Next-generation-sequencing-spectratyping reveals public T-cell receptor repertoires in pediatric very severe aplastic anemia and identifies a β chain CDR3 sequence associated with hepatitis-induced pathogenesis

Pina F. I. Krell,[1,2] Susanne Reuther,[1] Ute Fischer,[1] Thomas Keller,[3] Stephan Weber,[3] Michael Gombert,[1] Friedhelm R. Schuster,[1] Corinna Asang,[1] Polina Stepensky,[4] Brigitte Strahm,[5] Roland Meisel,[1] Jens Stoye,[2] and Arndt Borkhardt[1]

[1]Department of Pediatric Oncology, Hematology and Clinical Immunology, University Children's Hospital, Medical Faculty, Heinrich Heine University, Duesseldorf, Germany; [2]Genome Informatics, Faculty of Technology and Center for Biotechnology (CeBiTec), Bielefeld University, Bielefeld, Germany; [3]Acomed Statistics, Leipzig, Germany; [4]Department of Pediatric Hematology-Oncology, Hadassah Hebrew University Medical Center, Jerusalem, Israel; [5]Division of Pediatric Hematology and Oncology, Department of Pediatrics and Adolescent Medicine, University of Freiburg, Freiburg, Germany

ABSTRACT

Current diagnostic approaches that characterize T-cell deficiency by analyzing diversity of T-cell receptor sequences effectuate limited informational gain about the actual restrictiveness. For deeper insight into T-cell receptor repertoires we developed next-generation-sequencing-spectratyping, which employs high coverage Roche/454 sequencing of T-cell receptor (β)-chain amplicons. For automated analysis of high-throughput-sequencing data, we developed a freely available software, the TCR profiler. Gene usage, length, encoded amino acid sequence and sequence diversity of the complementarity determining region 3 were determined and comprehensively integrated into a novel complexity score. Repertoires of $CD8^+$ T cells from children with idiopathic or hepatitis-induced very severe aplastic anemia (n=7), children two months after bone marrow transplantation (n=7) and healthy controls (children n=5, adults n=5) were analyzed. Complexity scores clearly distinguished between healthy and diseased, and even between different immune deficiency states. The repertoire of aplastic anemia patients was dominated by public (i.e. present in more than one person) T-cell receptor clonotypes, whereas only 0.2% or 1.9% were public in normal children and adults, respectively. The CDR3 sequence ASSGVGFSGANVLT was highly prevalent in 3 cases of hepatitis-induced anemia (15-32% of all sequences), but was only low expressed in idiopathic aplastic anemia (2-5%, n=4) or healthy controls (<1%). Fifteen high frequent sequences were present exclusively in aplastic anemia patients. Next-generation-sequencing-spectratyping allows in-depth analysis of T-cell receptor repertoires and their restriction in clinical samples. A dominating clonotype was identified in hepatitis-induced anemia that may be associated with disease pathogenesis and several aplastic-anemia-associated, putatively autoreactive clonotypes were sequenced.

Introduction

Very severe aplastic anemia (vSAA, <200 granulocytes/μl blood) is an acquired bone marrow failure syndrome resulting in pancytopenia.[1] The etiology is idiopathic in the majority of cases,[1] but a subset of patients develops SAA a few months after an episode of acute hepatitis.[2,3] Pathological autoimmune responses are causative for the bone marrow failure in both types of SAA. Oligoclonally expanded populations of cytotoxic T cells diminish the patient's own hematopoietic $CD34^+$ stem cells.[1] The T-cell receptor (TCR) of the autoreactive T cells recognizes self-antigens presented by $CD34^+$ cells. The antigenic determinant is still unknown and may differ between patients. In hepatitis-associated SAA $CD34^+$ cells may present secondary targets of an immune response initially targeting the liver.[4]

A normal adult polyclonal T-cell compartment comprises an estimated number of 2.5×10^7 different T-cell clones each expressing a unique T-cell receptor.[5] Children harbor higher absolute numbers of T cells and express a more diverse TCR repertoire than adults. During aging both the number of T cells as well as the diversity decreases.[6,7] In 95% of T cells the TCR is a heterodimer composed of an and a protein chain that contain a constant and a variable region. Within the variable region the hypervariable, complementarity-determining region 3 (CDR3) is involved in antigen recognition. Its diversity is generated by somatic recombination of variable (V), joining (J) and in case of the β-chain also diversity (D) genes, termed V(D)J recombination. Random trimming and addition of non-template nucleotides at the junction sites (N-diversity mechanisms) greatly increases diversity further.[8,9] Theoretically, a repertoire of approximately 10^{18} different TCRs could be generated in humans.

Allelic exclusion leads to the expression of only one TCR β-chain protein in mature T cells.[10] Therefore, TCR β-chains provide a molecular fingerprint of T-cell clones. 48 different

 Haematologica 2013;98. doi:10.3324/haematol.2012.069708.
The online version of this paper has a Supplementary Appendix. PFIK, SR, UF, TK contributed equally to this manuscript.
Manuscript received on May 7, 2012. Manuscript accepted on May 2, 2013.
Correspondence: ute.fischer@med.uni-duesseldorf.de

functional TRBV, 13 TRBJ and 2 TRBD genes and additionally 1 TRBJ and 18 TRBV pseudogenes can potentially be rearranged. Commonly employed TCR repertoire determination by flow cytometric analysis[11] is thus strongly restricted by the availability and the specificity of anti-TRB antibodies and no TCR sequence information can be gained. An immunoscope technique, CDR3 spectratyping that analyzes CDR3 length polymorphisms, is currently more widely used.[8,12-16] This technique is limited by the fact that it requires numerous PCR reactions employing primers targeting all possible V and J genes for the amplification of individual β-chains. Delayed data availability, inability to detect diversity of TRB-chain sequences of equal length and the need to clone PCR products to obtain sequence information are further drawbacks.

More comprehensive approaches that accurately assess TCR diversity in patient samples could present powerful diagnostic tools. Important applications are monitoring of immune reconstitution, disease monitoring, control of therapeutic responses and the identification of pathogenic T-cell clonotypes in autoimmune disease.

Here, we developed next-generation-sequencing-spectratyping (NGS-S) that employs high coverage, massive parallel Roche/454-sequencing of TRB-chain amplicons as a fast and accurate method to determine T-cell diversity in clinical samples. For automated analysis of the generated NGS-S data we developed a novel software that allowed us to determine, for each TRB chain analyzed, the exact length and sequence of the CDR3, to identify the rearranged TRBV, TRBD and TRBJ genes, to analyze non-template nucleotides added at the junction sites and to exclude non-functional transcripts. Employing this technique we investigated the impairment of the T-cell compartment in pediatric patients with idiopathic and hepatitis induced vSAA and analyzed the expansion of specific potentially autoreactive T-cell clonotypes in these patients.

