



# Sex bias in MHC I-associated shaping of the adaptive immune system

Tilman Schneider-Hohendorf<sup>a</sup>, Dennis Görlich<sup>b</sup>, Paula Savola<sup>c</sup>, Tiina Kelkka<sup>c</sup>, Satu Mustjoki<sup>c</sup>, Catharina C. Gross<sup>a</sup>, Geoffrey C. Owens<sup>d</sup>, Luisa Klotz<sup>a</sup>, Klaus Dornmair<sup>e</sup>, Heinz Wiendl<sup>a</sup>, and Nicholas Schwab<sup>a,1</sup>

<sup>a</sup>Department of Neurology, University of Muenster, 48149 Muenster, Germany; <sup>b</sup>Institute of Biostatistics and Clinical Research, University of Muenster, 48149 Muenster, Germany; <sup>c</sup>Hematology Research Unit Helsinki, Department of Clinical Chemistry and Hematology, University of Helsinki and Helsinki University Hospital Comprehensive Cancer Center, 00029 Helsinki, Finland; <sup>d</sup>Department of Neurosurgery, David Geffen School of Medicine at the University of California Los Angeles, Los Angeles, CA 90095; and <sup>e</sup>Institute of Clinical Neuroimmunology, University Hospital and Biomedical Center, Ludwig-Maximilians University Munich, 80539 Munich, Germany

Edited by K. Christopher Garcia, Stanford University, Stanford, CA, and approved January 12, 2018 (received for review September 14, 2017)

**HLA associations, T cell receptor (TCR) repertoire bias, and sex bias have independently been shown for many diseases. While some immunological differences between the sexes have been described, they do not fully explain bias in men toward many infections/cancers, and toward women in autoimmunity. Next-generation TCR variable beta chain (TCRBV) immunosequencing of 824 individuals was evaluated in a multiparametric analysis including HLA-A -B/MHC class I background, TCRBV usage, sex, age, ethnicity, and TCRBV selection/expansion dynamics. We found that HLA-associated shaping of TCRBV usage differed between the sexes. Furthermore, certain TCRBVs were selected and expanded in unison. Correlations between these TCRBV relationships and biochemical similarities in HLA-binding positions were different in CD8 T cells of patients with autoimmune diseases (multiple sclerosis and rheumatoid arthritis) compared with healthy controls. Within patients, men showed higher TCRBV relationship Spearman's rhos in relation to HLA-binding position similarities compared with women. In line with this, CD8 T cells of men with autoimmune diseases also showed higher degrees of TCRBV perturbation compared with women. Concerted selection and expansion of CD8 T cells in patients with autoimmune diseases, but especially in men, appears to be less dependent on high HLA-binding similarity than in CD4 T cells. These findings are consistent with studies attributing autoimmunity to processes of epitope spreading and expansion of low-avidity T cell clones and may have further implications for the interpretation of pathogenic mechanisms of infectious and autoimmune diseases with known HLA associations. Reanalysis of some HLA association studies, separating the data by sex, could be informative.**

T cell receptor repertoire | immunological sex bias | autoimmunity | multiple sclerosis | rheumatoid arthritis

The T cell receptor (TCR) repertoire is highly diverse (1) and can generate up to  $10^{15}$  theoretical TCR sequences (2, 3) with the effective size of the CD8 TCR repertoire in vivo being made up of  $10^5$  to  $10^8$  sequences (2, 4–6). The repertoire starts developing prenatally (7) and is subject to continuous adaptation during the course of a lifetime due to antigen exposure and other environmental influences (8, 9). While its complementary determining region 3 (CDR3) is hypervariable and binds to the MHC-presented antigen (10), the binding sites of the TCR to the MHC are germline encoded within the CDR1 and CDR2 of the TCR variable chains (11, 12). Many studies over the last decades have shown that evolutionarily conserved anchor amino acids in these CDRs determine how single TCR constructs bind to their respective MHC (13). Concerning the TCR variable beta chains (TCRBVs), 31 individual genes are known to encode a functional, nonpseudogenic, nonsegregated TCR, each with an individual CDR1 and CDR2 (14, 15). The HLA background is genetically predetermined and invariable (16), leading to the accepted notion that the HLA shapes the TCR repertoire during their continuous interaction (17–20). Very recently, it was shown that mutations in

TCR binding amino acids of MHC molecules specifically influence the binding to and usage of TCR variable alpha (TCRAV) and TCRBV chains (19), and that the presence of specific TCR rearrangements is an indication of the HLA background of the host and also of previous exposure to pathogens (20). Dysregulation in the TCR–antigen–MHC complex can lead to autoimmunity (21) in the case of uncontrolled immune cell expansion or cross-reactivity between foreign and autoantigens. Hypoexpansions or low TCR avidity might lead to susceptibility to infections. Additionally, to combat tumor-associated immune system suppression, cancer patients can now be treated with T cell-relevant immune checkpoint inhibitors such as blockade of CTLA-4 (22) or PD-1 (23).

