Pathogenic Th17 cells in arthritogenic mice share T cell receptor features with Wild-type 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[2,9], 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 authors:

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; mara.llamas@biken.osaka-u.ac.jp

Abstract

Mutations in the ZAP-70 gene that cause moderate attenuation of T cell receptor (TCR) signaling in mice, can result in autoimmune manifestations akin to rheumatoid arthritis (RA). Here, we characterized the single-cell gene expression profiles and TCR repertoires of conventional (Tconv) and regulatory (Treg) $CD4^+$ T cells of arthritic (ZAC), poised (SKG) ZAP-70 mutant, and wild-type (WT) mice. We identified two Th17 cell subtypes in the joints of ZAC mice that were characterized by distinct transcriptional profiles and TCR repertoires, one of which exhibited a pathogenic signature and occurred exclusively in inflamed joints. Such a pathogenic signature was also uniquely detected in $CD4^+$ T cells obtained from inflamed joints of human RA patients. The TCR repertoire of pathogenic Th17 cells showed signs of increased intra-repertoire similarity (convergence) and was skewed toward the WT Treg rather than the WT Tconv repertoire. In addition, the overall similarity between the Treg and Tconv repertoires was severely reduced in arthritic mice. Our results support a model where, upon moderate 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.

Introduction

The development of autoimmune diseases is influenced by various factors, with the presence of self-reactive conventional T cells (Tconvs) being a critical element[1]. During T cell

maturation, thymocytes expressing T cell receptors (TCRs) with high affinity to self-peptide-MHC complexes are typically eliminated from the repertoire via negative selection or diverted to the T regulatory (Treg) cell lineage[2,3]. Nonetheless, a subset of self-reactive cells persists in the periphery of healthy individuals, usually kept in check by the suppressive activity of Tregs[4,5]. Antigen recognition by the TCR is key for both Treg development in the thymus and for their function in peripheral tissues[6-8]; and, it is generally accepted that the Treg repertoire is biased to self-recognition[9-12]. Therefore, disruption of the normal balance between Treg and Tconv repertoires can bias the immune system toward autoimmunity.

It has been previously reported that specific mutations in the ZAP-70 gene that moderately weaken downstream TCR signaling upon antigen recognition, altering the Treg-Tconv repertoire balance, predispose mice to an autoimmune disease closely resembling rheumatoid arthritis (RA)[13-15]. SKG mice, which are characterized by a W163C mutation in the ZAP-70 gene that causes a fifteen-fold reduction in ZAP-70 affinity for phosphorylated CD3ζ[13], develop disease upon secondary stimulation of innate immune responses (i.e. by mannan injection). 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[15]. The underlying mechanism has been ascribed to disruption of the normal thymic selection process that allows T cells with moderately to strongly self-reactive TCRs to escape negative selection[14,15]. As a consequence, in ZAP-70 mutant mice, some T cells harboring TCRs that would normally be observed in the Treg repertoire instead

develop into self-reactive Tconvs. Here, we refer to this mechanism as the "repertoire shift" hypothesis.

In order to test this hypothesis, we examined the transcriptome and immune repertoires of $CD4^+$ T cells in WT, non-arthritic SKG and arthritic ZAC mice at the single-cell level. We found migratory and pathogenic populations of Th17 cells in arthritic mice. Interestingly, pathogenic cells highly expressed the osteoclastogenic factor *Tnfsf11* (encoding for RANKL) and uniquely localize in inflamed joints but were absent in WT and SKG mice. A group of cells with a similar pathogenic signature was uniquely identified in T cells from human RA joints. Furthermore, the TCR repertoire of pathogenic Th17 cells was unique, convergent, and overlapped with the repertoire of WT Tregs. Finally, the overall similarity between the Treg and Tconv repertoires was severely reduced in arthritic mice, suggesting that a depletion of critical components of the Treg repertoire accompanies expansion of Tconvs bearing TCRs that are biased toward self-reactivity. Our data thus supports the repertoire shift hypothesis and suggests that the autoimmunity observed upon moderate reduction of TCR signaling strength results from T cells bearing self-reactive TCRs that evade negative selection. As a consequence, TCRs typical of Tregs are expressed in Tconv cells, and this, along with a depletion of specific receptors in the Treg compartment, lead to autoimmunity.