Design and Methods

Patients and controls

T-cell diversity of 7 children diagnosed with vSAA as previously described[17,18] and enrolled in the multicenter BFM-SAA 94 therapy study was prospectively analyzed (Table 1). The TCR repertoire of 7 age-matched children who underwent haploidentical bone marrow transplantation was investigated 2-4 months after transplantation. The conditioning regimen has been described previously.[19] The control group consisted of 5 healthy adult donors and 5 healthy age-matched pediatric donors. The local ethics committee approved the research and all participants or their parents gave written informed consent.

Sample preparation

Mononuclear cells were derived from bone marrow (all pediatric patients and controls) or peripheral blood (all healthy adult donors) by Ficoll (Biochrom AG, Berlin, Germany) density centrifugation. $CD8^+$ T lymphocytes were selected by using magnetic beads (Dynabeads, Dynal Biotech ASA, Oslo, Norway). Total RNA of 10^5 cells was extracted using TRIzol reagent (Invitrogen, Carlsbad, CA, USA) and complementary DNA was synthesized by standard procedures.

Spectratype analysis

Spectratyping was performed as described previously.[20] A brief description is provided in the *Online Supplementary Design and Methods* section.

Amplicon library preparation and next-generation-sequencing of TCR β-chains

Each amplicon library was generated by a single PCR. cDNA was amplified using the ACCUPrime Supermix I (Invitrogen, Carlsbad, CA, USA). Two custom-made degenerated V region primer mixtures (VP1: GCIITKTIYTGGTAYMGACA, VP2: CTITKTWTTGGTAYCIKCAG, 0.2 pmol each), designed to cover most TRBV genes and a reverse primer (CP1: GCACCTCCTTC-CCATTCAC, 0.2 pmol) covering both TRBC genes (*Online Supplementary Figure S1B*) were used.[21] PCR products were purified using the Gel Extraction Kit (Qiagen, Hilden, Germany) and high-throughput sequencing was performed according to the manufacturer's instructions (Roche/454, Basel, Switzerland).

CDR3 coverage and error estimation are described in the Online *Supplementary Design and Methods* section.

Analysis of T-cell diversity and development of a novel score to present TCR diversity

Individual amplicons were characterized by alignment-based identification of the rearranged germ line genes (TRBV, TRBD and TRBJ) and by determination of the CDR3 length, sequence, productivity and diversity (*Online Supplementary Figure S1C*). This information was used for the calculation of a next-generation-sequencing spectratype complexity score (CS_{NGS}), as described in the *Online Supplementary Appendix*. The TCR profiler software tool and a User's Manual are available from: *http://bibiserv.techfak.uni-bielefeld.de/tcrprofiler.*

Nomenclature

TCR β-chain germ line reference sequences were derived from the IMGT/GENE-DB database (*http://imgt.cines.org*).[22] The nomenclature provided by IMGT was used for the designation of TRBV, TRBD and TRBJ genes.[23,24] Correspondence between different TCR gene nomenclatures is available from:

http://imgt.org/IMGTrepertoire/index.php?section=LocusGenes&repertoire=nomenclatures&species=human&group=TRBV

Results

Skewing of complementarity determining region 3 length polymorphisms determined by next-generation-sequencing-spectratyping

The TCR repertoire of children at diagnosis of vSAA (n=7) was analyzed by next-generation-sequencing-spectratyping (NGS-S) as well as conventional CDR3 length spectratyping and compared to children who underwent haploidentical BMT (n=7) and healthy persons (adults (n=5) and children (n=5)) (Table 1). The CDR3 length spectratyping approach used analyzed polymorphic PCR products of 26 individual TRBV possibly rearranged with 13 TRBJ genes. By counting the number of fluorescent peaks generated after separation of PCR products on a gel matrix, the CDR3 length polymorphisms for each TRBV/TRBJ combination could be investigated (Figure 1A, upper panels). Skewing of CDR3 sequence length was observed in both the vSAA and the hBMT patient cohort, but hardly at all in healthy control cohorts. Complexity scores based on spectratyping results ($CS_{SPECTRATYPING}$) were obtained for each patient and healthy adult individuals included in this study (Table 1). Lowest scores indicating lowest complexity were derived for samples from patients after bone marrow transplantation (41±21, mean value±standard deviation), followed by samples from vSAA patients (61±44). Within the SAA cohort, scores

were lower for HAAA (29±4) than for IAA patients (86±44). Remarkably, 2 IAA patients (P6 and P7) not responding to immunosuppressive therapy had almost normal spectratyping complexity scores (124 and 123, respectively). This finding corresponds well to our earlier observations.[20] Based on CSSPECTRATYPING, a distinction between diseased cohorts was not possible because of the associated high standard deviations. Healthy adult controls demonstrated high complexity scores with low standard deviation (127±8). As described earlier,[20] bone marrow from healthy children had a high $CS_{SPECTRATYPING}$ similar to adult controls.

For TCR profiling using NGS-S (Figure 1A, lower panels), the CDR3 region was defined as starting at the first codon after the cysteine residue encoded by the CASS motif of the rearranged TRBV gene and ending with the phenylalanine of the conserved FGXG amino acid sequence motif in the TRBJ gene. The length of productive CDR3 sequences was analyzed for every TRBV/TRBD/TRBJ gene combination. Skewing of the CDR3 length distribution can be evaluated for the recombination of each gene separately or combined over all TRBV and TRBJ genes as an average distribution (Figure 2A). The length of the CDR3 sequences varied from 24 to 60 nucleotides coding for 8 to 20 amino acids in all cohorts. Individual-specific and average cohort analyses of CDR3 lengths (Figure 2A) in general show a bell-shape like distribution for healthy controls. Most CDR3 sequences in the healthy adult cohort were 12 amino acids long. This shifted to an average length of 15 amino acids in the pediatric control cohort. Patients after hBMT and vSAA patients showed a clear skewing of the CDR3 distribution. The CDR3 sequence length in vSAA patients reached a maximum at 16 amino acids. This was due to a dominant expansion in the HAAA patients P1, P2 and P3 only; the IAA patients had an almost normal CDR3 length profile.

T-cell pathologies are reliably distinguished by application of a new complexity score to next-generation-sequencing data

To analyze and compare the severity of TCR repertoire restriction in different T-cell pathologies we developed several new TCR complexity scores based on deep sequencing (*Online Supplementary Design and Methods*). Similar to the $CS_{SPECTRATYPING}$ score generated by analysis of conventional spectratyping results, the developed CDR3 Length Complexity Score CSL is a measure of the CDR3 length polymorphisms detected by NGS-S in each TRBV family. The CSL score readily distinguished samples from healthy (167±21, mean value±standard deviation) and diseased (vSAA: 116±15; hBMT: 111±25) individuals, but could not clearly differentiate between the two cohorts with constriction in their T-cell repertoire. To enhance discrimination between children with different T-cell pathologies, we developed the Combination Complexity Score CSC that was initially based on combinations of different TRBV and TRBJ genes identified in each sample (Figure 1B). However, we found a very high interrelationship between CSC and CSL, indicating that the numbers of different gene combinations and CDR3 length polymorphisms detected in each sample are dependent on one another. Thus, the CSC score provided basically the same information as CSL. Robust and com-

Table 1. Characteristics of the patients enrolled in TCR repertoire profiling.