Most immune system dysfunctions show strong sex bias: men are more susceptible to many infectious diseases and cancers of nonreproductive organs, whereas autoimmune diseases are much more common in women (24, 25). Even in psychological disorders such as major depression (26) and in neurodegenerative diseases such as Alzheimer's, recent studies have shown immune system involvement (27) and sex bias (28). Many studies have been conducted concerning either sex (25), HLA (29, 30), or TCR (31) associations with diseases. However, even though some sex differences in immune regulation have been reported (25, 32), it is still not completely clear why this bias exists. Multiparametric analysis of the influence of sex on HLA-mediated shaping of the TCR repertoire has been hindered by

## Significance

**Our analysis shows that sex can be associated with the degree to which HLA molecules propagate selection and expansion of T cells as characterized by their T cell receptor variable beta chain (TCRBV). Furthermore, CD8 T cells, especially in men with autoimmune diseases such as multiple sclerosis or rheumatoid arthritis, are capable of expanding in unison with other CD8 T cells, even without expressing TCRBVs with biochemical similarity in pivotal HLA-binding regions. Our findings add to the understanding of sex bias in diseases with immune system involvement: autoimmunity, infection, and cancer. These results also reveal pathology-associated TCRBVs of interest for future studies and support the argument for sex-separated analysis of HLA disease associations in general.**

Author contributions: T.S.-H., H.W., and N.S. designed research; T.S.-H. and N.S. performed research; D.G. and N.S. contributed new reagents/analytic tools; T.S.-H., D.G., and N.S. analyzed data; and T.S.-H., D.G., P.S., T.K., S.M., C.C.G., G.C.O., L.K., K.D., H.W., and N.S. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Published under the PNAS license.

<sup>1</sup>To whom correspondence should be addressed. Email: nicholas.schwab@ukmuenster.de.

This article contains supporting information online at [www.pnas.org/lookup/suppl/doi:10.1073/pnas.1716146115/-DCSupplemental](http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1716146115/-DCSupplemental).

Published online February 12, 2018.

low subject numbers in most studies. The advent of next-generation and high-throughput TCR immunosequencing in large cohorts of HLA-typed subjects now allows for a combined analysis of the TCR repertoire dependence on sex and HLA alleles (20).

## Results

**Influence of HLA-A and -B on TCRBV Usage Depends on the Biological Sex.** To assess whether the HLA background influences the frequency of T cells expressing certain TCRBVs in a sex-associated manner, TCRBV immunosequencing data from 673 individuals were analyzed (for details see [Datasets S1](#) and [S2](#)). The resulting dataset, therefore, contained the following parameters: sex (2; male, female), HLA-A and HLA-B genes (19; 8 HLA-A, 7 HLA-B, 4 most common haplotypes in the Caucasian population), and 31 TCRBV genes represented as percentage of the productive TCR repertoire in a sample. Only HLA with at least 40 individuals in the dataset or a balanced female-to-male ratio (percentage of female donors between 45% and 55%) and only functional TCRBV genes according to the International ImMunoGeneTics (IMGT) nomenclature (14, 15) were included. Nonfunctional TCRBV genes, pseudogenes, segregating, or open-reading frame genes (with potential mutations in the leader sequence) were not analyzed. A multiparametric analysis using multivariate linear models assessing the interaction of HLA and sex showed that the manner in which HLA genes influence TCRBV usage can differ between men and women (Fig. 1). Controlled for the known immunosequencing covariates, we found sex-associated differences in the propagation of TCRBV selection and expansion by the HLA [raw data in [Dataset S2](#), *P* values in [Dataset S3](#), and effect sizes (partial eta squared) in [Dataset S4](#)]. With the number of individuals in our study, 42 HLA/TCRBV combinations ( $P < 0.05$ ) were found in this dataset, of which three remained significant after correcting for multiple comparisons among those 589 assessed combinations, where TCRBV usage by a specific HLA in one sex was elevated compared with the other sex, independent of a potential influence of HLA or sex alone (Fig. S1 A–C). The Bonferroni-corrected sex-specific TCRBV expressions were challenged in a sensitivity analysis, where only healthy donors or only donors with determined (not inferred) HLA were assessed. The sex-specific differences could also be observed, when age or ethnicity (which was not available for all donors) or other relevant hidden covariates [as determined by

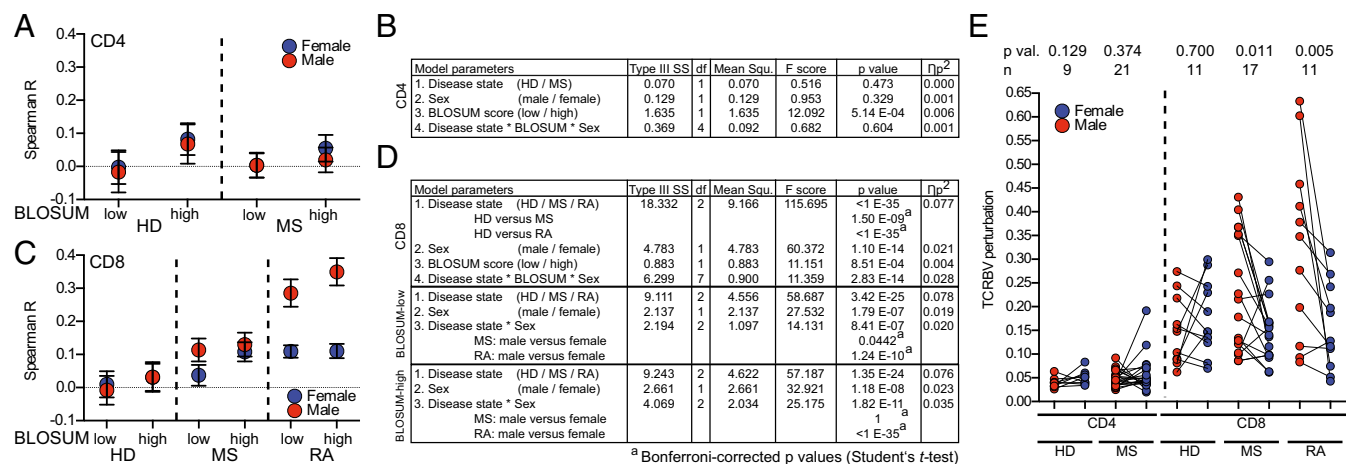
probabilistic estimation of expression residuals (PEER) analysis] (33, 34) were included in the model (Fig. S1 D and E).

