MHC-II alleles shape the CDR3 repertoires of conventional and regulatory naïve CD4$^+$ T cells

Nadezhda N. Logunova[a,1], Valeriia V. Kriukova[b,c,1], Pavel V. Shelyakin[b], Evgeny S. Egorov[b,d], Alina Pereverzeva[b], Nina G. Bozhanova[e,f], Mikhail Shugay[b,c,d], Dmitrii S. Shcherbinin[d], Mikhail V. Pogorelyy[b,d], Ekaterina M. Merzlyak[b,d], Vasiliy N. Zubov[b], Jens Meiler[e,f,g], Dmitriy M. Chudakov[b,c,d], Alexander S. Apt[a,2,3], and Olga V. Britanova[b,2,3]

[a]Laboratory for Immunogenetics, Central Institute for Tuberculosis, Moscow, 107564, Russia; [b]Department of Genomics of Adaptive Immunity, Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, Moscow, 117997, Russia; [c]Center of Life Sciences, Skoltech, Moscow, 143026, Russia; [d]Institute of Translational Medicine, Pirogov Russian National Research Medical University, Moscow, 117997, Russia; [e]Center for Structural Biology, Vanderbilt University, Nashville, TN 37235; [f]Department of Chemistry, Vanderbilt University, Nashville, TN 37235; and [g]Institute for Drug Discovery, Leipzig University, Leipzig, SAC 04103, Germany

Edited by Philippa Marrack, National Jewish Health, Denver, CO, and approved April 27, 2020 (received for review February 21, 2020)

T cell maturation and activation depend upon T cell receptor (TCR) interactions with a wide variety of antigenic peptides displayed in a given major histocompatibility complex (MHC) context. Complementarity-determining region 3 (CDR3) is the most variable part of the TCRα and -β chains, which govern interactions with peptide–MHC complexes. However, it remains unclear how the CDR3 landscape is shaped by individual MHC context during thymic selection of naïve T cells. We established two mouse strains carrying distinct allelic variants of *H2-A* and analyzed thymic and peripheral production and TCR repertoires of naïve conventional CD4$^+$ T (T_{conv}) and naïve regulatory CD4$^+$ T (T_{reg}) cells. Compared with tuberculosis-resistant C57BL/6 (H2-A^b) mice, the tuberculosis-susceptible H2-A^j mice had fewer CD4$^+$ T cells of both subsets in the thymus. In the periphery, this deficiency was only apparent for T_{conv} and was compensated for by peripheral reconstitution for T_{reg}. We show that H2-A^j favors selection of a narrower and more convergent repertoire with more hydrophobic and strongly interacting amino acid residues in the middle of CDR3α and CDR3β, suggesting more stringent selection against a narrower peptide–MHC-II context. H2-A^j and H2-A^b mice have prominent reciprocal differences in CDR3α and CDR3β features, probably reflecting distinct modes of TCR fitting to MHC-II variants. These data reveal the mechanics and extent of how MHC-II shapes the naïve CD4$^+$ T cell CDR3 landscape, which essentially defines adaptive response to infections and self-antigens.

TCR repertoire landscape | MHC-II | naïve CD4$^+$ T cells | regulatory T cells

The interaction of peptide–major histocompatibility complex (p-MHC) with T cell receptors (TCRs) plays a central role in positive and negative selection of T lymphocytes in the thymus as well as subsequent homeostasis of naïve, primed, and effector-memory T cells in the periphery (1). Even subtle shifts in p-MHC–TCR interactions may profoundly affect T cell responses (2–4) and in extreme cases, can result in immunological disorders (5–7). The theoretical diversity of TCRα/β variants initially produced by recombination in the thymus exceeds 10^{15} for mice (8) and 10^{19} for humans [per our current estimate (9)]. However, not all TCRs effectively interact with p-MHC; only ∼5% of T cells successfully pass through positive selection in the thymus, and TCR repertoires are further narrowed by negative selection (reviewed in ref. 10). Selection continues in the periphery, where recent thymic emigrants acquire the functional properties of mature naïve T cells—which are only capable of providing an antigen-specific response—after exhaustive screening against self p-MHCs (11, 12), enforcing MHC restriction. Subsequently, tonic TCR signaling—induced by interaction with self p-MHC—supports long-term survival of mature naïve T cells (13). Thus, the individual repertoire of naïve TCRs is strongly shaped by self p-MHC complexes, which determine the allowed range of affinities and angles of interaction (4, 14, 15). The resulting individual diversity of a functional TCRα/β repertoire gains about 2×10^6 TCRα/β variants per 2×10^7 cells in a mouse spleen (16). For a human, individual naïve TCRα/β diversity may reach 10^8 variants (17).

Binding of TCRα and -β chains to the p-MHC-II complex is largely determined by their complementarity-determining regions (CDRs). CDR1 and CDR2, encoded by a set of germline T cell receptor variable α (*TRAV*) and T cell receptor variable β (*TRBV*) genes, primarily make contact with α-helical domains of MHC-II, while CDR3, which originates from the hypervariable combination of variable, diversity and joining (V-(D)-J) segments, predominantly contacts antigenic peptides docked within the MHC-II grove (18). It is firmly established that the usage of TCRα and TCRβ V genes depends upon *MHC* allelic variations

Significance

The landscape of naïve T cell receptors (TCRs) essentially determines the ability of the host to respond to vaccination, infection, and cancer and to control autoimmunity and inflammation. The diversity of naïve T cells is generated by semirandom recombination, followed by tight shaping during thymic selection against individual pools of self-peptide major histocompatibility complexes (MHCs). We characterize TCR repertoires of naïve conventional and regulatory CD4$^+$ T cells produced in two different MHC-II contexts. We reveal profound differences in the diversity, convergence, and physicochemical properties of their antigen-interacting regions that indicate the distinct mode of TCR interaction with peptide–MHC-II. Such differences are likely to influence individual susceptibility to infections and autoimmunity and have implications for the future development of TCR-based diagnostics and therapies.

Author contributions: N.N.L., D.M.C., A.S.A., and O.V.B. designed research; N.N.L., V.V.K., E.S.E., A.P., and E.M.M. performed research; N.G.B., M.S., D.S.S., M.V.P., and J.M. contributed new reagents/analytic tools; V.V.K., P.V.S., and V.N.Z. analyzed data; and N.N.L., V.V.K., N.G.B., D.S.S., D.M.C., A.S.A., and O.V.B. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

Data deposition: The sequence data reported in this paper have been deposited in the National Center for Biotechnology Information (NCBI) Sequence Read Archive (accession no. PRJNA600142).

[1]N.N.L. and V.V.K. contributed equally to this work.

[2]A.S.A. and O.V.B. contributed equally to this work.

[3]To whom correspondence may be addressed. Email: alexapt0151@gmail.com or olbritan@gmail.com.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2003170117/-/DCSupplemental.

First published June 1, 2020.

(19–21). However, there is still a substantial gap in our understanding of how allelic variability in the MHC Class II locus shapes the intrinsic properties of naïve TCRα and TCRβ CDR3 repertoires. It is also unclear whether these effects differ substantially for conventional $CD4^+$ T (T_{conv}) and regulatory $CD4^+$ T (T_{reg}) cells, for which the thymic and peripheral selection process is thought to differ profoundly (22, 23). Nonsynonymous amino acid substitutions within the binding groove of an MHC-II molecule would be predicted to profoundly affect CDR3 repertoires. Such substitutions may alter the surface of an MHC-II molecule involved in interaction with the TCR, the conformation of antigenic peptides, and the whole repertoire of presented peptides, thus affecting TCR binding and T cell activation (24, 25).

Previously, we demonstrated that the rare MHC-II allelic variant *H2-A*^{j} is associated with increased susceptibility to tuberculosis infection (26). Both chains of H2-A^j have a number of potentially important substitutions located in α-helices that form the peptide binding groove. To investigate this association, we established a panel of *H2*-congenic recombinant mouse strains bearing different segments of the *H2* region from tuberculosis-susceptible I/St (*H2*^{j}) mice on the genetic background of the tuberculosis-resistant C57BL/6 (B6, *H2*^{b}) strain. One of the strains from this B6.I-*H2* panel, B6.I-9.3, differs from the B6 parent by the *H2-A* allele of the classical *H2* gene complex, which bears genetic material of I/St origin within the 30.90- to 34.34-Mb interval of chromosome 17. Both B6 and B6.I-9.3 are H2-E–negative strains; thus, the H2-A molecule is the only classical MHC-II product potentially influencing $CD4^+$ T cell repertoires in B6 and B6.I-9.3 mice. B6.I-9.3 and B6 mice also differ by an unusual allelic polymorphism in the nonclassical MHC-II *H2-Oa* and *H2-Ob* genes of the *H2*^{j} haplotype (26, 27), which are involved in peptide loading onto MHC-II (28, 29). These features make the B6 and B6.I-9.3 pair an attractive tool for studying the influence of MHC-II allelic variations on T cell differentiation and maturation in general and TCR repertoires in particular.

Here, we report that mice of the B6.I-9.3 (*H2-A*^{j}) strain differ from their congenic B6 (*H2-A*^{b}) counterparts by a profound deficiency of both T_{conv} and T_{reg} cells in the thymus. The T_{reg} population in H2-A^j animals was restored in the periphery, but T_{conv} cells remained deficient. Using deep TCR repertoire profiling, we evaluated the impact of the MHC-II polymorphism on the formation of TCRα and TCRβ CDR3 repertoires in naïve conventional $CD4^+$ T (nT_{conv}) and naïve regulatory $CD4^+$ T (nT_{reg}) cells. We found that the *H2-A*^{j} MHC context produces a restricted TCR repertoire with a higher level of sequence convergence and increased MHC-restricted publicity. A relatively higher number of homologous CDR3 variants formed a denser network of larger TCR clusters. Furthermore, we show that this MHC-II locus polymorphism dramatically altered key characteristics of both CDR3α and CDR3β repertoires. The potential strength of both CDR3α and CDR3β interactions with p-MHC-II complexes increases in the *H2-A*^{j} context—presumably compensating for the hampered $CD4^+$ T cell selection—and this feature was much more prominent for the nT_{reg} cells. Taken together, our results demonstrate that the structure and composition of the TCR CDR3α and CDR3β repertoires may be prominently shaped by allelic polymorphisms in the MHC-II gene complex, with notable differences between T_{conv} and T_{reg} cells that should essentially influence individual susceptibility to infectious and autoimmune diseases and cancer.

Results

Selection of $CD4^+$ T Lymphocytes Is Altered in *H2-A*^{j} Context. The I/St-derived *H2* segment present in B6.I-9.3 mice contains not just the two genes encoding the H2-A dimer but also, a few additional proximal genes—including the nonclassical MHC-II genes *H2-Ma*, *H2-Mb1*, *H2-Mb2*, *H2-Oa*, and *H2-Ob* (26). We therefore added the parental I/St strain to our analysis as a control that allows separation of bona fide *H2-A*^{j}–controlled phenotypes from possible effects of allelic variants in other genes located in the proximal part of the relevant chromosomal region. Furthermore, unlike B6 and B6.I-9.3 congenic mice, I/St mice express the second MHC-II molecule, H2-E, which also participates in $CD4^+$ cell selection.

