Th17 cells with a pathogenic signature in joints of ZAP-70 mutant arthritic mice harbor unique TCR repertoires that share features with WT Tregs

Mara A. Llamas-Covarrubias[1,2,3], Atsushi Tanaka[4,5,6], Martin Loza[7], Diego Diez[8], Zichang Xu[1], Ee Lyn Lim[5], Shunsuke Teraguchi[9], Shimon Sakaguchi[4,5], Daron M. Standley[1,2]†

[1]Laboratory of Systems Immunology, WPI Immunology Frontier Research Center, Osaka University, Suita, Japan; [2]Department of Genome Informatics, Research Institute for Microbial Diseases (RIMD), Osaka University, Suita, Japan; [3]Departamento de Biología Molecular y Genómica, Centro Universitario de Ciencias de la Salud, Universidad de Guadalajara, Guadalajara, México; [4]Department of Experimental Pathology, Institute for Frontier Medical Sciences, Kyoto University, Kyoto, Japan; [5]Laboratory of Experimental Immunology, WPI Immunology Frontier Research Center, Osaka University, Suita, Japan; [6]Department of Frontier Research in Tumor Immunology, Center of Medical Innovation and Translational Research, Graduate School of Medicine, Osaka University, Osaka, Japan; [7]Laboratory of functional analysis in silico, Human Genome Center, The University of Tokyo, Tokyo, Japan; [8]Quantitative Immunology Research Unit, Immunology Frontier Research Center, Osaka University, Suita, Japan; [9]Faculty of Data Science, Shiga University, Hikone, Japan

†Corresponding author:
Laboratory of Systems Immunology, WPI Immunology Frontier Research Center, Osaka University, Suita, Japan. Tel: +81-6-6879-8366; Fax: +81-6-6879-8366. Email: standley@biken.osaka-u.ac.jp

Abstract

Autoimmune diseases such as rheumatoid arthritis (RA) represent a significant source of worldwide morbidity and disability. Despite extensive research, the molecular mechanisms underlying the development of self-reactive T cells and the breach of tolerance remain elusive. Specifically, the intricate gene expression patterns and T cell receptors (TCRs) responsible for autoimmune diseases are still largely unknown. Mutation in the ZAP-70 gene (SKG, ZAC) that cause moderate attenuation of TCR signaling, can result in autoimmune manifestations akin to RA in mice. This study aims to characterize the gene expression profiles and TCR repertoires of distinct autoreactive T cell populations within the joints, lymph nodes, and spleens of arthritic ZAC mice at the single-cell level, comparing them with those of wild-type (WT) and non-arthritic ZAP-70 mutant mice (SKG). Our analysis revealed two distinct populations of Th17 cells (Th17 *Il22*+ and Th17 *Tnfsf11*+) in ZAC mice, each exhibiting unique transcriptional signatures, tissue distributions, and TCR repertoires. Notably, Th17 *Tnfsf11*+ cells, active in pathogenic transcriptional programs, are uniquely localized in inflamed joints and absent in WT or SKG mice. Furthermore, their TCR repertoire shared features with WT Treg repertoires, suggesting that reduced TCR signaling enables T cells bearing self-reactive TCRs to evade negative selection, leading to a repertoire shift from Treg to Tconv. This repertoire shift not only enables self-reactive Tconv development but also depletes specific Treg reactivities. These findings provide a comprehensive understanding of Treg and Tconv transcriptomes and repertoires in ZAP-70 mutant mice, contributing to a mechanistic understanding of the break in tolerance in autoimmune diseases.

Introduction

The development of autoimmune diseases is influenced by various factors, with the presence of self-reactive conventional T cells (Tconv) being a critical element. Equally important is the failure of regulatory T cells (Treg) to suppress self-reactivity and inflammation (Pisetsky

2023). Several mechanisms for Treg suppression, executed by cell-cell contact with target cells or by the production of soluble factors, have been reported in the literature, including production of inhibitory cytokines, cytolysis, metabolic disruption and indirect suppression by modulation of antigen presenting cell (APC) maturation or function (Vignali et al., 2008). T cell receptor (TCR) recognition of antigen is key for Treg development in the thymus and for their function in peripheral tissues (Levine et al., 2014; Li et al., 2016; Pohar et al., 2018); for example, Treg receptor specificity allows for Treg-Tconv co-localization, which enables additional regulatory mechanisms leading to Tconv suppression (Liu et al., 2015; Akkaya et al., 2019). Moreover, it is generally accepted that the Treg repertoire is biased to self-recognition (Jordan et al., 2001; Hsieh et al., 2006; Aschenbrenner et al., 2007; Lee et al., 2012a). Therefore, disruption of the normal balance between Treg and Tconv repertoires and activity can bias the immune system toward autoimmunity.

Specific mutations in the ZAP-70 gene in mice that weaken downstream TCR signaling upon antigen recognition and alter the Treg-Tconv repertoire balance have been previously described (Sakaguchi et al., 2003, Tanaka et al., 2023). Moderate weakening predisposes mice to an autoimmune disease closely resembling rheumatoid arthritis (RA). The SKG mouse model is characterized by a W163C mutation in the ZAP-70 gene that causes a fifteen-fold reduction in ZAP-70 affinity for phosphorylated CD3ζ (Sakaguchi et al. 2003) and necessitates secondary stimulation to induce the disease. In contrast, ZAC mice, which carry an H165A mutation that causes only an eight-fold reduction, spontaneously develop chronic autoimmune arthritis, even under specific pathogen-free (SPF) conditions (Tanaka et al., 2023). The underlying mechanism is likely attributed to the disruption of the normal thymic selection process, by allowing T cells with moderately to strongly self-reactive TCRs to escape negative selection (Tanaka et al. 2010, Tanaka et al. 2023); as a consequence, some T cells harboring TCRs that would normally be observed in Treg repertoires would be expected to develop into self-reactive Tconvs.

In order to test this hypothesis, we examined the T cells of WT, SKG and ZAC mice at the single-cell level. Interestingly, we observed two Th17 cell subtypes in the joints of ZAC mice that were characterized by distinct transcriptional profiles and TCR repertoires, one of which occurred exclusively in inflamed joints. In addition, a group of public TCRs were identified that were expanded, highly convergent in their complementarity-determining region (CDR) sequences and over-represented in Th17 cells. We observed a striking overlap between ZAC Tconv and WT Treg repertoires accompanied by a reduced intra-Treg-Tconv overlap within ZAC mice, suggesting that a depletion of critical components of the Treg TCR repertoire accompanies the expansion of Tconvs with a Treg-like repertoire in ZAC mice. Our results support of a model where, upon ZAP-70-mediated signal weakening, T cells that would normally develop into Tregs instead develop into self-reactive Tconvs, resulting in a breakdown in self-tolerance and susceptibility to autoimmune arthritis.

Results

Arthritic ZAC mice harbor two groups of Th17 Tconv cells in their joints with distinct gene programs

To study the transcriptomes of self-reactive Tconv cells in-depth, we sorted cells from the joints, draining lymph nodes (dLN), and spleen of ZAC mice and performed RNA+V(D)J single-cell sequencing. Gene expression analysis revealed a considerable overlap in the transcriptional profiles of spleens and dLN but a separate profile for joint cells (Figure S1a). Unsupervised clustering delineated nine gene expression clusters with similar distribution in replicates across tissues (Supplementary text 1, Figure 1a-d, S1c-d). Two clusters (1 and 4) showed canonical Th17 markers (*Rorc, Il17a, Il17f*), but were distinct in their gene expression programs and tissue abundances (Figure 1c-d, S1c-e). Cluster 1 was abundant in joints, representing ~50% of cells obtained from this tissue, but was also detected in the spleen and dLN at smaller but similar proportions (~15-17%) (Figure 1d, S1c). Relative to cluster 4, cells in cluster 1 differentially expressed *Klf2*, *S1pr1* and chemokine receptors such

as *Ccr2*, *Ccr4* and *Ccr6*; suggesting a phenotype of migratory cells, possibly recently emigrated, or ready to emigrate from lymphoid tissues (Figure 1e, Table S4). Consistently, pathway analysis revealed an enrichment in the regulation of T cell migration (GO: 20000404) and cell chemotaxis (GO: 0060326) GO Terms (Figure 1f, Table S5). In contrast, cluster 4 was almost exclusively observed in joints, representing approximately 27% of joint cells but less than 2% of dLN cells and less than 1% of Spleen cells. The top DEG in cluster 4 was *Tnfsf11* (encoding RANKL), a critical factor involved in osteoclast differentiation and activation that plays a central role in bone destruction in RA (Komatsu et al., 2022; Sims et al., 2015) (Figure 1e). Two other osteoclastogenic genes (*Csf1*, *Anxa2*) were highly expressed in this cluster as well; overall, we observed a significant enrichment in the osteoclast differentiation (GO:0030316) GO term in cluster 4 as compared to cluster 1 (Figure 1f, Table S5). In addition, transcripts involved in the response to TNF stimuli (for example: *Tnfrsf18*, *Tnfrsf9*, *Tnfsf8*, *Tnfrsf4*, *Lta*, *Tnf*), and the cellular response to tumor necrosis factor (GO:0072356) GO terms were enriched in cluster 4. Finally, this cluster uniquely expressed transcription factors associated with a pathogenic Th17 phenotype: *Rbpj* (Meyer Zu Horste et al., 2016), *Crem* (Ohl et al., 2018), *Atf3* (Glal et al., 2018), *Bhlhe40* (Lin et al., 2016), and *Igfbp7* (Di Toro et al., 2020) (Table S4). Finally, the transcription factor *Klf2*, a known inhibitor of NF-κB-dependent genes which also prevents Tfh differentiation and RORγt expression (Jha et al., 2017; Lee et al., 2015) was clearly absent from cluster 4 (Figure S1b). These results indicate that cluster 4 contains pathogenic Th17 cells with an increased potential to activate osteoclasts and thus promote bone damage. Interestingly, these two Th17 subtypes could be clearly distinguished by their expression of *Il22* or *Tnfsf11* (Figure 1g); thus, we annotated clusters 1 and 4 as Th17 *Il22*+ and Th17 *Tnfsf11*+, respectively. In order to further explore the dynamics of the identified cell states, we implemented Slingshot's pipeline for pseudotime and trajectory inference (Street et al., 2018) with the naïve cluster of cells as the starting point. We found that the highest pseudotime values were concentrated in Th17 *Tnfsf11*+, Trm and Tfh cells (Figure 1h), suggesting that these cells represent endpoint lineage states, which is consistent with their effector phenotype. Thus, we have identified the presence of two distinct groups of Th17 cells in arthritic ZAC mice: Th17 *Il22*+ cells,

characterized by a transcriptional profile indicative of migratory capabilities and the expression of the transcription factor *Klf2*; and Th17 *Tnfsf11*+ cells, which exhibit a pathogenic transcriptome and are predominantly localized within inflamed joints.

To contextualize our findings in arthritic ZAC mice tissues, we replicated the analyses in WT and SKG mice. Given that the WT and SKG mice were nonarthritic, we sampled only their spleens and integrated ZAC spleen data in order to facilitate direct comparison of clusters (Supplementary Text 2). Naïve cells were abundant in WT, and to a lesser extent SKG, but were almost absent in ZAC mice. Intermediate and Th17 cells together represented nearly 80% of ZAC Tconv cells, 11-40% of SKG and 8-33% of WT Tconv cells (Figure S2b). Importantly, Th17 cells represented only 1 and 6% of Tconvs in each WT replicate (Figure S2b). Our results thus indicate that Th17 or pre-Th17 cells (Intermediate), as well as detectable pathology, increased in frequency within a moderate range of TCR signaling weakening from WT to SKG to ZAC. We next compared the transcriptional programs of Tconv cells from spleens of WT and SKG with those of Th17 cells identified in ZAC mice tissues by calculating gene module scores for the transcriptional signatures from ZAC Th17 *Il22*+ and Th17 *Tnfsf11*+ cells (Supplementary Table 8). The Th17 *Il22*+ signature was more strongly expressed than the Th17 *Tnfsf11*+ program, in general, and was higher in the Intermediate and Th17 clusters (Figure S2h). Thus, Th17 cells are present in WT and SKG mice as well, but their transcriptional program is closer to ZAC Th17 *Il22*+ cells suggesting that Th17 *Tnfsf11*+ cells are unique to inflamed joint tissue.

Th17 Tconv cells in the joints of arthritic ZAC mice were expanded, and contained both private and public clones

Having identified two groups of potentially self-reactive Th17 cells with different transcriptomic features, we next asked whether their repertoires are also distinct. First, we examined the degree of clonal expansion, as evaluated by simple count of clonotypes, and found that all replicates and tissues exhibited clonal expansion, but that it was most prominent

in the joints and dLN (Figure S3a). Next, we calculated the proportion of a repertoire represented by each clone (clonal proportion) in each replicate and classified clones as: small (clonal proportion $<1x10^{-3}$), medium (clonal proportion from $1x10^{-3} – 1x10^{-2}$) or large (clonal proportion $>1x10^{-2}$). Large and medium clones were abundant in both Th17 cell groups, Intermediate-2, Trm and proliferating cells (Figure 2a-b, S3b-c). Interestingly, we also observed an increased frequency of singletons in Th17 *Tnfsf11*+ cells compared to Th17 *Il22*+ cells. Repertoire overlap between phenotype clusters at the exact paired CDR3 amino acid sequence level as evaluated by the Morisita-Horn overlap index (Chiffelle et al., 2020) revealed a significant overlap between Intermediate-2, Th17 *Il22*+, Trm and Proliferating cells (Figure 2c). By comparison, the overlap between Th17 *Tnfsf11*+ cells and other cell phenotypes was smaller, implying the existence of a unique repertoire in Th17 *Tnfsf11*+ cells. Despite the difference in clonal overlap, the overall diversity of the Th17 *Tnfsf11*+ cluster was similar to that of Intermediate-2, Trm and Proliferating clusters (Figure S3d). Together, these results indicate that although both Th17 *Il22*+ and Th17 *Tnfsf11*+ clusters were expanded, their repertoires were distinct, implying the targeting of different antigens. Nevertheless, no statistical differences were found in gene usage preferences (Figure S3e-h).