## CD8 T Cells in Men with Autoimmune Diseases Are Less Dependent on Biochemical CDR1/2 Similarity for Concerted Selection/Expansion.

Having shown that the HLA background can alter TCRBV usage in a sex-dependent manner and given that MHC molecules interact with the TCR mainly via the CDR1 and -2, we hypothesized that there could be a functional relationship between how similar TCRBV chains are to one another (in these regions) and how they are selected in the thymus or expanded in the periphery via their HLA interaction. To assess this, the germline-encoded CDR1 and -2 sequences of the 31 functional TCRBV transcripts were compared via a Blocks Substitution Matrix (BLOSUM62) to estimate biochemical amino acid similarities (35). Resulting alignment matrices are used to score single amino acids at a specific position in two different sequences with respect to their biophysical and biochemical similarities. The scores of each pairing are then computed and indicate how similar two sequences are to each other. The current nomenclature groups TCRBV chains according to the nucleotide similarity of their complete DNA sequence. As our question did not relate to evolutionary homology per se, but rather to biological function with regard to TCR–MHC binding, we only assessed the known binding amino acids (position 28, and 29 in CDR1 as well as 46, 48, and 54 in CDR2) with the BLOSUM62 matrix (13) (Fig. S2A). The resulting 465 individual similarity scores between –9 and 29 were color coded (values listed in [Dataset S5](#)) and clustered using the Ward (36) method. This resulted in four optimal clusters [calculated via gap statistic (37)] (Fig. S2B), grouping TCRBV gene segments into the known homology-based subgroups or “families” (14) (i.e., subgroups 4, 5, 7, and 10, but not TCRBV06). However, the analysis also suggested relationships, which were not expected from the nomenclature (i.e., TCRBV05–08 and the TCRBV07 subgroup). To evaluate the in vivo relevance of these relationships in health and disease, the TCRBV usage by CD4 and CD8 T cells from age-matched male and female healthy donors (HD) or patients suffering from autoimmune diseases [multiple sclerosis (MS) and rheumatoid arthritis (RA)] was analyzed for potential correlations in all possible TCRBV chain combinations (raw data in [Dataset S6](#)) resulting in 465 TCRBV relationship Spearman’s rank correlation coefficients (TCRBV rhos, [Dataset S5](#)). To allow for a categorical