Flow cytometry analysis of single-positive (SP) $CD4^+$ and $CD8^+$ cell populations recovered from the thymi of B6.I-9.3 (*H2-A*^{j}) mice demonstrated a profound reduction in the proportion of $CD4^+$ cells compared with parental mouse strains B6 (*H2-A*^{b}) and I/St (*H2-A*^{j} and *H2-E*^{j}), producing an unusually inverted CD4/CD8 ratio (Fig. 1*A*). The total cellularity of B6 and B6.I-9.3 thymi was similar ($207 \pm 5.4 \times 10^6$ and $219 \pm 33 \times 10^6$, respectively; $P > 0.05$) but was significantly higher compared with parental I/St mice ($144 \pm 12 \times 10^6$; $P < 0.005$). Both positive and negative selection may contribute to an unusual CD4/CD8 ratio in the thymi of B6.I-9.3 mice. We first assessed possible alterations in positive selection by measuring the expression of activation markers CD69 and CD5 in double-positive (DP) thymocytes, both of which are known to be up-regulated in response to p-MHC–TCR interactions during positive selection. As shown in Fig. 1*B*, the proportion of $CD5^+CD69^+$ DP thymocytes was lower in B6.I-9.3 mice compared with both parental strains. The recovery of the $CD4^+$ subset in B6 × B6.I-9.3 F1 hybrids confirmed that impaired thymic selection of $CD4^+$ T cells was governed by the *H2-A*^{j} allele; codominant expression of the *H2-A*^{b} allele restored the $CD4^+$ subset and normalized the population of $CD5^+CD69^+$ DP cells in thymi (Fig. 1 *A* and *B*). Together, these results clearly indicate that positive thymic selection of $CD4^+$ T lymphocytes is hampered in the context of *H2-A*^{j} compared with selection on H2-A^b or H2-E^j MHC-II molecules.

In the periphery, $CD4^+$ deficiency was also observed in the spleens and lymph nodes of B6.I-9.3 mice, leading to a dramatic decrease in the CD4/CD8 ratio compared with the parental strains (Fig. 1 *C* and *D*). F1 hybrids showed a distribution closer to that of B6 mice (spleen $0.85 \pm 0.15/1.35 \pm 0.3$). Expression of H2-$A^j$ on the surface of B cells in B6.I-9.3 mice was not impaired (*SI Appendix*, Fig. S1) and therefore, could not explain this defect in $CD4^+$ T cell selection. Variability in the CD4/CD8 ratio in peripheral lymphoid tissues has been described for various mouse strains, and a predominance of $CD4^+$ over $CD8^+$ T cells is common for most inbred mice (*SI Appendix*, Table S1). Accordingly, both experimental inbred mouse strains and *H2*-congenic recombinant mice, which have different haplotypes and only a single MHC-II molecule (H2-A), have a CD4/CD8 ratio ≥1. B6.I-9.3 represents an example of an inverted CD4/CD8 ratio that is <1. The fact that a typical CD4/CD8 ratio >1 persists in parental I/St mice that also bear *H2*^{j} alleles argues against the influence of genes proximal to *H2-A* on this CD4/CD8 inversion and instead, suggests the compensatory contribution of an H2-E–selected $CD4^+$ pool to the total $CD4^+$ population, as has been previously reported in a genetic study (30). Thus, total homeostatic proliferation in the periphery was unable to compensate for impaired thymic selection of naïve $CD4^+$ T cells in B6.I-9.3 mice. Profound changes in CD4/CD8 ratios indicate intrinsic differences in TCR selection in the context of distinct *H2-A* allelic variants.

T_{reg} but Not T_{conv} Populations Are Reconstituted in the Spleens of B6.I-9.3 Mice. TCR–p-MHC-II interactions determine activation thresholds during selection of T_{conv} and T_{reg} cells in the thymus. Importantly, the allowed affinities and the process of selection itself differ for these two populations (22, 31, 32). To assess how these differences might influence the TCR repertoires for nT_{conv} and nT_{reg} cells developing in B6 and B6.I-9.3 mice, we generated the B6.I-9.3.Foxp3^{GFP} mouse strain (hereafter referred to as H2-A^j

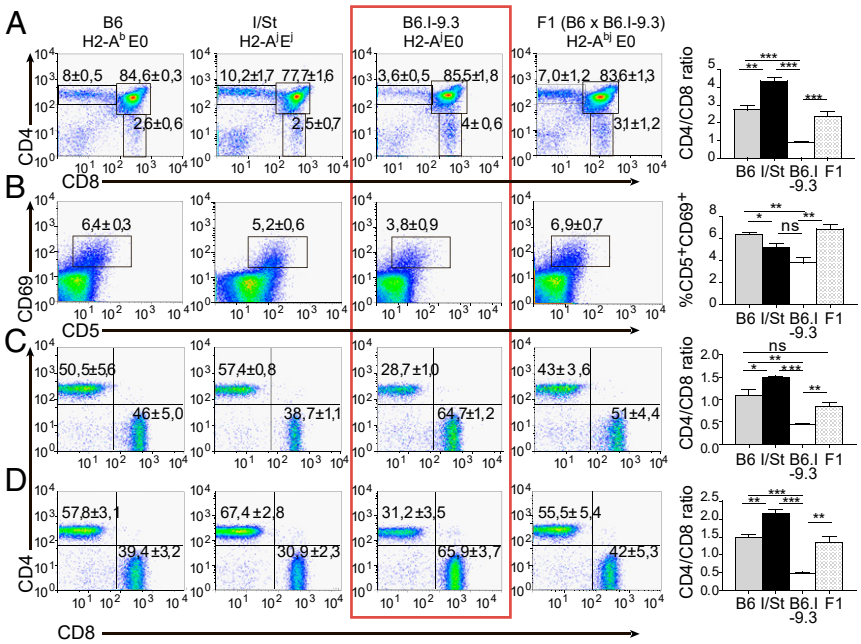

Fig. 1. B6.I-9.3 mice display an inverted CD4/CD8 ratio in the thymus and periphery. The percentage of $CD4^+$ and $CD8^+$ T cells in the (*A*) thymus, (*C*) spleen, and (*D*) peripheral lymph nodes as assessed by flow cytometry. Ratios are summarized as bar graphs in *Right*. (*B*) Thymocytes undergoing selection processes are defined as quadruple-positive $CD4^+CD8^+CD5^+CD69^+$ cells. (*C* and *D*) Results for gated $CD3^+$ T cells are shown. Representative data from one of three to five similar independent experiments are displayed, with total statistics for 6 to 12 mice per group provided in *Right* as mean ± SD (unpaired *t* test). ns, not significant; $*P < 0.05$; $**P < 0.01$; $***P < 0.001$.

mice) and compared it with the B6.Foxp3^{GFP} strain (hereafter referred to as H2-A^b mice) described previously (33). Both strains produce fluorescently labeled T_{reg} cell populations. Evaluation of CD4/CD8 ratios in lymphoid organs of H2-A^j and H2-A^b Foxp3^{GFP} reporter mice confirmed defective $CD4^+$ selection in the *H2-A^j* genetic context (Fig. 2 *A* and *B*).

We next assessed whether the generation and/or maintenance of T_{reg} and T_{conv} $CD4^+$ subsets differs in H2-A^b and H2-A^j mice by counting the content and proportion of $CD4^+$ T cells in the spleen and thymus. The proportion of $Foxp3^+$ T_{reg} cells among SP $CD4^+$ thymocytes in H2-A^j and H2-A^b mice appeared to be similar (Fig. 2*C*). However, due to reduced total $CD4^+$ content, the size of the T_{reg} population in thymi of H2-A^j mice was significantly smaller (Fig. 2*D*). At the same time, in the spleens of H2-A^j mice, the total T_{reg} (Fig. 2*E*) and nT_{reg} (Fig. 2*F*) populations recovered to an extent that they matched the counts

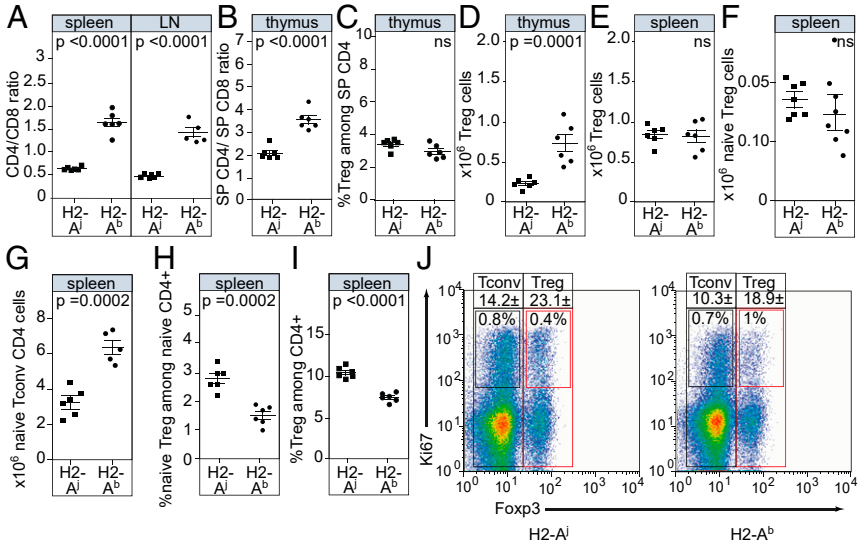

Fig. 2. $CD4^+$ T cell populations in lymphoid organs of H2-A^j and H2-A^b mice. CD4/CD8 T cell ratios in (*A*) spleens and lymph nodes (LN) and (*B*) thymi for H2-A^b ($n = 6$) and H2-A^j ($n = 6$) mice. (*C*) The percentage of nT_{reg} cells in the population of thymic SP $CD4^+$ T cells is equal in H2-A^j and H2-A^b mice. The total number of nT_{reg} cells in H2-A^b ($n = 6$) and H2-A^j ($n = 6$) mice differs for the SP $CD4^+$ thymic population (*D*) but not for the spleen (*E*). (*F*) The splenic content of nT_{reg} cells is equal in H2-A^j and H2-A^b mice, whereas (*G*) the size of the nT_{conv} cell population is significantly smaller in H2-A^j. The percentage of (*H*) nT_{reg} and (*I*) total T_{reg} cells is higher in H2-A^j compared with H2-A^b spleens. (*J*) Proliferation of splenic $CD4^+$ T cells from sorted T_{reg} cells gated for Foxp3 expression. All data are representative of at least two experiments ($n = 6$ females per group, unpaired *t* test, $P = 0.006$ for T_{conv} and $P = 0.003$ for T_{reg}). ns, not significant.

detected in H2-A^b mice, while the nT$_{conv}$ population remained significantly reduced (Fig. 2*G*). Thus, in the spleens of H2-A^j mice, the proportion of nT$_{reg}$ (Fig. 2*H*) and primed (Fig. 2*I*) T$_{reg}$ cells was increased compared with H2-A^b mice.

