EBioMedicine 68 (2021) 103429

Contents lists available at ScienceDirect

EBioMedicine

journal homepage: www.elsevier.com/locate/ebiom

Research paper

TCR repertoire diversity in Multiple Sclerosis: High-dimensional bioinformatics analysis of sequences from brain, cerebrospinal fluid and peripheral blood

Roberta Amoriello[a,1], Maria Chernigovskaya[b,1], Victor Greiff[b], Alberto Carnasciali[a], Luca Massacesi[c], Alessandro Barilaro[c], Anna M. Repice[c], Tiziana Biagioli[d], Alessandra Aldinucci[d], Paolo A. Muraro[e], David A. Laplaud[f], Andreas Lossius[g,*], Clara Ballerini[a,*]

[a] Dipartimento di Medicina Sperimentale e Clinica (DMSC), Laboratory of Neuroimmunology, University of Florence, Viale Pieraccini 6, 50139 Florence, Italy
[b] Department of Immunology, University of Oslo, Oslo, Norway
[c] Dipartimento di Neuroscienze, Psicologia, Area del Farmaco e Salute del Bambino (NEUROFARBA), University of Florence, Italy
[d] Laboratorio Generale, Careggi University Hospital, Florence, Italy
[e] Wolfson Neuroscience Laboratory, Department of Brain Sciences, Imperial College London, London, United Kingdom
[f] CRTI-Inserm U1064, CIC0004 and Service de Neurologie, CHU de Nantes, Hôpital Nord Laënnec, Nantes, France
[g] Institute of Clinical Medicine, University of Oslo, Postboks 1105, Blindern 0317 Oslo, Norway

ARTICLE INFO

Article History:
Received 25 November 2020
Revised 12 May 2021
Accepted 19 May 2021
Available online 11 June 2021

Keywords:
Multiple Sclerosis
High-throughput sequencing
System immunology
T-cell repertoire diversity
Cerebrospinal fluid
Brain

ABSTRACT

Background: T cells play a key role in the pathogenesis of multiple sclerosis (MS), a chronic, inflammatory, demyelinating disease of the central nervous system (CNS). Although several studies recently investigated the T-cell receptor (TCR) repertoire in cerebrospinal fluid (CSF) of MS patients by high-throughput sequencing (HTS), a deep analysis on repertoire similarities and differences among compartments is still missing.
Methods: We performed comprehensive bioinformatics on high-dimensional TCR Vβ sequencing data from published and unpublished MS and healthy donors (HD) studies. We evaluated repertoire polarization, clone distribution, shared CDR3 amino acid sequences (CDR3s-a.a.) across repertoires, clone overlap with public databases, and TCR similarity architecture.
Findings: CSF repertoires showed a significantly higher public clones percentage and sequence similarity compared to peripheral blood (PB). On the other hand, we failed to reject the null hypothesis that the repertoire polarization is the same between CSF and PB. One Primary-Progressive MS (PPMS) CSF repertoire differed from the others in terms of TCR similarity architecture. Cluster analysis splits MS from HD.
Interpretation: In MS patients, the presence of a physiological barrier, the blood-brain barrier, does not impact clone prevalence and distribution, but impacts public clones, indicating CSF as a more private site. We reported a high Vβ sequence similarity in the CSF-TCR architecture in one PPMS. If confirmed it may be an interesting insight into MS progressive inflammatory mechanisms. The clustering of MS repertoires from HD suggests that disease shapes the TCR Vβ clonal profile.
Funding: This study was partly financially supported by the Italian Multiple Sclerosis Foundation (FISM), that contributed to Ballerini-DB data collection (grant #2015 R02).

1. Introduction

Multiple Sclerosis (MS), the most widespread neurological disease among young adults [1], is an inflammatory and autoimmune disorder characterized by the disruption of myelin in the central nervous system (CNS). Despite the establishment of several risk factors including genetics, environment and infections, MS etiology remains unknown [2]. MS can manifest mainly in two forms: the most common (85% of all patients) is Relapsing-Remitting MS (RRMS), which progresses with relapses interspersed with periods of partial or complete recovery. The other forms are progressive, including Primary- (PPMS) and Secondary-Progressive MS (SPMS) [3].

* Corresponding authors.
E-mail addresses: andreas.lossius@medisin.uio.no (A. Lossius), clara.ballerini@unifi.it (C. Ballerini).
[1] Both the authors contributed equally to this work.

https://doi.org/10.1016/j.ebiom.2021.103429

Research in Context

Evidence before this study

The cerebrospinal fluid in Multiple Sclerosis, an immune-mediated demyelinating, inflammatory, chronic disease of the central nervous system, has been deeply investigated as a precious source of information about brain tissue and disease inflammatory status. T lymphocytes are usually present in Multiple Sclerosis cerebrospinal fluid, and a still unanswered question is whether the central nervous system T-cell receptor repertoire quantitatively and qualitatively differs from the periphery and whether these differences correlate with the pathogenic mechanism. The study of the T-cell receptor repertoire has incredibly improved after the advent of modern techniques of sequencing, and T-cell receptor diversity has been addressed with numerous bioinformatics approaches.

Added value of this study

Concerning T-cell receptor repertoire investigation, the use of different bioinformatics approaches and the small sample size of studies performed so far in Multiple Sclerosis cerebrospinal fluid, have hampered the identification of relevant information that may bridge molecular data and disease pathological mechanisms. We overcome these limits by performing a statistical analysis on pooled sequences carefully checked for quality. We found that clone distribution and prevalence were similar between Multiple Sclerosis cerebrospinal fluid and peripheral blood, but distinguished the T-cell receptor repertoire of Multiple Sclerosis patients and healthy donors. We found that cerebrospinal fluid shows more public clones across repertoires, whereas peripheral blood repertoires show higher overlap with public T-cell receptor databases (McPAS-TCR and VDJdb). Through network analysis to evaluate T-cell receptor similarity architecture, we detected a different repertoire structure in one active progressive Multiple Sclerosis patient, characterized by a high sequence similarity: in the future, it would be worthwhile to deepen this finding in progressive Multiple Sclerosis patients.

Implications of all the available evidence

Our analysis allowed us to detect, with an acceptable approximation, repertoire similarities and differences characterizing Multiple Sclerosis and Multiple Sclerosis compartments (cerebrospinal fluid, and peripheral blood). We believe that our approach will support advances in the future analysis of the T-cell receptor repertoire in Multiple Sclerosis.

MS pathogenesis is complex and heterogeneous and it is in part mediated by autoreactive T lymphocytes that migrate from the periphery to the CNS across the blood-brain barrier (BBB) and attach myelin. In their path to the CNS, T cells recirculate through the cerebrospinal fluid (CSF), which is a compartment of choice for MS diagnosis (e.g., detection of oligoclonal bands) and biomarkers investigation [4,5]. The abnormal presence of T cell repertoire within the CSF of MS patients was first reported in studies mainly performed by spectratyping technology and/or flow cytometry [6,7], and in recent times thanks to the advent of high-throughput sequencing (HTS), that revolutionized the study of the TCR [8] The latest investigations confirmed the presence of a compartmentalized enrichment of T cells in the CSF of MS patients [9], and showed that some T-cell clones seem to be selectively shared only among MS patients when compared to patients affected with other neurological disorders [10,11]. The comparison of *in vitro* expanded CSF-derived T-cell clones with other compartments such as peripheral blood and brain lesions of MS patients showed that some of these clones are shared with memory CD4+ *T* cells in the peripheral blood and in brain lesions [12].

Despite these interesting results, the investigation of the TCR repertoire by HTS in heterogeneous diseases such as MS remains a young field and the scientific community still lacks common criteria for the bioinformatics analysis of results that may underlie differences and similarities between compartments in MS patients. Furthermore, HTS analysis provides an enormous amount of TCR sequencing data that may be challenging to understand and to be linked with the clinical and biological aspects of the disease [13]

For these reasons, we decided to investigate the MS TCR repertoire, performing a statistical analysis on HTS-TCR sequencing data previously obtained by other groups [9–11], together with a newly collected database (Ballerini database). TCR data were derived from CSF and peripheral blood (PB) of MS patients. Our aim was to pool different MS databases increasing the statistical power of the analysis, to identify common features and differences in the TCR repertoire of CSF and PB of MS patients.

2. Methods

2.1. Selection of T-cell receptor databases

In the present work, we selected the most recent TCR Vβ sequencing studies performed by HTS from MS patients starting from 2014, when the HTS technology was used for the first time to investigate CSF T cells. For finding studies, we adopted the following keywords: TCR repertoire, MS, HTS/NGS, CSF. The included researches are reported among references and encompass TCR Vβ sequencing data from PB, CSF and brain lesions [9–11]. All non-HTS studies were excluded from this analysis [6,14,15] We also excluded a recent study [12] that reports TCR Vβ sequencing data only from MS brain lesions.

We added to the analysis our unpublished TCR Vβ sequences, termed hereafter Ballerini database (Ballerini-DB), according to inclusion criteria.

