



VJ Segment Usage of TCR-Beta Repertoire in Monozygotic Cystic Fibrosis Twins

Sebastian Fischer^{1*}, Frauke Stanke^{1,2} and Burkhard Tümmler^{1,2}

¹ Clinic for Pediatric Pneumology, Allergology and Neonatology, Hannover Medical School, Hannover, Germany, ² Biomedical Research in Endstage and Obstructive Lung Disease Hannover (BREATH), German Center for Lung Research, Hannover Medical School, Hannover, Germany

OPEN ACCESS

Edited by:

Stephanie Gras,
Monash University, Australia

Reviewed by:

Ilgar Z. Mamedov,
Institute of Bioorganic Chemistry
(RAS), Russia
Emma Grant,
La Trobe University, Australia

*Correspondence:

Sebastian Fischer
fischer.sebastian@mh-hannover.de

Specialty section:

This article was submitted to
T Cell Biology,
a section of the journal
Frontiers in Immunology

Received: 26 August 2020

Accepted: 28 January 2021

Published: 23 February 2021

Citation:

Fischer S, Stanke F and Tümmler B
(2021) VJ Segment Usage of
TCR-Beta Repertoire in Monozygotic
Cystic Fibrosis Twins.
Front. Immunol. 12:599133.
doi: 10.3389/fimmu.2021.599133

Sixteen monozygotic cystic fibrosis (CF) twin pairs of whom 14 pairs were homozygous for the most common p.Phe508del *CFTR* mutation were selected from the European Cystic Fibrosis Twin and Sibling Study Cohort. The monozygotic twins were examined in their T cell receptor (TCR) repertoire in peripheral blood by amplicon sequencing of the CDR3 variable region of the β -chain. The recruitment of TCR J and V genes for recombination and selection in the thymus showed a strong genetic influence in the CF twin cohort as indicated by the shortest Jensen-Shannon distance to the twin individual. Exceptions were the clinically most discordant and/or most severely affected twin pairs where clonal expansion probably caused by recurrent pulmonary infections overshadowed the impact of the identical genomic blueprint. In general the Simpson clonality was low indicating that the population of TCR β clonotypes of the CF twins was dominated by the naïve T-cell repertoire. Intrapair sharing of clonotypes was significantly more frequent among monozygotic CF twins than among pairs of unrelated CF patients. Complete nucleotide sequence identity was observed in about 0.11% of CDR3 sequences which partially should represent persisting fetal clones derived from the same progenitor T cells. Complete amino acid sequence identity was noted in 0.59% of clonotypes. Of the nearly 40,000 frequent amino acid clonotypes shared by at least two twin siblings 99.8% were already known within the immuneACCESS database and only 73 had yet not been detected indicating that the CDR3 β repertoire of CF children and adolescents does not carry a disease-specific signature but rather shares public clones with that of the non-CF community. Clonotypes shared within twin pairs and between unrelated CF siblings were highly abundant among healthy non-CF people, less represented in individuals with infectious disease and uncommon in patients with cancer. This subset of shared CF clonotypes defines CDR3 amino acid sequences that are more common in health than in disease.

Keywords: T cell receptor repertoire, immunotyping, VJ usage, cystic fibrosis, twins, CDR3, TCRB

INTRODUCTION

Cystic fibrosis (CF) is a systemic disorder of exocrine glands that is caused by mutations in the *Cystic Fibrosis Transmembrane Conductance Regulator (CFTR)* gene (1, 2). The basic defect manifests in an impaired luminal transport of chloride and bicarbonate leading in the lungs to compromised innate immunity, airway obstruction by mucus plugging and inhomogeneous ventilation (3). Chronic airway infections are the major cause for morbidity and mortality of this life-limiting autosomal recessive trait. Triggered by recurrent viral infections, the upper and lower airways are colonized with bacterial opportunistic pathogens, namely *Staphylococcus aureus* and – later in life – *Pseudomonas aeruginosa* (4). Bacterial persistence in biofilms and immigration of immune effector cells sustain a vicious cycle of chronic infection, inflammation, lung damage and remodeling which leads to ongoing deterioration of pulmonary function and finally to respiratory failure (5).

Massive neutrophil influx and subsequent release of neutrophil extracellular traps are key events that result in the accumulation of extracellular DNA and enzymes such as elastase and myeloperoxidase causing persistent CF lung injury (6). Neutrophilic infiltration of the airways occurs secondary to the presence of cytokines and chemokines. Notably, early signs of lung inflammation can precede colonization and infection suggesting that inflammation caused by mucus plugging of airways is the primary event in CF lung disease.

T lymphocytes with appropriate antigenic specificity to luminal pathogens are present in high numbers in the CF airway submucosa, alveolar septa, and draining lymph nodes, but there is a striking paucity of T lymphocytes in the lumen itself (7, 8). Neutrophils and T lymphocytes are compartmentalized in CF airways, with neutrophils accumulating in the lumen, whereas T cells stay in the submucosa and lymph nodes and are excluded from the lumen. Enhanced granule release from the neutrophils, the concomitant release of arginase and pro-apoptotic factors in the CF airway fluid are probably responsible for this exclusion/early blockade of T cell expansion in the CF airway lumen (9). This dysregulation of the T cell response may play an important role in the virus-related exacerbation of bacterial infections in the CF airways (6).

The T-cell receptor (TCR) on the surface of T lymphocytes recognizes fragments of antigen as peptides bound to major histocompatibility complex (MHC) molecules (10). T lymphocytes mature from thymus cells called thymocytes (11, 12). Each thymocyte has a distinct T-cell receptor. Thymocytes will die by apoptosis if their T-cell receptor reacts too weakly or not at all with a peptide/MHC ligand of the epithelial cells in the cortex or too strongly to self antigens in the medulla of the thymus (11–15). Less than one percent of thymocytes survive the positive and negative selection in the thymus and emigrate into the periphery as naïve T cells where during an infection the high-affinity recognition of an antigen on a MHC molecule will initiate an immune response and the clonal T cell proliferation.

TCRs are heterodimers, composed for most T cells of an α - and a β -chain with each T cell clone expressing a unique

combination. Peptide specificity of $\alpha\beta$ T cells is primarily determined by the amino acid sequence encoded in the third complementarity-determining region (CDR3) loops of the α - and β -chain variable domains. In the β -chain locus the CDR3 regions are formed by somatic recombination of paralogous variable (V), diversity (D) and joining (J) gene segments, with additional removal and random addition of nucleotides at the V-J, V-D and D-J junctions thus permitting an enormous combinatorial diversity in receptor composition (16). We were curious to know how the inherited CF condition affects the TCR profiles that are selected from a diverse, naive repertoire.

Considering the magnitude of theoretically possible diversity of 10^{18} TCR α - β pairs (17), the sharing of TCR sequences among individuals should be very rare. The frequency of α - β pairs has indeed been shown to be a stochastic product of individual germline encoded V and J gene segment usage (18). However, a subset of so-called “public” TCR sequences is frequently shared between individuals (19, 20). Studies on the TCR repertoire in monozygotic twins moreover identified a genetic influence on sharing of identical TCR sequences (18, 21–24) occurring at a normalized rate of 10^{-7} (21).

These studies on healthy twin pairs suggest that genetic determinants impose constraints on the human TCR repertoire. Cystic fibrosis is the most common monogenic disorder in Caucasian populations that predisposes to chronic life-limiting infection. This inherited disease offers the opportunity to examine the constraints on the TCR repertoire in a compromised host susceptible to infection who does not harbor an inborn error in the innate or adaptive immune system. Thanks to the support of the European Cystic Fibrosis Society we could recruit and examine monozygotic twins with cystic fibrosis who are homozygous for the most common disease-causing *CFTR* mutation p.Phe508del (25).

Here we report on the sequence diversity of the antigen-binding *TCR β* CDR3 region in peripheral blood of 16 twin pairs, the data of which have already been communicated in another context for three pairs (26). Monozygotic twins have identical genes encoding the MHC I alleles and the TCR recombination machinery. Hence, based on data published for healthy twins (18, 21, 23, 24) we expected a more similar distribution of V and J segments in a CF twin pair than between unrelated CF individuals. This assumption turned out to be true for the majority of twin pairs but a minority showed a more individual expansion of segments in their TCR repertoire.

MATERIALS AND METHODS

European Cystic Fibrosis Twin and Sibling Study Cohort (25)

530 CF patient pairs were enrolled from 158 CF clinics in 14 European countries who provided information about date of birth, gender, zygosity status of twins, and the most recent height, weight and spirometry (FEV1, forced expiratory volume in 1 s). The anthropometric parameter wfh%, percentage weight for height predicted, was calculated from the comparison of the measured height and weight to the predicted weight of a

healthy individual of the same height, age and gender extracted from the centile distribution of the Zürich study on growth and development (27). To correct for the CF-specific decline of FEV1%pred with age, CF specific disease centiles for FEV1, called FEV1%pred, were determined by mapping the patient's FEV1%pred referring to a non-CF population (28) onto the age- and gender-dependent centile distributions of FEV1%pred based on lung function data of 25,667 European CF patients compiled in the ERCF Annual Report 1998 (29).

The complete cohort of 1,060 patients was sorted by rank numbers x_i for wfh% and y_i for the FEV1 disease centile FEV1%pred, whereby the rank 1 indicated the most severely affected state.

By giving the rank numbers for anthropometry and lung function the same weight, the CF disease severity of an individual i in the cohort was defined by

$$(x_i^2 + y_i^2)^{1/2}$$

and the intrapair discordance Δ between the two individuals i and j of a twin and sibpair by

$$\Delta = [(x_i^2 - x_j^2) + (y_i^2 - y_j^2)]^{1/2}.$$

Rank numbers for Δ were assigned to the complete twin and sibpair cohort, whereby the rank 1 indicated the most concordant pair.

The reported zygosity status of 41 twin pairs was tested by multiplex PCR of nine short tandem repeat loci (D3S1358, D5S818, D7S820, D8S1179, D13S317, D18S51, D21S11, FGA, and vWA) (30).