HLA	N	M	F	Ratio	0201	0301	0401	0402	0501	0504	0505	0506	0508	0601	0605	0606	0702	0706	0707	0708	0709	0901	1002	1003	1102	1301	1401	1501	1801	1901	2001	2501	2701	2801	2901
A11	57	30	27	0.47	0.472	0.868	0.869	0.475	0.978	0.994	0.79	0.777	0.033	0.616	0.553	0.819	0.725	0.276	0.91	0.96	0.727	0.231	0.072	0.357	0.863	0.604	0.697	0.4	0.545	0.23	0.856	0.653	0.19	0.258	0.256
A23	25	12	13	0.52	0.354	0.761	0.102	0.003	0.185	0.303	0.806	0.299	0.086	0.103	0.649	0.824	0.005	0.801	0.782	0.148	0.795	0.651	0.728	0.664	0.337	0.441	0.969	0.029	0.019	0.121	0.811	0.428	0.855	0.832	0.053
A24	124	62	62	0.5	0.267	0.863	0.9	0.084	0.611	0.019	0.293	0.308	0.614	0.249	0.806	0.059	0.801	0.441	0.019	0.182	0.58	0.636	0.553	0.122	0.369	0.913	0.527	0.113	0.48	0.353	0.32	0.379	0.441	0.078	0.369
A25	23	12	11	0.48	0.643	0.607	0.533	0.813	0.62	0.564	0.102	0.674	0.707	0.99	0.359	0.709	0.488	0.789	0.815	0.898	0.578	0.658	0.9	0.055	0.489	0.665	0.47	0.666	0.982	0.676	0.557	0.727	0.35	0.61	0.73
A26	49	30	19	0.39	0.853	0.026	0.19	0.156	5.4E-05	0.957	0.585	0.814	0.151	0.771	0.235	0.981	0.68	0.581	0.639	0.018	0.919	0.679	0.945	0.057	0.402	0.68	0.263	0.338	0.791	0.489	0.348	0.727	0.749	0.271	0.937
A29	42	27	15	0.36	0.98	0.424	0.548	0.772	0.479	0.893	0.505	0.549	0.906	0.416	0.361	0.272	0.75	0.493	0.009	0.6	0.3	0.883	0.156	0.043	0.72	0.935	0.573	0.406	0.705	0.143	0.459	0.929	0.722	0.173	0.932
A32	40	19	21	0.53	0.174	0.053	0.526	0.155	0.413	0.013	0.71	0.056	0.552	0.08	0.93	0.964	0.925	0.056	0.069	0.619	0.084	0.464	0.228	0.306	0.53	0.415	0.474	0.756	0.328	0.588	0.958	0.099	0.373	0.051	0.949
A68	44	21	23	0.52	0.68	0.835	0.844	0.34	0.149	0.248	6.0E-06	0.653	0.657	0.136	0.241	0.003	0.834	0.89	0.809	0.578	0.961	0.674	0.823	0.915	0.752	0.254	0.628	0.769	0.77	0.875	0.724	0.419	0.633	0.002	0.91
B13	23	12	11	0.48	0.085	0.106	0.596	0.928	0.898	0.489	0.956	0.921	0.356	0.719	0.199	0.466	0.526	0.76	2.8E-04	0.459	0.77	0.449	0.874	0.334	0.118	0.233	0.679	0.515	0.786	0.836	0.271	0.829	0.575	0.274	0.577
B14	38	20	18	0.47	0.223	0.096	0.895	0.431	0.516	0.897	0.14	0.674	0.504	0.103	0.234	0.891	0.926	0.821	0.93	0.582	0.912	0.789	0.891	0.915	0.963	0.576	0.869	0.415	0.708	0.71	0.817	0.419	0.483	0.821	0.807
B15	42	16	26	0.62	0.539	0.858	0.032	0.89	0.357	0.174	0.167	0.116	0.204	0.913	0.192	0.879	0.425	0.635	0.138	0.587	0.274	0.993	0.11	0.016	0.373	0.74	0.807	0.453	0.249	0.295	3.0E-06	0.186	0.2	0.008	0.931
B18	58	26	32	0.55	0.25	0.308	0.428	0.967	0.126	0.841	0.143	0.134	0.588	0.432	0.838	0.734	0.15	0.238	0.986	0.757	0.412	0.96	0.907	0.232	0.466	0.373	0.074	0.364	0.741	0.315	0.916	0.033	0.641	0.566	0.05
B27	46	24	22	0.46	0.02	0.884	0.045	0.642	0.701	0.311	0.375	0.136	0.458	0.201	0.103	0.004	0.622	0.679	0.636	0.038	0.074	0.069	0.798	0.503	0.128	0.317	0.192	0.283	0.159	0.005	0.222	0.051	0.914	0.419	0.331
B35	110	54	56	0.51	0.899	0.107	0.98	0.072	0.447	0.623	0.277	0.641	0.194	0.79	0.043	0.154	0.144	0.856	0.073	0.026	0.519	0.915	0.566	0.486	0.487	0.688	0.96	0.531	0.283	0.959	0.176	0.949	0.634	0.135	0.338
B51	73	32	41	0.56	0.73	0.615	0.932	0.465	0.132	0.71	0.87	0.251	0.062	0.855	0.813	0.518	0.506	0.559	0.713	0.388	0.347	0.173	0.168	0.061	0.078	0.979	0.361	0.827	0.62	0.816	0.199	0.72	0.677	0.026	0.425
A01B08	106	54	52	0.49	0.849	0.821	0.028	0.92	0.631	0.721	0.562	0.214	0.394	0.621	0.706	0.448	0.152	0.053	0.866	0.335	0.466	0.668	0.782	0.397	0.495	0.35	0.294	0.128	0.683	0.02	0.569	0.719	0.486	0.791	0.164
A02B07	72	42	30	0.42	0.8	0.407	0.029	0.281	0.353	0.803	0.994	0.257	0.134	0.294	0.115	0.971	0.798	0.551	0.499	0.35	0.74	0.009	0.963	0.42	0.33	0.253	0.809	0.468	0.672	0.827	0.831	0.189	0.925	0.353	0.815
A02B44	85	47	38	0.45	0.272	0.824	0.786	0.224	0.453	0.628	0.683	0.336	0.099	0.997	0.471	0.403	0.673	0.514	0.696	0.004	0.513	0.76	0.505	0.206	0.525	0.034	0.454	0.017	0.453	0.929	0.783	0.635	0.364	0.76	0.272
A03B07	74	36	38	0.51	0.506	0.134	0.866	0.743	0.554	0.622	0.232	0.064	0.73	0.424	0.895	0.562	0.746	0.02	0.621	0.044	0.076	0.651	0.182	0.743	0.731	0.352	0.254	0.027	2.1E-04	0.717	0.973	0.813	0.668	0.08	0.569

**Fig. 1.** Sex bias in HLA-associated shaping of the TCRBV usage. Shown are uncorrected *P* values from a multivariate linear regression model (covariates in [Dataset S2](#)), assessing the interaction of HLA and sex on TCRBV usage. HLA/TCRBV combinations highlighted in red indicate  $P < 0.05$  after Bonferroni correction for multiple comparisons; yellow color suggests additional, potentially sex-specific HLA/TCRBV combinations, which did not pass Bonferroni correction. Left columns indicate the total number (N) as well as the male (M) and female (F) subject numbers for each HLA (raw data in [Dataset S3](#)), and green color indicates that this inclusion criterion was passed (one criterion was enough for inclusion).



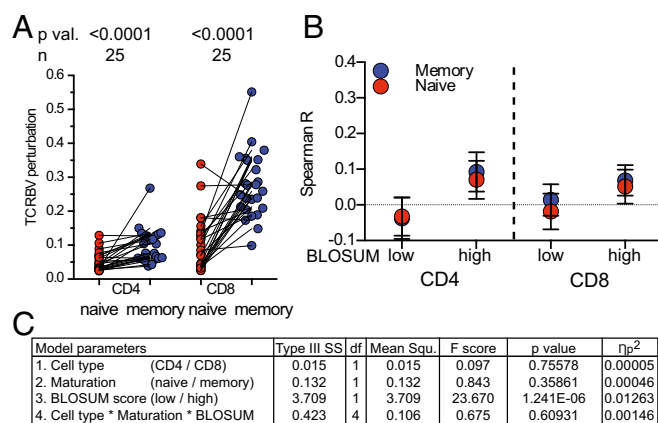
**Fig. 2.** CD8 T cells in men with autoimmune diseases are less dependent on biochemical CDR1/2 similarity for concerted selection/expansion. (A) TCRBV relationships Spearman's correlation coefficients (TCRBV rhos) of CD4 T cells of healthy donors (HD) and MS patients grouped according to BLOSUM-low and BLOSUM-high scores and divided by sex [male (red) versus female (blue)]. Shown are mean and 95% confidence interval. (B) Parameters of the linear model for CD4 T cells. Shown are the type III sum of squares (type III SS), degrees of freedom (df), mean square (mean squ.), F score, significance (P value), and partial eta squared ( $\eta^2$ ) as effect size. (C) TCRBV rhos of CD8 T cells of healthy donors (HD), and MS and RA patients grouped according to A. (D) Parameters of the linear model for CD8 T cells according to B. Two nested models are shown (BLOSUM-low and BLOSUM-high) within CD8 T cells. Pairwise comparison (superscript a) of the mean is shown for groups of interest (Bonferroni-corrected Student's *t* test). (E) TCRBV perturbation, measured as productive clonality. Shown are paired comparisons (Wilcoxon matched pairs signed rank test) between age-matched male and female individuals (CD4 and CD8 T cells; HD, MS, or RA).