Results

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

The ZAP-70 mutant mice (ZAC, SKG) used in this study are characterized by the spontaneous development of autoimmune arthritis (Fig. 1a), with disease progression typically accelerated in the SKG model through the stimulation of innate immunity. Unlike other widely used mouse models of arthritis, such as collagen-induced arthritis (CIA)[16], the arthritis in ZAP-70 mutant mice is genuinely autoimmune, mediated primarily by T-cells, and closely resembles human RA [13-15]. Therefore, ZAP-70 mutant mouse models are particularly well-suited for investigating self-reactive TCR repertoires. To study the transcriptomes and immune repertoires of self-reactive Tconv ($CD3^+CD4^+Foxp3^-$) cells in-depth, we sorted cells from the joints, draining lymph nodes (dLN), and spleens of arthritic 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 dLNs but a distinct profile for joint cells (Supplementary Fig. 1a). Unsupervised clustering delineated nine gene expression clusters with similar distribution in each tissue across replicates (Supplementary Note 1, Fig. 1b-c, Supplementary Fig. 1b). Two clusters (1 and 4) showed canonical Th17 markers (*Rorc*, *Il17a*, *Il17f*), but were distinct in their gene expression programs and tissue abundances (Fig. 1b-d, Supplementary Fig. 1b-d). Cluster 1 was abundant in joints, representing ~50% of cells obtained from this tissue, but was also detected in the spleen and dLNs at smaller proportions (~15-17%). 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 (Fig. 1e, Supplementary Table 4). Consistently, pathway analysis revealed an enrichment in the regulation of T cell migration (GO: 20000404) and cell chemotaxis (GO: 0060326) gene ontology (GO) Terms (Fig. 1f, Supplementary Table 5). 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 (Fig. 1c, Supplementary Fig.1b). The top differentially expressed gene (DEG) in cluster 4 relative to cluster 1 was *Tnfsf11*, which codes for RANKL, a critical factor involved in osteoclast differentiation and activation that plays a central role in bone destruction in RA[17] (Fig. 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 (Supplementary Table 5). In addition, transcripts involved in the response to TNF stimuli (for example: *Tnfrsf18*, *Tnfrsf9*, *Tnfrsf4*, and *Tnf*; encoding for GITR, 4-1BB, OX40, and TNF-α respectively), and the cellular response to tumor necrosis factor (GO:0072356) GO terms were enriched in cluster 4. Moreover, this cluster uniquely expressed transcription factors associated with a pathogenic Th17 phenotype: *Rbpj*[18], *Crem*[19], *Atf3*[20], *Bhlhe40*[21], and *Igfbp7*[22] (Fig. 1e). Finally, the transcription factor *Klf2*, a known inhibitor of NF-κB-dependent genes that also prevents Tfh differentiation and RORγt expression[23,24] was absent from cluster 4 (Supplementary Fig. 1e). These results indicate that cluster 4 contains pathogenic Th17 cells with an increased potential to activate osteoclasts and thus promote bone damage. Interestingly, cluster 1 and cluster 4 could be clearly distinguished by their differential expression of *Il22* or *Tnfsf11* (Fig. 1g). Finally, we annotated clusters 1 and 4 as

migratory and pathogenic Th17, respectively. In order to explore the dynamics of the different Tconv cell states, we performed trajectory inference setting the naïve cluster of cells as the starting point in pseudotime. This resulted in the identification of the pathogenic Th17 cluster as one of the endpoint lineage states (Fig. 1h), consistent with their effector phenotype. A population of pathogenic Th17 cells characterized by the expression of Th1 cell markers (*Ifng*; *Tbx21*, encoding for T-bet; and *Cxcr3;* Th17.1 cells)[25], and a group of osteoclastogenic Th17 cells derived from $CD4^+$ T cells previously expressing FoxP3 (exFoxP3) and characterized by the expression of *Sox4*, *Ccr6*, *Ccl20*, *Il23r* and *Tnfsf11*[26] have been previously reported to be enriched in arthritic joints; however, other than *Tnfsf11* we did not detect clear and specific expression of these markers in cluster 4, suggesting that it represents a distinct group of cells (Supplementary Fig. 1f-g).

To contextualize our findings in arthritic ZAC mice tissues, we replicated the analyses in WT and non-arthritic SKG mice. For this, we sampled Tconv from the spleens of SKG and WT mice, and integrated the ZAC spleen data in order to facilitate direct comparison of clusters (Supplementary Note 2, Supplementary Fig. 2a-g). This approach revealed that the general Th17 cell abundance increased progressively from WT (less than 1%) to SKG (2-9%) to ZAC (~36%) mice (Supplementary Fig. 2d). To analyze whether the transcriptional programs of these Th17 cells are similar to the transcriptional programs of the two Th17 cells identified in ZAC tissues, we calculated gene module scores for the signature genes from migratory and pathogenic Th17 cells (Supplementary Table 6). Expression of the migratory Th17 signature was clearly detected and gradually increased from WT, to SKG to ZAC splenic

Th17 cells; however, the pathogenic Th17 signature was largely undetected (Fig. 2a-b, Supplementary Fig. 2h). This result indicates that pathogenic Th17 cells are uniquely present in the inflamed joints of arthritic mice.

In summary, we demonstrated the presence of two distinct groups of Th17 cells in arthritic ZAC mice: migratory Th17 cells, expressing *Il22* and the transcription factor *Klf2*; and, pathogenic Th17 cells, which highly express *Tnfsf11* and uniquely localize within inflamed joints.