Since the reduction in the T$_{reg}$ population that we observed in the thymi of H2-A^j mice was restored in the spleen, we questioned whether this was due to increased proliferation of T$_{reg}$ cells in order to achieve homeostatic replenishment. We estimated the proliferation rate of splenic CD4$^+$ T cells using a Ki67 expression staining. This analysis revealed an increased proliferation rate for both T$_{conv}$ and T$_{reg}$ splenic subsets in H2-A^j compared with H2-A^b mice (Fig. 2*J*). However, T$_{reg}$ cells restored their peripheral niche more efficiently, which is consistent with other observation (34).

TCR Repertoires of H2-A^j Mice Are Distinct from H2-A^b and Characterized by Decreased Diversity and High Publicity and Convergence. The initial diversity of preimmune repertoires of naïve T cells is generated by the recombination machinery as well as thymic and peripheral selection (10). To explore the influence of MHC-II allelic variation on the TCR landscape, we analyzed repertoires of nT$_{conv}$ and nT$_{reg}$ cells isolated from splenic cell populations of homozygous H2-A^j and H2-A^b mice. We sorted 5,600 to 65,000 nT$_{reg}$ and 100,000 nT$_{conv}$ cells per mouse, with the controlled purity of cell fractions >98%. TCRα and -β complementary DNA (cDNA) libraries were obtained using rapid amplification of 5'-cDNA ends (5'-RACE) with unique molecular identifiers (UMIs) and TCR repertoires extracted using MIGEC (35) and MiXCR (36) software. Out of each sample, we extracted CDR3 repertoires for at least 2,000 nT$_{reg}$ TCRα, 8,000 nT$_{conv}$ TCRα, 5,000 nT$_{reg}$ TCRβ, and 14,000 nT$_{conv}$ TCRβ unique UMI-labeled cDNA molecules (*SI Appendix*, Table S2*A*).

For comparison of general repertoire similarity, we performed clustering analysis based on the number of amino acid CDR3 clonotypes shared between the 800 most frequent clonotypes of each pair of samples. This analysis clearly clustered the functional (in-frame, without stop codons) TCRα and TCRβ repertoires of nT$_{conv}$ CD4$^+$ subsets separately for H2-A^j and H2-A^b mice (Fig. 3*A*), suggesting that the distinct p-MHC-II context essentially shapes the structure of the TCR repertoires at the stage of thymic selection. This was not the case for nonfunctional TCR repertoires resulting from unsuccessful recombination events. Such TCRs have not been expressed and thus, preserve

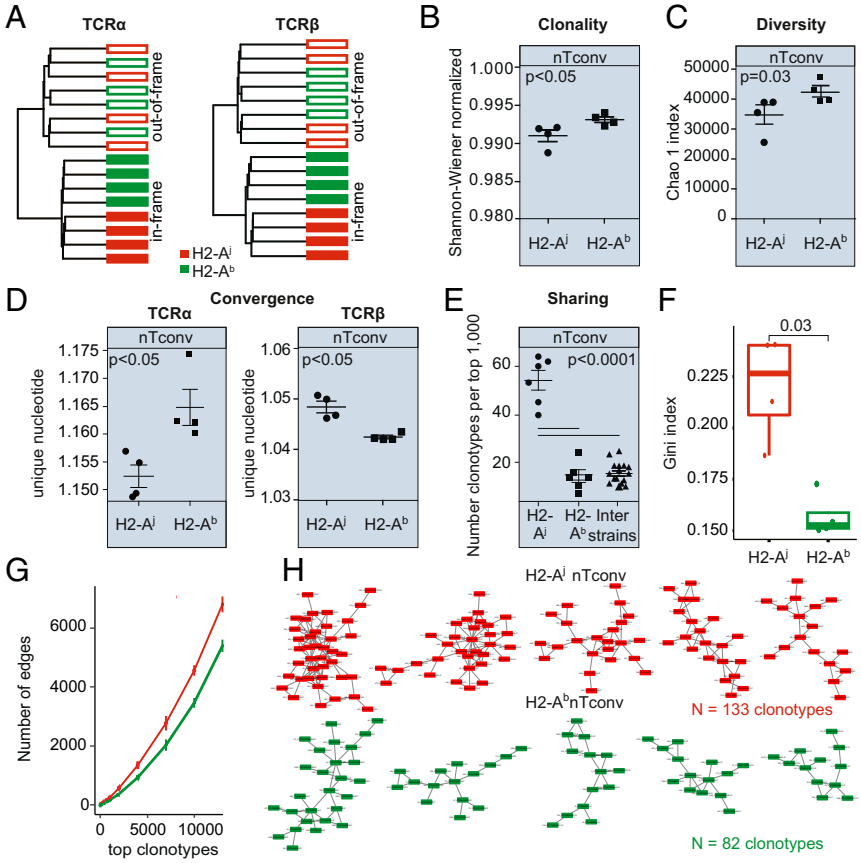

Fig. 3. TCR repertoire profiling. (*A*) Cluster analysis of the TCRα and TCRβ repertoires for nT$_{conv}$ isolated from spleens of H2-A^j (red; $n = 4$) and H2-A^b (green; $n = 4$) mice. Datasets were normalized based on the observed diversity of the smallest sample. Using a pairwise similarity metric (based on the number of identical amino acid sequences of clonotypes), we performed cluster analysis of CDR3β in-frame and nonfunctional variants for the top 800 clonotypes using VDJtools. (*B–E*) Properties of the TCRβ repertoire of nT$_{conv}$ from spleens. *P* values were computed using unpaired *t* test. (*B*) Assessment of clonality, where clonotype evenness (expressed as Shannon–Wiener index) was normalized by clonotype number. (*C*) Analysis of repertoire diversity using the Chao1 index for nT$_{conv}$ cells. We collected 14,000 UMIs per sample for CDR3β. (*D*) Convergence measurement as a ratio of unique nucleotide number encoding the same amino acid. (*E*) Sharing of TCR repertoires between normalized samples, based on the number of identical CDR3β amino acid sequences in the top 1,000 clonotypes. (*F*) The Gini index for the top 2,500 CDR3β amino acid clonotypes in the nT$_{conv}$ repertoires of H2-A^b and H2-A^j mice reflects inequality in clonotype size. *P* values were computed using Mann–Whitney *U* test. (*G*) Connectivity of repertoires in H2-A^j and H2-A^b mice. The average number of edges per node of homologous TCR clusters grows along with repertoire depth. (*H*) The top five clonotype clusters from the repertoires of a single H2-A^b and H2-A^j mouse. TCR networks were built from amino acid CDR3β sequences of the 2,500 most abundant clonotypes.

the repertoire features produced by the recombination machinery before thymic selection (37, 38). Clustering analysis did not separate the nonfunctional CDR3 repertoires of H2-A^j and H2-A^b mice, indicating that segregation of the functional repertoires was a result of distinct selection.

We assessed the normalized Shannon–Wiener and Chao1 diversity indices at the amino acid level (39) after normalization of samples by down-sampling to equal counts of unique UMI-labeled cDNA molecules. These two statistical metrics assess different aspects of the TCR repertoire; normalized Shannon–Weiner reflects the evenness of the clonotype frequency distribution, whereas Chao1 provides the lower boundary of total clonotype diversity as estimated based on the relative counts of low-frequency clonotypes. As could be expected, the drastic reduction in the count of naïve CD4$^+$ cells concomitant with elevated proliferation was reflected in the decreased diversity and increased clonality of naïve TCR repertoires. In comparison with H2-A^b mice, the amino acid TCRβ repertoires of nT$_{conv}$ cells in H2-A^j showed significantly lower Chao1 and higher Gini index, which are related to unequal distribution of clonotype size (Fig. 3 *B*, *C*, and *G*). We observed a similar tendency in the TCRβ repertoires of nT$_{reg}$, but the number of clonotypes that we obtained was too small for these data to be statistically representative (*SI Appendix*, Fig. S2 *A*–*C*).

Next, we assessed the relative convergence of TCRβ repertoires, as measured by the normalized abundance of CDR3 nucleotide sequences encoding the same amino acid sequence, which results in the appearance of identical frequently produced TCR clonotypes within and between animals (40). Remarkably, the CDR3β repertoire of nT$_{conv}$ cells in H2-A^j mice was characterized by a prominently higher degree of convergent recombination events (*t* test, $P < 0.005$) (Fig. 3*D*) in comparison with H2-A^b, while the CDR3α repertoire in H2-A^j contained less converged and rather diverse TCR variants (Fig. 3*D*). We then estimated the extent of TCR clonotype sharing (i.e., publicity) at the amino acid level within and between H2-A^j and H2-A^b mouse groups. The cohort of H2-A^j mice shared a significantly higher proportion of CDR3β clonotypes than the H2-A^b mice: on average, 54 vs. 14 shared amino acid clonotypes from the 1,000 most frequent clonotypes of each repertoire (Fig. 3*E*). Finally, network analysis of TCRβ homology clusters based on amino acid similarity showed denser network connectivity of TCRβ repertoires in H2-A^j mice (Fig. 3 *F*–*H*). Together, these observations clearly indicate that the *H2-A^j* genetic context favors selection of a narrower naïve TCRβ repertoire, characterized by high publicity, constrained diversity, and increased convergence.

Prominent and Distinct Impact of *H2-A^j* on CDR3α and CDR3β Characteristics. As described in ref. 41, we assessed average characteristics of the CDR3 repertoire, such as length, number of added nucleotides, and the physicochemical properties of amino acids located in the middle of CDR3 (*SI Appendix*, Fig. S3), which are usually in tight contact with p-MHC-II. Interestingly, the TCRβ repertoire of H2-A^j mice was skewed toward longer CDR3 loops in both nT$_{conv}$ and nT$_{reg}$ CD4$^+$ subsets (at average, weighted per-clonotype frequencies) (Fig. 4*A*). This increase in length was associated with a higher number of non-germline added nucleotides (*SI Appendix*, Fig. S3*A*). Remarkably, we observed an opposite tendency in the TCRα repertoire, with decreased CDR3 length (Fig. 4*A*). Additionally, nT$_{reg}$ cells in H2-A^j and H2-A^b mice possessed significantly shorter CDR3s in both α- and β-chains compared with nT$_{conv}$ cells. This is consistent with previous findings from other studies (42) as well as our current findings for human nT$_{reg}$ cells. The H2-A^j repertoire was also characterized by a lower average number of bulky amino acid residues based on the "volume" parameter of VDJtools (including W, R, K, Y, and F) in the middle of CDR3β but a higher number of such amino acids in CDR3α (Fig. 4*B*). We hypothesize that such reciprocal shifting of CDR3α and CDR3β characteristics might reflect distinct TCR fitting with p-MHC complex in H2-A^j mice.