Monozygotic Twins

Sixteen monozygotic twin pairs were recruited for the investigation of the TCR β -chain diversity of whom DNA or blood for DNA extraction, clinical data and in-house verified information about *CFTR* genotype and zygosity status were available. Fourteen exocrine pancreatic insufficient pairs of non-consanguineous marriage are homozygous for the most common CF causing mutation p.Phe508del (31). As non-congruent outliers of *CFTR* genotype two further pairs were included, i.e., (a) an exocrine pancreatic sufficient twin pair who is compound heterozygous for p.Phe508del and the class V splice mutation c.3717+12191 C-T which confers some residual wild type *CFTR* activity (32) and (b) an exocrine pancreatic insufficient twin pair of first cousin parents carrying the ultra-rare class I splice mutation c.3469-2 A-G on both CF alleles. First cousin progeny has been reported to carry homozygous segments in 11% of their genome with an average segment length of 26 cM (33) (Table 1).

Determination of TCR β CDR3 Sequence Diversity

DNA was extracted from peripheral blood cells with phenol/chloroform (34) and stored in TE buffer at 4°C. **Supplementary Table 1** lists the amount and concentration of DNA in the stock solutions. High-throughput multiplex amplicon sequencing of the CDR3 variable region of the β -chain of T cell receptors, defined according to IMGT (35)

TABLE 1 | CF twin pair cohort.

Twin pair	Gender	<i>CFTR</i> genotype	Age at assessment [yrs]
53	female	p.Phe508del/p.Phe508del	5
97	female	p.Phe508del/p.Phe508del	9
116	male	p.Phe508del/p.Phe508del	19.8
118	male	c.3469-2 A-G/c.3469-2 A-G	14.9
156	female	p.Phe508del/c.3717+12191 C-T	36.4
180	male	p.Phe508del/p.Phe508del	9.2
237	female	p.Phe508del/p.Phe508del	8.3
238	female	p.Phe508del/p.Phe508del	8.2
244	male	p.Phe508del/p.Phe508del	17.4
291	female	p.Phe508del/p.Phe508del	30.2
292	female	p.Phe508del/p.Phe508del	12.7
294	female	p.Phe508del/p.Phe508del	8.8
309	female	p.Phe508del/p.Phe508del	2
401	male	p.Phe508del/p.Phe508del	9.5
403	male	p.Phe508del/p.Phe508del	6.1
512	female	p.Phe508del/p.Phe508del	5.2

was performed with the immunoSEQ immune profiling system according to the manufacturer's instructions (Adaptive Biotechnologies, Seattle, WA) (36, 37). Barcoded sequencing libraries were prepared from three μ g of fragmented DNA by bridge amplification with 15 cycles of PCR. Sequencing of the immune repertoire was performed on an Illumina NextSeq system (Illumina) followed by the data evaluation on the immunoSEQ Analyzer platform (<https://clients.adaptivebiotech.com>). The amplicon sequences of the TCR β CDR region obtained for the CF twins were searched for shared sequences of other projects deposited in the immuneACCESS database as of August 6, 2020 and in the VDJdb database as of December 18, 2020 (38).

Diversity of the TCR β sequences was estimated by the measure of Simpson clonality, i.e., the square root of Simpson's index of in-frame productive rearrangements. The generation probability of TCR β sequences was computed with the OLGA algorithm (39). To compare the clonotype V and J gene usage distribution in pairs of individuals, we applied Zvyagin's and coworkers' approach to calculate the Jensen-Shannon distance (21). The Jensen-Shannon divergence (40–42) based on Jensen's inequality and Shannon entropy computes the similarity between two probability distributions, in this case the equally-weighted segment usage in two individuals.

The Jensen-Shannon distance is defined as the Jensen-Shannon divergence divided by the mean entropy of two distributions (21).

[Background (42): $M_+^1(A)$ is a set of probability distributions where A is a set provided with some σ -algebra.

Jensen-Shannon divergence JSD: $M_+^1(A) \times M_+^1(A) \rightarrow [0, \infty]$ is a symmetrized and smoothed version of the Kullback-Leibler divergence $D(P || Q)$. It is defined as

$JSD(P \parallel Q) = \frac{1}{2} D(P \parallel M) + \frac{1}{2} D(Q \parallel M)$ with $M = \frac{1}{2}(P + Q)$

For more information please consult the Python package for discrete information theory at https://dit.readthedocs.io/en/latest/measures/divergences/jensen_shannon_divergence.html.

RESULTS

CF Twin Clinical Characteristics

Ten female and six male monozygotic CF twin pairs aged 2.0–36.4 years (median 9.1 years) were selected for the study (Table 1). The group had on average worse anthropometry and lung function than the dizygous CF sibpairs who had been recruited during the same period from the CF centers. Figure 1 compares the disease severity of study participants within the whole twin and sibpair cohort as indicated by their individual rank for two parameters sensitive for course and prognosis, i.e., weight-for-height and disease centile FEV1 (25). The monozygotic twins shared a more similar nutritional and pulmonary status than the CF sibpairs as shown by lower rank number differences in nutritional and pulmonary status (Figure 1). The lower intra-pair difference of the monozygotic twins should reflect their genome, gender and age identity and lifelong sharing of domicile, health care system and local CF care.

Motivation to Analyze the TCR β CDR3 Region in CF Monozygotic Twins

Each twin pair had been exposed since conception to closely related living conditions including the exposure to antigens. Monozygotic twins share the same MHC I allele profile (43) and the genes coding for the TCR recombination machinery. In analogy to sequence data of healthy monozygotic twins' TCR repertoires (18, 21), we thus expected the sharing of TCR clonotypes and a higher similarity of V and J segment usage in CF

monozygotic twins than seen among unrelated CF individuals. By sequencing the variable region of the TCR β chain with the standardized immunoSEQ protocol developed by Robins et al. (36), we were curious to know whether we could identify a CF-associated signature in the TCR repertoire that is distinct from the sequences of healthy controls (37) stored in the publicly available immuneACCESS database.

TCR β -chain diversity was assessed in peripheral blood of the 16 CF twin pairs. Table 2 summarizes the features of the 32 CF twins' CDR3 sequence repertoire described in the forthcoming chapters.

Frequency Distribution of Clonotypes in the Individual Twins' TCR β CDR3 Repertoire

Simpson clonality as a measure of the evenness of the distribution of clonotypes varied between 0.0049 and 0.0490 within the range observed for healthy individuals (37) (Table 2). The median value of 0.0095 is lower than that of 0.021 reported for healthy adults consistent with the fact that all but two of our pairs were children or adolescents at the day of sampling and that clonality increases with age reflecting the episodes of clonal expansion due to antigen exposure occurring over time (Figure 2). The low clonality values indicate that the population of TCR β clonotypes is dominated by the naïve T-cell repertoire. Thus CF twins were in general harboring an almost evenly distributed TCR β repertoire.

Five CF patients, two pairs (180, 294) and a singleton (244B), however, showed a higher clonality than expected for their age (Table 2). The singleton belonged to the twin pair with the worst lung function of the cohort and the two pairs represent the twins with the highest intra-pair discordance in lung function, i.e., one patient suffered from an already highly advanced stage of CF pulmonary disease while the other twin's pulmonary function was in the typical CF range of this age. In other words, of the five individuals with the highest Simpson clonality, three belonged to the quartet with the highest annual rate of

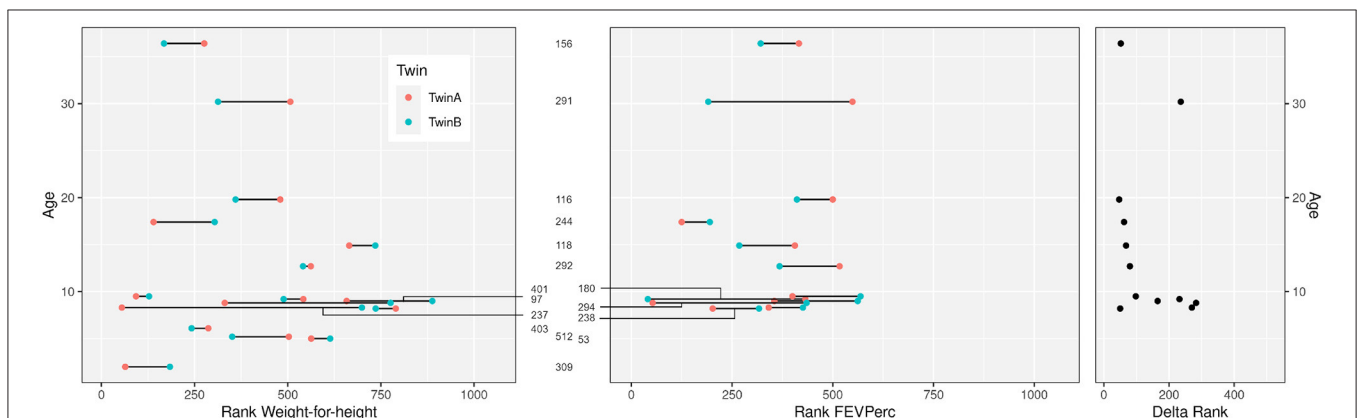


FIGURE 1 | Clinical data of the 16 CF twin pairs discerned by their sample number ID in the European Cystic Fibrosis Twin and Sibling Study (see Table 1). The figure shows the age of the 32 monozygotic twins and their ranks within the 1,060 participants of the European Cystic Fibrosis Twin and Sibling Study for the clinical parameters “Weight-for-height” (left) and “FEV1” (middle) at the day of blood sampling. A low rank number indicates a severe outcome. Patients had to be older than 6 years to perform a robust FEV1 measurement. The delta rank (right) indicates the intrainpair discordance of the twin pairs. A low delta rank indicates a comparable disease manifestation of the twins. In contrast, large delta values indicate divergent clinical phenotypes.

TABLE 2 | Features of the CDR3 amino acid sequences of the beta chain T cell receptors (TCR β) in monozygotic p.Phe508del homozygous CF twin pairs.