assessment, BLOSUM values were divided at the median into 232 low- and 233 high-similarity scores. Subsequently the TCRBV rhos were used as the dependent variable in linear models for CD4 (Fig. 2A and B) and CD8 T cells (Fig. 2C and D) with the main effects "disease state (HD/MS/RA)," "sex (male/female)," and "BLOSUM score (high/low)." As hypothesized, TCRBV rhos of both T cell subsets were influenced by BLOSUM scores. However, only CD8 T cells were additionally influenced by disease state, as well as sex: (i) CD8 T cells from patients with both autoimmune diseases (MS or RA) differed from CD8 T cells from healthy donors. (ii) With regard to sex, men with autoimmune diseases showed significantly higher rhos in TCRBV combinations with low BLOSUM scores than women, which in the case of RA could also be observed in TCRBV combinations with high BLOSUM scores. To further explore the hypothesis that an MHC-TCR stimulus of the same affinity could lead to higher CD8 expansions of similar TCRBV chains in men with autoimmune diseases than in women, TCRBV chain perturbation was assessed (4, 38). In an age-matched, paired analysis, male subjects with autoimmune diseases (MS or RA) presented with higher "productive clonality" (a measurement of TCRBV perturbation), than female subjects in the CD8 T cell compartment (Fig. 2E).

**In Vivo Relationships Between TCRBV Transcripts Are Cause and Not Result of TCRBV Perturbations.** Twenty-five pairs of naive and memory CD4 and CD8 T cells from healthy donors (100 samples in total, raw data in Dataset S7) were analyzed to address the following questions: (i) Is the higher TCRBV perturbation in CD8 T cells in men cause or effect of the relationships seen in Fig. 2? (ii) Can a high degree of TCRBV perturbation, which would be expected in memory T cells, influence the relationships between TCRBV gene transcripts? As expected, memory T cells presented with higher TCRBV perturbation compared with naive T cells (Fig. 3A). However, this neither influenced the relationship between BLOSUM scores and TCRBV rhos in CD4 T cells nor CD8 T cells (Fig. 3B and C). Additionally, the relationships between TCRBV transcripts found in naive cells could also be found in memory cells (Fig. S3A and B, TCRBV rhos listed in Dataset S5).

## Discussion

The most surprising finding of our study was that the biological sex of an individual can influence the HLA association with T cell selection and expansion. Most autoimmune diseases more frequently affect women (25), whereas men are more susceptible to many infectious diseases, e.g., leishmaniasis or hepatitis A (39). HLA associations with cancer and the fact that men develop cancer in nonreproductive organs much more frequently than women has also been known for a long time (24). As cancer can also be viewed as a failure of the immune system to control neoplastic tissue, low expression in certain TCRBV families



**Fig. 3.** In vivo relationships between TCRBV chains are similar in naive and memory T cell subsets. (A) TCRBV perturbation quantified by the sequencing parameter productive clonality of naive and memory CD4 and CD8 T cell subsets from 25 healthy subjects. (B) TCRBV rhos of naive (red) and memory (blue) CD4/CD8 T cells grouped according to low and high BLOSUM scores. Shown is mean and 95% confidence interval (TCRBV rhos in Dataset S5 and raw data in Dataset S7). (C) Parameters of the linear model with TCRBV rhos as dependent and cell type, maturation, BLOSUM scores as main effects with the added interaction of those three main effects.



could be an indication of this defect as well. Apart from susceptibilities and disease incidences, general sex differences in parameters of the adaptive immune system have been found before. Women have higher amounts of circulating IgMs (40) and more CD4 T cells (41). These differences are thought to be mainly mediated by steroid hormones, with estrogens often promoting and androgens suppressing immune responses during infections, after vaccination, or in case of autoimmunity (reviewed in refs. 42–44). Recent studies also suggest an influence of glucocorticoid receptors in the suppression of autoimmunity in pregnant MS patients (32). Therefore, the observed sex differences in HLA-mediated shaping of the TCR repertoire could in part be caused or facilitated by the steroid hormone milieu, influencing lymphocyte activation during immune responses (32, 45), but also their maturation during thymic selection (46). Our data support these studies by providing evidence that the observed relationships between TCRBV transcripts play a similar role in thymic selection (reflected in naïve cells), as well as during expansion in the periphery due to antigenic stimuli (reflected in memory cells).