Patient	Gender	Age (years)	Diagnosis	Treatment	Sample	$CS_{SPEC-TRA-TYPING}$	Leukocytes (per μL)	$CD3^+/CD4^+/CD8^+$ (per μL)
P1	m	1	vSAA, hepatitis-associated	IST, R	BM	32	600	NA
P2	m	14	vSAA, hepatitis-associated	IST, R	BM	30	1500	NA
P3	m	10	vSAA, hepatitis-associated	IST, R	BM	24	NA	NA
P4	m	10	vSAA, idiopathic	IST, R	BM	56	2100	NA
P5	m	14	vSAA, idiopathic	IST, R	BM	39	1400	NA
P6	m	3	vSAA, idiopathic	IST, NR	BM	124	NA	NA
P7	m	5	vSAA, idiopathic	IST, NR	BM	123	1000	NA
P8	f	14	AUL, del(12p)	hBMT	BM	39	NA	408/48/328
P9	f	6	AML, t(6;11)	hBMT	BM	28	NA	1680/150/1440
P10	m	12	AML	hBMT	BM	39	NA	151/34/95
P11	f	1	AML, t(7;12)	hBMT	BM	82	NA	412/57/341
P12	m	14	ALL	hBMT	BM	19	NA	168/163/4
P13	m	11	ALL, t(12;21)	hBMT	BM	28	NA	NA
P14	f	17	T-ALL	hBMT	BM	53	NA	NA
N15-N19	NA	> 18 years	Healthy adults	Blood bank donors	PB	127 ± 8	Norm values: 4500-10000	Norm values: 1450/830/560
N20-N24	m f f f m	6 8 3 5 12	Healthy children	Matched sibling donors	BM	Norm values: Schuster *et al.*[20]	Norm values: 4500-30000	Norm values: 1800-4050 980-2530 720-1390

All patients were pediatric patients, either newly diagnosed with very severe aplastic anemia (vSAA, n=7) or undergoing haploidentical bone marrow transplantation (hBMT, n=7). vSAA patients received immunosuppressive therapy (IST). Response (R) or non-response (NR) to this treatment was evaluated after 6 months. Bone marrow samples for TCR repertoire profiling were collected at diagnosis of vSAA or two to three months after hBMT (when expanded T cells were detectable), respectively. Normal controls include bone marrow samples from age-matched children and peripheral blood samples from healthy adults (m: male; f: female; AML: acute myeloid leukemia; ALL: acute lymphoblastic leukemia; AUL: acute undifferentiated leukemia; TALL: T-lineage acute lymphoblastic leukemia; del: deletion; t: translocation; BM: bone marrow; PB: peripheral blood; CSSPECTRATYPING complexity score derived from conventional CDR3 spectratype analysis; NA: not available).

plete separation of the diseased cohorts was achieved when the information of both scores was combined and diversity was further addressed on the basis of the amino acid sequence of individual CDR3 sequences (Figure 2B and *Online Supplementary Figure S2*). The calculated novel CS_{NGS} complexity score showed that the repertoire of healthy children is more complex compared to healthy adults (Figure 2B, left panel). CS_{NGS} further showed significant differences not only between healthy and diseased individuals, but also between children with vSAA and after hBMT (Figure 2B, middle panel), and also discriminated between vSAA cohorts with a different pathogenesis (HAAA and IAA, Figure 2B right panel). CS_{NGS} values were significantly higher for children suffering from vSAA than for children during immune reconstitution after hBMT. This indicates that the constraint of the T-cell compartment was more severe in the children three months after hBMT than in the children newly diagnosed with vSAA. The comparison of patients with hepatitis-induced *versus* idiopathic vSAA indicated that the T-cell repertoire was more restricted in IAA patients.

Restricted usage of variable and joining T-cell receptor β-chain genes in severe aplastic anemia

The identification of recombined V genes in a next-generation-sequencing approach can be hampered due to the shortness of sequenced reads, the high homology of TRBV genes, and the addition of non-template nucleotides at the gene boundaries.[25] Thus, we employed a previously reported method to generate amplicon libraries that span the entire Vβ-chain.[21] After long-read next-generation-sequencing of these libraries, we could unambiguously identify every TRBV gene rearranged in each of the identified CDR3 sequences. We amplified and sequenced most of the 48 possible TRBV genes (Figure 1B). The overview of TRBV sequences recombined with each TRBJ gene detected clearly

