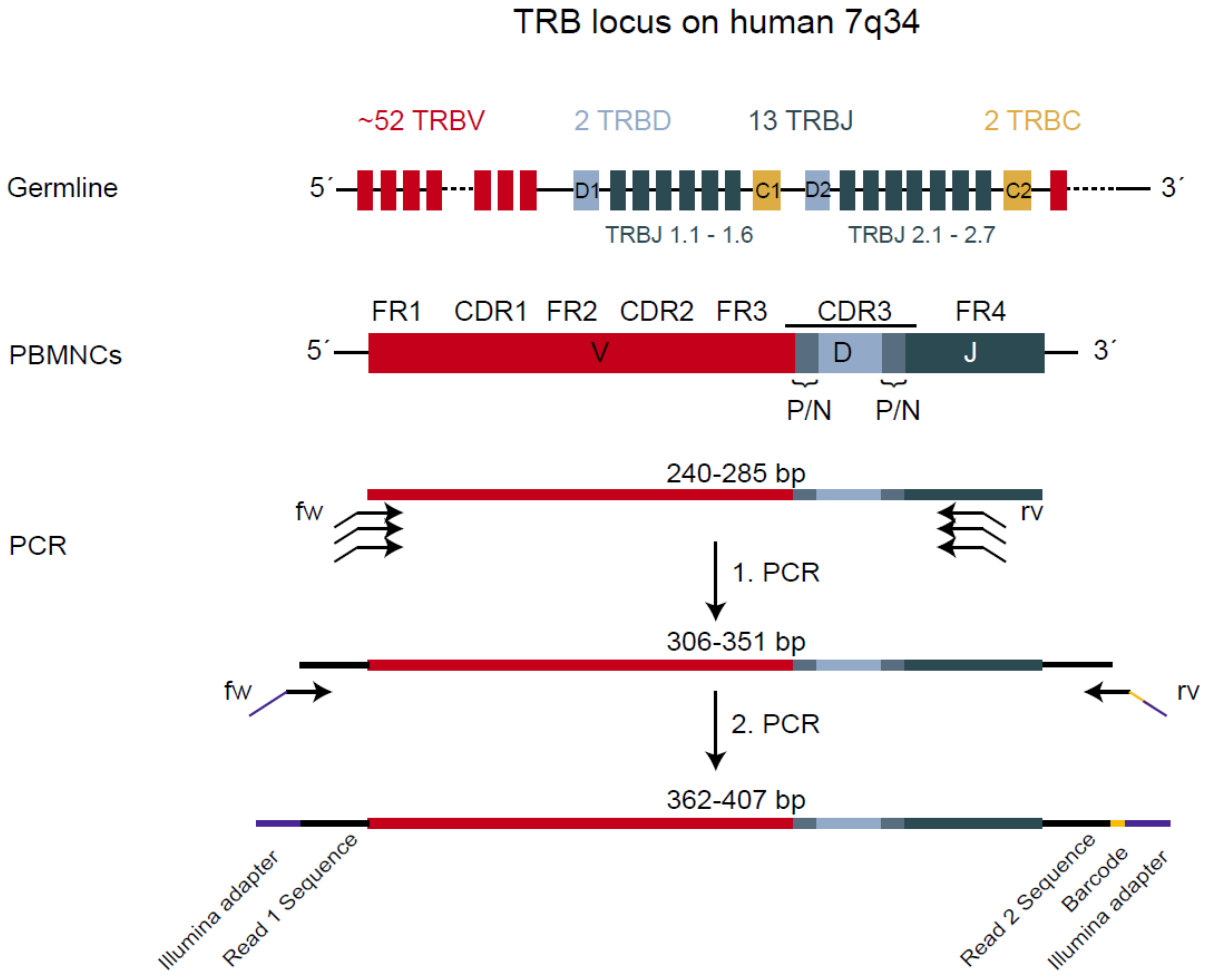


Figure 4. Scheme of V(D)J amplification from peripheral blood mononuclear cell (PBMNC) DNA. Distribution of germline TRB V, D, J and C (constant region) genes on the long arm of chromosome 7 according to IMGT (http://www.imgt.org). *CDR, complementarity-determining region; FR, frame-work region; fw, forward; N, non-template encoded nucleotides; P, palindromic nucleotides (not in germline); rv, reverse.*

Different approaches are used to perform bulk TRB sequencing, also termed antigen immune receptor repertoire sequencing (AIRR seq), from genomic DNA (gDNA) or total RNA including multiplex-PCR, target enrichment methods, 5´-RACE (rapid amplification of cDNA ends), template-switch and nested PCR [81]. The available starting material has an

influence on which method can be used for AIRR seq. Genomic DNA is more stable and makes the analysis of single clones more feasible since only one template, as opposed to multiple transcripts, is present in one cell. However, using RNA the entire V and J region can be captured and information on expression levels can be extracted. Moreover, the amount of starting material as well as sequencing depth play a crucial role and if different samples are to be compared, both parameters should be comparable [82].

In this work, sequencing was performed using templates generated with multiplex-PCR and gDNA extracted from peripheral blood cells (Figure 4). The applied primers were developed and standardized within a European BIOMED-2 study [83]. Our group further optimized the primers to increase binding and specificity [84, 85] (Primer sequences are listed in the Attachment).

1.2.2 ANALYSIS OF AIRR SEQ DATA

NGS sequencing data is stored in a file format called FASTQ. Besides the text-based information on nucleotide sequence (as stored in FASTA) this file format provides information on the corresponding quality score for each letter using ASCII (American Standard Code for Information Interchange) character. During de-multiplexing the sequencing machine assigns sample names to the raw data according to a pre-designed sample sheet which links each barcode to the corresponding sample. The compiled FASTQ files represent the data source for several bioinformatical tools and pipelines specifically designed for AIRR analysis, i.e. the MiXCR pipeline [86]. It consists of three main processing steps: (1) sequencing reads are aligned against reference V, D and J genes available from the international ImMunoGeneTics information system (IMGT) database (http://www.imgt.org) or NCBI Gene (https://www.ncbi.nlm.nih.gov/gene), (2) clonotypes are assembled based on identical CDR3 nucleotide sequence and (3) a table with

clonotypes, frequency distribution, gene usage and CDR3 length is exported to a human-readable text-file, which is used for the majority of subsequent analyses.

Repertoire metrics include the analysis of richness, diversity and clonality/evenness (Figure 5) of the AIRR and provide information on its clonal architecture. The Shannon diversity index [87] is a commonly used measure of diversity and can be calculated as follows:

(1) $$H = \sum_{i=1}^{S} p_i log_2 p_i$$

where *i* indexes the distinct clone, S is the number of species/clones (= richness) and p is the relative abundance of the i^{th} clone within the repertoire. p is calculated from n/N, where n = read count of the i^{th} clone and N = sum of all reads in the repertoire.

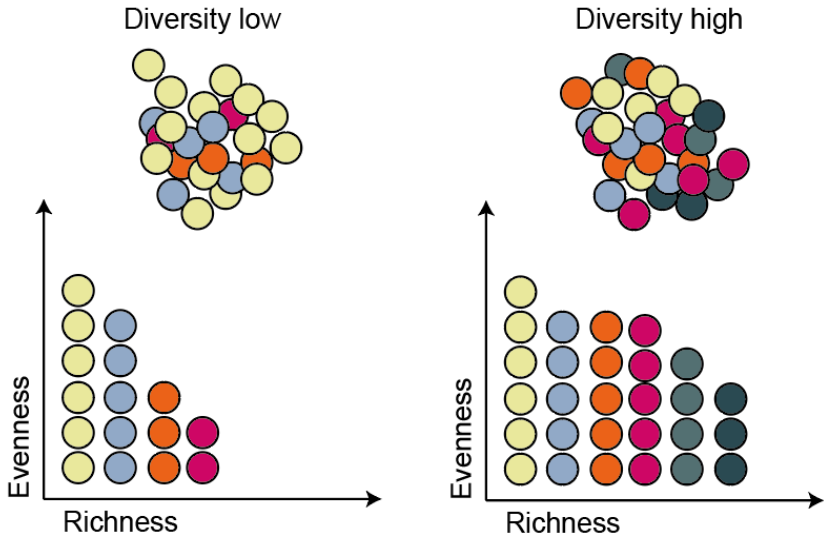


Figure 5. Schematic representation of repertoire diversity and its components richness and evenness.

The clonality index represents the reciprocal value of evenness (clonality = 1 - evenness), where evenness is calculated from H/H_{max} with H_{max} being the maximal possible value of H, if every clone in the repertoire was present at the same frequency. Indices of clonality or evenness provide information on the distribution of clonal frequencies within the

repertoire. Single clones, which are selected and proliferate during immune responses, will shift the clonality index of the repertoire towards 1. In contrast, the clonality index would be equal to zero at maximal diversity where each clone within the repertoire is present at the same frequency. Repertoire metrics are error-prone because the richness of a repertoire is directly influenced by fluctuations of read depth and amount of input material. For multi-repertoire comparisons, this problem can be minimized by sequencing the same amounts of DNA (= same amount of target genomes) with identical read depths and by normalizing read counts before metrics calculation. Notably, the evenness measure represents a rather stable exception since it also integrates the absolute number of clones present within a given repertoire.

1.2.2.1 GROUPING OF LYMPHOCYTE INTERACTION BY PARATOPE HOTSPOTS (GLIPH)

The GLIPH algorithm compiled by the research group of Mark Davis clusters TCR CDR3 sequences that potentially share antigen specificity based on conserved amino acid motifs as well as global similarity [88, 89].

As described before, the CDR3 is the most hypervariable region and thereby mainly accounts for the specificity of the TCR. Based on solved TCR-pMHC crystal structures it could be demonstrated that the probability to be within close proximity (threshold: 5 Å (= 500pm)) from the peptide antigen is highest for the CDR3β and especially for the positions 107-116 (IMGT nomenclature) as opposed to positions 104-106, 117 and 118 [90].

Basically, the GLIPH algorithm is set out to identify T cell clones that are overall similar by global similarity (differ only by 2-4 amino acids) and/or share conserved motifs within the high-contact probability CDR3β region (local similarity).

To do so, the algorithm calculates the probability of all motifs found within a dataset and compares these against a reference dataset of naïve and unselected $CD4^+$ and $CD8^+$ TCR sequences. The threshold for motifs to be defined as antigen specific was determined with data from tetramer experiments. Thereby, it was found that shared amino acid motifs were enriched in tetramer sorted T cell clones by >10-fold with a probability of <0.001 to find these motifs in the reference dataset. The GLIPH algorithm is therefore a powerful tool to identify TCRs that potentially share antigen specificity, especially towards foreign antigens, from bulk AIRR seq.

1.2.2.2 ANALYSIS OF THE GENERATION PROBABILITY OF TCRβ

Each TCR is generated stochastically through the recombination of V, D and J gene segments with a theoretical diversity of 10^{61} different sequences in humans [91]. It is estimated that only ~3% of T cells pass thymic selection [92]. Despite the apparent low probability to find the same TCR sequence in different individuals, some clonotypes are commonly shared between individuals, thus termed public clonotypes or public T cell response [93-96]. The presence of these public TCR seems to depend on a process called convergent recombination. These TCRs are produced more often because their amino acid sequence can be encoded by several different nucleotide sequences and these nucleotide sequences can be generated through the recombination of several different gene segments [96-100]. In other words, public or abundant TCRs seem to have an above average generation probability and therefore exist in the unselected repertoire. The likelihood for the generation of a TCR can be calculated using algorithms like IGoR (Inference and Generation of Repertoires, [101]) and OLGA (Optimized Likelihood estimate of immunoGlobulin Amino-acid sequences, [102]). Of note, it is assumed that 10-20% of the total human T cell receptor repertoire is public [100].

1.3 T CELLS AND AGE

The term senescence literally means “to grow old”. The state of senescence describes a stress-induced and durable cell-cycle arrest. Increasing numbers of these cells are observed in the aging human potentially providing an evolutionary protection against cancer [103]. However, the accumulation of somatic senescent cells (SSC) with increasing age contributes to impaired tissue function as well as the induction of senescence-associated secretory phenotype (SASP). SASP is characterized by pro-inflammatory secretion patterns, which help to maintain a systemic inflammatory milieu [104, 105].

The aging immune system is mainly characterized by two states, inflammation (inflamm-aging) and immunosenescence (insufficiency). The term inflamm-aging is coined during the turn of the 20th century and refers to a concept which associates aging with a sterile and chronic low-grade inflammation. Immunosenescence is thought to be of detrimental nature, meaning it involves only one-directional hypo-responsiveness of the immune system. However, through the ineffective clearance of accumulated SSC, immunosenescence also contributes to chronic inflammation [106], thereby maintaining an inflammatory milieu created by SASP.

Hence, inflamm-aging and immunosenescence, seem to present two sides of the same coin rather than opposing phenotypes. In the following, several factors that help to induce and maintain these states of immunity in the aging individual are highlighted.

Ultimately, all lymphocytes are derived from HSCs, therefore one explanation for the decreased immune responsiveness with increasing age can be the aging of the HSC. The diminished potential to self-renew as well as an imbalance favoring the generation of myeloid over lymphoid precursors are hallmarks of the HSC aging [107]. These events are shown to be genetically controlled by several quantitative trait loci [108-112]. The genes affecting HSC aging are for example responsible for encoding proteins involved in response to environmental factors or DNA repair. Changes in their expression or the activity of their

products seem to impair the quality of HSC [109]. Moreover, the decreased potential of lymphoid progenitors to survive and proliferate is thought to contribute to the diminished immunity with age. One explanation might be that the production of growth-hormone decreases in elderly individuals leading to a displacement of active bone marrow by adipose tissue, which is not able to support hematopoiesis [113, 114].

The age-related impairment of HSC and its progeny is accompanied by an involution of the primary lymphoid organ for T cell maturation, the thymus. The replacement of parenchymal tissue by adipose tissue starts under the influence of sex hormones after puberty in humans and progresses with increasing age. Although some functional thymic tissue seems to remain in individuals over 60 years of age [115, 116], the establishment of central and peripheral tolerance is rather ineffective.

The formation of central tolerance in the thymus is of great importance for the maintenance of the integrity of the organism. It is constituted of two processes, namely negative selection or deletion of self-reactive T cell clones and generation of T_{reg} which are responsible for keeping immune responses in the periphery towards self and non-self in check. The pool of circulating T_{reg} cells comprises T_{regs} induced in the periphery (pT_{reg}), however, 80-95% are supposed to derive directly from the thymus (tT_{reg}) [117-119]. Conflicting evidence exists about how tT_{reg} generation changes in the atrophied thymus with increasing age.

A study in *Rag*-GFP transgenic mice showed that, after being activated in the periphery, tT_{reg} cells home back to the thymus where they inhibit the IL-2 dependent generation of new tT_{reg} cells via IL-2 adsorption resulting in decreased tT_{reg} cell generation with increasing age [120]. The accumulation of peripherally activated T_{reg} cells in the thymus was also confirmed in humans. The activation of T_{reg} cells in the periphery through pMHC leads to their proliferation and thus enrichment within the peripheral T_{reg} repertoire. The homing mechanism of activated T_{regs} is thought to tag them as functionally relevant while

simultaneously preventing their dilution in the periphery with newly generated T_{regs} of yet unproven specificity. These findings point towards decreased numbers of T_{regs} in elderly as explanation for the inefficient clearance of the chronic inflammatory state.

Opposing to this, a study by Oh and colleagues also based on transgenic mouse models, found that tT_{reg} cell generation was relatively enhanced in the atrophied thymus mediated through a shift in TCR signal strength [121]. The authors interpreted this finding as an attempt of the atrophied thymus to balance out for the defective negative selection. During negative selection a T cell receives only proliferative signals if it exhibits a weak or intermediate TCR signal after encountering an MHC II presented self-peptide (self-pMHCII), thereby becoming a conventional T cell (T_{con}) or T_{reg} cell, respectively. The TCR signal strength is determined by the avidity of the TCR for the whole self-pMHCII complex as well as the total number of interactions that occur simultaneously. Age-dependent decreased expression of Aire and MHCII lead to a decreased number of self-pMHCII complexes on the surface of mTEC [121, 122]. T cells undergoing maturation and interacting with these mTECs during negative selection are not able to exhibit a “strong” TCR signal, since the number of interactions *per se* is smaller. This in turn is suggested to result in an overall shift of TCR signal to either intermediate or weak TCR signals only. Thereby, strongly reactive T cell clones are diverted to the tT_{reg} lineage and tT_{reg} cell-biased clones are diverted to the T_{con} lineage, ultimately leading to an increased number of self-reactive T cells that are released into circulation [106, 121]. The divergence from T_{reg} to T_{con} phenotype was also confirmed by a study comparing $Aire^{-/-}$ and $Aire^{+/+}$ mice [123]. Prostate antigen specific T_{con} cells in $Aire^{-/-}$ mice preferentially expressed TCRs found in prostate antigen specific T_{reg} cells in $Aire^{+/+}$ mice, indicating that Aire maintains immune tolerance by directing autoreactive clonotypes in the thymus to the T_{reg} cell lineage.

Overall, the shift in TCR signal strength allowing for more self-reactive T_{con} to leave the thymus may potentially explain the increased rates of autoimmunity in the elderly.

The discrepancies between the studies regarding the numbers of T_{regs} seem to partially arise from methodological differences. The first group compared newly generated tT_{regs} with peripherally activated tT_{regs} and newly generated T_{con} with peripherally activated T_{con} in young and old mice, respectively, not allowing for the interpretation of changes in the ratio between T_{reg} and T_{con}. However, the finding that peripherally activated T_{regs} inhibit *de novo* tT_{reg} generation might represent a potential negative feedback loop conveying information to the thymus about immunoregulation in the periphery and thereby adapting T cell differentiation to peripheral immune responses.

Besides the dysfunctional establishment of central and peripheral tolerance with increasing age the atrophied thymus is also characterized by a decreased output of naïve T cells [124-127]. Together with an amplified peripheral expansion of memory T cells this leads to decreased TCR diversity [128]. Diversity is influenced by the number of different T cell clones available (richness) and their relative abundance within the repertoire (as expressed by evenness or clonality indices). A repertoire with many unique T cell clones which are evenly abundant would exhibit a high diversity and vice versa.

Taken together, the state of immunosenescence in the T cell compartment comprises of decreased thymic output of naïve T cells and increased accumulation of memory T cells in the periphery resulting in a less diverse T cell repertoire. Additionally, the inverted CD4/CD8 ratio as well as altered immune response phenotypes in activated T cells (in favor of T helper cell 2 response) lead to dampened immune responses to new infection or vaccination and diminished anti-tumor immunosurveillance in the elderly [129]. Moreover, a chronic inflammatory phenotype is maintained by increased output of self-reactive T cells from the thymus and less effective clearing of SSC.

1.4 T CELLS AND CANCER

Altered or misled immune responses have significant impact on the initiation and progression of cancer [130]. Age-related immunosenescence as well as chronic inflammatory conditions are associated with increased risk for cancer development [131, 132]. Diverse myeloid and lymphoid cells are found within tumors exhibiting tumor-inhibiting but also tumor-promoting properties. The main players executing tumor-specific immune responses are $CD8^+$ cytotoxic T cells (CTLs). Tumor antigens displayed on MHCI molecules on DCs in lymph nodes allow for the selection of antigen-specific T cells. Upon binding of the TCR with high affinity/avidity to the presented pMHCI as well as co-stimulation of CD28 via CD80/CD86 in the DC-derived cytokine milieu (IL-12, type I interferons, and IL-15), naïve $CD8^+$ T cells are activated and differentiate into CTLs and expand clonally [133]. During this differentiation, the cell undergoes transcriptional, epigenetic and metabolic reprogramming and acquires effector features like production of cytotoxic mediators, e.g. granzymes A and B or perforin. Moreover, the cell increases the expression of αEβ7, an adhesion receptor, allowing its migration into the tumor [134]. Under "normal" circumstances the majority of activated T cells die after antigen clearance and only a small subset persists to differentiate into memory (T_{mem}) cells in order to enable rapid CTL generation upon antigen re-exposure [135]. In this state, the effector functions are downregulated and the cell acquires, similarly to stem cells, an ability to survive independent of antigen, undergoing slow homeostatic self-renewal driven by IL-7 and IL-15 [136]. During cancer as well as chronic infections, however, the effective differentiation to T_{mem} fails mainly due to persisting antigen stimulation resulting in exhausted T cells (T_{ex}) [137]. T_{ex} differ from other states of T cell dysfunction, like anergic or senescent T cells and are characterized by a sustained increased expression of inhibitory receptors (IR), like programmed cell death protein 1 (PD-1), cytotoxic T lymphocyte antigen 4 (CTLA-4), T cell immunoglobulin domain and mucin domain-3 (Tim-3), lymphocyte activation gene 3 (LAG-

3) or T cell tyrosine-based inhibitory motif (ITIM) domain (TIGIT) [138-144]. Moreover, altered transcriptional, epigenetic and metabolic profiles lead to a loss of effector functions. The severity of exhaustion is determined by the types and amount of IR expressed [145, 146].

Besides the influence of age and cancer itself, also treatment-induced impairments of the T cell repertoire need to be considered. Chemotherapy with its wide-ranging adverse effects on non-target tissues, including the immune system, remains the standard of care for many cancer entities. These often highly cytotoxic regimens induce lymphopenia resulting in increased susceptibility to infection as well as impaired anti-tumor immune control. Moreover, long-term effects on the immune system have to be considered which can affect future therapy options, like immune stimulating therapies, as well as vaccination success in these individuals.

It is well established, that immune cell populations recover at different rates after chemotherapy, whereby innate immune cells recover more rapidly as opposed to B or T cells [147, 148]. $CD8^+$ T cells seem to recover shortly after chemotherapy, whereas naïve $CD4^+$ T cells remain reduced at least up to 12 months resulting in a significantly altered CD4/CD8 ratio. The rapid recovery of the $CD8^+$ T cells pool remains controversial with data pointing towards a replenishment by expansion of terminally differentiated, senescent $CD28^-$ effector cells [149-151]. However, more recent data rather indicates increased thymic activity in lymphopenic environments [147]. In line with the latter, a faster repopulation of the $CD4^+$ naïve and tT_{reg} T cell subsets was observed in patients with lymphoma who presented thymic hyperplasia within 12 months after chemotherapy as compared to patients without thymic hyperplasia [152].

To achieve full functionality of the adaptive immunity both the size as well as the complexity/diversity need to be restored.

1.5 T CELLS AND AUTOIMMUNITY

Autoimmunity represents a state where the immune system is not able to distinguish between self and non-self and directs specific effector responses against non-harmful self-antigens, thereby injuring the host. Autoimmune diseases (AID), the manifestation of autoimmune processes, affect about 5% of the population and are a heterogeneous group of disorders with distinct symptoms and pathomechanisms involving specific organs or multiple organ systems within the host [153]. Autoimmune processes are also present in physiological circumstances but effectively quenched by immunoregulatory mechanisms, such as central and peripheral tolerance. These mechanisms are responsible for controlling and eliminating self-reactive lymphocytes and thereby preventing the progression into AID. There is consensus, that the permanent failure of one or several tolerance mechanisms is the major commonality of AID. The constellation of genetic predisposition (specific HLA- and non-HLA genes), environmental factors and/or derailed immunoregulatory processes facilitate the persistence and activation of self-reactive T and B cell clones ultimately leading to organ damage [154].

In the following, we will focus on central and peripheral tolerance mechanisms and elucidate how escaped, self-reactive T cells represent a prerequisite for AID development. From thymic recruitment to the expression of a functional αβ receptor, T lymphocytes mature independently from MHC proteins or antigens. However, all subsequent development decisions during positive and negative selection, also known as central tolerance, depend on the interaction with pMHC complexes encountered in the thymic stroma [155]. Overall, 90-95% of all T cell precursors die during positive selection (through death by neglect) and 50-70% of positively selected T cells are subject to negative selection; numbers that impressively reflect the stringent selection processes responsible for shaping the developing T cell repertoire [156-161]. The basis for negative selection is the response of immature T cells to stimulation by antigen. As described earlier, the

"peripheral self" is mirrored in the thymus by mTECs under the control of Aire, thereby antigens of virtually all parenchymal organs are presented via MHC molecules to the developing T cell. The TCR signal strength then determines whether the cell receives survival (weak to intermediate signal strength) or apoptotic signals (strong signal strength), a process also known as clonal deletion [162]. Moreover, cells exhibiting high TCR signal strength towards self-antigens can be directed to the T_{reg} lineage responsible for maintaining peripheral tolerance [163]. This is very important since central tolerance, although very efficient, is not perfect, thereby permitting some auto-reactive T cells (with low or high avidity) to reach the periphery in a functional state. Once these T cells encounter their respective autoantigen, for example myelin basic protein or glutamic acid decarboxylase, they can be activated, adopt effector functions and contribute to the initiation of AIDs, like in these cases, multiple sclerosis or autoimmune diabetes, respectively [30, 164]. Under physiological conditions, however, forkhead box P3 postive (FoxP3+) T_{reg} cells are able to suppress these self-directed immune responses in the periphery. T_{regs} can exert their functions through several mechanisms namely anergy, clonal deletion and immunoregulation in a (1) cell contact-dependent manner via CD39, CD73 and LAG-3, or granzyme and perforin-mediated killing of responder T cells, (2) by secretion of immunosuppressive cytokines such as IL-10, transforming growth factor beta (TGF-β), IL-35 and galectin-1, or (3) deprivation of cytokines like IL-2, which are necessary for the proliferation and survival of responder T cells [165-167].

In pathological circumstances, decreased numbers or diminished suppressor functions of T_{regs} are shown to be associated with autoimmune processes, eventually leading to AID [168-171].

1.5.1 AUTOIMMUNE CYTOPENIAS

1.5.1.1 PATHOGENESIS

The term autoimmune cytopenias (AIC) encompasses autoimmune hemolytic anemia (AIHA), immune thrombocytopenia (ITP), autoimmune neutropenia (AIN) as well as immune mediated maturation defects in the bone marrow, pure red cell aplasia (PRCA) and acquired amegakaryocytic thrombocytopenia (AATP). These disorders are characterized by the production of autoantibodies against epitopes on erythrocytes (AIHA), thrombocytes (ITP), neutrophils (AIN) or erythroblasts (PRCA) and megakaryocytes (AATP) [172-176]. Moreover, combination syndromes like Evan´s syndrome, which is an AIHA associated with ITP, are described [177]. The production of these autoantibodies can either be idiopathic (primary disease) or associated with an underlying malignancy (e.g. chronic lymphocytic leukemia, CLL), systemic autoimmune disease or may be drug-induced [178]. In the majority of cases, blood cells coated with autoantibody are recognized by phagocytes via membrane bound Fcγ-receptors and cleared in the reticuloendothelial system in spleen and liver [178, 179]. Several disruptions in the normal immune response are proposed as mechanisms underlying the production of autoantibodies in AIC. The reduced control of autoreactive B cell clones by T_{regs} (breakdown of peripheral tolerance) is an important factor leading to this pathogenesis [180, 181]. Moreover, autoreactive T cells, especially of $CD8^+$ phenotype, can also mediate destruction of platelets and impair the production of platelets in the bone marrow [182, 183].

1.5.1.2 DIAGNOSIS & TREATMENT

AIHAs are divided into warm and cold type, depending on the reactivity (at 37°C or 4°C, respectively) and the predominant class of the autoantibodies (IgG (warm) and IgM (cold)). A positive direct Coomb's or antiglobulin test (DAT) which detects antibody (usually IgG) and complement (usually C3d) on the surface of erythrocytes, is characteristic for AIHA and

therefore used for the diagnosis [184]. Treatment of warm AIHA with corticosteroids, as opposed to cold AIHA, leads to an improvement of anemia in 70-80% of patients, however, only a small fraction of up to 20% achieve complete response [178]. Besides splenectomy, which again is ineffective in cold AIHA, treatment with the monoclonal antibody rituximab (anti-CD20) has shown great efficiency in warm as well as cold AIHA with durable responses of up to 3 years [185].

Isolated thrombocytopenia with platelet counts $<30x10^9$/L and symptoms of bleeding define ITP. Moreover, the diagnosis of ITP relies on the exclusion of alternative disorders like HIV or Hepatitis C infection and myelodysplasia to name a few [186]. Autoantibodies in ITP are most commonly directed against glycoprotein (Gp)IIb–IIIa, GpIb–IX, GpIb or GpIIIa. Similarly to AIHA, first line treatment include the administration of corticosteroids, however, expecting only about 30% of patients to achieve lasting remission [178]. Second line treatments include splenectomy, intravenous immunoglobulin (IVIg) as well as rituximab, which has been shown to be efficient in over 50% of patients with refractory ITP [187].