Sample name	Total templates (Sum) ^a	Total productive templates (Sum) ^b	Productive fraction ^c	Rearrangements ^d	Productive rearrangements ^e	Shared NT clonotypes (no./%) ^f	Shared AA clonotypes (no./%) ^g	Productive clonality ^h	Max productive frequency (%) ⁱ
53_A	74,359	60,394	0.8122	58,295	47,444	90/0.16	791/1.67	0.0056	0.13
53_B	54,633	43,850	0.8026	40,931	32,998	90/0.22	791/2.39	0.0086	0.43
97_A	55,365	46,496	0.8398	48,126	40,752	100/0.21	614/1.52	0.0082	0.29
97_B	34,980	29,393	0.8403	30,528	25,923	100/0.33	614/2.4	0.0115	0.85
116_A	22,749	17,803	0.7826	19,238	15,115	13/0.07	126/0.83	0.0102	0.45
116_B	16,668	13,310	0.7985	16,472	13,186	13/0.08	126/0.95	0.0089	0.14
118_A ^l	27,188	22,096	0.8127	22,821	18,477	61/0.27	217/1.17	0.0102	0.35
118_B ^l	24,862	20,075	0.8075	22,178	17,871	61/0.28	217/1.21	0.0086	0.22
156_A ^k	13,389	10,931	0.8164	11,430	9,379	1/<0.01	76/0.83	0.016	0.99
156_B ^k	21,483	17,831	0.83	16,154	13,211	1/<0.01	76/0.58	0.0206	1.14
180_A	9,519	7,653	0.804	8,057	6,399	8/0.09	58/0.91	0.0435	3.03
180_B	19,253	15,353	0.7974	15,764	12,581	8/0.05	58/0.46	0.0339	2.22
237_A	15,161	12,500	0.8245	14,950	12,332	34/0.23	155/1.25	0.0091	0.06
237_B	22,341	18,568	0.8311	21,835	18,162	34/0.16	155/0.85	0.0077	0.12
238_A	52,383	44,480	0.8491	44,272	37,627	36/0.08	684/1.82	0.0068	0.20
238_B	51,082	43,265	0.847	49,265	41,763	36/0.07	684/1.64	0.005	0.05
244_A	17,564	14,737	0.839	16,622	13,977	0	6/0.04	0.0095	0.23
244_B	1,736	1,446	0.8329	1,589	1,321	0	6/0.46	0.0341	1.94
291_A	33,749	26,884	0.7966	30,442	24,532	31/0.10	457/1.86	0.0119	0.57
291_B	43,361	34,600	0.798	35,092	28,112	31/0.09	457/1.63	0.0097	0.38
292_A	83,484	66,033	0.791	59,097	46,808	27/0.05	614/1.3	0.0075	0.48
292_B	38,969	30,983	0.7951	37,940	30,147	27/0.07	614/2.02	0.0059	0.05
294_A	28,112	22,486	0.7999	21,865	17,398	35/0.16	362/2.07	0.049	3.58
294_B	47,793	37,288	0.7802	42,678	33,621	35/0.08	362/1.07	0.0221	2.00
309_A	77,263	60,094	0.7778	62,948	49,303	42/0.07	819/1.66	0.0101	0.68
309_B	50,585	39,846	0.7877	45,482	35,787	42/0.09	819/2.28	0.0061	0.14
401_A	41,606	34,505	0.8293	35,877	30,220	11/0.03	460/1.52	0.0077	0.27
401_B	42,413	35,237	0.8308	37,819	31,687	11/0.03	460/1.45	0.0089	0.48
403_A	12,679	10,791	0.8511	11,916	10,167	44/0.37	336/3.36	0.0104	0.19
403_B	59,694	51,267	0.8588	54,272	46,735	44/0.08	336/0.73	0.0049	0.07
512_A	46,689	37,641	0.8062	42,584	34,320	4/0.01	342/0.99	0.0058	0.06
512_B	30,713	24,890	0.8104	29,149	23,641	4/0.01	342/1.44	0.0074	0.23

^aNumber of raw sequencing reads.^bNumber of in-frame coding reads.^cRatio of in-frame coding reads to raw sequencing reads.^dNumber of the sequences with rearranged V β , D β , and J β gene segments.^eNumber of functional in-frame rearrangements.^fThe numbers refer to within pair shared clonotypes with identical nucleotide sequence.^gThe numbers refer to within pair shared clonotypes with identical amino acid sequence.^hSquare root of Simpson's index of in-frame productive rearrangements.ⁱThe maximal frequency of a specific productive rearrangement among all productive rearrangements within a sample calculated as the templates for a specific rearrangement divided by the sum of productive templates for a sample.^lFirst cousin offspring homozygous for c.3469-2 A-G.^kCompound heterozygous for p.Phe508del/c.3717+12191 C-T.

pulmonary exacerbations and the other two individuals were the other monozygotic twin. These five patients harbored the within sample most abundant clonotypes that made up 2–4% of their individual TCR β repertoire (**Table 3**). Of these eight sequences, two sequences were not present in the immuneACCESS database,

three sequences were detected at a frequency of <1% and another three sequences were common clones identified in 7–11% of samples deposited in the database (**Table 3**). In summary, the individually most abundant clonotypes in our data set represent an ensemble of private, rare and common public clones.

Intrapair Sharing of Clonotypes at the Nucleotide and Amino Sequence Levels

A twin shared on the average about 30 clonotypes of identical nucleotide sequence with the twin sibling (inner quartiles: 11–42; range 0–100) (Table 2). Intra-pair sharing occurred for in total 0.11% of all CDR3 sequences. For comparison, we determined the sharing of clonotypes within CF twin pairs arbitrary pairs of unrelated CF twins generated from the same data-sets (Figure 3A; Supplementary Table 2). The inter- vs. intrapair comparison revealed that the number of shared nucleotide sequences was significantly higher within monozygotic CF twin

pairs ($P = 0.000026$, Mann-Whitney rank test). The almost identical genomic blueprint and the exposure to the same environment since conception had led to a small repertoire of shared TCR β clonotypes with identical nucleotide sequence. Within our small cohort we did not identify any association between the number of shared CDR3 sequences and age, concordance and severity of CF disease.

Next, we investigated the sharing of TCR β clonotypes at the amino acid sequence level in our CF twin data set. The CF twins did not share CDR3 peptides more often than arbitrary pairs assembled from the 32 CF twins ($P = 0.766$) (Figure 3B; Supplementary Table 2). The numbers of shared amino acid sequences increased with cloneset size consistent with previous reports on healthy monozygotic twins (21).

Simpson clonality is a statistic measure of how much of the repertoire is made up of expanded clones (see above). The intra-pair variation of Simpson clonality of the CF twins was low (median 0.0019, range 0.0012–0.0055). We observed a non-significant trend that twin pairs show a lower clonality difference than unrelated patient samples (Figure 4) suggesting that genetic constraints and the shared environmental exposure may favor a higher similarity of the TCR β repertoire.