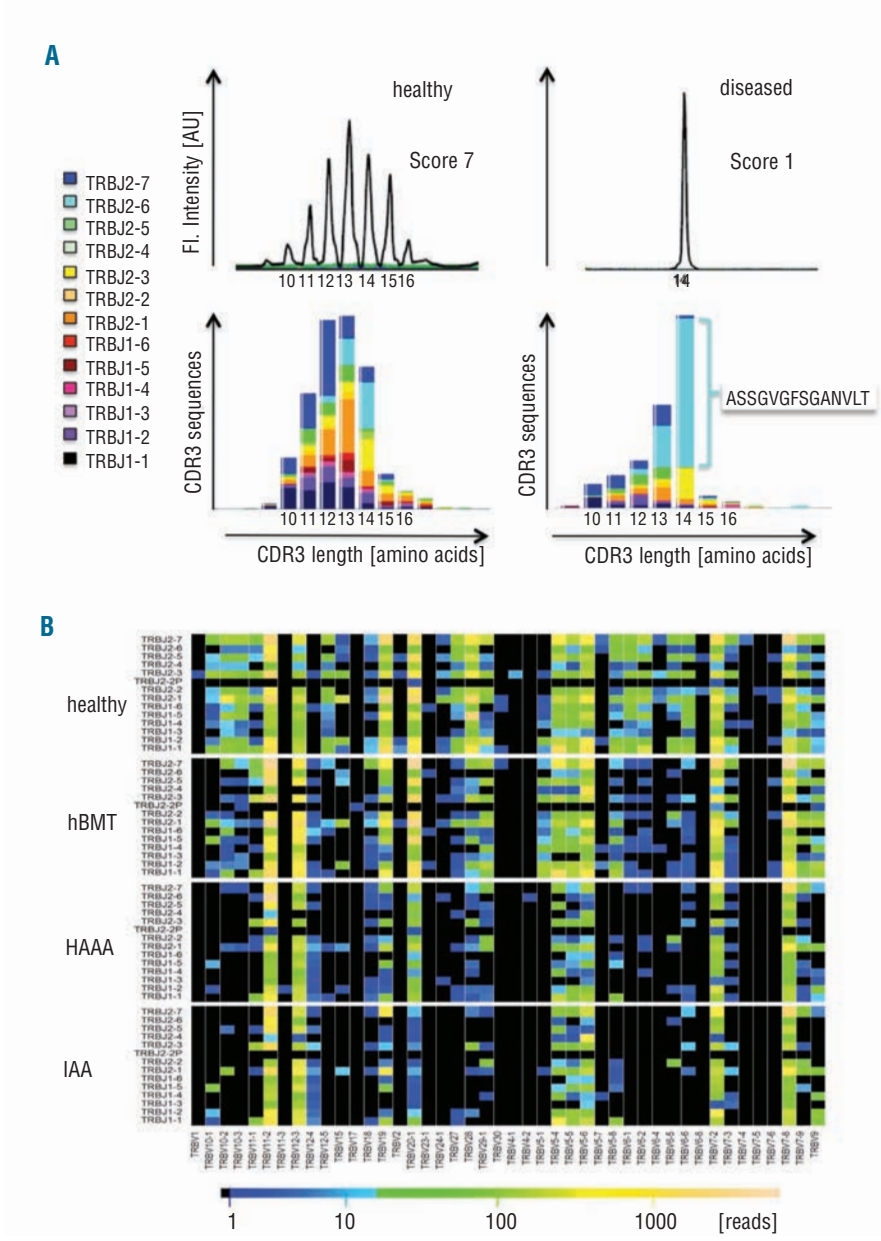

Figure 1. Analysis of TCR diversity of healthy controls, vSAA and hBMT patients by conventional spectratyping and NGS-S. (A) Upper panels: Representative spectratypes of a rearranged TRBV gene. cDNA from $CD8^+$ lymphocytes derived from a healthy control (left) and a diseased patient (right) was amplified by PCR using a gene family specific TRBV and a TRBC primer. Run-off reactions were performed employing fluorescently labeled TRBJ primers and analyzed on an ABI 3730 DNA sequencer. Fluorescence intensity in arbitrary units (AU) and the size of the products in amino acids are shown and were analyzed using GeneMapper v.4.0 software (life technologies, Grand Island, NY, USA). In samples derived from healthy persons (left) the length distribution of CDR3 sequences typically assumes a bell shaped curve with the highest peak formed by DNA fragments of 39 nucleotides (translated in 13 amino acids). Due to a skewed TCR repertoire the sample derived from a vSAA patient (right) presents only one predominant peak. The individual score for these single spectratypes is indicated. The spectratype complexity score $CS_{SPECTRATYPE}$ is generated by summing up all scores for each TRBV gene. Lower panels: patient samples were amplified using two degenerated wobble TRBV primer mixtures and a TRBC primer. After next-generation-sequencing the TCR profiler identifies the CDR3 length, the rearranged TRBJ genes, the nucleotide sequence and analyzes the functionality of the amplified CDR3. The total number of CDR3 sequences analyzed, the length in nucleotides, and the TRBJ gene rearranged with TRBV11-2 are presented. In the vSAA sample a CDR3 comprising TRBV11-2, TRBD2 and TRBJ2-6 was highly amplified indicating clonal expansion. The deduced amino acid sequence is shown. (B) Frequencies of specific TRBV and TRBJ gene pairing in healthy controls, vSAA and hBMT patients are depicted. TRBJ genes are indicated on the left, TRBV genes at the bottom of the panels. Total reads of a given pairing are indicated by color code. BMT, Hepatitis-induced vSAA (HAAA) and idiopathic vSAA (IAA) patients show restricted usage of TRBV and TRBJ genes.

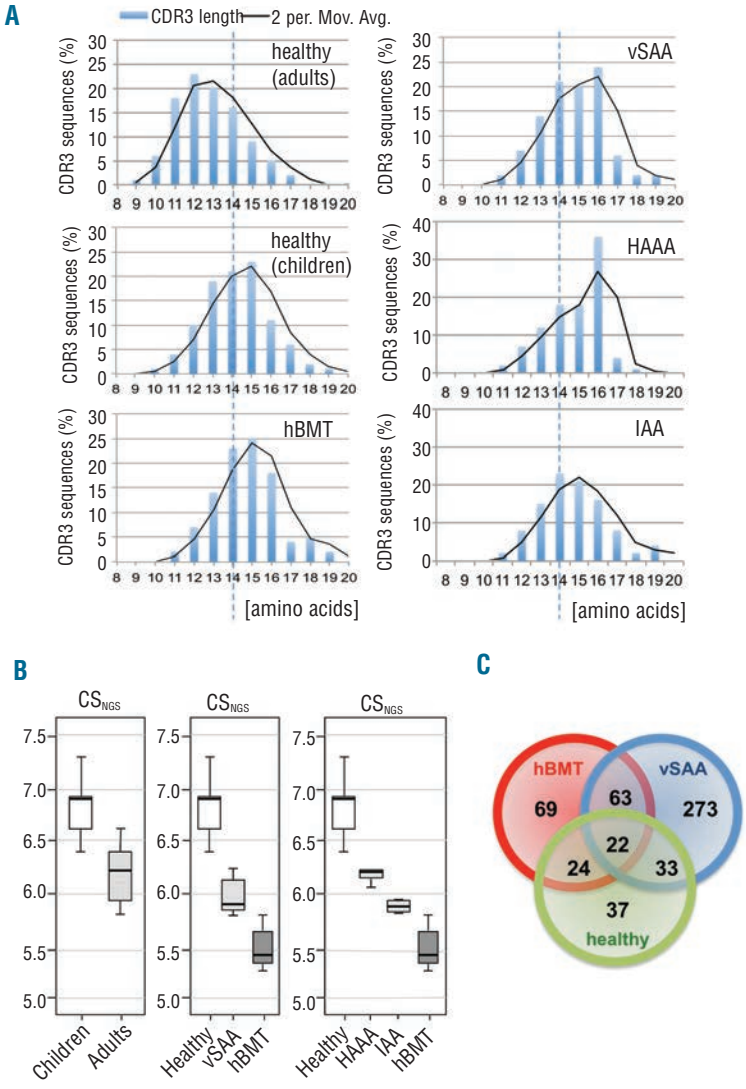

Figure 2. (A) Analysis of CDR3 length polymorphism in the analyzed cohorts by NGS-S. The percentage of CDR3 sequences of different length (presented in amino acids) is depicted as average values for the respective cohorts. The healthy cohorts show typical bell-shaped distributions of CDR3 sequences of different length with a maximum value at 12 amino acids (healthy adults) and 15 amino acids (healthy children), respectively. The CDR3 distribution of vSAA (all SAA patients), HAAA, IAA and hBMT pediatric patients is skewed. The HAAA cohort shows a shift of the maximum to longer sequences. (B) Comparison of complexity scores based on NGS-S. Box plots present the results gained by scoring TCR diversity in the five analyzed cohorts: healthy controls (adults and children), vSAA patients combined or split into the two subgroups HAAA and IAA, as well as hBMT patients. The left panel presents the NGS-S-based score (CS_{NGS}) results for the comparison of healthy adults and children. The middle panel depicts C_{SNGS} for the healthy children, the combined vSAA cohort and the hBMT patients. The right panel presents in addition the scores for the HAAA and IAA patients, separately. C_{SNGS} achieved a robust separation not only between healthy and diseased, but also between all cohorts with different T-cell pathologies. (Distribution and median values are shown.) (C) Expansion of public CDR3 amino acid sequences in vSAA patients. The Venn diagram presents the total number of CDR3 amino acid sequences that were shared between the different groups.

shows the reduced complexity of the TCR repertoires from children with vSAA (hepatitis induced as well as idiopathic) and children after hBMT compared to the healthy control group (Figure 1B). In samples derived from pediatric vSAA patients, TRBV11-2 was predominantly rearranged: 44% of all functional CDR3 sequences in HAAA patients, 26% in IAA, and 16-18% in healthy controls and hBMT patients, respectively. In the cohorts of healthy controls and children after hBMT, TRBV20-1 was rearranged most frequently. TRBV20-1 was used only in 3.9% of CDR3 sequences in the HAAA and in 0.4% in the IAA cohort.