Based on statistical mechanical computations of pairwise interactions between amino acids (43, 44), a set of generally strongly interacting amino acid residues has been selected. In essence, those include hydrophobic and aromatic residues that can mediate diverse types of interaction (45) and contribute greatly to TCR–p-MHC binding affinity (46). We estimated relative strength of CDR3 interaction with diverse (i.e., not necessarily bearing cognate antigen) p-MHC complexes based on the average number of such strongly interacting amino acid residues (F, I, L, M, V, W, and Y) within the central part of the CDR3 using the "strength" parameter of VDJtools. Previously, this parameter has been shown to distinguish TCR repertoires of T$_{reg}$ cells and CD4$^+$ T$_{conv}$ cells in mice (39, 47), in keeping with the higher TCR affinity allowed during thymic selection of T$_{reg}$ population (31, 32). It has been hypothesized that T$_{reg}$ cells that possess high-affinity TCRs can compete with conventional CD4$^+$ cells in the periphery for targets on antigen-presenting cells (23), preventing spontaneous proliferation of T$_{conv}$ cells. High CDR3 strength also distinguishes human nT$_{reg}$ and effector T$_{reg}$ cells vs. other helper subsets, CD8$^+$CXCR3$^+$ vs. CD8$^+$CXCR3$^-$ naive T cells (48). Furthermore, CDR3 strength and hydrophobicity are low in experienced memory B cells (49). To some extent, high CDR3 strength and hydrophobicity can be interpreted as T or B cell potential to broad, cross-reactive, and self-reactive interactions (44, 50).

In agreement with previous studies and our current data for human nT$_{reg}$ cells, we found that the average number of strongly interacting amino acids among the five residues in the middle of CDR3 was higher in nT$_{reg}$ compared with nT$_{conv}$ repertoires in both mouse strains, and for both CDR3α and CDR3β (Fig. 4*C*). An increased average number of strongly interacting amino acids in the nT$_{reg}$ CDR3s suggests the preferential selection of TCRs with high affinity to self-antigens for generation of thymic T$_{reg}$ population (22, 31, 32, 51). We also noted that the relative strength was significantly higher for H2-A^j mice compared with H2-A^b mice in the CDR3α and -β repertoires of both nT$_{reg}$ and nT$_{conv}$ cells. Furthermore, this effect was significantly more prominent in the nT$_{reg}$ subset. As a result, the difference between nT$_{reg}$ and nT$_{conv}$ cells was substantially higher in H2-A^j mice (Fig. 4*D*), which might potentially result in generally increased capacity of T$_{reg}$ cells to suppress activated T cells in H2-A^j mice.

Using cluster connectivity analysis, we searched for the specific characteristics of CDR3β amino acid sequence variants enriched in the repertoire of nT$_{reg}$ cells selected on different MHC-II molecules. We separately pooled all clonotypes for H2-A^b and H2-A^j mice and searched for similar amino acid TCR clonotypes within mouse strains. These similar TCRs may be represented as a network of CDRβ sequences having one amino acid mismatch between nodes. Considering only sequences that have more neighbors in one strain rather than in another, we found clonotypes that might carry an imprint of selection on a certain MHC-II allele. Notably, we detected distinct dominant CDR3β motifs (Fig. 4*E*). A CDR3β motif identical to that found in H2-A^j context was previously reported among public TCR clonotypes responding to respiratory syncytial virus infection (52). It is also interesting that the same CDR3β motif was previously found among cross-reactive TCRs recognizing both MHC-I and -II (53).

TCR repertoire analysis of all samples using high-dimensional features enabled us to clearly separate CDR3α and CDR3β repertoires by both functionality and genetic background (Fig. 4*F* and *SI Appendix*, Fig. S4). This demonstrates the prominent differences in the TCR CDR3 repertoire landscapes of both nT$_{reg}$ and nT$_{conv}$ subsets between H2-A^j and H2-A^b mice.

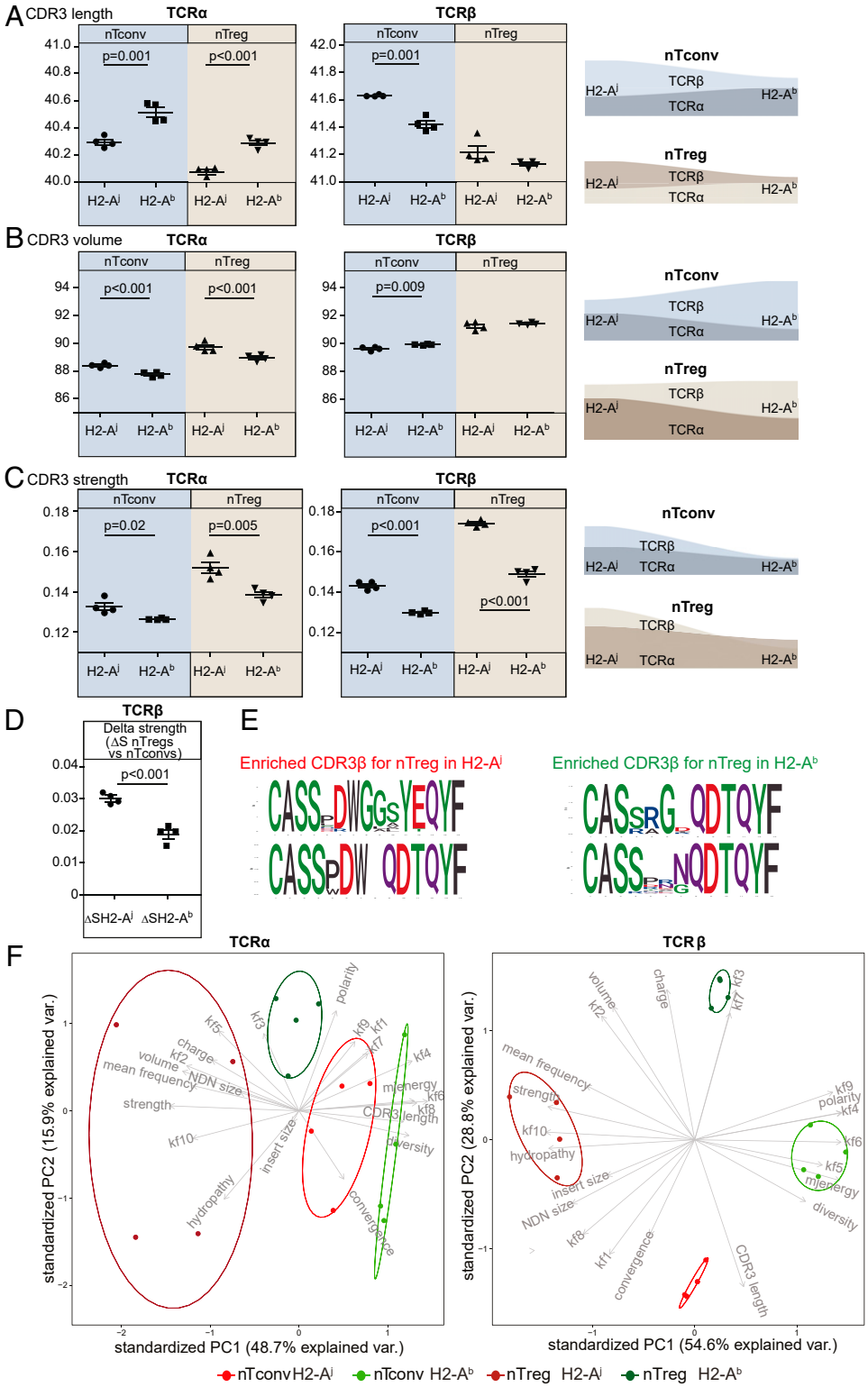

Fig. 4. Characteristics of CDR3α and CDR3β loops in naïve TCR repertoires. Average count of (*A*) CDR3 length and (*B*) volume based on the presence of bulky amino acid side chains for CDR3α and CDR3β in nT_{conv} and nT_{reg} cells from H2-A^b or H2-A^j. *Right* shows schematic illustrations of the differences in CDR3α and CDR3β length and volume parameters in mouse strains. (*C*) Predicted strength of interaction for bulk CDR3α and CDR3β repertoires of nT_{reg} and nT_{conv} cells in H2-A^j and H2-A^b mice. $n = 4$ mice in each group. *Right* shows schematic illustrations of the differences in the CDR3α and CDR3β interaction strength in the two strains. (*D*) The difference in strength values (ΔS) of CDR3β for nT_{reg} and nT_{conv} subsets in H2-A^b vs. H2-A^j. *P* values were computed using Benjamin Hochberg-corrected *t* test, based on false discovery rate <0.05. (*E*) Representative logo plots visualizing amino acid motifs of TCR CDR3β that were significantly enriched as a result of similar recombination events in nT_{reg} cells. (*F*) Cumulative contribution of 22 parameters (e.g., physicochemical properties, Kidera Factors, diversity, NDN size (NDN, added nucleotides flanking the D gene segment), and CDR3 length) in principal component analysis for the TCR repertoires of nT_{reg} and nT_{conv}. Shown here are two-dimensional principal component analysis projections of 2,000 CDR3α and 5,000 CDR3β UMIs that were randomly selected for each sample. PC1, principal component 1; PC2, principal component 2. All parameters were scaled to a similar range using *Z*-score normalization.

Skewed Usage of TCR V Segments and Impact of CDR1 and CDR2 on p-MHC-II–TCR Interaction. The germline-encoded CDR1/CDR2 regions of the TCRα and TCRβ variable domains are believed to contribute mainly to the basic interaction of TCR with MHC-II, modulating the binding angle of the hypervariable region relative to p-MHC (15, 54). Previously, it has been reported that TRBV and TRAV segment usage is strongly associated with the expressed allelic variant of MHC-II (19, 55, 56). The observation that point substitutions in conserved stretches of amino acids in MHC-II alter TRBV and TRAV segment usage in CD4 cells but do not affect the $CD4^+$ numerical count (25) provides further, direct evidence for preselection of TCR repertoires in MHC recognition.

In our experiments, H2-A^j and H2-A^b mice could be separated into clusters of V segment usage depending on MHC context (*SI Appendix*, Fig. S2*D*). Specifically, the TRAV12.2, TRAV4.3, and TRAV12.1 gene segments were used significantly more often by nT_{conv} $CD4^+$ subsets in H2-A^j mice compared with H2-A^b (*SI Appendix*, Fig. S2*D*), whereas the usage of several TRAV segments (TRAV14D.1, TRAV14.3, TRAV7.4) dropped drastically in H2-A^j mice. The observed changes in V segment usage should reflect distinct preferences in selection of those gene segments in H2-A^j mice. In other words, distinct V gene segments would be expected to provide optimal affinity for the *H2-A^j* haplotype of MHC-II, increasing the probability of passing both positive and negative selection.