The increased risk for infections in neutropenia is due to decreased numbers of neutrophils in the blood. An absolute neutrophil count of <1500/µL defines the condition of neutropenia. However, in order to detect antibodies directed against human neutrophil-specific antigens (HNA, located on Fcγ receptor IIIb) and diagnose AIN the combination of granulocyte agglutination test and the granulocyte immunofluorescence test (GIFT) is recommended [188]. Treatment initiation depends on the degree of neutropenia as well as if the patient is symptomatic. As opposed to AIHA and ITP, corticosteroids, splenectomy and rituximab are ineffective or even risky for AIN patients and therefore not recommended. Granulocyte colony-stimulating factor (G-CSF) represents the treatment of choice as well as administration of broad-spectrum antibiotics if the neutrophil count is below 500/µL [189].

2 PUBLICATIONS AND CONTRIBUTIONS

Publications arising from this thesis

T cell receptor next-generation sequencing reveals cancer-associated repertoire metrics and reconstitution after chemotherapy in patients with hematological and solid tumors.

Simnica D, Akyüz N, Schliffke S, Mohme M, Wenserski v. L, Mährle T, Fanchi F. L, Lamszus K, Binder M. *OncoImmunology*, July 2019.

doi.org/10.1080/2162402X.2019.1644110

High-throughput immunogenetics reveals a lack of physiological T cell clusters in patients with Autoimmune Cytopenias.

Simnica D, Schliffke S, Schultheiß C, Bonzanni N, Fanchi F. L, Akyüz N, Gösch B, Casar C, Thiele B, Schlüter J, Lohse W. A, Binder M. *Frontiers in Immunology*. August 2019.

doi.org/10.3389/fimmu.2019.01897

2.1 T Cell Receptor Next-Generation Sequencing Reveals Cancer-Associated Repertoire Metrics And Reconstitution After Chemotherapy In Patients With Hematological And Solid Tumors

Authors

Simnica D, Akyüz N, Schliffke S, Mohme M, Wenserski v. L, Mährle T, Fanchi F. L, Lamszus K, Binder M.

Published in

OncoImmunology, July 2019. doi.org/ 10.1080/2162402X.2019.1644110

ORIGINAL RESEARCH

T cell receptor next-generation sequencing reveals cancer-associated repertoire metrics and reconstitution after chemotherapy in patients with hematological and solid tumors

Donjete Simnica[a], Nuray Akyüz[b], Simon Schliffke[b], Malte Mohme[c], Lisa v.Wenserski[a], Thorben Mährle[b], Lorenzo F. Fanchi[d], Katrin Lamszus[c], and Mascha Binder[a]

[a]Department of Internal Medicine IV, Oncology/Hematology, Martin-Luther-University Halle-Wittenberg, Halle (Saale), Germany; [b]Department of Oncology and Hematology, BMT with section Pneumology, University Cancer Center Hamburg, University Medical Center Hamburg-Eppendorf (UKE), Hamburg, Germany; [c]Department of Neurosurgery, University Medical Center Hamburg-Eppendorf (UKE), Hamburg, Germany; [d]ENPICOM B.V., 's-Hertogenbosch, The Netherlands

ABSTRACT

The dynamics of immunoaging and the onset of immunoparesis in healthy individuals and cancer patients has been controversially discussed. Moreover, the role of chemotherapy on T cell regeneration needs further elucidation in light of novel immunotherapies that have become standard of care for many elderly cancer patients. We used next-generation immunosequencing to study T cell receptor (TCR) repertoire metrics on 346 blood samples from healthy individuals and cancer patients producing a dataset with around 8.8 million TCR reads. This analysis showed that decline of T cell diversity and increase in T cell clonality is a continuous process beginning in healthy individuals over 40 years of age. Untreated patients with both hematological and solid tumors showed blood TCR repertoires with significantly lower diversity and higher clonality as compared to healthy individuals across all decades. Loss in T cell diversity was essentially driven by a loss in richness in aging healthy individuals, while in cancer patients a loss in repertoire evenness was an additional contributing factor. Interestingly, chemotherapy did not impair the regeneration of blood TCR repertoire diversity to pre-treatment age-specific levels. Surprisingly, even patients over the age of 70 years receiving highly T cell toxic therapies reestablished their pre-treatment T cell diversity suggesting rebound thymic activity rather than recovery of T cell counts by peripheral expansion only. Taken together, these data suggest that human TCR repertoire metrics gradually deteriorate in the aging individual, but age-specific TCR metrics are restored after T cell depleting therapy even in elderly cancer patients.

ARTICLE HISTORY
Received 14 June 2019
Revised 8 July 2019
Accepted 8 July 2019

KEYWORDS
Immunoaging; immune reconstitution; NGS; TCR; diversity

Introduction

A decrease in immune function over time is thought to explain increasing rates of infection, autoimmunity and cancer with age .[1] T cells are traditionally connected with this concept, since the thymus that generates the naïve T cell repertoire from bone marrow derived hematopoietic stem cells undergoes a visible decline from child- to adulthood .[2] This thymic involution is associated with a disorganization of the thymic tissue architecture, a decrease of the thymic stroma and its replacement by adipose tissue.[3,4] Due to this visible, age-dependent decline, thymus-dependent T cell recovery has been assumed to be severely limited in adults. However, age-associated T cell immunoparesis is not only restricted to thymic involution since it also involves aging of hematopoietic stem cells, lymphoid progenitors and mature lymphocytes in secondary lymphoid organs.[5]

T cell reconstitution is a critical feature of the recovery of the adaptive immune response and has two main components: thymic output of new T cells and peripheral homeostatic expansion of preexisting T cells. To explore immune repertoire metrics over time, extrapolations from animal models are not useful since – despite many immunological principles being shared across species – the size of the T cell system and the lifespan of typical animal models are fundamentally different from the human system.[6] Studies on immunoaging and immune reconstitution in humans gave inconsistent results partly due to the use of methods that should be interpreted with caution, such as T cell receptor excision circle (TREC) measurements.[7–9] In 2014, a next-generation sequencing (NGS) study by Britanova et al. has analyzed an extensive set of 39 healthy individuals setting the benchmark for studies on age-dependent TCR repertoire immunodynamics. As opposed to previous studies using less precise technologies, this group found a linear decrease of T cell diversity over time with significant reduction already apparent at age 40.[10]

Unlike in the healthy population, it is still a matter of debate if cancer patients are able to restore a functional TCR repertoire after cytotoxic and notably often T cell toxic

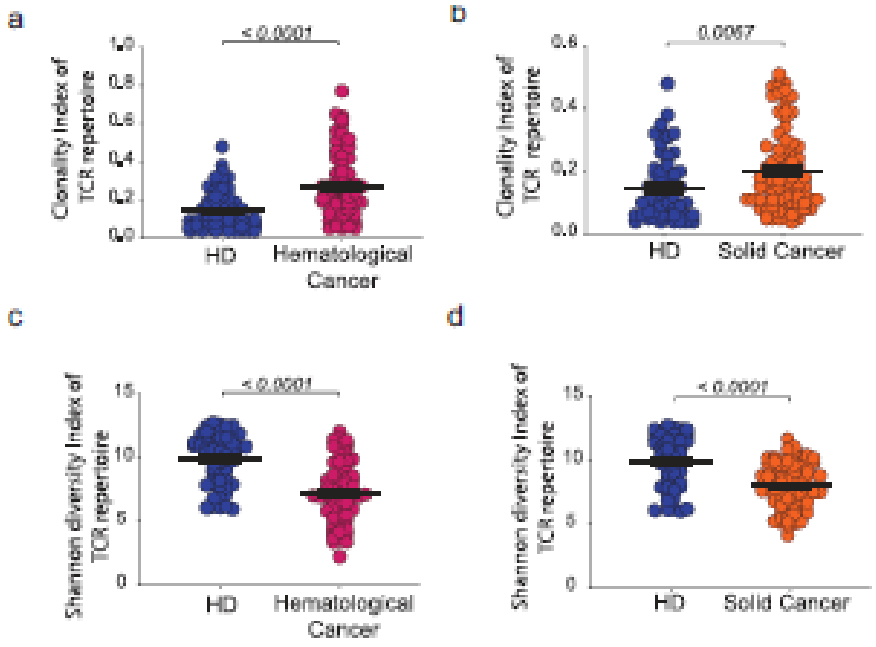


Figure 1. Comparison of T cell repertoire metrics between patients with cancer and healthy individuals.
Clonality index with mean ±SEM of PB TCR repertoire of patients with hemC ($n = 70$) compared to age-matched HDs ($n = 68$) (a) and patients with solC ($n = 89$) compared to age-matched HDs ($n = 58$) (b). Shannon diversity index with mean ±SEM of PB TCR repertoire of patients with hemC ($n = 70$) compared to age-matched HDs ($n = 68$) (c) and patients with solC ($n = 89$) compared to age-matched HDs ($n = 58$) (d). Statistical test: unpaired two-sided t-test.

therapy. In this comprehensive analysis of 346 TCR repertoires we found clear evidence for premature immunoaging in cancer patients independent of their treatment. To our surprise, however, we found that even elderly patients undergoing T cell toxic therapies largely reconstituted their age-specific TCR repertoire arguing strongly in favor of the hypothesis that thymic output may be reactivated driven by treatment induced lymphopenia.

Methods

Study approval

Informed consent was obtained from all patients and healthy donors (HD) for the use of their peripheral blood (PB) as approved by the ethics commission Hamburg (Ethikkommission der Ärztekammer Hamburg, Germany, project number PV4767). The study has been performed in accordance with the declaration of Helsinki of 1975.

Patients and samples

We included a cohort of 218 cancer patients, 94 of which with hematological cancers (hemC), 124 with solid cancers (solC) and 95 healthy donors (HD) as a reference cohort. In all investigated age groups (≤30, 31 to 40, 41 to 50, 51 to 60, 61 to 70 and ≥70) we included a minimum of 3 individuals. The analyses presented in Figures 5 and 6 include two different subcohorts: (i) matched patients according to age and tumor type who were untreated or received chemotherapy (labeled as „samples from different age-matched patients"). (ii) patients who had paired samples taken prior to and after chemotherapy (labeled as „paired samples from pre-post chemotherapy"). Patients with blood involvement by their disease were excluded from the analysis. Information on patient and sample characteristics are summarized in Supplementary Table 1.

Multiplex PCR of T cell receptor beta (TRB) repertoire for Illumina targeted next-generation sequencing (NGS)

The rearranged *TRB* receptor sequence containing the entire *V*, *D* and *J* gene segments was amplified using a mixed primer *TRB*-A/B pool and a touch-down PCR protocol from peripheral blood (PB) genomic DNA.[11] In two consecutive PCR reactions, amplicons were tagged with Illumina adapters and indices as previously described.[12–16] PCRs were performed using Phusion HS II (Thermo Fisher Scientific Inc., Germany). Amplicons were purified after agarose gel electrophoresis using the NucleoSpin® Gel and PCR Clean-up kit (Macherey-Nagel, Germany). Before being subjected to NGS, the concentration and quality of the amplicons/libraries was determined using Qubit (QIAGEN, Germany) and Agilent 2100 Bioanalyzer (Agilent technologies, Germany), respectively.

Illumina next-generation sequencing (NGS)

NGS and de-multiplexing was performed on an Illumina MiSeq sequencer (600-cycle, single indexed, paired-end run). Analysis of the *TRB* locus was computed using the MiXCR analysis tool V2.12 and V3.0.5.[15,17] Each unique complementarity-

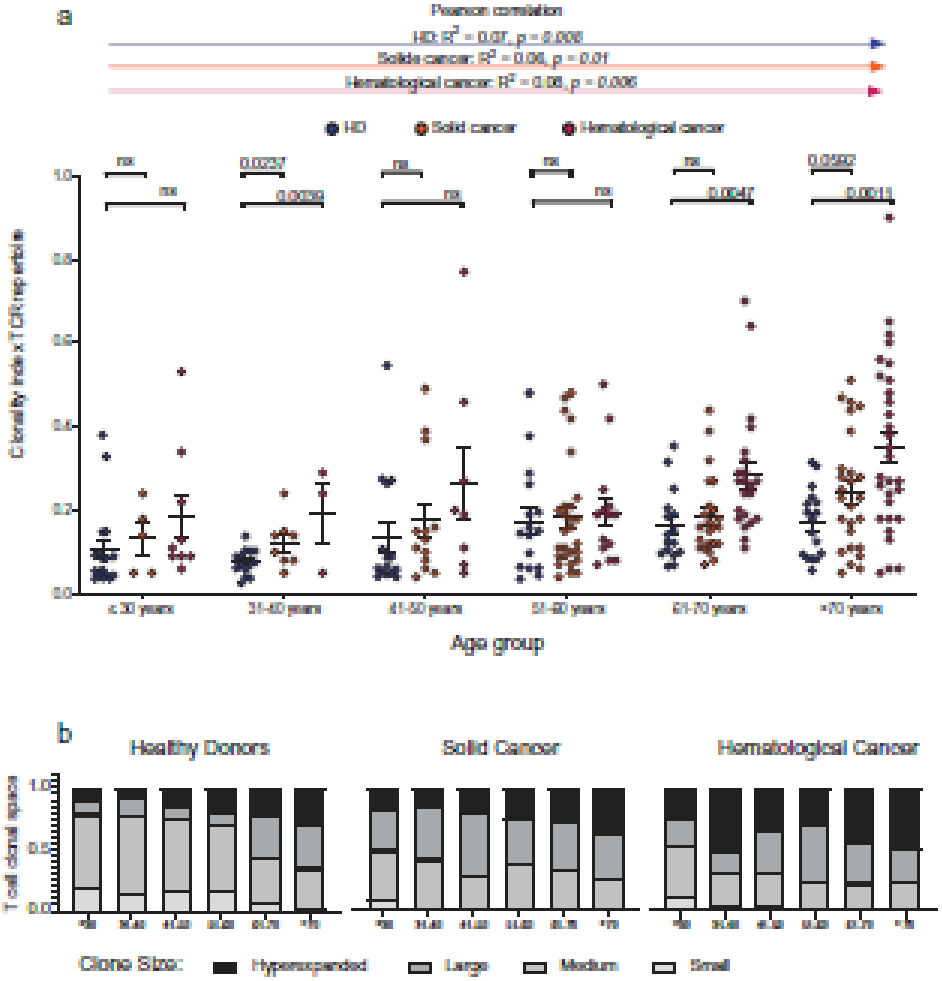


Figure 2. Age-dependent TCR repertoire clonality in healthy individuals and patients with cancer.

Clonality index with mean ±SD of PB TCR repertoire of HD, solC and hemC patients, plotted according to age group (a). Mean clonal space distribution of PB TCR repertoire in HD, solC patients and hemC patients per age group (b). Clone sizes are defined as hyperexpanded (0.01 < x ≤ 1), large (0.001 < x ≤ 0.01), medium (1E-04 < x ≤ 0.001), small (1E-05 < x ≤ 1E-04). Number of individuals per age group: HD/solC/hemC, ≤30: 17/5/9, 31–40: 14/8/5, 41–50: 15/14/8, 51–60: 16/28/15, 61–70: 16/23/24, ≥70: 17/29/37. Statistical test: unpaired, two-sided *t*-test between two subgroups, Pearson correlation between age and clonality index.

determining region 3 (CDR3) nucleotide sequence was defined as one clone and all clones of one sample/timepoint were defined as one TCR repertoire. Only productive sequences with a read count ≥ 2 were included in the analysis.

Calculation of repertoire metrics

For this work, a number of indices reflecting broad repertoire metrics were used:

The Shannon index (H) is a commonly used measure of diversity,[18] which can be calculated as follows:

$$H = \sum_{i=1}^{S} p_i \log_2 p_i \qquad (1)$$

where S is the number of species/clones (richness) and p is the proportion of each clone within the repertoire. p = n/N, n = read count of each individual clone and N = the sum of all reads in the repertoire.

Since the Shannon index is a diversity measure that weighs small and large clones relatively evenly, we plotted diversity curves with alpha-modulated sensitivity for the relatively rare clones (alpha-parameterized diversity.[19] As α increases, high frequency clones are weighed more. We generated diversity profile curves for α = 0 to α = 5, in steps of 0.2 using the R script kindly provided by Dr. V. Greiff of University of Oslo, Institute of Clinical Medicine.

Evenness is calculated from H and H_{max} (evenness = $\frac{H}{H_{max}}$) with H_{max} being the maximal possible value of H, if every clone in the repertoire was present at the same frequency.

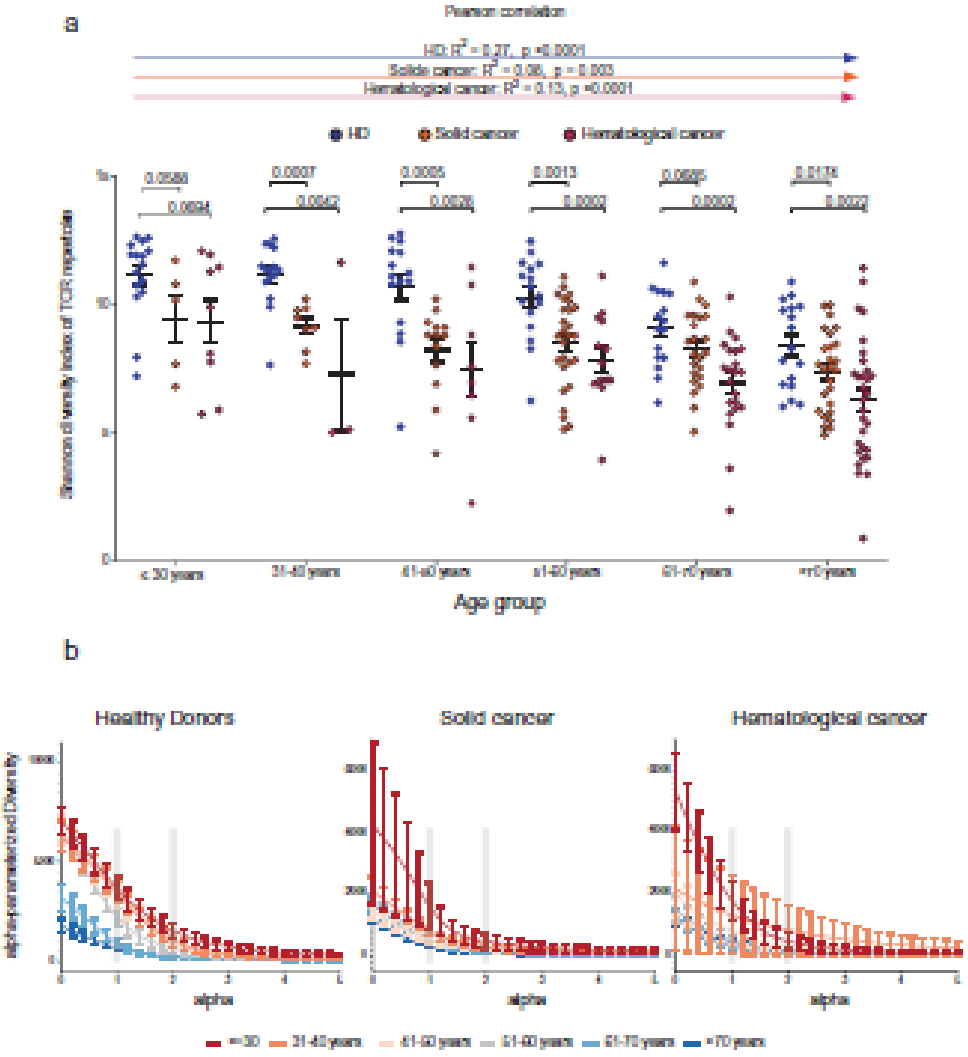


Figure 3. Age-dependent TCR repertoire diversity in healthy individuals and patients with cancer.
Shannon diversity index with mean ±SD of PB TCR repertoire of HD, solC and hemC patients, plotted according to age group (a). Mean ±SEM diversity curves of PB TCR repertoire in HD, solC patients and hemC patients per age group (b). Grey boxes indicate alpha = 1 resembles Shannon index, alpha = 2 resembles Simpson's index. Number of individuals per age group: HD/solC/hemC, ≤30: 17/8/9, 31–40: 14/8/3, 41–50: 15/14/8, 51–60: 16/28/15, 61–70: 16/23/24, ≥70: 17/20/37. Statistical test: unpaired, two-sided t-test between two subgroups, Pearson correlation between age and Shannon index.

The clonality index is the reciprocal value of evenness (clonality = 1 – evenness). It assumes a value between 1 and 0, with 1 being a repertoire consisting of only one clone and 0 being a repertoire of maximal evenness.

Generation probability calculation

We investigated the generation probability of the T cell clones in our paired patient samples before and after chemotherapy treatment using the IGoR algorithm.[20] All values are depicted on a square root scale for plotting purposes. T cell clones with low generation probability are presumed to be peripherally expanded by antigen pressure, whereas T cell clones exhibiting a high generation probability accumulate by random V(D)J recombination. Therefore, a T cell repertoire, which has seen antigens over a lifetime is expected to have a lower mean generation probability compared to a newly formed T cell repertoire (e.g. after lymphodepleting chemotherapy).

Data analysis

Analyses were carried out and data plotting was performed using R (version 3.4.4)[21] and the package tcR[22] as well as GraphPad Prism 7 (San Diego, CA). A P value of <0.05 was considered statistically significant.

Data availability

The datasets generated for this study can be found in the European Nucleotide Archive (ENA). ID: PRJEB33490

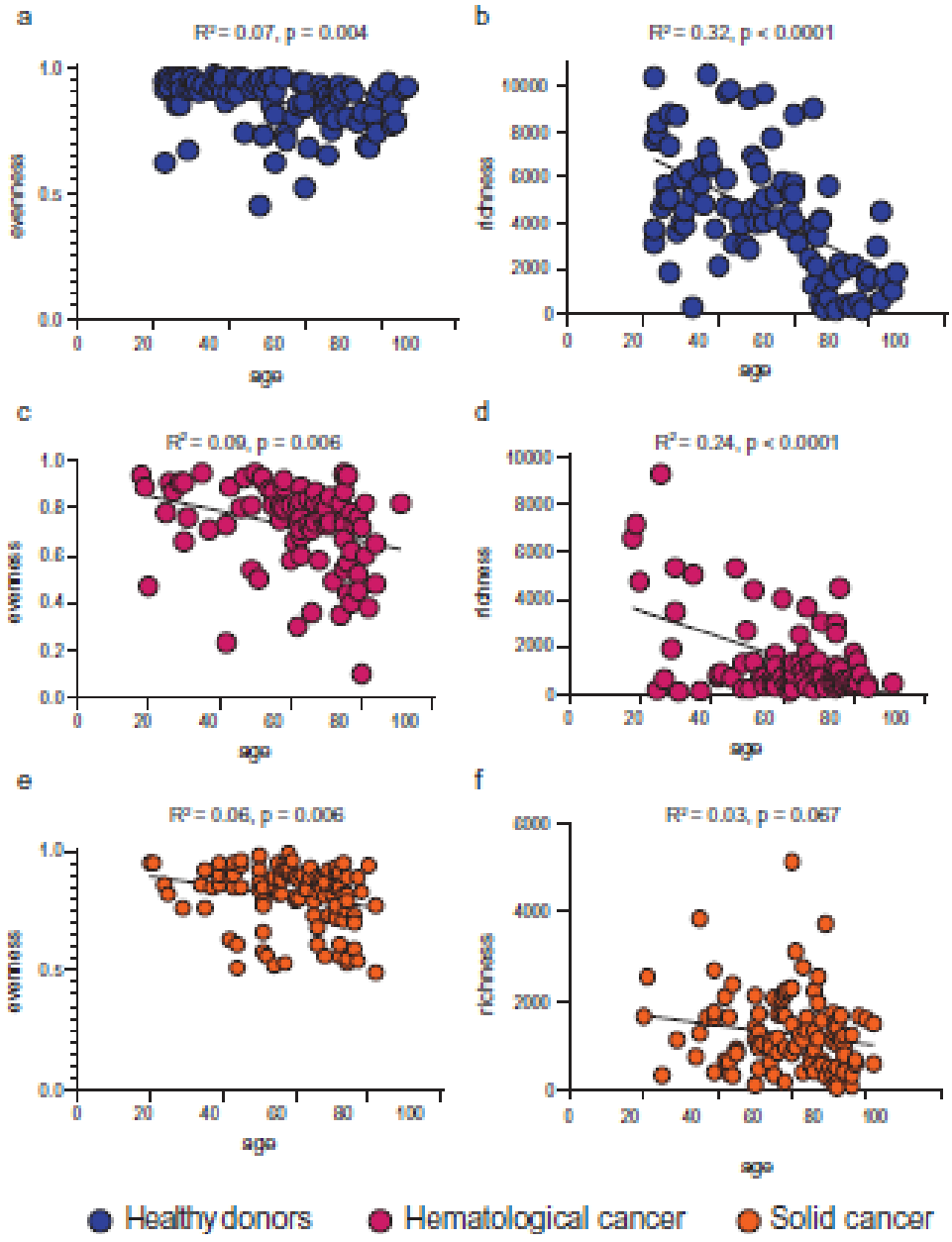


Figure 4. Correlation between age and TCR repertoire evenness/richness.

Pearson correlation between age [years] and T cell repertoire evenness in HD (a), hemC (c) and solC (e) patients. Pearson correlation between age [years] and T cell repertoire richness in HD (b), hemC (d) and solC (f). Black line indicates regression model.

Results

Broad immune metrics in healthy individuals vs. cancer patients

A cohort of 85 untreated patients with hematological cancers (hemC) and 108 patients with solid tumors (solC) was subjected to peripheral blood T cell repertoire profiling by NGS and their T cell spaces were compared to 95 healthy individuals (HD). To account for differences in the immune repertoire due to age differences, we compared subcohorts of patients (hemC and solC, respectively) to age-matched subcohorts of HDs. Untreated patients with hematological (but non-T-cell) cancers showed significantly higher clonality and lower blood T cell diversity ($p < .0001$) compared to HDs (Figure 1a,c). This was expected since these patients had infiltration of primary and/or secondary lymphoid organs potentially impairing emigration of hematopoietic precursors to the thymus and/or peripheral T cell expansion. Yet, also

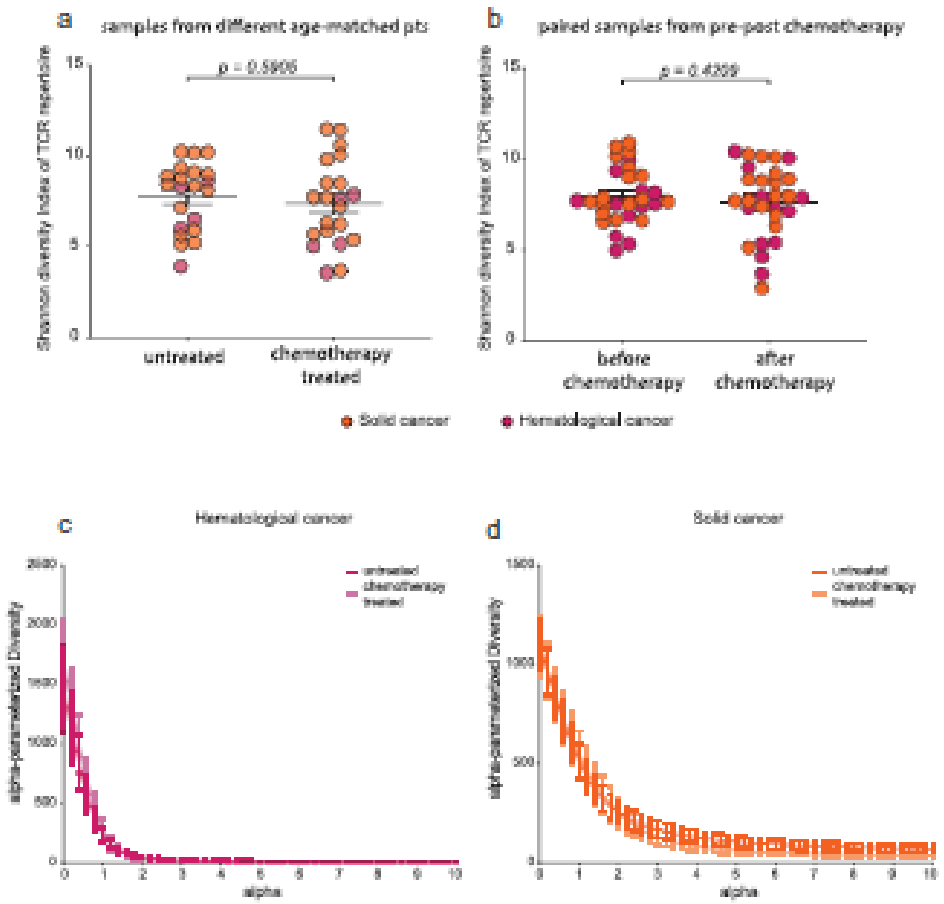


Figure 5. Effect of chemotherapy on T cell diversity in cancer patients.