Next, we searched for the presence of public clones, defined as those shared in two or more individual mice. We identified a total of 9 unique public clones occurring in all tissues (Figure 2d, Table S9). Most of these had medium clonal proportions in at least one replicate and were most commonly associated with Intermediate-2 or Th17 *Il22*+ phenotypes. However, 5 of the 9 public clones had medium or large clonal proportions in at least one replicate and were associated with the Th17 *Tnfsf11*+ phenotype (Figure S3i). Interestingly, one clone (TRAV13-1, TRAJ40, CAMEPGNYKYVF, TRBV16, TRBJ1-1, CASSLDNTEVFF) occurred in 3 out of 4 replicates with medium or large clonal proportion. This clone was identified in all tissues and was abundant in Intermediate-2 and Th17 *Il22*+ clusters; however, it was also present in the Th17 *Tnfsf11*+ cluster in one replicate (Figure S3i). In terms of gene usage preferences of public clones, we observed a skewing towards TRAJ18, TRBV1 and TRBJ2-4 (Figure 2e). When considering paired gene usage, we

observed 2 public clones sharing TRAJ18 and TRBV19 (Figure S3j, Table S9). Remarkably, we identified a public CDR3b sequence (CASSLGGQNTLYF) previously documented in the VDJdb, a database of TCR sequences with established antigen specificities (Shugay et al., 2018). This sequence corresponds to a murine cytomegalovirus (MCMV) epitope recognized by a CD8 T clone (HGIRNASFI, Table S10).

In order to evaluate TCR convergence independent of gene usage, we used TCRdist (Dash et al., 2017), which measures the distance between TCRs using CDR sequences. A bias in the distribution of distances towards lower values was interpreted as an indication of TCR convergence. For this analysis, we calculated TCR distances for the two Th17 clusters and for the public clones, using the naïve cluster as a reference. The mean distance of public clones was significantly smaller compared to the Th17 or naïve clusters, indicating that public clones were convergent, presumably due to their ability to recognize similar antigens (Figure 2f); in addition, an S.v.N motif (where: S=S; v=VILM; N=N; .=any amino acid) was identified in beta chains of public clones (Figure 2g, S3k), notably, this motif was observed in the beta chain of the highly public clone described before. In the Th17 and naive groups, the average distances were not significantly different; surprisingly, in ZAC Th17 repertoires, TCR distances were in general larger than in WT Tconv naïve repertoires (Figure S3l), highlighting the polyclonal nature of Th17 cells in ZAC mice.

Together, these results indicate that although Th17 response in arthritic ZAC mice is highly polyclonal, a number of public clones were observed in inflamed tissue. Moreover, these public clones were broadly expanded, and shown greater TCR convergence than other Th17 cells, suggesting that they likely play a role in the recognition of related self-antigens. On the other hand, the repertoire of joint specific Th17 *Tnfsf11*+ cells was overall less similar than that of other Tconvs, with a higher number of TCR singletons.

Treg cells in the Joints of arthritic ZAC mice are abundant in tissue-resident memory phenotypes, but also in intermediate (not fully effector) phenotypes

Autoimmune diseases develop in susceptible hosts due to the combination of many factors. The presence of self-reactive T cells is critical, but the failure in Treg suppression of self-reactivity and inflammation is also required. In order to examine the Tconv-Treg landscape, we sorted Treg cells from the samples previously described and performed single cell RNA and V(D)J sequencing. Broad overlap between dLN and Spleen Treg transcriptomes was observed, whereas joint transcriptomes were less diverse (Figure S4a). Unsupervised clustering identified seven gene expression clusters (Figure 3a). By examining canonical Treg markers, top cluster markers and enriched GO terms (Supplementary text 3, Figure 3b-c, Table S11- S12), we identified naïve (*Il7r, Ccr7, Sell*), Trm (*Il7r, Sell, Cd44, Cd69, Itgae*), and proliferating (*Mki67*) Treg cells. In addition, by examining predicted cell trajectories, we annotated two additional populations of cells as Intermediate-1 (cluster 3) and Intermediate-2 (cluster 0) (Figure S4d). The transcriptional signature of cluster 1 was highly similar to cluster 0 and only a few genes were differentially expressed between these two clusters (Figure S4e-f); in addition, cluster 1 showed no clear expression of typical Treg activation markers (Figure 3b), so we labeled it Intermediate-3. Interestingly, the Trm group displayed the highest expression of gene markers associated with Treg activation and effector function (*Ccr2, Cxcr3, Ikzf2, Icos, Gzmb*) (Figure 3b). In addition, this cluster displayed the highest expression of *Cd5*, a marker typically used as a measure of TCR signal intensity (Mandl et al., 2013; Persaud et al., 2014) (Figure S4b); thus, we considered this group as likely recognizing tissue antigens with high affinity and exerting the highest suppressive function. Trm and Intermediate-2 cells were abundant in joints, whereas Intermediate-1 Tregs was abundant in dLN, and Intermediate-3 Tregs in the spleen; moreover, naïve Tregs were nearly absent in joints (Figure 3d, S4c). Clonal expansion was observed in all tissues at similar levels (Figure S4h-i). As expected, naïve cells exhibited the least expansion, while Intermediate-3 Tregs exhibited the highest degree of expansion, with a higher proportion of expanded clones than Trm cells (Figure 3e-f, S4j). There was a significant degree of similarity in the repertoires of cells in Intermediate-2 and Intermediate-3 cell groups, and, to a lesser degree, Trm (Figure 3g). In addition, only one public clone was identified in ZAC

Tregs (Supplementary table 13). Taken together, the Treg landscape in tissues of arthritic ZAC mice included a population of Trm cells with high suppressive potential mostly concentrated in joints. Nevertheless, intermediate phenotypes with low expression of effector markers were abundant as well, indicating an impaired Treg response. In addition, we found that public clones were much less frequent in the Treg compartment as compared to the Tconv compartment in ZAC mice.

In WT and SKG mice Naïve Treg cells were abundant (Supplementary text 4, Figure S5b, S5f). A group of cells annotated as effector *Gzmb*+ cells which were equivalent to Intermediate-2 and Intermediate-3 clusters in ZAC tissues (Figure S5e), was identified and represented about 50% of ZAC Spleen Treg cells while they were much less abundant in WT and SKG samples (SKG: 11.6%; WT: 7%). In addition, even though the abundance of Trm cells was low in general, this group was almost exclusively observed in ZAC samples (ZAC: 5-7%; SKG: 0-0.8%; WT: 0.08-0.8%) (Figure S5b, S5f). Thus, our results indicate that Trm Treg cells were exclusive of ZAC mice.

We next explored Treg-Tconv cell-cell crosstalk across the different cell phenotypes by predicting cell communication events using Cellphone DB (Vento-Tormo et al., 2018). Overall, we found that the number of predicted interactions between Th17 *Tnfsf11*+ with the Treg Intermediate and activated subtypes (Treg clusters 0, 1, 3 and 4) was only slightly higher than for Th17 *Il22*+ cells (Figure S4k), indicating that the overall ability of the two Th17 subtypes to interact with Tregs was not substantially different. We next explored the qualitative nature of these predicted cell communication events by looking for interactions representing Treg suppressive mechanisms (as explained in methods), costimulatory or coinhibitory interactions. We did not detect significant interactions from Treg towards Tconv inducing suppression; however, we found Th17 *Tnfsf11*+ cells to be uniquely predicted to interact with Tregs in two potentially suppressive ways (Figure 3h). First, a significant FASLG-FAS prediction, with the potential to induce Naïve, Intermediate and Trm Treg cell death was observed. Second, TNFSF4-TNFRSF4 (OX40L-OX40) crosstalk, which has been

reported to suppress Treg function (Fu et al., 2020), was uniquely predicted for Th17 *Tnfsf11*+ with Naïve, Intermediate and Trm Treg cells. Further examination of expression of these transcripts highlights the uniqueness of these potential interactions (Figure S4l). Such predicted interactions were not observed in WT or SKG spleens, and the expression of *Fasl* and *Tnfsf4* was very limited in Th17 cells from spleen (Figure S5h). Thus, our data suggest that Treg effector function is disrupted in ZAC mice and that Tconv cells, in particular Th17 *Tnfsf11*+, may be able to counteract Treg activity by inducing Treg cell death and inactivation.

TCR repertoires of self-reactive Th17 are more Treg-like than Tconv-like

Self-reactive Th17 cells would likely have been deleted or developed into Tregs under conditions of normal TCR signaling. We next asked if the TCR repertoires of self-reactive Th17 cells in ZAP-70 mutant mice were more Treg-like than Tconv-like. When we looked for CDR3 matches (as described in Methods) at the paired sequencing data of our WT mice, we found too little overlap with ZAC mice to draw any conclusions (Figure 4a, S6a). This is unsurprising because the number of T cells in each sample (total Treg and Tconv number per sample: 436 – 17,258; Table S1) is many orders of magnitude smaller than the number of possible TCRs (~10^8 in mice) (Vrisekoop et al., 2014). In order to increase the sensitivity, we next studied the overlap of paratope sequences, which have been shown to be sensitive and specific to targeted antigen and epitope in the case of BCRs (Richardson et al., 2021; Saputri et al. 2023). To this end, we performed paratope clustering of the whole dataset and calculated Morisita-Horn indices by considering paratope clusters as receptor identities. However, even though the overall magnitude of overlap was higher, we only observed slightly increased overlaps for ZAC Tconv-WT Treg repertoires as compared to ZAC Tconv-WT Tconv repertoires for beta chains (Figure 4b, S6b-c). Next, in order to expand the number of WT TCRs for comparison, we turned to previously published Tconv or Treg TCR reference datasets (Logunova et al., 2020; Lu et al., 2020; Ko et al., 2020; Wolf et al., 2016), with the Logunova set being at least an order of magnitude larger than the rest (Table S16).

For each query TCR in an experimental dataset (WT_Tconv, WT_Treg, SKG_Tconv, SKG_Treg, ZAC_Tconv, ZAC_Treg) we computed the number of Tconv and Treg hits in a given reference dataset (Logunova, Lu, Ko, Wolf). Based on the number of Tconv and Treg hits, along with the number of reference Tconvs and Tregs, we deemed the query TCR Tconv-like, Treg-like or neither using a hypergeometric p-value cutoff of 0.001. We then expressed the total number of Tconv-like and Treg-like hits as a fraction of the original query ZAP-70 repertoire. Considering the largest (Logunova) reference dataset, we found that the fraction Treg-like TCRs within the ZAC Tconv repertoires was dramatically higher than the fraction of Tconv-like TCRs (Figure 4c). The raw fractions of Tconv-like or Treg-like TCRs varied with the reference dataset; however, when we expressed these fractions relative to the WT Tconv query repertoire, we observed a consistent trend that both ZAC mutant TCRs were 2-12 times more Treg-like than Tconv-like (Figure 4d, S6d). In addition, when exploring the Th17 subgroups in detail, we noticed that the highest ratio of WT Treg:WT Tconv fractional repertoire were observed in the ZAC Th17 *Tnfsf11*+ cluster for three out of four studies (Logunova, 2020; Lu, 2020 and Wolf, 2016). Interestingly, in two of these studies (Lu, 2020 and Wolf, 2016), no WT Tconv overlap was observed, while WT Treg overlap was clearly detected (Figure 4e-f, S6e). Thus, our analysis indicates that, consistent with the repertoire shift hypothesis, ZAP-70 mutant Tconv repertoires are more Treg-like than Tconv-like, and this similarity is even more pronounced in pathogenic Th17 *Tnfsf11*+ cells that concentrate in inflamed tissue than for Tconv cells in general.

The analysis above involves a number of computational steps. We next sought a simpler procedure that could be used to independently validate the overall conclusions reached. Lagattuta and co-workers have recently developed a machine learning tool that enables quantification of the regulatory potential of a TCR (TCR-intrinsic regulatory potential, TiRP) based on TCR beta chain features (Lagatutta et al., 2022). We next calculated TiRP scores for WT, SKG and ZAC repertoires in order to ascertain systematic differences. As expected, our Treg datasets exhibited higher TiRP scores than Tconv datasets, in general. However, more importantly, and consistent with the above analysis, ZAP-70 mutant Tconvs had

significantly higher TiRP scores than WT Tconvs (Figure 4g). In addition, we note that TiRP scores increased with disease susceptibility, especially in the Treg repertoires (i.e. higher in ZAC than SKG). When considering the ZAC dataset phenotypes in detail, we found that Th17 *Il22*+ cell TiRP scores were not significantly different from WT Tconvs, but Th17 *Tnfsf11*+ scores were significantly higher, again indicating a bias towards a more Treg-like repertoire in Th17 *Tnfsf11*+ cells than in normal WT Tconv cells (Figure 4h). Together, these results support the Tconv repertoire shift model as a consequence of ZAP-70 mutations, which would allow a set of highly Treg-like TCRs to mature as Tconvs and become pathogenic Th17 cells that accumulate in joint tissue upon activation.

Evidence of missing TCR specificities in ZAP-70 mutant WT Treg repertoires

In line with the repertoire shift hypothesis, the weakening of TCR signaling has the potential to also impact the Treg repertoire by generating a deficit in certain TCR specificities. It has been reported that missing specificities in the Treg repertoire have a large impact on the suppressive function against self-reactive T cells (Ooi et al., 2017; Pohar et al., 2018). This raises the possibility that an absence of Treg specificities within ZAP-70 mutant mice repertoires contribute to a loss in tolerance toward specific self-antigens. To explore this hypothesis, we analyzed the Treg-Tconv repertoire similarity within each mouse and compared it across mouse model groups (WT, SKG, ZAC). As with the Tconv analysis above, we first used the Morisita-Horn index for exact CDR3 overlap and found a notably reduced Treg-Tconv overlap in ZAC mice relative to WT or SKG, especially in single chain repertoires (Figure 5a). When we focused on Tconv subgroups, we noticed that in WT and SKG mice, paired overlap within the Treg repertoire mostly occurred in intermediate, Th17 and Tfh groups, suggesting a suppressive potential for shared Tregs towards activated T cells. Such enrichment was not found in ZAC mice (Figure 5b). We next examined the Treg-Tconv overlap of paratope sequences in each sample and found consistently smaller overlaps in ZAC Tconv repertoires with their respective Tregs as compared to WT or SKG mice (Figure 5c). To further test these observations, we employed the TCRdist method and calculated repertoire-repertoire correlations for pooled WT, SKG or ZAC repertoires; here, more similar

repertoires will exhibit higher correlation coefficients (represented in yellow). We found that WT Treg and WT Tconv repertoires were well correlated, whereas SKG and ZAC Tconv repertoires exhibited higher correlation internally than with their corresponding Treg repertoire. Indeed, the ZAC Treg repertoire was poorly correlated in general, indicating that it was distant from the other tested repertoires (Figure 5d). The Treg-Tconv overlapping clones were generally distributed across the intermediate, Th17 and Tfh Tconv phenotypes and the intermediate and Gzmb+ Treg phenotypes in WT and SKG mice (Figure 5e-f, S7a-b) and tended to represent the highest proportions in the Th17 repertoires. In ZAC mice, overlapping clones were distributed mainly across Intermediate-2, Th17 *Il22*+ and Trm Tconv cells (Figure 5g, S7c-d) and even though some overlapping clones were also detected in the Th17 *Tnfsf11*+ phenotype, they only represented a small proportion of the repertoire in this group. In the Treg compartment of ZAC mice, overlapping clones were concentrated in Intermediate and Trm subtypes. In general, in all mouse models, the proportion represented by the overlapping clones in the Treg subtypes was much smaller than in Tconv, presumably related with the generally observed reduced expansion of Treg cells (Figure 3e-f). Together, these results are consistent with the scenario where ZAP-70 mutant repertoires are depleted in certain Treg TCR specificities, which in turn, may contribute to a loss of tolerance and the development of autoimmune arthritis.