2.2. Ballerini database: patient enrollment and ethics statement

CSF samples were obtained from incoming patients undergoing the diagnostic lumbar puncture and enrolled at Neurological Clinic (University of Florence, Italy). All patients signed informed consent and Local Ethical Committee authorized the study (#2014/003,601). Patients characteristics are summarized in Table 2. Final MS diagnosis was performed according to McDonald criteria.[4] Recruited patients had the following characteristics: age between 18 and 55 years; no immunomodulatory therapy since at least six months; no immunesuppressive therapy since at least 12 months; no corticosteroid administration in the previous month. Exclusion criteria: cognitive decline preventing informed consent release; concomitant diseases that may interact with disease course or with sample processing (e.g. neoplasms or infections as hepatitis or Human Immunodeficiency Virus [HIV]) or induce any risk for the patient; pregnancy or breastfeeding. Recruited patients were clinically assessed every six months. All patient material was encoded to avoid identification of the individual's name or identity during material processing.

2.3. Ballerini database: peripheral blood and cerebrospinal fluid collection

Whole PB was collected in heparin-containing tubes. Peripheral blood mononuclear cells (PBMCs) were collected by density gradient centrifugation using Pancoll (density: 1.077 g/mL; PAN-Biotech, Germany) at 1500 rpm, RT, for 30 min within six hours from blood collection. PBMCs were cryopreserved in 10% dimethyl sulfoxide (DMSO) in aliquots containing 20×10^6 cells and stored in a liquid nitrogen freezer until used.

We collected, as routine, 12 mL of CSF from each patient by lumbar puncture. The spinal tap was carried out using a Sprotte type 19 or 20 G conical elliptic shaped spinal atraumatic needle (Pajuk, Germany), inserted through a four cm long, cutting tip needle as intruder needle. CSF samples were analyzed immediately after lumbar puncture for cell counts, quantitative (IgM and IgG indices) and qualitative (oligoclonal bands) analysis of intrathecal Ig synthesis, and albumin quotient as an indicator of the BBB status using standard methods. We included CSF samples (12 mL each) with >0.8 cell/μl, in order to achieve a total of 10,000 cells and to avoid a sample bias that may have affected the TCR repertoire profile.[16] Cells were isolated by centrifuge at 800 rpm for 15 min, RT, and destined to RNA isolation, as described below.

2.4. Ballerini database: T-cell isolation and sorting, RNA extraction and TCR Vβ sequencing

CD3+ cells were isolated from PBMCs of patients. PBMCs were thawed in defrosting medium (10% Fetal Bovine Serum in Dulbecco's phosphate-buffered saline), centrifuged at 1300 rpm, for 10 min, RT, then counted and destined to CD3+ cells isolation by negative immunomagnetic depletion using the Pan T Cell Isolation Kit human (Miltenyi Biotec, Germany), following manufacturer's protocol. CD3+ cells fraction was eluted through the LS column placed on a suitable MACS separator (Miltenyi Biotec, Germany).

From CD3+ purified cells, we sorted EM and CM CD4 and CD8 T cell subpopulations. CD3+ cells were washed once in MACS buffer (PBS 1X with 0.5% bovine serum albumin and 2 mM ethylenediaminetetraacetic acid) at 1500 rpm, RT, for 5 min, and then resuspended in RPMI 1640 medium and counted. To identify EM and CM CD4 and CD8 T cells, CD3+ cells were divided in two tubes and labelled for 20 min, RT, in the dark, with the following human fluorescent antibodies: one tube with anti-CD4 APC (clone: OKT4), anti-CD45RA FITC (clone: HI100), anti-CCR7 PE (clone: 3D12); the other tube with anti-CD8 PE Cy7 (clone: RPA-T8), anti-CD45RA FITC (clone: HI100), anti-CCR7 PE (clone: 3D12). All the antibodies are from eBioscence, USA. Memory T-cell subpopulations were sorted by FACSAria II flow cytometer (BD Biosciences, USA).

From CSF cells and paired peripheral EM and CM CD4 and CD8 T cells, RNA was extracted immediately after sorting by the RNeasy Plus Micro Kit (Qiagen, Netherlands), according to the manufacturer's protocol. RNA purity was assessed by NanoDrop ND-1000 spectrophotometer (EuroClone, Italy) evaluating the 260/280 nm absorbance. RNA concentration (pg/μL), rRNA ratio [28S/18S], and RNA Integrity Number (RIN) were evaluated using the RNA 6000 Pico Kit (Agilent, USA) and the Agilent BioAnalyzer 2100, according to the manufacturer's protocols.

RNA was destined to TCR Vβ sequencing. TCR Vβ libraries generation and HTS were performed by iRepertoire Inc. (Huntsville, AL, USA). TCR Vβ libraries were prepared by the patented tem-PCR (for target enriched multiplex PCR technology; iRepertoire Inc.). After purification, amplified TCR Vβ chain libraries were sequenced by MiSeq Illumina platform, using 250 PER (paired-end read) primers, which cover framework 1 to C-region.

3. Data statement

The TCR Vβ sequencing data will be deposited upon publication and made available by the authors, without undue reservation, to any qualified researcher.

3.1. Data visualization

Statistical analysis of TCR Vβ sequencing data was performed using the programming environment R [17]. Graphics were generated using the R packages ggplot2 [18], ggpubr [19] igraph [20] and ComplexHeatmap [21]. Hierarchical clustering of evenness profiles was performed based on correlation-based distance and visualized by heatmaps using the NMF R package; [22] according to NMF package default, average linkage clustering was used. Repertoire architecture networks were visualized using Cytoscape [23].

3.2. Statistics

In order to compare two compartments (CSF and PB), we evaluated a set of statistics for each TCR Vβ repertoire — Shannon-Evenness (see below), public clones percentage, public databases overlap, connected clones percentage.

To adjust for database differences, we performed mean-centering normalization within each database, similar to Nygaard et al. [24]. Briefly, we subtracted the mean value of the statistics in one database from all statistics in that database. After the normalization, we compared the statistics between two compartments (CSF and PB) using a two-sample Mann–Whitney test.

For public databases overlap analysis, we performed one-tailed Fisher's exact test for each disease category (see Supplementary Fig. 3 for details). The *p*-values were adjusted for multiple testing using the Bonferroni correction.

Statistical significance was considered when adjusted *p*-value<0.05.

3.3. Shannon-Evenness and clonal expansion profiles

Shannon–Evenness (S-E) is defined as the quotient of the exponential of the Shannon entropy and species richness (SR: number of unique CDR3s in a given TCR database) [25,26].

Shannon-Entropy = $-\sum f_i \; log \; f_i$, where *fi* is the frequency of the *i*th clone in a given TCR database.

Shannon-Evenness = $exp(-\sum f_i \; log \; f_i)/SR$

Shannon–Evenness is one if all clones in a repertoire have the same frequency (an "even" repertoire), or it converges to zero if very few clones dominate in the repertoire (a "polarized" repertoire), that is, if very few clones have very high frequency and a lot of clones have a very low frequency.

Clonal expansion profiles (also called Evenness profiles) represent Hill diversity profiles scaled by the repertoire SR. Hill diversity profiles are defined as:

$$^{q}D = \left(\sum_{i=1}^{SR} f_i^q\right)^{\frac{1}{1-q}},$$

where *fi* is the frequency of the *i*th clone in a given TCR database. Clonal expansion profiles are then defined as ^{q}D/SR, where q ranges from zero to ten with a step size of 0.2. The parameter q determines the importance of high-frequency clones in the determination of the *q*-parameterized evenness; the higher *q*, the more high-frequency clones are weighted [26].

4. Definition of private and public Vβ clones

A TCR Vβ clone (V-J-CDRβ3 a.a.) was considered private when present in the repertoires of only one individual and public when shared across repertoires of at least two individuals or shared with T cell Vβ clones present in one of the public databases below [25].

5. Clone network analysis

Networks of TCR Vβ repertoires were constructed as follows: Briefly, for each repertoire, a network was drawn using the R package igraph [20] based on the top 10,000 (or less if fewer clones were found in a patient's repertoire) CDR3s-a.a. (top 10,000 represented for all repertoires with more than 90% of sequencing reads). Within

each network, each node is a CDR3-a.a., and links between nodes are drawn if the internode Levenshtein distance (LD) is maximally one (=one a.a. sequence change) [27–29]

6. Publicly available T-cell receptor Vβ databases used

Our TCR Vβ data were compared with other databases briefly described in the following. The McPAS-TCR database is manually curated and stores TCR Vβ sequences associated with various pathologies and antigens based on published literature [30]. The McPAS-TCR data here analyzed were downloaded on February 3, 2020.

VDJdb is a database that stores and aggregates the results of published T-cell specificity assays coupling antigen specificities with TCR Vβ sequences [31]. The VDJdb data here analyzed were downloaded on June 20, 2019.

Soto-DB is a recently published TCR Vβ database of PB repertoire from healthy donors (HD) [32]. The database is publicly available and was downloaded on September 1, 2020.

7. Role of funding source

This study was partly financially supported by the Italian Multiple Sclerosis Foundation (FISM), that contributed to Ballerini-DB data collection (grant #2015 R02).

8. Results

8.1. Database characterization

According to study selection criteria (see Methods), we collected TCR Vβ sequencing data from Lossius-DB [9], Laplaud-DB [10] and Muraro-DB, [11] upon authorization by the authors. The characteristics of selected studies are summarized in Table 1. Raw data were used for all the analyses and included sample ID, CDR3 amino acid Vβ sequences (CDR3s-a.a.) with paired reads number. In Table 2, we reported patient characteristics of the included databases.

To assess TCR Vβ sequencing data quality, we first investigated whether the number of reads (#Reads), which is the total number of CDR3 amino acid Vβ sequences (CDR3s-a.a.) in a TCR repertoire, adequately reflected the clonal distribution by calculating the Pearson correlation coefficient (r) between Shannon-Evenness (S-E; see Methods) and #Reads for repertoires within each database (Fig. 1a) in CSF and PB compartments (Fig. 1b).