Using the same approach, we further identified all 13 TRBJ genes. Use of the TRBJ genes was similar in all cohorts analyzed with a descending usage of TRBJ2-7, TRBJ2-1, TRBJ1-1 TRBJ2-3, 1-2 TRBJ2-5, TRBJ1-5 and TRBJ2-2 (*Online Supplementary Figure S3*). An expansion of sequences with rearranged TRBJ2-6 was obvious only in the patients P1, P2 and P3, who had acquired vSAA after an episode of hepatitis (Figure 3A and B). In the vSAA patient P2, these sequences accounted for 42% of all functional CDR3 sequences detected (vSAA P3: 30%, vSAA P1: 20%). The vast majority of these sequences had TRBJ2-6 rearranged to TRBV11-2 and coded for the amino acid sequence ASSGVGFS-GANVLT (P1: 15%, P2: 32%, P3: 24% of all productive CDR3 sequences) (Figure 3C). Such an expansion was not observed in the investigated vSAA patients who acquired their bone marrow failure without preceding hepatitis.

Patients suffering from very severe aplastic anemia share a large number of complementarity determining region 3 amino acid sequences

Most of the CDR3 amino acid sequences analyzed in healthy adult controls were non-public, i.e. so-called private sequences that were not shared with other individuals. Only 37 (1.9%) of all unique CDR3 amino acid sequences were shared between healthy adult individuals and even less between healthy and diseased individuals (Figure 2C). Healthy children shared only 0.2% of CDR3 sequences. None of the CDR3 amino acid sequences shared between healthy individuals, children or adults, were exclusively expressed in this cohort.

In contrast, T cells from vSAA patients expressed a large number of CDR3 sequences that were present in all individuals of this cohort: 273 individual sequences, 17.2% of all CDR3 amino acid sequences detected in this cohort. Within this cohort, HAAA patients shared more sequences with one another than IAA patients. Frequent sequences exclusively expressed in the vSAA

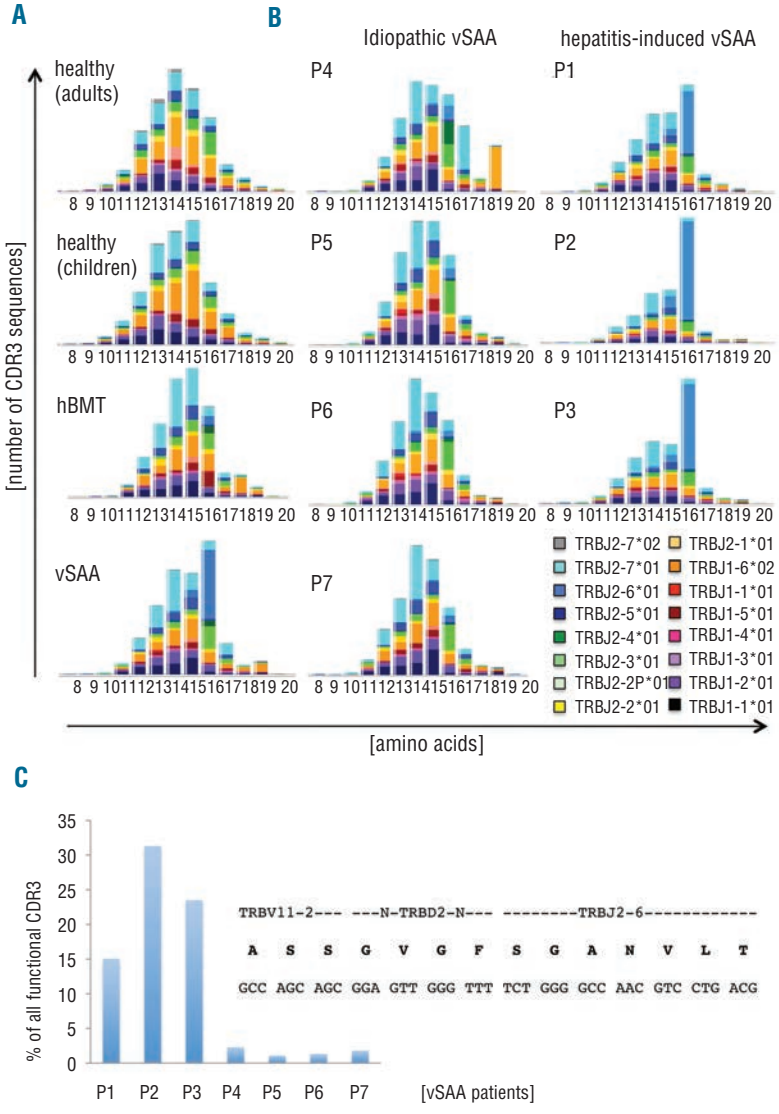

Figure 3. (A) Analysis of CDR3 length polymorphisms and TRBJ gene usage by NGS-S as an average for the respective cohort. For each cohort (healthy adults and children, hBMT, vSAA) included in the study CDR3 lengths determined by NGS-S are presented in amino acids on the abscissa. In addition the number of reads using specific TRBJ genes for each CDR3 length are annotated. The legend provides the color code for the respective TRBJ genes. (B) Patient specific analysis of CDR3 length and TRBJ gene usage in patients with hepatitis induced (P1 - P3) and idiopathic vSAA (P4 - P7) are shown. (C) A rearrangement of TRBV11-2, TRBD2 and TRBJ2-6 encodes the sequence ASSGVGFSGANVLT that is expressed in all analyzed cohorts ("*public*" sequence). The number of sequenced reads encoding the amino acid sequence ASSGVGFSGANVLT is presented in the bar diagram for individual vSAA patients (in % of all functional CDR3 sequences). Sequence expansion is most prominent in the vSAA patients P1, P2 and P3 with hepatitis-associated etiology.

cohort may be considered as candidate sequences relevant for the pathogenesis of the disease (Table 2). vSAA patients shared more CDR3 sequences with hBMT patients (4%) than with healthy controls (2%). Of the clonotypes shared between all hBMT patients (5.9%), most were also expressed in vSAA patients and none was specifically expressed in this cohort.