Discussion

Central T cell tolerance is established during lymphocyte development in the thymus through the process of thymic selection. This process ensures the elimination of self-reactive Tconv cells while promoting the generation of Treg cells, which play a crucial role in suppressing unwanted inflammatory responses in peripheral tissues (Kronenberg et al., 2005; Goodnow et al., 2005; Ashby et al., 2023). In peripheral tissues, Treg and Tconv cells operate in a delicate balance, facilitating protective immunity while suppressing autoaggressive immune reactions. Our study aimed to enhance our understanding of autoimmunity by providing a detailed characterization of Tconv and Treg cell phenotypes and repertoires in a mouse model of RA.

One crucial aspect of autoimmune diseases involves identifying self-reactive T cells and characterizing their target auto-antigens. Extensive literature documents the pivotal role of IL-17 in the pathogenesis of various autoimmune diseases, including RA (Lubberts et al., 2002; Nakae et al., 2003; Corneth et al., 2014), and Th17 cells constitute the primary source of IL-17A and IL-17F in the immune system. Here, we identified two transcriptionally distinct populations of Th17 cells within inflamed joints of arthritic ZAC mice: Th17 *Il22*+ and Th17 *Tnfsf11*+. Apart from differentially expressing *Il22* and *Tnfsf11*, these cells exhibited clear differences in their transcriptional profiles. Th17 Il22+ cells were enriched in pathways associated with T cell migration, whereas Th17 *Tnfsf11*+ cells displayed a pathogenic Th17 cell signature with high osteoclastic potential. However, it has been previously shown that Th17 cells differentiated *in vitro* by the stimulation with anti CD3 mAb and Il-23 cannot induce osteoclastogenesis alone despite expressing RANKL, the product of *Tnfsf11* gene (Sato et al., 2006). Previous studies have highlighted the impact of cytokine combinations present in the microenvironment during Th17 cell differentiation on their phenotype and effector function. For instance, a combination of IL-23 and IL-6 induces the most pathogenic phenotype (Lee et al., 2012b), while IL-6 is crucial for inducing a population of Th17 cells with direct osteoclastogenic abilities (Komatsu et al., 2014). However, our study did not detect the exFoxP3 signature in any of the Th17 groups, and the osteoclastogenic ability of Th17 *Tnfsf11*+ cells was not further assessed. Future *in vitro* and *in vivo* experiments are warranted to confirm the origin and pathogenic potential of Th17 *Tnfsf11*+ cells as their full phenotype might differ from Th17 cells previously analyzed in bulk. These investigations will provide valuable insights into the role of Th17 cells in autoimmune diseases, particularly rheumatoid arthritis. Furthermore, trajectory prediction suggests that Th17 *Tnfsf11*+ cells represent a terminal differentiation state. This observation aligns with the research conducted by Glaublomme et al. (Glaublomme et al., 2015), who analyzed cell heterogeneity and trajectory progression of Th17 cells in mice afflicted with experimental autoimmune encephalomyelitis (EAE). Their study documented a progression from a self-renewing cellular state, primarily enriched in lymph nodes, to what they identified as an effector Th17 pre-Th1-like phenotype found in both lymph nodes and the central

nervous system (CNS). This pre-Th1-like phenotype expressed chemokine receptors akin to our Th17 *Il22*+ population. Subsequently, the Th17/Th1-like effector phenotype, concentrated in the CNS, exhibited a pathogenic signature and high expression of *Tnfsf11*. Intriguingly, in vitro experiments confirmed that the Th17/Th1-like effector phenotype arises only under pathogenic conditions (IL-1β, IL-6, IL-23).

In our study, we also identified a group of Th17 cells in splenic Tconvs from WT or non-arthritic SKG mice. However, their transcriptomes were more similar to the Th17 *Il22*+ phenotype, and the Th17 *Tnfsf11*+ signature was observed to a lesser extent. It is noteworthy that the repertoire of Th17 *Il22*+ and Th17 *Tnfsf11*+ populations exhibited relatively low degrees of similarity and despite the expansion of Th17 *Tnfsf11*+ cells, Th17 *Il22*+ cells underwent a more pronounced expansion. This result indicates the existence of substantial differences in the repertoires of these two Th17 populations. However, with our current approach, we were unable to identify clear preferences in differential gene usage, TCR motifs, or repertoire convergence. The phenomenon of clonal expansion among CD4+ T cells in RA patients or mouse models of autoimmune diseases has been extensively documented (Schmidt et al., 1996, Wagner et al., 2002, Argyriou et al., 2022). Nevertheless, studies suggest that clonal dynamics change throughout disease progression, with evidence indicating increased diversity and reduced clonal dominance in established disease stages (Klarenbeek et al., 2012; Al Khabouri et al., 2021). Additionally, investigations into the contribution of high-affinity versus lower-affinity T cell specificities to disease establishment and progression, as observed in EAE, have revealed the critical role of high-affinity autoreactive cells in initiating disease, which can be maintained by lower-affinity cells once established (Yi et al., 2022). In our study, it is plausible that the Th17 *Il22*+ population comprises cells involved in the early stages of disease, given their predicted trajectory, increased expansion relative to Th17 *Tnfsf11*+, and their similarity to Th17 cells in WT or non-arthritic SKG mice. However, elucidating their specific role in disease progression and their affinity for self-antigens falls outside the scope of this paper and warrants further investigation in subsequent studies.

Remarkably, we observed a significant number of public clones within Tconv cells of ZAC mice, particularly concentrated within both Th17 phenotypes, and these were notably expanded. In comparison to ZAC Th17 *Il22*+, Th17 *Tnfsf11*+, or Naïve cells, public clones exhibited significant convergence, suggesting their potential involvement in the recognition of similar self-antigens. Indeed, the significance of public clones, along with the specific contribution of TCRβ in the development of autoimmune conditions, has been previously documented in the literature (Menezes et al., 2007; Zhao et al., 2016; Lu et al., 2022). Furthermore, certain genes appeared to be preferentially utilized by these public clones, including TRAJ18, TRBV1, and TRBJ2-4, with the presence of an S.v.N motif in their beta chains, which was notably detected in the most prevalent public clone. However, despite our thorough analysis, none of the CDR3 sequences within our public clones were found in public antigen-receptor databases such as VDJdb. Consequently, the identity of any potential epitopes or amino acid motifs remains unknown.

The mechanisms governing the loss of T cell central tolerance remain incompletely understood, yet it has been postulated that defective TCR signaling may disrupt the T cell selection process, allowing T cells bearing highly self-reactive TCRs to evade negative selection (Tanaka et al., 2010, Tanaka et al., 2023). Under normal TCR signaling conditions, these cells would typically undergo deletion or selection into the Treg repertoire. Consequently, the TCRs of self-reactive ZAP-70 mutant Tconv cells would exhibit greater similarity to those of WT Tregs than to those of WT Tconv cells. Our study presented evidence supporting an increased similarity between the ZAC Tconv repertoire and the WT Treg repertoire, particularly among pathogenic Th17 cells. These findings are in line with the observation that T cells infiltrating prostatic autoimmune lesions in Aire$^{-/-}$ mice express receptors preferentially found in Tregs from Aire$^{+/+}$ mice. Notably, potentially self-reactive receptors were observed in the Aire$^{-/-}$ Treg repertoire but at significantly reduced frequencies (Malchow et al., 2016). It is worth noting that achieving sufficient resolution in our analysis required contrasting our data with large public WT TCR datasets, underscoring the challenges associated with obtaining comprehensive TCR coverage in repertoire analyses,

particularly with single-cell sequencing methods (Valkiers et al., 2022). However, the additional calculation of TCR-intrinsic regulatory potential (Lagatutta et al., 2022) provided further evidence of increased similarity between ZAP-70 mutant Tconv repertoires and Treg repertoires, especially evident in pathogenic Th17 cells.

Evidence suggests that the affinity of the TCR not only influences the selection of Tregs but also shapes their phenotypes and functional properties in the periphery (Wyss et al., 2016; Zemmour et al., 2018). In our study, we identified a population of Trm cells enriched in joint tissue of ZAC mice, a population largely absent in WT or SKG mice. These Trm cells exhibited high expression levels of markers associated with Treg activation and function. However, it has been clearly established that Tregs from ZAC mice are more severely affected by the mutation in ZAP-70 compared to Tconvs. Despite normal or even elevated expression of typical factors associated with Treg activity, both in vitro and in vivo studies have demonstrated impaired function in ZAC Tregs (Tanaka et al., 2023).

Furthermore, our prediction of possible cell-cell crosstalk events between Treg and Tconv cells in ZAC mice identified two potential mechanisms of Treg suppression uniquely mediated by Th17 *Tnfsf11*+ cells. One mechanism involves a FASLG-FAS interaction from Th17 *Tnfsf11*+ cells toward intermediate phenotypes of ZAC Treg, potentially inducing Treg cell death (O' Reilly et al., 2009). The second mechanism implicates the delivery of OX40 (encoded by *Tnfrsf4*) signals from Th17 *Tnfsf11*+ cells through OX40L (encoded by *Tnfsf4*) in intermediate and Trm cells. The OX40-OX40L axis has been implicated in providing costimulatory signals to conventional T cells, enhancing their function. However, its effect on Treg cells is contrary, as it suppresses their differentiation and activity (Croft et al., 2009). While Treg-Tconv cell contact has been shown to be necessary for efficient suppression of Tconv proliferation in vitro (Collison et al., 2009), in the context of ZAC mice, this interaction could potentially favor the counter-regulation of Tregs by pathogenic Th17 cells, exacerbating the baseline deficient Treg activity in these mice. However, the viability of these mechanisms and their significance in vivo require further examination.

Overall, we observed that the degree of expansion in the Treg compartment was lower than in Tconvs across all mice, and only one public clone identified in ZAC mice. This finding

aligns with previous research indicating that the proliferative capacity of Tregs from ZAC mice is reduced (Tanaka et al., 2023). Additionally, the accumulation of public Treg clones in tumor tissues has been documented (Sainz-Perez et al., 2012). Therefore, our results may reflect not only the compromised proliferation of ZAC Tregs but also disturbances in repertoire composition.

The repertoire shift hypothesis also opens the possibility of losing certain TCR specificities within the Treg compartment. Considering that antigen recognition by the Treg TCR appears to be less degenerated compared to Tconv cells (Kieback et al., 2016; Pohar et al., 2018; Leonard et al., 2017), the loss of certain Treg specificities could profoundly impact tolerance towards specific self-antigens (Ooi et al., 2017). While it is generally acknowledged that Treg and Tconv repertoires are distinct, a small proportion of their repertoires typically overlap (Hsieh et al., 2004; Pacholczyk et al., 2006). In our study, we observed a reduced overlap in ZAC Treg-Tconv repertoires compared to SKG or WT mice, particularly with a noticeable loss of overlap with activated Tconv cell groups. Additionally, repertoire correlation between Treg and Tconv compartments was diminished in ZAP70 mutants relative to WT mice. Therefore, it is plausible to speculate that ZAP-70 mutation may not only influence the loss of tolerance by shifting the Tconv repertoire but also by depriving the Treg repertoire of specificities. The identity of such specificities remains to be discovered but could offer insights into the development of transgenic TCR Treg immunotherapy for the treatment of autoimmune diseases.

In summary, our study identified two distinct Th17 subtypes in the joints of ZAC arthritic mice, characterized by unique transcriptional profiles and repertoires. We also provided evidence of a Treg-Tconv repertoire shift in ZAP-70 mutant mice, as evidenced by the increased similarity of ZAC Tconv repertoires to WT Treg repertoires, particularly among pathogenic ZAC Th17 cells. Additionally, we proposed the possibility of a loss of certain Treg specificities associated with the repertoire shift induced by weakened TCR signaling. These processes, coupled with the decrease in Treg activity, likely contribute to the development of autoimmunity in ZAC mice. Our research expands the current understanding of tolerance breakdown and proposes novel avenues for future research in the field.

Methods

Mice

Animal experiments were approved by the institutional review board of Osaka University, and animals were treated in accordance with the institutional guidelines of the Immunology Frontier Research Center, Osaka University. BALB/c and SKG mice were acquired from CLEA Japan. SKG (Sakaguchi et al., 2003), ZAC (Tanaka et al., 2023) and FIG (Foxp3-IRES-GFP knock-in; Ohkura et al., 2012) mice were previously described. In this study, FIG BALB/c WT, FIG SKG and FIG ZAC mice were used and generated as previously reported (Tanaka et al., 2023). All animals were maintained in Specific pathogen-free (SPF) facilities.

Droplet-based single cell RNA sequencing.

$CD4^+Foxp3^-$ or $Foxp3^+$ cells from FIG BALB/c WT, FIG SKG or FIG ZAC mice were sorted by FACSAria SORP (BD Biosciences). For FIG BALB/c WT and FIG SKG mice, cells were obtained from spleens, processed with the Chromium Single Cell 5' Library & Gel bead kit (10x Genomics), loaded onto a Chromium Single Cell G Chip, and encapsulated in a Chromium single cell controller (10x Genomics) to generate single-cell gel beads in the emulsion (GEMs) following manufacturer's protocol. After encapsulation, reverse transcription and cDNA amplification were performed, followed by target enrichment of V(D)J segments and the independent construction of V(D)J and Gene expression libraries as recommended by the manufacturer's instructions.

For FIG ZAC mice, cells from inflamed joint, draining popliteal lymph node (dLN) and spleen from each replicate were barcoded with TotalSeq hashtag antibodies (BioLegend) and then pooled by tissue and cell type (i.e., Joint Tconv, Joint Treg, etc.) before the generation of GEMs. Next, RNA and V(D)J libraries were produced as described above and an

additional library for the barcoded antibodies was prepared. HiSeq3000 and NovaSeq6000 were used for sequencing.

Single cell RNA-seq data processing

Raw reads were aligned to the mouse reference genome (GRCm38 and GRCm38 VDJ v5.0.0, from 10× Genomics), filtered, and demultiplexed to create gene expression matrices by Cell Ranger (v.5.0). The resulting data was analyzed in R software (v 4.1.2) with Seurat v4.0 package (Hao et al., 2021; Stoeckius et al., 2018; Satija et al., 2015).