We found the Pearson correlation coefficients to be negative in all databases ($r_{Ballerini\text{-}DB}=-0.27$; $r_{Lossius\text{-}DB}=-0.35$; $r_{Muraro\text{-}DB}=-0.21$; $r_{Laplaud\text{-}DB}=-0.61$) (Fig. 1a) and in both CSF and PB compartments ($r_{CSF}=0.27$; $r_{PB}=-0.11$) (Fig. 1b), thereby showing sufficient sequencing depth. [25] The number of reads for each analyzed database is reported in Supplementary Fig 1.

8.2. T-cell receptor repertoire diversity in cerebrospinal fluid and peripheral blood of Multiple Sclerosis patients

In order to describe the TCR repertoire diversity and clonality in terms of Vβ clone frequency and polarization, we compared normalized (see Methods) S–E values across repertoires of different compartments (Fig. 1c). Results suggest that CSF does not significantly differ from PB, in terms of repertoire polarization.

TCR Vβ clones are named “public” when shared in at least two repertoires (see Methods). In Fig. 1d we show that CSF-TCR repertoires have a higher normalized percentage of public clones compared to PB (p-value=0.002; two-sample Mann–Whitney test) (Fig. 1d).

8.3. Sequence overlap with publicly available T-cell receptor databases

To gain insight into the possible diversity of antigen-driven TCR repertoires among CSF and PB, we compared our repertoires for the overlap with two public TCR sequences databases: McPAS-TCR, that collects Vβ CDR3s-a.a. from literature by disease type [30], and VDJdb [31], that collects Vβ CDR3s-a.a. by antigen species (the two databases are characterized in Supplementary Fig. 2a and 2b).

McPAS-TCR classifies sequences by diseases, e.g. autoimmune diseases, and infections, e.g. cytomegalovirus (CMV) and Epstein-Barr virus (EBV). We first applied one-tailed Fisher's exact test (see Methods and Supplementary Fig. 3) to test for overrepresented disease-associated CDR3s-a.a. from McPAS-TCR database. We also calculated standardized Pearson residual (reported in Fig. 2a) for disease-associated CDR3s-a.a. detected in our MS data. The p-values were corrected for multiple testing using the Bonferroni correction. Five diseases showed significant overrepresentation in our MS data: Celiac disease (CD) (adjusted p-value=2.75^{-13}), Influenza (adjusted p-value=2.61^{-14}), *M. tuberculosis* (adjusted p-value=0.0004), and Yellow Fever virus (YFV) (adjusted p-value=7.86^{-11}). Then, for each significant disease, we compared CSF and PB compartments using two-sample Mann-Whitney test and mean-normalization (Fig. 2a). The p-values were corrected for multiple testing using the Bonferroni correction. Compartment comparison within these categories showed that the overlap is significantly higher in PB compared to CSF for CD (adjusted p-value=0.001) and YFV (adjusted p-value=0.004).

VDJdb classifies sequences by antigen species. Using the same method as above, we compared compartments CSF and PB in diseases with overrepresented disease-associated CDR3s-a.a. (Fig. 2b) — Dengue virus (DENV) type 1 (adjusted p-value=2.80^{-06}) and 3/4 (adjusted p-value=4.50^{-05}), and YFV (adjusted p-value=1.10^{-05}). In all these diseases, the sequence overlap is significantly higher in PB compared to CSF (adjusted $p\text{-value}_{DENV1}=0.027$; adjusted $p\text{-value}_{DENV3/4}=0.019$; adjusted $p\text{-value}_{YFV}=0.04$).

In order to focus on mutual interactions between CSF and brain, we narrowed the analysis to those CDR3s-a.a. Vβ sequences shared between MS-CSF TCR repertoires and brain TCR repertoires, available

Table 1
Examined TCR Vβ databases.

Database	Year	Disease	N. of patients	Source	Tot n. of CDR3 Vβ amino acid sequences	Average n. of CDR3 Vβ amino acid sequences per patient
Ballerini-DB[A]	–	MS	4	CSF, PB	4 395 408	219 770,4
Laplaud-DB [10]	2015	MS	3	brain lesions, CSF, PB	159 723 129	794 343,9
Lossius-DB [9]	2014	MS	10	CSF, PB	46 406 403	7 780 137,9
Muraro-DB [11]	2016	MS	5	CSF, PB	29 251 026	2 925 102,6

[A] unpublished data.

Table 2
MS patients characteristics.

Database	Patient	Sex	Diagnosis	Age (years)	Treatments	Disease duration (months)	OCB[A]	CSF cells count/μl	HLA-DR	BBB[B] damage	IgG index alteration
Ballerini-DB	CSF2	M	PPMS	51	none	60	positive	4	DRB1*03:01	no	yes
	CSF3	F	RRMS	45	none	60	positive	2	DRB1*13:01	no	yes
	CSF4	M	RRMS	25	none	72	positive	2	DRB1*15:01	no	yes
	CSF5	F	RRMS	37	none	48	negative	6	DRB1*15:01	no	yes
Database	Patient	Sex	Diagnosis	Age (years)	Treatments	Disease duration (months)	OCB[A]	CSF cells count/μl	HLA-DR		
Lossius-DB	MS-1	F	RRMS	31	none	2	positive	25	HLA-DRB1*07,08		
	MS-2	F	RRMS	38	none	1	positive	7	HLA-DRB1*15		
	MS-3	M	RRMS	39	none	8	positive	15	HLA-DRB1*13,15		
	MS-4	F	RRMS	20	none	11	positive	2	HLA-DRB1*04,13		
	MS-5	F	CIS[C]	45	none	9	positive	7	HLA-DRB1*15		
	MS-6	F	RRMS	29	none	16	positive	12	HLA-DRB1*15		
	MS-7	M	RRMS	37	none	6	positive	15	HLA-DRB1*11		
	MS-8	M	RRMS	29	none	4	positive	11	HLA-DRB1*03,15		
	MS-9	F	RRMS	33	none	12	positive	5	HLA-DRB1*07,15		
	MS-10	F	RRMS	32	none	60	positive	10	HLA-DRB1*04,15		
Database	Patient	Sex	Diagnosis	Age (years)	Treatments	Disease duration (months)	Last treatment-sampling interval	CSF, brain and blood collection	HLA-DR	Death-sampling interval	Cause of death
Laplaud-DB	MS-1	M	SPMS	45	Azathioprine-Cyclophosphamide Mitoxantrone-Interferon Methotrexate	12	24 months	*postmortem*	DRB1*0102/1501	12 h	Lung cancer
	MS-2	F	PPMS	66	none	10	treatment naïve	*postmortem*	DRB1*1101/1302	6 h	Pulmonary infection
	MS-3	F	PPMS	54	Mitoxantrone	23	120 months	*postmortem*	DRB1*0301(0350)/ 1301	8 h	Pulmonary infection
Database	Patient	Sex	Diagnosis	Age (years)	Treatments	CSF IgG OCB	WBC[D] in CSF	CSF volume (mL)			
Muraro-DB	MS-1	M	RRMS	25	none	negative	2.2×10^5	11			
	MS-2	F	RRMS	29	none	positive	2×10^5	11			
	MS-3	F	RRMS	33	none	positive	1.3×10^5	10			
	MS-4	F	RRMS	28	none	positive	1.3×10^5	11			
	MS-5	M	RRMS	41	none	positive	1×10^5	10			

[A] OCB: oligoclonal bands.
[B] BBB: blood-brain barrier.
[C] Clinically isolated syndrome.
[D] White blood cells.