The transcriptional signature of pathogenic Th17 cells is present in inflamed joints of human RA but not in other autoimmune diseases or non-autoimmune arthritis

To investigate the presence of cells resembling pathogenic Th17 cells in human tissues, we calculated migratory and pathogenic Th17 gene signature module scores in publicly available single-cell RNAseq datasets. We analyzed data obtained from inflamed synovia[27] and PBMCs[28] of RA patients; and, for comparison, we also analyzed data from the synovia and infrapatellar fat pad (IPFP) of osteoarthritis (OA) patients[29] as well as peripheral CD4+ T cells from healthy controls and patients with other autoimmune diseases (myasthenia gravis, multiple sclerosis, and systemic lupus erythematosus)[30]. Detailed descriptions of the datasets are provided in Supplementary Table 7.

Our analysis revealed that the migratory Th17 signature was clearly detected in $CD4^+$T cells from inflamed synovia or peripheral blood from RA patients (Fig. 2c-d), as well as in $CD4^+$T

cells obtained from joint tissues of OA patients (Fig. 2e), and in circulating Th17 cells in healthy controls and patients with other autoimmune diseases (Fig. 2f).

In contrast, expression of the pathogenic Th17 signature was nearly undetectable in all samples (Fig. 2 d-f) except in $CD4^+$ T cells derived from inflamed joints of RA patients (Fig. 2c) where not only the highest expression of this gene signature was detected, but its expression was found to be non-overlapping to the expression of the migratory Th17 signature, mirroring our observations in ZAC joints. Notably, as in ZAC joints, the expression of Th17.1 or exFoxP3 markers was not specifically detected in the cluster of cells expressing the pathogenic Th17 gene signature in human RA joints (Supplementary Fig. 3a-d).

OA is the most common form of arthritis and is mainly provoked by the mechanical wear and tear of joints, and, even though inflammation is involved in its pathogenesis, it is not a disease of autoimmune origin[31]. Thus, our results indicate that cells with a pathogenic Th17 transcriptional signature are likely to be specifically involved in the pathogenesis of human autoimmune arthritis.

Pathogenic Th17 cells are expanded, convergent, and contain public clones

Having identified two groups of potentially self-reactive Th17 cells with characteristic transcriptomic features, we next asked whether their TCR repertoires were also distinct. First, we examined the degree of clonal expansion by calculating the proportion of a whole

repertoire that each clone represented (clonal proportion) in each replicate, and subsequently classified clones as: small, medium or large (methods). We found that large and medium clones were abundant in both Th17 cell groups (Fig. 3a-b, Supplementary Fig. 4a). Next, we evaluated the repertoire similarity of both Th17 groups to other Tconv populations and between each other by using the Morisita-Horn index[32]. At the exact paired CDR3 amino acid sequence level, the repertoire of migratory Th17 cells was largely similar to other Tconv repertoires from ZAC tissues (Fig. 3c). On the other hand, the pathogenic Th17 repertoire was less similar to other Tconv cells. Consistently, the similarity between the two Th17 cell repertoires was rather low (Fig. 3c). These results suggest that pathogenic Th17 cells possess a unique repertoire, presumably due to the targeting of a specific set of antigens.

Given the unique repertoire of pathogenic Th17 cells, we next evaluated intra-repertoire TCR similarity as a measure of repertoire convergence. For this purpose, we used TCRdist[33], which calculates the distance between the CDR1, CDR2 and CDR3 amino acid sequences of single or paired TCR chains. For this analysis, we independently calculated the TCR distances among all receptors in the two Th17 clusters, as well as the naïve cluster for reference. A bias in the distribution of the distances to lower values indicates a higher similarity of its receptors, and therefore can be interpreted as an indication of repertoire convergence. We found that the TCR distance distribution of the pathogenic Th17 repertoire trended lower in paired and single chain analysis and was significantly lower in the alpha chain (Fig. 3d). This result indicates that the repertoire of pathogenic Th17 cells is more convergent than that of migratory Th17 cells, especially in the alpha chain. Given their

specific location in joints, and the evidence of similar cells occurring in human RA but not OA joints, our results suggest that pathogenic Th17 may recognize a particular self-antigen in the affected tissue and that the alpha chain may play a key role in such recognition.

Finally, we identified a total of 9 unique clones shared in two or more individuals (public clones) (Fig. 3e, Supplementary Table 8). Interestingly, public clone 1 occurred in 3 out of 4 replicates, suggesting that this clone is likely to be involved in the recognition of relevant arthritogenic self-antigens (Fig. 3e).

Together, these results indicate that although the Th17 response in arthritic ZAC mice is polyclonal, the repertoire of tissue-specific pathogenic Th17 cells is unique and shows high TCR convergence, suggesting the recognition of a limited set of self-antigens. Moreover, a number of highly expanded public clones was observed within the inflamed tissue.