Shannon diversity index with mean ±SEM of PB TCR repertoire of untreated and chemotherapy treated age-matched solC (n = 5) and hemC (n = 16) patient samples (a) and of paired solC (n = 12) and hemC (n = 17) patient samples before and after chemotherapy (b). Mean ±SEM diversity curves of PB TCR repertoire of untreated and chemotherapy treated samples from hemC patients (n = 16 + 17) (c) and solC patients (n = 5 + 12) (d). Statistical test: unpaired, two-sided t-test.]

untreated patients with solid malignancies showed significantly more clonal ($p = .0067$) and less diverse ($p < .0001$) repertoires compared to their respective age-matched subgroup of HDs (Figure 1b,d). Taken together, HDs showed more diverse and more evenly distributed peripheral blood TCR repertoires as compared to patients with hematological or solid cancers.

Age-dependent immune metrics in healthy individuals and cancer patients

Patients and HDs were binned in to age groups spanning ten years ranging from below 30 to over 70 years of age. As rough estimates of age-dependent repertoire metrics, T cell clonality and diversity were calculated (Figures 2a and 3a). In patients with cancer, the increase in repertoire clonality started at earlier age groups as compared to HDs (Figure 2a). Mean clonal space distributions showed that in elderly HDs up to an age of 60 years, small clones still make up approximately 15% of the T cell space while the repertoire fraction of small T cell clones is neglectable in cancer patients over 30 years of age (Figure 2b). In line with previously published data, [10] healthy donors showed a clear contraction of their T cell repertoires with increasing age as indicated by the decreased diversity, beginning approximately beyond age 40 (correlation coefficient $R^2 = 0.27$, $p < .0001$, Figure 3a,b). For patients with solid or hematological malignancies a similar age-dependent decline of T cell repertoire diversity was observed, however, at much steeper slope (correlation coefficient solid cancer $R^2 = 0.08$, $p = .003$, hematological cancer $R^2 = 0.13$, $p < .0001$, Figure 3a,b). We dissected the single parameters of diversity (richness and evenness) to better understand how they account for the decline in diversity over age (Figure 4). The decreased diversity in healthy aging appeared to be most explained by a loss of richness since the repertoires stayed relatively even over time (Figure 4a,b). In the aging cancer patient, however, both losses in richness and evenness contributed to the loss of T cell diversity (Figure 4c,d).

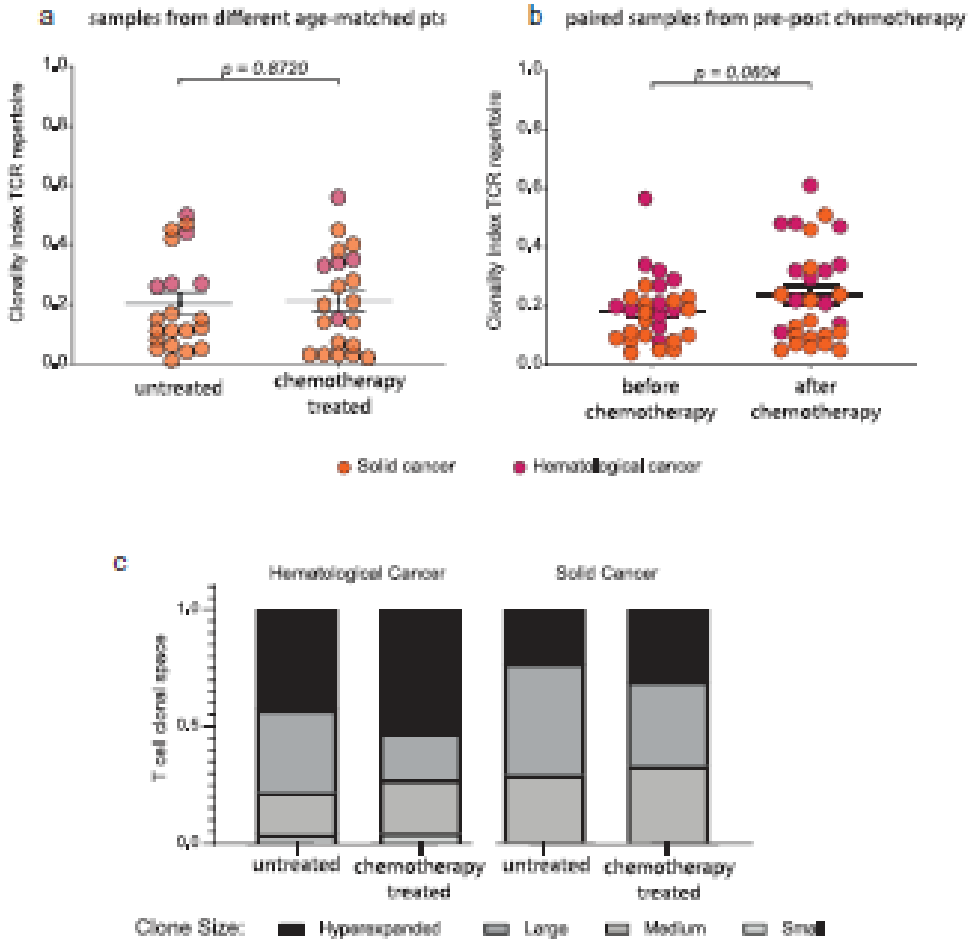


Figure 6. Effect of chemotherapy on T cell clonality in cancer patients.

Clonality index with mean ±SD of PB TCR repertoire of untreated and chemotherapy treated age-matched solC (n=5) and hemC (n=16) patient samples (a) and of paired solC (n=12) and hemC (n=17) patient samples before and after chemotherapy (b). Mean clonal space distribution of PB TCR repertoire of untreated and chemotherapy treated samples from solC patients (n=5+12) and hemC patients (n=16+17) (c). Clone sizes are defined as hyperexpanded ($0.01 < x \leq 1$), large ($0.001 < x \leq 0.01$), medium ($1E\text{-}04 < x \leq 0.001$), small ($1E\text{-}05 < x \leq 1E\text{-}04$). Statistical test: unpaired, two-sided t-test.

Reconstitution of age-specific T cell repertoire metrics after chemotherapy

Next, we investigated T cell regenerative potential after chemotherapy in patients with cancer and a median age of 60 years. In these patients, disease- and age-specific T cell repertoire differences were already discernible before treatment initiation. We investigated a cohort of cancer patients (predominantly with solid tumors) after various types of chemotherapy and compared them with a control cohort of disease- and age-matched untreated cancer patients (Figures 5a and 6a). Interestingly, there were no significant differences between treated and untreated patients in terms of T cell repertoire diversity or clonality (Figures 5a and 6a). In addition, we wished to assess repertoire changes through chemotherapy in individual patients (paired subcohort) over time. This analysis – carried out in patients with chronic lymphocytic leukemia (CLL) receiving fludarabine- or bendamustine-based therapy and in brain tumor patients receiving temozolomide-based therapy – suggested that the T cell repertoire largely recovered to the age-specific pre-treatment diversity/clonality, even in elderly patients undergoing T cell toxic treatment (Figures 5b and 6b). Only in the fludarabine-treated cohort of haematological patients with CLL a trend towards higher clonality after treatment was recognizable (Figure 6b), but diversity measures remained stable (Figure 5b). Taking together both cohorts, no drastic changes in diversity (Figure 5c,d) or clonality (Figure 6c) were apparent after chemotherapy.

Rebound thymic activity rather than peripheral T cell expansion leads to reconstitution of the T cell repertoire after chemotherapy

Repertoire overlap analysis in the two cohorts (glioblastoma, CLL) with paired pre- and post-treatment blood samples showed almost no clonal overlap whereas control TCR repertoires of HDs recorded over several months to years presented a clear clonal overlap even after years of re-sampling (Figure 7). This corroborated our assumption that full reconstitution of TCR repertoire diversity and evenness after highly hematotoxic

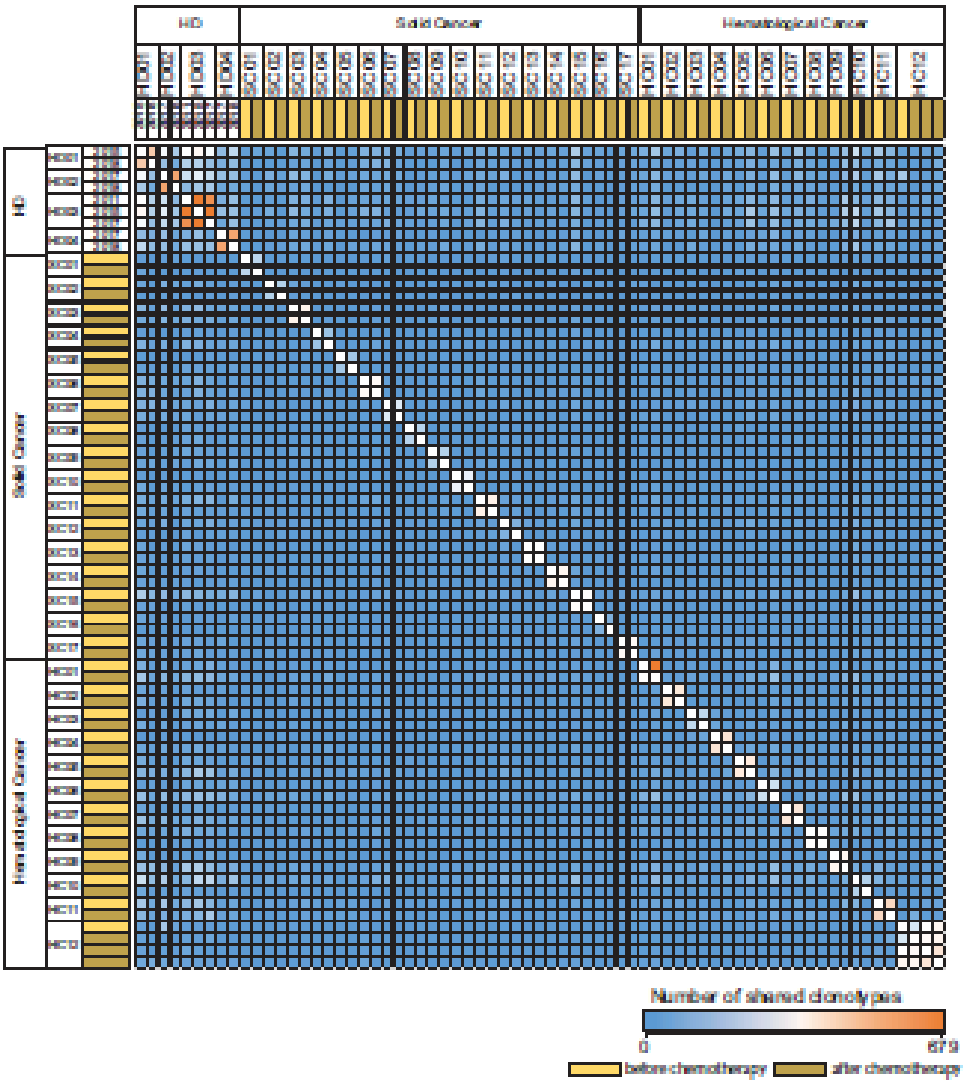


Figure 7. T cell repertoire overlap in healthy individuals and patients with cancer pre and post chemotherapy.
TCR repertoire overlap matrix based on CDR3 amino acid sequence of long-term follow-up samples of HD ($n = 4$) (indicated by year dates) and of paired solC patient ($n = 17$) and hemC patient ($n = 12$) samples before and after chemotherapy, respectively. CDR3 – complementarity determining region 3

therapies was due to rebound thymic activity rather than peripheral T cell expansion. To confirm this further, we subjected these repertoires to an algorithm, which calculates the generation probability for each recombined receptor sequence found in the productive repertoire. Since TCR repertoires shaped over time by clonal peripheral expansion contain higher numbers of clones with lower generation probabilities (selected by specific antigens), we hypothesized that rebound thymic output after chemotherapy may lead to repertoires shifted towards clones with higher generation probabilities. Generally, we observed lower generation probabilities in T cells of untreated cancer patients as compared to healthy subjects, pointing to more peripheral expansion of clones in the cancer patients. Moreover, in line with our hypothesis, we observed a clear shift (hemC: $p = 1.3E-15$, solC: $p < 2.2E-16$) towards an increased median generation probability over the treatment interval in the cancer patient cohorts (Figure 8). This data suggested that hematotoxic chemotherapies eradicate T cell clones in the periphery followed by repertoire reconstitution through rebound thymic output.

Discussion

Age-, disease- and treatment-associated impairment of T cell immunity in cancer patients is clinically relevant as it determines susceptibility to infection (especially viral reactivations and opportunistic infections) as well as anti-tumor immune control. It may also be a relevant factor that influences response to immune-activating antibodies that have become one of the mainstays of systemic cancer treatment over the

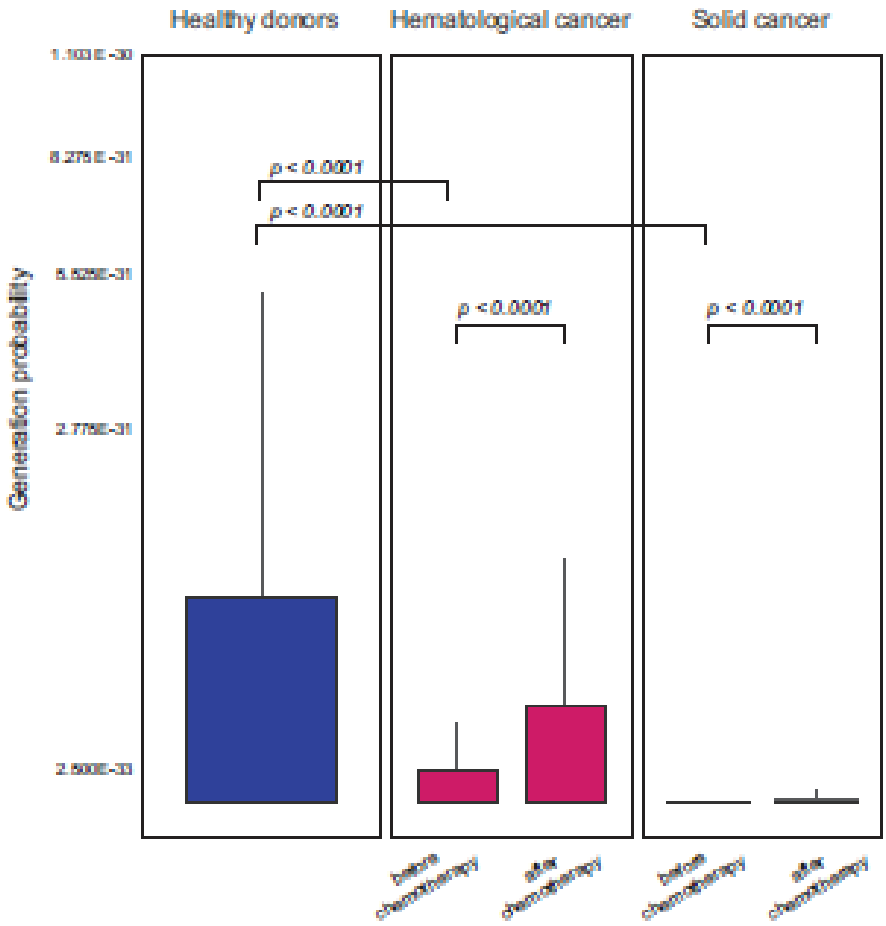


Figure 8. Generation probability of TCR repertoires.
Boxplot of generation probability of long-term follow-up TCR repertoire of HD ($n = 4$) and paired samples of patients with hemC ($n = 12$) and patients with solC ($n = 17$) before and after chemotherapy, respectively. Statistical test: Wilcoxon test.

past years. NGS immunosequencing is an advanced technology for T cell repertoire metrics analysis with the potential to precisely assess repertoire changes in immunoaging, cancer-associated immunosuppression or post-treatment immune reconstitution. It allows simultaneous identification of tens of thousands to millions of T cell receptor (TCR) rearrangements from a single tissue sample. In this way, this technology makes it possible to characterize large repertoires in depth and at high throughput, to monitor repertoires over time and to integrate all information to derive quantitative and reliable repertoire metrics such as clonality and diversity indices or clonal space distribution.

Here, we report the first study applying this approach to an unprecedented number of peripheral blood samples derived from chemotherapy-treated and -untreated patients with solid and hematological malignancies as well as healthy control subjects. The NGS data deposited with this manuscript may serve as the so far most extensive publically accessible database on T cell immune repertoires and may be exploited by the scientific community for further analyzes. Bioinformatic analyzes presented in this manuscript may be condensed to three essential findings: i) the T cell space qualitatively deteriorates with age as evidenced by loss of diversity and increasing clonality, ii) the age-dependent impairment in T cell metrics is more pronounced in patients with cancer than in healthy subjects and iii) age-specific repertoire metrics can be fully restored after lymphotoxic treatment even in elderly patients with cancer. These findings open up interesting new perspectives on immunoaging, carcinogenesis and immune reconstitution in cancer patients as well as they may influence the way we think about treatment algorithms of immunological and cytotoxic cancer treatment. One essential aspect of this (which is in contrast to previously published work[23,24]) is our NGS data suggest there is a significant contribution by the thymus to immune reconstitution after T cell toxic chemotherapy in adults, even in the age group above 70 years. The full recovery of T cell diversity after treatment in this age group was rather unexpected and – while acknowledging that this data does not allow to draw any conclusions about immunological memory – it suggests that we may not have

to fear drastic and uncompensated diversity losses through chemotherapy that would endanger the success of subsequent administered immune-stimulating antibodies in the elderly population.

Moreover, we were surprised to see clear-cut signs of premature immunoaging in treatment-naïve cancer patients. Of note, the considerably impaired diversity and clonality indices in cancer patients were not a confounding result of infiltration of lymphatic organs with subsequent displacement of lymphopoiesis since also patients with solid tumors without infiltration of lymphatic organs showed the same impairment. Clonality is not the counterpart of diversity in that increases in clonality do not necessarily reduce the overall diversity of the repertoire. Therefore, while increased blood T cell clonality in patients with cancer may simply reflect ongoing anti-tumor immune responses, the significantly lower age-adjusted diversity in these T cell repertoires may also point to pre-existing T cell repertoire defects that could be seen not only as a consequence, but also as a potential basis favoring carcinogenesis.

As TCR sequencing becomes more and more available to the scientific community and a growing number of research questions in the field of autoimmunity, vaccination and infection are addressed, our data may add a new level of evidence for age-specific biases in TCR repertoires. Studies on T cell repertoires need to be aware of these biases described in our study and therefore need to rigorously work with age-matched cohorts.

Taken together, our work gives valuable insight into immunoaging in health and disease and provides a reference NGS dataset of TCR repertoires across all age groups that may serve the scientific community for future studies.

Acknowledgments

We thank Barbara Goesch for expert technical assistance and Janina Schlüter for processing of blood samples. We like to acknowledge Prof. Zeller for the allocation of healthy donor genomic DNA samples used in this study. Finally, we are grateful to all healthy donors and patients for their valuable blood donation.

Disclosure statement

Author IF is employed by ENPICOM B.V. All other authors declare no competing interests.

Funding

This work was supported by the Deutsche Forschungsgemeinschaft [SFB841 to MB].

References

1. Montecino-Rodriguez E, Berent-Maoz B, Dorshkind K. Causes, consequences, and reversal of immune system aging. J Clin Invest. 2013; 123(3):958–965. Epub 2013/03/05. doi:10.1172/JCI64096. PubMed PMID: 23454758; PMCID: PMC3582124.
2. Simon AK, Hollander GA, McMichael A. Evolution of the immune system in humans from infancy to old age. Proc Biol Sci. 2015; 282(1821):20143085. Epub 2015/12/25. doi:10.1098/rspb.2014.3085. PubMed PMID: 26702035; PMCID: PMC4707740.
3. Famili F, Wiekmeijer AS, Staal FJ. The development of T cells from stem cells in mice and humans. Future Sci OA. 2017; 3(3): Fso186. Epub 2017/09/09. doi:10.4155/fso-2016-0095. PubMed PMID: 28883990; PMCID: PMC5583695.
4. Yanes RE, Gustafson CE, Weyand CM, Goronzy JJ. Lymphocyte generation and population homeostasis throughout life. Semin Hematol. 2017; 54(1):33–38. Epub 2017/01/17. doi:10.1053/j.seminhematol.2016.10.003. PubMed PMID: 28088985; PMCID: PMC5260809.
5. Linton PJ, Dorshkind K. Age-related changes in lymphocyte development and function. Nat Immunol. 2004; 5(2):133–139. Epub 2004/01/30. doi:10.1038/ni1033. PubMed PMID: 14749784.
6. Den Braber I, Mugwagwa T, Vrisekoop N, Westera L, Mogling R, de Boer AB, Willems N, Schrijver EH, Spierenburg G, Gaiser K, et al. Maintenance of peripheral naive T cells is sustained by thymus output in mice but not humans. Immunity. 2012; 36(2):288–297. Epub 2012/03/01. doi:10.1016/j.immuni.2012.02.006. PubMed PMID: 22365666.
7. Mitchell WA, Lang PO, Aspinall R. Tracing thymic output in older individuals. Clin Exp Immunol. PubMed PMID: WOS:000280909300012. 2010;161(3):497–503. doi:10.1111/j.1365-2249.2010.04209.x.
8. Douek DC, Vescio RA, Betts MR, Brenchley JM, Hill BJ, Zhang L, Berenson JR, Collins RH, Koup RA. Assessment of thymic output in adults after haematopoietic stem-cell transplantation and prediction of T-cell reconstitution. Lancet. 2000; 355(9218):1875–1881. Epub 2000/06/24. doi:10.1016/S0140-6736(00)02293-5. PubMed PMID: 10866444.
9. Williams KM, Hakim FT, Gress RE. T cell immune reconstitution following lymphodepletion. Semin Immunol. 2007; 19(5):318–330. Epub 2007/11/21. doi:10.1016/j.smim.2007.10.004. PubMed PMID: 18023361; PMCID: PMC2180244.
10. Britanova OV, Putintseva EV, Shugay M, Merzlyak EM, Turchaninova MA, Staroverov DB, Bolotin DA, Lukyanov S, Bogdanova EA, Mamedov IZ, et al. Age-related decrease in TCR repertoire diversity measured with deep and normalized sequence profiling. J Immunol. 2014; 192(6):2689–2698. Epub 2014/02/11. doi:10.4049/jimmunol.1302064. PubMed PMID: 24510963.
11. van Dongen JJ, Langerak AW, Brüggemann M, Evans PA, Hummel M, Lavender FL, Delabesse E, Davi F, Schuuring E, Garcia-Sanz R, et al. Design and standardization of PCR primers and protocols for detection of clonal immunoglobulin and T-cell receptor gene recombinations in suspect lymphoproliferations: report of the BIOMED-2 Concerted Action BMH4-CT98-3936. Leukemia. PubMed PMID: 14671650. 2003;17(12):2257–2317. doi:10.1038/sj.leu.2403202.
12. Akyüz N, Brandt A, Stein A, Schliffke S, Mahrle T, Quidde J, Goekkurt E, Loges S, Hanick T, Ford CT, et al. T-cell diversification reflects antigen selection in the blood of patients on immune checkpoint inhibition and may be exploited as liquid biopsy biomarker. Int J Cancer. 2017; 140(11):2535–2544. Epub 2016/12/08. doi:10.1002/ijc.30549. PubMed PMID: 27925177.
13. Mährle T, Akyüz N, Fuchs P, Bomanni N, Simnica D, Germing U, Asemissen AM, Jann J-C, Nolte F, Hofmann W-K, et al. Deep sequencing of bone marrow microenvironments of patients with del(5q) myelodysplastic syndrome reveals imprints of antigenic selection as well as generation of novel T cell clusters as response pattern to lenalidomide2019. Haematologica. 2019 Jul;104(7):1355–1364.
14. Mohme M, Schliffke S, Maire CL, Runger A, Glau L, Mende KC, Matschke J, Gehbauer C, Akyuz N, Zapf S, et al. Immunophenotyping of newly diagnosed and recurrent glioblastoma defines distinct immune exhaustion profiles in peripheral and tumor-infiltrating lymphocytes. Clin Cancer Res. 2018;24(17):4187–4200. Epub 2018/02/16. doi:10.1158/1078-0432.CCR-17-2617. PubMed PMID: 29444930.
15. Schliffke S, Akyuz N, Ford CT, Mahrle T, Thanhaeuser T, Krohn-Grimberghe A, Knop S, Bokemeyer C, Binder M. Clinical response to ibrutinib is accompanied by normalization of the T-cell environment in CLL-related autoimmune cytopenia. Leukemia. 2016;30(11):2232–2234. Epub 2016/11/03. doi:10.1038/leu.2016.157. PubMed PMID: 27220665.
16. Schliffke S, Caembia A, Akyüz N, Thiele B, Herkel J, Binder M. T-cell repertoire profiling by next-generation sequencing reveals tissue migration dynamics of TRBV13-family clonotypes in a common experimental autoimmune encephalomyelitis mouse model. J Neuroimmunol. 2019;330:49–56. doi:10.1016/j.jneuroim.2019.03.014.
17. Bolotin DA, Poslavsky S, Mitrophanov I, Shugay M, Mamedov IZ, Putintseva EV, Chudakov DM. MiXCR: software for comprehensive adaptive immunity profiling. Nat Methods. 2015;12(5):380–381. Epub 2015/04/30. doi:10.1038/nmeth.3364. PubMed PMID: 25924071.
18. Shannon CE. A Mathematical Theory of Communication. Bell Syst Tech J. PubMed PMID: WOS:A1948UH0300001. 1948;27(3):379–423. doi:10.1002/j.1538-7305.1948.tb01338.x.
19. Greiff V, Bhat P, Cook SC, Menzel U, Kang W, Reddy ST. A bioinformatic framework for immune repertoire diversity profiling enables detection of immunological status. Genome Med. 2015 May 28;7(1):49. doi:10.1186/s13073-015-0169-8. PubMed PMID: 26140055; PubMed Central PMCID: PMC4489130.
20. Marcou Q, Mora T, Walczak AM. High-throughput immune repertoire analysis with IGoR. Nat Commun. 2018; 9(1):561. Epub 2018/02/10. doi:10.1038/s41467-018-02832-w. PubMed PMID: 29422654; PMCID: PMC5805751.
21. Team RC. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2018. Available from: https://www.R-project.org.
22. Nazarov VI, Pogorelyy MV, Komech EA, Zvyagin IV, Bolotin DA, Shugay M, Chudakov DM, Lebedev YB, Mamedov IZ. tcR: an R package for T cell receptor repertoire advanced data analysis. BMC Bioinf. 2015;16:175. Epub 2015/05/29. doi:10.1186/s12859-015-0613-1. PubMed PMID: 26017500; PMCID: PMC4445501.
23. Hakim FT, Memon SA, Cepeda R, Jones EC, Chow CK, Kasten-Sportes C, Odom J, Vance BA, Christensen BL, Mackall CL, et al. Age-dependent incidence, time course, and consequences of thymic renewal in adults. J Clin Invest. 2005; 115(4):930–939. Epub 2005/03/19. doi:10.1172/jci23942. PubMed PMID: 15776111; PMCID: PMC1064081.
24. Hare M, McAdams HP, Vredenburgh JJ, Herndon JE, Patz EF. Thymic hyperplasia after high-dose chemotherapy and autologous stem cell transplantation: incidence and significance in patients with breast cancer. Am J Roentgenol. 1999;173(5):1341–1344. doi:10.2214/ajr.173.5.10541115.

CONTRIBUTIONS TO THIS WORK

The manuscript “T cell receptor next-generation sequencing reveals cancer-associated repertoire metrics and reconstitution after chemotherapy in patients with hematological and solid tumors” was accepted for publication in July 2019 in *Oncoimmunology*. This project was initiated and designed by Mascha Binder. We aimed to deepen our understanding of the dynamics of the aging immune system as well as the onset of immunoparesis in healthy individuals and cancer patients using TRB repertoire NGS data of more than 200 individuals. The age of the studied subjects ranged from under 30 years to over 70 years.