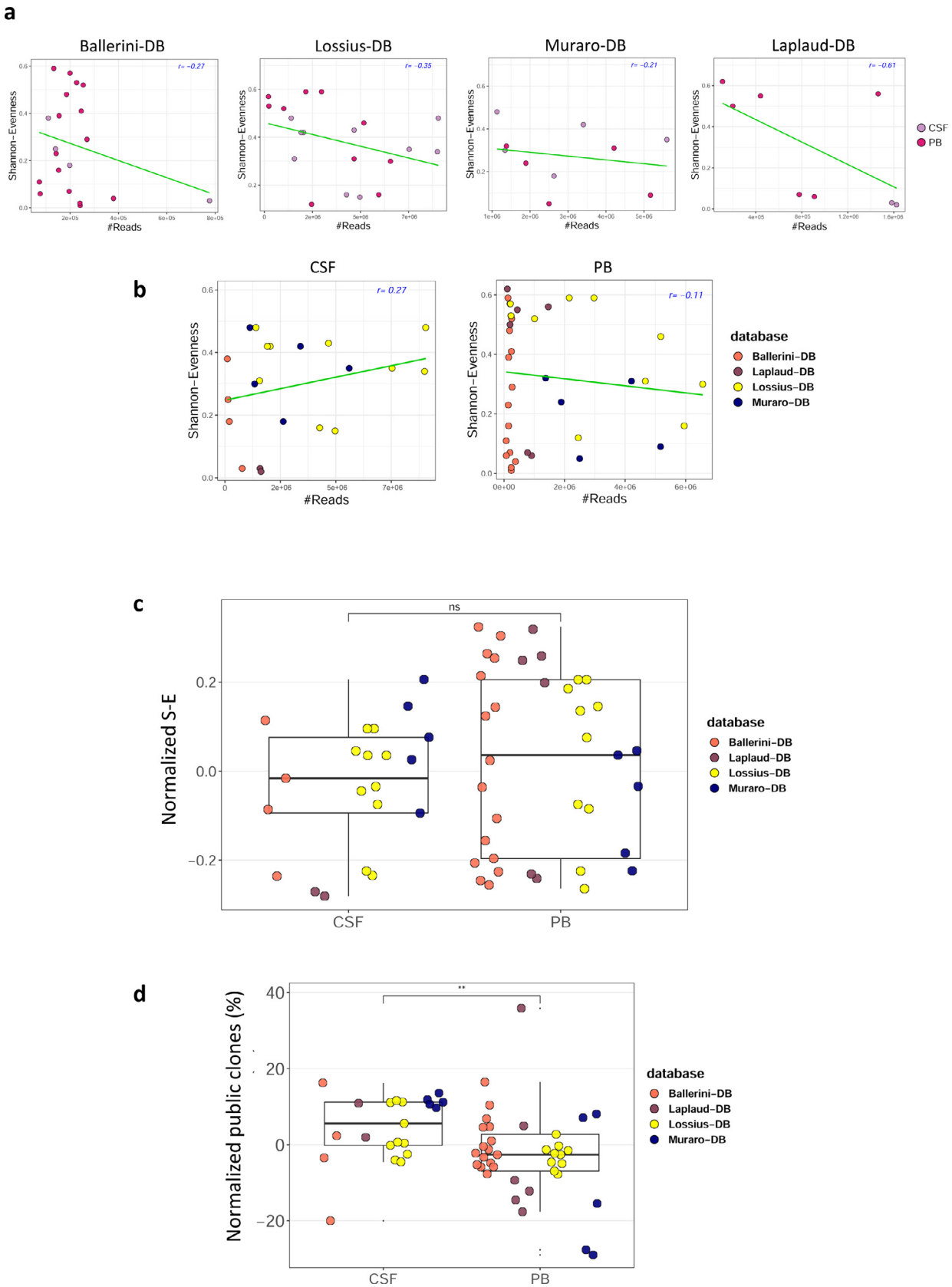

Fig. 1. TCR Vβ sequencing data quality assessment. (a) Pearson correlation (r) between Shannon-Evenness and #Reads (number of total amino acid CDR3 Vβ sequences, CDR3s-a.a.) in TCR repertoires reported by a database (from left to right: Ballerini-DB, Lossius-DB, Muraro-DB and Laplaud-DB). The correlation coefficient value is reported on the upper right within each graph. Each dot represents a patient TCR repertoire and is colored by compartment (cerebrospinal fluid [CSF] and peripheral blood [PB]). (b) Pearson correlation (r) between Shannon-Evenness and #Reads in TCR repertoires reported by compartment (CSF, left plot; PB, right plot). The correlation coefficient value is reported on the upper right within each graph. Each dot represents a patient TCR repertoire and is colored by database. (c) Repertoire statistics of normalized Shannon-Evenness (S-E) of the analyzed databases reported by compartment (CSF and PB). Each dot represents a TCR repertoire and is colored based on the database. (d) Public clones normalized percentage (%) across databases, reported by compartments. Two-sample Mann-Whitney test was used ($^{**}p<0.01$).

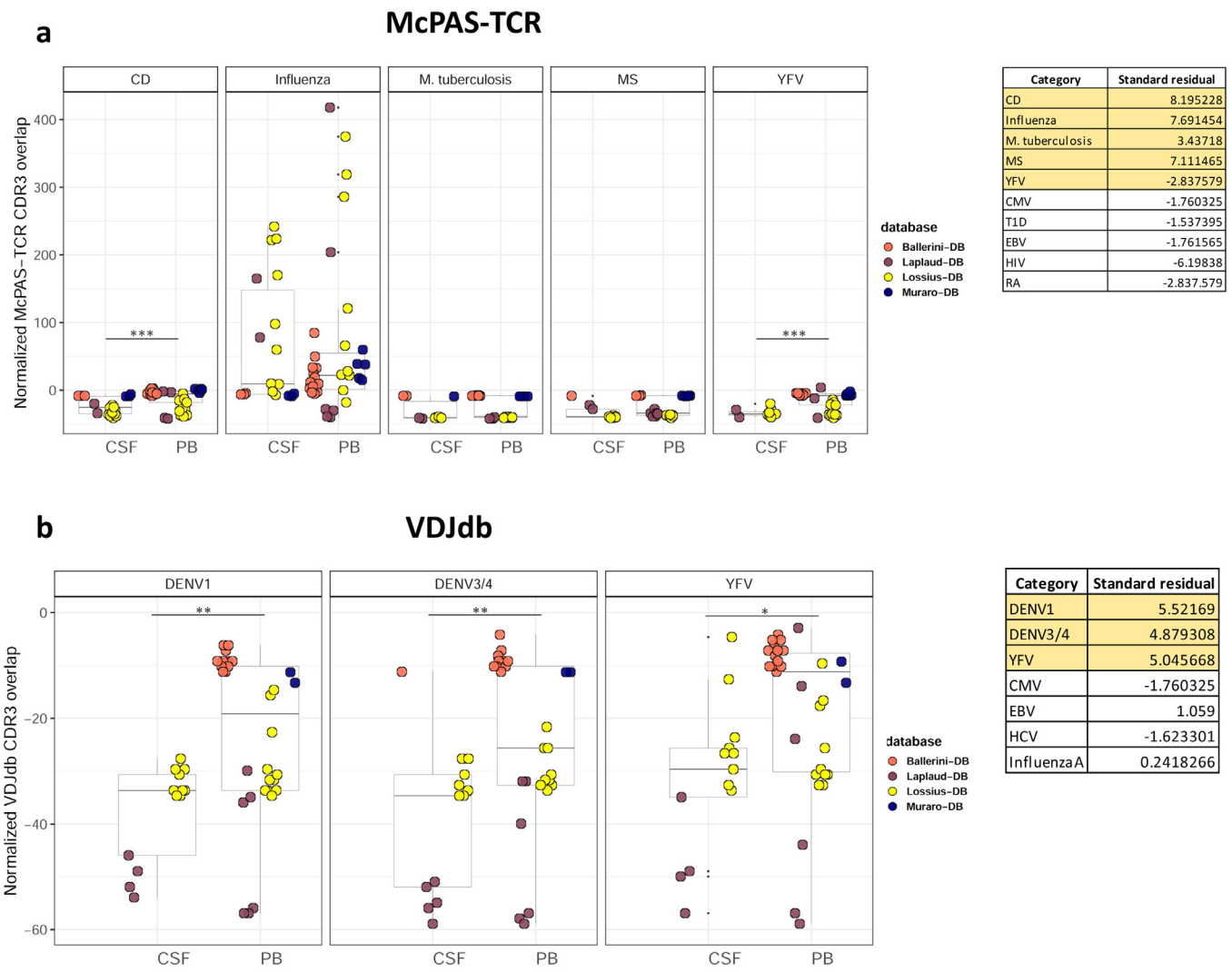

Fig. 2. CDR3 Vβ amino acid sequences overlap between the analyzed TCR databases and the public databases McPAS-TCR and VDJdb. (a) Graph reports the normalized CDR3s-a.a. overlap (absolute number) between the analyzed TCR database compartments (CSF and PB; x-axis) and McPAS-TCR by disease category. (b) Normalized CDR3s-a.a. overlap (absolute number) between the analyzed TCR database compartments (CSF and PB; x-axis) and VDJdb by antigen species. The standardized Pearson residual for disease-associated CDR3s-a.a. detected in the MS data for each disease category is reported on the right side of both (a) and (b) graphs. One-tailed Fisher's exact test with Bonferroni correction and two-sample Mann-Whitney test for CSF and PB comparison were used (*$p<0.05$; **$p<0.01$; ***$p<0.001$) (CD=Celiac disease; CMV=Cytomegalovirus; EBV=Epstein Barr virus; HCV=Hepatitis C virus; HIV=Human Immunodeficiency virus; MS=Multiple Sclerosis; RA=Rheumatoid arthritis; T1D=Type 1 Diabetes; YFV=Yellow Fever virus; DENV=Dengue virus).

only in Laplaud-DB. In our sample, CSF and brain repertoires share a total of 7150 CDR3s-a.a. (Fig. 3a). Therefore, we tested for overrepresented disease-associated CDR3s-a.a. from public databases as performed for Fig. 2, and p-values were corrected for multiple testing using the Bonferroni correction. We found significant overlap with the following categories: CD and MS from McPAS-TCR (adjusted p-value$_{CD}$=1.12^{-05}; adjusted p-value$_{MS}$=4.36^{-35}) (Fig. 3b) and DENV-1 from VDJdb (adjusted p-value=0.0012) (Fig. 3c).

8.4. Sequence overlap with publicly available healthy donor T-cell receptor databases

In order to evaluate the sequence overlap between MS and healthy donors (HD) TCR repertoire, we utilized a recently published database [32] reporting HD TCR repertoire (Soto-DB) derived from peripheral T cells (Fig. 4a). Soto-DB is characterized in Supplementary Fig. 2c.

We found that PB shares more CDR3s-a.a. with HD TCR repertoires compared to CSF (p-value=0.007; two-sample Mann-Whitney test) (Fig. 4a).