Effector Treg cells are abundant in the joints of arthritic ZAC mice but potentially suppressed by cell crosstalk with self-reactive Tconvs

As discussed previously, the failure of Tregs to suppress self-reactive T cells and subsequent inflammation is a key factor in the development of autoimmune diseases. To investigate the Treg landscape in ZAP-70 mutant mice, we sorted Treg cells from the joints, dLN, and spleens of arthritic ZAC mice as well as from spleens of SKG and WT mice and performed single-cell RNA and V(D)J sequencing. In the tissues of ZAC mice, we identified seven gene

expression clusters, including two groups of effector cells (*Ccr2, Cxcr3*, *Gzmb*) and one cluster of cells with a tissue resident memory (Trm) signature (*Il7r, Sell, Cd44, Cd69, Itgae*) (Supplementary Fig. 5a-c), which were notably expanded and abundant in joints (Supplementary Note 3, Supplementary Fig. 5a-g, Supplementary Table 10, 11). Conversely, in WT and SKG Tregs, we observed limited effector and Trm cells, while naïve cells were abundant (Supplementary Note 4, Supplementary Fig. 6a-f). Remarkably, despite the evident expansion in the Treg compartment of ZAC mice, only one public clone could be identified (Supplementary Table 12). These findings suggest that Trm and effector Treg cells undergo expansion in the tissues of arthritic ZAC mice.

Despite the abundance of effector Treg cells in the sampled tissues, ZAC mice were overly autoreactive. To elucidate the disparities in cell activity among WT, SKG, and ZAC effector Tregs, we performed DEG analysis across mouse models. Our analysis revealed that while the transcriptional programs of WT and SKG Treg cells were largely similar, several genes involved in TNF signaling were uniquely expressed in ZAC effector Tregs (Supplementary Fig. 7a-b, Supplementary Table 13). While the exact effect of TNF signaling on Treg effector function remains unclear, studies have suggested its suppressive effects on Treg function in RA patients[34].

Tregs can exert their suppressive functions by various mechanisms, including physical interactions with target cells or secretion of soluble factors[35,36]. Therefore, we sought to identify potentially significant cell-cell crosstalk events between Tconv Th17 cells and effector Tregs cells by predicting cell communication events using Cellphone DB[37].

Consistent with the increased TNF signaling activity we observed in ZAC mice, there was a gradual increase in the number of predicted cell communication events involving members of the TNF family from WT, SKG, and ZAC Th17 cells toward Treg cells (Supplementary Fig. 7c). Thus, our findings suggest that despite the abundance of effector Treg cells in ZAC mice, their function is impaired and one of the mechanisms may involve cell crosstalk with pathogenic Th17 cells through members of the TNF signaling family.

TCR repertoires of ZAP-70 mutant Tconv, particularly pathogenic Th17, cells are more WT Treg- than WT Tconv-like

Because T cells harboring self-reactive TCRs would likely be deleted or develop into Tregs in WT mice, the repertoire shift hypothesis postulates that ZAP-70 hypomorphic mutations would lead to the appearance of TCR repertoires in Tconvs that would be typically found in WT Tregs, or normally deleted[14,15]. A natural way of testing this hypothesis would be to compare the similarity of the repertoires of WT Tregs and SKG or ZAC Tconvs. If ZAP-70 mutant Tconvs were more similar to WT Tregs than to WT Tconvs, it would support the repertoire shift hypothesis. To address this question, we turned to previously published TCR reference datasets[38-41] in order to consolidate WT Tconv and WT Treg reference data (Supplementary Table 14). For the calculation, we treated each of our experimentally determined (WT Tconv, WT Treg, SKG Tconv, SKG Treg, ZAC Tconv, or ZAC Treg) TCRs as a "query" that we used to search a reference WT Tconv or WT Treg dataset (Logunova[38], Lu[39], Ko[40], Wolf[41]). For each of these queries, we computed the number of reference WT

Tconv and WT Treg CDR1, CDR2 and CDR3 sequence matches (methods) and, based on the number of matches along with the total size of each reference dataset, deemed a given query as “WT Treg-like”, “WT Tconv-like” or “not significant” using a hypergeometric p-value cutoff of 0.001. We then quantified the total number of matches of each type in terms of the query repertoire size. We then expressed these values relative to the WT Tconv query group (Fig. 4a). Considering first the largest (Logunova[38]) reference dataset, we found that the fraction of WT Treg-like TCRs within SKG and ZAC Tconv queries was approximately twice as high as the fraction in WT Tconvs, and even higher than the fraction in query WT Tregs (Fig. 4b). In addition, the fraction of WT Tconv-like TCRs was notably reduced within ZAP-70 mutant Tconv queries (Fig. 4b). Similar trends were observed in the other reference datasets (Supplementary Fig. 8a). Furthermore, the ZAC pathogenic Th17 repertoire showed an increased fraction of WT Treg-like TCRs and consistently displayed the smallest fraction of WT Tconv-like TCRs (Fig. 4c, Supplementary Fig. 8b), indicating that the TCRs of pathogenic Th17 cells were largely absent from WT Tconv repertoires, but similar to WT Treg repertoires. Of note, pathogenic Th17 cells from ZAC mice did not show any significant WT Treg or WT Tconv matches within the Ko et al., 2020[40] dataset (Supplementary Fig. 8b). These observations support the repertoire shift hypothesis: ZAP-70 mutant Tconv repertoires are more similar to WT Treg repertoires than to WT Tconv repertoires, and this trend is most pronounced in the pathogenic Th17 cells.