Where necessary, hashtag demultiplexing was performed by the *HTODemux* function from Seurat and only data from cells classified as singlets was kept for further steps. Next, low-quality data was removed from all datasets, this means cells with unique feature counts less than 200 or over 2500, and cells in which mitochondrial reads represented >5% of reads. Only cells annotated to a single alpha and a single beta chain in the V(D)J result were retained for further analysis. Normalization was performed by the "LogNormalize" method from Seurat and the top 2,000 variable genes were identified in each dataset independently.

Batch correction and data integration were performed in combinations of data as follows: 1) Tconvs from ZAC tissues; 2) Tregs from ZAC tissues; 3) Tconvs from spleen (including WT, SKG and ZAC samples); 4) Tregs from spleen (including WT, SKG and ZAC samples). Batch correction was performed with the Canek R package (Loza et al., 2022) using the intersecting variable genes for the datasets to be integrated in each run.

After integration and QC, the final datasets for analysis consisted of: 6,509 cells for ZAC Tconv, 5,306 cells for ZAC Treg, 24,743 cells for Spleen Tconv and 25,227 cells for Spleen Treg (Table S1). The uniform manifold approximation and projection (UMAP) were used to visualize gene expression profiles and clusters from the principal components (PCs) representation of the integrated datasets; the number of PCs were independently selected on each dataset as the elbow suggested in variance plot of PCs.

Single cell gene expression analysis

Clustering and cell-type annotation

The identification of cell clusters based on gene expression data was performed in the PCs representation of integrated datasets integrated datasets with the *FindClusters* function from Seurat using a resolution parameter of 0.5. Clusters were annotated by the examination of expression of canonical Treg or Tconv markers together with the inspection of markers defining each cluster and enriched GO Biological Process Terms (Supplementary text 1-4). Additional exclusion of artifactual populations was performed after clustering: cells lacking expression of *Cd3d*, *Cd3e*, and *Cd3g*, invariant NKT (iNKT) cells, and clusters highly expressing Hemoglobin genes (*Hba-a1, Hba-a2, Hbb-bt, Hbb-bs*).

DEG and gene set enrichment analysis

Identification of cluster marker genes was performed with *FindAllMarkers* function in Seurat, and significant markers were considered if fold change (FC) ≥1.2 and adjusted p value ≤0.05. When the comparison of two populations was required, DEGs were identified with the *FindMarkers* function with the min.pct parameter set at 0.25. Genes with adjusted p value ≤0.05 where considered significantly changed and were classified as upregulated if FC ≥1.5 or downregulated if FC ≤0.67.

Gene set enrichment analysis was performed with enrichR (v3.1) package (Chen et al., 2013) using the *GO_Biological_Process_2021* database as reference.

Trajectory analysis

Trajectory analyses were performed by using the Slingshot Bioconductor package (v2.10.0) with the PCs representation of integrated datasets and their respective cluster labels as inputs. We set the parameter *start.clus='naive'* to use the cluster of Naive cells as the starting point of the inferred trajectories.

Module score calculation

Calculation of expression levels from selected ZAC Tconv programs in Spleen datasets was performed by the *AddModuleScore* function from Seurat using the specified gene lists (Table S8) and default parameters.

Cell-cell crosstalk prediction

We used Cellphone DB (Vento-Tormo et al., 2018; Efremova et al., 2020) to predict cell-cell communication. After converting mouse gene IDs to human orthologues, we ran CellPhoneDB v4.0.0 in statistical analysis mode, then, results were visualized by the ktplots R package (Troulé et al., 2023). We manually searched for specific interactions reported in literature to be important for Treg function (Schmidt et al., 2012; Jiang et al., 2021; Dikiy et al., 2023): CTLA4, LAG3, TGFB, GZMB, IL10, GITR, TIGIT, ICOS, PD1, TRAIL, and DR5. In addition, we searched for coinhibitory and costimulatory interactions by using the *gene_family* parameter in ktplots. Only statistically significant interactions with a reasonable biological meaning are shown in detail in the results.

Single TCR V(D)J sequencing analysis

Clonal expansion, diversity and repertoire overlap

Basic repertoire analysis was performed with Immunarch (v0.9.0) R package (ImmunoMindTeam, 2019) and included repertoire diversity as evaluated by Inverse Simpson's index and paired CDR3 aa overlap calculated with the Morisita-Horn index. The Morisita-Horn index is a typical metric used in ecology to calculate the similitude of two ecosystems in terms of species distribution and abundance, and it has been introduced as a metric for repertoire overlap analysis; its values go from 0 to1, where 1 represents a perfect overlap (Chiffelle et al., 2020). Repertoire overlap comparisons were designed in several different ways depending on the purpose of analysis: 1) for comparison of repertoire similitude in ZAC Tconv and Treg gene expression clusters separately; 2) for examination of ZAP-70 mutant Tconv repertoires with WT Treg repertoire; and 3) for inspection of intra Treg-Tconv repertoire similitude in each subject.

Clonal expansion was evaluated in two ways: 1) by the simple examination of clone counts in a specific repertoire 2) by the calculation of clonal proportions. Clonal proportion calculation was used to classify clones by their magnitude of expansion, and indicates the proportion of a repertoire represented by each clone in each replicate. Clones were classified as: small (clonal proportion $<1x10^{-3}$), medium (clonal proportion from $1x10^{-3} – 1x10^{-2}$) and large (clonal proportion $>1x10^{-2}$). Given that only a small sample size could be recovered for some replicates, which inflates clonal proportions of lowly abundant clones, we labeled singletons (clone size=1) separately. Gene usage frequencies were calculated considering unique clonotypes and not taking its counts into account (unwheighted).

Clonal publicity

To examine the presence of public clones, we identified TCRs sharing paired gene usage and CDR3 amino acid sequences; and publicity was considered when the same clone existed in more than one subject.

TCR distances, repertoire correlation and identification of motifs

In order to evaluate if the overall features of selected repertoires show convergence, we calculated receptor distances with TCRdist. TCRdist allows for the calculation of a distance measure between TCRs which considers pMHC binding structural information (Dash et al., 2017), then, a neighbor distance metric consisting of the weighted average of the distance of a given TCR to the nearest 25th percentile the other TCRs in the repertoire can be calculated and the distribution of this metric was plotted and used to examine repertoire convergence of public and Th17 clones in ZAC mice. CDR3 motifs were analyzed by TCRdist algorithm which looks for sequence patterns that occur more often in the epitope-specific datasets than in a background set of sequences and measures significance by a chi-square test. Expected is judged based on two populations of background TCR sequences arranged to have same V and J gene frequencies as the epitope specific set. One background set is generated by a random model of TCR rearrangement and the other one is drawn from next generation sequencing (NGS) data. The value of expected used to calculate significance is the maximum

of the expected values based on these two background sets. Finally, the repertoire-repertoire correlation metric from TCRdist algorithm was computed for assessment of intra Treg-Tconv repertoire similitude.

Paratope clustering

The residues that interact with antigens are known as paratopes, and residues are majorly located in or near three complementarity-determining regions (CDRs). We named the amino acid of CDR1-3 as paratope sequence. We implemented the paratope clustering using Interclone method (Wilamowski et al., 2022). The source code was deposited in https://gitlab.com/sysimm/interclone and the clustreing parameter using the default setting (Sequence identity ≥ 80% and Coverage ≥ 90%). The above method was utilized to cluster the WT, SKG and ZAC TCR repertoire. Next, Morisita-Horn overlap indices were computed across WT, SKG and ZAC Tconv and Treg repertoires vs WT Treg or WT Tconv repertoires, considering paratope clusters as receptor identities.

TiRP score calculation

Lagattuta et al proposed a Treg-propensity scoring system for the TCR (TiRP) method to quantify the likelihood of Treg cells based on the TCR feature on Treg fate. (Lagattuta et al., 2022). The source code was obtained from https://github.com/immunogenomics/TiRP. The TiRP score of this study was calculated using the default parameter. For each TCR sequence, the v-gene and amino acid of CDR3 were utilized as input information.

Published sequence preparation

The published mouse Treg and Tconv TCR beta chain sequence data were obtained from four studies (Logunova et al., 2020; Lu et al., 2020; Ko et al., 2020; Wolf et al., 2016). The detailed accession ID and references were summarized in Table S16. The annotation of raw sequences was performed by mixcr (3.0.13) (Bolotin et al., 2015) with align, assemble, assembleContigs, and exportClones process. The sequences that have complete sequences of frameworks and CDRs regions were utilized in this study. In summary, we collected 16889

Tconv and 13715 Treg TCRs from Lu et al, 5203 Tconv and 3802 Treg TCRs from Ko et al, 375190 Tconv and 127179 Treg TCRs from Logunova et al, and 9743 Tconv and 4219 Treg from Wolf et al.

Sequence searching and comparison

We implemented the TCR sequence searching using Interclone method (Wilamowski et al., 2022). The source code was deposited in https://gitlab.com/sysimm/interclone and the searching parameter using the default setting. In brief, InterClone regards the CDR1,2,3 of one TRB seq as a pseudo sequence and considers one pair of hitting using the following CDR similarity and coverage thresholds: CDR1 ≥ 90%, CDR2 ≥ 90%, CDR3 ≥ 80%, and Coverage ≥ 90%. We first searched Tconv and Treg TCR templates from the same paper using identical query inputs. Then we compare the corresponding hitting number with Tconv and Treg templates for each query and calculate the significant Tconv or Treg-biased query sequence by the hypergeom distribution.

Figure legends

Figure 1. Two transcriptionally different groups of Th17 cells exist in inflamed joints of ZAC mice. a) UMAP plot of Tconv cells obtained from joints, draining lymph node (dLN) and spleen of arthritic ZAC mice. Unsupervised clustering was applied by the Louvain algorithm and clusters were annotated as described in the text. **b)** Relative expression of canonical Tconv markers across gene expression clusters. Circle color and intensity represents average gene expression whereas circle size represents the percentage of cells from each cluster expressing the gene. **c)** Expression of top 5 gene markers defining each cluster; clusters were subsampled to 100 cells for visualization. **d)** Distribution of gene expression clusters across sampled tissues. **e)** Differentially expressed genes (DEG) among cluster 1 (Th17 *Il22*+) and cluster 4 (Th17 *Tnfsf11*+). **f)** Gene Ontology (GO) Terms enriched in DEGs from cluster 1 (Th17 *Il22*+) and cluster 4 (Th17 *Tnfsf11*+), the intensity of the bar colors represents the odds ratio (OR) of enrichment. **g)** UMAP plot of Tconv cells from tissues of arthritic ZAC mice, cells are colored according to their expression of *Il22* (Upper panel) or *Tnfsf11* (Bottom panel). **h)** UMAP plot of Tconv cells from tissues of arthritic ZAC mice, cells are colored according to their inferred pseudotime values and clusters are connected according to the trajectory predicted by Slingshot.

Figure 2. TCR repertoires of Th17 *Il22*+ and Th17 *Tnfsf11*+ cells are substantially different, but contain convergent public clones. a) UMAP plot of Tconv cells from tissues of arthritic ZAC mice, cells are colored according to clonal proportion classes. **b)** Distribution of clonal proportion classes across gene expression clusters of ZAC Tconv cells. **c)** Morisita-Horn index across ZAC Tconv gene expression clusters (y axis); each box shows median overlap indices of each cluster (x axis) calculated to a reference repertoire (plot subtitles). **d) d)** Public clones and their clonal proportion classification in each mouse replicate. **e)** TRBV (green), TRBJ (blue), TRAV (purple) and TRAJ (red) gene usage preferences in public clones. **f)** Distribution of neighbor distances to the nearest 25th percentile (nbr.25; methods) in the repertoires of Naïve, Th17 *Il22*+, Th17 *Tnfsf11*+ and public clones (y axis). Paired

(left) and single alpha (middle) and beta (right) TCR chain distributions are shown. **g)** CDR3 motif in TCRβ of public clones, the color bars between the two logos indicate the inferred nucleotide source for the corresponding positions: light gray for V region; black for D region; dark gray for J region; and red for N insertions. Motif amino acid key: S=S; v=VILM; N=N; .=any amino acid.

Figure 3. Inflamed joints of ZAC mice are enriched in Trm and Intermediate Treg cells. **a)** UMAP plot of Treg cells obtained from joints, draining lymph node (dLN) and spleen of arthritic ZAC mice. Unsupervised clustering was applied by the Louvain algorithm and clusters were annotated as described in the text. **b)** Relative expression of canonical Treg markers across gene expression clusters. Circle color and intensity represents average gene expression whereas circle size represents the percentage of cells from each cluster expressing the gene. **c)** Expression of top 5 gene markers defining each cluster; clusters were subsampled to 100 cells for visualization. **d)** Distribution of gene expression clusters across sampled tissues. **e)** UMAP plot of Treg cells from tissues of arthritic ZAC mice, cells are colored according to clonal proportion classes. **f)** Distribution of clonal proportion classes across gene expression clusters of ZAC Treg cells. **g)** Morisita-Horn index across ZAC Treg gene expression clusters (y axis); each box shows median overlap indices of each cluster (x axis) calculated to a reference repertoire (plot subtitles). **h)** Circos plot of predicted ligand-receptor interactions between ZAC Tconv and ZAC Treg cell clusters. The direction of the interaction is represented by the arrow point in the plot and goes from Tconv (provide the ligand) to Treg (provide the receptor) cells.

Figure 4. ZAC Tconv repertoires have an increased similarity to WT Treg repertoires. Morisita-Horn index (y axis) of WT, SKG and ZAC Tconv and Treg repertoires with WT Tconv (magenta bars) or WT Treg (blue bars) repertoires used as reference: **(a)** calculated at the exact CDR3 amino acid level and are shown for paired (left) and single alpha (middle) and beta (right) chains; (**b)** calculated by considering paratope TRB clusters as receptor

identities. **c)** Fraction of WT, SKG and ZAC Tconv and Treg repertoires hitting with reference WT Tconv (magenta section) or WT Treg repertoires (blue section) obtained from Logunova et al., 2020. **d)** Fold change of the ratio of fractional repertoire hitting with public WT Tregs vs the fractional repertoire hitting with public WT Tconv fraction for SKG and ZAC repertoires, relative to the same ratio for WT repertoires. The results of the four public datasets used as reference are shown (Logunova et al., 2020; Lu et al., 2020; Ko et al., 2020; Wolf et al., 2016). **e)** Fraction of ZAC Naïve, Intermediate-2, Th17 *Il22*+ and Th17 *Tnfsf11*+ repertoires hitting with reference WT Tconv (magenta section) or WT Treg (blue section) repertoires obtained from Logunova et al., 2020. **f)** Fold change of the ratio of fractional repertoire hitting with public WT Tregs vs the fractional repertoire hitting with public WT Tconv fraction for ZAC Naïve, Intermediate-2, Th17 *Il22*+ and Th17 *Tnfsf11*+ repertoire, relative to the same ratio for WT repertoires. The results of the four public datasets used as reference are shown (Logunova et al., 2020; Lu et al., 2020; Ko et al., 2020; Wolf et al., 2016). **g)** Left: distribution of TiRP scores for Tconv and Treg repertoires of WT, SKG and ZAC mice; Right: Median TiRP score change of Tconv and Treg repertoires of WT, SKG and ZAC mice relative to WT Tconv. **h)** Left: distribution of TiRP scores for WT Tconv, ZAC Th17 *Il22*+ and ZAC Th17 *Tnfsf11*+ repertoires; Right: Median TiRP score change of ZAC Th17 *Il22*+ and ZAC Th17 *Tnfsf11*+ repertoires relative to WT Tconv.