Cluster analysis of clonal expansion across repertoires gives the possibility to link similarity of TCR repertoire polarization by the use of established Hill-based evenness profiles [25,26]. The heatmap in Fig. 4b reports the hierarchical clustering of clonal polarization comparing MS group with HD from Soto-DB, independently from the compartment. This result suggests we may differentiate MS from HD on the basis of homogeneity of clonal polarization values.

8.5. T-cell receptor sequence similarity is higher in cerebrospinal fluid compared to peripheral blood

We next analyzed the TCR repertoire architecture using Vβ clonal networks (see Methods). Repertoire architecture directly reflects antigen recognition breadth, based on sequence similarity within a TCR repertoire: degree of similarity is in indirect proportion with antigen recognition breadth [25,28,29]. For each patient, we determined the Vβ CDR3s-a.a. similarity at one a.a. level (Fig. 5a–d).

We found that CSF repertoires of MS patients share a similar clone architecture and connectivity (Fig. 5a, left network in 5b and 5c), whereas in one PPMS CSF (CSF2, Ballerini-DB) TCR Vβ clones are in most part (98%) connected (Fig. 5b, right network). Of note, the other

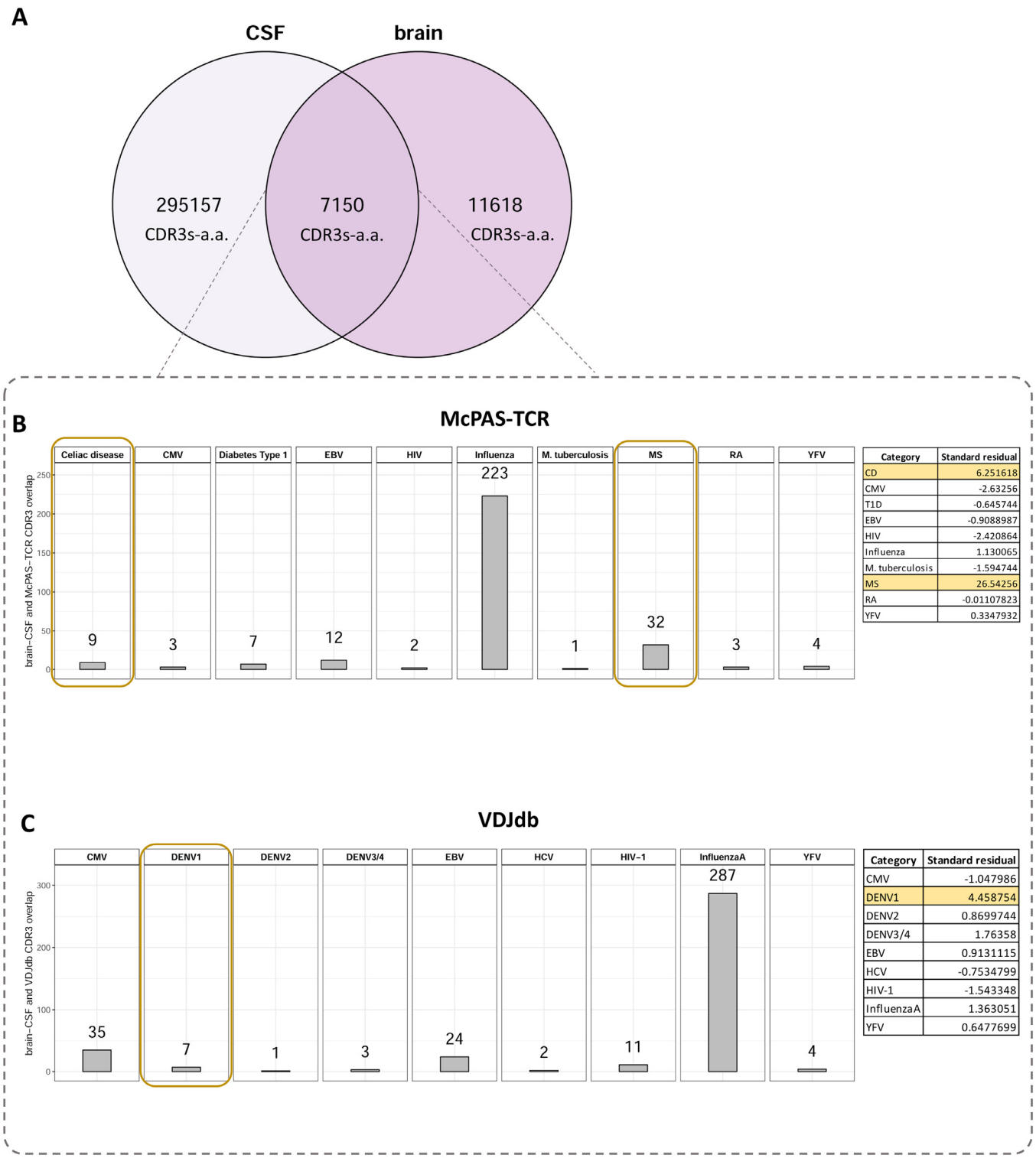

Fig. 3. Shared CDR3 Vβ amino acid sequences between CSF and brain repertoires and overlap with public databases by disease category and antigen species. (a) Venn diagram shows the absolute number of shared CDR3s-a.a. (N = 7150) between CSF (295,157 CDR3s-a.a.) and brain (11,618 CDR3s-a.a.) TCR repertoires. (b, c) CDR3s-a.a. overlap (absolute number) between selected CSF-brain shared CDR3s-a.a. and McPAS-TCR (b) or VDJdb (c) by disease category or by antigen species, respectively. The exact number of shared CDR3s-a.a. for each category is reported over bars. The standardized Pearson residual for disease-associated CDR3s-a.a. detected in the MS data for each disease category is reported below each (b) and (c) graphs. One-tailed Fisher's exact test and Bonferroni correction were used (CMV=Cytomegalovirus; EBV=Epstein Barr virus; HIV=Human Immunodeficiency virus; MS=Multiple Sclerosis; RA=Rheumatoid arthritis; T1D=Type 1 Diabetes; YFV=Yellow Fever virus; DENV=Dengue virus; HCV=Hepatitis C virus).

PPMS patient (Laplaud-DB) (Fig. 5d) is more similar, in terms of TCR architecture, to RRMS patients rather than the PPMS patient from Ballerini-DB.

In order to quantify the clone sequence similarity architecture similarity, in Fig. 5E we reported the percentage of connected clones. This percentage is significantly higher in CSF repertoires compared to PB (p-value=0.0001; two-sample Mann-Whitney test). We, therefore, conclude that broad antigen recognition breadth is a repertoire quality that differs among compartments, distinguishing CSF from PB.