Oligoclonal expansions of public T-cell receptor sequences are predominant in patients with hepatitis-induced very severe aplastic anemia

Strikingly, the most abundantly expressed sequences in vSAA patients belonged to the group of 22 clonotypes that were truly public sequences, i.e. present in more than one person. These were commonly expressed in all cohorts (Figure 2C and *data not shown*). The clonotype ASSGVGFSGANVLT was most highly expanded in the subcohort of HAAA patients P1, P2 and P3. This sequence constitutes a public sequence that was frequently found in all healthy controls and constituted one of the top twelve expressed sequences in 4 of 5 healthy individuals. In addition to this sequence, we found several public clonotypes expressed in all vSAA patients but in none of the control cohorts (Table 3). Due to dominance of the ASSGVGFSGANVLT sequence, other CDR3 sequences are rarer in the HAAA than in the IAA cohort. The finding of a vSAA repertoire dominated by public clonotypes may indicate the generation of TCR protein surfaces that recognize one or more common, maybe permanently present, antigens.

Discussion

The TCR repertoire of patients with aplastic anemia has so far mainly been analyzed by conventional CDR3 length spectratyping or flow cytometry. Skewing of the distribution of CDR3 lengths in TRBV families has been observed in aplastic anemia using spectratyping.[26-30] However, a deviation from the bell-shaped distribution typical of healthy individuals is commonly found also in normal immune responses and has been reported to become more frequent with age.[31-35] When we analyzed our patient cohort of 7 children newly diagnosed with vSAA by spectratyping, the CDR3 length distribution was severely skewed only in 5 of 7 patients. Interestingly, the 2 patients with normal $CS_{SPECTRATYPING}$ complexity scores were not responding to immunosuppressive therapy. The

Table 2. CDR3 sequences present exclusively in vSAA patients.

CDR3 sequences	HAAA				IAA		
	P1	P2	P3	P4	P5	P6	P7
ASSPWRTQYF	0.24					0.20	
ASGSSYNEQFF	0.24	0.45					
ASSAGLSNYGYTF	0.30						
ASSFQVGSTEAFF		0.18					
ASSLSGTTYGYTF			0.19				
ASSSLNQPQHF			0.23				
SARDGLPSYNEQFF			0.51				
ASSLSGTTYGYTF				0.16			
ASSLANLVTQYF				0.21			
ASSLGGEQPQHF				0.26			
ASKAVRGGVHQETQYF				0.42		0.25	
ASSLGWSTEAFF					0.45		
ASSSGSYNSPLHF						0.22	
ASSLVSGTQYF						0.22	
ASSLGVGLNYGYTF						0.25	

Frequent CDR3 sequences exclusively detected in vSAA samples. Sequence and frequency (% of all productive CDR3 amino acid sequences per patient) are shown for each individual patient. These sequences were not detected in the healthy control cohorts or the cohort of transplanted children. HAAA: hepatitis associated aplastic anemia; IAA: idiopathic aplastic anemia; P1-7: patient numbers, see Table 1.

TCR repertoire is skewed during immune reconstitution after BMT[36-38] and this could be shown also in the cohort of children that we analyzed shortly after bone marrow transplantation. Large variations between the individuals of this cohort were observed. Conventional CDR3 spectratyping did not allow us to judge the severity of the T-cell constriction or to distinguish between a vSAA- or hBMT-induced defective repertoire. Furthermore, to obtain sequence information of potentially pathogenic T-cell clones, a multitude of individual PCR products would have to be cloned and sequenced.

With the advent of next-generation-sequencing technologies, it has now become feasible to directly sequence a massive number of TCR β chains in parallel. We applied NGS to TCR spectratyping and developed a high throughput method to analyze the TCR repertoire in clinical samples that we termed NGS-Spectratyping (NGS-S). For every patient sample, an average of 15,000 to 30,000 reads were sequenced and analyzed by our newly designed software. Currently, only 3844 mRNA sequences of the TCR β chain are published in GenBank and only a very few have already been associated with specific antigens. On average, we identified 3,800-9,000 unique functional CDR3 sequences per patient, most of them with an as yet unknown antigen specificity.

A side-by-side comparison with conventional CDR3 spectratyping showed that 'virtual' NGS-S based spectratypes contained all the information presented by conventional spectratypes. But NGS-S was clearly superior to spectratyping in terms of sensitivity. In many V gene families, virtual spectratyping revealed additional CDR3 sequences not detected by conventional spectratyping. A recent study evaluating TCR diversity in 2 healthy adults based on next-generation-sequencing and spectratyping similarly showed improved sensitivity using NGS.[39] Moreover, after implementing our newly developed CS_{NGS} score we were able to reliably distinguish different T-cell abnormalities from one another.

The strength of the long-sequencing read method used is the unambiguous identification of all rearranged TRBV and TRBJ genes due to the length of the accomplished sequencing reads. The combination of TRBV and TRBJ genes was not random and differed between the cohorts analyzed. Specific genes were used and recombined more frequently than others. Similar to a previous report,[25] we observed a dominance of the TRBV20-1 gene in healthy individuals but also in transplanted patients (24% of all CDR3 sequences in both cohorts). In contrast, only 2% of CDR3 sequences in vSAA patients used this gene. Some reports have observed a dominant usage of certain TRBV genes in SAA patients, such as TRBV629 or TRBV24.[40-42] Most obvious in the present study was a severely skewed CDR3 profile, and dominance and expansion of sequences using TRBV11-2 combined with TRBJ2-6 in the patients with hepatitis-associated vSAA that was not observed in patients with idiopathic pathogenesis. IAA patients had an almost normal CDR3 length profile in the NGS analysis and showed no expansion of CDR3 sequences using specific TRBV genes. All HAAA patients, but only 2 of the IAA patients, responded well to immunosuppressive therapy after 6-12 months. The pattern of TRBJ gene usage in our study was other-

Table 3. Overall and subgroup-specific average frequency of public TCR β chain sequences present in all SAA patients.

Public CDR3 sequences	vSAA (n=7)	HAAA (n=3)	IAA (n=4)
ASSGVGFSGANVLT	10.83	23.17	1.57
ASSASGIAYEQYF	4.07	3.54	4.47
ASTSLGSASTDTQYF	3.12	2.60	3.51
ASSTGQAYEQYF	1.69	1.51	1.82
ASSIDLWASASSYNEQFF	0.96	0.12	1.59
ASSLAADFQETQYF	0.89	0.73	1.01
ASSPRDRGLAGGEQYF	0.89	0.06	1.52
ASSSHRVPNTEAFF	0.92	0.66	1.11
ASSLAWREGGNIQYF	0.84	0.39	1.18

Frequent CDR3 amino acid sequences shared by all SAA cohorts (public sequences) are depicted. The frequency of these sequences in the cohorts of all SAA (vSAA), the HAAA and IAA patients, respectively, is given in % of all productively rearranged CDR3 sequences per cohort.

wise similar to previous reports describing T-cell compartments of healthy individuals using NGS-based CDR3 sequence analysis.[25,39] With the exception of the preferred TRBD2 gene usage in the HAAA patients associated with the expansion of the ASSGVGFS-GANVLT sequence, we did not detect any differences in TRBD gene usage between the cohorts.