Figure 5. Intra-Treg-Tconv similitude in WT, SKG and ZAC repertoires. Morisita-Horn index of intra Tconv and Treg overlap in WT, SKG and ZAC repertoires: **(a)** calculated at the exact CDR3 amino acid level and are shown for paired (left) and single alpha (middle) and beta (right) chains; (**c)** calculated by considering paratope TRA and TRB clusters as receptor identities. **b)** Morisita-Horn index of intra Tconv and Treg overlap in WT, SKG and ZAC repertoires split by Tconv phenotype. **d)** repertoire-repertoire neighbor distance (nbrdist) scores correlations for WT, SKG and ZAC Tconv and Treg. Color scale ranges from blue (low correlation) to yellow (high correlation), and higher correlation scores indicate closer repertoires. Representative clonotype tracking plots of intra-Treg-Tconv shared clones across Treg and Tconv gene expression clusters for WT **(e)**, SKG **(f)** and ZAC **(g)** mice.

References

Akkaya, B., Oya, Y., Akkaya, M., Al Souz, J., Holstein, A. H., Kamenyeva, O., Kabat, J., Matsumura, R., Dorward, D. W., Glass, D. D., & Shevach, E. M. (2019). Regulatory T cells mediate specific suppression by depleting peptide-MHC class II from dendritic cells. *Nature immunology*, 20(2), 218–231. https://doi.org/10.1038/s41590-018-0280-2

Al Khabouri, S., Benson, R. A., Prendergast, C. T., Gray, J. I., Otto, T. D., Brewer, J. M., & Garside, P. (2021). TCRβ Sequencing Reveals Spatial and Temporal Evolution of Clonal CD4 T Cell Responses in a Breach of Tolerance Model of Inflammatory Arthritis. *Frontiers in immunology*, *12*, 669856. https://doi.org/10.3389/fimmu.2021.669856

Argyriou, A., Wadsworth, M. H., 2nd, Lendvai, A., Christensen, S. M., Hensvold, A. H., Gerstner, C., van Vollenhoven, A., Kravarik, K., Winkler, A., Malmström, V., & Chemin, K. (2022). Single cell sequencing identifies clonally expanded synovial $CD4^+$ T_{PH} cells expressing GPR56 in rheumatoid arthritis. *Nature communications*, *13*(1), 4046. https://doi.org/10.1038/s41467-022-31519-6

Aschenbrenner, K., D'Cruz, L. M., Vollmann, E. H., Hinterberger, M., Emmerich, J., Swee, L. K., Rolink, A., & Klein, L. (2007). Selection of Foxp3+ regulatory T cells specific for self antigen expressed and presented by Aire+ medullary thymic epithelial cells. *Nature immunology*, 8(4), 351–358. https://doi.org/10.1038/ni1444

Ashby, K. M., & Hogquist, K. A. (2024). A guide to thymic selection of T cells. *Nature reviews. Immunology*, *24*(2), 103–117. https://doi.org/10.1038/s41577-023-00911-8

Bolotin, D. A., Poslavsky, S., Mitrophanov, I., Shugay, M., Mamedov, I. Z., Putintseva, E. V., & Chudakov, D. M. (2015). MiXCR: software for comprehensive adaptive immunity profiling. *Nature methods*, 12(5), 380–381. https://doi.org/10.1038/nmeth.3364

Chen, E. Y., Tan, C. M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G. V., Clark, N. R., & Ma'ayan, A. (2013). Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC bioinformatics*, 14, 128. https://doi.org/10.1186/1471-2105-14-128

Chiffelle, J., Genolet, R., Perez, M. A., Coukos, G., Zoete, V., & Harari, A. (2020). T-cell repertoire analysis and metrics of diversity and clonality. *Current opinion in biotechnology*, 65, 284–295. https://doi.org/10.1016/j.copbio.2020.07.010

Collison, L. W., Pillai, M. R., Chaturvedi, V., & Vignali, D. A. (2009). Regulatory T cell suppression is potentiated by target T cells in a cell contact, IL-35- and IL-10-dependent

manner. *Journal of immunology (Baltimore, Md. : 1950)*, *182*(10), 6121–6128. https://doi.org/10.4049/jimmunol.0803646

Corneth, O. B., Mus, A. M., Asmawidjaja, P. S., Klein Wolterink, R. G., van Nimwegen, M., Brem, M. D., Hofman, Y., Hendriks, R. W., & Lubberts, E. (2014). Absence of interleukin-17 receptor a signaling prevents autoimmune inflammation of the joint and leads to a Th2-like phenotype in collagen-induced arthritis. *Arthritis & rheumatology (Hoboken, N.J.)*, *66*(2), 340–349. https://doi.org/10.1002/art.38229

Croft, M., So, T., Duan, W., & Soroosh, P. (2009). The significance of OX40 and OX40L to T-cell biology and immune disease. *Immunological reviews*, *229*(1), 173–191. https://doi.org/10.1111/j.1600-065X.2009.00766.x

Dash, P., Fiore-Gartland, A. J., Hertz, T., Wang, G. C., Sharma, S., Souquette, A., Crawford, J. C., Clemens, E. B., Nguyen, T. H. O., Kedzierska, K., La Gruta, N. L., Bradley, P., & Thomas, P. G. (2017). Quantifiable predictive features define epitope-specific T cell receptor repertoires. *Nature*, 547(7661), 89–93. https://doi.org/10.1038/nature22383

DiToro, D., Harbour, S. N., Bando, J. K., Benavides, G., Witte, S., Laufer, V. A., Moseley, C., Singer, J. R., Frey, B., Turner, H., Bruning, J., Darley-Usmar, V., Gao, M., Conover, C., Hatton, R. D., Frank, S., Colonna, M., & Weaver, C. T. (2020). Insulin-Like Growth Factors Are Key Regulators of T Helper 17 Regulatory T Cell Balance in Autoimmunity. *Immunity*, *52*(4), 650–667.e10. https://doi.org/10.1016/j.immuni.2020.03.013

Dikiy, S., & Rudensky, A. Y. (2023). Principles of regulatory T cell function. *Immunity*, 56(2), 240–255. https://doi.org/10.1016/j.immuni.2023.01.004

Efremova, M., Vento-Tormo, M., Teichmann, S. A., & Vento-Tormo, R. (2020). CellPhoneDB: inferring cell-cell communication from combined expression of multi-subunit ligand-receptor complexes. *Nature protocols*, 15(4), 1484–1506. https://doi.org/10.1038/s41596-020-0292-x

Fu, Y., Lin, Q., Zhang, Z., & Zhang, L. (2020). Therapeutic strategies for the costimulatory molecule OX40 in T-cell-mediated immunity. *Acta pharmaceutica Sinica. B*, *10*(3), 414–433. https://doi.org/10.1016/j.apsb.2019.08.010

Glal, D., Sudhakar, J. N., Lu, H. H., Liu, M. C., Chiang, H. Y., Liu, Y. C., Cheng, C. F., & Shui, J. W. (2018). ATF3 Sustains IL-22-Induced STAT3 Phosphorylation to Maintain

Mucosal Immunity Through Inhibiting Phosphatases. *Frontiers in immunology*, *9*, 2522. https://doi.org/10.3389/fimmu.2018.02522

Gaublomme, J. T., Yosef, N., Lee, Y., Gertner, R. S., Yang, L. V., Wu, C., Pandolfi, P. P., Mak, T., Satija, R., Shalek, A. K., Kuchroo, V. K., Park, H., & Regev, A. (2015). Single-Cell Genomics Unveils Critical Regulators of Th17 Cell Pathogenicity. *Cell*, *163*(6), 1400–1412. https://doi.org/10.1016/j.cell.2015.11.009

Goodnow, C. C., Sprent, J., Fazekas de St Groth, B., & Vinuesa, C. G. (2005). Cellular and genetic mechanisms of self tolerance and autoimmunity. *Nature*, *435*(7042), 590–597. https://doi.org/10.1038/nature03724

Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W. M., 3rd, Zheng, S., Butler, A., Lee, M. J., Wilk, A. J., Darby, C., Zager, M., Hoffman, P., Stoeckius, M., Papalexi, E., Mimitou, E. P., Jain, J., Srivastava, A., Stuart, T., Fleming, L. M., Yeung, B., Rogers, A. J., … Satija, R. (2021). Integrated analysis of multimodal single-cell data. *Cell*, 184(13), 3573–3587.e29. https://doi.org/10.1016/j.cell.2021.04.048

Hsieh, C. S., Liang, Y., Tyznik, A. J., Self, S. G., Liggitt, D., & Rudensky, A. Y. (2004). Recognition of the peripheral self by naturally arising CD25+ CD4+ T cell receptors. *Immunity*, *21*(2), 267–277. https://doi.org/10.1016/j.immuni.2004.07.009

Hsieh, C. S., Zheng, Y., Liang, Y., Fontenot, J. D., & Rudensky, A. Y. (2006). An intersection between the self-reactive regulatory and nonregulatory T cell receptor repertoires. *Nature immunology*, *7*(4), 401–410. https://doi.org/10.1038/ni1318

ImmunoMind Team. (2019). immunarch: An R Package for Painless Bioinformatics Analysis of T-Cell and B-Cell Immune Repertoires.
Zenodo. http://doi.org/10.5281/zenodo.3367200

Jha, P., & Das, H. (2017). KLF2 in Regulation of NF-κB-Mediated Immune Cell Function and Inflammation. *International journal of molecular sciences*, *18*(11), 2383. https://doi.org/10.3390/ijms18112383

Jiang, Q., Yang, G., Liu, Q., Wang, S., & Cui, D. (2021). Function and Role of Regulatory T Cells in Rheumatoid Arthritis. *Frontiers in immunology*, 12, 626193. https://doi.org/10.3389/fimmu.2021.626193

Jordan, M. S., Boesteanu, A., Reed, A. J., Petrone, A. L., Holenbeck, A. E., Lerman, M. A., Naji, A., & Caton, A. J. (2001). Thymic selection of CD4+CD25+ regulatory T cells

induced by an agonist self-peptide. *Nature immunology*, 2(4), 301–306. https://doi.org/10.1038/86302

Kieback, E., Hilgenberg, E., Stervbo, U., Lampropoulou, V., Shen, P., Bunse, M., Jaimes, Y., Boudinot, P., Radbruch, A., Klemm, U., Kühl, A. A., Liblau, R., Hoevelmeyer, N., Anderton, S. M., Uckert, W., & Fillatreau, S. (2016). Thymus-Derived Regulatory T Cells Are Positively Selected on Natural Self-Antigen through Cognate Interactions of High Functional Avidity. *Immunity*, *44*(5), 1114–1126. https://doi.org/10.1016/j.immuni.2016.04.018

Klarenbeek, P. L., de Hair, M. J., Doorenspleet, M. E., van Schaik, B. D., Esveldt, R. E., van de Sande, M. G., Cantaert, T., Gerlag, D. M., Baeten, D., van Kampen, A. H., Baas, F., Tak, P. P., & de Vries, N. (2012). Inflamed target tissue provides a specific niche for highly expanded T-cell clones in early human autoimmune disease. *Annals of the rheumatic diseases*, *71*(6), 1088–1093. https://doi.org/10.1136/annrheumdis-2011-200612

Ko, A., Watanabe, M., Nguyen, T., Shi, A., Achour, A., Zhang, B., Sun, X., Wang, Q., Zhuang, Y., Weng, N. P., & Hodes, R. J. (2020). TCR Repertoires of Thymic Conventional and Regulatory T Cells: Identification and Characterization of Both Unique and Shared TCR Sequences. *Journal of immunology (Baltimore, Md. : 1950)*, *204*(4), 858–867. https://doi.org/10.4049/jimmunol.1901006

Komatsu, N., Okamoto, K., Sawa, S., Nakashima, T., Oh-hora, M., Kodama, T., Tanaka, S., Bluestone, J. A., & Takayanagi, H. (2014). Pathogenic conversion of Foxp3+ T cells into TH17 cells in autoimmune arthritis. *Nature medicine*, *20*(1), 62–68. https://doi.org/10.1038/nm.3432

Komatsu, N., & Takayanagi, H. (2022). Mechanisms of joint destruction in rheumatoid arthritis - immune cell-fibroblast-bone interactions. *Nature reviews. Rheumatology*, *18*(7), 415–429. https://doi.org/10.1038/s41584-022-00793-5

Kronenberg, M., & Rudensky, A. (2005). Regulation of immunity by self-reactive T cells. *Nature*, *435*(7042), 598–604. https://doi.org/10.1038/nature03725

Lagattuta, K. A., Kang, J. B., Nathan, A., Pauken, K. E., Jonsson, A. H., Rao, D. A., Sharpe, A. H., Ishigaki, K., & Raychaudhuri, S. (2022). Repertoire analyses reveal T cell antigen receptor sequence features that influence T cell fate. *Nature immunology*, *23*(3), 446–457. https://doi.org/10.1038/s41590-022-01129-x

Lee, H. M., Bautista, J. L., Scott-Browne, J., Mohan, J. F., & Hsieh, C. S. (2012a). A broad range of self-reactivity drives thymic regulatory T cell selection to limit responses to self. *Immunity*, *37*(3), 475–486. https://doi.org/10.1016/j.immuni.2012.07.009

Lee, Y., Awasthi, A., Yosef, N., Quintana, F. J., Xiao, S., Peters, A., Wu, C., Kleinewietfeld, M., Kunder, S., Hafler, D. A., Sobel, R. A., Regev, A., & Kuchroo, V. K. (2012b). Induction and molecular signature of pathogenic TH17 cells. *Nature immunology*, *13*(10), 991–999. https://doi.org/10.1038/ni.2416

Lee, J. Y., Skon, C. N., Lee, Y. J., Oh, S., Taylor, J. J., Malhotra, D., Jenkins, M. K., Rosenfeld, M. G., Hogquist, K. A., & Jameson, S. C. (2015). The transcription factor KLF2 restrains $CD4^+$ T follicular helper cell differentiation. *Immunity*, 42(2), 252–264. https://doi.org/10.1016/j.immuni.2015.01.013