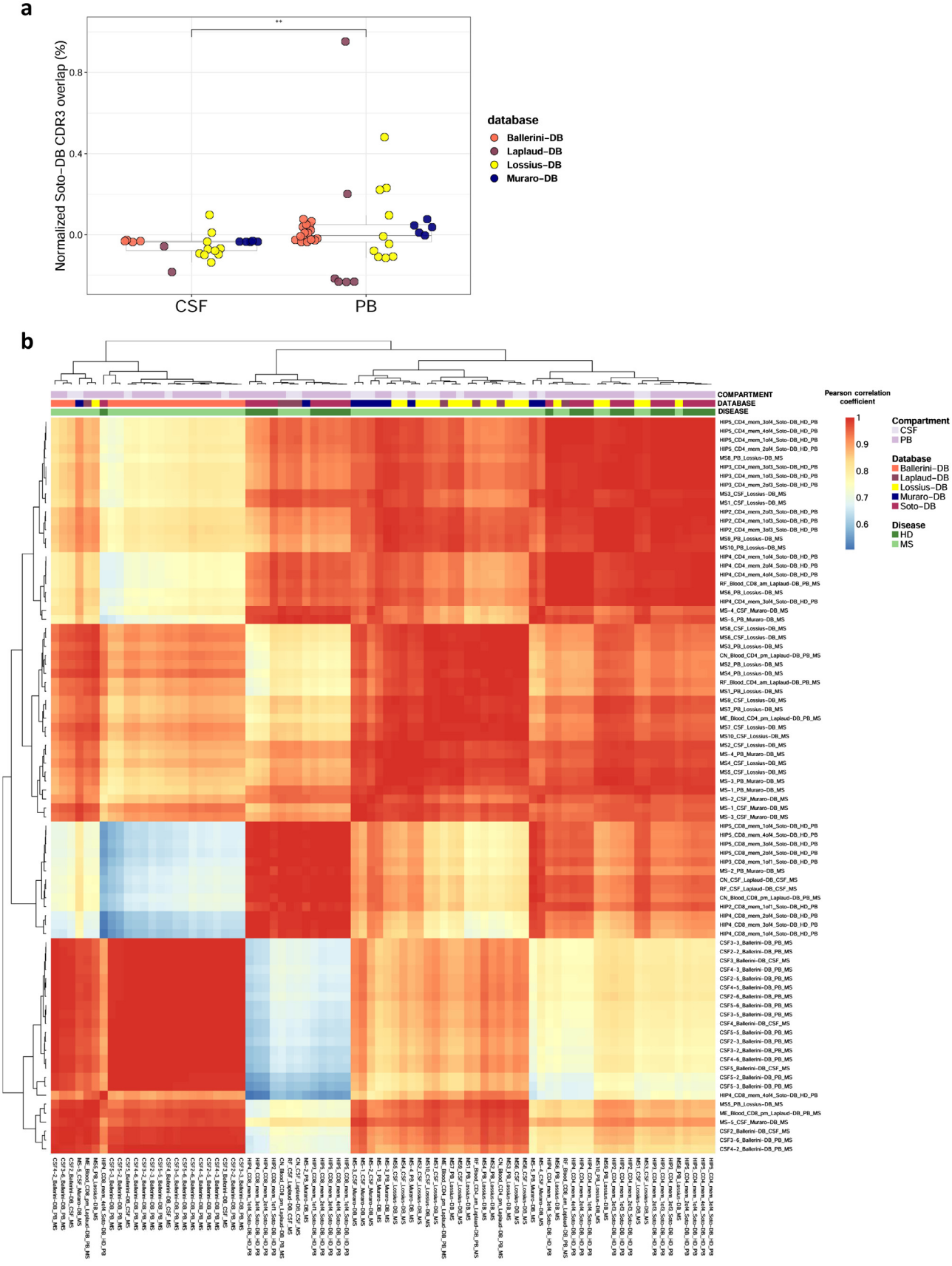

Fig. 4. CDR3 Vβ amino acid sequence overlap between the TCR databases and healthy donors and TCR clonal expansion across Multiple Sclerosis and healthy donors' repertoires. (a) Normalized CDR3s-a.a. overlap percentage (%) between the analyzed TCR databases reported by compartment (CSF and PB) and healthy donors (HD) TCR database (Soto-DB). Two-sample Mann-Whitney test was used (**$p<0.01$). (b) Heatmap reports the pairwise Pearson correlation between Shannon-Evenness profiles (or "evenness profiles") of TCR repertoires across MS and HD (Soto-DB) databases. Pearson correlation values range from ≈0.5 (blue) to ≈1 (red). Color bars on the top of the heatmap indicate compartment (different shades of pink for CSF and PB), database (each database is reported in a different color) and disease group (dark green for HD and light green for MS). Hierarchical clustering of evenness profiles was performed using correlation-based distance. Average linkage clustering was used by default. The x- and y-axis of the heatmap report the complete list of all TCR repertoires.

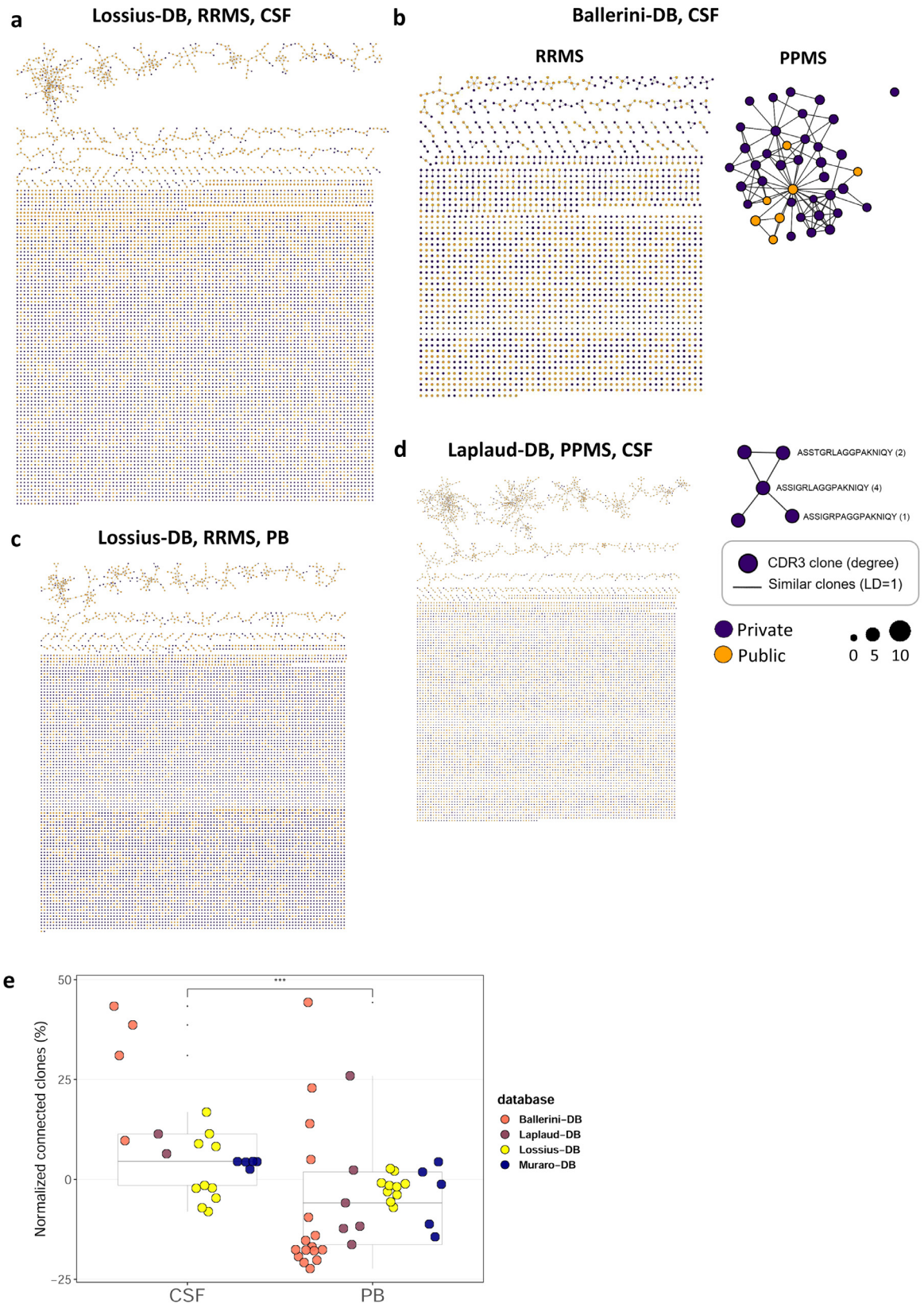

Fig. 5. TCR repertoire architecture represented as clonal networks and statistics. Figure shows five representative TCR repertoires displayed as clonal networks: (a) CSF-TCR repertoire of one Relapsing-Remitting MS (RRMS) patient from Lossius-DB; (b) CSF-TCR repertoire of two MS patients from Ballerini-DB, one diagnosed with RRMS (left network) and the other diagnosed with Primary Progressive MS (PPMS) (right network); (c) PB-TCR repertoire of a patient with RRMS from Lossius-DB; (d) PPMS patient CSF-TCR repertoire from Laplaud-DB. In all clonal networks, each dot represents a clone (a single CDR3-a.a. sequence) and connections are made between clones that differ for only 1 a.a. (Levenshtein distance [LD] = 1). Private clones are blue and public clones are yellow. Dots size depends on clone frequency logarithm. (e) Normalized connected clones percentage (%) in TCR repertoires from the databases visualized by compartment (CSF and PB). Two-sample Mann-Whitney test was used ($^{***}p<0.001$).

9. Discussion

We have recently developed a multidimensional, statistical analysis of TCR Vβ sequencing data that allowed us to identify differences among peripheral repertoire in MS patients under two successful treatments [25] Our primary aim in the present work was to apply the same method to investigate the TCR repertoire in different compartments (CSF and PB) in MS patients, in order to establish whether there is a detectable molecular signature due to the combination of disease/compartment. To this end, we collected and pooled TCR repertoire Vβ sequences obtained by HTS methods derived from three independent studies performed on CSF and PB in MS patients [9–11]. We also included a recently developed TCR repertoire database from healthy donors (HD) [32]. Among the few studies performed by HTS on MS, we included those with comparable patient characteristics (age, gender ratio, HLA type, treatments, etc.), in order to obtain a homogeneous cohort of patients. We excluded a study on TCR repertoire (performed by HTS) investigating MS and pregnancy [33], as the included population was different from the others. By our strategy, we collected a total of 239 775 966 CDR3-a.a. Vβ sequences from a cohort of 22 MS patients. Our methodological approach is based on pooling TCR Vβ sequences together [34], if the quality of the data is satisfactory. Therefore, after we assessed that data quality was good (Fig. 1a, 1b), we proceeded with the statistics.

First, we did not find any significant difference between CSF and PB in S-E, reflecting TCR Vβ clonal distribution and prevalence (Fig. 1c). This result may point out that taking into account exclusively the clonal expansion parameter to track repertoire polarization, as it has been done in previous studies [11,35], may be a misleading approach, promoting the idea that CSF and PB in MS could be different in terms of clonal expansion, which is not exact.

We then described the presence of public clones within patients in the two compartments (Fig. 1d). Indeed, one striking result of the human TCR repertoires deep-sequencing approaches is the detection of public TCR Vβ clones, with a rate of sharing around an average of 15% between two individuals [36]. This unexpected reduction of the inter-individual TCR Vβ diversity allows speculating about the role of public clones in immunity, suggesting that these sequences may predominate in the response against relevant pathogens. We compared the percentage of shared Vβ sequences ("public clones") within our samples and, interestingly, results showed that CSF repertoires are more public compared to PB repertoires (Fig. 1d). This may indicate that, in MS patients, CSF-TCR repertoire shares the same TCR Vβ chain more frequently than it happens in the periphery. Whether this is linked to a specific role in the disease pathology of these Vβ clones, or it is due to the specific compartment, should be investigated comparing brain repertoires derived from other neurological diseases, inflammatory and not. Of note, whether the presence of public clones indicates common antigen specificities is still under debate. Furthermore, the increase of shared sequences in CSF may depend on many different reasons [37], and could be enriched by bystander inflammatory mechanisms or by the state of the BBB, independently from the recognition of an antigen. It is worth mentioning that our calculation of public clones among repertoires was performed on the total of CDR3-a.a. Vβ sequences from each repertoire, without any sub-selection of sequencing data based on clonal expansion or clonal frequency degree: we believe that investigating a disease of unknown etiology, such as MS, does not allow to prioritize a Vβ clone rather than another because, as far as we know, each clone may carry the same weight in terms of biological meaning and link to the disease pathogenesis. Furthermore, we acknowledge that further investigations into approaches to comparing samples of diversity and T-cell subpopulation structure are necessary [38].