I contributed to the DNA extraction of PBMNC samples as well as TRB PCRs and NGS library preparation. Furthermore, I was responsible for data mining, which included but was not limited to the storage and organization of raw data and sample information like demographics, diagnosis and treatment. I was responsible for data analyses of the NGS data, which included the conversion of raw NGS data (FASTQ-files) to human-readable text-files using our in-house bioinformatics pipeline. Furthermore, I performed data analyses using established R scripts and R scripts that I wrote myself. Finally, in close collaboration with my supervisor Mascha Binder, I interpreted the results and contributed to the first draft of the manuscript as well as all revisions made until final publication.

_____

Prof. Dr. med. Mascha Binder

2.2 High-Throughput Immunogenetics Reveals A Lack Of Physiological T Cell Clusters In Patients With Autoimmune Cytopenias

Authors

Simnica D, Schliffke S, Schultheiß C, Bonzanni N, Fanchi F. L, Akyüz N, Gösch B, Casar C, Thiele B, Schlüter J, Lohse W. A, Binder M.

Published in

Frontiers in Immunology. August 2019. doi.org/10.3389/fimmu.2019.01897

High-Throughput Immunogenetics Reveals a Lack of Physiological T Cell Clusters in Patients With Autoimmune Cytopenias

*Donjete Simnica[1], Simon Schliffke[2], Christoph Schultheiß[1], Nicola Bonzanni[3], Lorenzo F. Fanchi[3], Nuray Akyüz[2], Barbara Gösch[2], Christian Casar[4,5], Benjamin Thiele[2], Janina Schlüter[2], Ansgar W. Lohse[4] and Mascha Binder[1]**

[1] Department of Internal Medicine IV, Oncology and Hematology, Martin-Luther-University Halle-Wittenberg, Halle, Germany, [2] Department of Oncology and Hematology, BMT With Section Pneumology, University Cancer Center Hamburg, University Medical Center Hamburg-Eppendorf (UKE), Hamburg, Germany, [3] ENPICOM B.V., 's-Hertogenbosch, Netherlands, [4] Department of Gastroenterology With Sections Infectious and Tropical Diseases, University Medical Center Hamburg-Eppendorf (UKE), Hamburg, Germany, [5] Bioinformatics Core, University Medical Center Hamburg-Eppendorf (UKE), Hamburg, Germany*

OPEN ACCESS

Edited by:
Laurence Morel, University of Florida, United States

Reviewed by:
Yu Nee Lee, Shaba Medical Center, Israel
Anna Vardi, Institute of Applied Biosciences - The Centre for Research & Technology (CERTH), Greece

***Correspondence:**
Mascha Binder
mascha.binder@uk-halle.de

Specialty section:
This article was submitted to Autoimmune and Autoinflammatory Disorders, a section of the journal Frontiers in Immunology

Received: *21 May 2019*
Accepted: *26 July 2019*
Published: *21 August 2019*

Citation:
Simnica D, Schliffke S, Schultheiß C, Bonzanni N, Fanchi LF, Akyüz N, Gösch B, Casar C, Thiele B, Schlüter J, Lohse AW and Binder M (2019) High-Throughput Immunogenetics Reveals a Lack of Physiological T Cell Clusters in Patients With Autoimmune Cytopenias. Front. Immunol. 10:1897. doi: 10.3389/fimmu.2019.01897

Autoimmune cytopenias (AIC) such as immune thrombocytopenia or autoimmune hemolytic anemia are claimed to be essentially driven by a dysregulated immune system. Using next-generation immunosequencing we profiled 59 T and B cell repertoires (*TRB* and *IGH*) of 25 newly diagnosed patients with primary or secondary (lymphoma-associated) AIC to test the hypothesis if these patients present a disease-specific immunological signature that could reveal pathophysiological clues and eventually be exploited as blood-based biomarker. Global *TRB* and *IGH* repertoire metrics as well as *VJ* gene usage distribution showed uniform characteristics for all lymphoma patients (high clonality and preferential usage of specific *TRBV*- and *TRBJ* genes), but no AIC-specific signature. Since T cell immune reactions toward antigens are unique and polyclonal, we clustered TCRβ clones *in-silico* based on target recognition using the GLIPH (grouping of lymphocyte interactions by paratope hotspots) algorithm. This analysis revealed a considerable lack of physiological T cell clusters in patients with primary AIC. Interestingly, this signature did not discriminate between the different subentities of AIC and was also found in an independent cohort of 23 patients with active autoimmune hepatitis. Taken together, our data suggests that the identified T cell cluster signature could represent a blood biomarker of autoimmune conditions in general and should be functionally validated in future studies.

Keywords: autoimmunity, AIC, CLL, GLIPH, immunosequencing, NGS, biomarker

INTRODUCTION

Autoimmune cytopenias (AIC) mainly comprise autoimmune hemolytic anemia (AIHA), immune thrombocytopenia (ITP), and autoimmune neutropenia (AIN). These disorders are driven by immune mediated destruction of mature hematopoietic cells in the periphery. Additionally, in rare cases AIC can be due to an immune mediated maturation defect in an otherwise

normal bone marrow manifesting as pure red cell aplasia (PRCA) or acquired amegakaryocytic thrombocytopenia (AATP). The clinical course of AIC is highly diverse, ranging from subclinical presentation with abnormal laboratory findings to severe anemia, bleeding, or infection (1–3). While in some cases a clinical trigger cannot be identified, a considerable fraction of cases presents as secondary diseases in the context of an underlying malignancy (e.g., Non-Hodgkin's Lymphoma, NHL) or systemic autoimmune disease. Patients with chronic lymphocytic leukemia (CLL) have an especially high risk of developing AIC which is commonly attributed to their dysregulated immune system (4). The risk in CLL is reported to be higher in patients with adverse cytogenetics, unmutated *IGHV* of the malignant CLL clone and specific stereotyped B cell receptor configurations (5–10). Classical AIC cases rarely represent a diagnostic challenge for the hematologist, but the diagnosis may also be more intricate e.g., in cases showing overlapping patterns of immune reactivity with precursor and mature blood cells (11–16). When diagnosed, most cases of AIC—whether primary or secondary—are primarily treated with steroids (17). Second line treatments are much less defined, including other immunomodulatory drugs, growth factor receptor agonists, or splenectomy (18–20). Despite the plurality of therapeutic options, the rate of durable remissions is rather unsatisfactory to date (3). Moreover, reliable response predictive biomarkers are lacking.

Proposed mechanisms underlying AIC are: (i) Breakdown of central and/or peripheral tolerance resulting in T cell and autoantibody production against precursor and mature blood cells, (ii) complement-dependent cytotoxicity (CDC) of these autoantibodies causing blood cell clearance by macrophages in the reticuloendothelial system in spleen and liver, (iii) cellular cytotoxicity mediated by CD8+ T cells and natural killer cells (13, 21–23).

In order to gain new insights into the pathogenesis of AIC and identify signatures of autoreactive immune repertoires that could potentially be used in the diagnosis and/or follow-up of these patients, we used state-of-the-art immunosequencing technology. This let us dissect composition, dynamics and diversity of the T and B cell immune repertoire in a cohort of 25 patients with primary and secondary AIC with available follow-up biosamples. We found blood T cell signature in AIC patients consisting in a lack of physiological T cell clusters, which we confirmed in a different autoimmune cohort of 23 patients with active autoimmune hepatitis (AIH). Future studies will need to define the functional role of these clusters.

MATERIALS AND METHODS

Study Approval

Informed consent was obtained from all patients and healthy donors (HD) for the use of their peripheral blood (PB) as approved by the ethics commission Hamburg (Ethikkommission der Ärztekammer Hamburg, Germany, project numbers PV4767 and PV4081). The study has been performed in accordance with the declaration of Helsinki of 1975.

Patient Characteristics

Patient PB samples were collected between July 2012 and September 2018. The cohort comprised 25 patients with AIC. The diagnosis and response assessment was based on the criteria reported previously in the guidelines of the American Society of Hematology. In line with the distribution pattern of AIC, our cohort consisted essentially of patients with ITP ($n = 12$) and AIHA ($n = 8$), while only few patients had AIN, PRCA, or AATP (**Table 1**). Eleven of our patients had primary AIC, while 14 patients had secondary disease with an underlying lymphoid malignancy; the majority thereof was associated with CLL. As a reference and control cohorts, 43 PB samples of HD (Mean/Median Age: 56/59 years, 22 female), 14 samples from untreated patients with CLL without AIC (Mean/Median Age: 65/63 years, 5 female) and 23 PB samples of patients with active autoimmune hepatitis (AIH) were used (**Table 1**, **Supplementary Table 1**).

Multiplex PCR for T Cell Receptor Beta (TRB) and Immunoglobulin Heavy Chain (IGH) Repertoire Amplification for Illumina Targeted Next-Generation Sequencing (NGS)

The *TRB* and *IGH* genes containing the entire *V*, *D*, and *J* gene segments were amplified with *TRB*-A and –B and *IGH* primer pools, respectively, from 250 ng PB genomic DNA, which corresponds to ~37,500 genome equivalents (24). Compared to our previously published protocol, the *TRB* PCR assay was refined by using a touch-down PCR protocol and a mixed primer *TRB*-A/B tube. In two consecutive PCR reactions amplicons were tagged with Illumina adapters and indices as previously described (25, 26). PCRs were performed using Phusion HS II (Thermo Fisher Scientific Inc., Germany). Amplicons were purified after agarose gel electrophoresis using the NucleoSpin® Gel and PCR Clean-up kit (Macherey-Nagel, Germany). Before being subjected to NGS the concentration and quality of the amplicons/libraries was determined using Qubit (QIAGEN, Germany) and Agilent 2100 Bioanalyzer (Agilent technologies, Germany), respectively.

Illumina Next-Generation Sequencing (NGS) and Data Analysis

NGS and de-multiplexing was performed on an Illumina MiSeq sequencer (600–cycle single indexed, paired-end run). Analysis of the *TRB* and *IGH* locus was computed using the MiXCR analysis tool versions 2.1.12 and 3.0.6, respectively (27, 28). At the alignment step the default MiXCR library was used for *TRB* sequences and the external IMGT library v3 (29) was used for *IGH* sequences. Each unique complementarity-determining region 3 (CDR3) nucleotide sequence was defined as one clone. Only productive sequences with a read count ≥2 were included in the analysis. Analyses were carried out and data plotting was performed using R (version 3.4.4) (30) and packages tidyverse (31), tcR (32), ade4 (33), as well as GraphPad Prism 7 (San Diego, CA). A *P*-value of < 0.05 was considered statistically significant.

TABLE 1 | Patient and sample characteristics.

Patient	Age [y]	Diagnosis	Baseline	Follow-up at persistent disease	Follow-up at remission/inactive disease
			(no Tx)	Number of samples	
IMMUNE THROMBOCYTOPENIA (ITP)					
ITP 1	21–25	ITP	1	0	0
ITP 2	86–90	CLL + ITP	1	0	0
ITP 3	71–75	CLL + ITP	1	3 on Ibrutinib 2 on RB 1 w/o treatment	0
ITP 4	71–75	CLL + ITP	1	0	0
ITP 5	81–85	NHL* + ITP	1	0	0
ITP 6	40–45	ITP	1	0	0
ITP 7	76–80	ITP	1	0	0
ITP 8	71–75	CLL + ITP	1	1 on steroids	0
ITP 9	81–85	ITP	1	0	1 on steroids
ITP 10	81–85	ITP	1	1 on steroids	0
ITP 11	46–50	ITP	1	1 on steroids	0
ITP 12	81–85	ITP	1	0	1 on steroids
AUTOIMMUNE HEMOLYTIC ANEMIA (AIHA)					
AIHA 1	81–85	AIHA	1	0	1 on steroids
AIHA 2	81–85	AIHA	1	1 on steroids	0
AIHA 3	81–85	CLL + AIHA	1	1 on Ibrutinib	0
AIHA 4	81–85	CLL + AIHA	1	0	0
AIHA 5	56–60	CLL + AIHA	1	0	1 on RB
AIHA 6	76–80	CLL + AIHA	1	1 on steroids 1 on rituximab	1 on venetoclax
AIHA 7	46–50	CLL + AIHA	1	0	0
AIHA 8	81–85	CLL + AIHA	1	0	0
EVANS SYNDROME					
AIHA/ITP	41–45	AIHA + ITP	1	1 on steroids	0
PURE RED CELL APLASIA (PRCA)					
PRCA 1	81–85	CLL + PRCA	1	0	7 on Ibrutinib
AUTOIMMUNE NEUTROPENIA (AIN)					
AIN 1	71–75	CLL + AIN	1	0	5 on Ibrutinib
AQUIRED AMEGAKARYOTIC THROMBOCYTOPENIC PURPURA (AATP)					
AATP 1	95–100	AATP	1	0	0
AUTOIMMUNE HEPATITIS (AIH)					
AIH 2	46–50	AIH + PSC	0	1 on Pred/Aza	0
AIH 3	56–60	AIH + PSC	0	1 on Azathioprin and Budesonid	0
AIH 4	46–50	AIH	0	1 on Pred/Aza	0
AIH 5	51–55	AIH	0	1 on Prednisolon and Infliximab	0
AIH 6	56–60	AIH	0	1 on Pred/Aza	0
AIH 7	31–35	AIH	0	1 on Prednisolon	0
AIH 9	20–25	AIH	0	1 on Pred/Aza	0
AIH 13	46–50	AIH + PSC	0	1 on Pred/Aza	0
AIH 14	66–70	AIH + PSC	1	0	0
AIH 16	21–25	AIH	1	0	0
AIH 18	45–50	AIH + PSC	0	1 on Pred/Aza	0
AIH 22	76–80	AIH	0	1 on Pred/Aza	0
AIH 25	31–35	AIH	0	1 on Pred/Aza	0
AIH 27	76–80	AIH	0	1 on Azathioprin	0
AIH 28	71–75	AIH	0	1 on Pred/Aza	0
AIH 29	31–35	AIH	0	1 on Pred/Aza	0

(Continued)

TABLE 1 | Continued

Patient	Age [y]	Diagnosis	Baseline	Follow-up at persistent disease	Follow-up at remission/inactive disease
			(no Tx)	Number of samples	
AIH31	61–65	AIH	0	1 on Pred/Aza and Metformin	0
AIH32	26–30	AIH	0	1 on Pred/Aza	0
AIH35	71–75	AIH + PBC	0	1 on Pred/Aza	0
AIH36	66–70	AIH + PBC	0	1 on Pred/Aza	0
AIH37	61–65	AIH	1	0	0
AIH40	36–40	AIH	0	1 on Prednisolon and Mercaptopurinin	
AIH41	46–50	AIH+PBC	0	1 on Ursodeoxycholic acid (UDCA)	0

*CLL, Chronic Lymphocytic Leukemia; NHL, Non-Hodgkin Lymphoma; PBC, Primary Biliary Cholangitis; Pred/Aza, Prednisolon/Azathioprin; PSC, Primary Sclerosing Cholangitis; RB, Rituximab/Bendamustine; Tx, Treatment; *low-grade Lymphoma, not otherwise specified.*

Calculation of Repertoire Metrics

We plotted diversity curves with alpha-modulated sensitivity [alpha-parameterized diversity (34)] to model scenarios where the relatively rare clones are weighed differentially as opposed to single diversity measures (see below) that have a defined sensitivity for clones in the repertoire. Diversity profiles of each cohort were calculated using the following equation:

$$ {}^{\alpha}D\ (f) = \left(\sum_{i=1}^{n} f_i^{\alpha}\right)^{\frac{1}{1-\alpha}} $$

As α increases, high frequency clones are weighed more. We generated diversity profile curves for $\alpha = 0$ to $\alpha = 5$, in steps of 0.2 using the R script kindly provided by Dr. V. Greiff of University of Oslo, Institute of Clinical Medicine. Two special cases of alpha are Shannon index ($\alpha = 1$) and the Simpson's index ($\alpha = 2$) – highlighted by gray bars in Figure 1.

The Shannon index (H) is a commonly used measure of diversity (18), which can be calculated as follows:

$$ H = \sum_{i=1}^{S} p_i \log_2 p_i $$

Where, S is the number of species/clones (richness) and p is the proportion of each clone within the repertoire. $p = n/N$, n = read count of each individual clone and N = the sum of all reads in the repertoire.

Evenness is calculated from H and Hmax (evenness = H/Hmax) with Hmax being the maximal possible value of H, if every clone in the repertoire was present at the same frequency.

The Simpson index (D) (35) is another measure used to quantify biological diversity in a given sample and is calculated as follow:

$$ D = \sum \left(\frac{n}{N}\right)^2 $$

where n is the read count of each individual clone and N = the sum of all reads in the repertoire.

Calculation of Physio-Chemical Properties of *IGH* Data

The grand average of hydropathy index [GRAVY (36)] was calculated using BRepertoire (37). All GRAVY indices were ordered in ascending fashion for each cohort, respectively, and plotted against the cumulative relative frequency of each CDR3 amino acid sequence (1/total number of clones within cohort).

PCA of Combinatorial VJ Usage

The combinatorial *VJ* usage was calculated using the gene.usage function of the tcR package (32). The dudi.pca function of the ade4 package (33) transformed the frequencies of each *VJ* combination ($n = 650$) within two cohorts tested to principal components to find linear association of different *VJ* combinations, that separate different individuals corresponding to different disease states.

In silico GLIPH (Grouping of Lymphocyte Interactions by Paratope Hotspots)

We applied the GLIPH algorithm on our NGS generated AIC, HD, CLL, and AIH *TRB* dataset as described elsewhere (38). Briefly, we analyzed the following sets of NGS *TRB* data matched by age and read depth: primary AIC vs. HD, secondary AIC vs. CLL and active AIH vs. HD. Each cluster basically represents a consensus sequence of the CDR3 amino acid sequences it consists of. The majority of clusters found in each of the scenarios were shared between the respective cohorts analyzed. In order to compare the differences of "cluster size," we calculated the Log_2 fold-change of the mean CDR3 amino acid frequency. Log_2 fold-change for clusters which were present exclusively in one cohort were set to artificial Log_2 values for plotting purposes. We used the Wilcoxon-rank test to test if the mean frequencies were different between cohorts. A P-value of <0.05 was considered statistically significant.

RESULTS

Immunosequencing of AIC Patients and Control Cohorts

We asked if patients with AIC showed a disease-specific immune signature of their PB T and B cells and if this potential signature

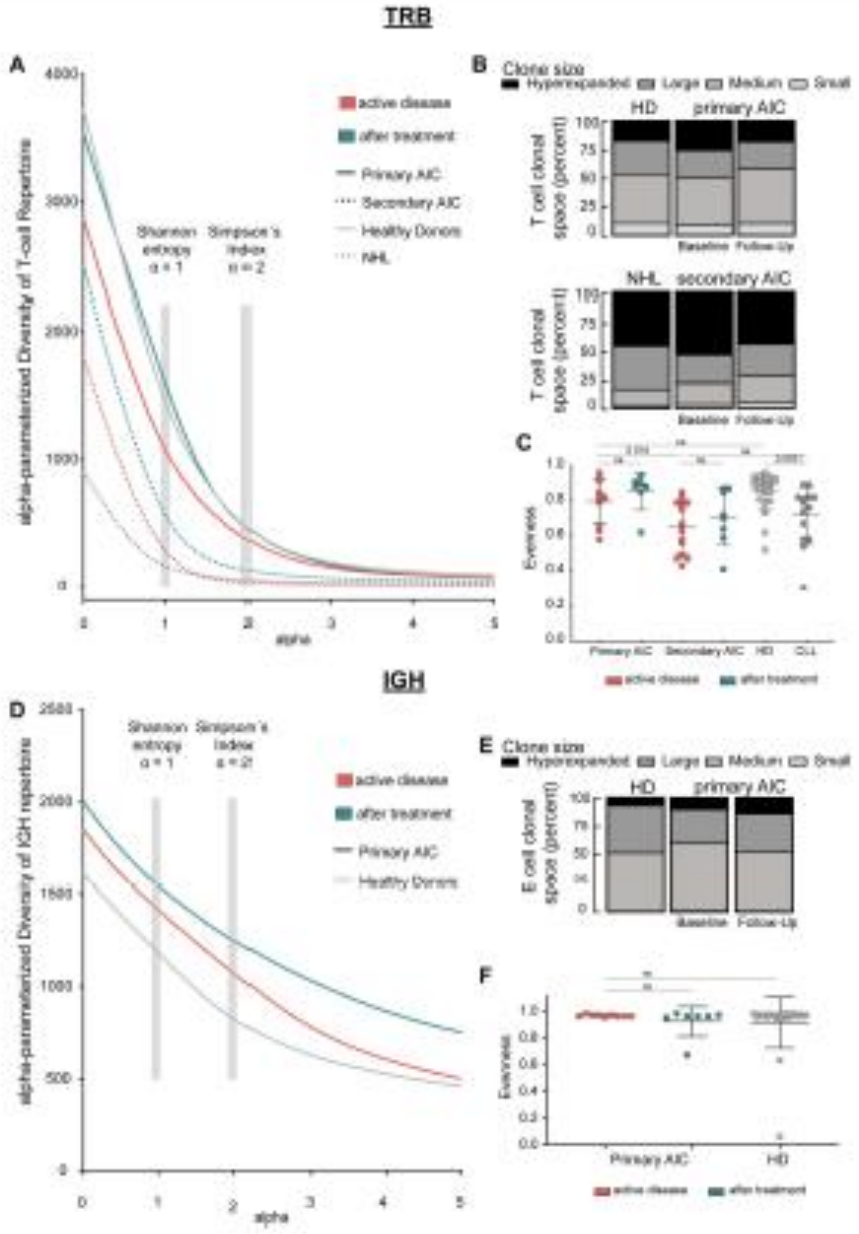


FIGURE 1 | Immune repertoire metrics. **(A)** Alpha-parameterized diversity of *TRB* repertoire. The mean diversity profile curves of *TRB* repertoires at active disease and after treatment samples of patients with primary and secondary AIC, respectively, and HD and patients with CLL are shown. **(B)** Clonal T cell space of patients with primary and secondary AIC at active disease and after treatment, respectively, HD and patients with CLL. The distribution of small, medium, large, and hyperexpanded clones in % within the clonal T cell space is shown. **(C)** *TRB* Evenness (mean ± standard deviation) of patients with primary and secondary AIC at active disease and after treatment, respectively, HD and patients with CLL. **(D)** Alpha-parameterized diversity of *IGH* repertoire. The mean diversity profile curves of *IGH* repertoires at active disease and after treatment samples of patients with primary AIC and HD are shown. **(E)** Clonal B cell space of patients with primary AIC at active disease and after treatment and HD. The distribution of small, medium, large, and hyperexpanded clones in % within the clonal B cell space is shown. **(F)** *IGH* Evenness (mean ± standard deviation) of patients with primary AIC at active disease and after treatment and HD. Clone sizes are defined as hyperexpanded ($0.01 < x \leq 1$), large ($0.001 < x \leq 0.01$), medium ($1E\text{-}04 < x \leq 0.001$), small ($1E\text{-}05 < x \leq 1E\text{-}04$). Gray boxes highlight alpha = 1, which resembles Shannon-Index, alpha = 2, which resembles Simpson's Index. Statistical analysis: Two-sided unpaired *t*-test. Sample sizes in all panels are: Active disease primary AIC ($n = 11$), secondary AIC ($n = 14$); after treatment primary AIC ($n = 7$), secondary AIC ($n = 8$); HD ($n = 25$), and patients with CLL ($n = 14$).

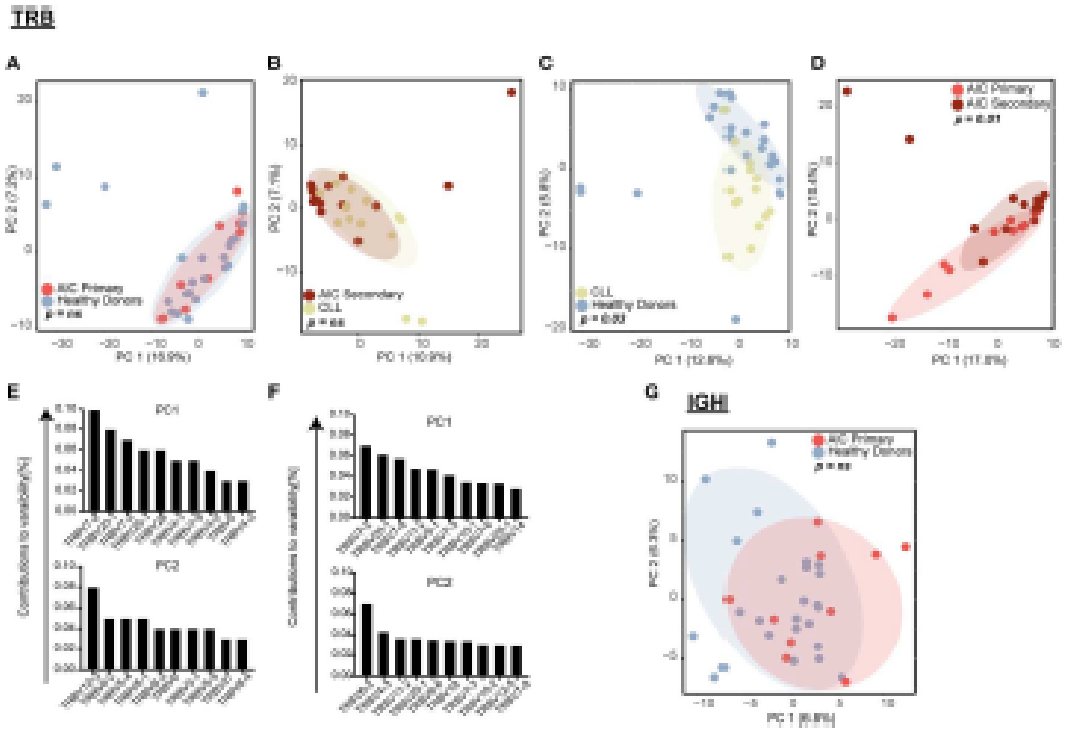


FIGURE 2 | Principal component analysis (PCA) of *VJ* gene usage. *TRB* VJ PCA of **(A)** primary AIC vs. HD, **(B)** secondary AIC vs. CLL, **(C)** HD vs. CLL, and **(D)** primary vs. secondary AIC. Ten highest contributing variables for PCA of HD vs. CLL **(E)** and primary vs. secondary AIC **(F)** for each principal component in descending order are shown. **(G)** *IGH* VJ PCA for primary AIC vs. HD. Statistical analysis: Pillai-Bartlett test of MANOVA of principal components 1 and 2. AIC, Autoimmune Cytopenia; *IGHV*, Immunoglobulin heavy-chain V gene; CLL, Chronic Lymphocytic Leukemia; PC, Principal Component; *TRBV*, T cell receptor beta V gene.

was associated with response to immunomodulatory treatment. Therefore, we performed *TRB* and *IGH* NGS on 25 AIC patients with active disease and on 34 follow-up samples in the course of immunomodulatory treatment. As a control group, we used 25 age-matched HD for the primary AIC cohort and 14 age-matched patients with CLL for the secondary AIC cohort. We included an independent cohort of 23 patients with active AIH as control cohort for the GLIPH analysis. The median number of distinct *TRB* and *IGH* clonotypes per sample was 2,350 and 1,388, respectively. The samples were sequenced with a median sequencing depth of 51 575 reads per sample for *TRB* NGS (total of 7.7 million reads) and 26 915 for *IGH* NGS (total of 1.5 million reads), overview of sequencing results is shown in **Supplementary Table 2**.

Broad Repertoire Metrics Analysis

Since clonal expansion and diversity represent important indicators for ongoing immune responses, we analyzed measures of T and B cell repertoire diversity and evenness in our AIC cases—before and on immunomodulatory treatment—as well as in age-matched healthy and CLL control cases. We generated diversity profile curves as described in the methods section to show the whole diversity spectrum of the repertoires. Generally, high diversity samples present higher y-values for each given alpha. We found that patients with primary AIC and healthy control cases had a more diverse PB *TRB* repertoire compared to patients with secondary AIC or the respective CLL control cases. This was reflected by a higher repertoire richness (at alpha = 0) as well as diversity profiles which overall ran at a higher level in primary AIC and healthy controls (**Figure 1A**). Moreover, the relatively balanced distribution of small, medium, large and hyperexpanded clones within the T cell clonal space of primary AIC and healthy controls confirmed the significantly more even repertoires in these individuals compared to secondary AIC and CLL control cases (**Figures 1B,C**). Overall, the *TRB* repertoire shifts mostly reflected a lymphoma bias and did not appear to be associated with autoimmunity (**Figures 1A–C**).