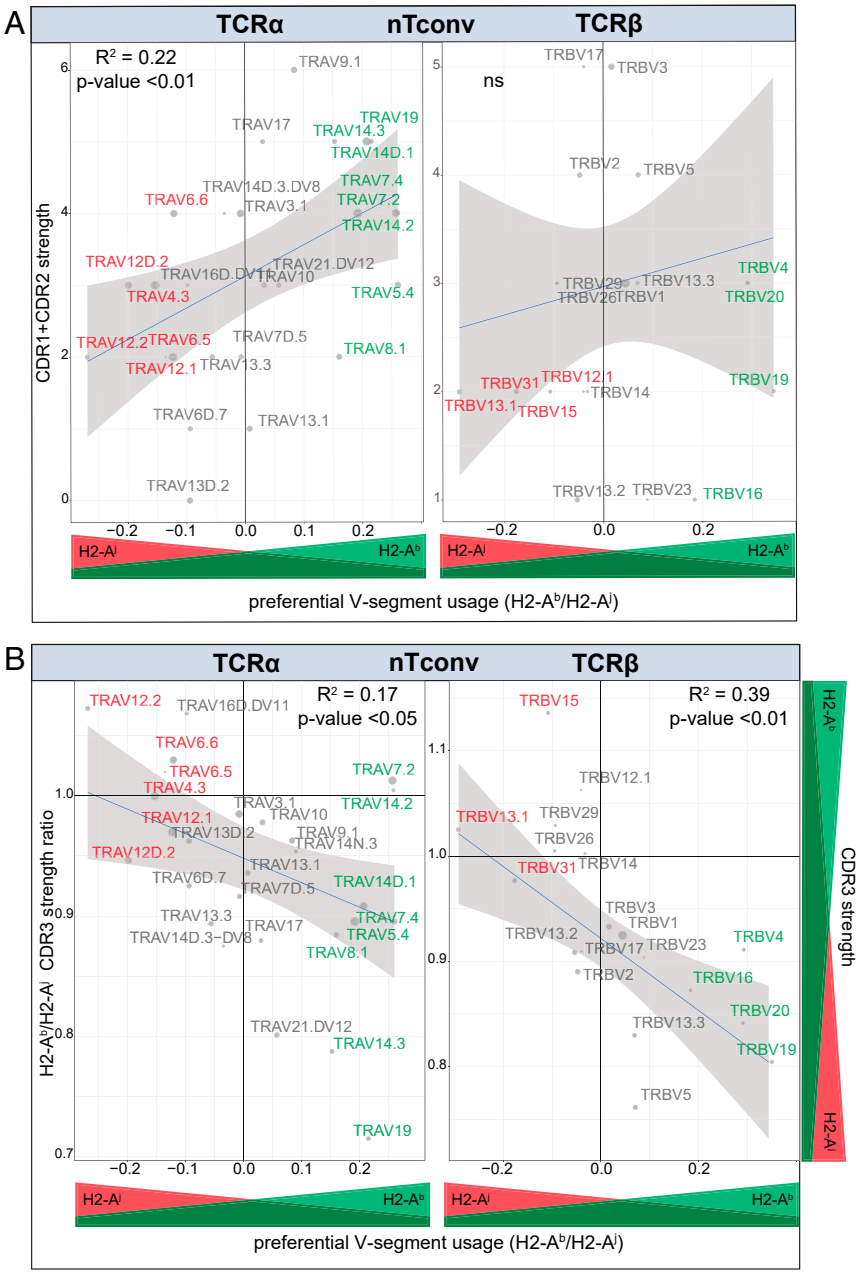

Fig. 5. V segment usage in TCR repertoires of nT_{conv} cell subsets in H2-A^b and H2-A^j mice. (*A*) Correlation of unweighted cumulative (CDR1 + CDR2) strength to ratio of average usage frequency of pooled TRBV or TRAV segments from the nT_{conv} repertoire of H2-A^b vs. H2-A^j mice. (*B*) Correlation between the ratio of weighted CDR3 strength (H2-A^b vs. H2-A^j) and ratio of average usage frequency for TRAV or TRBV segments between the nT_{conv} repertoires of H2-A^b vs. H2-A^j mice. Green and orange colors denote TRBV segments preferentially used in H2-A^b and H2-A^j repertoire, respectively.

Taking into account that TCR CDR1 and CDR2 sequences are identical in both mouse strains, we calculated the cumulative interaction strength based on the average number of strongly interacting amino acids constituting these regions. The cumulative interaction strength of CDR1 and CDR2 for each TRAV segment correlated with the fold change of its usage frequency in H2-A^b nT_{conv} cells relative to H2-A^j (Fig. 5*A*). Preferential selection of defined TRAV segments (e.g., TRAV12.2, TRAV12.1, TRAV4.3, and TRAV6.5) on H2-A^j molecule required low numbers of strongly interacting amino acid residues in the middle of CDR1 and CDR2, while H2-A^b favored the selection of TRAV segments bearing CDR1 and CDR2 with high cumulative interaction strength (e.g., TRAV19, TRAV14.3, TRAV7.4) (Fig. 5*A*).

Certain TRBV segments were also enriched in H2-A^j and H2-A^b contexts but were not clearly related to the strength of corresponding CDR1 and CDR2 regions (Fig. 5*A*), suggesting preferable TCR–MHC interactions mediated by particular complementary structures. Interestingly, the TRBV segments enriched in either *H2-A^j* or *H2-A^b* context carried CDR3β variants with lower interaction strength, on average (Fig. 5*B*). This may indicate that some optimal modes of preencoded CDR1 and CDR2 binding to corresponding MHC are sufficient for these TRBV segment variants, such that strongly interacting CDR3β motifs are not required for successful thymic selection. In the nT_{reg} population, we observed similar shifts in TRBV usage (*SI Appendix*, Fig. S5). However, in contrast to nT_{conv} cells, only TRAV6.6 was specifically enriched in H2-A^j nT_{reg} repertoire, which could reflect the influence of distinct thresholds in the course of nT_{reg} and nT_{conv} cells selection. A similar trend was observed for TRAV segments, where segments with weakly interacting CDR3α loops were generally preferentially selected for any allelic variants of MHC-II (Fig. 5*B*).

We can therefore conclude the following. First, distinct structural features of H2-A influence selection preferences for particular complementary TRAV and TRBV segments. Second, for TRAV segments, these differences are associated with the relative abundance of strongly interacting amino acid residues in CDR1 and CDR2 loops. Finally, in the context of highly complementary interactions between V segments and MHC, the average number of strongly interacting amino acids within the CDR3 region reciprocally decreases, potentially reflecting cumulative fitting to the optimal selection window of interaction strength required to pass thymic selection.

Nonclassical MHC-II Molecule H2-O Does Not Have Essential Impact on Selecting TCR Repertoire. We have previously observed that H2-O^j has several nonconservative substitutions in the *H2-Oa* and *H2-Ob* genes in the *H2^j* haplotype compared with H2-O^b (26) that could make the H2-O^j chaperone dysfunctional (27). To elucidate whether this nonclassical MHC-II molecule contributes to CD4 selection and TCR repertoire shaping, we analyzed repertoires of naïve CD4 subsets in H2-$A^{b\ Ob-/-}$ mice, which lack the H2-O molecule (27) (*SI Appendix*, Table S2*B*). H2-$A^{b\ Ob-/-}$ thymus and spleen had a CD4/CD8 ratio similar to H2-A^b (*SI Appendix*, Fig. S6 *A–C*). This further supports the notion that the inverted CD4/CD8 ratio in H2-A^j mice results from impaired $CD4^+$ T cell selection on the H2-A^j molecule. We assessed physicochemical properties of TCR CDR3 regions of naive $CD4^+$ T cell subsets in H2-$A^{b\ Ob-/-}$, H2-A^b, and H2-A^j. In contrast to H2-A^j mice, both H2-$A^{b\ Ob-/-}$ and H2-A^b nT_{conv} cells had similar numbers of strongly interacting amino acid residues within the central part of CDR3 in both TCRα and TCRβ chains (*SI Appendix*, Fig. S6*D*). In keeping with the results described above, H2-A^j but not H2-$A^{b\ Ob-/-}$ mice tended to have more bulky amino acids in the center of CDR3α and smaller ones in CDR3β. H2-$A^{b\ Ob-/-}$ nT_{conv} cells clearly have CDR3α amino acid side-chain size equal to the congenic H2-A^b strain (*SI Appendix*, Fig. S6*E*). Taken together, these results indicate that differences in the repertoire landscape of nT_{conv} in H2-A^j mice arise from TCR selection on structurally different classical MHC-II molecules but not from allelic variants of H2-O.

Discussion

The preimmune, naïve TCR repertoire landscape formed by thymic and peripheral selection (12) essentially determines an organism's capacity to protect against diverse pathogenic (55, 57) and cancerous antigens (58, 59) and to achieve a balanced network of regulatory interactions between immune cells (23). Convincing data show that the interplay of the TCR repertoire with certain MHC-II allelic variants can predispose individuals to autoimmune diseases (7). The individual MHC context determines the selection of $CD4^+$ T cells and shapes the TCR repertoire, influencing both V and J segment usage (19, 25) and frequencies of particular TCR CDR3 variants (60).

Here, we have focused on identifying how selection in particular MHC-II contexts determines the structure of the TCR repertoire of nT_{conv} and nT_{reg} cells. We used mouse strains expressing different allelic variants of a single classical MHC-II while preserving the natural complexity of the TCR repertoires and obtained several intriguing findings. Confirming previous observations (19, 20, 25), we found prominent differences in usage of V gene segments correlating with the role of CDR1 and CDR2 in p-MHC interaction. Our data on the preferable selection of TCRs bearing defined TRAV and TRBV gene segments by each MHC allele support the germline-encoded model of TCR engagement with p-MHC (18, 61). Furthermore, the CDR3 repertoires of nT_{conv} and nT_{reg} subsets in H2-A^j mice (B6.I-9.3.$Foxp3^{GFP}$) strikingly differed from H2-A^b mice (B6.$Foxp3^{GFP}$) in terms of CDR3α and CDR3β amino acid composition and length.