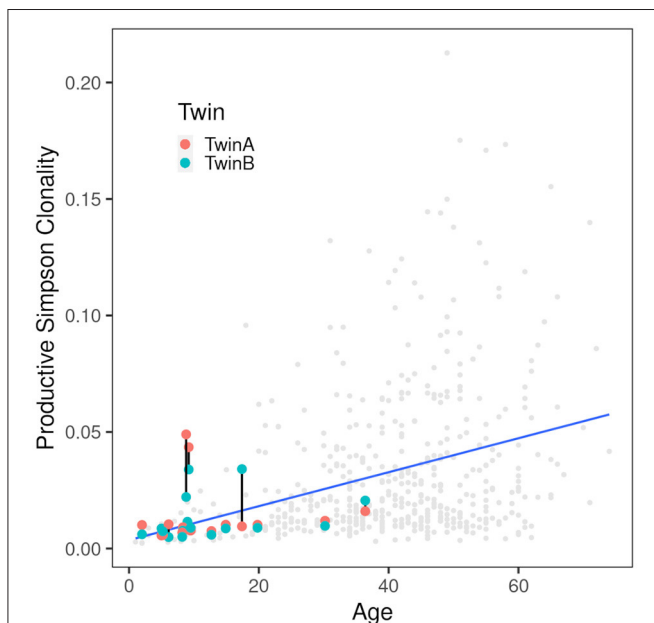


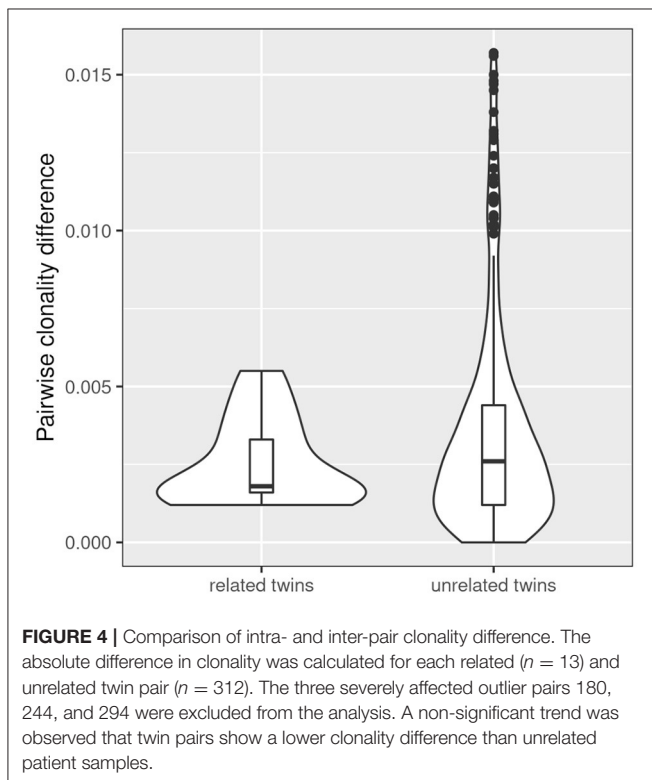
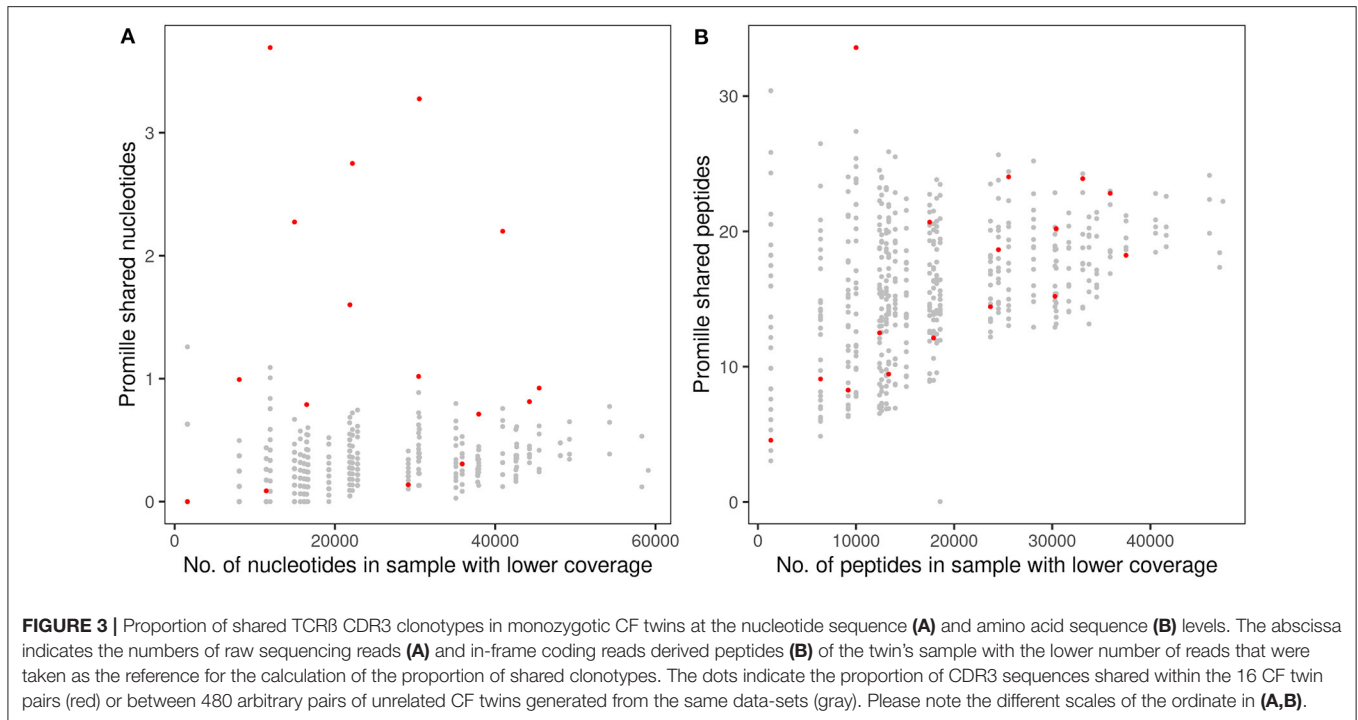
FIGURE 2 | Correlation of productive Simpson clonality with age. Productive clonality is plotted for each twin. Twin pairs are connected with a black line. The gray dots indicate the trend of the controls. As expected, the clonality increases with age. Of the twins only the severely affected patients are above the trend line indicating that the CF samples are comparable to those from healthy controls.

Intrapair Sharing of V and J Gene Segment Usage in CF Twins

Monozygotic twins share the TCR recombination machinery. Hence, based on data published on healthy twins (18, 21), we expected a more similar distribution of V and J gene segments within a CF twin pair than between unrelated CF individuals. The Jensen-Shannon divergence divided by mean entropy (21, 42) was calculated as an equally weighted measure of the similarity of the TCR β populations of V and J gene usage for in-frame TCR β clonotypes of all possible pairs of individuals (Figure 5). The shortest Jensen-Shannon distance between the populations of V and J gene segments was mutually to the other CF twin individual for 13 of 16 twin pairs in case of the V usage and for 8 of the 16 twin pairs in case of the J usage. This data verify the observation in healthy monozygotic twins (21) that genetic constraints are involved in the selection of V and J gene segments.

TABLE 3 | The most frequent non-shared TCR β clonotypes in the CF twin samples.

Sample	CDR3 amino acid sequence	Count	Frequency in twin sample (%)	Presence in immuneACCESS database	
				No of samples	Frequency (%)
294_A	CASSLPREGNEQYF	804	3.58	4	0.04
294_A	CASSPGDTIYF	443	1.97	662	7.04
180_A	CASSYHRSTAEAFF	232	3.03	0	0
180_A	CASSFAGALQETQYF	150	1.96	63	0.67
180_B	CASSLWGWEQYV	341	2.22	0	0
180_B	CASSLRGGSYEQYV	324	2.11	55	0.59
294_B	CASSLSRGEAFF	745	2	718	7.64
244_B	CASRQGSQPQHF	28	1.94	1051	11.18



In contrast to the observation in the majority of twin pairs, individual signatures of V and J usage were also noted. In case of the five pairs 180, 244, 291, 292, and 294 the J sequence

population was most similar from one twin to the twin sibling, but not vice versa, and both twins of the pairs 156, 403, and 512 shared a more similar distribution of J sequences to unrelated CF individuals than to the other twin. Similarly, the twin pair 180 and the twins 244A and 294B had a more similar V usage with unrelated CF patients than with their twin sibling. These three twin pairs also demonstrated the most uneven distribution of clonotypes and the poorest age-corrected lung function (see **Figure 1**).

Figure 6 visualizes the individual V segment usage in the 180 twin pair who for his age was most affected by severe pulmonary disease in our cohort. The box plots depict the ratio of copy numbers of V segments of twin 180A or twin 180B to those of all other members of the CF twin cohort. Of the 63 V segment classes, the usage was similar within a factor of ± 2 for about 80% of classes among all study participants. However, four V gene segments were more than 2-fold more frequently detected in the TCR β chain repertoire of both twins than in the unrelated individuals. Moreover twin 180A harbored nine segments and twin 180B carried five segments two- to more than hundred fold more frequently in his TCR β repertoire than all other study participants including his twin brother. In summary, twins 180A and 180B exhibited a more individual and more skewed usage of V gene segments than all other examined monozygotic CF twins.

Sharing of Clonotypes at the Amino Acid Sequence Level in the CF Twin Cohort

39,812 clonotypes were detected twice or more in our data-set. **Figure 7** depicts the number of shared clonotypes in the CF twin samples. The frequency of detection of the presence of a sequence in the sample set shows a concave deviation from a logarithmic

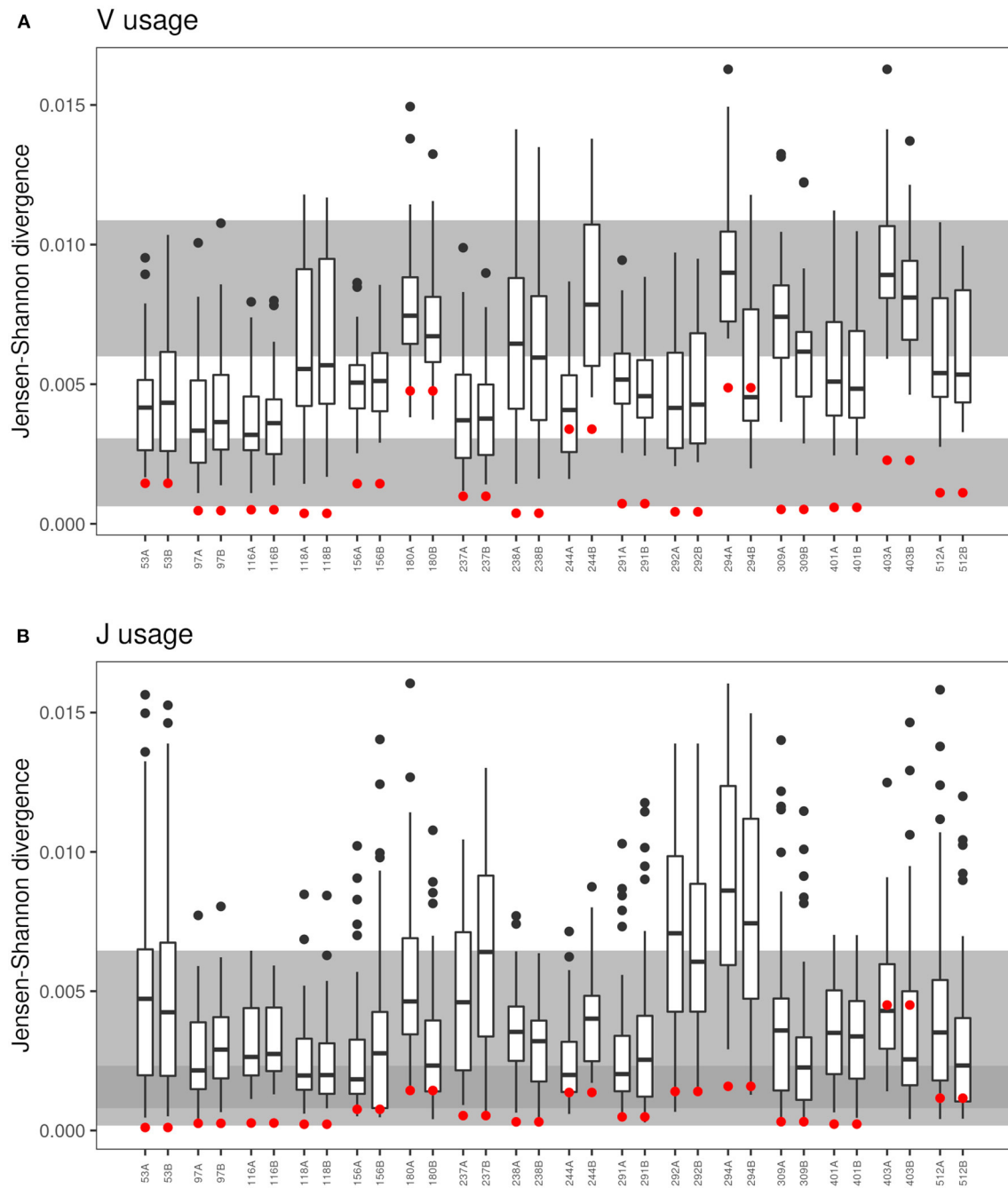


FIGURE 5 | Comparison of VJ-usage in related and unrelated twin pairs. The boxplot presentation shows the Jensen-Shannon divergence (40–42) for the comparison of each sample to the twin sibling and to the 30 unrelated twins for the **(A)** V-segment and **(B)** J-segment usage of the CDR3. Red dots represent the Jensen-Shannon divergence to the twin sibling. The gray background areas indicate the reference values for healthy unrelated twins (upper gray panel) and related healthy twins (lower gray panel) (21). In case of the J-segment, the reference panels overlap indicated by the dark gray area.

decline consistent with our knowledge that the generation probability of human CDR3 amino acid sequences varies over 20 orders of magnitude (39) and that the most common sequences are present in the TCR β repertoire of virtually any individual.