IGH repertoire analysis was only feasible for patients with primary AIC and healthy control cases since the repertoires of the secondary AIC and CLL cohorts were largely biased toward the malignant B cell clone. Mean PB *IGH* repertoire diversity of patients with primary AIC was numerically higher than in healthy control cases, however not significantly (**Figure 1D**). The B cell clonal space distribution and evenness as well as CDR3 length distribution and hydrophobicity properties of primary AIC were comparable to healthy controls (**Figures 1E,F** and **Supplementary Figure 1**).

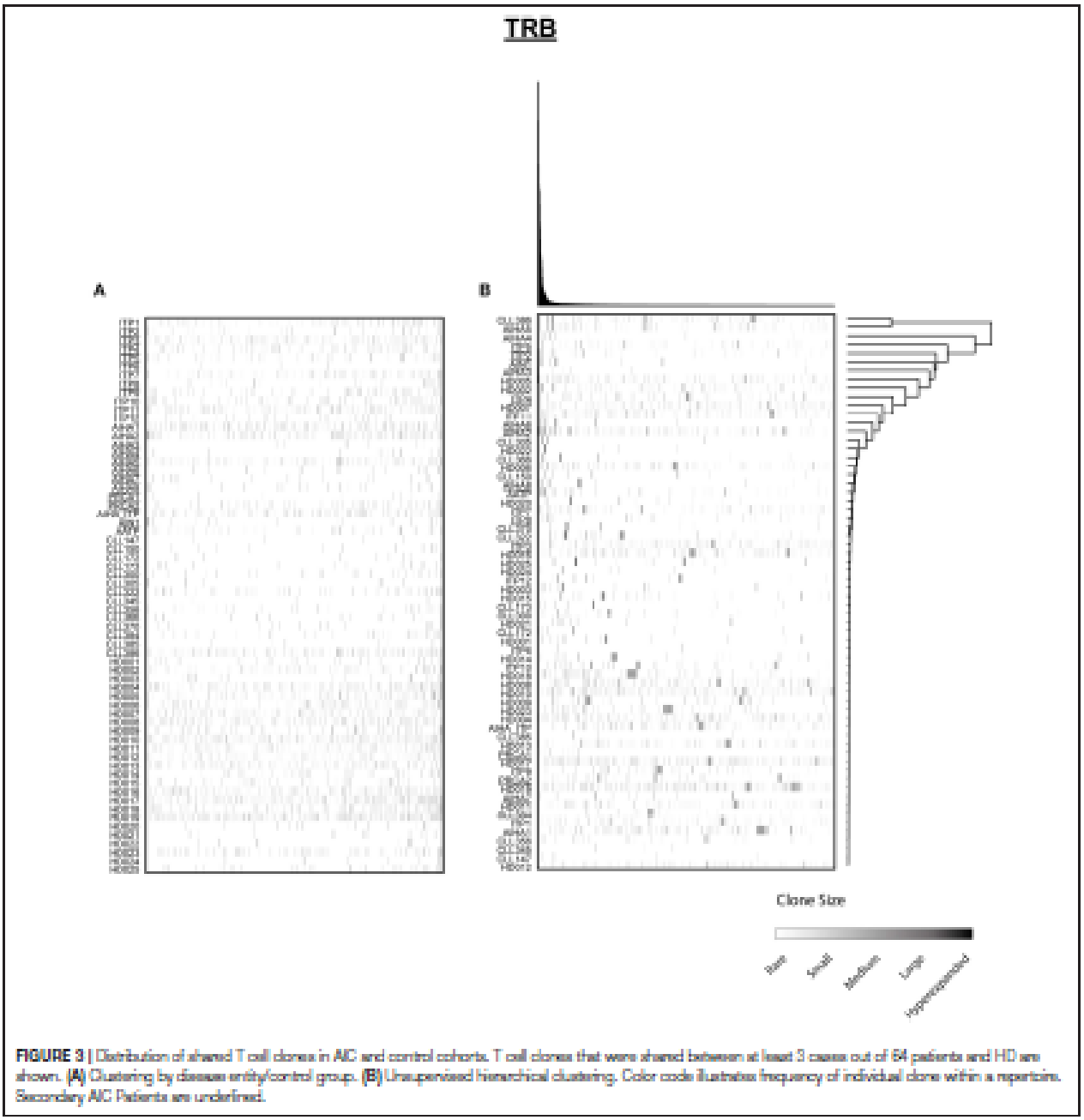


FIGURE 3 | Distribution of shared T cell clones in AIC and control cohorts. T cell clones that were shared between at least 3 cases out of 64 patients and HD are shown. **(A)** Clustering by disease entity/control group. **(B)** Unsupervised hierarchical clustering. Color code illustrates frequency of individual clone within a repertoire. Secondary AIC Patients are underlined.

VJ Usage of Peripheral Blood T and B Cells in Patients With AIC

While global repertoire metrics appeared to point rather toward lymphoma-biased repertoire differences between patients with primary or secondary AIC compared to respective control cohorts, we set out to screen all immune repertoires for the presence of individual expanded T and B cell receptor configurations present in patients with active AIC. Therefore, the combinatorial VJ gene usage of the T and B cell repertoire for patients with primary and secondary AIC and respective control groups were subjected to principal component analyses (**Figure 2**). Primary AIC cases clustered together with the healthy cohort, suggesting no specific *TRB VJ* usage signature (**Figure 2A**). TRB VJ gene combinations were clearly biased in patients with lymphoma vs. all other cohorts (**Figures 2B–D**). Some of the major biases in repertoire distributions of lymphoma patients consisted in a preferential usage of *TRBV7*, *TRBV6* and *TRBV12-3* genes (**Figures 2E,F**). Also *IGH VJ* usage did not

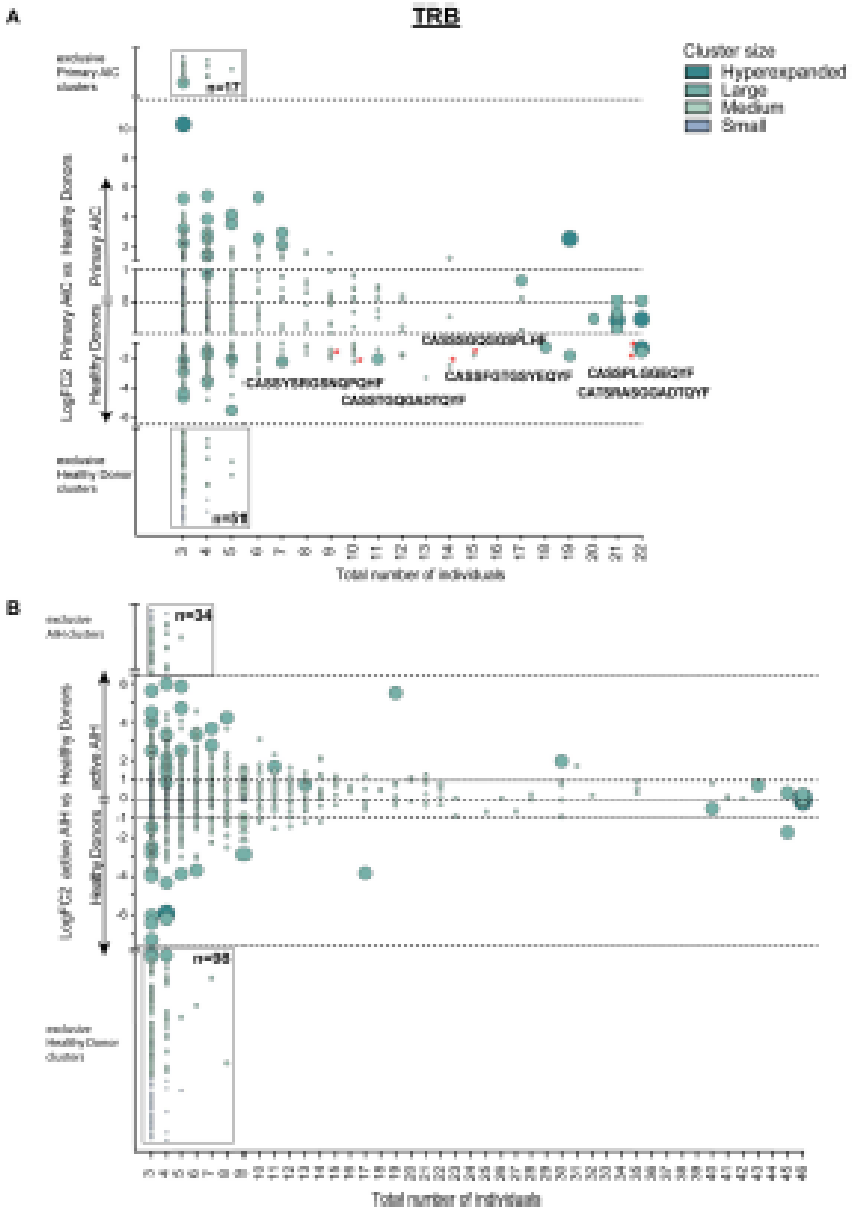


FIGURE 4 | Cluster analysis of autoimmune *TRB* repertoires. Logarithmic fold-change of clusters that were present in at least 3 individuals in **(A)** patients with primary AIC vs. HD and **(B)** patients with active AIH vs. HD. Cluster size and color represents mean frequency of CDR3 clones that constitute the respective cluster. Cluster sizes are defined as hyperexpanded ($0.01 < x \leq 1$), large ($0.001 < x \leq 0.01$), medium ($1E\text{-}04 < x \leq 0.001$), small ($1E\text{-}05 < x \leq 1E\text{-}04$). Significantly different abundant cluster were marked with *and their respective cluster sequences. Statistical test: Wilcoxon-rank test. *$p \leq 0.05$.

reveal any specificities of primary AIC (Figure 2G). Secondary cases could not be analyzed due to the dominance of the malignant *IGH* clone in these repertoires.

Clustering of T Cells With Common Antigen Recognition

Next, we mined all repertoires for shared T cell receptor sequences characteristic for AIC in general or specific AIC subgroups. Out of the 64 patient or control cases, we identified a total number of 1066 CDR3 amino acid sequences shared between at least three individuals (Figure 3A). Many of these clones were shared by different cohorts and some of these represented known public clones that had been previously annotated in public databases (Sequences in **Supplementary Table 3**). As expected, unsupervised hierarchical

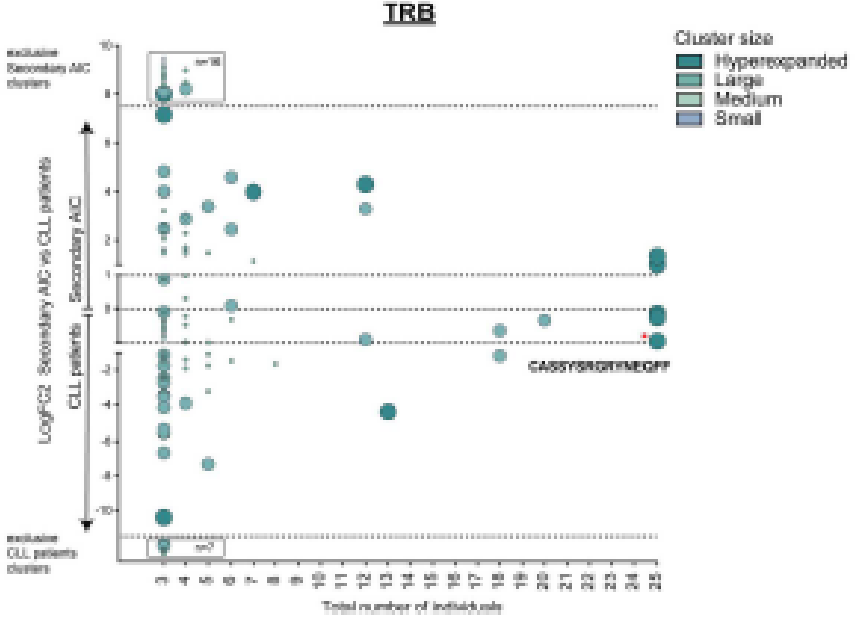


FIGURE 5 | Cluster analysis of lymphoma-associated *TRB* repertoires. Logarithmic fold-change of clusters that were present in at least 3 individuals in patients with secondary AIC vs. patients with CLL. Cluster size and color represents mean frequency of CDR3 clones that constitute the respective cluster. Cluster sizes are defined as hyperexpanded ($0.01 < x \leq 1$), large ($0.001 < x \leq 0.01$), medium ($1E\text{-}04 < x \leq 0.001$), small ($1E\text{-}05 < x \leq 1E\text{-}04$). Significantly different abundant cluster were marked with *and their respective cluster sequence. Statistical test: Wilcoxon-rank test. *$p \leq 0.05$.

clustering of the shared clonotypes did not reveal any biases (**Figure 3B**).

Since T cell immune reactions toward antigens are unique and polyclonal we clustered TCRβ clones based on target recognition using GLIPH (grouping of lymphocyte interactions by paratope hotspots) to identify disease-specific clusters involved in the pathogenesis of AIC. Primary AIC and HDs were analyzed together, and 657 clusters present in at least three individuals were found (**Figure 4A**). We identified clusters that were significantly overrepresented in HDs indicating that these were physiological *TRB* clusters (**Figure 4A**, marked with red asterisks). The numbers of clusters exclusively found in AIC or HD were 17 vs. 51, respectively, suggesting that patients with an autoimmune condition lack physiological T cell clusters compared to HDs (**Figure 4A** black-rimmed, Sequences in **Supplementary Table 4**).

To confirm our finding, we performed the T cell clustering analysis on an independent group of patients with active autoimmune hepatitis (AIH) using the T cell repertoires of 23 age- and sex-matched HDs as a reference. The analysis revealed 919 clusters that were present in at least 3 individuals (**Figure 4B**). Strikingly, and analogous to the analysis in primary AIC, HD presented almost a 3-fold higher number of exclusive *TRB* clusters than AIH (**Figure 4B**, black-rimmed, Sequences in **Supplementary Table 4**).

This signature of "physiological *TRB* cluster lack" was not seen when comparing patients with secondary AIC to patients with CLL (**Figure 5**), indicating a high similarity of these repertoires that was likely driven by the underlying lymphoma.

DISCUSSION

Autoimmune cytopenias (AIC) are driven by dysregulated immune interactions resulting in the production of blood cell and/or precursor cell directed autoantibodies and T cells. Pathophysiologically these conditions are complex and it remains unclear to what extent certain immune abnormalities are causative in disease pathogenesis or epiphenomena in the inflammatory processes. Also, initiating and perpetuating factors in these diseases are insufficiently understood to date.

Immunosequencing has opened up new avenues for in-depth analysis of immune repertoires in autoimmune conditions (39). We used this technology to study peripheral blood T and B cell repertoires in a cohort of 25 patients with active AIC and in the course of treatment. We found that baseline *IGH* and *TRB* immune metrics such as diversity measures as well as *VJ* gene usage are not informative in discriminating cases with primary AIC from healthy controls or patients with secondary AIC from CLL patients (40, 41). Yet, in individual patients that responded to treatment, we observed increases in repertoire diversity and evenness reflecting the replacement of preexisting by novel small clones upon immunomodulation (28, 42). As expected, we found a clear lymphoma signature consisting in a highly contracted T cell repertoire and preferential *TRB VJ* gene usage in secondary AIC and CLL controls compatible with previous reports (43).

The clustering of our *TRB* data into groups of T cell clones with presumably identical antigen recognition revealed a significant lack or underrepresentation of several *TRB* clusters found physiologically in healthy individuals compared to cases with AIC. This suggests that the lack of certain T cell clusters—potentially of tolerogenic nature–could be involved in the pathogenesis of these autoimmune disorders. The functional properties of these T cell clusters, however, need to be confirmed in future experiments in order to shed further light on the pathophysiology of AIC or other autoimmune conditions. Interestingly, only patients with primary AIC could be differentiated from healthy controls by use of the clustering algorithm while the secondary cases were overall very similar to the lymphoma cohort. This suggested that the strong lymphoma-associated signature superimposed potential AIC- or autoimmunity-specific changes.

Since our immune signature of AIC was very consistent across the different AIC subentities, we reasoned that this immune configuration should be permissive for the development of immunologically-mediated cytopenias in general, without being lineage-specific. Also clinically, patients may develop combined ITP and AIHA (Evan's syndrome), suggesting common immunological grounds that may facilitate the development and persistence of autoreactive immune cell clones directed against one or more blood lineages or even blood precursor cells. Autoimmunity in general is assumed to arise when mechanisms conferring immune tolerance break down (44) and the paradox association of autoimmune disorders with immune deficiency and infections (45) implies underlying common factors or immune signatures, that contribute to alterations in immune response that could result in completely different autoimmune conditions. With this in mind, we tested if the lack of physiological T cell clusters holds true for other autoimmune conditions as well, where the immunological attack is directed against non-blood cell antigens. Intriguingly, we confirmed our findings in patients with AIH, an autoimmune condition with an often chronic T cell directed immune response against hepatic antigens. This finding corroborates the idea of an autoimmune specific T cell signature spanning disease entities.

Taken together, our data suggests an immunological signature of autoimmunity that may be derived from a simple blood test using next-generation immunosequencing. Our *in-silico* approach needs further corroboration especially to fully understand the functional role of the clusters. As a future perspective we propose that this signature should be prospectively studied in clinical trials.

DATA AVAILABILITY

The datasets generated for this study can be found in the European Nucleotide Archive (ENA). http://www.ebi.ac.uk/ena/data/view/PRJEB33806.

AUTHOR CONTRIBUTIONS

MB, DS, and SS designed study and wrote manuscript. DS, SS, CS, BG, and NA performed DNA extraction and PCR amplification. NB and LF performed GLIPH analyses. DS and CC wrote R scripts for data analysis. MB, DS, and AL analyzed and interpreted the data. NA and BG optimized PCR amplification. JS collected and processed blood samples. AL and BT provided samples. All authors critically reviewed and approved the manuscript.

FUNDING

This work was supported by the Deutsche Forschungsgemeinschaft (grant BI 1711-4-1 to MB) and the SFB 841 (project no. 80750187 to MB).

ACKNOWLEDGMENTS

We would like to thank Malik Alawi and Sebastian Kral from the Bioinformatics Core Facility, UKE Hamburg, for their support with data processing. Moreover, we thank Ann-Kathrin Ozga for consultations on statistics. We like to acknowledge Prof. Zeller and Prof. Addo for the allocation of healthy donor genomic DNA samples used in this study.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2019.01897/full#supplementary-material

Supplementary Figure 1 | Physio-chemical properties of *IGH* repertoire in primary AIC and Healthys. Mean length distribution [+ standard deviation] **(A)** and Grand Average of hydropathy Index (GRAVY) of IGH CDR3 sequences in primary AIC with active disease and healthy controls. Statistical tests: Multiple *t*-tests **(A)** and Kolmogorov–Smirnov test **(B)**.

Supplementary Table 1 | CLL clonotypes of secondary AIC patients and CLL control patients.

Supplementary Table 2 | Overview of sequencing results.

Supplementary Table 3 | Shared CDR3 amino acid sequences.

Supplementary Table 4 | Exclusive TRB clusters of healthy donors.

REFERENCES

1. Audia S, Mahevas M, Samson M, Godeau B, Bonnotte B. Pathogenesis of immune thrombocytopenia. *Autoimmun Rev.* (2017) 16:620–32. doi: 10.1016/j.autrev.2017.04.012
2. Youinou P, Jamin C, Le Pottier L, Renaudineau Y, Hillion S, Pers J, et al. Diagnostic criteria for autoimmune neutropenia. *Autoimmun Rev.* (2014) 13:574–6. doi: 10.1016/j.autrev.2014.01.001
3. Barcellini W, Fattizzo B, Zaninoni A, Radice T, Nichele I, Di Bona E, et al. Clinical heterogeneity and predictors of outcome in primary autoimmune hemolytic anemia: a GIMEMA study of 308 patients. *Blood.* (2014) 124:2930–6. doi: 10.1182/blood-2014-06-583021
4. Rogers KA, Wojcik JA. Secondary autoimmune cytopenias in chronic lymphocytic leukemia. *Semin Oncol.* (2016) 43:300–10. doi: 10.1053/j.seminoncol.2016.02.011

5. Visco C, Barcellini W, Maura F, Neri A, Corticelezzi A, Rodeghiero F. Autoimmune cytopenias in chronic lymphocytic leukemia. *Am J Hematol.* (2014) 89:1055–62. doi: 10.1002/ajh.23785
6. Moreno C, Hodgson K, Ferrer G, Elena M, Filella X, Pereira A, et al. Autoimmune cytopenia in chronic lymphocytic leukemia: prevalence, clinical associations, and prognostic significance. *Blood.* (2010) 116:4771–6. doi: 10.1182/blood-2010-05-286500
7. Visco C, Giaretta I, Ruggeri M, Madeo D, Tonetto A, Rodeghiero F. Un-mutated IgVH in chronic lymphocytic leukemia is associated with a higher risk of immune thrombocytopenia. *Leukemia.* (2007) 21:1092–3. doi: 10.1038/sj.leu.2404582
8. Visco C, Nowella E, Peotta E, Paolini R, Giaretta I, Rodeghiero F. Autoimmune hemolytic anemia in patients with chronic lymphocytic leukemia is associated with IgVH status. *Haematol-Hematol J.* (2010) 95:1230–2. doi: 10.3324/haematol.2010.022079
9. Visco C, Maura F, Tuana G, Agnelli L, Lionetti M, Fabris S, et al. Immune thrombocytopenia in patients with chronic lymphocytic leukemia is associated with stereotyped B-cell receptors. *Clin Cancer Res.* (2012) 18:1870–8. doi: 10.1158/1078-0432.CCR-11-3019
10. Maura F, Visco C, Falini E, Reda G, Fabris S, Agnelli L, et al. B-cell receptor configuration and adverse cytogenetics are associated with autoimmune hemolytic anemia in chronic lymphocytic leukemia. *Am J Hematol.* (2013) 88:32–6. doi: 10.1002/ajh.23342
11. Toyokawa Y, Kingetsu I, Yasuda C, Yasuda I, Yoshida K, Kuroiwa D, et al. A case of pure red cell aplasia complicated by Evans syndrome. *Mod Rheumatol.* (2007) 17:333–7. doi: 10.1007/s10165-007-0584-9
12. Sekiguchi Y, Shimada A, Imai H, Wakabayashi M, Sugimoto K, Nakamura N, et al. A case of recurrent autoimmune hemolytic anemia during remission associated with acute pure red cell aplasia and hemophagocytic syndrome due to human parvovirus B19 infection successfully treated by steroid pulse therapy with a review of the literature. *Int J Clin Exp Pathol.* (2014) 7:2624–35.
13. Zufferey A, Kapur R, Semple JW. Pathogenesis and therapeutic mechanisms in immune thrombocytopenia (ITP). *J Clin Med.* (2017) 6:E16. doi: 10.3390/jcm6020016
14. Conley CL, Lippman SM, Ness PM, Petz LD, Branch DR, Gallagher MT. Autoimmune hemolytic anemia with reticulocytopenia and erythroid marrow. *N Engl J Med.* (1982) 306:281–6. doi: 10.1056/NEJM198202043060507
15. Saha M, Ray S, Kundu S, Chakrabarti P. Pure red cell aplasia following autoimmune hemolytic anemia: an enigma. *J Postgrad Med.* (2013) 59:51–3. doi: 10.4103/0022-3859.109495
16. Brown GE, Babiker HM, Cantu CL, Yeager AM, Krishnadasan R. "Almost Bleeding to Death" The conundrum of acquired amegakaryocytic thrombocytopenia. *J Case Rep Hematol.* (2014) 2014:806541. doi: 10.1155/2014/806541
17. Nixon CP, Sweeney JD. Autoimmune cytopenias: diagnosis & management. *R I Med J.* (2013) 99:36–40.
18. Arbach O, Funck R, Seibt F, Salama A. Erythropoietin may improve anemia in patients with autoimmune hemolytic anemia associated with reticulocytopenia. *Transfus Med Hemother.* (2012) 39:221–3. doi: 10.1159/000339260
19. Bowen DA, Call TG, Shanafelt TD, Kay NE, Schwager SM, Reinalda MS, et al. Treatment of autoimmune cytopenia complicating progressive chronic lymphocytic leukemia/small lymphocytic lymphoma with rituximab, cyclophosphamide, vincristine, and prednisone. *Leuk Lymphoma.* (2010) 51:620–7. doi: 10.3109/10428191003682767
20. Ekbote M, Marsh J. Autoimmune cytopenias. *Medicine.* (2009) 37:159–63. doi: 10.1016/j.mpmed.2008.12.008
21. Barcellini W. New insights in the pathogenesis of autoimmune hemolytic anemia. *Transfus Med Hemother.* (2015) 42:287–93. doi: 10.1159/000439002
22. Olsson B, Andersson PO, Jernas M, Jacobsson S, Carlsson B, Carlsson LM, et al. T-cell-mediated cytotoxicity toward platelets in chronic idiopathic thrombocytopenic purpura. *Nat Med.* (2003) 9:1123–4. doi: 10.1038/nm921
23. Fagiolo E. Immunological tolerance: loss vs. erythrocyte self antigens and cytokine network disregulation in autoimmune hemolytic anaemia. *Autoimmun Rev.* (2004) 3:53–9. doi: 10.1016/S1568-9972(03)00085-5
24. van Dongen JJ, Langerak AW, Brüggemann M, Evans PA, Hummel M, Lavender FL, et al. Design and standardization of PCR primers and protocols for detection of clonal immunoglobulin and T-cell receptor gene recombinations in suspect lymphoproliferations: report of the BIOMED-2 Concerted Action BMH4-CT98-3936. *Leukemia.* (2003) 17:2257–317. doi: 10.1038/sj.leu.2403202
25. Oberle A, Brandt A, Voigtlaender M, Thiele B, Radloff J, Schulenkorf A, et al. Monitoring multiple myeloma by next-generation sequencing of V(D)J rearrangements from circulating myeloma cells and cell-free myeloma DNA. *Haematologica.* (2017) 102:1105–11. doi: 10.3324/haematol.2016.161414
26. Rossi A, Voigtlaender M, Ianjetovic S, Thiele B, Alawi M, Marx M, et al. Mutational landscape reflects the biological continuum of plasma cell dyscrasias. *Blood Cancer J.* (2017) 7:e537. doi: 10.1038/bcj.2017.19
27. Bolotin DA, Poslavsky S, Mitrophanov I, Shugay M, Mamedov IZ, Putintseva EV, et al. MiXCR: software for comprehensive adaptive immunity profiling. *Nat Methods.* (2015) 12:380–1. doi: 10.1038/nmeth.3364
28. Schliffke S, Akyuz N, Ford CT, Mahrle T, Thenhausen T, Krohn-Grimberghe A, et al. Clinical response to ibrutinib is accompanied by normalization of the T-cell environment in CLL-related autoimmune cytopenia. *Leukemia.* (2016) 30:2232–34. doi: 10.1038/leu.2016.157
29. Lefranc MP. IMGT, the international ImMunoGeneTics information system. *Cold Spring Harbor Prot.* (2011) 2011:595–603. doi: 10.1101/pdb.top115
30. Team RC. *R: A Language and Environment for Statistical Computing.* Vienna: R Foundation for Statistical Computing (2018).
31. Wickham H. *tidyverse: Easily Install and Load the 'Tidyverse'.* R package version 1.2.1 (2017). Available online at: https://tidyverse.tidyverse.org/
32. Nazarov VI, Pogorelyy MV, Komech EA, Zvyagin IV, Bolotin DA, Shugay M, et al. tcR: an R package for T cell receptor repertoire advanced data analysis. *BMC Bioinform.* (2015) 16:175. doi: 10.1186/s12859-015-0613-1
33. Dray S, Dufour AB. The ade4 package: implementing the duality diagram for ecologists. *J Stat Soft.* (2007). 22:1–20. doi: 10.18637/jss.v022.i04
34. Greiff V, Bhat P, Cook SC, Menzel U, Kang W, Reddy ST. A bioinformatic framework for immune repertoire diversity profiling enables detection of immunological status. *Genome Med.* (2015) 7:49. doi: 10.1186/s13073-015-0169-8
35. Simpson EH. Measurement of diversity. *Nature.* (1949) 163:688. doi: 10.1038/163688a0
36. Kyte J, Doolittle RF. A simple method for displaying the hydropathic character of a protein. *J Mol Biol.* (1982) 157:105–32. doi: 10.1016/0022-2836(82)90515-0
37. Margreitter C, Lu HC, Townsend C, Stewart A, Dunn-Walters D, Fraternali F. IRepertoire: a user-friendly web server for analysing antibody repertoire data. *Nucleic Acids Res.* (2018) 46:W264–70. doi: 10.1093/nar/gky276
38. Glanville J, Huang H, Nau A, Hatton O, Wagar LE, Rubelt F, et al. Identifying specificity groups in the T cell receptor repertoire. *Nature.* (2017) 547:94–8. doi: 10.1038/nature22976
39. Woodsworth DJ, Castellarin M, Holt RA. Sequence analysis of T-cell repertoires in health and disease. *Genome Med.* (2013) 5:98. doi: 10.1186/gm502
40. Vardi A, Vlachonikola E, Karypidou M, Stalika E, Bikos V, Gemenetzi K, et al. Restrictions in the T-cell repertoire of chronic lymphocytic leukemia: high-throughput immunoprofiling supports selection by shared antigenic elements. *Leukemia.* (2017) 31:1555–61. doi: 10.1038/leu.2016.362
41. Vardi A, Agathangelidis A, Stalika E, Karypidou M, Siorenta A, Anagnostopoulos A, et al. Antigen selection shapes the T-cell repertoire in chronic lymphocytic leukemia. *Clin Cancer Res.* (2016) 22:167–74. doi: 10.1158/1078-0432.CCR-14-3017
42. Akyüz N, Brandt A, Stein A, Schliffke S, Mährle T, Quidde J, et al. T-cell diversification reflects antigen selection in the blood of patients on immune checkpoint inhibition and may be exploited as liquid biopsy biomarker. *Int J Cancer.* (2017) 140: 2535–44. doi: 10.1002/ijc.30549
43. Vardi A, Agathangelidis A, Stalika E, Marincevic M, Karypidou M, Anagnostopoulos A, et al. T cell receptor gene repertoire restriction in chronic lymphocytic leukemia with stereotyped IGHV4-34/IGKV2-30 antigen receptors. *Blood.* (2012) 120:3908.
44. Goodnow CC. Multistep pathogenesis of autoimmune disease. *Cell.* (2007) 130:25–35. doi: 10.1016/j.cell.2007.06.033
45. Feuille EJ, Anoushiravani N, Sullivan KE, Fulcihan RL, Cunningham-Rundles C. Autoimmune cytopenias and associated conditions in CVID: a report from the USIDNET registry. *J Clin Immunol.* (2018) 38:28–34. doi: 10.1007/s10875-017-0456-9

Conflict of Interest Statement: NE and LF are employed by company ENPICOM B.V.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

CONTRIBUTIONS TO THIS WORK

The manuscript “High-Throughput Immunogenetics Reveals a Lack of Physiological T Cell Clusters in Patients with Autoimmune Cytopenias” was accepted for publication in August 2019 in *Frontiers of Immunology*. Mascha Binder and Simon Schliffke initiated this project based on their former work where they studied two patients with chronic lymphocytic leukemia (CLL) and secondary autoimmune cytopenia (AIC) from this cohort and showed a normalization of the peripheral T cell clonal space upon ibrutinib treatment. Therefore, we aimed to shed light into the immune repertoire profile of patients with AIC using NGS in order to understand if autoimmune biases in the TCR repertoire of these patients were present and if these can be exploited as blood-based biomarkers.