It has previously been shown that the HLA background shapes the TCR repertoire by favoring or neglecting certain T cell subpopulations represented by TCRBV usage (19). Certain HLAs are better at presenting auto- or foreign antigens to TCRBV-specific subgroups of T cells. Over time this can lead to an individually diverse “immunological fingerprint” shaped by genetics and exposure to antigens. This better or worse HLA–TCR interaction will have consequences for infectious diseases (susceptibility, progression, adverse effects), and also for anti-tumor responses, autoimmunity, or adverse medical reactions. Unexpectedly, we found differences between men and women with respect to the influence of HLA genes on TCRBV transcripts. In light of these findings, it makes sense that the TCR repertoire at birth is similar between cohorts with vastly different HLA backgrounds (47), but is then shaped in the course of life due to exposure to and HLA presentation of different antigens as shown in twin studies (17, 48). Consistent with this idea, it has recently been shown that T cell expansions show high degrees of homology on an amino acid level, but not on a DNA sequence level in individuals with similar HLA backgrounds (49). One possible explanation might have been a different expression level of the HLA molecules on the cells of men and women, but a recently published large study showed that at least the soluble form of the  $\beta$ -2 microglobulin in serum of men and women is comparable (50), which would fit the above-mentioned hypothesis that the observed differences could be a hormone-mediated downstream effect taking place after the MHC–TCR binding.

The second finding that CD8 T cell subpopulations with low biochemical similarities in their TCRBV chains correlate more strongly in patients with autoimmune diseases (MS and RA) in general, but particularly in men, is consonant with the overall higher CD8 T cell TCRBV perturbation in men compared with women. We hypothesize that a defined antigen stimulus would thereby lead to more clonal expansions in CD8 T cells in men, as the presentation by a given HLA molecule could also trigger T cell expansions with low binding affinity between TCR and MHC. Additionally, it is known that CD4 and CD8 in their function as coreceptors bind to the MHC molecules with different binding affinities (reviewed in ref. 51): CD4 has a weak binding affinity to MHC class II (52), whereas CD8 has a high binding affinity to MHC class I (53). These studies have also suggested that MHC class II ligands modulate TCR signaling directly via the TCR, whereas MHC class I ligands do not influence TCR signaling via the TCR, but via CD8 itself (54), and CD8 T cells are less dependent on continuous antigen presentation (55). Therefore, our data showing that low CDR similarity can still lead to parallel expansions in unrelated TCRBVs fit the hypothesis that CD8 T cells are less dependent on a strong

MHC–TCR binding. Furthermore, antigen-triggered expansions could also occur in TCRs despite suboptimal binding to MHC as long as the CD8 binding affinity is strong enough (56). Many hypotheses concerning autoimmunity involve an immune reaction against self-antigens via antigen spreading, molecular mimicry, or because low-avidity clones become activated, even though the threshold for activation should not have been reached, especially in cases of dual TCR expression by CD8 T cells (ref. 57 and reviewed in ref. 58). Our finding that CD8 T cells of MS patients in general, but particularly in men with MS expand in concert in regions of low CDR1/2 similarity is consistent with the finding that men present with higher disease burden, if they suffer from MS (59). We found similar correlations in CD8 T cells of patients with RA, but in contrast to MS, the clinical picture in RA is not as clear concerning sex differences and disease severity, potentially due to the fact that the measurement of disease activity in RA involves secondary outcome parameters with inherent sex bias (60). Our study provides evidence that biological sex modifies associations between HLA molecules and T cell TCRBV subpopulations and might, therefore, have implications regarding prior studies of HLA associations. These studies could possibly be revisited and male and female participants analyzed separately to ensure that some significant results are not lost due to statistical adjustments.

## Materials and Methods

**Study Cohorts.** This study included samples from five cohorts: (cohort 1) healthy subjects (20); (cohort 2) subjects suffering from epilepsy, multiple sclerosis, or a common cold, as well as corresponding healthy subjects (ref. 61 and raw data in [Datasets S2](#) and [S6](#)); (cohort 3) subjects suffering from rheumatoid arthritis and corresponding healthy subjects (62); (cohort 4) healthy subjects (63); and (cohort 5) healthy subjects (64). Details of these cohorts and their contribution to the analyses are listed in [Dataset S1](#).

**Study Approval.** Studies on human samples were approved by the local ethics committee (University of Muenster: Ethik-Kommission der Ärztekammer Westfalen-Lippe und der Medizinischen Fakultät der Westfälischen Wilhelms-Universität, registration number: 2010–245-f-5). Written informed consent was given by all participants in the study and all experiments were performed according to the Declaration of Helsinki.

**Sequencing.** TCR immunosequencing was done according to the immunoSEQ assay (Adaptive Biotechnologies) as previously described (38, 61, 65); for details please see [SI Materials and Methods](#).

**Statistics.** In Fig. 1, the interaction of sex and HLA on TCRBV usage was analyzed with multivariate linear regression models with adjustment for covariates influencing TCRBV usage. Eight covariates (see below and [Dataset S2](#)) were used for all comparisons. The linear regression model included 31 TCRBV transcripts as dependents, the interaction of sex and one HLA as fixed factors, together with covariates (TCRBV = HLA + sex + HLA  $\times$  sex + covariates; one model per HLA, resulting in 31 models with 589 HLA  $\times$  sex interactions). Significance was corrected for multiple comparisons according to the Bonferroni method (66) using R (function `p.adjust`). Significances not passing Bonferroni correction, but exhibiting an uncorrected  $P$  value  $\leq 0.05$ , are considered exploratory. Covariates are as follows: fraction productive, the fraction of in-frame templates; productive rearrangements, count of unique and functional rearrangements; productive clonality, normalized TCRBV repertoire diversity measure using log (base 2) of the total number of unique productive rearrangements to normalize the entropy and subtracting the result from the value 1; max productive frequency, the highest frequency of a rearrangement; productive entropy, diversity measure for the TCRBV repertoire according to Shannon's entropy (67), calculated over all productive rearrangements; cell type, isolated cell subset (PBMC, CD4, or CD8 T cells); sequencing source material (cgDNA), genomic DNA (gDNA), or cDNA (cDNA); and batch, sequencing batch as assessed by time of sequencing (24 batches). Covariates “age” (of donor at time point of sampling) and “ethnicity” (self-reported ethnicity of a donor, Caucasian, or non-Caucasian) were included in a sensitivity analysis ([Fig. S1E](#)), as these variables were not available for all donors. The sensitivity assessment also included one analysis performed only with “healthy” individuals and one only with individuals, where the HLA background was determined, not “inferred.” HLA-A/B background was fully available for cohorts 1 and 3 and