Polymorphisms near or in the peptide binding groove in MHC-II alter the peptide repertoire and the overall TCR–p-MHC interaction interface (55, 62, 63) and consequently, cause perturbations of the TCR repertoire that successfully passes the first thymic selection. At present, no crystallographic data have been obtained for the p-H2-A^j complex, but molecular modeling of H2-A^j with CLIP peptide has suggested changes in the structure and size of binding pockets P1, P4, P6, P7, and P9 (26). Furthermore, the volume reduction of the H2-A^j P1 pocket, which is implicated in binding to H2-M (DM in humans) (64, 65), might contribute to exchange efficiency of invariant peptide (e.g., CLIP) for self-peptides in MHC-II. The root-mean-square deviation (RMSD) analysis of the molecular dynamics trajectories (*SI Appendix*, Fig. S7) supports the hypothesis of lower binding capacity of H2-A^j for CLIP than H2-A^b. The ability of H2-A^j to form a stable complex with peptides can be affected by the lack of one of the highly conserved H bonds between the peptide backbone and A74, which replaces the highly conserved E74 in the β-chain. These modeling results suggest that diversity of endogenous peptides loaded and stably presented on H2-A^j may be limited, thus hampering the positive selection of diverse CDR3 repertoire. Further comparative studies of endogenous peptide repertoires presented by H2-A^j vs. H2-A^b are required to verify this hypothesis. The most noticeable structural changes in the α-helixes of H2-A^j are the unique substitution Q61K, located in the α-chain, and the very rare substitution R70Q in the β-chain, both of which alter the polarity of interactions between chains. The position αQ61 has been reported to be a contact point between the H2-Aα chain and TCRβ CDR1, CDR2, and CDR3 (66), with targeted substitution significantly reducing the TCR activation threshold (67). Overall, all of these alterations can lead to formation of p-H2-A^j complexes that could affect interaction with TCR and influence the selection of $CD4^+$ T cells.

Based on our results, we speculate that mutual adjustment of the length of the TCRα and TCRβ CDR3 loops is mediated by the distinct docking mode of TCR–p-MHC interaction during thymic selection, such that shortening of the first CDR3 loop is

compensated for by elongation of the second one. One of the notable physicochemical characteristics of CDR3 is the number of strongly interacting (i.e., mainly hydrophobic and aromatic) amino acid residues (43) in the middle of TCRα and TCRβ CDR3. These parts of the TCR interact most closely with presented peptides (41). As shown previously, high strength is characteristic of T_{reg} CDR3 repertoires, reflecting their permissible higher affinity to self p-MHC during thymic selection (22, 31, 32). We observed a remarkable increase in the average representation of strongly interacting residues in both TCRα and TCRβ CDR3 repertoires in H2-A^j mice. The effect was more prominent in T_{reg} cells (Fig. 4*D*), which perhaps, reflects the competitive nature of thymic T_{reg} generation (51). This feature of the TCRs in the nT_{reg} population of H2-A^j mice and its numerical advantage over T_{conv} in the periphery might potentially contribute to susceptibility to tuberculosis via suppression of an efficient and timely Th1 response (68, 69)

H2-A^j genetic context was associated with profound deficiency of $CD4^+$ cells in the thymus and spleen (Fig. 2). At the same time, we found that the TCR repertoire of H2-A^j mice was characterized by narrowed diversity, increased publicity across animals of the same strain, and higher convergence—that is, increased average number of nucleotide CDR3 sequence variants per amino acid sequence variant (Fig. 3 and *SI Appendix*, Fig. S2). Cumulatively, these data led us to conclude that the *H2-A^j* context, due either to the features of the H2-A^j surface or to the restricted pool of H2-A^j–presented peptides (see below), imposes more rigorous selection of an essentially narrowed pool of TCR variants compared with H2-A^b, which also results in numerical deficiency of naive $CD4^+$ cells produced.

The number and diversity of naïve T cells are considered to be crucial for efficient and specific response to a broad range of pathogens (70). Therefore, the increased "gaps" in the narrowed repertoire of $CD4^+$ naive T cells in the *H2-A^j* context could contribute to susceptibility to particular infections, including tuberculosis. However, the diversity of antigenic molecules covered by a TCR repertoire can also be increased through higher cross-reactivity. In other words, an efficient immune response against a wide range of pathogens can be provided either by high diversity of the TCR repertoire and/or by broad TCR cross-reactivity (71–75). Accounting for specific or degenerate TCR recognition is a complex process that depends on the interaction strength of TCR with p-MHC (76). The specific feature of the H2-A^j repertoire is an increased number of strongly interacting amino acid residues in the middle of both CDR3α and CDR3β in nT_{conv} and nT_{reg} repertoires (Fig. 4*C*). Most strongly interacting amino acid residues belong to hydrophobic residues, which form high-affinity binding contacts within protein–protein interfaces (77). TCR cross-reactivity correlates with the frequency of strongly interacting residues in CDR3β (50), and we therefore cannot exclude the possibility that restricted diversity of the TCR repertoire in H2-A^j mice might be offset by potentially enhanced cross-reactivity. This could allow adaptive immunity to recognize comparable diversity of antigens with a narrower repertoire of TCRs, at the cost of potential nonspecific immune response and inflammation. More generally, such a feature of a narrowed repertoire may reflect the natural balance of diversity and cross-reactivity in terms of MHC context selection. It comes in line with the study shown that skewing properties of TCR CDR3 toward increased hydrophobicity was detected in the cases of altered T cell selection (50, 78).

The H2-A^j genome also contains nonclassical MHC-II genes of the *H2^j* haplotype, encoding H2-M and H2-O. These genes are highly conserved across organisms and participate in peptide loading onto the classical MHC-II molecule. Together, H2-O and H2-M modulate the exchange of CLIP peptide with self-peptides, thus influencing peptide repertoire editing (79, 80). Nevertheless, the exact functional role of H2-O in TCR shaping is still elusive (28, 29). H2-O expression is mainly associated with increased CLIP–MHC-II expression on the cell surface (28), and according to recent mass spectrometry data, lack of H2-O expression may alter the diversity of the presented peptide repertoire (81). Evidence suggests that the activity of the H2-O^j allelic variant is reduced, leading to enhanced H2-M function, as is observed in H2-Ob knockout mice (27). However, it has been demonstrated here and in another study (82) that the number of $CD4^+$ T cells in H2-A^b mice lacking H2-O was not diminished, and according to our data, H2-O does not have any prominent effect on generation and features of TCR repertoires of nT_{conv} cells. Thus, we can conclude that polymorphism of H2-A^j, but not H2-O^j, determines the striking differences in the selection of $CD4^+$ T cells.

The association of the *H2-A^j* allele with significantly reduced $CD4^+$ cell numbers in the thymus and spleen might be linked to certain differences in the binding groove of H2-A^j that result in less efficient positive selection of naïve $CD4^+$ T cells. However, the low count of $CD4^+$ cells may not necessarily lead to poor infection control in organisms. The composition and diversity of MHC-bound peptides have a strong effect on positive selection. Previous studies of genetically modified mice, including H2-M- or Ii (CLIP)-deficient mice, have demonstrated that limited multiplicity of low-abundance self-peptides and loose binding of peptides to MHC-II (83–85) result in reduced $CD4^+$ T cell counts. We speculate that the spectrum of peptides that are efficiently presented in H2-A^j might be more limited compared with H2-A^b. The requirement for immunodominant epitope presentation was shown in the elegant experiments by Nepal et al. (86), which also demonstrated that low $CD4^+$ cell counts in the periphery are not critical for controlling *Mycobacterium tuberculosis* (Mtb) dissemination, but proper H2-M function determines the efficiency of the immune response against Mtb infection. Thus, the inability of H2-A^j to bind and present some range of peptides might be related to Mtb susceptibility.

To summarize, in this work we show that distinct MHC-II contexts may prominently shape the features of the selected landscape for both nT_{conv} and nT_{reg} CDR3 repertoires. This influence is measurable not only as a distinct frequency of TRAV and TRBV gene segment usage but also at the level of CDR3α and CDR3β repertoire features. The reciprocal changes in CDR3α and CDR3β repertoire characteristics could reflect the distinct preferences of TCR–p-MHC accommodation. At the same time, the increased number of hydrophobic and strongly interacting amino acid residues in the middle of CDR3α and CDR3β may indicate the necessity of increased average strength of TCR–p-MHC interaction for successful positive selection. This is supported by the narrowed diversity of the more focused naive repertoire produced in the *H2-A^j* context and the decreased counts of naive $CD4^+$ T cells. Thus, we demonstrate that TCR repertoire profiling can provide a "digital" view of T cell selection. Further progress in our understanding of these processes—and ultimately, in our ability to predict the antigen-specific properties of individual TCR repertoire landscapes—should have direct implications for the development of vaccines and T cell-based immunotherapies for cancer and autoimmune disorders.

Materials and Methods

Mice and Genotyping. Mice of inbred strains I/StSnEgYCit (abbreviation I/St, H2^i), C57BL/6JCit (B6, H2^b), H2-recombinant congenic strain B6.I-9.3.19.8 (abbreviation B6.I-9.3) (26), and (B6.I-9.3xB6) F1 hybrid were bred and maintained under conventional, nonspecific pathogen-free (non-SPF) conditions at the Animal Facilities of the Central Institute for Tuberculosis in accordance with the guidelines from Russian Ministry of Health #755 and under the NIH Office of Laboratory Animal Welfare Assurance #A5502-11. Animals were killed using carbon dioxide. Transgenic mouse strain B6.Foxp3^{GFP} (abbreviated here as H2-A^b) was provided by A. Rudensky, Memorial Sloan Kettering Cancer Center, New York, NY (33). The protocol of genotyping is in *SI Appendix*. C57BL/6J *H2-Ob$^{-/-}$* mice were provided

by T. Golovkina, University of Chicago, Chicago, IL. Crossing B6.Foxp3^{GFP} and C57BL/6J *H2-Ob*$^{-/-}$, we generated F1 hybrids that were intercrossed to obtain C57BL/6 Foxp3^{GFP} *H2-Ob*$^{-/-}$ (H2-A$^{b\ Ob-/-}$) mouse strain.

Cell Preparations and Flow Cytometry. As described in detail in *SI Appendix*, spleen, thymus, and lymph nodes of 2 to 3 month-old mice were removed. Suspensions of thymus, spleen, and lymph node cells were obtained using routine procedures. The following monoclonal antibodies for fluorescence-activated cell sorting (FACS) analysis were purchased from eBiosciences, Biolegend, BD Biosciences, or Miltenyi Biotec: CD3-FITC (clone 145-2C11), CD3-PE-Cy5 (clone 17A2), CD4-PerCP or CD4-APC-cy7 (clone GK1.5), CD8-APC (or PE or PE-Cy5; clone 53-6.7), B220-APC (clone RA3-6B2), CD62L-PE (clone Mel-14), CD44-eFluor 450 (clone IM7), MHC class II I-E (clone 14-4-4S), and Foxp3-PE (clone FJK-16s). MHC class II I-A-Alexa Fluor 488 (clone 10.2-16) cross-reactive to the β-chain of H2-A^{j} was provided by T. Golovkina. Intracellular staining of Foxp3 and Ki67 (clone Sp6; Thermo Scientific) was done according to the manufacturer's protocol for Foxp3/Transcription Factor Staining Buffer Set (eBioscience). Data were acquired using a BD Biosciences FACSCalibur flow cytometer and BD Biosciences FACSCARIA III and analyzed using FlowJo software (Tree Star).