I organized all patient data, extracted DNA from frozen PBMNCs and performed the PCRs as well as all necessary quality measurements and prepared the libraries for NGS.

Moreover, I was responsible for data analyses of the NGS data, which included the conversion of raw NGS data (FASTQ-files) to human-readable text-files using our in-house bioinformatics pipeline. Furthermore, I conducted different analyses of the immune repertoire using R Studio. Through literature review, I came across other meaningful analysis modules, like the alpha-parameterized diversity curves as well as the VJ principal-component analysis and established these R scripts for this project.

In close collaboration with my supervisor Mascha Binder I interpreted the results, finalized the figures and tables for publication. Together, we wrote the first draft and revised the manuscript until final publication.

Prof. Dr. med. Mascha Binder

3 DISCUSSION

3.1 FRONTIERS OF T CELL REPERTOIRE PROFILING: FROM SANGER SEQUENCING TO NGS AND STATE-OF-THE-ART BIOINFORMATICS

T cells recognize antigens presented on MHC molecules via their antigen receptor, also named T cell receptor (TCR). The TCR consists of two polypeptide chains (αβ or γδ) and is encoded by the recombination of V(D)J gene segments on the respective loci. This process results in a highly diverse repertoire of TCRs, whereby each possesses a unique antigen binding profile. The large diversity of the repertoire offers recognition of, and therefore immunological protection against, a wide range of antigens for the host. Once T cells encounter their cognate antigen they proliferate and form a clone of cells expressing identical receptors. After successful clearance of the antigen, the majority of cells undergo apoptosis apart from a fraction of cells that form the immunological memory. This dynamic process of clonal expansion, migration and apoptosis indicates that the clonal composition of the immune repertoire is ever evolving and reflects the current state as well as the antigen exposure history of an individual [195].

Before high-throughput technologies were available, immune receptors were analyzed using Sanger Sequencing. Therefore, each sequence had to be cloned into a plasmid vector and one reaction generated one sequence read in the machine. With this low throughput it was practically impossible to grasp the large diversity and clonal architecture of the

immune receptor repertoire [190]. Using Sanger sequencing, the study of hundreds of immune receptor sequences presented the higher end of throughput that was possible [191]. Spectratyping or Immunoscope were other methods that were used to study immune repertoire diversity [192]. Using V and J gene specific primer pairs the CDR3 length of TCRs was analyzed after gel electrophoresis separation. These approaches are less time consuming and costly than Sanger sequencing, however, do not provide information on the nucleotide sequence. Moreover, flow cytometry can be used on protein level to analyze the clonal composition of a repertoire [193, 194]. Here TCR Vβ and Vα-specific monoclonal antibodies are used either individually or in a multiplex format to characterize the repertoire composition. Again, this approach lacks nucleotide information and does not precisely depict the clonal architecture of the T cell repertoire.

Despite these technological limitations, a number of studies performed in the pre-NGS era allowed meaningful repertoire analysis and a few examples of such research deserve to be mentioned: This research aimed for example at understanding the origin of T_{regs} in the tumor environment. By combining different methods, like a transgenic mouse model (TCR^{mini}, with a limited size of rearranged TCR repertoire) with spectratyping and Sanger sequencing, it could be shown that T_{regs} in the tumor environment are rather of thymic origin than peripherally induced and are even more diverse than the naïve T cell pool [191, 195]. Moreover, research reports started to accumulate around biased T cell repertoires (skewed usage of certain V(D)J combinations) and public clonotypes shared by several individuals who were exposed to the same antigen, e.g. during infection with influenza A virus [94], Epstein-Barr virus (EBV) [196] or human cytomegalovirus (CMV) [197]. Additionally, biases in the T and B cell repertoire V gene usage in malignancy were reported [198-201] and closely homologous CDR3 amino acid sequences, termed stereotyped receptors, in CLL were identified and correlated with clinical outcome [202]. In the field of autoimmunity,

immune receptor analyses revealed shared TCR alpha and beta chains in systemic lupus as well as TCR biases in multiple sclerosis [203-205].

The emergence of NGS greatly facilitated the study of the AIRR since it provided all of the repertoire relevant information at once, at high-throughput and low cost. The Roche 454 system was the earliest NGS platform and could generate 500 bp reads thereby covering the entire recombined V(D)J region. The zebrafish antibody repertoire was the first to be sequenced at high-throughput with the Roche 454 in 2009 [206]. This represented a turning point in the study of immune repertoires. Since then, research has focused on characterizing immune repertoires and thereby deepening our understanding of immune dynamics in health and disease. Besides assessing repertoire metrics like clonal richness and diversity, we can track clone members between tissues and across subsets of T cells. Moreover, the longitudinal tracking of clones - e.g. for detection of minimal residual disease (MDR) during treatment of lymphoid malignancy - has proven useful and much more sensitive than traditional techniques [207].

One aspect of AIRR seq, with special relevance for TCR repertoires, is the characterization of so called public clonotypes. The concept of public clonotypes has emerged in a time when sequencing was restricted to public clonotypes that share the same amino acid sequence of their CDR3, however, without necessarily using the same V(D)J recombination. Sharing of amino acid CDR3 sequences in repertoires of different individuals has been suggested to be impossible due to the recombinatorial breadth of the TCR repertoire and the limited number of T cells generated per individual. However, it is now well accepted that public clonotypes make up approximately 10-20% of the human peripheral blood T cell repertoire enabling the generation of similar antigen-specific TCRs against ubiquitous pathogens in a population [100]. Their generation is a result of both recombinatorial biases

(certain V(D)J rearrangements are more likely to occur than others) and convergent recombination (different nucleotide sequences encode the same amino acid sequence) [97, 208]. Both mechanisms ultimately lead to higher generation probabilities (pGEN) for certain sequences to be found in the TCR repertoire than for others [100]. Beside the presence of public clonotypes in infections with common viruses as reported earlier, using NGS, several groups could show that public clonotypes preferentially expand in EBV [209], CMV [210, 211], hepatitis C virus [212] or influenza [213] infections implying an important role for these clonotypes during the immune response towards these viruses. NGS technology has largely fostered our understanding of these clones, yet many questions remain unanswered e.g. their contribution to other immune reactions such as tumor antigens or to vaccination.

While quite an extensive fraction of the repertoire is taken up by public clonotypes, the private (not intra-individually shared) clonotypes will likely contribute to the majority of immune responses in infection, cancer and autoimmunity. There is scientific consensus that – despite the private character of TCR sequences contributing to an immune response – there are overarching paratope features that may be used to group TCR sequences according to identical antigen recognition. The development of such algorithms has been instrumental to immune repertoire research in that it allows to reduce the complexity of the individual datasets to allow functional groups of TCRs to be studied. Several algorithms have been published that approach this issue [88, 89, 214, 215]. GLIPH (Grouping lymphocyte interaction by paratope hotspots) belongs to the first algorithms set out to group T cell clones by antigen specificity and as elaborated in the introduction, TCR sequences are clustered based on conserved motifs found within the CDR3 high-contact probability region and based on global similarity of the sequences. Using GLIPH, Yost et al [216] noted that clones belonging to one T cell cluster presented overlaps between $CD8^+$

phenotypes including cells with memory or activated states. A major finding of this study was the observation that while in most T cell clusters these phenotypic transitions were seen (even T cell clusters containing $CD4^+$ and $CD8^+$ T cells), there were no clusters containing pre-treatment exhausted and non-exhausted $CD8^+$ T cells, indicating that the phenotypic transition capacity is rather limited in exhausted T cells. Together, this suggested that during PD-1 immune checkpoint blockade in cancer treatment, it is not the T_{ex} that are "un-leashed" as it was assumed when immune checkpoint therapy was first introduced but rather new T cell clones that enter the tumor site that eliminate the malignant cells.

In the research thesis presented here, the aim was to apply 1) NGS technology combined with 2) state-of-the-art bioinformatics data analysis to understand the contributions this approach may have to the immunological understanding of complex immunological diseases – with autoimmune cytopenias chosen as one example for a complex group of diseases - as well as immunoaging in health and in cancer. In the following two paragraphs, I will discuss the major findings in the context of the respective literature.

3.2 THE AGING T CELL REPERTOIRE IN HEALTH AND CANCER

The improvements in life quality, i.e. availability of food, quality of housing, improved hygiene and improved medical care etc. all contributed to an increased life span in humans [217]. This in turn led, especially in western countries to progressive transformation of population demographics. Overall an increase of the aged population is notable, especially of individuals above 65 years of age who currently make up 15-20% of the population in Western Europe, a 2-fold increase during the last 60 years [218]. Although longevity seems

like a victory for the human race, it comes with a price, because the effectiveness of the immune system diminishes over time and a higher cancer incidence and mortality rates are evident in these age groups [219]. Moreover, autoimmune diseases as well as infectious disease pose an increased threat to this population group [220]. Therefore, healthy aging has become a public health priority, focusing research and medical efforts on understanding how the immune system changes with increasing age and optimizing vaccine success in the elderly [221].

To understand physiological changes of the T cell repertoire over the life span of a human, we analyzed the TCR repertoire of 95 healthy individuals spanning age groups from below 30 to above 70 years [222]. Our data indicate that the diversity of the peripheral TCR repertoire decreases with increasing age beginning at the fourth decade of life. Contracted TCR diversity is associated with decreased immune responsiveness observed in the elderly such as the reduced ability to fight new infections [223-226].

Interestingly, we observed that the contracted diversity was rather explained by a significant decrease in richness (number of unique clonotypes) than changes in evenness which was in line with findings of earlier studies [46, 222, 227-229].

From four donors also longitudinal samples collected over 2 to 6 years were available. The analysis of these samples demonstrated the presence of a “core-immune repertoire” that is unique to each person and is maintained over very long periods of time if no major intervention, like e.g. cytotoxic therapy, takes place. A recent paper demonstrates that different individuals and even identical twins can be discriminated by repertoire data from as little as 10.000 T cells (~10µL blood) with a false positive and false negative rate of $<10^{-6}$ [230]. Not only does this work corroborate the idea of a core-immune repertoire or an “immune fingerprint” which can be exploited to understand a person´s immune

history and the functional state of their repertoire. It also puts our handling (storage and sharing) of repertoire data up for debate since it clearly demonstrates how personal repertoire data can be.

Our data could also have implications for optimal timing of vaccination in the course of life. Vaccination regimes ensure protection against common infections which pose a major threat in the elderly [220]. Most vaccines that are currently in use are developed empirically, and we know very little about how they activate the immune system [231]. Antibody titers of antigen specific antibodies are used as surrogate markers for vaccine success. However, besides B cells that produce these antibodies, T cells play a major role during the generation of high-affinity antibodies as well as providing immunological memory [231]. In T-cell-inducing vaccines the magnitude as well as the breadth of T cell response are correlated with vaccine protection [232]. Therefore, besides the composition of antigens administered with the vaccine also the diversity of the host´s TCR repertoire needs to be considered when designing prophylactic vaccines [232]. Our data indicate decreases in TCR diversity starting from 40 years of age. Given that some of the vaccination regimes are indicated especially for elderly risk groups (e.g. pneumococcal or herpes zoster vaccination) and that vaccination outcomes are often poor in these cohorts, our and other people's findings on repertoire alterations across age groups could inspire the conception of earlier vaccination timepoints to induce strong and long-lived memory.

In addition to the data provided on immunoaging of healthy individuals, our project delivered valuable insight into repertoire restriction in patients with cancer. Tumor immunosurveillance is a well described theory that states that the immune system has the potential to control malignant cells and thereby prevent tumor outgrowth even over several years [233]. Although direct proof of this theory is difficult to obtain, several observations

support the immunological control of tumors. Moreover, it is well established that an impaired immune system is associated with increased cancer risk.

We analyzed TCR repertoires of more than 200 patients with hematological or solid cancers before and after treatment with chemotherapy and used the repertoire data of the healthy aging cohort as reference. Our data points towards pre-mature immunoaging in patients with cancer. Pre-treatment TCR repertoires of patients with cancer were less diverse and contained more expanded clonotypes responsible for an uneven distribution of clones within the repertoire as compared to age-matched healthy controls. Moreover, subcohort analyses on samples before and after chemotherapy (from patients matched for age, sex and tumor type or paired pre/post chemotherapy samples) revealed that the T cell repertoire diversity and clonality recovered almost completely to pretreatment levels. This indicated no long-term loss of T cell diversity after chemotherapy. The reconstitution of diversity can happen in two ways, which consist of peripheral homeostatic expansion of existing clones or the generation of new T cell clones. The former, would come with the price of increased clonality, since the homeostatic expansion would eventually lead to an uneven distribution of clones within the repertoire. To test how diversity was recovered, we assessed repertoire similarity between pre- and post-therapy samples. As opposed to healthy individuals who exhibited great overlaps (=high number of identical clones) in repertoires sampled over several months to years, the repertoires of patients after chemotherapy had little to no overlap with the repertoire before chemotherapy.

It remains elusive, if the new T cell clones we “see” by sequencing are newly generated by thymic output or rather represent rare clones of the peripheral repertoire or even tissue resident clones which re-entered circulation and were not captured by sequencing at the first time. A recent review on tissue resident memory T cells (T_{RM}) however argues that T_{RM}

lack markers and receptors to re-enter blood stream and are “permanently lodged” and committed to the tissue of residence [234], which would be in favor of the immune reconstitution by thymic output. Moreover, our findings of a maintained T cell repertoire diversity and evenness after chemotherapy are in line with findings from a recent study which also attributes T cell rebound to increased thymic activity [147]. By using mass cytometry in combination with T cell receptor excision circles (TRECs[1]) measurement and TCR sequencing the authors show in a cohort of patients with breast cancer that one year after adjuvant chemotherapy total lymphocyte counts as well as the compositional complexity of most major innate and adaptive immune cell subsets are recovered. Naïve $CD4^+$ and $CD8^+$ T cell subsets of patients exhibited an extensive TCR diversity even after chemotherapy as well as increased TRECs compared to age- and sex-matched healthy controls.

Taken together, it seems reasonable to assume that the post-chemotherapy repertoire constitutes newly generated clones from thymic output which is really promising especially in the era of new cancer treatments including immune stimulatory agents like immune checkpoint blockade (ICB).

The analyzes performed in our immunoaging project at this point only represent an arbitrary selection of investigations that can be performed on our valuable NGS resource of TCR repertoires of healthy individuals and patients with cancer of all ages. To allow other research groups with different focus to access our data, we deposited this together with our manuscript to be used by the scientific community to help characterize age-dependent changes of the TCR repertoire especially in the future when new algorithms and models will be available. However, in future studies multi-axis approaches are still warranted

[1] TRECs are found in newly generated T cells and are diluted out during clonal expansion

including phenotypic analyses of the different cell populations and matching sub-repertoire or even single cell analyzes.

3.3 THE TCR REPERTOIRE IN AUTOIMMUNE CYTOPENIAS

The second publication which arose from this work focuses on the characterization of immune repertoires of patients with autoimmune diseases arising as primary disease as well as a complication of an underlying hematological malignancy.

The high rate of refractory disease in AIC, despite the many therapeutic options, motivated us to study the immune repertoires of patients with primary and secondary AIC [235]. We hypothesized that underlying immune biases lead to the manifestation of the disorder even after initial successful treatment.

Interestingly, the T and B cell clonal architecture (i.e. richness, diversity and VJ-gene usage) of patients with primary AIC is comparable to age- and sex-matched healthy donors (HD). Moreover, we could not detect disease specific shared clonotypes between patients with AIC. In contrast, we recently published characteristic VJ-gene usage shifts in the T cell repertoire of patients with autoimmune hepatitis (AIH) [236], a chronic autoimmune disease of the liver with high relapse rates. Interestingly, these shifts persisted even after complete biochemical response to immunomodulatory treatment, indicating that T cell functionality, rather than the pathological T cell architecture is targeted by the treatment.

In order to reduce the complexity of the repertoires and identify groups of T cells with potentially common antigen recognition involved in the pathogenesis of AIC, we performed GLIPH on the peripheral TCR repertoire of primary AIC patients (n=11) and HD (n=11). In total we identified 657 T cell clusters that were present in at least three individuals. Six of

these clusters were significantly overrepresented in HDs, however no cluster was overrepresented in patients with AIC – indicating that disease specific T cell clones are not discernable based solely on their sequence. Moreover, we identified clusters that were exclusively present in HDs or patients with AIC, respectively, and noted an increased number of HD exclusive clusters (n=51) as opposed to AIC exclusive clusters (n=17). We could verify this finding in an independent cohort of patients with active AIH, where the number of HD exclusive clusters was also approximately 3-times higher than that of AIH exclusive clusters.

Since autoimmune diseases develop as a consequence of complex mechanisms including the breakdown of tolerance checkpoints, we interpreted our finding as a lack of physiological T cell clusters in patients with an autoimmune condition. These clusters could potentially be of tolerogenic nature since it is well established that the breakdown of peripheral tolerance, especially dysfunctional and decreased numbers of T_{regs} play a causative role in autoimmune diseases [237]. However, these interpretations remain speculative since our data does not provide any information on the cellular phenotype.

In the context of autoimmunity, it is worth discussing the degenerative or "promiscuous" nature of the TCR repertoire. TCR degeneracy permits efficient T cell responses of a single T cell clone to a large number of different peptides presented on MHCs but also of many different T cell clones to a single pMHC [238]. These features enable a TCR repertoire to recognize any potential pathogen, however, still providing a high enough specificity to respond selectively to non-self and presenting minimal cross-reaction with self-derived peptides (since only a minority of the population suffers from AID).

Moreover, the majority of clones bearing high-affinity TCRs towards self-antigen undergo negative selection in the thymus (see section T CELL DEVELOPMENT). Consequently, T cell responses against self-antigens as seen in systemic autoimmunity or cancer have to be

enriched for low-affinity TCRs. This together with the natural degeneracy of T cells, would imply that T cell responses towards self are highly heterogeneous, potentially even lacking global and/or local similarity and thereby representing special challenges for algorithms like GLIPH.

Taken together, the analysis of the TCR repertoire can provide insights into the pathobiology of autoimmune diseases. However, it seems that identifying groups of self-antigen directed T cells from bulk sequencing currently represents a challenge which needs further elucidation.

3.4 T CELL REPERTOIRE PROFILING: OVERCOMING LIMITATIONS AND OUTLOOK

Despite the great opportunities of NGS T cell repertoire profiling, several limitations do remain. Some of which are essential in our view to compare different studies and also to draw functional clues from these massive amounts of data.

Our own analyses, including repertoire data acquired following at least two different library amplification protocols and our cross-study comparisons, show that one current essential limitation is the standardization of protocols. There is a big need for standardization of AIRR seq, since as of today, a myriad of preparation protocols for sequencing libraries are in use and data analysis can be performed with many different tools, which were shown to produce non-consistent results [239]. The latter was mainly explained by the different reference gene libraries that are currently available. Indeed, the Germline Database

Working Group of the AIRR Community works on the development of complete and accurate sets of reference germline genes for TCR as well as Immunoglobulins.

Since our analyses presented here have been confined to the analysis of blood, I will discuss the major aspects relevant to the standardization needs for this specific immune compartment. Peripheral blood samples are easily accessible and not invasive and therefore widely used for immunosequencing. Moreover, their potential to reflect tissue immune architecture (e.g. tumor) and response dynamics has been demonstrated [240]. Several confounding factors of this method need to be considered during sample preparation, sequencing and analysis of immune repertoire data:

3.4.1 SAMPLING ERROR

As described above the entire human T cell repertoire is estimated to consist of 10^8 to 10^{10} different T cell clones. Therefore, a small amount of blood or tissue from which a portion of DNA or RNA is used for repertoire sequencing represents only a fraction of the real diversity. Indeed, when assessing the repertoire composition of the same sample repeatedly, approximately 25% of the clones do not overlap [240]. This indicates the huge diversity of T cell clones in the sample. Repertoire metrics like richness and diversity provide information on the clonal architecture of the sample and are rather stable measures also in repeated sampling. However, they depend directly on the number of lymphocytes sampled as well as read depth applied to the sample [240]. Therefore, these factors need to be accounted for when comparing repertoire metrics of different samples and/or individuals.

3.4.2 READ DEPTH

To ensure that the real clonal diversity of the sample is captured, several quality standards need to be established.

The optimal read depth can be estimated as follows: If 500 ng of peripheral blood mononuclear cells (PBMNCs) DNA is used for AIRR seq, which corresponds to 75 000 cells [241], a minimal read depth of 116 000 reads would be required to pass quality control. This assumes that PBMNC are composed of 70-90% lymphocytes and of those approximately 70-85% are T cells [242]. Hence, 500 ng PBMNC DNA contain approximately 37 000 – 58 000 T cell genomes. At a minimal coverage of 2 reads per genome, at least 116 000 reads are required to ensure coverage of every T cell genome in the sample.

Additional flow cytometry analyses can provide more precise information on the composition of PBMNCs and hence the number of T cells in the sample.

3.4.3 ANALYSIS OF AIRR SEQ DATA

Once the sample has passed quality control for read depth, the raw data are passed through bioinformatic pipelines which annotate the sequences with the corresponding V, D and J genes. Currently multiple pipelines are available to perform this task, however, a recent paper which compared three widely used pipelines found that only 40% of reference germline sequences were shared by the tools [239]. Subsequently, results generated with different tools are not comparable. To tackle these discrepancies and standardize the generation, analysis and sharing of immune repertoire data, a community around AIRR data (AIRR Community) has evolved from 2014 and develops white papers on these topics (e.g. MiAIRR metadata standards on AIRR data submission).

Another feature which can be helpful to assess quality of the data is the proportion of sequences in the repertoire with a high generation probability (pGEN). Using mathematical models, Elhanati and colleagues estimated the degree of publicness of TCR sequences

and found that in humans as well as mice 10-20% of the T cell repertoire represent clones with a pGEN larger than 1/number of total T cell clones in the body [100].

In addition to the undeniable need for more standardization, a second major limitation of TCR profiling by NGS is the (current) unfeasibility to derive the cognate antigen from bulk receptor sequencing data. However, this information is key to most research projects addressing immunological questions, especially in the field of autoimmunity and cancer. The fact that antigen recognition strongly depends on genetic background, environmental factors as well as the antigen presenting machinery complicates this problem even more. The GLIPH algorithm addresses this issue by clustering T cells with overlapping specificities based on sequence motifs and structural considerations derived from known receptor-antigen interactions [88, 89].

However, grouping T cells with similar antigen recognition is only an attempt of trying to make sense of the antigen or epitope which is being recognized by the receptor. Especially in autoimmunity this would help to decipher the biology of the disorder. What in the hosts body is being presented in a way that their own immune system is not able to distinguish it as self? Knowing the target antigen could lead to biomarker discovery which will greatly facilitate diagnosis and treatment response evaluation in AID.

In cancer, on the other hand, the identification of tumor specific antigens can pave the way to new clinical treatments, that will specifically target cancer cells that present these antigens without harming healthy cells.

Currently dozens of databases exist, where AIRR seq data is stored together with patient and disease metainformation and in some cases also protein structure [243-250]. The increasing volume of NGS generated data as well as their complexity calls for structured ways to store and more importantly for novel computational and analytical approaches in order to allow antigen specificity prediction of immune receptors, an endeavor that is currently seen as the “holy grail” of immune repertoire research.

Databases like VDJdb, McPAS-TCR and IEDB are yet to mature, however provide first approaches towards prediction and analysis of immune epitopes [251-253]. The VDJdb and McPAS-TCR are curated databases which currently hold more than 80 000 TCR sequences drawn from the literature that are associated with pathological conditions. These sequences mainly originate from homo sapiens or mus musculus and are searchable via the web interface which also holds the information on the cognate antigen epitope. However, these databases are mainly limited to TCR sequences potentially directed towards viral antigens like CMV, influenza and EBV and only a minor fraction of TCR sequences are derived from high confidence experiments, i.e. tetramer binding and/or T cell stimulation assays.

Finally, the majority of studies using AIRR seq, including the papers presented in this work, are primarily of descriptive nature. In order to reveal deeper biological and clinical meaning, AIRR profiling will need to be linked to transcriptome and epigenetic data [74] (i.e. in single-cell sequencing approaches). The vast amount of information provided by AIRR seq combined with multi-dimensional profiling strategies will allow insight into the transcriptional and phenotypic profiles of the cells and therefore may facilitate the prediction of TCR or antibody specificity. The integration of these approaches calls for sophisticated computational algorithms, i.e. machine learning, which are currently being used and refined [254-257].

Taken together AIRR seq has helped to deepen our knowledge of lymphocyte biology. Once we overcome the current technical and computational challenges this technique will fully realize its potential offering significant contributions to the development of immunodiagnostics, vaccines, cancer immunotherapy, and antibody engineering.