was inferred for part of samples of cohort 2, based on HLA-A/B-specific CDR3 sequences according to Emerson (20). To determine possible hidden factors in addition to known covariates, we used the PEER package (33, 34). From 25 hidden factors provided by the algorithm, 9 were included in the sensitivity analysis, as these were in the  $\alpha_c$  range of the known covariates ( $\alpha_c < 100$ ) (Fig. S1D). The SPSS syntax used for Fig. 1 is given in Dataset S2.

In Fig. 2, all combinations of 31 TCRBVs resulted in 465 scores ranging from  $-9$  to  $26$ , each of them the sum of the CDR1 and  $-2$  binding sites' BLOSUM62 scores (Fig. S2). BLOSUM62 scores were calculated using the function "pairwiseAlignment" from the Biostrings package within Bioconductor (68, 69) and are presented as a heatmap clustered according to Ward (36). Optimal number of clusters was determined with gap statistic (37) by using the functions "clusGap" and "fviz\_gap\_stat" from the factoextra package. The 465 scores were divided at the median value of 3 into two groups of low (BLOSUM-low, score  $-9$ – $2$ ,  $n = 232$ ) and high (BLOSUM-high, score  $3$ – $26$ ,  $n = 233$ ) biochemical similarity. Multivariate linear models were applied to CD4 and CD8 datasets with TCRBV rhos as dependent and the main effects "sex," "disease state," and "BLOSUM score," as well as their interaction. Subgroup analyses (nested models) were applied to BLOSUM-low and BLOSUM-high with the main effects of sex and disease state, as well as their interaction.