Although predicting antigen binding from the receptor sequence is challenging [39,40], the recent development and public availability of TCR Vβ sequence databases such as VDJdb and McPAS-TCR [30,31] allowed us to check for the presence of antigen/disease-associated TCR Vβ sequences (Fig. 2). First, we used a Fisher's test approach to compare expected and observed sequence data between the databases and our MS repertoires. Six sequence categories were found to be significantly enriched in our MS repertoires: three viruses-associated ones (DENV, Influenza and YFV), *M. tuberculosis*, and two disease-associated (CD and MS). Indeed, there is no previous report on a possible association among DENV, YFV in the MS European population: this is probably due to the possibility to have shared sequences in our TCR repertoire that are not necessarily associated with one exclusive antigen specificity. We have to underline that YFV repertoires are missing in two of our MS databases, this may impact significance. Interestingly, in a Taiwan cohort of MS patients, DENV has been associated with autoimmune and neurological diseases [41] and DENV infections have been suggested to lead to various diseases in humans [42]. On the other hand, finding sequences associated with CD is in agreement with the idea that MS shares the genetic basis with other complex diseases (e.g., rheumatoid arthritis, diabetes, CD), involving coincidence of variants involved in the immune response [43,44]. Furthermore, a recent report investigated bacterial infections, e.g. *M. tuberculosis*, as a component of a "multiple-hit" model of neurodegeneration in several diseases, MS included [45]. Of note, also MS-associated sequences are not randomly distributed in our database: this underlines the importance to increase, in the future, our knowledge on MS-associated and shared sequences, in order to detect common antigen-driven pathogenic responses. Second, we compared these sequences between CSF and PB repertoires, although the absence of some of them in CSF Ballerini-DB and CSF Muraro-DB may impact significance. The overlap of CD, YFV, DENV-1 and DENV3/4-associated sequences was increased in PB compared to CSF. Overall, our interpretation is that the two compartment repertoires are not exactly mirroring each other for several aspects, including the presence of shared sequences associated with antigenic responses. Focusing on shared sequences between CSF and brain lesion TCR repertoires (the last from Laplaud-DB) (Fig. 3), we identified the presence of sequences associated with autoimmune diseases (MS and CD) (Fig. 3b) and with DENV-1 (Fig. 3c). This finding is intriguing: the presence of these sequences in white matter brain lesions, an hallmark of MS disease, stresses the fact that at least a part of the antigen driven repertoire during an autoimmune reaction is public, as it was observed in the immune response directed against certain pathogens [46]. In the future, sharing TCR Vβ sequences with the scientific community will be a valid approach to furtherly increase the number of sequences available for a comprehensive analysis.

Leveraging a recent publication [32], we had the chance to compare our data with the TCR repertoire of HD (characterized in Supplementary Fig. 2c). We exploited this freely available database to check the CDR3-a.a. Vβ sequences overlap between HD and MS compartments (Fig. 4a). As we reported for the overlap analysis with McPAS-TCR and VDJdb databases, CSF-TCR repertoire overlap with HD is lower compared to PB. This finding may depend on the compartment itself, on the disease, or both. Previously, the TCR repertoire of HD has been compared to MS by hierarchical cluster analysis [34], distinguishing HD from MS in terms of Vβ clone overlap and frequencies. Here, we performed hierarchical clustering analysis by correlating S-E values across repertoires (Fig. 4b). We showed that the TCR Vβ clone distribution and polarization across MS repertoires may be distinguished from HD on the basis of the disease, and not on the compartment. It would be of interest to widen this analysis and to check whether treatments, that successfully induce TCR repertoire reconstitution [47] (e.g., autologous hematopoietic stem cell transplantation, anti-CD52 monoclonal antibody, cladribine), impact MS patients clustering. Furthermore, it would be interesting to study whether other autoimmune diseases, that do not involve the CNS, behave in a similar way or whether this is a MS-specific feature.

Finally, we analyzed the TCR repertoire similarity architecture. This analysis allows investigating the TCR repertoire's antigen recognition breadth (Fig. 5), since similar Vβ clones are suggested to recognize the same antigens.[25,28,29] We found that the CSF-TCR similarity architecture of one PPMS patient's repertoire (patient CSF2 from Ballerini-DB; Fig. 5b, right network) differs from other CSF-TCR repertoires, including another PPMS from Laplaud-DB (Fig. 5d) and RRMS patients (Fig. 5a and left network in Fig. 5b). The PPMS patient CSF2 is characterized by higher Vβ clone connectivity (about 98% of all his Vβ clones are connected, thus similar) and a low number (41) of unique CDR3s-a.a. The other PPMS patient (Laplaud-DB, Fig. 5d) does not show the same TCR similarity architecture: this may be due to a different HLA haplotype, disease duration, or timing of CSF sampling (at diagnosis or *post-mortem*) (see Table 2). Furthermore, the PPMS patient CSF2 experienced a clinical relapse after the lumbar puncture, and may therefore be defined "active". It would be worthwhile to verify whether a similar TCR architecture may be found in other PPMS patients. As previously shown [10,11], CSF and CNS parenchyma share TCR Vβ sequences. However, we did not find any Vβ sequence overlap between CSF and paired PB of the patient CSF2.

In conclusion, our comprehensive analysis allowed us to identify those characteristics of the TCR repertoire that may be shaped by MS and by the anatomical compartment within MS cohorts. We stress the fact that a *consensus* in TCR repertoire analysis is needed, in order to ensure scientific discussion on a common basis [48,49].

Contributors

R.A.: data analysis, data interpretation, figures, and writing; M.C.: data analysis, statistical methods, manuscript revision; V.G.: data analysis and manuscript revision; A.C.: data analysis; L.M.: clinical data collection; A.B.: clinical data collection; A.M.R.: clinical data collection; T.B.: data collection and analysis; A.A.: cell sorting; P.A.M.: data collection; D.L.: data collection; A.L.: study conceptualization and data collection; C.B.: study conceptualization, data interpretation, and writing.

Data sharing statement

The R scripts used for the statistical analysis and data visualisation are stored in the following Github repository: https://github.com/roberta91/TCR_MS_diversity. Ballerini-DB TCR sequencing data was deposited under the following doi: https://doi.org/10.5281/zenodo.3703311. [25] Other raw data is available upon request.

Declaration of Competing Interest

A.L. reports grants from Sanofi Genzyme, outside the submitted work. D.L. reports grants from EDMUS Foundation, from ARSEP Foundation and from ANR, and personal fees from Biogen, BMS, Alexion, Merck, Sanofi and Roche, outside the submitted work. P.A.M. reports personal fees from Jasper Therapeutics and from Magenta Therapeutics, outside the submitted work. V.G. declares advisory board positions in aiNET GmbH and Enpicom B.V.

Acknowledgments

We acknowledge the Italian Multiple Sclerosis Foundation (FISM) that financially supported Ballerini-DB data collection (grant #2015 R02).

Supplementary materials

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.ebiom.2021.103429.