ABBREVIATIONS

A

AATP – acquired amegakaryocytic thrombocytopenia

AIC – autoimmune cytopenias

AID – autoimmune disease

AIH – autoimmune hepatitis

AIHA – autoimmune hemolytic anemia

AIN – autoimmune neutropenia

Aire – autoimmune regulator

AIRR seq – adaptive immune receptor repertoire sequencing

APC – antigen presenting cell

B

Bcl-2 – B-cell lymphoma 2

C

CD – cluster of differentiation

CD^{hi} – cluster of differentiation, highly expressed

CD^{lo} – cluster of differentiation, low expression

CDR – complementarity determining region

CLL – chronic lymphocytic leukemia

CLP – common lymphoid progenitor cell

CMJ – corticomedullary junction

CMV –cytomegalovirus

cTEC – cortical thymus epithelial cell

CTL – cytotoxic T cell

CTLA-4 – cytotoxic T lymphocyte antigen 4

CXCR4 – C-X-C chemokine receptor type 4

D

DAT – direct antiglobulin test

DC – dendritic cells

DN – double-negative

dNTP – Deoxyribonucleotide triphosphate

DP – double-positive

E

EBV – Epstein-Barr virus

Erk – extracellular signal-regulated kinases

F

FoxP3 – forkhead box P3

FR – framework region

G

G-CSF – granulocyte colony-stimulating factor

gDNA – genomic deoxyribonucleic acid

GIFT – granulocyte immunofluorescence test

GLIPH – grouping of lymphocyte interaction by paratope hotspots

H

HD – healthy donor

HGP – Human Genome Project

HLA – Human Leukocyte Antigen

HMGB – high mobility group box protein

HNA – human neutrophil-specific antigens

HSC – hematopoietic stem cells

I

IAN –immune-associated nucleotide-binding protein

ICB – immune checkpoint blockade

IgG/IgM – immunoglobulin G/M

IgOR – Inference and Generation of Repertoires

IL-7R – interleukin 7 receptor

IMGT – the international ImMunoGeneTics information system database

IR – inhibitory receptor

ITAM – immunoreceptor tyrosine-based activation motifs

ITP – immune thrombocytopenia

IVIg – intravenous immunoglobulin

L

LAG-3 – lymphocyte activation gene 3

LTβR – lymphotoxin-β receptor

M

MAP kinases – mitogen-activated protein kinases

MHC – Major Histocomaptibility Complex

mTEC – medullary thymus epithelial cell

N

NGS – Next-generation Sequencing

NHEJ – nonhomologous end joining

NKT – natural killer T cell

O

OLGA – Optimized Likelihood estimate of immunoGlobulin Amino-acid sequences

P

PBMNC – peripheral blood mononuclear cell

PCR – Polymerase chain reaction

PD-1 – programmed cell death protein 1

PDK – phosphoinositide-dependent kinase

pGEN – generation probability

PKCθ – protein kinase C theta

pMHC – peptide-MHC complex

polμ/polλ – polymerase μ/λ

PRCA – pure red cell aplasia

pTα – germline-encoded pre-TCRα

PTK – protein tyrosine kinases

R

5´-RACE – rapid amplification of complementary DNA ends

RAG – recombination-activating gene

RNA – ribonucleic acid

RSS – recombination signa sequence

S

SASP – senescence-associated secretory phenotype

SP – single positive

SSC – somatic senescent cells

T

T_{con} – conventional T cell

TCR – T cell receptor

TdT – Terminal deoxynucleotidyl transferase

T_{ex} – exhausted T cell

TGF-β – transforming growth factor beta

TIGIT – T cell tyrosine-based inhibitory motif domain

Tim-3 – T cell immunoglobulin domain and mucin domain-3

T_{mem} – memory T cell

TRA – T cell receptor alpha gene loci

TRB – T cell receptor beta gene loci

TRD – T cell receptor delta gene loci

TREC – T cell receptor excision circles

T_{reg} – regulatory T cell

TRG – T cell receptor gamma gene loci

T_{RM} – tissue resident memory T cells

TSLP – thymic stromal lymphopoietin

V

V(D)J – V, D and J gene segments of adaptive immune receptor gene loci

X

XLF – XRCC4-like factor

XRCC4 – X-ray repair cross-complementing protein 4

REFERENCES

1. Lancaster, J.N., Y. Li, and L.I.R. Ehrlich, *Chemokine-Mediated Choreography of Thymocyte Development and Selection.* Trends Immunol, 2018. **39**(2): p. 86-98.
2. Lind, E.F., et al., *Mapping precursor movement through the postnatal thymus reveals specific microenvironments supporting defined stages of early lymphoid development.* J Exp Med, 2001. **194**(2): p. 127-34.
3. Ladi, E., et al., *Thymic microenvironments for T cell differentiation and selection.* Nat Immunol, 2006. **7**(4): p. 338-43.
4. Porritt, H.E., et al., *Heterogeneity among DN1 prothymocytes reveals multiple progenitors with different capacities to generate T cell and non-T cell lineages.* Immunity, 2004. **20**(6): p. 735-45.
5. Petrie, H.T. and J.C. Zuniga-Pflucker, *Zoned out: functional mapping of stromal signaling microenvironments in the thymus.* Annu Rev Immunol, 2007. **25**: p. 649-79.
6. Rothenberg, E.V., *T cell lineage commitment: identity and renunciation.* J Immunol, 2011. **186**(12): p. 6649-55.
7. Shah, D.K. and J.C. Zuniga-Pflucker, *An overview of the intrathymic intricacies of T cell development.* J Immunol, 2014. **192**(9): p. 4017-23.
8. Wilson, A., W. Held, and H.R. MacDonald, *Two waves of recombinase gene expression in developing thymocytes.* J Exp Med, 1994. **179**(4): p. 1355-60.
9. Capone, M., R.D. Hockett, Jr., and A. Zlotnik, *Kinetics of T cell receptor beta, gamma, and delta rearrangements during adult thymic development: T cell receptor rearrangements are present in CD44(+)CD25(+) Pro-T thymocytes.* Proc Natl Acad Sci U S A, 1998. **95**(21): p. 12522-7.
10. Livak, F., et al., *Characterization of TCR gene rearrangements during adult murine T cell development.* J Immunol, 1999. **162**(5): p. 2575-80.
11. Petrie, H.T., et al., *T cell receptor gene recombination patterns and mechanisms: cell death, rescue, and T cell production.* J Exp Med, 1995. **182**(1): p. 121-7.
12. Massa, S., et al., *Critical role for c-kit (CD117) in T cell lineage commitment and early thymocyte development in vitro.* Eur J Immunol, 2006. **36**(3): p. 526-32.
13. Huang, J., et al., *Propensity of adult lymphoid progenitors to progress to DN2/3 stage thymocytes with Notch receptor ligation.* J Immunol, 2005. **175**(8): p. 4858-65.
14. Dudley, E.C., et al., *T cell receptor beta chain gene rearrangement and selection during thymocyte development in adult mice.* Immunity, 1994. **1**(2): p. 83-93.
15. Saint-Ruf, C., et al., *Analysis and expression of a cloned pre-T cell receptor gene.* Science, 1994. **266**(5188): p. 1208-12.
16. Egerton, M., K. Shortman, and R. Scollay, *The kinetics of immature murine thymocyte development in vivo.* Int Immunol, 1990. **2**(6): p. 501-7.
17. Tourigny, M.R., et al., *T cell receptor (TCR)-beta gene recombination: dissociation from cell cycle regulation and developmental progression during T cell ontogeny.* J Exp Med, 1997. **185**(9): p. 1549-56.
18. Penit, C., B. Lucas, and F. Vasseur, *Cell expansion and growth arrest phases during the transition from precursor (CD4-8-) to immature (CD4+8+) thymocytes in normal and genetically modified mice.* J Immunol, 1995. **154**(10): p. 5103-13.
19. Petrie, H.T., et al., *Development of immature thymocytes: initiation of CD3, CD4, and CD8 acquisition parallels down-regulation of the interleukin 2 receptor alpha chain.* Eur J Immunol, 1990. **20**(12): p. 2813-5.
20. Petrie, H.T., et al., *Multiple rearrangements in T cell receptor alpha chain genes maximize the production of useful thymocytes.* J Exp Med, 1993. **178**(2): p. 615-22.

21. Penit, C., *Localization and phenotype of cycling and post-cycling murine thymocytes studied by simultaneous detection of bromodeoxyuridine and surface antigens.* J Histochem Cytochem, 1988. **36**(5): p. 473-8.
22. Prockop, S.E. and H.T. Petrie, *Regulation of thymus size by competition for stromal niches among early T cell progenitors.* J Immunol, 2004. **173**(3): p. 1604-11.
23. Tanigaki, K., et al., *Regulation of alphabeta/gammadelta T cell lineage commitment and peripheral T cell responses by Notch/RBP-J signaling.* Immunity, 2004. **20**(5): p. 611-22.
24. Speiser, D.E., et al., *Positive and negative selection of T cell receptor V beta domains controlled by distinct cell populations in the thymus.* J Exp Med, 1989. **170**(6): p. 2165-70.
25. Germain, R.N., *T-cell development and the CD4-CD8 lineage decision.* Nat Rev Immunol, 2002. **2**(5): p. 309-22.
26. Yu, Q., et al., *Cytokine signal transduction is suppressed in preselection double-positive thymocytes and restored by positive selection.* J Exp Med, 2006. **203**(1): p. 165-75.
27. Kurobe, H., et al., *CCR7-dependent cortex-to-medulla migration of positively selected thymocytes is essential for establishing central tolerance.* Immunity, 2006. **24**(2): p. 165-77.
28. Porritt, H.E., K. Gordon, and H.T. Petrie, *Kinetics of steady-state differentiation and mapping of intrathymic-signaling environments by stem cell transplantation in nonirradiated mice.* J Exp Med, 2003. **198**(6): p. 957-62.
29. Sprent, J., E.K. Gao, and S.R. Webb, *T cell reactivity to MHC molecules: immunity versus tolerance.* Science, 1990. **248**(4961): p. 1357-63.
30. Kyewski, B. and J. Derbinski, *Self-representation in the thymus: an extended view.* Nat Rev Immunol, 2004. **4**(9): p. 688-98.
31. Heath, W.R. and F.R. Carbone, *Cross-presentation, dendritic cells, tolerance and immunity.* Annu Rev Immunol, 2001. **19**: p. 47-64.
32. Anderson, M.S., et al., *Projection of an immunological self shadow within the thymus by the aire protein.* Science, 2002. **298**(5597): p. 1395-401.
33. Anderson, M.S., et al., *The cellular mechanism of Aire control of T cell tolerance.* Immunity, 2005. **23**(2): p. 227-39.
34. Buhlmann, J.E., S.K. Elkin, and A.H. Sharpe, *A role for the B7-1/B7-2:CD28/CTLA-4 pathway during negative selection.* J Immunol, 2003. **170**(11): p. 5421-8.
35. Foy, T.M., et al., *An essential role for gp39, the ligand for CD40, in thymic selection.* J Exp Med, 1995. **182**(5): p. 1377-88.
36. Yu, Q., et al., *In vitro evidence that cytokine receptor signals are required for differentiation of double positive thymocytes into functionally mature CD8+ T cells.* J Exp Med, 2003. **197**(4): p. 475-87.
37. Linette, G.P., et al., *Bcl-2 is upregulated at the CD4+ CD8+ stage during positive selection and promotes thymocyte differentiation at several control points.* Immunity, 1994. **1**(3): p. 197-205.
38. Hancock, W.W., et al., *Costimulatory function and expression of CD40 ligand, CD80, and CD86 in vascularized murine cardiac allograft rejection.* Proc Natl Acad Sci U S A, 1996. **93**(24): p. 13967-72.
39. Nitta, T., et al., *IAN family critically regulates survival and development of T lymphocytes.* PLoS Biol, 2006. **4**(4): p. e103.
40. Fontenot, J.D. and A.Y. Rudensky, *A well adapted regulatory contrivance: regulatory T cell development and the forkhead family transcription factor Foxp3.* Nat Immunol, 2005. **6**(4): p. 331-7.
41. Hsieh, C.S., et al., *An intersection between the self-reactive regulatory and nonregulatory T cell receptor repertoires.* Nat Immunol, 2006. **7**(4): p. 401-10.
42. Watanabe, N., et al., *Hassall's corpuscles instruct dendritic cells to induce CD4+CD25+ regulatory T cells in human thymus.* Nature, 2005. **436**(7054): p. 1181-5.
43. Jiang, Q., et al., *Delayed functional maturation of natural regulatory T cells in the medulla of postnatal thymus: role of TSLP.* BMC Immunol, 2006. **7**: p. 6.

44. Boehm, T., et al., *Thymic medullary epithelial cell differentiation, thymocyte emigration, and the control of autoimmunity require lympho-epithelial cross talk via LTbetaR.* J Exp Med, 2003. **198**(5): p. 757-69.
45. Roth, D.B., *V(D)J Recombination: Mechanism, Errors, and Fidelity.* Microbiol Spectr, 2014. **2**(6).
46. Qi, Q., et al., *Diversity and clonal selection in the human T-cell repertoire.* Proc Natl Acad Sci U S A, 2014. **111**(36): p. 13139-44.
47. Lythe, G., et al., *How many TCR clonotypes does a body maintain?* J Theor Biol, 2016. **389**: p. 214-24.
48. Ye, B., et al., *High-throughput sequencing of the immune repertoire in oncology: Applications for clinical diagnosis, monitoring, and immunotherapies.* Cancer Lett, 2018. **416**: p. 42-56.
49. Schatz, D.G. and P.C. Swanson, *V(D)J recombination: mechanisms of initiation.* Annu Rev Genet, 2011. **45**: p. 167-202.
50. Hesse, J.E., et al., *Extrachromosomal DNA substrates in pre-B cells undergo inversion or deletion at immunoglobulin V-(D)-J joining signals.* Cell, 1987. **49**(6): p. 775-83.
51. Hesse, J.E., et al., *V(D)J recombination: a functional definition of the joining signals.* Genes Dev, 1989. **3**(7): p. 1053-61.
52. Oettinger, M.A., et al., *RAG-1 and RAG-2, adjacent genes that synergistically activate V(D)J recombination.* Science, 1990. **248**(4962): p. 1517-23.
53. van Gent, D.C., et al., *Stimulation of V(D)J cleavage by high mobility group proteins.* EMBO J, 1997. **16**(10): p. 2665-70.
54. McBlane, J.F., et al., *Cleavage at a V(D)J recombination signal requires only RAG1 and RAG2 proteins and occurs in two steps.* Cell, 1995. **83**(3): p. 387-95.
55. Hiom, K. and M. Gellert, *Assembly of a 12/23 paired signal complex: a critical control point in V(D)J recombination.* Mol Cell, 1998. **1**(7): p. 1011-9.
56. Lieber, M.R., *The mechanism of human nonhomologous DNA end joining.* J Biol Chem, 2008. **283**(1): p. 1-5.
57. Ma, Y., et al., *Hairpin opening and overhang processing by an Artemis/DNA-dependent protein kinase complex in nonhomologous end joining and V(D)J recombination.* Cell, 2002. **108**(6): p. 781-94.
58. Yamtich, J. and J.B. Sweasy, *DNA polymerase family X: function, structure, and cellular roles.* Biochim Biophys Acta, 2010. **1804**(5): p. 1136-50.
59. Kreslavsky, T., et al., *alphabeta versus gammadelta fate choice: counting the T-cell lineages at the branch point.* Immunol Rev, 2010. **238**(1): p. 169-81.
60. Zhao, Y., C. Niu, and J. Cui, *Gamma-delta (gammadelta) T cells: friend or foe in cancer development?* J Transl Med, 2018. **16**(1): p. 3.
61. Hayes, S.M., L. Li, and P.E. Love, *TCR signal strength influences alphabeta/gammadelta lineage fate.* Immunity, 2005. **22**(5): p. 583-93.
62. Haks, M.C., et al., *Attenuation of gammadeltaTCR signaling efficiently diverts thymocytes to the alphabeta lineage.* Immunity, 2005. **22**(5): p. 595-606.
63. Chapman, J.C., F.M. Chapman, and S.D. Michael, *The production of alpha/beta and gamma/delta double negative (DN) T-cells and their role in the maintenance of pregnancy.* Reprod Biol Endocrinol, 2015. **13**: p. 73.
64. Smith-Garvin, J.E., G.A. Koretzky, and M.S. Jordan, *T cell activation.* Annu Rev Immunol, 2009. **27**: p. 591-619.
65. Isakov, N., *Immunoreceptor tyrosine-based activation motif (ITAM), a unique module linking antigen and Fc receptors to their signaling cascades.* J Leukoc Biol, 1997. **61**(1): p. 6-16.
66. Letourneur, F. and R.D. Klausner, *Activation of T cells by a tyrosine kinase activation domain in the cytoplasmic tail of CD3 epsilon.* Science, 1992. **255**(5040): p. 79-82.
67. Gorentla, B.K. and X.P. Zhong, *T cell Receptor Signal Transduction in T lymphocytes.* J Clin Cell Immunol, 2012. **2012**(Suppl 12): p. 5.
68. Sanger, F., S. Nicklen, and A.R. Coulson, *DNA sequencing with chain-terminating inhibitors.* Proc Natl Acad Sci U S A, 1977. **74**(12): p. 5463-7.

69. Lander, E.S., et al., *Initial sequencing and analysis of the human genome.* Nature, 2001. **409**(6822): p. 860-921.
70. Venter, J.C., et al., *The sequence of the human genome.* Science, 2001. **291**(5507): p. 1304-51.
71. Wetterstrand, K.A. *DNA Sequencing Costs: Data from the NHGRI Genome Sequencing program (GSP)*. 16.03.2020]; Available from: www.genome.gov/sequencingcostsdata.
72. Buermans, H.P. and J.T. den Dunnen, *Next generation sequencing technology: Advances and applications.* Biochim Biophys Acta, 2014. **1842**(10): p. 1932-1941.
73. Metzker, M.L., *Sequencing technologies - the next generation.* Nat Rev Genet, 2010. **11**(1): p. 31-46.
74. Stubbington, M.J.T., et al., *T cell fate and clonality inference from single-cell transcriptomes.* Nat Methods, 2016. **13**(4): p. 329-332.
75. Eltahla, A.A., et al., *Linking the T cell receptor to the single cell transcriptome in antigen-specific human T cells.* Immunol Cell Biol, 2016. **94**(6): p. 604-11.
76. Davodeau, F., et al., *Dual T cell receptor beta chain expression on human T lymphocytes.* J Exp Med, 1995. **181**(4): p. 1391-8.
77. Padovan, E., et al., *Normal T lymphocytes can express two different T cell receptor beta chains: implications for the mechanism of allelic exclusion.* J Exp Med, 1995. **181**(4): p. 1587-91.
78. Steinel, N.C., et al., *Posttranscriptional silencing of VbetaDJbetaCbeta genes contributes to TCRbeta allelic exclusion in mammalian lymphocytes.* J Immunol, 2010. **185**(2): p. 1055-62.
79. Rybakin, V., et al., *Allelic exclusion of TCR alpha-chains upon severe restriction of Valpha repertoire.* PLoS One, 2014. **9**(12): p. e114320.
80. Niederberger, N., et al., *Allelic exclusion of the TCR alpha-chain is an active process requiring TCR-mediated signaling and c-Cbl.* J Immunol, 2003. **170**(9): p. 4557-63.
81. Rosati, E., et al., *Overview of methodologies for T-cell receptor repertoire analysis.* BMC Biotechnol, 2017. **17**(1): p. 61.
82. Mamedov, I.Z., et al., *Preparing unbiased T-cell receptor and antibody cDNA libraries for the deep next generation sequencing profiling.* Front Immunol, 2013. **4**: p. 456.
83. van Dongen, J.J., et al., *Design and standardization of PCR primers and protocols for detection of clonal immunoglobulin and T-cell receptor gene recombinations in suspect lymphoproliferations: report of the BIOMED-2 Concerted Action BMH4-CT98-3936.* Leukemia, 2003. **17**(12): p. 2257-317.
84. Akyuz, N., et al., *T-cell diversification reflects antigen selection in the blood of patients on immune checkpoint inhibition and may be exploited as liquid biopsy biomarker.* Int J Cancer, 2017. **140**(11): p. 2535-2544.
85. Schliffke, S., et al., *Clinical response to ibrutinib is accompanied by normalization of the T-cell environment in CLL-related autoimmune cytopenia.* Leukemia, 2016. **30**(11): p. 2232-2234.
86. Bolotin, D.A., et al., *MiXCR: software for comprehensive adaptive immunity profiling.* Nat Methods, 2015. **12**(5): p. 380-1.
87. Shannon, C.E., *A Mathematical Theory of Communication.* Bell System Technical Journal, 1948. **27**(3): p. 379-423.
88. Glanville, J., et al., *Identifying specificity groups in the T cell receptor repertoire.* Nature, 2017. **547**(7661): p. 94-98.
89. Huang, H., et al., *Analyzing the Mycobacterium tuberculosis immune response by T-cell receptor clustering with GLIPH2 and genome-wide antigen screening.* Nat Biotechnol, 2020.
90. Rudolph, M.G., R.L. Stanfield, and I.A. Wilson, *How TCRs bind MHCs, peptides, and coreceptors.* Annu Rev Immunol, 2006. **24**: p. 419-66.
91. Mora, T. and A. Walczak, *Quantifying lymphocyte receptor diversity*, in *Systems Immunology: An Introduction to Modeling Methods for Scientists*, J. Das and C. Jayaprakash, Editors. 2019, CRC Press, Taylor and Francis: Boca Raton, FL. p. 1-10.

92. Shortman, K., et al., *The generation and fate of thymocytes.* Semin Immunol, 1990. **2**(1): p. 3-12.
93. Argaet, V.P., et al., *Dominant selection of an invariant T cell antigen receptor in response to persistent infection by Epstein-Barr virus.* J Exp Med, 1994. **180**(6): p. 2335-40.
94. Moss, P.A., et al., *Extensive conservation of alpha and beta chains of the human T-cell antigen receptor recognizing HLA-A2 and influenza A matrix peptide.* Proc Natl Acad Sci U S A, 1991. **88**(20): p. 8987-90.
95. Casanova, J.L., et al., *H-2-restricted cytolytic T lymphocytes specific for HLA display T cell receptors of limited diversity.* J Exp Med, 1992. **176**(2): p. 439-47.
96. Venturi, V., et al., *The molecular basis for public T-cell responses?* Nat Rev Immunol, 2008. **8**(3): p. 231-8.
97. Quigley, M.F., et al., *Convergent recombination shapes the clonotypic landscape of the naive T-cell repertoire.* Proc Natl Acad Sci U S A, 2010. **107**(45): p. 19414-9.
98. Venturi, V., et al., *The role of production frequency in the sharing of simian immunodeficiency virus-specific CD8+ TCRs between macaques.* J Immunol, 2008. **181**(4): p. 2597-609.
99. Venturi, V., et al., *Sharing of T cell receptors in antigen-specific responses is driven by convergent recombination.* Proc Natl Acad Sci U S A, 2006. **103**(49): p. 18691-6.
100. Elhanati, Y., et al., *Predicting the spectrum of TCR repertoire sharing with a data-driven model of recombination.* Immunol Rev, 2018. **284**(1): p. 167-179.
101. Marcou, Q., T. Mora, and A.M. Walczak, *High-throughput immune repertoire analysis with IGoR.* Nat Commun, 2018. **9**(1): p. 561.
102. Sethna, Z., et al., *OLGA: fast computation of generation probabilities of B- and T-cell receptor amino acid sequences and motifs.* Bioinformatics, 2019. **35**(17): p. 2974-2981.
103. He, S. and N.E. Sharpless, *Senescence in Health and Disease.* Cell, 2017. **169**(6): p. 1000-1011.
104. Freund, A., et al., *Inflammatory networks during cellular senescence: causes and consequences.* Trends Mol Med, 2010. **16**(5): p. 238-46.
105. Coppe, J.P., et al., *The senescence-associated secretory phenotype: the dark side of tumor suppression.* Annu Rev Pathol, 2010. **5**: p. 99-118.
106. Thomas, R., W. Wang, and D.M. Su, *Contributions of Age-Related Thymic Involution to Immunosenescence and Inflammaging.* Immun Ageing, 2020. **17**: p. 2.
107. Sharp, A., T. Kukulansky, and A. Globerson, *In vitro analysis of age-related changes in the developmental potential of bone marrow thymocyte progenitors.* Eur J Immunol, 1990. **20**(12): p. 2541-6.
108. de Haan, G., et al., *A genetic and genomic analysis identifies a cluster of genes associated with hematopoietic cell turnover.* Blood, 2002. **100**(6): p. 2056-62.
109. Geiger, H. and G. Van Zant, *The aging of lympho-hematopoietic stem cells.* Nat Immunol, 2002. **3**(4): p. 329-33.
110. Henckaerts, E., et al., *Genetically determined variation in the number of phenotypically defined hematopoietic progenitor and stem cells and in their response to early-acting cytokines.* Blood, 2002. **99**(11): p. 3947-54.
111. Morrison, S.J., et al., *A genetic determinant that specifically regulates the frequency of hematopoietic stem cells.* J Immunol, 2002. **168**(2): p. 635-42.
112. Morrison, S.J., et al., *The aging of hematopoietic stem cells.* Nat Med, 1996. **2**(9): p. 1011-6.
113. Lamberts, S.W., A.W. van den Beld, and A.J. van der Lely, *The endocrinology of aging.* Science, 1997. **278**(5337): p. 419-24.
114. Linton, P.J. and K. Dorshkind, *Age-related changes in lymphocyte development and function.* Nat Immunol, 2004. **5**(2): p. 133-9.
115. Haynes, B.F., et al., *The role of the thymus in immune reconstitution in aging, bone marrow transplantation, and HIV-1 infection.* Annu Rev Immunol, 2000. **18**: p. 529-60.
116. Jamieson, B.D., et al., *Generation of functional thymocytes in the human adult.* Immunity, 1999. **10**(5): p. 569-75.

117. Klein, L., E.A. Robey, and C.S. Hsieh, *Central CD4(+) T cell tolerance: deletion versus regulatory T cell differentiation.* Nat Rev Immunol, 2019. **19**(1): p. 7-18.
118. Pohar, J., Q. Simon, and S. Fillatreau, *Antigen-Specificity in the Thymic Development and Peripheral Activity of CD4(+)FOXP3(+) T Regulatory Cells.* Front Immunol, 2018. **9**: p. 1701.
119. Lathrop, S.K., et al., *Antigen-specific peripheral shaping of the natural regulatory T cell population.* J Exp Med, 2008. **205**(13): p. 3105-17.
120. Thiault, N., et al., *Peripheral regulatory T lymphocytes recirculating to the thymus suppress the development of their precursors.* Nat Immunol, 2015. **16**(6): p. 628-34.
121. Oh, J., et al., *Capacity of tTreg generation is not impaired in the atrophied thymus.* PLoS Biol, 2017. **15**(11): p. e2003352.
122. Fletcher, A.L., et al., *Reduced thymic Aire expression and abnormal NF-kappa B2 signaling in a model of systemic autoimmunity.* J Immunol, 2009. **182**(5): p. 2690-9.
123. Malchow, S., et al., *Aire Enforces Immune Tolerance by Directing Autoreactive T Cells into the Regulatory T Cell Lineage.* Immunity, 2016. **44**(5): p. 1102-13.
124. Yanes, R.E., et al., *Lymphocyte generation and population homeostasis throughout life.* Semin Hematol, 2017. **54**(1): p. 33-38.
125. Simon, A.K., G.A. Hollander, and A. McMichael, *Evolution of the immune system in humans from infancy to old age.* Proc Biol Sci, 2015. **282**(1821): p. 20143085.
126. Rezzani, R., et al., *Thymus and aging: morphological, radiological, and functional overview.* Age (Dordr), 2014. **36**(1): p. 313-51.
127. Palmer, D.B., *The effect of age on thymic function.* Front Immunol, 2013. **4**: p. 316.
128. Nikolich-Zugich, J., *Aging of the T cell compartment in mice and humans: from no naive expectations to foggy memories.* J Immunol, 2014. **193**(6): p. 2622-9.
129. Chou, J.P. and R.B. Effros, *T cell replicative senescence in human aging.* Curr Pharm Des, 2013. **19**(9): p. 1680-98.
130. Palucka, A.K. and L.M. Coussens, *The Basis of Oncoimmunology.* Cell, 2016. **164**(6): p. 1233-1247.
131. Thun, M.J., S.J. Henley, and T. Gansler, *Inflammation and cancer: an epidemiological perspective.* Novartis Found Symp, 2004. **256**: p. 6-21; discussion 22-8, 49-52, 266-9.
132. Campisi, J., et al., *Cellular senescence: a link between cancer and age-related degenerative disease?* Semin Cancer Biol, 2011. **21**(6): p. 354-9.
133. Steinman, R.M., *Decisions about dendritic cells: past, present, and future.* Annu Rev Immunol, 2012. **30**: p. 1-22.
134. Le Floc'h, A., et al., *Alpha E beta 7 integrin interaction with E-cadherin promotes antitumor CTL activity by triggering lytic granule polarization and exocytosis.* J Exp Med, 2007. **204**(3): p. 559-70.
135. Appay, V., D.C. Douek, and D.A. Price, *CD8+ T cell efficacy in vaccination and disease.* Nat Med, 2008. **14**(6): p. 623-8.
136. Cui, W. and S.M. Kaech, *Generation of effector CD8+ T cells and their conversion to memory T cells.* Immunol Rev, 2010. **236**: p. 151-66.
137. McLane, L.M., M.S. Abdel-Hakeem, and E.J. Wherry, *CD8 T Cell Exhaustion During Chronic Viral Infection and Cancer.* Annu Rev Immunol, 2019. **37**: p. 457-495.
138. Doering, T.A., et al., *Network analysis reveals centrally connected genes and pathways involved in CD8+ T cell exhaustion versus memory.* Immunity, 2012. **37**(6): p. 1130-44.
139. Schietinger, A. and P.D. Greenberg, *Tolerance and exhaustion: defining mechanisms of T cell dysfunction.* Trends Immunol, 2014. **35**(2): p. 51-60.
140. Wherry, E.J., et al., *Molecular signature of CD8+ T cell exhaustion during chronic viral infection.* Immunity, 2007. **27**(4): p. 670-84.
141. Day, C.L., et al., *PD-1 expression on HIV-specific T cells is associated with T-cell exhaustion and disease progression.* Nature, 2006. **443**(7109): p. 350-4.
142. Chauvin, J.M., et al., *TIGIT and PD-1 impair tumor antigen-specific CD8(+) T cells in melanoma patients.* J Clin Invest, 2015. **125**(5): p. 2046-58.
143. Woo, S.R., et al., *Immune inhibitory molecules LAG-3 and PD-1 synergistically regulate T-cell function to promote tumoral immune escape.* Cancer Res, 2012. **72**(4): p. 917-27.