- Kappler J, et al. (1983) The major histocompatibility complex-restricted antigen receptor on T cells in mouse and man: Identification of constant and variable peptides. *Cell* 35:295–302.
- Arstila TP, et al. (1999) A direct estimate of the human alphabeta T cell receptor diversity. *Science* 286:958–961.
- Lythe G, Callard RE, Hoare RL, Molina-París C (2016) How many TCR clonotypes does a body maintain? *J Theor Biol* 389:214–224.
- Robins HS, et al. (2010) Overlap and effective size of the human CD8+ T cell receptor repertoire. *Sci Transl Med* 2:47ra64.
- Qi Q, et al. (2014) Diversity and clonal selection in the human T-cell repertoire. *Proc Natl Acad Sci USA* 111:13139–13144.
- Li HM, et al. (2016) TCR $\beta$  repertoire of CD4+ and CD8+ T cells is distinct in richness, distribution, and CDR3 amino acid composition. *J Leukoc Biol* 99:505–513.
- Haynes BF, Heinly CS (1995) Early human T cell development: Analysis of the human thymus at the time of initial entry of hematopoietic stem cells into the fetal thymic microenvironment. *J Exp Med* 181:1445–1458.
- Batliwalla F, Monteiro J, Serrano D, Gregersen PK (1996) Oligoclonality of CD8+ T cells in health and disease: Aging, infection, or immune regulation? *Hum Immunol* 48:68–76.
- Britanova OV, et al. (2014) Age-related decrease in TCR repertoire diversity measured with deep and normalized sequence profiling. *J Immunol* 192:2689–2698.
- Dai S, et al. (2008) Crossreactive T cells spotlight the germline rules for alphabeta T cell-receptor interactions with MHC molecules. *Immunity* 28:324–334.
- Moss PAH, Bell JI (1995) Sequence analysis of the human alpha beta T-cell receptor CDR3 region. *Immunogenetics* 42:10–18.
- Yin L, et al. (2011) A single T cell receptor bound to major histocompatibility complex class I and class II glycoproteins reveals switchable TCR conformers. *Immunity* 35:23–33.
- Marrack P, Scott-Browne JP, Dai S, Gapin L, Kappler JW (2008) Evolutionarily conserved amino acids that control TCR-MHC interaction. *Annu Rev Immunol* 26:171–203.
- Lefranc MP, et al. (1998) IMGT, the International ImMunoGeneTics database. *Nucleic Acids Res* 26:297–303.
- Dean J, et al. (2015) Annotation of pseudogenic gene segments by massively parallel sequencing of rearranged lymphocyte receptor loci. *Genome Med* 7:123.
- van Leeuwen A, Termijtelen A, Shaw S, van Rood JJ (1982) Recognition of a polymorphic monocyte antigen in HLA. *Nature* 298:565–567.
- Hoover ML, et al. (1986) Insulin-dependent diabetes mellitus is associated with polymorphic forms of the T-cell receptor beta chain gene. *Hum Immunol* 17:176 (abstr).
- Akolkar PN, Gulwani-Akolkar B, McKinley M, Fisher SE, Silver J (1995) Comparisons of T cell receptor (TCR) V  $\beta$  repertoires of lamina propria and peripheral blood lymphocytes with respect to frequency and oligoclonality. *Clin Immunol Immunopathol* 76:155–163.
- Sharon E, et al. (2016) Genetic variation in MHC proteins is associated with T cell receptor expression biases. *Nat Genet* 48:995–1002.
- Emerson RO, et al. (2017) Immunosequencing identifies signatures of cytomegalovirus exposure history and HLA-mediated effects on the T cell repertoire. *Nat Genet* 49:659–665.
- García KC, et al. (1998) Structural basis of plasticity in T cell receptor recognition of a self peptide-MHC antigen. *Science* 279:1166–1172.
- Ramagopal UA, et al. (2017) Structural basis for cancer immunotherapy by the first-in-class checkpoint inhibitor ipilimumab. *Proc Natl Acad Sci USA* 114:E4223–E4232.
- Le DT, et al. (2017) Mismatch repair deficiency predicts response of solid tumors to PD-1 blockade. *Science* 357:409–413.
- Clocchiatti A, Cora E, Zhang Y, Dotto GP (2016) Sexual dimorphism in cancer. *Nat Rev Cancer* 16:330–339.
- Whitacre CC, Reingold SC, O'Looney PA (1999) A gender gap in autoimmunity. *Science* 283:1277–1278.
- Labonté B, et al. (2017) Sex-specific transcriptional signatures in human depression. *Nat Med* 23:1102–1111.
- Hong S, et al. (2016) Complement and microglia mediate early synapse loss in Alzheimer mouse models. *Science* 352:712–716.
- Mosconi L, et al. (2017) Sex differences in Alzheimer risk: Brain imaging of endocrine vs chronological aging. *Neurology* 89:1382–1390.
- Gough SC, Simmonds MJ (2007) The HLA region and autoimmune disease: Associations and mechanisms of action. *Curr Genomics* 8:453–465.
- Blackwell JM, Jamieson SE, Burgner D (2009) HLA and infectious diseases. *Clin Microbiol Rev* 22:370–385.
- Kirsch I, Vignali M, Robins H (2015) T-cell receptor profiling in cancer. *Mol Oncol* 9:2063–2070.
- Engler JB, et al. (2017) Glucocorticoid receptor in T cells mediates protection from autoimmunity in pregnancy. *Proc Natl Acad Sci USA* 114:E181–E190.
- Stegle O, Parts L, Piipari M, Winn J, Durbin R (2012) Using probabilistic estimation of expression residuals (PEER) to obtain increased power and interpretability of gene expression analyses. *Nat Protoc* 7:500–507.
- Stegle O, Parts L, Durbin R, Winn J (2010) A Bayesian framework to account for complex non-genetic factors in gene expression levels greatly increases power in eQTL studies. *PLoS Comput Biol* 6:e1000770.
- Henikoff S, Henikoff JG (1992) Amino acid substitution matrices from protein blocks. *Proc Natl Acad Sci USA* 89:10915–10919.
- Ward JH (1963) Hierarchical grouping to optimize an objective function. *J Am Stat Assoc* 58:236–244.
- Tibshirani R, Walther G, Hastie T (2001) Estimating the number of clusters in a data set via the gap statistic. *J R Stat Soc Series B Stat Methodol* 63:411–423.
- Robins HS, et al. (2009) Comprehensive assessment of T-cell receptor beta-chain diversity in alphabeta T cells. *Blood* 114:4099–4107.
- Guerra-Silveira F, Abad-Franch F (2013) Sex bias in infectious disease epidemiology: Patterns and processes. *PLoS One* 8:e62390.
- Grundbacher FJ (1972) Human X chromosome carries quantitative genes for immunoglobulin M. *Science* 176:311–312.
- Amadori A, et al. (1995) Genetic control of the CD4/CD8 T-cell ratio in humans. *Nat Med* 1:1279–1283.
- Giefing-Kröll C, Berger P, Lepperdinger G, Grubeck-Loebenstien B (2015) How sex and age affect immune responses, susceptibility to infections, and response to vaccination. *Aging Cell* 14:309–321.
- Fish EN (2008) The X-files in immunity: Sex-based differences predispose immune responses. *Nat Rev Immunol* 8:737–744.
- Klein SL (2012) Sex influences immune responses to viruses, and efficacy of prophylaxis and treatments for viral diseases. *BioEssays* 34:1050–1059.
- Grimaldi CM, Cleary J, Dagtas AS, Moussai D, Diamond B (2002) Estrogen alters thresholds for B cell apoptosis and activation. *J Clin Invest* 109:1625–1633.
- Zhu M-L, et al. (2016) Sex bias in CNS autoimmune disease mediated by androgen control of autoimmune regulator. *Nat Commun* 7:11350.
- Ramakrishnan NS, Grunewald J, Janson CH, Wigzell H (1992) Nearly identical T-cell receptor V-gene usage at birth in two cohorts of distinctly different ethnic origin: Influence of environment in the final maturation in the adult. *Scand J Immunol* 36:71–78.
- Zyagin IV, et al. (2014) Distinctive properties of identical twins' TCR repertoires revealed by high-throughput sequencing. *Proc Natl Acad Sci USA* 111:5980–5985.
- Ruggiero E, et al. (2015) High-resolution analysis of the human T-cell receptor repertoire. *Nat Commun* 6:8081.
- Juraschek SP, et al. (2013) Comparison of serum concentrations of  $\beta$ -trace protein,  $\beta$ 2-microglobulin, cystatin C, and creatinine in the US population. *Clin J Am Soc Nephrol* 8:584–592.
- Stone JD, Chervin AS, Kranz DM (2009) T-cell receptor binding affinities and kinetics: Impact on T-cell activity and specificity. *Immunology* 126:165–176.
- Krummel MF, Sjaastad MD, Wülfing C, Davis MM (2000) Differential clustering of CD4 and CD3zeta during T cell recognition. *Science* 289:1349–1352.