Because the query-reference matching procedure described above requires a number of ad hoc normalization steps, we next sought a simpler procedure that could be used as an

independent check. To this end, we took advantage of a machine learning tool that enables the quantification of the regulatory potential of a TCR (TCR-intrinsic regulatory potential, TiRP) based on TCR beta chain features[42]. As expected, our experimentally determined Treg datasets exhibited higher TiRP scores than WT Tconv cells, and their respective Tconv compartments (Supplementary Fig. 8c-d). Importantly, the ZAP-70 mutant Tconvs also exhibited significantly higher TiRP scores than WT Tconvs, and higher than WT Tregs (Fig. 4d-e). Furthermore, when considering the ZAC Th17 phenotypes, we found that pathogenic Th17 TiRP scores were significantly higher compared to those of WT Tconvs (Fig. 4f-g). The higher TiRP scores in self-reactive T cells[42] suggest that the TCRs in Tconv cells from ZAP-70 mutant mice, particularly those in pathogenic Th17 cells, are biased towards self-reactivity. These results agree with those of the query-reference matching procedure, and, again, support the repertoire shift hypothesis as an explanation for how ZAP-70 mutation-mediated TCR signal weakening leads to the emergence of self-reactive Th17 cells that develop a pathogenic phenotype and accumulate in joints upon activation.

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

A second consequence of the repertoire shift hypothesis is that a weakening of TCR signaling impacts the Treg compartment by generating “holes” in the normal Treg repertoire. It has been reported that missing specificities in the Treg repertoire have a large impact on the suppressive function against self-reactive T cells[8,43]. This raises the possibility that an absence of Treg specificities within ZAP-70 mutant mice repertoires contributes 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 groups (WT, SKG, ZAC). The Morisita-Horn index for exact CDR3 overlap showed a notably reduced Treg-Tconv similarity in ZAC mice relative to WT or SKG, especially at the single chain level (Fig. 5a). The Treg-Tconv overlap was generally distributed across effector phenotypes (Tconv: intermediate, Th17 and Tfh. Treg: intermediate and effector) in non-arthritic mice (Supplementary Fig. 9a-h), suggesting a suppressive potential for shared Tregs towards activated Tconvs; however, such enrichment was not observed in ZAC mice. To further test these observations, we again employed TCRdist[33] and calculated repertoire-repertoire correlations (methods) for pooled WT, SKG and ZAC repertoires: here, a pair of similar repertoires exhibit high correlation coefficients. We found that WT Treg and WT Tconv repertoires were well correlated, whereas SKG and ZAC Tconv repertoires exhibited lower correlation with their Treg repertoires. Indeed, the ZAC Treg repertoire was poorly correlated with other groups in general, indicating a reduced similarity to any other tested repertoires (Fig. 5b). Together, these results are consistent with a scenario wherein ZAP-70 mutant were unable to select certain Treg TCR specificities, normally present in WT mice and capable of suppressing autoimmune arthritis; which, along with a more self-reactive Th17 population, contributes to a loss of tolerance and the development of autoimmune arthritis.

Discussion

In this study, we provide a comprehensive characterization of the Treg and Tconv phenotypes and repertoires of WT, non-arthritic SKG and arthritic ZAC mice at the single cell level. We

identified two transcriptionally distinct populations of Th17 cells within tissues of arthritic ZAC mice: migratory Th17 with high expression of *Il22*, and pathogenic Th17 cells, highly expressing *Tnfsf11*. Pathogenic Th17 cells have a unique transcriptional signature, and are almost exclusively detected in joints. In addition, pathogenic cells may target a restricted set of antigens, as they harbor unique and convergent TCRs with a greater similarity to the TCRs of WT Treg cells, which are generally understood to be biased toward self-reactivity[9-12].

Investigations into the contribution of high-affinity versus lower-affinity T cell specificities to autoimmune disease establishment and progression, as observed in experimental autoimmune encephalomyelitis (EAE), have revealed a critical role of high-affinity autoreactive cells in initiating disease that can be maintained by lower-affinity cells once established[44]. In our study, it is possible that the pathogenic Th17 population comprises cells critical for the establishment of disease, given their unique repertoire and tissue location; while migratory Th17 cells may contribute to disease progression and maintenance and represent a prior differentiation state for cells that later acquire a pathogenic Th17 phenotype. Such dynamics would be in line with increased diversity and reduced clonal dominance in established disease stages[45,46].

Consistently, our trajectory prediction suggests that pathogenic Th17 cells represent a terminal differentiation state, preceded by the migratory Th17 phenotype. This observation also aligns with previous research showing that in EAE Th17 cells progress from a self-renewing cellular state primarily enriched in lymph nodes, to an effector pre-Th1-like

phenotype found in both lymph nodes and the central nervous system (CNS) and characterized by the expression of chemokine receptors, to a Th17/Th1-like effector phenotype that is concentrated in the CNS and exhibits a pathogenic signature with high expression of *Tnfsf11*[47].