Cell Sorting. For cell sorting, CD4^{+} T cells isolated from spleen of individual H2-A^{b} and H2-A^{j} mice were enriched prior to FACS sorting by microbeads technology using negative selection with CD4^{+} T cell Isolation Kit II (Miltenyi Biotec). Purified CD4^{+} T cells were stained with CD4, CD62L, and CD44 antibodies and sorted into the two following populations: nT$_{reg}$ (CD4^{+}Foxp3 GFP^{+}CD62L^{+}CD44^{-}), nT$_{conv}$ (CD4^{+}Foxp3 GFP^{-}CD62L^{+}CD44^{-}). Each cell population was sorted at >98% purity. All cell populations were collected directly into RLT buffer (Qiagen) and stored at −70 °C.

RNA Isolation and cDNA Library Preparation. Total RNA was isolated using RNeasy mini kit (Qiagen) according to the manufacturer's instructions. In all experiments described here, cDNA libraries were obtained using 5′-RACE with UMIs (35, 41) using the Mouse TCR profiling kit (MiLaboratory LLC). Details of the procedure can be found in *SI Appendix*.

Raw Sequencing Data Analysis. We performed two independent sequencing runs. In the first run, we sequenced in parallel cDNA libraries obtained from four H2-A^{b} and four H2-A^{j} mice. In the second run, we performed sequencing of samples obtained from three H2-A^{b}, three H2-A^{j}, and five H2-A$^{b\ Ob-/-}$ mice (87). Sequencing data were analyzed using MiGEC (35), MiXCR, and VDJtools software (88). Details of the analysis can be found in *SI Appendix*.

Statistical Analysis. Statistical analysis was performed using Prism 5.0 (GraphPad) and programming language R. One-sided *t* test with unequal variances (Welch's test) was used for comparison of T cell population composition, and Mann–Whitney *U* test was used for comparison of repertoire properties. Details of the analysis can be found in *SI Appendix*.

Physicochemical CDR3 Properties and Correlation Analysis. CDR1 and CDR2 strength was calculated for mouse sequences according to http://imgt.org/ and the built-in VDJtools correspondence table (88). CDR3 strength was calculated as clonotype-weighted average, for the five amino acids in the middle of CDR3. All nonfunctional TCR sequences were excluded from analysis. Details of the analysis can be found in *SI Appendix*. The resulting *P* values were adjusted using Benjamini–Hochberg (BH) method with a false discovery rate not greater than 0.05.

Clustering Analysis. For graph construction, we take the top 2,500 most frequent clonotypes clustered by CDR3 nucleotide sequence identity out of each sample. Each node on the graph represents a single translated clonotype (clonotypes with same amino acid sequence but different nucleotide sequences do not collapse). Nodes were connected by edges if the Levenshtein distance between them was equal to zero or one and V segments of these clonotypes were identical. Further, we use only the top five biggest connected components out of the graph. Connectivity of repertoires was compared using Gini index and Mann–Whitney *U* test. Visualization of graph was made in Cytoscape.

For TCR motif detection, we separately pooled TCR clonotypes of all four H2-A^{b} and four H2-A^{j} mice and performed TCR neighborhood enrichment test, as described in ref. 89. Details of the analysis can be found in *SI Appendix*.

Principal Component Analysis. Data on the CDR3α and CDR3β repertoires were analyzed separately but using similar pipeline. For the principal component analysis we used physicochemical properties for the central five amino acids within CDR3 and other CDR3 parameters. Details of the analysis can be found in *SI Appendix*. All parameters were scaled to similar range using *Z*-score normalization; principal component analysis was performed and visualized using programming language R and package ggbiplot.

ACKNOWLEDGMENTS. We thank Prof. A. Rudensky (Sloan Kettering Institute) for providing the Foxp3^{GFP} reporter mice and T. Golovkina (The University of Chicago) for providing C57BL/6J *H2-Ob*$^{-/-}$ mice. We also thank M. Eisenstein for the English editing. Cell sorting experiments were carried out using the equipment provided by the Core facility in Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry. Sequencing and data analysis part of the work was supported by Russian Science Foundation [Grant 16-15-00149] while mouse breeding and flow cytometry data analysis was supported by Russian Foundation for Basic Research [Grant 19-015-00082 (to N.N.L.)].

1. K. A. Hogquist, S. C. Jameson, The self-obsession of T cells: How TCR signaling thresholds affect fate 'decisions' and effector function. *Nat. Immunol.* **15**, 815–823 (2014).
2. L. Racioppi, F. Ronchese, L. A. Matis, R. N. Germain, Peptide-major histocompatibility complex class II complexes with mixed agonist/antagonist properties provide evidence for ligand-related differences in T cell receptor-dependent intracellular signaling. *J. Exp. Med.* **177**, 1047–1060 (1993).
3. A. Sette, J. Alexander, K. Snoke, H. M. Grey, Antigen analogs as tools to study T-cell activation function and activation. *Semin. Immunol.* **8**, 103–108 (1996).
4. L. Yin, J. Scott-Browne, J. W. Kappler, L. Gapin, P. Marrack, T cells and their eons-old obsession with MHC. *Immunol. Rev.* **250**, 49–60 (2012).
5. J. A. Todd, J. I. Bell, H. O. McDevitt, HLA-DQ beta gene contributes to susceptibility and resistance to insulin-dependent diabetes mellitus. *Nature* **329**, 599–604 (1987).
6. Z. Hovhannisyan *et al*., The role of HLA-DQ8 beta57 polymorphism in the anti-gluten T-cell response in coeliac disease. *Nature* **456**, 534–538 (2008).
7. M. Relle, A. Schwarting, Role of MHC-linked susceptibility genes in the pathogenesis of human and murine lupus. *Clin. Dev. Immunol.* **2012**, 584374 (2012).
8. G. Lythe, R. E. Callard, R. L. Hoare, C. Molina-París, How many TCR clonotypes does a body maintain? *J. Theor. Biol.* **389**, 214–224 (2016).
9. T. Dupic, Q. Marcou, A. M. Walczak, T. Mora, Genesis of the αβ T-cell receptor. *PLOS Comput. Biol.* **15**, e1006874 (2019).
10. A. J. Yates, Theories and quantification of thymic selection. *Front. Immunol.* **5**, 13 (2014).
11. R. D. Kilpatrick *et al*., Homeostasis of the naive CD4+ T cell compartment during aging. *J. Immunol.* **180**, 1499–1507 (2008).
12. E. G. Houston Jr., P. J. Fink, MHC drives TCR repertoire shaping, but not maturation, in recent thymic emigrants. *J. Immunol.* **183**, 7244–7249 (2009).
13. I. Stefanová, J. R. Dorfman, R. N. Germain, Self-recognition promotes the foreign antigen sensitivity of naive T lymphocytes. *Nature* **420**, 429–434 (2002).
14. J. S. Bridgeman, A. K. Sewell, J. J. Miles, D. A. Price, D. K. Cole, Structural and biophysical determinants of αβ T-cell antigen recognition. *Immunology* **135**, 9–18 (2012).
15. P. Marrack, J. P. Scott-Browne, S. Dai, L. Gapin, J. W. Kappler, Evolutionarily conserved amino acids that control TCR-MHC interaction. *Annu. Rev. Immunol.* **26**, 171–203 (2008).
16. A. Casrouge *et al*., Size estimate of the alpha beta TCR repertoire of naive mouse splenocytes. *J. Immunol.* **164**, 5782–5787 (2000).
17. Q. Qi *et al*., Diversity and clonal selection in the human T-cell repertoire. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 13139–13144 (2014).
18. N. L. La Gruta, S. Gras, S. R. Daley, P. G. Thomas, J. Rossjohn, Understanding the drivers of MHC restriction of T cell receptors. *Nat. Rev. Immunol.* **18**, 467–478 (2018).
19. E. Sharon *et al*., Genetic variation in MHC proteins is associated with T cell receptor expression biases. *Nat. Genet.* **48**, 995–1002 (2016).
20. K. Gao *et al*., Germline-encoded TCR-MHC contacts promote TCR V gene bias in umbilical cord blood T cell repertoire. *Front. Immunol.* **10**, 2064 (2019).
21. W. S. DeWitt 3rd *et al*., Human T cell receptor occurrence patterns encode immune history, genetic background, and receptor specificity. *eLife* **7**, e38358 (2018).
22. C. S. Hsieh, Y. Zheng, Y. Liang, J. D. Fontenot, A. Y. Rudensky, An intersection between the self-reactive regulatory and nonregulatory T cell receptor repertoires. *Nat. Immunol.* **7**, 401–410 (2006).
23. M. Ono, R. J. Tanaka, Controversies concerning thymus-derived regulatory T cells: Fundamental issues and a new perspective. *Immunol. Cell Biol.* **94**, 3–10 (2016).
24. D. K. Cole *et al*., Human TCR-binding affinity is governed by MHC class restriction. *J. Immunol.* **178**, 5727–5734 (2007).
25. D. Silberman *et al*., Class II major histocompatibility complex mutant mice to study the germ-line bias of T-cell antigen receptors. *Proc. Natl. Acad. Sci. U.S.A.* **113**, E5608–E5617 (2016).
26. N. Logunova, M. Korotetskaya, V. Polshakov, A. Apt, The QTL within the H2 complex involved in the control of tuberculosis infection in mice is the classical class II H2-Ab1 gene. *PLoS Genet.* **11**, e1005672 (2015).