Leonard, J. D., Gilmore, D. C., Dileepan, T., Nawrocka, W. I., Chao, J. L., Schoenbach, M. H., Jenkins, M. K., Adams, E. J., & Savage, P. A. (2017). Identification of Natural Regulatory T Cell Epitopes Reveals Convergence on a Dominant Autoantigen. *Immunity*, *47*(1), 107–117.e8. https://doi.org/10.1016/j.immuni.2017.06.015

Levine, A. G., Arvey, A., Jin, W., & Rudensky, A. Y. (2014). Continuous requirement for the TCR in regulatory T cell function. Nature immunology, 15(11), 1070–1078. https://doi.org/10.1038/ni.3004

Li, M. O., & Rudensky, A. Y. (2016). T cell receptor signalling in the control of regulatory T cell differentiation and function. Nature reviews. Immunology, 16(4), 220–233. https://doi.org/10.1038/nri.2016.26

Lin, C. C., Bradstreet, T. R., Schwarzkopf, E. A., Jarjour, N. N., Chou, C., Archambault, A. S., Sim, J., Zinselmeyer, B. H., Carrero, J. A., Wu, G. F., Taneja, R., Artyomov, M. N., Russell, J. H., & Edelson, B. T. (2016). IL-1-induced Bhlhe40 identifies pathogenic T helper cells in a model of autoimmune neuroinflammation. *The Journal of experimental medicine*, *213*(2), 251–271. https://doi.org/10.1084/jem.20150568

Liu, Z., Gerner, M. Y., Van Panhuys, N., Levine, A. G., Rudensky, A. Y., & Germain, R. N. (2015). Immune homeostasis enforced by co-localized effector and regulatory T cells. *Nature*, *528*(7581), 225–230. https://doi.org/10.1038/nature16169

Logunova, N. N., Kriukova, V. V., Shelyakin, P. V., Egorov, E. S., Pereverzeva, A., Bozhanova, N. G., Shugay, M., Shcherbinin, D. S., Pogorelyy, M. V., Merzlyak, E. M., Zubov, V. N., Meiler, J., Chudakov, D. M., Apt, A. S., & Britanova, O. V. (2020). MHC-II alleles shape the CDR3 repertoires of conventional and regulatory naïve CD4+ T cells. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(24), 13659–13669. https://doi.org/10.1073/pnas.2003170117

Loza, M., Teraguchi, S., Standley, D. M., & Diez, D. (2022). Unbiased integration of single cell transcriptome replicates. NAR genomics and bioinformatics, 4(1), lqac022. https://doi.org/10.1093/nargab/lqac022

Lu, D. R., Wu, H., Driver, I., Ingersoll, S., Sohn, S., Wang, S., Li, C. M., & Phee, H. (2020). Dynamic changes in the regulatory T-cell heterogeneity and function by murine IL-2 mutein. *Life science alliance*, *3*(5), e201900520. https://doi.org/10.26508/lsa.201900520

Lu, C., Pi, X., Xu, W., Qing, P., Tang, H., Li, Y., Zhao, Y., Liu, X., Tang, H., & Liu, Y. (2022). Clinical significance of T cell receptor repertoire in primary Sjogren's syndrome. *EBioMedicine*, *84*, 104252. https://doi.org/10.1016/j.ebiom.2022.104252

Lubberts, E., Joosten, L. A., van de Loo, F. A., Schwarzenberger, P., Kolls, J., & van den Berg, W. B. (2002). Overexpression of IL-17 in the knee joint of collagen type II immunized mice promotes collagen arthritis and aggravates joint destruction. *Inflammation research : official journal of the European Histamine Research Society ... [et al.]*, *51*(2), 102–104. https://doi.org/10.1007/BF02684010

Malchow, S., Leventhal, D. S., Lee, V., Nishi, S., Socci, N. D., & Savage, P. A. (2016). Aire Enforces Immune Tolerance by Directing Autoreactive T Cells into the Regulatory T Cell Lineage. *Immunity*, *44*(5), 1102–1113. https://doi.org/10.1016/j.immuni.2016.02.009

Mandl, J. N., Monteiro, J. P., Vrisekoop, N., & Germain, R. N. (2013). T cell-positive selection uses self-ligand binding strength to optimize repertoire recognition of foreign antigens. *Immunity*, *38*(2), 263–274. https://doi.org/10.1016/j.immuni.2012.09.011

Menezes, J. S., van den Elzen, P., Thornes, J., Huffman, D., Droin, N. M., Maverakis, E., & Sercarz, E. E. (2007). A public T cell clonotype within a heterogeneous autoreactive repertoire is dominant in driving EAE. *The Journal of clinical investigation*, *117*(8), 2176–2185. https://doi.org/10.1172/JCI28277

Meyer Zu Horste, G., Wu, C., Wang, C., Cong, L., Pawlak, M., Lee, Y., Elyaman, W., Xiao, S., Regev, A., & Kuchroo, V. K. (2016). RBPJ Controls Development of Pathogenic Th17 Cells by Regulating IL-23 Receptor Expression. *Cell reports*, *16*(2), 392–404. https://doi.org/10.1016/j.celrep.2016.05.088

Nakae, S., Nambu, A., Sudo, K., & Iwakura, Y. (2003). Suppression of immune induction of collagen-induced arthritis in IL-17-deficient mice. *Journal of immunology (Baltimore, Md. : 1950)*, *171*(11), 6173–6177. https://doi.org/10.4049/jimmunol.171.11.6173

O' Reilly, L. A., Tai, L., Lee, L., Kruse, E. A., Grabow, S., Fairlie, W. D., Haynes, N. M., Tarlinton, D. M., Zhang, J. G., Belz, G. T., Smyth, M. J., Bouillet, P., Robb, L., & Strasser, A. (2009). Membrane-bound Fas ligand only is essential for Fas-induced apoptosis. *Nature*, *461*(7264), 659–663. https://doi.org/10.1038/nature08402

Ohkura, N., Hamaguchi, M., Morikawa, H., Sugimura, K., Tanaka, A., Ito, Y., Osaki, M., Tanaka, Y., Yamashita, R., Nakano, N., Huehn, J., Fehling, H. J., Sparwasser, T., Nakai, K., & Sakaguchi, S. (2012). T cell receptor stimulation-induced epigenetic changes and Foxp3 expression are independent and complementary events required for Treg cell development. *Immunity*, 37(5), 785–799. https://doi.org/10.1016/j.immuni.2012.09.010

Ohl, K., Nickel, H., Moncrieffe, H., Klemm, P., Scheufen, A., Föll, D., Wixler, V., Schippers, A., Wagner, N., Wedderburn, L. R., & Tenbrock, K. (2018). The transcription factor CREM drives an inflammatory phenotype of T cells in oligoarticular juvenile idiopathic arthritis. *Pediatric rheumatology online journal*, *16*(1), 39. https://doi.org/10.1186/s12969-018-0253-x

Ooi, J. D., Petersen, J., Tan, Y. H., Huynh, M., Willett, Z. J., Ramarathinam, S. H., Eggenhuizen, P. J., Loh, K. L., Watson, K. A., Gan, P. Y., Alikhan, M. A., Dudek, N. L., Handel, A., Hudson, B. G., Fugger, L., Power, D. A., Holt, S. G., Coates, P. T., Gregersen, J. W., Purcell, A. W., … Kitching, A. R. (2017). Dominant protection from HLA-linked autoimmunity by antigen-specific regulatory T cells. *Nature*, *545*(7653), 243–247. https://doi.org/10.1038/nature22329

Pacholczyk, R., Ignatowicz, H., Kraj, P., & Ignatowicz, L. (2006). Origin and T cell receptor diversity of Foxp3+CD4+CD25+ T cells. *Immunity*, 25(2), 249–259. https://doi.org/10.1016/j.immuni.2006.05.016

Persaud, S. P., Parker, C. R., Lo, W. L., Weber, K. S., & Allen, P. M. (2014). Intrinsic CD4+ T cell sensitivity and response to a pathogen are set and sustained by avidity for thymic and peripheral complexes of self peptide and MHC. *Nature immunology*, *15*(3), 266–274. https://doi.org/10.1038/ni.2822

Pisetsky D. S. (2023). Pathogenesis of autoimmune disease. *Nature reviews. Nephrology*, 19(8), 509–524. https://doi.org/10.1038/s41581-023-00720-1

Pohar, J., Simon, Q., & Fillatreau, S. (2018). Antigen-Specificity in the Thymic Development and Peripheral Activity of CD4+FOXP3+ T Regulatory Cells. *Frontiers in immunology*, *9*, 1701. https://doi.org/10.3389/fimmu.2018.01701

Richardson, E., Galson, J. D., Kellam, P., Kelly, D. F., Smith, S. E., Palser, A., Watson, S., & Deane, C. M. (2021). A computational method for immune repertoire mining that identifies novel binders from different clonotypes, demonstrated by identifying anti-pertussis toxoid antibodies. *mAbs*, *13*(1), 1869406. https://doi.org/10.1080/19420862.2020.1869406

Sainz-Perez, A., Lim, A., Lemercier, B., & Leclerc, C. (2012). The T-cell receptor repertoire of tumor-infiltrating regulatory T lymphocytes is skewed toward public sequences. *Cancer research*, 72(14), 3557–3569. https://doi.org/10.1158/0008-5472.CAN-12-0277

Sakaguchi, N., Takahashi, T., Hata, H., Nomura, T., Tagami, T., Yamazaki, S., Sakihama, T., Matsutani, T., Negishi, I., Nakatsuru, S., & Sakaguchi, S. (2003). Altered thymic T-cell selection due to a mutation of the ZAP-70 gene causes autoimmune arthritis in mice. *Nature*, *426*(6965), 454–460. https://doi.org/10.1038/nature02119

Saputri, D. S., Ismanto, H. S., Nugraha, D. K., Xu, Z., Horiguchi, Y., Sakakibara, S., & Standley, D. M. (2023). Deciphering the antigen specificities of antibodies by clustering their complementarity determining region sequences. *mSystems*, *8*(6), e0072223. https://doi.org/10.1128/msystems.00722-23

Satija, R., Farrell, J. A., Gennert, D., Schier, A. F., & Regev, A. (2015). Spatial reconstruction of single-cell gene expression data. *Nature biotechnology*, 33(5), 495–502. https://doi.org/10.1038/nbt.3192

Sato, K., Suematsu, A., Okamoto, K., Yamaguchi, A., Morishita, Y., Kadono, Y., Tanaka, S., Kodama, T., Akira, S., Iwakura, Y., Cua, D. J., & Takayanagi, H. (2006). Th17 functions as an osteoclastogenic helper T cell subset that links T cell activation and bone destruction. *The Journal of experimental medicine*, *203*(12), 2673–2682. https://doi.org/10.1084/jem.20061775

Schmidt, D., Goronzy, J. J., & Weyand, C. M. (1996). CD4+ CD7- CD28- T cells are expanded in rheumatoid arthritis and are characterized by autoreactivity. *The Journal of clinical investigation*, 97(9), 2027–2037. https://doi.org/10.1172/JCI118638

Schmidt, A., Oberle, N., & Krammer, P. H. (2012). Molecular mechanisms of treg-mediated T cell suppression. *Frontiers in immunology*, 3, 51. https://doi.org/10.3389/fimmu.2012.00051

Shugay, M., Bagaev, D. V., Zvyagin, I. V., Vroomans, R. M., Crawford, J. C., Dolton, G., Komech, E. A., Sycheva, A. L., Koneva, A. E., Egorov, E. S., Eliseev, A. V., Van Dyk, E., Dash, P., Attaf, M., Rius, C., Ladell, K., McLaren, J. E., Matthews, K. K., Clemens, E. B., Douek, D. C., … Chudakov, D. M. (2018). VDJdb: a curated database of T-cell receptor sequences with known antigen specificity. *Nucleic acids research*, *46*(D1), D419–D427. https://doi.org/10.1093/nar/gkx760

Sims, N. A., & Romas, E. (2015). Is RANKL inhibition both anti-resorptive and anabolic in rheumatoid arthritis?. *Arthritis research & therapy*, *17*, 328. https://doi.org/10.1186/s13075-015-0861-5

Stoeckius, M., Zheng, S., Houck-Loomis, B., Hao, S., Yeung, B. Z., Mauck, W. M., 3rd, Smibert, P., & Satija, R. (2018). Cell Hashing with barcoded antibodies enables multiplexing and doublet detection for single cell genomics. *Genome biology*, 19(1), 224. https://doi.org/10.1186/s13059-018-1603-1

Street, K., Risso, D., Fletcher, R. B., Das, D., Ngai, J., Yosef, N., Purdom, E., & Dudoit, S. (2018). Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *BMC genomics*, *19*(1), 477. https://doi.org/10.1186/s12864-018-4772-0

Tanaka, S., Maeda, S., Hashimoto, M., Fujimori, C., Ito, Y., Teradaira, S., Hirota, K., Yoshitomi, H., Katakai, T., Shimizu, A., Nomura, T., Sakaguchi, N., & Sakaguchi, S. (2010). Graded attenuation of TCR signaling elicits distinct autoimmune diseases by altering thymic T cell selection and regulatory T cell function. *Journal of immunology (Baltimore, Md. : 1950)*, *185*(4), 2295–2305. https://doi.org/10.4049/jimmunol.1000848

Tanaka, A., Maeda, S., Nomura, T., Llamas-Covarrubias, M. A., Tanaka, S., Jin, L., Lim, E. L., Morikawa, H., Kitagawa, Y., Akizuki, S., Ito, Y., Fujimori, C., Hirota, K., Murase, T., Hashimoto, M., Higo, J., Zamoyska, R., Ueda, R., Standley, D. M., Sakaguchi, N., … Sakaguchi, S. (2023). Construction of a T cell receptor signaling range for spontaneous development of autoimmune disease. *The Journal of experimental medicine*, *220*(2), e20220386. https://doi.org/10.1084/jem.20220386

Troulé, K., Petryszak, R., Prete, M., Cranley, J., Harasty, A., Tuong, Z. K., & Teichmann, S. A. (2023). CellPhoneDB v5: Inferring cell-cell communication from single-cell multiomics data. *ArXiv*. /abs/2311.04567. https://doi.org/10.48550/arXiv.2311.04567

Valkiers Sebastiaan, de Vrij Nicky, Gielis Sofie, Verbandt Sara, Ogunjimi Benson, Laukens Kris, Meysman Pieter. (2022). Recent advances in T-cell receptor repertoire analysis: Bridging the gap with multimodal single-cell RNA sequencing. *ImmunoInformatics*. 5(10009). https://doi.org/10.1016/j.immuno.2022.100009