By utilizing the OLGA algorithm recently published by Sethna and colleagues (39), we calculated the probability of generating

CDR3 amino acid sequences for various subsets of shared clonotypes. Of the CDR3 amino acid sequences exclusively identified in both siblings of one twin pair, the median probability of generating one of these clonotypes was computed to be 3.1×10^{-8} [inner quartiles 8.4×10^{-9} - 7.9×10^{-8} ; range 2.5×10^{-21} - 2.6×10^{-6}]. A 2-fold higher median generation probability of

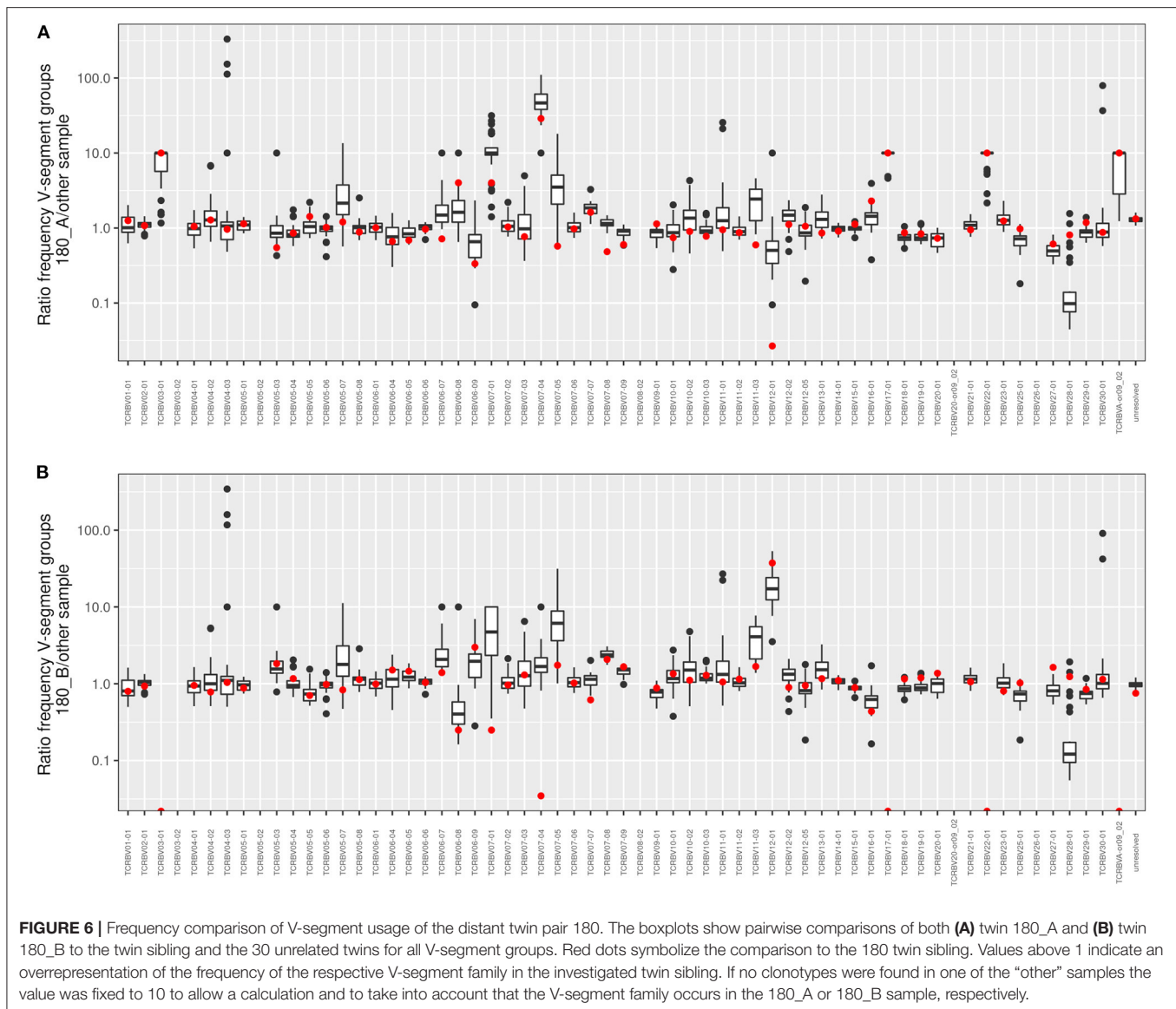


FIGURE 6 | Frequency comparison of V-segment usage of the distant twin pair 180. The boxplots show pairwise comparisons of both **(A)** twin 180_A and **(B)** twin 180_B to the twin sibling and the 30 unrelated twins for all V-segment groups. Red dots symbolize the comparison to the 180 twin sibling. Values above 1 indicate an overrepresentation of the frequency of the respective V-segment family in the investigated twin sibling. If no clonotypes were found in one of the “other” samples the value was fixed to 10 to allow a calculation and to take into account that the V-segment family occurs in the 180_A or 180_B sample, respectively.

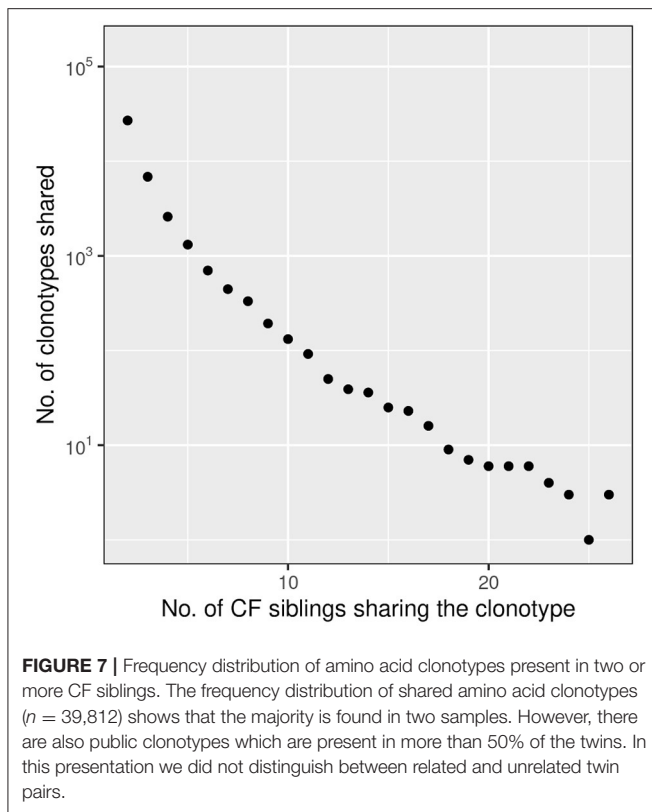
6.7×10^{-8} [2.3×10^{-8} - 1.6×10^{-7} ; 0.81×10^{-6}] was calculated for clonotypes that were only detected in unrelated patients and a 10-fold higher median probability of 2.8×10^{-7} [1.2×10^{-7} - 6.0×10^{-7} ; 3.1×10^{-15} - 6.8×10^{-6}] for the sequences identified in one to 12 twin pairs and singletons of the remaining pairs. Since generation probabilities larger than 10^{-10} (44)– 10^{-8} (45) have been estimated to be present in any individual we can safely assume that at least 75% of the detected shared clonotypes in our CF twin cohort are public clones commonly present in the human TCR repertoire.

Abundance of CF Twin Shared Amino Acid Sequence Clonotypes in Health and Disease

We were curious to know whether and to which extent the shared CDR3 sequences had already been detected in healthy

individuals or patients with other diseases. The screening of the TCR β repertoire of 9,100 samples in the immuneACCESS database revealed that 39,739 clonotypes, i.e., 99.82% of total, had already been previously found in 1–5,896 samples (median: 587) (**Figure 8**).

Just 73 shared clonotypes had not been detected before in the other projects, 31 of which in a twin pair and 42 in unrelated CF patients. Their median generation probability was computed to be 7.8×10^{-11} (inner quartiles 1.2×10^{-11} - 2.9×10^{-10} ; range 2.5×10^{-21} - 2.8×10^{-8}). These values are substantially lower than calculated for the three sets of shared clonotypes described in the previous paragraph. Hence these shared CF twin clonotypes that yet have not been documented in the immuneAccess database are indeed not common in the human TCR β repertoire. A subset of these sequences may have emerged in the monozygotic CF twins during shared exposure to opportunistic pathogens as a consequence of their inherited



susceptibility to infection. Consistent with this interpretation none of these 73 clonotypes was listed in the VDJdb database (38) that contains clonotypes with known specificities. Conversely, 47 of the 84 clonotypes that were identified in more than half of the CF twin samples could be mapped to the epitopes listed for *Mycobacterium tuberculosis* (one entry) and one or more viruses (67 entries) in the VDJdb database (38). Forty clonotypes were assigned to CMV epitopes and 11 clonotypes to EBV epitopes, eight of which were mapped to both CMV and EBV epitopes. These 84 most abundant clonotypes in the CF twin cohort have a median generation probability of 1.4×10^{-6} [8.3×10^{-7} - 2.1×10^{-6} ; 2.7×10^{-7} - 5.6×10^{-6}] demonstrating that these CDR3 sequences present public clones of the non-CF community.