Interestingly, patients at diagnosis of vSAA shared an unexpectedly high number of CDR3 amino acid sequences (17.2%). In a previous study, approximately 1% of the CDR3 sequences were shared between 2 healthy adults.[39] This is very similar to our results: 1.9% were shared between 5 healthy adult individuals and only 0.2% between healthy children under 13 years of age. Transplanted patients showed only a moderate rise in the percentage of common sequences. Such shared sequences have been suggested to be nearer to the germ line sequence.[39] Alternatively, they may be strongly selected by common or disease-associated antigens resulting in semi-public repertoires of certain patient groups.[30] In addition, we identified 15 high frequent CDR3 sequences only present in vSAA patients. As we focused our analyses on marrow-derived $CD8^+$ T cells, some of these CDR3 sequences may indeed identify T-cell clones involved in the autoimmune pathogenesis of the disease. It will be interesting to test these candidate pathological T cells in further *in vitro* studies by co-cultivation with target $CD34^+$ stem cells or *in vivo* after adoptive transfer in mice.

During many chronic infections, as well as cancer, T-cell dysfunction occurs because of an exhaustion of T cells.[43] Persistent antigen presence leads to a functional decline in antigen specific effector T cells and eventually to their depletion.[44-46] Exhausted cells lose the ability to produce large amounts of INFγ. Interestingly, low or absent production of this cytokine by bone marrow T cells was observed in some SAA patients and was predictive of a lack of response to immunosuppressive therapy.[47] Recently, we consistently found that $CD34^+$ stem cells of individual non-responding vSAA patients lacked expression of typical IFNγ stimulated genes.[48] Taking into account our current finding of a mainly public repertoire in vSAA patients, it may be speculated that non-responding SAA could alternatively be a disease of chronic immune stimulation finally leading to an exhaustion and depletion of antigen-specific private T cells. In this respect, it is interesting to observe that the 2 non-responding cases included in our study (P6 and P7) had a repertoire characterized by an even lower frequency of residual clonotypes and a lack of clonal expansions compared to responding patients. More patients should be analyzed in a prospective clinical study to draw firm conclusions about the value of NGS-S in predicting response.

Acknowledgments

The authors wish to thank the colleagues in pediatric oncology centers and all parents who gave their consent to use the biological material from their minors. AB was supported by grants from the "Bundesministerium für Forschung und Technologie" (BMBF).

Authorship and Disclosures

Information on authorship, contributions, and financial & other disclosures was provided by the authors and is available with the online version of this article at www.haematologica.org.

References

1. Young NS. Hematopoietic cell destruction by immune mechanisms in acquired aplastic anemia. Semin Hematol. 2000;37(1):3-14.
2. Lorenz E, Quaiser K. [Panmyelopathy following epidemic hepatitis]. Wien Med Wochenschr. 1955;105(1):19-22.
3. Brown KE, Tisdale J, Barrett AJ, Dunbar CE, Young NS. Hepatitis-associated aplastic anemia. N Engl J Med. 1997;336(15):1059-64.
4. Lu J, Basu A, Melenhorst JJ, Young NS, Brown KE. Analysis of T-cell repertoire in hepatitis-associated aplastic anemia. Blood. 2004;103(12):4588-93.
5. Arstila TP, Casrouge A, Baron V, Even J, Kanellopoulos J, Kourilsky P. A direct estimate of the human alphabeta T cell receptor diversity. Science. 1999;286(5441):958-61.
6. Bofill M, Janossy G, Lee CA, MacDonald-Burns D, Phillips AN, Sabin C, et al. Laboratory control values for CD4 and CD8 T lymphocytes. Implications for HIV-1 diagnosis. Clin Exp Immunol. 1992; 88(2):243-52.
7. Garderet L, Dulphy N, Douay C, Chalumeau N, Schaeffer V, Zilber MT, et al. The umbilical cord blood alphabeta T-cell repertoire: characteristics of a polyclonal and naive but completely formed repertoire. Blood. 1998;91(1):340-6.
8. Pannetier C, Cochet M, Darche S, Casrouge A, Zoller M, Kourilsky P. 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 USA. 1993;90(9):4319-23.
9. Cabaniols JP, Fazilleau N, Casrouge A, Kourilsky P, Kanellopoulos JM. Most alpha/beta T cell receptor diversity is due to terminal deoxynucleotidyl transferase. J Exp Med. 2001;194(9):1385-90.
10. Brady BL, Steinel NC, Bassing CH. Antigen receptor allelic exclusion: an update and reappraisal. J Immunol. 2010;185(7):3801-8.
11. Maciejewski JP, O'Keefe C, Gondek L, Tiu R. Immune-mediated bone marrow failure syndromes of progenitor and stem cells: molecular analysis of cytotoxic T cell clones. Folia Histochem Cytobiol. 2007;45 (1):5-14.
12. Prevost-Blondel A, Ostankovitch M, Melle J, Pannetier C, Macintyre E, Dreyfus F, et al. CDR3 size analysis of T cell receptor V beta transcripts: follow-up study in a patient with T cell acute lymphoblastic leukemia. Leukemia. 1995;9(10):1711-7.
13. Gorski J, Piatek T, Yassai M, Gorski J, Maslanka K. Improvements in repertoire analysis by CDR3 size spectratyping. Bifamily PCR. Ann N Y Acad Sci. 1995; 756(99-102.
14. Even J, Lim A, Puisieux I, Ferradini L, Dietrich PY, Toubert A, et al. T-cell repertoires in healthy and diseased human tissues analysed by T-cell receptor beta-chain CDR3 size determination: evidence for oligoclonal expansions in tumours and inflammatory diseases. Res Immunol. 1995; 146(2):65-80.
15. Gorochov G, Neumann AU, Kereveur A, Parizot C, Li T, Katlama C, et al. Perturbation of CD4+ and CD8+ T-cell repertoires during progression to AIDS and regulation of the CD4+ repertoire during antiviral therapy. Nat Med. 1998;4(2):215-21.
16. Mugnaini EN, Syversen AM, Sannes M, Freng A, Brinchmann JE. Normal CD4 T-cell receptor repertoire in tonsillar tissue despite perturbed repertoire in peripheral blood in HIV-1 infected individuals. Aids. 1999; 13(18):2507-13.
17. Fuhrer M, Rampf U, Baumann I, Faldum A, Niemeyer C, Janka-Schaub G, et al. Immunosuppressive therapy for aplastic anemia in children: a more severe disease predicts better survival. Blood. 2005;106(6): 2102-4.
18. Camitta BM, Rappeport JM, Parkman R, Nathan DG. Selection of patients for bone marrow transplantation in severe aplastic anemia. Blood. 1975;45(3):355-63.
19. Kolb HJ, Guenther W, Gyurkocza B, Hoetzl F, Simoes B, Falk C, et al. Tolerance and chimerism. Transplantation. 2003;75(9 Suppl):26S-31S.
20. Schuster FR, Hubner B, Führer M, Eckermann O, Gombert M, Dornmair K, et al. Highly skewed T-cell receptor V-beta chain repertoire in the bone marrow is associated with response to immunosuppressive drug therapy in children with very severe aplastic anemia. Blood Cancer J. 2011;1(3): e8.
21. Zhou D, Srivastava R, Grummel V, Cepok S, Hartung HP, Hemmer B. High throughput analysis of TCR-beta rearrangement and gene expression in single T cells. Lab Invest. 2006;86(3):314-21.
22. Giudicelli V, Chaume D, Lefranc MP. IMGT/GENE-DB: a comprehensive database for human and mouse immunoglobulin and T cell receptor genes. Nucleic Acids Res. 2005;33(Database issue):D256-61.
23. Folch G, Lefranc MP. The human T cell receptor beta variable (TRBV) genes. Exp Clin Immunogenet. 2000;17(1):42-54.
24. Folch G, Lefranc MP. The human T cell