Vento-Tormo, R., Efremova, M., Botting, R. A., Turco, M. Y., Vento-Tormo, M., Meyer, K. B., Park, J. E., Stephenson, E., Polański, K., Goncalves, A., Gardner, L., Holmqvist, S., Henriksson, J., Zou, A., Sharkey, A. M., Millar, B., Innes, B., Wood, L., Wilbrey-Clark, A., Payne, R. P., … Teichmann, S. A. (2018). Single-cell reconstruction of the early maternal-fetal interface in humans. *Nature*, *563*(7731), 347–353. https://doi.org/10.1038/s41586-018-0698-6

Vignali, D. A., Collison, L. W., & Workman, C. J. (2008). How regulatory T cells work. Nature reviews. Immunology, 8(7), 523–532. https://doi.org/10.1038/nri2343

Voisinne, G., Gonzalez de Peredo, A., & Roncagalli, R. (2018). CD5, an Undercover Regulator of TCR Signaling. *Frontiers in immunology*, 9, 2900. https://doi.org/10.3389/fimmu.2018.02900

Vrisekoop, N., Monteiro, J. P., Mandl, J. N., & Germain, R. N. (2014). Revisiting thymic positive selection and the mature T cell repertoire for antigen. *Immunity*, *41*(2), 181–190. https://doi.org/10.1016/j.immuni.2014.07.007

Wagner, U., Pierer, M., Kaltenhäuser, S., Wilke, B., Seidel, W., Arnold, S., & Häntzschel, H. (2003). Clonally expanded CD4+CD28null T cells in rheumatoid arthritis use distinct

combinations of T cell receptor BV and BJ elements. *European journal of immunology*, 33(1), 79–84. https://doi.org/10.1002/immu.200390010

Wilamowski Jan , Xu Zichang , Ismanto Hendra S , Li Songling , Teraguchi Shunsuke , Llamas- Covarrubias Mara Anais, Lu Xiuyuan , Yamasaki Sho , Standley Daron M. (2022). InterClone: Store, Search and Cluster Adaptive Immune Receptor Repertoires. *bioRxiv*. doi: https://doi.org/10.1101/2022.07.31.501809

Wolf, K. J., Emerson, R. O., Pingel, J., Buller, R. M., & DiPaolo, R. J. (2016). Conventional and Regulatory CD4+ T Cells That Share Identical TCRs Are Derived from Common Clones. *PloS one*, *11*(4), e0153705. https://doi.org/10.1371/journal.pone.0153705

Wyss, L., Stadinski, B. D., King, C. G., Schallenberg, S., McCarthy, N. I., Lee, J. Y., Kretschmer, K., Terracciano, L. M., Anderson, G., Surh, C. D., Huseby, E. S., & Palmer, E. (2016). Affinity for self antigen selects Treg cells with distinct functional properties. *Nature immunology*, 17(9), 1093–1101. https://doi.org/10.1038/ni.3522

Yi, J., Miller, A. T., Archambault, A. S., Jones, A. J., Bradstreet, T. R., Bandla, S., Hsu, Y. S., Edelson, B. T., Zhou, Y. W., Fremont, D. H., Egawa, T., Singh, N., Wu, G. F., & Hsieh, C. S. (2022). Antigen-specific depletion of $CD4^+$ T cells by CAR T cells reveals distinct roles of higher- and lower-affinity TCRs during autoimmunity. *Science immunology*, *7*(76), eabo0777. https://doi.org/10.1126/sciimmunol.abo0777

Zemmour, D., Zilionis, R., Kiner, E., Klein, A. M., Mathis, D., & Benoist, C. (2018). Single-cell gene expression reveals a landscape of regulatory T cell phenotypes shaped by the TCR. *Nature immunology*, *19*(3), 291–301. https://doi.org/10.1038/s41590-018-0051-0

Zhao, Y., Nguyen, P., Vogel, P., Li, B., Jones, L. L., & Geiger, T. L. (2016). Autoimmune susceptibility imposed by public TCRβ chains. *Scientific reports*, *6*, 37543. https://doi.org/10.1038/srep37543

Figure 1

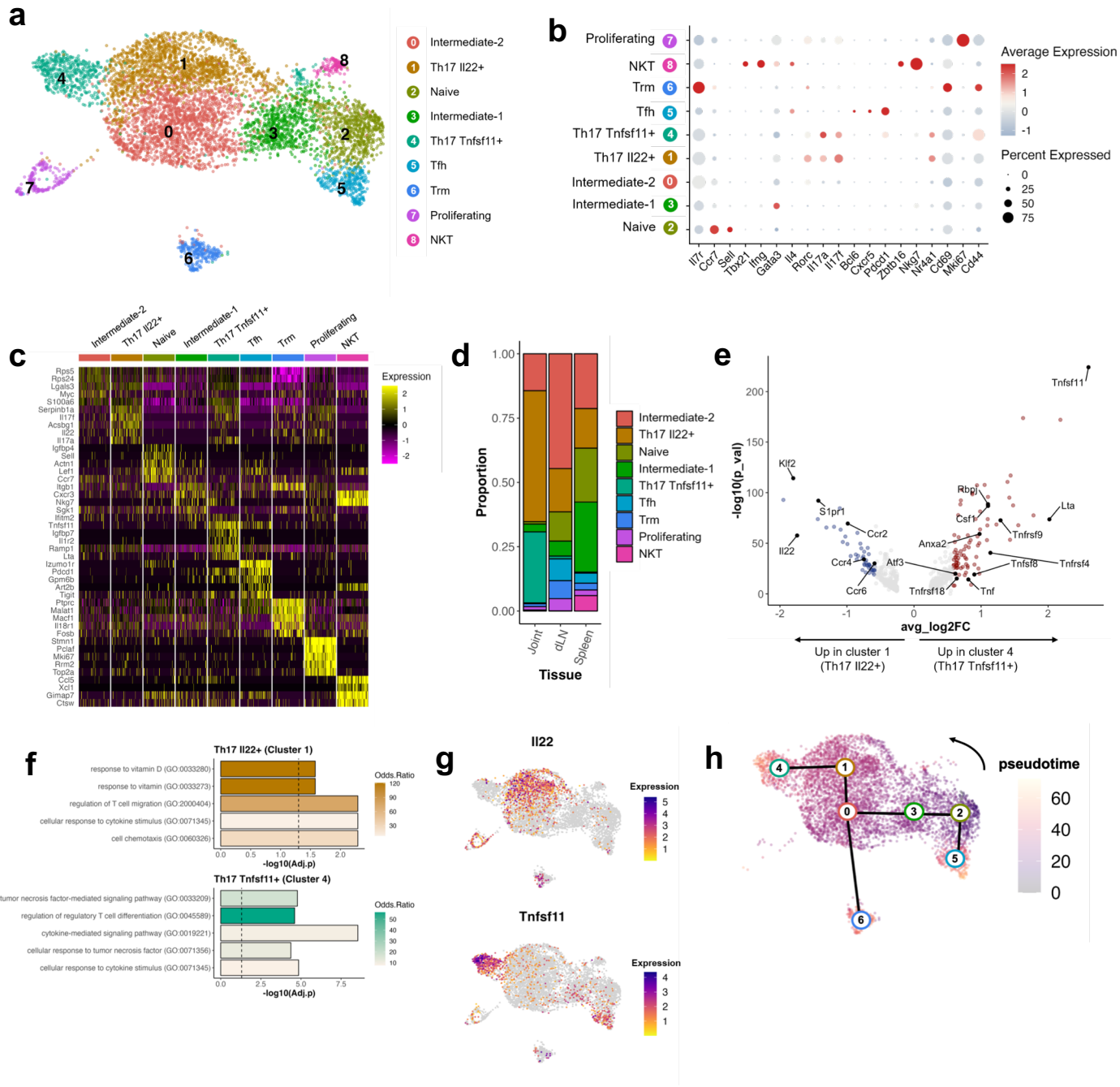

a
0 Intermediate-2
1 Th17 Il22+
2 Naive
3 Intermediate-1
4 Th17 Tnfsf11+
5 Tfh
6 Trm
7 Proliferating
8 NKT
b
Proliferating
NKT
Trm
Tfh
Th17 Tnfsf11+
Th17 Il22+
Intermediate-2
Intermediate-1
Naive
Il7r Ccr7 Sell Tbx21 Ifng Gata3 Il4 Rorc Il17a Il17f Bcl6 Cxcr5 Pdcd1 Zbtb16 Nkg7 Nr4a1 Cd69 Mki67 Cd44
Average Expression
Percent Expressed
c
Intermediate-2 Th17 Il22+ Naive Intermediate-1 Th17 Tnfsf11+ Tfh Trm Proliferating NKT
Expression
d
Proportion
Tissue
Joint dLN Spleen
e
-log10(p_val)
avg_log2FC
Tnfsf11
Klf2
S1pr1
Ccr2
Il22
Ccr4
Ccr6
Atf3
Rbpj
Csf1
Anxa2
Tnfrsf18
Tnf
Tnfsf8
Tnfrsf9
Tnfrsf4
Lta
Up in cluster 1 (Th17 Il22+)
Up in cluster 4 (Th17 Tnfsf11+)
f
Th17 Il22+ (Cluster 1)
response to vitamin D (GO:0033280)
response to vitamin (GO:0033273)
regulation of T cell migration (GO:2000404)
cellular response to cytokine stimulus (GO:0071345)
cell chemotaxis (GO:0060326)
Odds.Ratio
-log10(Adj.p)
Th17 Tnfsf11+ (Cluster 4)
tumor necrosis factor-mediated signaling pathway (GO:0033209)
regulation of regulatory T cell differentiation (GO:0045589)
cytokine-mediated signaling pathway (GO:0019221)
cellular response to tumor necrosis factor (GO:0071356)
cellular response to cytokine stimulus (GO:0071345)
g
Il22
Tnfsf11
Expression
h
pseudotime

Figure 2

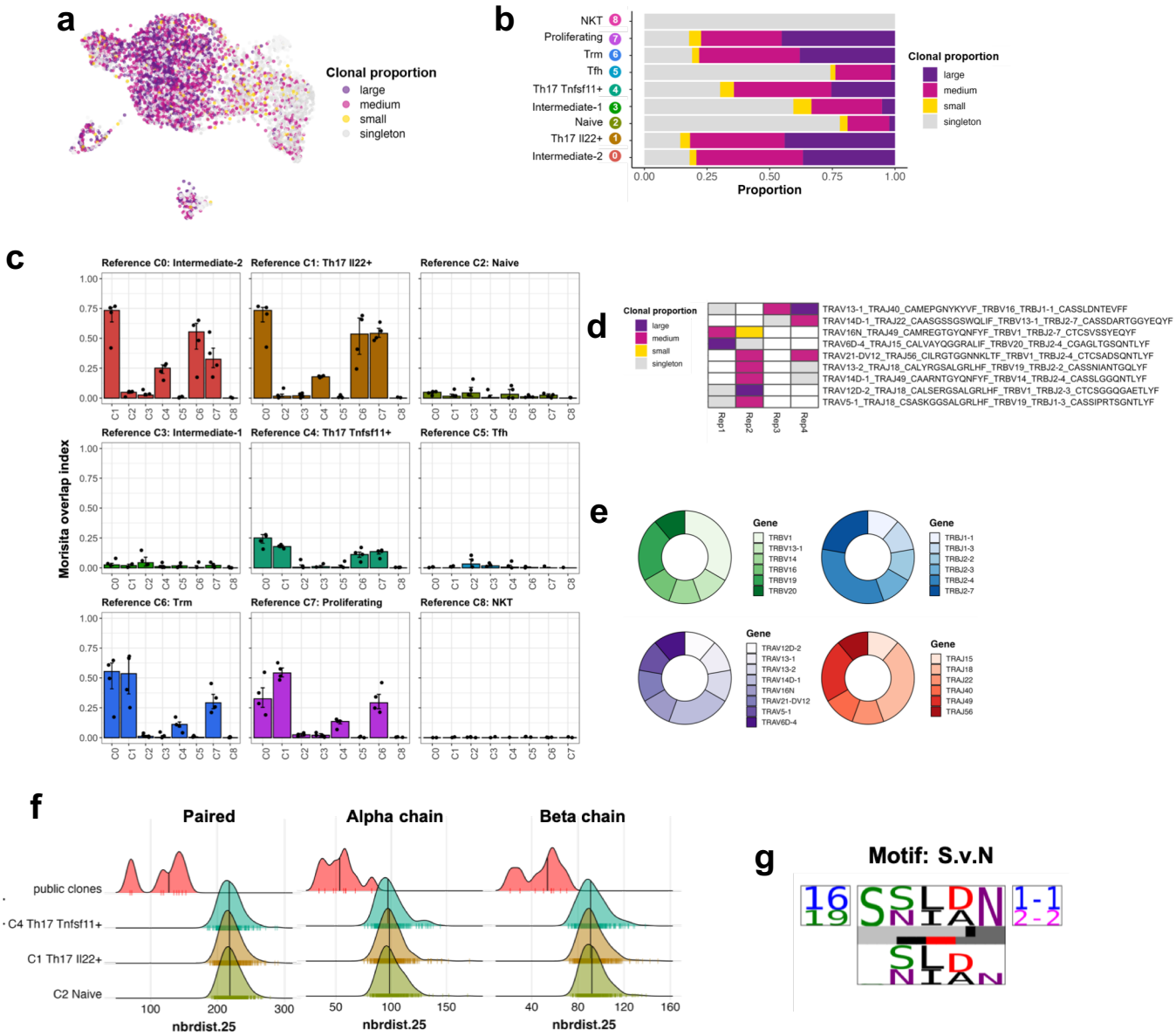

a
Clonal proportion
large
medium
small
singleton
b
NKT 8
Proliferating 7
Trm 6
Tfh 5
Th17 Tnfsf11+ 4
Intermediate-1 3
Naive 2
Th17 Il22+ 1
Intermediate-2 0
0.00 0.25 0.50 0.75 1.00
Proportion
Clonal proportion
large
medium
small
singleton
c
Reference C0: Intermediate-2
Reference C1: Th17 Il22+
Reference C2: Naive
Reference C3: Intermediate-1
Reference C4: Th17 Tnfsf11+
Reference C5: Tfh
Reference C6: Trm
Reference C7: Proliferating
Reference C8: NKT
Morisita overlap index
d
Clonal proportion
large
medium
small
singleton
TRAV13-1_TRAJ40_CAMEPGNYKYVF_TRBV16_TRBJ1-1_CASSLDNTEVFF
TRAV14D-1_TRAJ22_CAASGSSGSWQLIF_TRBV13-1_TRBJ2-7_CASSDARTGGYEQYF
TRAV16N_TRAJ49_CAMREGTGYQNFYF_TRBV1_TRBJ2-7_CTCSVSSYEQYF
TRAV6D-4_TRAJ15_CALVAYQGGRALIF_TRBV20_TRBJ2-4_CGAGLTGSQNTLYF
TRAV21-DV12_TRAJ56_CILRGTGGNNKLTF_TRBV1_TRBJ2-4_CTCSADSQNTLYF
TRAV13-2_TRAJ18_CALYRGSALGRLHF_TRBV19_TRBJ2-2_CASSNIANTGQLYF
TRAV14D-1_TRAJ49_CAARNTGYQNFYF_TRBV14_TRBJ2-4_CASSLGGQNTLYF
TRAV12D-2_TRAJ18_CALSERGSALGRLHF_TRBV1_TRBJ2-3_CTCSGGQGAETLYF
TRAV5-1_TRAJ18_CSASKGGSALGRLHF_TRBV19_TRBJ1-3_CASSIPRTSGNTLYF
Rep1
Rep2
Rep3
Rep4
e
Gene
TRBV1
TRBV13-1
TRBV14
TRBV16
TRBV19
TRBV20
Gene
TRBJ1-1
TRBJ1-3
TRBJ2-2
TRBJ2-3
TRBJ2-4
TRBJ2-7
Gene
TRAV12D-2
TRAV13-1
TRAV13-2
TRAV14D-1
TRAV16N
TRAV21-DV12
TRAV5-1
TRAV6D-4
Gene
TRAJ15
TRAJ18
TRAJ22
TRAJ40
TRAJ49
TRAJ56
f
Paired
Alpha chain
Beta chain
public clones
C4 Th17 Tnfsf11+
C1 Th17 Il22+
C2 Naive
100 200 300
50 100 150
40 80 120 160
nbrdist.25
nbrdist.25
nbrdist.25
g
Motif: S.v.N