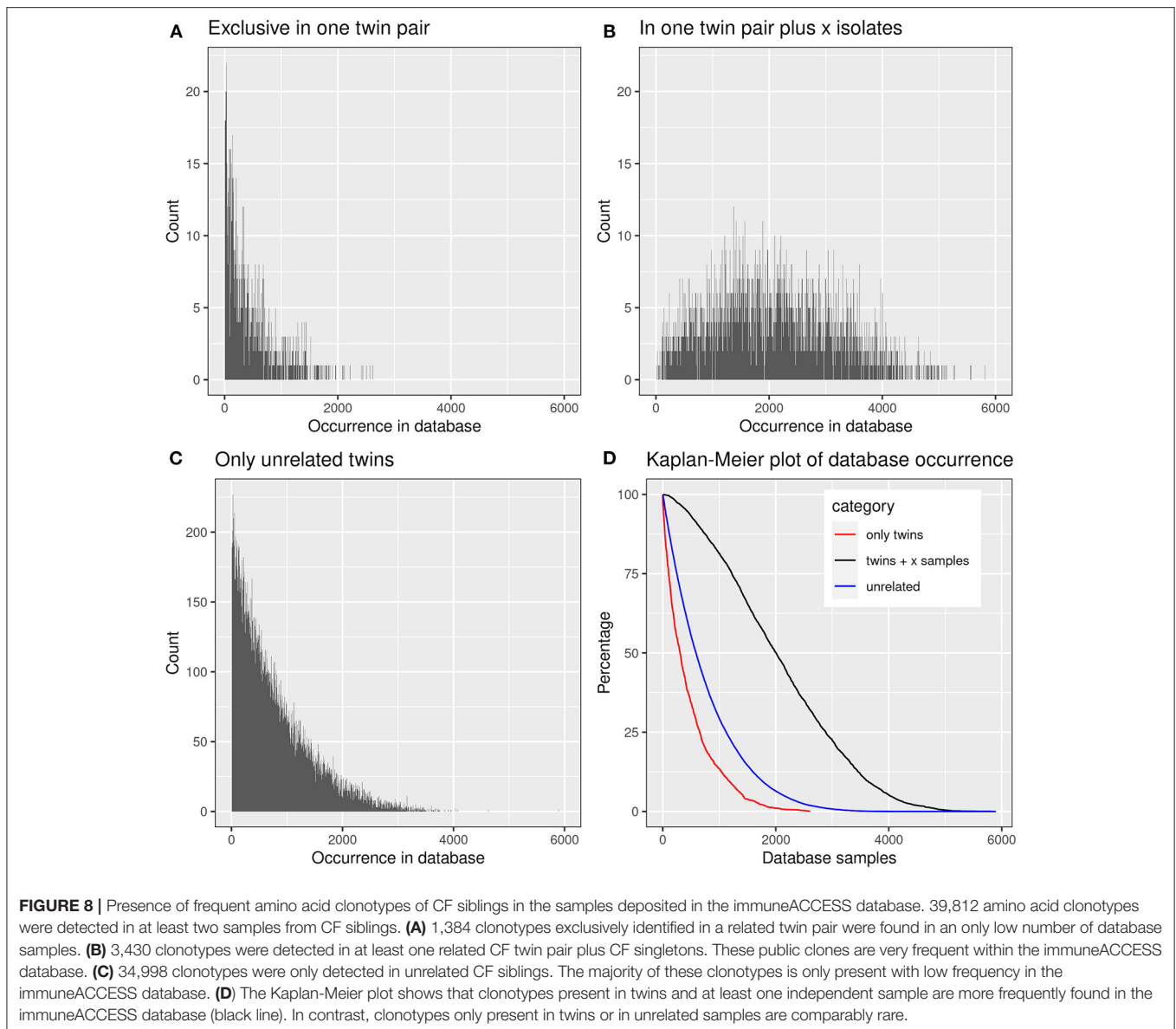
The three data subsets of shared clonotypes of our CF cohort (twin pair only, unrelated patients only, twin pairs and unrelated patients) were analyzed for their frequency distribution among the non-CF samples of the immuneACCESS database (Figure 8, Table 4). Interestingly, the empirical frequency distributions of the three subsets in the database matched with their distribution of TCR β generation probabilities predicted by the OLGA algorithm. If a clonotype had only been identified in the two sibs of a twin pair, it generally was infrequently present in the non-CF samples. 85% of the clonotypes were detected at a frequency of <10% in the immuneACCESS database. Similarly, clonotypes solely identified among unrelated CF patients were mainly detected at low frequency in the non-CF samples (Table 4). However, they were more common than the twin pair sequences indicated by a smaller portion of 68% of

sequences that were identified in <10% of non-CF samples. The frequency distribution of clonotypes from both these CF data sets was exponentially declining in the non-CF sample collection (Figures 8A,C). Conversely, if a clonotype had been identified in the two siblings of a twin pair and in addition in unrelated CF patients, the frequency among non-CF samples showed a skewed Gaussian distribution (Figure 8B). This sub-set of CF clonotypes mainly consists of widely distributed public clones (Table 4), i.e., a majority of 83% of the CF clonotypes were present in more than 10% of the non-CF samples. Since these frequent CDR3 sequences are shared by both monozygotic twin pairs and unrelated individuals, we hypothesized that they should represent a pool of sequences that are non-randomly selected by both global and individual genetic and environmental factors. The OLGA algorithm predicted a high median generation probability of 2.8×10^{-7} for this subset of CF clonotypes consistent with the empirical finding that the most abundant TCR β CDR3 sequences arise from the higher generation probability compared to the repertoire median (46). These conserved public sequences are central within TCR sequence-similarity networks (47, 48).

We checked the distribution of the “public” CF clonotypes (Figure 8B) in subsets of the immuneACCESS database, i.e., samples from healthy controls, patients with infectious disease or cancer. As seen in Figure 9, the largest abundance of these “public” CF clonotypes were observed among healthy non-CF people, less represented in individuals with infectious disease and uncommon in patients with cancer. Hence we would like to conclude that the clonotypes that are shared within twin pairs and between unrelated CF sibs represent public clones that are common in healthy humans. We speculate that in humans – at least in those of European descent – a subportion of the naïve T cell receptor repertoire is overrepresented that is not necessarily associated with disease as it had been deduced from murine animal models (49, 50).

DISCUSSION

This study examined the number of unique TCR β sequences in peripheral blood specimens collected from monozygotic p.Phe508del homozygous CF twins. In contrast to expectation we identified only a few yet unknown CDR3 sequences at amino acid sequence level. More than 99% of sequences had already been observed in healthy people or other diseases. The recurrent airway infections which had affected our twin cohort since early childhood had not introduced any disease-specific bias in the TCR β repertoire. We predominantly examined CF children and adolescents who were carrying individual CDR3 sequences at low copy numbers. Individual clonal expansions were primarily seen in the clinically most severely affected patients suggesting that clonality may be higher in CF adults with more advanced lung disease. Alternatively, the compartmentalization in CF airways, with neutrophils accumulating in the lumen, whereas T cells stay in the submucosa and lymph nodes and are excluded from the lumen (9), could prevent a more active role of T cells against the luminal pathogens and the neutrophil-driven chronic inflammation.



To our knowledge TCR β clonotypes were investigated in the yet largest cohort of monozygotic twins in humans with a monogenic life-limiting disease. Previous studies in healthy adults (18, 21, 51) have shown a genetic influence on the usage of V and to a lesser non-significant extent on the usage of J gene segments. In contrast we observed a similar genetic influence on the usage of both V and J segments in our monozygotic CF twin cohort which just may be the consequence of the fact that the previous analyses examined only three (21), five (51) or six pairs (18) whereas this study is based on 16 twin pairs. Since no genetic impact of gene segment pairs for both variable and joining regions beyond that for gene segment usage has been identified in TCR α - β pairs (18), we would like to conclude in accordance with the literature that recombination accounts for the genetic influence of V and J gene segment usage.

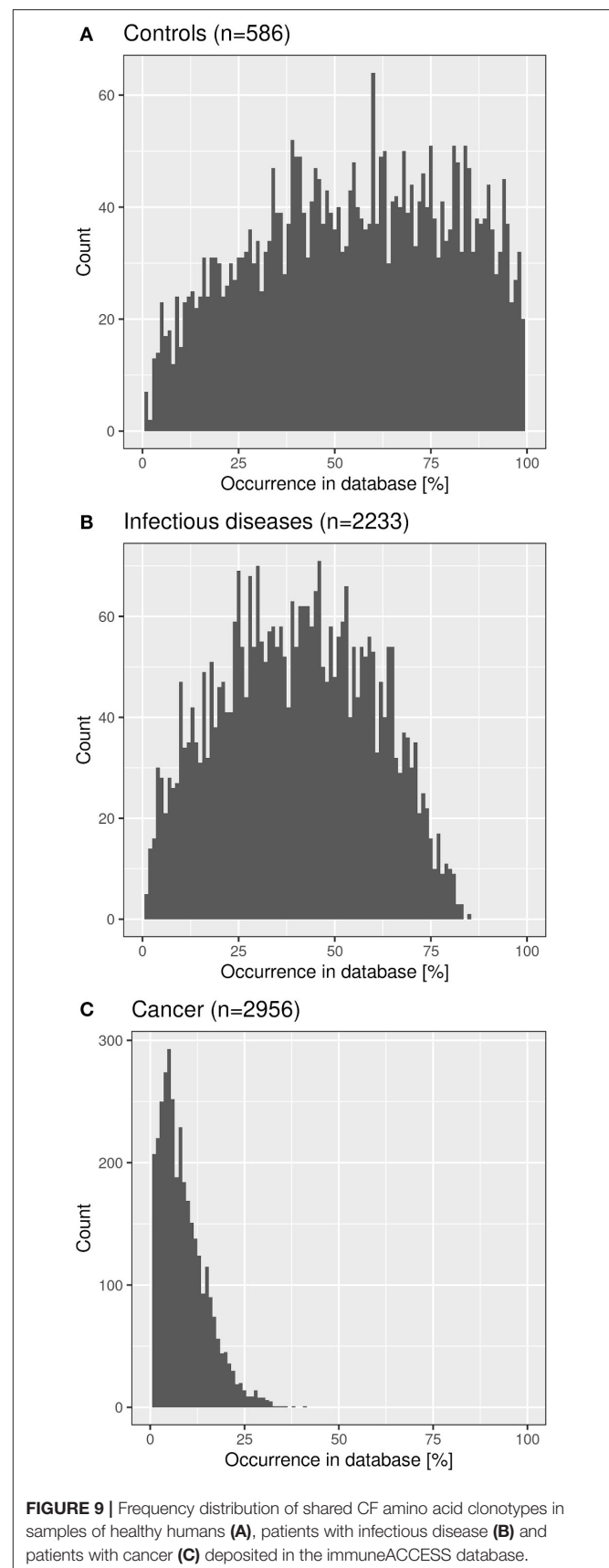
This genetic bias showed up in the Jensen-Shannon distance of V and J segment usage that was typically shortest to the twin individual (**Figure 7**). Exceptions were the clinically most discordant and/or most severely affected twin pairs where clonal expansion probably caused by recurrent pulmonary infections overshadowed the impact of the identical genomic blueprint. Recently the T-cell receptor repertoire has been reported for 28 monozygotic twin pairs of whom one or both twins were affected by the polyfactorial inflammatory bowel disease (IBD) (52). The Jensen-Shannon divergence of the V gene usage of TCR β was substantially higher in the IBD than in the CF twin cohort which we would not ascribe to the diverse disease etiologies but rather to the age difference that the IBD patients were mid-age adults and all but two CF twin pairs were children and adolescents.