In addition to these observations on the pathogenic phenotype of ZAC Th17 cells, we carried out the first systematic investigation at single cell resolution of the repertoire shift of conventional T cells upon TCR signal weakening. Our results are consistent with the observation that T cells infiltrating prostatic autoimmune lesions in Aire$^{-/-}$ mice express receptors preferentially found in Tregs from Aire$^{+/+}$ mice[48]. In addition, it was shown that exFoxP3 cells display a higher osteoclastogenic potential as compared to Th17 derived from naïve CD4^{+} T cells, presumably due to their higher affinity to self-antigens[26]. Together with our findings, these reports point to a key role of TCRs normally occurring in the Treg compartment in the recognition of self-antigens, and suggests that aberrant expression of such TCRs in Tconv cells may drive autoimmune disease. Finally, the unique identification of CD4^{+} T cells expressing the pathogenic Th17 gene signature in inflamed synovia from RA patients highlights their potential involvement in autoimmune arthritic inflammation, motivating further identification of their tissue-specific self-antigens.

Remarkably, we observed a number of public clones within ZAC Tconv cells, particularly concentrated within both Th17 phenotypes, and notably expanded. The significance of public clones, in the development of autoimmune conditions, has been previously documented[49,50].

Among the CDR3 sequences of public clones, we identified only one exact match to a TCR in the VDJdb database[51] of TCR sequences with known antigen specificities (Supplementary Table 9). Unexpectedly, the match corresponded to a CD8 TCR, and its target peptide was presented in a class I MHC. Although the presence of "class-mismatched" T cells has been reported in the literature[52-54], it remains unknown whether this phenomenon occurs in this or other self-reactive clones in ZAC mice. Consequently, the identity of any potential epitopes or amino acid motifs remains an open problem.

As migratory and pathogenic Th17 cells were predicted to be related both by trajectory analysis and by the existence of a limited number of shared clones, we speculate that the pathogenic Th17 phenotype is only acquired within joints and only by cells bearing self-reactive TCRs. The data supports a model wherein these cells are critical in triggering and establishing disease in target tissues, whereas migratory Th17 cells play a supporting inflammatory role, but are not strongly self-reactive.

Finally, the repertoire shift hypothesis raises the prospect of loss of certain TCR specificities within the Treg compartment. Considering that antigen recognition by the Treg TCR appears to be less degenerate than in Tconv cells[8,55,56], the loss of certain Treg specificities could profoundly impact tolerance towards specific self-antigens[43]. While it is generally acknowledged that Treg and Tconv repertoires are distinct, a small proportion of their repertoires typically overlap[46,57,58]. In this study, we observed a reduction in the similarity between ZAC Treg and Tconv repertoires, specially within activated Tconv cell groups,

suggesting a divergence in the Treg-Tconv repertoires that could result in the loss of critical Treg function. This result is consistent with the observation of a significantly reduced proportion of potentially self-reactive receptors in the Treg repertoires of autoimmune Aire$^{-/-}$ mice[48]. Therefore, it is possible to speculate that moderate TCR weakening by ZAP-70 mutation may not only shift the Tconv repertoire but also deplete the Treg repertoire of relevant TCRs. The functional identity of such critical TCRs remains unknown, and thus worthy of future study as a starting point for development of TCR-transgenic Treg immunotherapy for the treatment of autoimmune diseases[59].

Taken together, the consequences of weakened TCR signaling in the development of autoimmune disease are multilayered. Firstly, it allows cells with self-reactive TCRs to evade negative selection or deflection into the Treg repertoire by differentiation into self-reactive Tconvs. At the same time, Tregs are affected in at least two ways: on the one hand, the Treg repertoire may become deficient in some normally-occurring and potentially critical TCRs, resulting in not a shift, but also a bifurcation of the repertoire. On the other hand, Treg effector activity upon TCR weakening is significantly disrupted[15], rendering Tregs insufficient for the suppression of self-reactive Tconvs (Fig. 6).

It should be noted that our study has several limitations, such as the limited sample size and repertoire coverage that can be practically analyzed by current single cell sequencing approaches. In addition, the pathogenic potential of Th17 cells, their self-reactivity, as well as their potential to suppress Treg function have not been further experimentally assessed.

Moreover, the antigen specificity of public clones could only be predicted, which yielded only a single match to a functionally annotated TCR. This TCR was described as being specific to a viral epitope presented via MHC class I. Future *in vitro* and *in vivo* experiments will be required to elucidate the specific role and antigen specificities of Th17 cell populations in disease establishment and progression. In spite of these limitations, our research expands the cellular and molecular level understanding of the breakdown in tolerance during autoimmune disease and proposes a straightforward path to discovering specific TCR-antigen interactions involved in RA.