Figure 3

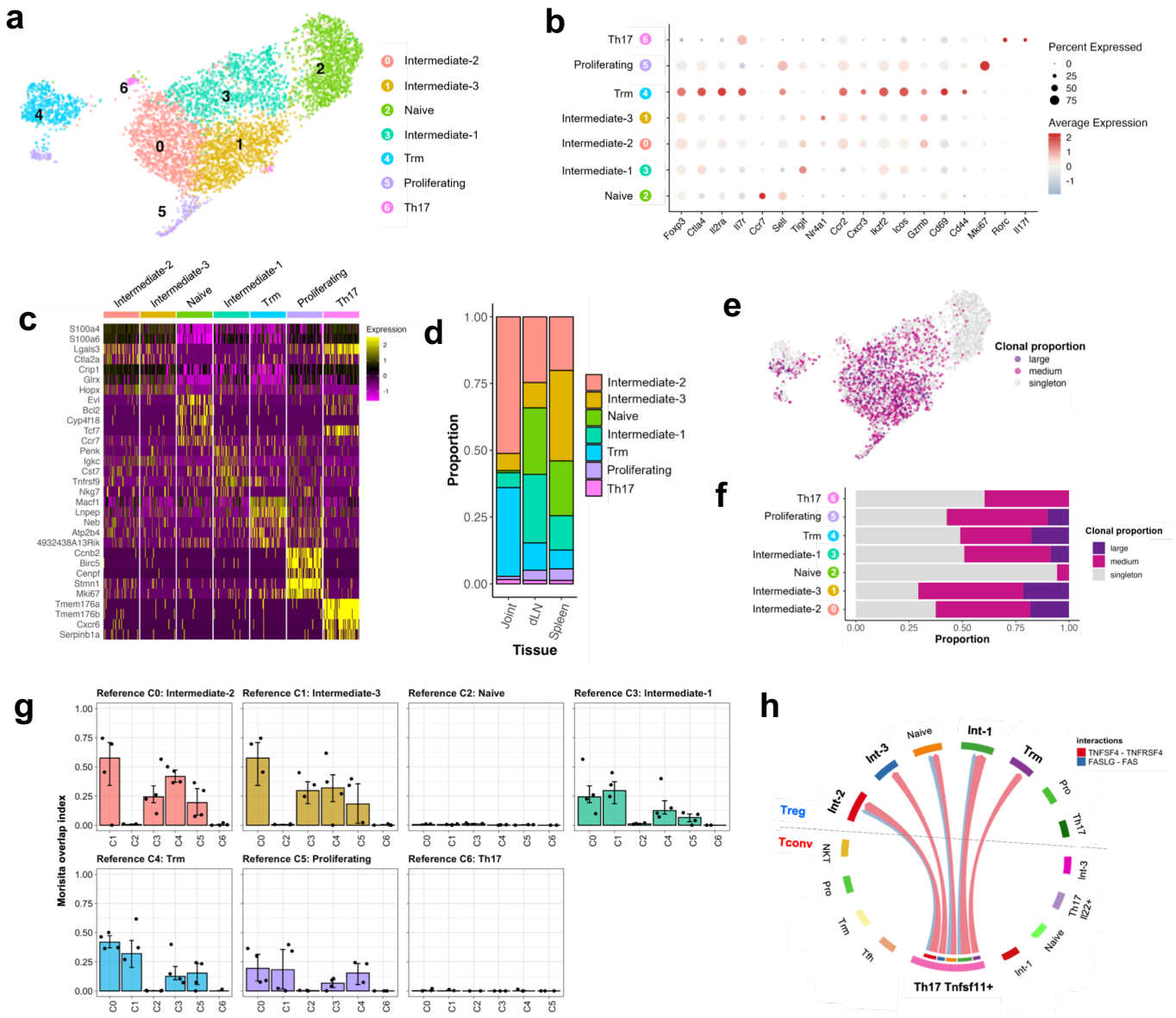

a
Intermediate-2
Intermediate-3
Naive
Intermediate-1
Trm
Proliferating
Th17
b
Percent Expressed
Average Expression
Foxp3
Ctla4
Il2ra
Il7r
Ccr7
Sell
Tigit
Nr4a1
Ccr2
Cxcr3
Ikzf2
Icos
Gzmb
Cd69
Cd44
Mki67
Rorc
Il17f
c
Expression
d
Proportion
Tissue
Joint
dLN
Spleen
e
Clonal proportion
large
medium
singleton
f
Proportion
g
Reference C0: Intermediate-2
Reference C1: Intermediate-3
Reference C2: Naive
Reference C3: Intermediate-1
Reference C4: Trm
Reference C5: Proliferating
Reference C6: Th17
Morisita overlap index
h
Treg
Tconv
interactions
TNFSF4 - TNFRSF4
FASLG - FAS
Th17 Tnfsf11+

Figure 4

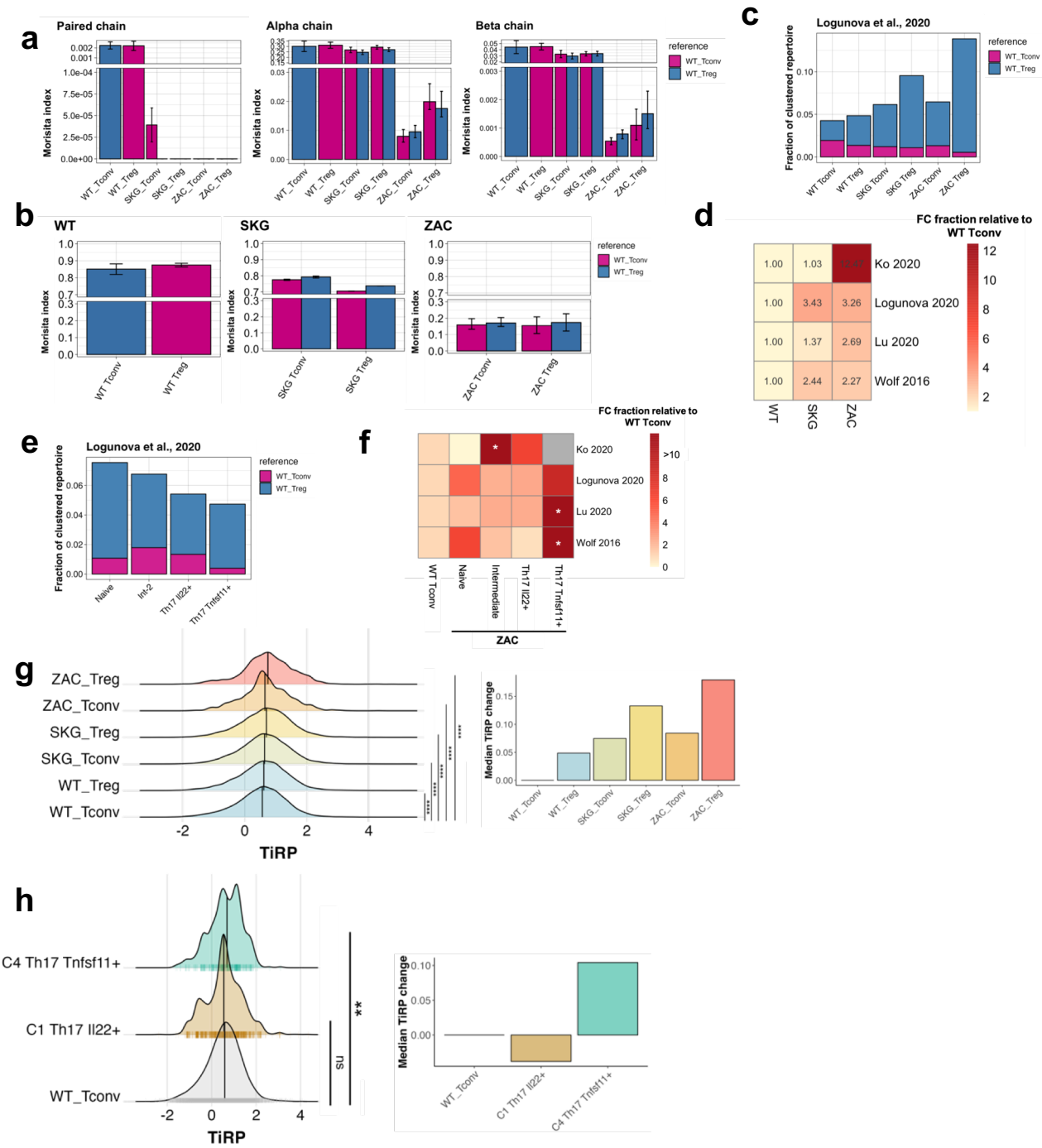
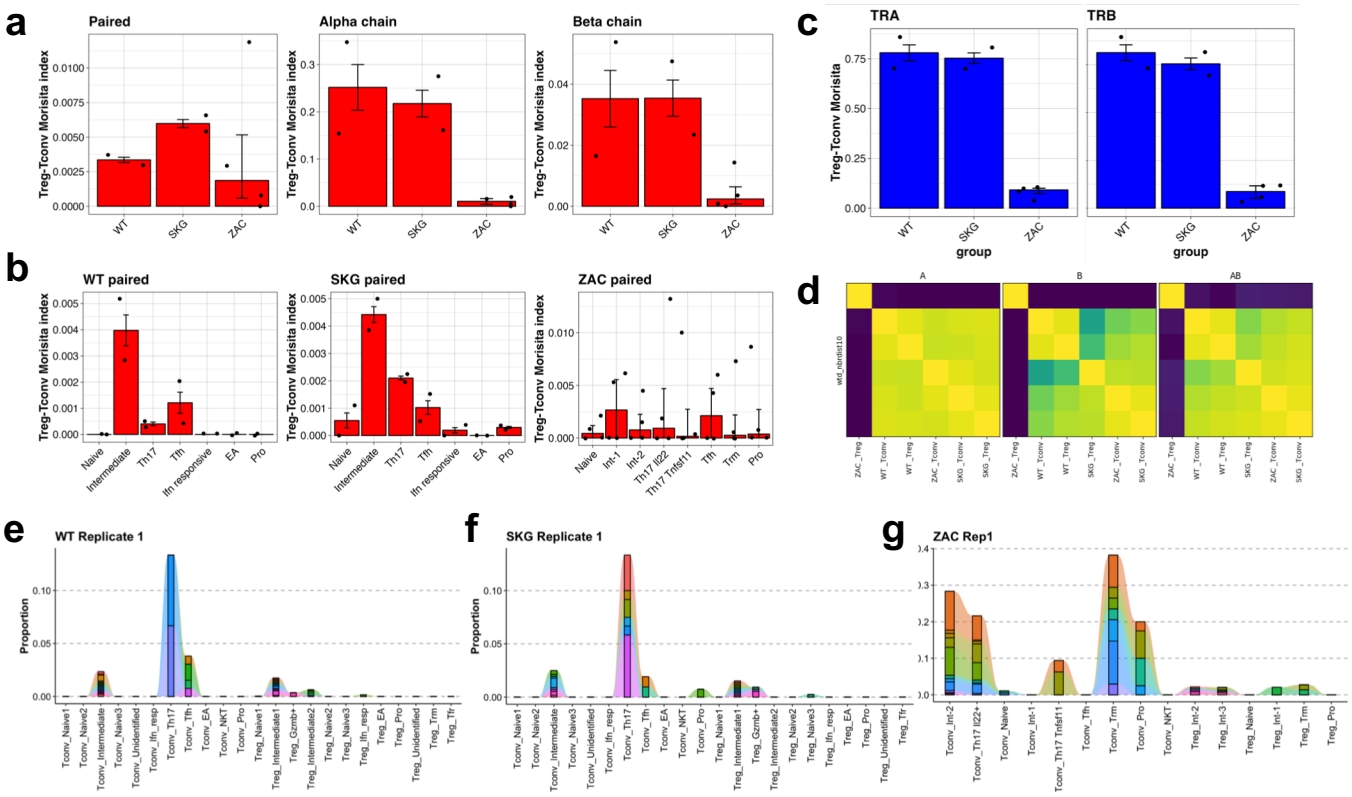

a
Paired chain
Alpha chain
Beta chain
Morisita index
reference
WT_Tconv
WT_Treg
SKG_Tconv
SKG_Treg
ZAC_Tconv
ZAC_Treg
b
WT
SKG
ZAC
c
Logunova et al., 2020
Fraction of clustered repertoire
d
FC fraction relative to WT Tconv
Ko 2020
Logunova 2020
Lu 2020
Wolf 2016
e
Naive
Int-2
Th17 Il22+
Th17 Tnfsf11+
f
Intermediate
g
TiRP
Median TiRP change
h
C4 Th17 Tnfsf11+
C1 Th17 Il22+
ns
**

Figure 5


a
Paired
Alpha chain
Beta chain
Treg-Tconv Morisita index
WT
SKG
ZAC
c
TRA
TRB
Treg-Tconv Morisita
group
b
WT paired
SKG paired
ZAC paired
Naive
Intermediate
Th17
Tfh
Ifn responsive
EA
Pro
Int-1
Int-2
Th17 Il22
Th17 Tnfsf11
Trm
d
A
B
AB
e
WT Replicate 1
f
SKG Replicate 1
g
ZAC Rep1
Proportion