TABLE 4 | Presence of the TCR β clonotypes found in two or more CF samples in the 9,100 non-CF samples of the immuneACCESS database.

Sharing of clonotypes in the CF twin data set		Frequency (%) in non-CF samples median (inner quartiles; range)
	Total no. of clonotypes	
Exclusive in one twin pair	1,384	3.2 (1.1–7.0; 0–27.8)
One twin pair + unrelated patients	3,430	21.3 (13.0–29.4; 0.05–61.9)
Only unrelated patients	34,998	6.3 (2.6–11.9; 0–62.7)

Intrapair sharing of clonotypes was significantly more frequent among monozygotic CF twins than among pairs of unrelated CF patients. Complete nucleotide sequence identity was observed in about 0.11% of CDR3 sequences. Deep sequencing of the TCR repertoire in cord blood and healthy twins and unrelated individuals of different age (23) have provided convincing evidence that a large portion of these shared clones are derived from the same progenitor T cells generated during fetal development. However, besides the persistence of fetal clones further mechanisms should exist that generated an increased repertoire of shared clonotypes among the monozygotic CF twin pairs.

More than 99.8% of the clonotypes of the CF twin cohort had already been detected at amino acid sequence level in a cohort of about 10,000 non-CF individuals whose TCR β repertoire had been investigated by matching protocols of CDR3 amplicon sequencing. The subset of shared clonotypes in the CF cohort showed divergent frequency distributions in the non-CF cohort. Clonotypes were typically re-identified in only few non-CF individuals if the sequence was shared by A) either the sibs of a twin pair or B) two or more unrelated CF patients (cf. **Figure 8**). The sharing of clones in these two scenarios can be ascribed to either identical genetic makeup including MHC I alleles and shared environmental exposure in case of the monozygotic CF twins (A) or to positive selection of a functional and self-tolerant T cell repertoire driven by recombination machinery, structural and functional constraints (15) in case of unrelated CF patients (B). In both scenarios (A) or (B) the frequency distribution of clonotypes in the non-CF cohort showed an exponential decline indicating that the probability is comparatively low that the same clonotype will re-emerge in an unrelated individual. In contrast, if both scenarios (A) and (B) applied, the respective clonotypes were widely distributed in the non-CF cohort (cf. **Figure 8B**). These CDR3 sequences are public clones in the narrow sense which survive thymic selection in healthy humans with high probability (**Figure 9**). The co-occurrence in monozygotic twins and unrelated individuals suggests that these clonotypes are positively selected by both genetic and functional factors. Public clones have been frequently detected in mouse models of autoimmune and infectious diseases, but the immuneACCESS database tells us that these clones are characteristic for healthy humans.

**FIGURE 9** | Frequency distribution of shared CF amino acid clonotypes in samples of healthy humans (A), patients with infectious disease (B) and patients with cancer (C) deposited in the immuneACCESS database.

In conclusion, this study uncovered in general a low Simpson clonality indicating that the population of TCR β clonotypes of the CF twins was dominated by the naïve T-cell repertoire. The genetic constraints of TCR repertoire formation showed up in highly similar distributions of the usage of J and V genes of in-frame clonotypes in a CF twin pair. Almost all CDR3 peptides shared by CF twins were public clones prevalent in health and disease, but the few shared CDR3 amino acid sequences that were found for the first time, probably represent CF – associated clonotypes.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are publicly available. This data can be found here: <https://clients.adaptivebiotech.com/pub/fischer-2021-fi>.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethics Committee of Hannover Medical School (no. 1300 and no. 3739, last update May 8th, 2014). Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

REFERENCES

- Cutting GR. Cystic fibrosis genetics: from molecular understanding to clinical application. *Nat Rev Genet.* (2015) 16:45–56. doi: 10.1038/nrg3849
- Elborn JS. Cystic fibrosis. *Lancet.* (2016) 388:2519–31. doi: 10.1016/S0140-6736(16)00576-6
- Stoltz DA, Meyerholz DK, Welsh MJ. Origins of cystic fibrosis lung disease. *N Engl J Med.* (2015) 372:351–62. doi: 10.1056/NEJMra1300109
- Blanchard AC, Waters VJ. Microbiology of cystic fibrosis airway disease. *Semin Respir Crit Care Med.* (2019) 40:727–36. doi: 10.1055/s-0039-1698464
- Ratjen F, Bell SC, Rowe SM, Goss CH, Quittner AL, Bush A. Cystic fibrosis. *Nat Rev Dis Primers.* (2015) 1:15010. doi: 10.1038/nrdp.2015.10
- Khan MA, Ali ZS, Swezey N, Grasemann H, Palaniyar N. Progression of cystic fibrosis lung disease from childhood to adulthood: neutrophils, neutrophil extracellular trap (NET) formation, and NET degradation. *Genes.* (2019) 10:183. doi: 10.3390/genes10030183
- Hubeau C, Lorenzato M, Couetil JP, Hubert D, Dusser D, Puchelle E, et al. Quantitative analysis of inflammatory cells infiltrating the cystic fibrosis airway mucosa. *Clin Exp Immunol.* (2001) 124:69–76. doi: 10.1046/j.1365-2249.2001.01456.x
- Regamey N, Tsartsali L, Hilliard TN, Fuchs O, Tan HL, Zhu J, et al. Distinct patterns of inflammation in the airway lumen and bronchial mucosa of children with cystic fibrosis. *Thorax.* (2012) 67:164–70. doi: 10.1136/thoraxjnl-2011-200585
- Ingersoll SA, Laval J, Forrest OA, Preininger M, Brown MR, Arafat D, et al. Mature cystic fibrosis airway neutrophils suppress T cell function: evidence for a role of arginase 1 but not programmed death-ligand 1. *J Immunol.* (2015) 194: 5520–8. doi: 10.4049/jimmunol.1500312
- Rudolph MG, Stanfield RL, Wilson IA. How TCRs bind MHCs, peptides, and coreceptors. *Annu Rev Immunol.* (2006) 24:419–66. doi: 10.1146/annurev.immunol.23.021704.115658

AUTHOR CONTRIBUTIONS

FS and BT planned the study. SF and BT analyzed the data. SF and BT wrote the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This research was funded by the Deutsche Forschungsgemeinschaft (DFG) (Grant Number: SFB900/3 - 158989968 - Z1) and the German Center for Lung Research at BREATH (Grant number: 82DZL002A1).

ACKNOWLEDGMENTS

We thank the patients, their parental guides, and our colleagues at the CF centers in Aberdeen, Amsterdam, Birmingham, Chemnitz, Dublin, Frankfurt, Genoa, Hannover, Innsbruck, Marseille, Münster, Paris, Rabka, Rotterdam, and Utrecht for their participation in the CF twin study. DNA extraction and sample processing by Silke Hedtfeld, Ahmed Farag, and Lutz Wiehlmann is gratefully acknowledged.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fimmu.2021.599133/full#supplementary-material>

- Boehm T. Thymus development and function. *Curr Opin Immunol.* (2008) 20:178–84. doi: 10.1016/j.coi.2008.03.001
- Park JE, Botting RA, Domínguez Conde C, Popescu DM, Lavaert M, Kunz DJ, et al. A cell atlas of human thymic development defines T cell repertoire formation. *Science.* (2020) 367:eaay3224. doi: 10.1126/science.aay3224
- Alawam AS, Anderson G, Lucas B. Generation and regeneration of thymic epithelial cells. *Front Immunol.* (2020) 11:858. doi: 10.3389/fimmu.2020.00858
- Fu G, Rybakina V, Brzostek J, Paster W, Acuto O, Gascoigne NR. Fine-tuning T cell receptor signaling to control T cell development. *Trends Immunol.* (2014) 35:311–8. doi: 10.1016/j.it.2014.05.003
- Klein L, Kyewski B, Allen PM, Hogquist KA. Positive and negative selection of the T cell repertoire: what thymocytes see (and don't see). *Nat Rev Immunol.* (2014) 14:377–91. doi: 10.1038/nri3667
- Turner SJ, Doherty PC, McCluskey J, Rossjohn J. Structural determinants of T-cell receptor bias in immunity. *Nat Rev Immunol.* (2006) 6:883–94. doi: 10.1038/nri1977
- Davis MM, Bjorkman PJ. T-cell antigen receptor genes and T-cell recognition. *Nature.* (1988) 334:395–402. doi: 10.1038/334395a0
- Tanno H, Gould TM, McDaniel JR, Cao W, Tanno Y, Durrett RE, et al. Determinants governing T cell receptor α/β -chain pairing in repertoire formation of identical twins. *Proc Natl Acad Sci USA.* (2020) 117:532–40. doi: 10.1073/pnas.1915008117
- Robins HS, Srivastava SK, Campregher PV, Turtle CJ, Andriesen J, Riddell SR, et al. Overlap and effective size of the human CD8+ T cell receptor repertoire. *Sci Transl Med.* (2010) 2:47ra64. doi: 10.1126/scitranslmed.3001442
- Venturi V, Quigley ME, Greenaway HY, Ng PC, Ende ZS, McIntosh T, et al. A mechanism for TCR sharing between T cell subsets and individuals revealed by pyrosequencing. *J Immunol.* (2011) 186:4285–94. doi: 10.4049/jimmunol.1003898
- Zvyagin IV, Pogorelyy MV, Ivanova ME, Komech EA, Shugay M, Bolotin DA, et al. Distinctive properties of identical twins' TCR repertoires revealed