144. Johnston, R.J., et al., *The immunoreceptor TIGIT regulates antitumor and antiviral CD8(+) T cell effector function.* Cancer Cell, 2014. **26**(6): p. 923-937.
145. Fourcade, J., et al., *Upregulation of Tim-3 and PD-1 expression is associated with tumor antigen-specific CD8+ T cell dysfunction in melanoma patients.* J Exp Med, 2010. **207**(10): p. 2175-86.
146. Fourcade, J., et al., *CD8(+) T cells specific for tumor antigens can be rendered dysfunctional by the tumor microenvironment through upregulation of the inhibitory receptors BTLA and PD-1.* Cancer Res, 2012. **72**(4): p. 887-96.
147. Gustafson, C.E., et al., *Immune cell repertoires in breast cancer patients after adjuvant chemotherapy.* JCI Insight, 2020. **5**(4).
148. Verma, R., et al., *Lymphocyte depletion and repopulation after chemotherapy for primary breast cancer.* Breast Cancer Res, 2016. **18**(1): p. 10.
149. Fagnoni, F.F., et al., *T-cell dynamics after high-dose chemotherapy in adults: elucidation of the elusive CD8+ subset reveals multiple homeostatic T-cell compartments with distinct implications for immune competence.* Immunology, 2002. **106**(1): p. 27-37.
150. Havenith, S.H., et al., *Rapid T cell repopulation after rabbit anti-thymocyte globulin (rATG) treatment is driven mainly by cytomegalovirus.* Clin Exp Immunol, 2012. **169**(3): p. 292-301.
151. Mackall, C.L., et al., *Distinctions between CD8+ and CD4+ T-cell regenerative pathways result in prolonged T-cell subset imbalance after intensive chemotherapy.* Blood, 1997. **89**(10): p. 3700-7.
152. Sun, D.P., et al., *Thymic hyperplasia after chemotherapy in adults with mature B cell lymphoma and its influence on thymic output and CD4(+) T cells repopulation.* Oncoimmunology, 2016. **5**(5): p. e1137417.
153. Goodnow, C.C., *Multistep pathogenesis of autoimmune disease.* Cell, 2007. **130**(1): p. 25-35.
154. Nagy, G., et al., *Selected Aspects in the Pathogenesis of Autoimmune Diseases.* Mediators Inflamm, 2015. **2015**: p. 351732.
155. Kyewski, B. and L. Klein, *A central role for central tolerance.* Annu Rev Immunol, 2006. **24**: p. 571-606.
156. Shortman, K., D. Vremec, and M. Egerton, *The kinetics of T cell antigen receptor expression by subgroups of CD4+8+ thymocytes: delineation of CD4+8+3(2+) thymocytes as post-selection intermediates leading to mature T cells.* J Exp Med, 1991. **173**(2): p. 323-32.
157. Huesmann, M., et al., *Kinetics and efficacy of positive selection in the thymus of normal and T cell receptor transgenic mice.* Cell, 1991. **66**(3): p. 533-40.
158. Ignatowicz, L., J. Kappler, and P. Marrack, *The repertoire of T cells shaped by a single MHC/peptide ligand.* Cell, 1996. **84**(4): p. 521-9.
159. Fukui, Y., et al., *Positive and negative CD4+ thymocyte selection by a single MHC class II/peptide ligand affected by its expression level in the thymus.* Immunity, 1997. **6**(4): p. 401-10.
160. van Meerwijk, J.P., et al., *Quantitative impact of thymic clonal deletion on the T cell repertoire.* J Exp Med, 1997. **185**(3): p. 377-83.
161. Tourne, S., et al., *Selection of a broad repertoire of CD4+ T cells in H-2Ma0/0 mice.* Immunity, 1997. **7**(2): p. 187-95.
162. Gascoigne, N.R., et al., *TCR Signal Strength and T Cell Development.* Annu Rev Cell Dev Biol, 2016. **32**: p. 327-348.
163. Li, M.O. and A.Y. Rudensky, *T cell receptor signalling in the control of regulatory T cell differentiation and function.* Nat Rev Immunol, 2016. **16**(4): p. 220-33.
164. Gotter, J., et al., *Medullary epithelial cells of the human thymus express a highly diverse selection of tissue-specific genes colocalized in chromosomal clusters.* J Exp Med, 2004. **199**(2): p. 155-66.
165. Tang, Q. and J.A. Bluestone, *The Foxp3+ regulatory T cell: a jack of all trades, master of regulation.* Nat Immunol, 2008. **9**(3): p. 239-44.

166. Shevach, E.M., *Mechanisms of foxp3+ T regulatory cell-mediated suppression.* Immunity, 2009. **30**(5): p. 636-45.
167. Vignali, D.A., L.W. Collison, and C.J. Workman, *How regulatory T cells work.* Nat Rev Immunol, 2008. **8**(7): p. 523-32.
168. Barath, S., et al., *Regulatory T cells in peripheral blood of patients with mixed connective tissue disease.* Scand J Rheumatol, 2006. **35**(4): p. 300-4.
169. Szodoray, P., et al., *Cells with regulatory function of the innate and adaptive immune system in primary Sjogren's syndrome.* Clin Exp Immunol, 2009. **157**(3): p. 343-9.
170. Liu, M.F., et al., *Decreased CD4+CD25+ T cells in peripheral blood of patients with systemic lupus erythematosus.* Scand J Immunol, 2004. **59**(2): p. 198-202.
171. Dejaco, C., et al., *Imbalance of regulatory T cells in human autoimmune diseases.* Immunology, 2006. **117**(3): p. 289-300.
172. Barcellini, W., et al., *Clinical heterogeneity and predictors of outcome in primary autoimmune hemolytic anemia: a GIMEMA study of 308 patients.* Blood, 2014. **124**(19): p. 2930-6.
173. Audia, S., et al., *Pathogenesis of immune thrombocytopenia.* Autoimmun Rev, 2017. **16**(6): p. 620-632.
174. Youinou, P., et al., *Diagnostic criteria for autoimmune neutropenia.* Autoimmun Rev, 2014. **13**(4-5): p. 574-6.
175. Ichimata, S., et al., *Acquired amegakaryocytic thrombocytopenia previously diagnosed as idiopathic thrombocytopenic purpura in a patient with hepatitis C virus infection.* World J Gastroenterol, 2017. **23**(35): p. 6540-6545.
176. Means, R.T., Jr., *Pure red cell aplasia.* Blood, 2016. **128**(21): p. 2504-2509.
177. Jaime-Perez, J.C., et al., *Evans syndrome: clinical perspectives, biological insights and treatment modalities.* J Blood Med, 2018. **9**: p. 171-184.
178. Elebute, M. and J. Marsh, *Autoimmune Cytopenias.* Medicine, 2009. **37**: p. 159-163.
179. Zufferey, A., R. Kapur, and J.W. Semple, *Pathogenesis and Therapeutic Mechanisms in Immune Thrombocytopenia (ITP).* J Clin Med, 2017. **6**(2).
180. Nishimoto, T. and M. Kuwana, *CD4+CD25+Foxp3+ regulatory T cells in the pathophysiology of immune thrombocytopenia.* Semin Hematol, 2013. **50 Suppl 1**: p. S43-9.
181. Fagiolo, E., *Immunological tolerance loss vs. erythrocyte self antigens and cytokine network disregulation in autoimmune hemolytic anaemia.* Autoimmunity Reviews, 2004. **3**(2): p. 53-59.
182. Olsson, B., et al., *T-cell-mediated cytotoxicity toward platelets in chronic idiopathic thrombocytopenic purpura.* Nat Med, 2003. **9**(9): p. 1123-4.
183. Khodadi, E., et al., *Bone marrow niche in immune thrombocytopenia: a focus on megakaryopoiesis.* Ann Hematol, 2016. **95**(11): p. 1765-76.
184. Hill, A. and Q.A. Hill, *Autoimmune hemolytic anemia.* Hematology Am Soc Hematol Educ Program, 2018. **2018**(1): p. 382-389.
185. Garvey, B., *Rituximab in the treatment of autoimmune haematological disorders.* Br J Haematol, 2008. **141**(2): p. 149-69.
186. George, J.N., *Definition, diagnosis and treatment of immune thrombocytopenic purpura.* Haematologica, 2009. **94**(6): p. 759-62.
187. Cooper, N., et al., *The efficacy and safety of B-cell depletion with anti-CD20 monoclonal antibody in adults with chronic immune thrombocytopenic purpura.* Br J Haematol, 2004. **125**(2): p. 232-9.
188. Bux, J. and J. Chapman, *Report on the second international granulocyte serology workshop.* Transfusion, 1997. **37**(9): p. 977-83.
189. Smith, M.A. and J.G. Smith, *Clinical experience with the use of rhG-CSF in secondary autoimmune neutropenia.* Clin Lab Haematol, 2002. **24**(2): p. 93-7.
190. Liu, X. and J. Wu, *History, applications, and challenges of immune repertoire research.* Cell Biol Toxicol, 2018. **34**(6): p. 441-457.
191. Hindley, J.P., et al., *Analysis of the T-cell receptor repertoires of tumor-infiltrating conventional and regulatory T cells reveals no evidence for conversion in carcinogen-induced tumors.* Cancer Res, 2011. **71**(3): p. 736-46.

192. Pannetier, C., et al., *The sizes of the CDR3 hypervariable regions of the murine T-cell receptor beta chains vary as a function of the recombined germ-line segments.* Proc Natl Acad Sci U S A, 1993. **90**(9): p. 4319-23.
193. McLean-Tooke, A., et al., *T cell receptor Vbeta repertoire of T lymphocytes and T regulatory cells by flow cytometric analysis in healthy children.* Clin Exp Immunol, 2008. **151**(1): p. 190-8.
194. van den Beemd, R., et al., *Flow cytometric analysis of the Vbeta repertoire in healthy controls.* Cytometry, 2000. **40**(4): p. 336-45.
195. Pacholczyk, R., et al., *Origin and T cell receptor diversity of Foxp3+CD4+CD25+ T cells.* Immunity, 2006. **25**(2): p. 249-59.
196. Price, D.A., et al., *Avidity for antigen shapes clonal dominance in CD8+ T cell populations specific for persistent DNA viruses.* J Exp Med, 2005. **202**(10): p. 1349-61.
197. Khan, N., et al., *Comparative analysis of CD8+ T cell responses against human cytomegalovirus proteins pp65 and immediate early 1 shows similarities in precursor frequency, oligoclonality, and phenotype.* J Infect Dis, 2002. **185**(8): p. 1025-34.
198. Rettig, M.B., et al., *VH gene usage is multiple myeloma: complete absence of the VH4.21 (VH4-34) gene.* Blood, 1996. **87**(7): p. 2846-52.
199. Lossos, I.S., et al., *Molecular analysis of immunoglobulin genes in diffuse large B-cell lymphomas.* Blood, 2000. **95**(5): p. 1797-803.
200. Salvi, S., et al., *Overexpression of the T-cell receptor beta-chain variable region TCRBV14 in HLA-A2-matched primary human melanomas.* Cancer Res, 1995. **55**(15): p. 3374-9.
201. Farina, C., et al., *Conserved TCR usage by HLA-Cw* 1601-restricted T cell clones recognizing melanoma antigens.* Int Immunol, 1996. **8**(9): p. 1463-6.
202. Stamatopoulos, K., et al., *Over 20% of patients with chronic lymphocytic leukemia carry stereotyped receptors: Pathogenetic implications and clinical correlations.* Blood, 2007. **109**(1): p. 259-70.
203. Luo, W., et al., *Analysis of the interindividual conservation of T cell receptor alpha- and beta-chain variable regions gene in the peripheral blood of patients with systemic lupus erythematosus.* Clin Exp Immunol, 2008. **154**(3): p. 316-24.
204. Oksenberg, J.R., et al., *Selection for T-cell receptor V beta-D beta-J beta gene rearrangements with specificity for a myelin basic protein peptide in brain lesions of multiple sclerosis.* Nature, 1993. **362**(6415): p. 68-70.
205. Wucherpfennig, K.W., et al., *Shared human T cell receptor V beta usage to immunodominant regions of myelin basic protein.* Science, 1990. **248**(4958): p. 1016-9.
206. Weinstein, J.A., et al., *High-throughput sequencing of the zebrafish antibody repertoire.* Science, 2009. **324**(5928): p. 807-10.
207. Martinez-Lopez, J., et al., *Analytical and clinical validation of a novel in-house deep-sequencing method for minimal residual disease monitoring in a phase II trial for multiple myeloma.* Leukemia, 2017. **31**(6): p. 1446-1449.
208. Li, H., et al., *Determinants of public T cell responses.* Cell Res, 2012. **22**(1): p. 33-42.
209. Gil, A., et al., *Epstein-Barr Virus Epitope-Major Histocompatibility Complex Interaction Combined with Convergent Recombination Drives Selection of Diverse T Cell Receptor alpha and beta Repertoires.* mBio, 2020. **11**(2).
210. Chen, G., et al., *Sequence and Structural Analyses Reveal Distinct and Highly Diverse Human CD8(+) TCR Repertoires to Immunodominant Viral Antigens.* Cell Rep, 2017. **19**(3): p. 569-583.
211. Trautmann, L., et al., *Selection of T cell clones expressing high-affinity public TCRs within Human cytomegalovirus-specific CD8 T cell responses.* J Immunol, 2005. **175**(9): p. 6123-32.
212. Miles, J.J., et al., *Antigen-driven patterns of TCR bias are shared across diverse outcomes of human hepatitis C virus infection.* J Immunol, 2011. **186**(2): p. 901-12.
213. Song, I., et al., *Broad TCR repertoire and diverse structural solutions for recognition of an immunodominant CD8(+) T cell epitope.* Nat Struct Mol Biol, 2017. **24**(4): p. 395-406.

214. Mölder, F., et al., *Rapid T cell receptor interaction grouping with ting*. 2020, bioRxiv.
215. Zhang, H., et al., *Investigation of Antigen-Specific T-Cell Receptor Clusters in Human Cancers.* Clin Cancer Res, 2020. **26**(6): p. 1359-1371.
216. Yost, K.E., et al., *Clonal replacement of tumor-specific T cells following PD-1 blockade.* Nat Med, 2019. **25**(8): p. 1251-1259.
217. Oeppen, J. and J.W. Vaupel, *Demography. Broken limits to life expectancy.* Science, 2002. **296**(5570): p. 1029-31.
218. Bank, W. *Population ages 65 and above (% of total population)* World Development Indicators 2019 19.07.2020]; Available from: https://data.worldbank.org/indicator/SP.POP.65UP.TO.ZS?locations=DE-FR-US-IT-PL-PT-ES-AT-CH-SE-NO-FI.
219. DeSantis, C.E., et al., *Cancer statistics for adults aged 85 years and older, 2019.* CA Cancer J Clin, 2019. **69**(6): p. 452-467.
220. Goronzy, J.J. and C.M. Weyand, *Understanding immunosenescence to improve responses to vaccines.* Nat Immunol, 2013. **14**(5): p. 428-36.
221. Teresa Aguado, M., et al., *Report on WHO meeting on immunization in older adults: Geneva, Switzerland, 22-23 March 2017.* Vaccine, 2018. **36**(7): p. 921-931.
222. Simnica, D., et al., *T cell receptor next-generation sequencing reveals cancer-associated repertoire metrics and reconstitution after chemotherapy in patients with hematological and solid tumors.* Oncoimmunology, 2019. **8**(11): p. e1644110.
223. Nikolich-Zugich, J., *The twilight of immunity: emerging concepts in aging of the immune system.* Nat Immunol, 2018. **19**(1): p. 10-19.
224. Messaoudi, I., et al., *Direct link between mhc polymorphism, T cell avidity, and diversity in immune defense.* Science, 2002. **298**(5599): p. 1797-800.
225. Davenport, M.P., D.A. Price, and A.J. McMichael, *The T cell repertoire in infection and vaccination: implications for control of persistent viruses.* Curr Opin Immunol, 2007. **19**(3): p. 294-300.
226. Chen, H., et al., *TCR clonotypes modulate the protective effect of HLA class I molecules in HIV-1 infection.* Nat Immunol, 2012. **13**(7): p. 691-700.
227. Boyd, S.D., et al., *Human lymphocyte repertoires in ageing.* Curr Opin Immunol, 2013. **25**(4): p. 511-5.
228. Yager, E.J., et al., *Age-associated decline in T cell repertoire diversity leads to holes in the repertoire and impaired immunity to influenza virus.* J Exp Med, 2008. **205**(3): p. 711-23.
229. Naylor, K., et al., *The influence of age on T cell generation and TCR diversity.* J Immunol, 2005. **174**(11): p. 7446-52.
230. Dupic, T., et al., *Immune fingerprinting.* bioRxiv, 2020: p. 2020.06.24.168484.
231. Siegrist, C.-A., *2 - Vaccine immunology*, in *Vaccines (Sixth Edition)*, S.A. Plotkin, W.A. Orenstein, and P.A. Offit, Editors. 2013, W.B. Saunders: London. p. 14-32.
232. Panagioti, E., et al., *Features of Effective T Cell-Inducing Vaccines against Chronic Viral Infections.* Front Immunol, 2018. **9**: p. 276.
233. Finn, O.J., *A Believer's Overview of Cancer Immunosurveillance and Immunotherapy.* J Immunol, 2018. **200**(2): p. 385-391.
234. Sasson, S.C., et al., *Local heroes or villains: tissue-resident memory T cells in human health and disease.* Cell Mol Immunol, 2020. **17**(2): p. 113-122.
235. Simnica, D., et al., *High-Throughput Immunogenetics Reveals a Lack of Physiological T Cell Clusters in Patients With Autoimmune Cytopenias.* Front Immunol, 2019. **10**: p. 1897.
236. Schultheiss, C., et al., *Next-generation immunosequencing reveals pathological T cell architecture in autoimmune hepatitis.* Hepatology, 2020.
237. Sakaguchi, S., et al., *FOXP3+ regulatory T cells in the human immune system.* Nat Rev Immunol, 2010. **10**(7): p. 490-500.
238. Wilson, D.B., et al., *Specificity and degeneracy of T cells.* Mol Immunol, 2004. **40**(14-15): p. 1047-55.
239. Smakaj, E., et al., *Benchmarking immunoinformatic tools for the analysis of antibody repertoire sequences.* Bioinformatics, 2020. **36**(6): p. 1731-1739.

240. Cowell, L.G., *The Diagnostic, Prognostic, and Therapeutic Potential of Adaptive Immune Receptor Repertoire Profiling in Cancer.* Cancer Res, 2020. **80**(4): p. 643-654.
241. Piovesan, A., et al., *On the length, weight and GC content of the human genome.* BMC Res Notes, 2019. **12**(1): p. 106.
242. Kleiveland, C.R., *Peripheral Blood Mononuclear Cells*, in *The Impact of Food Bioactives on Health: in vitro and ex vivo models*, K. Verhoeckx, et al., Editors. 2015: Cham (CH). p. 161-167.
243. Corrie, B.D., et al., *iReceptor: A platform for querying and analyzing antibody/B-cell and T-cell receptor repertoire data across federated repositories.* Immunol Rev, 2018. **284**(1): p. 24-41.
244. Miho, E. and S.T. Reddy. *SystimsDB*. [cited 2020; This is a database of preprocessed and annotated CDR3 clones from B and T cell receptor repertoire NGS data. The data present here is collected from publicly available raw sequences of immune repertoires analyzed in scientific articles. The database is searchable and results are downloadable as tabular and graphical output.]. Available from: https://www.systimsdb.ethz.ch.
245. Kovaltsuk, A., et al., *Observed Antibody Space: A Resource for Data Mining Next-Generation Sequencing of Antibody Repertoires.* J Immunol, 2018. **201**(8): p. 2502-2509.
246. Ohlin, M., et al., *Inferred Allelic Variants of Immunoglobulin Receptor Genes: A System for Their Evaluation, Documentation, and Naming.* Front Immunol, 2019. **10**: p. 435.
247. Swindells, M.B., et al., *abYsis: Integrated Antibody Sequence and Structure-Management, Analysis, and Prediction.* J Mol Biol, 2017. **429**(3): p. 356-364.
248. Chailyan, A., A. Tramontano, and P. Marcatili, *A database of immunoglobulins with integrated tools: DIGIT.* Nucleic Acids Res, 2012. **40**(Database issue): p. D1230-4.
249. Ferdous, S. and A.C.R. Martin, *AbDb: antibody structure database-a database of PDB-derived antibody structures.* Database (Oxford), 2018. **2018**.
250. Gowthaman, R. and B.G. Pierce, *TCR3d: The T cell receptor structural repertoire database.* Bioinformatics, 2019. **35**(24): p. 5323-5325.
251. Bagaev, D.V., et al., *VDJdb in 2019: database extension, new analysis infrastructure and a T-cell receptor motif compendium.* Nucleic Acids Res, 2020. **48**(D1): p. D1057-D1062.
252. Vita, R., et al., *The immune epitope database 2.0.* Nucleic Acids Res, 2010. **38**(Database issue): p. D854-62.
253. Tickotsky, N., et al., *McPAS-TCR: a manually curated catalogue of pathology-associated T cell receptor sequences.* Bioinformatics, 2017. **33**(18): p. 2924-2929.
254. Ostmeyer, J., et al., *Statistical classifiers for diagnosing disease from immune repertoires: a case study using multiple sclerosis.* BMC Bioinformatics, 2017. **18**(1): p. 401.
255. Ostmeyer, J., et al., *Biophysicochemical Motifs in T-cell Receptor Sequences Distinguish Repertoires from Tumor-Infiltrating Lymphocyte and Adjacent Healthy Tissue.* Cancer Res, 2019. **79**(7): p. 1671-1680.
256. Greiff, V., et al., *Learning the High-Dimensional Immunogenomic Features That Predict Public and Private Antibody Repertoires.* J Immunol, 2017. **199**(8): p. 2985-2997.
257. Priel, A., et al., *Network Representation of T-Cell Repertoire- A Novel Tool to Analyze Immune Response to Cancer Formation.* Front Immunol, 2018. **9**: p. 2913.

ATTACHEMENT

Table 1. List of the sequencing primers

		Primer	Sequence
IGH locus 1st PCR (inner)	FR1 fw	VH1-FR1_FW	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGGCCTCAGTGAAGGTCTCCTGCAAG
		VH2-FR1_FW	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGTCTGGTCCTACGCTGGTGAAACCC
		VH3-FR1_FW	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCTGGGGGGTCCCTGAGACTCTCCTG
		VH4-FR1_FW	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCTTCGGAGACCCTGTCCCTCACCTG
		VH5-FR1_FW	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCGGGGAGTCTCTGAAGATCTCCTGT
		VH6-FR1_FW	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTCGCAGACCCTCTCACTCACCTGTG
	FR1 rv	JH-consensus_RV	TGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCTTACCTGAGGAGACGGTGACC
TRB locus 1st PCR (inner)	TCRß-A/B fw	Vβ2_FW	ACACTCTTTCCCTACACGACGCTCTTCCGATCTAACTATGTTTTGGTATCGTCA
		Vβ4_FW	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCAtGATGTTCTGGTACCGTCAGCA
		Vβ5/1_FW	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCAGTGTGTCCTGGTACCAACAG
		Vβ6a/11_FW	ACACTCTTTCCCTACACGACGCTCTTCCGATCTAACCCTTTATTGGTACCGACA
		Vβ6b/25_FW	ACACTCTTTCCCTACACGACGCTCTTCCGATCTATCCCTTTaTTGGTACCgACAG
		Vβ6c_FW	ACACTCTTTCCCTACACGACGCTCTTCCGATCTAACCCTTTATTGGTATCAACAG
		Vβ7_FW	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCGCTATGTATTGGTACAAGCA
		Vβ8a_FW	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCTCCCTTTTCTGGTACAGACAGAC
		Vβ9_FW	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCGCTATGTATTGGTATAAACAG
		Vβ10_FW	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTTATGTTTACTGGTATCGTAAGAAGC
		Vβ11_FW	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCAAAATGTACTGGTATCAACAA
		Vβ12a/3/13a/15_FW	ACACTCTTTCCCTACACGACGCTCTTCCGATCTATACATGTACTGGTATCGACAAGAC
		Vβ13b_FW	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGGCCATGTACTGGTATAGACAAG
		Vβ13c/12b/14_FW	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGTATATGTCCTGGTATCGACAAGA
		Vβ16_FW	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTAAtCTTTATTGGTATCGACGTGT
		Vβ17_FW	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGGCCATGTACTGGTACCGACA
		Vβ18_FW	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTCATGTTTACTGGTATCGGCAG
		Vβ19_FW	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTTtTGTTTATTGGTATCAACAGAATCA
		Vβ21_FW	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTACCCTTTACTGGTACCGGCAG
		Vβ20_FW	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCAACCTATACTGGTACCGACA
		Vβ22_FW	ACACTCTTTCCCTACACGACGCTCTTCCGATCTATACTTCTATTGGTACAGACAAATCT
		Vβ23/8b_FW	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCACtGTCTACTGGTACCAGCA
		Vβ24_FW	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCGTCATGTACTGGTACCAGCA
	TCRß-A/B rv	Jβ1.1_RV	TGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCTTACCTACAACTGTGAGTCTGGTG
		Jβ1.2_RV	TGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCTTACCTACAACGGTTAACCTGGTC
		Jβ1.3_RV	TGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCTCACCTACAACAGTGAGCCAACTT
		Jβ1.4_RV	TGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCATACCCAAGACAGAGAGCTGGGTTC
		Jβ1.5_RV	TGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCTTACCTAGGATGGAGAGTCGAGTC
		Jβ1.6_RV	TGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCATACCTGTCACAGTGAGCCTG
		Jβ2.2_RV	TGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCTTACCCAGTACGGTCAGCCT
		Jβ2.6_RV	TGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCTCACCCAGCACGGTCAGCCT
		Jβ2.7_RV	TGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCTCACCTGTGACCGTGAGCCTG
		Jβ2.1_RV	TGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCCTTCTTACCTAGCACGGTGA
		Jβ2.3_RV	TGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCCCGCTTACCGAGCACTGTCA
		Jβ2.4_RV	TGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCCAGCTTACCCAGCACTGAGA
		Jβ2.5_RV	TGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGCCCGCGCTCACCGAGCACCA
2nd PCR (outer)	BC fw	Barcode_FW	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTC
	BC rv	Barcode_RV	CAAGCAGAAGACGGCATACGAGATAATCTTAGTGACTGGAGTTCAGACGTGTG

The primer sequences are colour coded to indicate annealing part on IGH (FR1 primer pool) or TRB (TCRß-A and –B pools) loci according to [83] (black; red letters indicate nucleotide changes compared to [83]), Illumina adapter sequencing primer sequence (green), Illumina adapter flow cell attachment sequence (blue) and barcode sequence (bordaux). fw, forward; rv, reverse. Adapted from [84, 85].