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[13], ZAC[15] and FIG (Foxp3-IRES-GFP knock-in[60]) mice were previously described. In this study, FIG BALB/c WT, FIG SKG and FIG ZAC mice were used and generated as previously reported [15]. 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 RNAseq 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[61,62].

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[63] 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 (Supplementary Table 1). Uniform manifold approximation and projection (UMAP) was 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.

Public single-cell RNAseq data was obtained as count matrices and pre-processed accordingly for QC, normalization, batch correction and data integration (when required).

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 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 Notes 1-4). Additional exclusion of artifactual populations was performed after clustering, including cells lacking expression of CD3 (*Cd3d*, *Cd3e*, and *Cd3g)*, invariant NKT (iNKT) cells, and clusters with high expression of 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. Significant markers were considered with a fold change (FC) ≥1.2 and adjusted p value ≤0.05, and were used for gene set enrichment analysis.

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[64] using the *GO_Biological_Process_2021* database as reference.

For the generation of ZAC migratory and pathogenic Th17 gene signatures for module score calculation, significant markers were considered if fold change (FC) ≥2 and adjusted p value ≤0.05.

Trajectory analysis

Trajectory analyses were performed by using the Slingshot Bioconductor package (v2.10.0)[65] 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 Th17 ZAC programs in spleen and public human scRNAseq datasets was performed by the *AddModuleScore* function from Seurat using the specified gene lists or their corresponding human orthologs (Supplementary Table 6) and default parameters.

For public scRNAseq datasets, clustering was performed as described above. Where it was possible to annotate Th17 cells, these were separated in an independent Seurat object and processed for module score calculation and visualization. If Th17 cells were not clearly identified, the whole $CD4^+$ T cell clusters were separated in independent Seurat objects for module score calculation.

Cell-cell crosstalk prediction

We used Cellphone DB[37] to predict cell-cell communication. After converting mouse gene IDs to human orthologues, we ran CellPhoneDB v4.0.0 in statistical analysis mode, for the subsets of effector Treg together with Intermediate or Th17 Tconvs. Results were visualized by the ktplots R package. We manually searched for all significant interactions in each mouse model. 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 and repertoire overlap

Clonal expansion was assessed by the calculation of clonal proportions, then, clones were classified by their size 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.

The Morisita-Horn index is a typical metric used in ecology to calculate the similarity 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[32]. The Morisita-Horn index was calculated between repertoires of each individual replicate, at the paired or single CDR3 amino acid level by the Immunarch (v0.9.0) R package.

The median Morisita-Horn indices were compared by a Wilcoxon test using a significance level of 0.05. In addition, clonotype tracking across Treg and Tconv phenotypes was performed by Immunarch.

Clonal publicity

Public clones were defined as TCRs sharing paired gene usage and CDR3 amino acid sequences and occurring in more than one subject.

TCR distances and repertoire-repertoire correlation

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 by comparing the CDR1, CDR2 and CDR3 amino acid sequences of single or paired TCR chains[33], then, a neighbor distance metric consisting of the weighted average of the distance of a given TCR to the nearest 25th percentile of the other TCRs in the repertoire can be calculated (weighted neighbor distance). Then, distance distributions of selected repertoires were compared and distributions biased toward lower distances were interpreted as an indication of repertoire convergence. In order to restrict the convergence analysis to the qualitative composition of the repertoires, we used unique clone sequences i.e., clonal counts where not considered. Median distances of Naïve, migratory and pathogenic Th17 cells were compared by a Wilcoxon test with a significance level of 0.05. Finally, the epitope-epitope (or repertoire-repertoire) neighbor distance score correlation metric from the TCRdist algorithm was computed for assessment of intra Treg-Tconv

repertoire similitude. Briefly, a score is assigned to each TCR in one repertoire which is the average distance to the nearest TCRs in a reference repertoire. Then, the sets of scores obtained by two different repertoires can be compared over the entire merged set of TCRs and the Pearson correlation score is calculated. In principle, two similar repertoires will be well correlated.

TiRP score calculation

Lagattuta et al proposed a Treg-propensity scoring system for the TCR to quantify the likelihood of Treg cells based on the TCR feature on Treg fate (TiRP score)[42]. 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. Median TiRP scores were compared across groups by a Wilcoxon test with a significance level of 0.05.

Published sequence preparation

The published mouse Treg and Tconv TCR beta chain sequence data were obtained from four studies[38-41]. The detailed accession ID and references were summarized in Supplementary Table 14. The annotation of raw sequences was performed by mixcr (3.0.13)[66] with align, assemble, assembleContigs, and exportClones process. The sequences that have complete sequences of frameworks and CDRs regions were utilized in this study.

Sequence searching and comparison

We implemented the TCR sequence searching using Interclone method[67]. 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 reference 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 using a hypergeometric p-value cutoff of 0.001. We then quantified the total number of matches in terms of the query repertoire size and normalized the results to the matches in the query WT Tconvs (Fig. 4a).