- by high-throughput sequencing. *Proc Natl Acad Sci USA*. (2014) 111:5980–5. doi: 10.1073/pnas.1319389111
22. Qi Q, Cavanagh MM, Le Saux S, NamKoong H, Kim C, Turgano E, et al. Diversification of the antigen-specific T cell receptor repertoire after varicella zoster vaccination. *Sci Transl Med*. (2016) 8:332ra46. doi: 10.1126/scitranslmed.aaf1725
 23. Pogorelyy MV, Elhanati Y, Marcou Q, Sycheva AL, Komech EA, Nazarov VI, et al. Persisting fetal clonotypes influence the structure and overlap of adult human T cell receptor repertoires. *PLoS Comput Biol*. (2017) 13:e1005572. doi: 10.1371/journal.pcbi.1005572
 24. Pogorelyy MV, Minervina AA, Touzel MP, Sycheva AL, Komech EA, Kovalenko EI, et al. Precise tracking of vaccine-responding T cell clones reveals convergent and personalized response in identical twins. *Proc Natl Acad Sci USA*. (2018) 115:12704–9. doi: 10.1073/pnas.1809642115
 25. Mekus F, Ballmann M, Bronsveld I, Bijman J, Veeze H, Tümmler B. Categories of deltaF508 homozygous cystic fibrosis twin and sibling pairs with distinct phenotypic characteristics. *Twin Res*. (2000) 3:277–93. doi: 10.1375/13690520032056256
 26. Schamschula E, Hagmann W, Assenov Y, Hedtfeld S, Farag AK, Roesner LM, et al. Immunotyping of clinically divergent p.Phe508del homozygous monozygotic cystic fibrosis twins. *J Cyst Fibros*. (2021) 20:149–53. doi: 10.1016/j.jcf.2020.06.009
 27. Prader A, Largo RH, Molinari L, Issler C. Physical growth of Swiss children from birth to 20 years of age. First Zurich longitudinal study of growth and development. *Helv Paediatr Acta Suppl*. (1989) 52:1–125.
 28. Knudson RJ, Lebowitz MD, Holberg CJ, Burrows B. Changes in the normal maximal expiratory flow-volume curve with growth and aging. *Am Rev Respir Dis*. (1983) 127:725–34.
 29. McKenzie SG, Chowdhury S, Strandvik B, Hodson ME. Investigators of the Epidemiologic Registry of Cystic Fibrosis. Dornase alfa is well tolerated: data from the epidemiologic registry of cystic fibrosis. *Pediatr Pulmonol*. (2007) 42:928–37. doi: 10.1002/ppul.20685
 30. Sacchetti L, Calcagno G, Coto I, Tinto N, Vuttariello E, Salvatore F. Efficiency of two different nine-loci short tandem repeat systems for DNA typing purposes. *Clin Chem*. (1999) 45:178–83. doi: 10.1093/clinchem/45.2.178
 31. Bobadilla JL, Macek M Jr, Fine JP, Farrell PM. Cystic fibrosis: a worldwide analysis of CFTR mutations—correlation with incidence data and application to screening. *Hum Mutat*. (2002) 19:575–606. doi: 10.1002/humu.10041
 32. Chiba-Falek O, Kerem E, Shoshani T, Aviram M, Augarten A, Bentur L, et al. The molecular basis of disease variability among cystic fibrosis patients carrying the 3849+10 kb C→T mutation. *Genomics*. (1998) 53:276–83. doi: 10.1006/geno.1998.5517
 33. Woods CG, Cox J, Springell K, Hampshire DJ, Mohamed MD, McKibbin M, et al. Quantification of homozygosity in consanguineous individuals with autosomal recessive disease. *Am J Hum Genet*. (2006) 78:889–96. doi: 10.1086/503875
 34. Gross-Bellard M, Dudet P, Chambon P. Isolation of high-molecular-weight DNA from mammalian cells. *Eur J Biochem*. (1973) 36:32–8. doi: 10.1111/j.1432-1033.1973.tb02881.x
 35. Yousfi Monod M, Giudicelli V, Chaume D, Lefranc MP. IMGT/JunctionAnalysis: the first tool for the analysis of the immunoglobulin and T cell receptor complex V-J and V-D-J JUNCTIONS. *Bioinformatics*. (2004) 20:i379–85. doi: 10.1093/bioinformatics/bth945
 36. Robins HS, Campregher PV, Srivastava SK, Wachter A, Turtle CJ, Khsai O, et al. Comprehensive assessment of T-cell receptor beta-chain diversity in alphabeta T cells. *Blood*. (2009) 114:4099–107. doi: 10.1182/blood-2009-04-217604
 37. Emerson RO, DeWitt WS, Vignali M, Gravley J, Hu JK, Osborne EJ, et al. Immunosequencing identifies signatures of cytomegalovirus exposure history and HLA-mediated effects on the T cell repertoire. *Nat Genet*. (2017) 49:659–65. doi: 10.1038/ng.3822
 38. Bagaev DV, Vroomans RMA, Samir J, Stervbo U, Rius C, Dolton G, et al. VDJDdb in 2019: database extension, new analysis infrastructure and a T-cell receptor motif compendium. *Nucleic Acids Res*. (2020) 48:D1057–62. doi: 10.1093/nar/gkz874
 39. Sethna Z, Elhanati Y, Callan CG, Walczak AM, Mora T. OLGA: fast computation of generation probabilities of B- and T-cell receptor amino acid sequences and motifs. *Bioinformatics*. (2019) 35:2974–81. doi: 10.1093/bioinformatics/btz035
 40. Lin J, Wong SKM. A new directed divergence measure and its characterization. *Int J General Systems*. (1990) 17:73–81. doi: 10.1080/03081079008935097
 41. Lin J. Divergence measures based on the Shannon entropy. *IEEE Trans Inform Theor*. (1991) 37:145–51. doi: 10.1109/18.61115
 42. Fuglede B, Topsøe F. Jensen-Shannon divergence and Hilbert space embedding. In: *Proceedings of the International Symposium on Information Theory*. Chicago, IL: IEEE (2004). p. 31.
 43. Goldrath AW, Bevan MJ. Selecting and maintaining a diverse T-cell repertoire. *Nature*. (1999) 402:255–62. doi: 10.1038/46218
 44. Lythe G, Callard RE, Hoare RL, Molina-Paris C. How many TCR clonotypes does a body maintain? *J Theor Biol*. (2016) 389:214–24. doi: 10.1016/j.jtbi.2015.10.016
 45. Qi Q, Liu Y, Cheng Y, Glanville J, Zhang D, Lee JY, et al. Diversity and clonal selection in the human T-cell repertoire. *Proc Natl Acad Sci USA*. (2014) 111:13139–44. doi: 10.1073/pnas.1409155111
 46. Murugan A, Mora T, Walczak AM, Callan CG Jr. Statistical inference of the generation probability of T-cell receptors from sequence repertoires. *Proc Natl Acad Sci USA*. (2012) 109:16161–6. doi: 10.1073/pnas.1212755109
 47. Madi A, Shifrut E, Reich-Zeliger S, Gal H, Best K, Ndifon W, et al. T-cell receptor repertoires share a restricted set of public and abundant CDR3 sequences that are associated with self-related immunity. *Genome Res*. (2014) 24:1603–12. doi: 10.1101/gr.170753.113
 48. Madi A, Poran A, Shifrut E, Reich-Zeliger S, Greenstein E, Zaretsky I, et al. T cell receptor repertoires of mice and humans are clustered in similarity networks around conserved public CDR3 sequences. *Elife*. (2017) 6:e22057. doi: 10.7554/eLife.22057
 49. Menezes JS, van den Elzen P, Thornes J, Huffman D, Droin NM, Maverakis E, et al. A public T cell clonotype within a heterogeneous autoreactive repertoire is dominant in driving EAE. *J Clin Invest*. (2007) 117:2176–85. doi: 10.1172/JCI28277
 50. Zhao Y, Nguyen P, Ma J, Wu T, Jones LL, Pei D, et al. Preferential use of public TCR during autoimmune encephalomyelitis. *J Immunol*. (2016) 196:4905–14. doi: 10.1093/jimmunol.1501029
 51. Rubelt F, Bolen CR, McGuire HM, Vander Heiden JA, Gadala-Maria D, Levin M, et al. Individual heritable differences result in unique cell lymphocyte receptor repertoires of naïve and antigen-experienced cells. *Nat Commun*. (2016) 7:11112. doi: 10.1038/ncomms11112
 52. Rosati E, Pogorelyy MV, Dowds CM, Moller FT, Sorensen SB, Lebedev YB, et al. Identification of disease-associated traits and clonotypes in the T cell receptor repertoire of monozygotic twins affected by inflammatory bowel diseases. *J Crohns Colitis*. (2020) 14:778–90. doi: 10.1093/ecco-jcc/jjz179

Conflict of Interest: The 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.

Copyright © 2021 Fischer, Stanke and Tümmler. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.