References

[1] Atlas of MS 2013: Mapping Multiple Sclerosis Around the World. London: Multiple Sclerosis international federation; 2013. MS International Federation.
[2] Compston A, Coles A. Multiple sclerosis - ScienceDirect. Mult Scler 2008;372:1502–17.
[3] Milo R, Kahana E. Multiple sclerosis: geoepidemiology, genetics and the environment. Autoimmun Rev 2010;9:A387–94.
[4] Carroll WM. McDonald MS diagnostic criteria: evidence-based revisions. Mult Scler J 2017;24:92–5 2018.
[5] Magliozzi R, Marastoni D, Rossi S, Castellaro M, Mazziotti V, Pitteri M, Gajofatto A, Monaco S, Benedetti MD, Calabrese M. Increase of CSF inflammatory profile in a case of highly active multiple sclerosis. BMC Neurol 2019;19:231.
[6] Skulina C, Schmidt S, Dornmair K, Babbe H, Roers A, Rajewsky K, Wekerle H, Hohlfeld R, Goebels N. Multiple sclerosis: brain-infiltrating CD8+ T cells persist as clonal expansions in the cerebrospinal fluid and blood. Proc Natl Acad Sci 2004;101:2428–33.
[7] Jacobsen M, Cepok S, Quak E, Happel M, Gaber R, Ziegler A, Schock S, Oertel WH, Sommer N, Hemmer B. Oligoclonal expansion of memory CD8+ T cells in cerebrospinal fluid from multiple sclerosis patients. Brain 2002;125:538–50.
[8] Rosati E, Dowds CM, Liaskou E, Henriksen EKK, Karlsen TH, Franke A. Overview of methodologies for T-cell receptor repertoire analysis. BMC Biotechnol 2017;17(1):61.
[9] Lossius A, Johansen JN, Vartdal F, Robins H, Jūratė Šaltytė B, Holmøy T, Olweus J. High-throughput sequencing of TCR repertoires in multiple sclerosis reveals intrathecal enrichment of EBV-reactive CD8 + T cells. Clin Immunol *Eur J Immunol* 2014;44:3439–52.
[10] Salou M, Garcia A, Michel L, Gainche-Salmon A, Loussouarn D, Nicol B, Guillot F, Hulin P, Nedellec S, Baron D, et al. Expanded CD8 T-cell sharing between periphery and CNS in multiple sclerosis. Ann Clin Trans Neurol 2015;2:609–22.
[11] Sousa A, Johnson KR, Nicholas R, Darko S, Price DA, Douek DC, Jacobson S, Muraro PA. Intrathecal T-cell clonal expansions in patients with multiple sclerosis. Ann Clin Transl Neurol 2016;3:422–33.
[12] Planas R, Metz I, Martin R, Sospedra M. Detailed characterization of T cell receptor repertoires in Multiple Sclerosis brain lesions. Front Immunol 2018;9:509.
[13] Brown AJ, Snapkov I, Akbar R, Pavlović M, Miho E, Sandve GK, Greiff V. Augmenting adaptive immunity: progress and challenges in the quantitative engineering and analysis of adaptive immune receptor repertoires. Mol Syst Des Eng 2019;4:701–36.
[14] Junker A, Ivanidze J, Malotka J, Eiglmeier I, Lassmann H, Wekerle H, Meinl E, Hohlfeld R, Dornmair K. Multiple Sclerosis: T-cell receptor expression in distinct brain regions. Brain 2007;130(Pt 11):2789–99.
[15] van Nierop GP, van Luijn MM, Michels SS, Melief MJ, Janssen M, Langerak AW, Ouwendijk WJD, Hintzen RQ, Verjans GMGM. Phenotypic and functional characterization of T cells in white matter lesions of multiple sclerosis patients. Acta Neuropathol 2017;134(3):383–401.
[16] Kaplinsky J, Arnaout R. Robust estimates of overall immune-repertoire diversity from high-throughput measurements on samples. Nat Commun 2016;7:11881.
[17] Core Team R. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2018. Available at https://www.r-project.org/.
[18] Hadley W. ggplot2: elegant graphics for data analysis. New York, NY: Springer-Verlag New York; 2016.
[19] Kassambara A. ggpubr: "ggplot2" based publication ready plots. Available at https://rpkgs.datanovia.com/ggpubr/index.html.
[20] Csardi G, Nepusz T. The igraph software package for complex network research. Interjournal Complex Syst 2006;1695 Available at http://igraph.org.
[21] Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics 2016;32:2847–9.
[22] Gaujoux R, Seoighe C. A flexible R package for nonnegative matrix factorization. BMC Bioinform 2010;11:367.
[23] Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 2003;13:2498–504.
[24] Nygaard V, Rødland EA, Hovig E. Methods that remove batch effects while retaining group differences may lead to exaggerated confidence in downstream analyses. Biostatistics 2016;17(1):29–39.
[25] Amoriello R, Greiff V, Aldinucci A, Bonechi E, Carnasciali A, Peruzzi B, Repice AM, Mariottini A, Saccardi R, Mazzanti B, Massacesi L, Ballerini C. The TCR repertoire reconstitution in Multiple Sclerosis: comparing one-shot and continuous immunosuppressive therapies. Front Immunol 2020;11:559.
[26] Greiff V, Bhat P, Cook SC, Menzel U, Kang W, Reddy ST. A bioinformatic framework for immune repertoire diversity profiling enables detection of immunological status. Genome Med 2015;7:49.
[27] Greiff V, Miho E, Menzel U, Reddy ST. Bioinformatic and statistical analysis of adaptive immune repertoires. Trends Immunol 2015;36:738–49.
[28] Miho E, Roškar R, Greiff V, Reddy ST. Large-scale network analysis reveals the sequence space architecture of antibody repertoires. Nat Commun 2019;10:1321.
[29] Madi A, Poran A, Shifrut E, Reich-Zeliger S, Greenstein E, Zaretsky I, Arnon T, Laethem FV, Singer A, Lu J, et al. T cell receptor repertoires of mice and humans are clustered in similarity networks around conserved public CDR3 sequences. Elife 2017;6:e22057.
[30] Tickotsky N, Sagiv T, Prilusky J, Shifrut E, Friedman N. McPAS-TCR: a manually curated catalogue of pathology-associated T cell receptor sequences. Bioinformatics 2017;33:2924–9.

[31] Shugay M, Bagaev DV, Zvyagin IV, Vroomans RM, Crawford JC, Dolton G, Komech EA, Sycheva AL, Koneva AE, Egorov ES, et al. VDJdb: a curated database of T-cell receptor sequences with known antigen specificity. Nucleic Acids Res 2018;46:D419–27.
[32] Soto C, Bombardi RG, Kozhevnikov M, Sinkovits RS, Chen EC, Branchizio A, Kose N, Day SB, Pilkinton M, Gujral M, Mallal S, Crowe Jr. JE. High frequency of shared clonotypes in human T cell receptor repertoires. Cell Rep 2020;32(2):107882.
[33] Ramien C, Yusko EC, Engler JB, Gamradt S, Patas K, Schweingruber N, Willing A, Rosenkranz SC, Diemert A, Harrison A, Vignali M, Sanders C, Robins HS, Tolosa E, Heesen C, Arck PC, Scheffold A, Chan K, Emerson RO, Friese MA, Gold SM. T cell repertoire dynamics during pregnancy in Multiple Sclerosis. Cell Rep 2019;29(4):810–5 e4.
[34] Shugay M, Bagaev DV, Turchaninova MA, Bolotin DA, Britanova OV, Putintseva EV, Pogorelyy MV, Nazarov VI, Zvyagin IV, Kirgizova VI, Kirgizov KI, Skorobogatova EV, Chudakov DM. VDJtools: unifying post-analysis of T cell receptor repertoires. PLoS Comput Biol 2015;11(11):e1004503.
[35] Muraro PA, Robins H, Malhotra S, Howell M, Phippard D, Desmarais C, de Paula Alves, Sousa A, Griffith LM, Lim N, Nash RA, Turka LA. T cell repertoire following autologous stem cell transplantation for multiple sclerosis. J Clin Invest 2014;124(3):1168–72.
[36] Elhanati Y, Sethna Z, Callan Jr CG, Mora T, Walczak AM. Predicting the spectrum of TCR repertoire sharing with a data-driven model of recombination. Immunol Rev 2018;284(1):167–79.
[37] Thomas PG, Crawford JC. Selected before selection: a case for inherent antigen bias in the T-cell receptor repertoire. Curr Opin Syst Biol 2019;18:36–43.
[38] Rempala GA, Seweryn M, Ignatowicz L. Model for comparative analysis of antigen receptor repertoires. J Theor Biol 2011;269(1):1–15.
[39] Bradley P, Thomas PG. Using T cell receptor repertoires to understand the principles of adaptive immune recognition. Annu Rev Immunol 2019;37:547–70.
[40] Greiff V, Yaari G, Cowell L. Mining adaptive immune receptor repertoires for biological and clinical information using machine learning. Curr Opin Syst Biol 2020;24:109–19.
[41] Li HM, Huang YK, Su YC, Kao CH. Increased risk of autoimmune diseases in dengue patients: a population-based cohort study. J Infect 2018;77(3):212–9.
[42] Dey L, Mukhopadhyay A. A graph-based approach for finding the dengue infection pathways in humans using protein-protein interactions. J Comput Biol 2020;27(5):755–68.
[43] Baranzini SE, Oksenberg JR. The genetics of Multiple Sclerosis: from 0 to 200 in 50 Years. Trends Genet 2017;33(12):960–70.
[44] Orrù V, Steri M, Sidore C, Marongiu M, Serra V, Olla S, et al. Complex genetic signatures in immune cells underlie autoimmunity and inform therapy. Nat Genet 2020;52(10):1036–45.
[45] Patrick KL, Bell SL, Weindel CG, Watson RO. Exploring the "Multiple-Hit Hypothesis" of neurodegenerative disease: bacterial infection comes up to bat. Front Cell Infect Microbiol 2019;9:138.
[46] Cassotta A, Goldstein JD, Durini G, Jarrossay D, Baggi Menozzi F, Venditti M, Russo A, Falcone M, Lanzavecchia A, Gagliardi MC, Latorre D, Sallusto F. Broadly reactive human CD4+ T cells against Enterobacteriaceae are found in the naïve repertoire and are clonally expanded in the memory repertoire. Eur J Immunol 2021;51(3):648–61.
[47] Lünemann JD, Ruck T, Muraro PA, Bar-Or A, Wiendl H. Immune reconstitution therapies: concepts for durable remission in multiple sclerosis. Nat Rev Neurol 2020;16:56–62.
[48] Arnaout RA, Luning Prak ET, Schwab N, Rubelt F, The adaptive immune receptor repertoire community. The future of blood testing is the immunome. Front Immunol 2021;12:228.
[49] Pavlović M., Scheffer L., Motwani K., Kanduri C., Kompova R., Vazov N. et al. immuneML: an ecosystem for machine learning analysis of adaptive immune receptor repertoires. bioRxiv 2021.03.08.433891; https://doi.org/10.1101/2021.03.08.433891.