27. L. K. Denzin *et al*., Neutralizing antibody responses to viral infections are linked to the non-classical MHC class II gene H2-Ob. *Immunity* **47**, 310–322.e7 (2017).
28. Y. Gu, P. E. Jensen, X. Chen, Immunodeficiency and autoimmunity in H2-O-deficient mice. *J. Immunol.* **190**, 126–137 (2013).
29. E. D. Mellins, L. J. Stern, HLA-DM and HLA-DO, key regulators of MHC-II processing and presentation. *Curr. Opin. Immunol.* **26**, 115–122 (2014).
30. B. Yalcin *et al*., Commercially available outbred mice for genome-wide association studies. *PLoS Genet.* **6**, e1001085 (2010).
31. M. O. Li, A. Y. Rudensky, T cell receptor signalling in the control of regulatory T cell differentiation and function. *Nat. Rev. Immunol.* **16**, 220–233 (2016).
32. M. S. Jordan *et al*., Thymic selection of CD4+CD25+ regulatory T cells induced by an agonist self-peptide. *Nat. Immunol.* **2**, 301–306 (2001).
33. J. D. Fontenot *et al*., Regulatory T cell lineage specification by the forkhead transcription factor foxp3. *Immunity* **22**, 329–341 (2005).
34. K. Attridge, L. S. Walker, Homeostasis and function of regulatory T cells (Tregs) in vivo: Lessons from TCR-transgenic Tregs. *Immunol. Rev.* **259**, 23–39 (2014).
35. M. Shugay *et al*., Towards error-free profiling of immune repertoires. *Nat. Methods* **11**, 653–655 (2014).
36. D. A. Bolotin *et al*., MiXCR: Software for comprehensive adaptive immunity profiling. *Nat. Methods* **12**, 380–381 (2015).
37. B. J. Manfras, D. Terjung, B. O. Boehm, Non-productive human TCR beta chain genes represent V-D-J diversity before selection upon function: Insight into biased usage of TCRBD and TCRBJ genes and diversity of CDR3 region length. *Hum. Immunol.* **60**, 1090–1100 (1999).
38. E. V. Putintseva *et al*., Mother and child T cell receptor repertoires: Deep profiling study. *Front. Immunol.* **4**, 463 (2013).
39. M. Izraelson *et al*., Comparative analysis of murine T-cell receptor repertoires. *Immunology* **153**, 133–144 (2018).
40. M. F. Quigley *et al*., Convergent recombination shapes the clonotypic landscape of the naive T-cell repertoire. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 19414–19419 (2010).
41. E. S. Egorov *et al*., The changing landscape of naive T cell receptor repertoire with human aging. *Front. Immunol.* **9**, 1618 (2018).
42. A. S. Bergot *et al*., TCR sequences and tissue distribution discriminate the subsets of naïve and activated/memory Treg cells in mice. *Eur. J. Immunol.* **45**, 1524–1534 (2015).
43. S. Miyazawa, R. L. Jernigan, Residue-residue potentials with a favorable contact pair term and an unfavorable high packing density term, for simulation and threading. *J. Mol. Biol.* **256**, 623–644 (1996).
44. A. Kosmrlj, A. K. Jha, E. S. Huseby, M. Kardar, A. K. Chakraborty, How the thymus designs antigen-specific and self-tolerant T cell receptor sequences. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 16671–16676 (2008).
45. S. Chakrabarti, S. H. Bryant, A. R. Panchenko, Functional specificity lies within the properties and evolutionary changes of amino acids. *J. Mol. Biol.* **373**, 801–810 (2007).
46. N. K. Singh *et al*., Emerging concepts in TCR specificity: Rationalizing and (maybe) predicting outcomes. *J. Immunol.* **199**, 2203–2213 (2017).
47. Y. Feng *et al*., A mechanism for expansion of regulatory T-cell repertoire and its role in self-tolerance. *Nature* **528**, 132–136 (2015).
48. G. De Simone *et al*., CXCR3 identifies human naive $CD8^+$ T cells with enhanced effector differentiation potential. *J. Immunol.* **203**, 3179–3189 (2019).
49. O. Grimsholm *et al*., The interplay between CD27(dull) and CD27(bright) B cells ensures the flexibility, stability, and resilience of human B cell memory. *Cell Rep.* **30**, 2963–2977.e6 (2020).
50. B. D. Stadinski *et al*., Hydrophobic CDR3 residues promote the development of self-reactive T cells. *Nat. Immunol.* **17**, 946–955 (2016).
51. Z. Dembic, On integrity in immunity during ontogeny or how thymic regulatory T cells work. *Scand. J. Immunol.* **90**, e12806 (2019).
52. P. Billam *et al*., T Cell receptor clonotype influences epitope hierarchy in the CD8+ T cell response to respiratory syncytial virus infection. *J. Biol. Chem.* **286**, 4829–4841 (2011).
53. N. N. Logunova *et al*., Restricted MHC-peptide repertoire predisposes to autoimmunity. *J. Exp. Med.* **202**, 73–84 (2005).
54. L. Deng, R. J. Langley, Q. Wang, S. L. Topalian, R. A. Mariuzza, Structural insights into the editing of germ-line-encoded interactions between T-cell receptor and MHC class II by Vα CDR3. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 14960–14965 (2012).
55. X. Chen, L. Poncette, T. Blankenstein, Human TCR-MHC coevolution after divergence from mice includes increased nontemplate-encoded CDR3 diversity. *J. Exp. Med.* **214**, 3417–3433 (2017).
56. P. Marrack *et al*., The somatically generated portion of T cell receptor CDR3α contributes to the MHC allele specificity of the T cell receptor. *eLife* **6**, e30918 (2017).
57. W. Luo *et al*., Limited T cell receptor repertoire diversity in tuberculosis patients correlates with clinical severity. *PLoS One* **7**, e48117 (2012).
58. E. Strønen *et al*., Targeting of cancer neoantigens with donor-derived T cell receptor repertoires. *Science* **352**, 1337–1341 (2016).
59. J. H. Cui *et al*., TCR repertoire as a novel indicator for immune monitoring and prognosis assessment of patients with cervical cancer. *Front. Immunol.* **9**, 2729 (2018).
60. A. K. Bentzen, S. R. Hadrup, Evolution of MHC-based technologies used for detection of antigen-responsive T cells. *Cancer Immunol. Immunother.* **66**, 657–666 (2017).
61. J. P. Scott-Browne, J. White, J. W. Kappler, L. Gapin, P. Marrack, Germline-encoded amino acids in the alphabeta T-cell receptor control thymic selection. *Nature* **458**, 1043–1046 (2009).
62. K. Falk, O. Rötzschke, S. Stevanović, G. Jung, H. G. Rammensee, Allele-specific motifs revealed by sequencing of self-peptides eluted from MHC molecules. *Nature* **351**, 290–296 (1991).
63. Z. Zavala-Ruiz, I. Strug, M. W. Anderson, J. Gorski, L. J. Stern, A polymorphic pocket at the P10 position contributes to peptide binding specificity in class II MHC proteins. *Chem. Biol.* **11**, 1395–1402 (2004).
64. C. A. Painter, L. J. Stern, Structural insights into HLA-DM mediated MHC II peptide exchange. *Curr. Top. Biochem. Res.* **13**, 39–55 (2011).
65. L. Yin, L. J. Stern, HLA-DM focuses on conformational flexibility around P1 pocket to catalyze peptide exchange. *Front. Immunol.* **4**, 336 (2013).
66. B. D. Stadinski *et al*., A role for differential variable gene pairing in creating T cell receptors specific for unique major histocompatibility ligands. *Immunity* **35**, 694–704 (2011).
67. B. D. Stadinski *et al*., Effect of CDR3 sequences and distal V gene residues in regulating TCR-MHC contacts and ligand specificity. *J. Immunol.* **192**, 6071–6082 (2014).
68. J. P. Scott-Browne *et al*., Expansion and function of Foxp3-expressing T regulatory cells during tuberculosis. *J. Exp. Med.* **204**, 2159–2169 (2007).
69. M. Kursar *et al*., Cutting edge: Regulatory T cells prevent efficient clearance of Mycobacterium tuberculosis. *J. Immunol.* **178**, 2661–2665 (2007).
70. M. P. Davenport, D. A. Price, A. J. McMichael, The T cell repertoire in infection and vaccination: Implications for control of persistent viruses. *Curr. Opin. Immunol.* **19**, 294–300 (2007).
71. H. A. Van Den Berg, C. Molina-París, A. K. Sewell, Specific T-cell activation in an unspecific T-cell repertoire. *Sci. Prog.* **94**, 245–264 (2011).
72. L. F. Su, B. A. Kidd, A. Han, J. J. Kotzin, M. M. Davis, Virus-specific CD4(+) memory-phenotype T cells are abundant in unexposed adults. *Immunity* **38**, 373–383 (2013).
73. S. C. Clute *et al*., Broad cross-reactive TCR repertoires recognizing dissimilar Epstein-Barr and influenza A virus epitopes. *J. Immunol.* **185**, 6753–6764 (2010).
74. M. E. Birnbaum *et al*., Deconstructing the peptide-MHC specificity of T cell recognition. *Cell* **157**, 1073–1087 (2014).
75. D. Mason, A very high level of crossreactivity is an essential feature of the T-cell receptor. *Immunol. Today* **19**, 395–404 (1998).
76. Y. Sun *et al*., Specificity, privacy, and degeneracy in the CD4 T cell receptor repertoire following immunization. *Front. Immunol.* **8**, 430 (2017).
77. A. A. Bogan, K. S. Thorn, Anatomy of hot spots in protein interfaces. *J. Mol. Biol.* **280**, 1–9 (1998).
78. J. Henwood *et al*., Human T cell repertoire generation in the absence of MHC class II expression results in a circulating CD4+CD8- population with altered physicochemical properties of complementarity-determining region 3. *J. Immunol.* **156**, 895–906 (1996).
79. L. K. Denzin, N. F. Robbins, C. Carboy-Newcomb, P. Cresswell, Assembly and intracellular transport of HLA-DM and correction of the class II antigen-processing defect in T2 cells. *Immunity* **1**, 595–606 (1994).
80. A. B. Vogt, H. Kropshofer, G. Moldenhauer, G. J. Hämmerling, Kinetic analysis of peptide loading onto HLA-DR molecules mediated by HLA-DM. *Proc. Natl. Acad. Sci. U.S.A.* **93**, 9724–9729 (1996).
81. P. P. Nanaware, M. M. Jurewicz, J. D. Leszyk, S. A. Shaffer, L. J. Stern, HLA-DO modulates the diversity of the MHC-II self-peptidome. *Mol. Cell. Proteomics* **18**, 490–503 (2019).
82. M. Liljedahl *et al*., Altered antigen presentation in mice lacking H2-O. *Immunity* **8**, 233–243 (1998).
83. G. Kenty, W. D. Martin, L. Van Kaer, E. K. Bikoff, MHC class II expression in double mutant mice lacking invariant chain and DM functions. *J. Immunol.* **160**, 606–614 (1998).
84. S. Viville *et al*., Mice lacking the MHC class II-associated invariant chain. *Cell* **72**, 635–648 (1993).
85. C. E. Grubin, S. Kovats, P. deRoos, A. Y. Rudensky, Deficient positive selection of CD4 T cells in mice displaying altered repertoires of MHC class II-bound self-peptides. *Immunity* **7**, 197–208 (1997).
86. R. M. Nepal, B. Vesosky, J. Turner, P. Bryant, DM, but not cathepsin L, is required to control an aerosol infection with Mycobacterium tuberculosis. *J. Leukoc. Biol.* **84**, 1011–1018 (2008).
87. N. N. Logunova *et al*., Data from Bioproject: Sequencing of TCR repertoires in mice with MHC-II polymorphism. National Center for Biotechnology Information Sequence Read Archive. https://www.ncbi.nlm.nih.gov/bioproject/PRJNA600142. Deposited 9 January 2020.
88. M. Shugay *et al*., VDJtools: Unifying post-analysis of T cell receptor repertoires. *PLOS Comput. Biol.* **11**, e1004503 (2015).
89. M. V. Pogorelyy, M. Shugay, A framework for annotation of antigen specificities in high-throughput T-cell repertoire sequencing studies. *Front. Immunol.* **10**, 2159 (2019).