receptor beta diversity (TRBD) and beta joining (TRBJ) genes. Exp Clin Immunogenet. 2000;17(2):107-14.
25. Freeman JD, Warren RL, Webb JR, Nelson BH, Holt RA. Profiling the T-cell receptor beta-chain repertoire by massively parallel sequencing. Genome Res. 2009;19(10):1817-24.
26. Kook H, Risitano AM, Zeng W, Wlodarski M, Lottemann C, Nakamura R, et al. Changes in T-cell receptor VB repertoire in aplastic anemia: effects of different immunosuppressive regimens. Blood. 2002; 99(10):3668-75.
27. Risitano AM, Kook H, Zeng W, Chen G, Young NS, Maciejewski JP. Oligoclonal and polyclonal CD4 and CD8 lymphocytes in aplastic anemia and paroxysmal nocturnal hemoglobinuria measured by V beta CDR3 spectratyping and flow cytometry. Blood. 2002;100(1):178-83.
28. Plasilova M, Risitano A, Maciejewski JP. Application of the molecular analysis of the T-cell receptor repertoire in the study of immune-mediated hematologic diseases. Hematology. 2003;8(3):173-81.
29. Zeng W, Maciejewski JP, Chen G, Young NS. Limited heterogeneity of T cell receptor BV usage in aplastic anemia. J Clin Invest. 2001;108(5):765-73.
30. Risitano AM, Maciejewski JP, Green S, Plasilova M, Zeng W, Young NS. In-vivo dominant immune responses in aplastic anaemia: molecular tracking of putatively pathogenetic T-cell clones by TCR beta-CDR3 sequencing. Lancet. 2004;364(9431): 355-64.
31. Gorski J, Yassai M, Zhu X, Kissela B, Kissella B, Keever C, et al. Circulating T cell repertoire complexity in normal individuals and bone marrow recipients analyzed by CDR3 size spectratyping. Correlation with immune status. J Immunol. 1994;152(10): 5109-19.
32. van den Beemd R, Boor PP, van Lochem EG, Hop WC, Langerak AW, Wolvers-Tettero IL, et al. Flow cytometric analysis of the Vbeta repertoire in healthy controls. Cytometry. 2000;40(4):336-45.
33. Posnett DN, Sinha R, Kabak S, Russo C. Clonal populations of T cells in normal elderly humans: the T cell equivalent to "benign monoclonal gammapathy". J Exp Med. 1994;179(2):609-18.
34. Shen DF, Doukhan L, Kalams S, Delwart E. High-resolution analysis of T-cell receptor beta-chain repertoires using DNA heteroduplex tracking: generally stable, clonal CD8+ expansions in all healthy young adults. J Immunol Methods. 1998;215(1-2):113-21.
35. Degauque N, Boeffard F, Foucher Y, Ballet C, Brouard S, Soulillou JP. The blood of healthy individuals exhibits CD8 T cells with a highly altered TCR Vb repertoire but with an unmodified phenotype. PLoS One. 2011;6(6):e21240.
36. Verfuerth S, Peggs K, Vyas P, Barnett L, O'Reilly RJ, Mackinnon S. Longitudinal monitoring of immune reconstitution by CDR3 size spectratyping after T-cell-depleted allogeneic bone marrow transplant and the effect of donor lymphocyte infusions on T-cell repertoire. Blood. 2000;95(12):3990-5.
37. Akatsuka Y, Cerveny C, Hansen JA. T cell receptor clonal diversity following allogeneic marrow grafting. Hum Immunol. 1996; 48(1-2):125-34.
38. Fallen PR, McGreavey L, Madrigal JA, Potter M, Ethell M, Prentice HG, et al. Factors affecting reconstitution of the T cell compartment in allogeneic haematopoietic cell transplant recipients. Bone Marrow Transplant. 2003;32(10):1001-14.
39. Robins HS, Campregher PV, Srivastava SK, Wacher A, Turtle CJ, Kahsai O, et al. Comprehensive assessment of T-cell receptor beta-chain diversity in alphabeta T cells. Blood. 2009;114(19):4099-107.
40. Nakao S, Takamatsu H, Yachie A, Itoh T, Yamaguchi M, Ueda M, et al. Establishment of a CD4+ T cell clone recognizing autologous hematopoietic progenitor cells from a patient with immune-mediated aplastic anemia. Exp Hematol. 1995;23(5):433-8.
41. Zeng W, Nakao S, Takamatsu H, Yachie A, Takami A, Kondo Y, et al. Characterization of T-cell repertoire of the bone marrow in immune-mediated aplastic anemia: evidence for the involvement of antigen-driven T-cell response in cyclosporine-dependent aplastic anemia. Blood. 1999;93(9): 3008-16.
42. Nakao S, Takami A, Takamatsu H, Zeng W, Sugimori N, Yamazaki H, et al. Isolation of a T-cell clone showing HLA-DRB1*0405-restricted cytotoxicity for hematopoietic cells in a patient with aplastic anemia. Blood. 1997;89(10):3691-9.
43. Wherry EJ. T cell exhaustion. Nat Immunol. 2011;12(6):492-9.
44. Lin MY, Welsh RM. Stability and diversity of T cell receptor repertoire usage during lymphocytic choriomeningitis virus infection of mice. J Exp Med. 1998;188(11):1993-2005.
45. Zajac AJ, Blattman JN, Murali-Krishna K, Sourdive DJ, Suresh M, Altman JD, et al. Viral immune evasion due to persistence of activated T cells without effector function. J Exp Med. 1998;188(12):2205-13.
46. Gallimore A, Glithero A, Godkin A, Tissot AC, Pluckthun A, Elliott T, et al. Induction and exhaustion of lymphocytic choriomeningitis virus-specific cytotoxic T lymphocytes visualized using soluble tetrameric major histocompatibility complex class I-peptide complexes. J Exp Med. 1998;187 (9):1383-93.
47. Sloand E, Kim S, Maciejewski JP, Tisdale J, Follmann D, Young NS. Intracellular interferon-gamma in circulating and marrow T cells detected by flow cytometry and the response to immunosuppressive therapy in patients with aplastic anemia. Blood. 2002; 100(4):1185-91.
48. Fischer U, Ruckert C, Hubner B, Eckermann O, Binder V, Bakchoul T, et al. CD34+ gene expression profiling of individual children with very severe aplastic anemia indicates a pathogenic role of integrin receptors and the proapoptotic death ligand TRAIL. Haematologica. 2012;97(9): 1304-11.