Statistical analysis

For identification of cluster markers, significant markers were considered with a FC ≥1.2 and adjusted p value ≤0.05. For DEG analysis between two cell populations, 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. The median Morisita-Horn indices, median distances of Naïve, migratory and pathogenic Th17 repertoires and median TiRP scores were compared by a Wilcoxon test using a significance level of 0.05. For TCR sequence searching and comparison, we calculated the significant Tconv or Treg-biased query sequence using a hypergeometric p-value cutoff of 0.001.

Data availability

The datasets generated during the current study have been deposited in the DDBJ Sequence Read Archive database (accession numbers DRA010311, DRA010223), and NCBI Gene Expression Omnibus (accession number GSE180432).

References

1 Richards, D. M., Kyewski, B. & Feuerer, M. Re-examining the Nature and Function of Self-Reactive T cells. *Trends Immunol* **37**, 114-125 (2016). https://doi.org:10.1016/j.it.2015.12.005

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

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

4 Moon, J. J. *et al.* Quantitative impact of thymic selection on Foxp3+ and Foxp3- subsets of self-peptide/MHC class II-specific CD4+ T cells. *Proc Natl Acad Sci U S A* **108**, 14602-14607 (2011). https://doi.org:10.1073/pnas.1109806108

5 Cebula, A. *et al.* Dormant pathogenic CD4(+) T cells are prevalent in the peripheral repertoire of healthy mice. *Nat Commun* **10**, 4882 (2019). https://doi.org:10.1038/s41467-019-12820-3

6 Levine, A. G., Arvey, A., Jin, W. & Rudensky, A. Y. Continuous requirement for the TCR in regulatory T cell function. *Nat Immunol* **15**, 1070-1078 (2014). https://doi.org:10.1038/ni.3004

7 Li, M. O. & Rudensky, A. Y. T cell receptor signalling in the control of regulatory T cell differentiation and function. *Nat Rev Immunol* **16**, 220-233 (2016). https://doi.org:10.1038/nri.2016.26

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

9 Jordan, M. S. *et al.* Thymic selection of CD4+CD25+ regulatory T cells induced by an agonist self-peptide. *Nat Immunol* **2**, 301-306 (2001). https://doi.org:10.1038/86302

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

11 Aschenbrenner, K. *et al.* Selection of Foxp3+ regulatory T cells specific for self antigen expressed and presented by Aire+ medullary thymic epithelial cells. *Nat Immunol* **8**, 351-358 (2007). https://doi.org:10.1038/ni1444

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

13 Sakaguchi, N. *et al.* Altered thymic T-cell selection due to a mutation of the ZAP-70 gene causes autoimmune arthritis in mice. *Nature* **426**, 454-460 (2003). https://doi.org:10.1038/nature02119

14 Tanaka, S. *et al.* Graded attenuation of TCR signaling elicits distinct autoimmune diseases by altering thymic T cell selection and regulatory T cell function. *J Immunol* **185**, 2295-2305 (2010). https://doi.org:10.4049/jimmunol.1000848

15 Tanaka, A. *et al.* Construction of a T cell receptor signaling range for spontaneous development of autoimmune disease. *J Exp Med* **220** (2023). https://doi.org:10.1084/jem.20220386

16 Brand, D. D., Latham, K. A. & Rosloniec, E. F. Collagen-induced arthritis. *Nat Protoc* **2**, 1269-1275 (2007). https://doi.org:10.1038/nprot.2007.173

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

18 Meyer Zu Horste, G. *et al.* RBPJ Controls Development of Pathogenic Th17 Cells by Regulating IL-23 Receptor Expression. *Cell Rep* **16**, 392-404 (2016). https://doi.org:10.1016/j.celrep.2016.05.088

19 Ohl, K. *et al.* The transcription factor CREM drives an inflammatory phenotype of T cells in oligoarticular juvenile idiopathic arthritis. *Pediatr Rheumatol Online J* **16**, 39 (2018). https://doi.org:10.1186/s12969-018-0253-x

20 Glal, D. *et al.* ATF3 Sustains IL-22-Induced STAT3 Phosphorylation to Maintain Mucosal Immunity Through Inhibiting Phosphatases. *Front Immunol* **9**, 2522 (2018). https://doi.org:10.3389/fimmu.2018.02522

21 Lin, C. C. *et al.* IL-1-induced Bhlhe40 identifies pathogenic T helper cells in a model of autoimmune neuroinflammation. *J Exp Med* **213**, 251-271 (2016). https://doi.org:10.1084/jem.20150568

22 DiToro, D. *et al.* Insulin-Like Growth Factors Are Key Regulators of T Helper 17 Regulatory T Cell Balance in Autoimmunity. *Immunity* **52**, 650-667 e610 (2020). https://doi.org:10.1016/j.immuni.2020.03.013

23 Jha, P. & Das, H. KLF2 in Regulation of NF-kappaB-Mediated Immune Cell Function and Inflammation. *Int J Mol Sci* **18** (2017). https://doi.org:10.3390/ijms18112383

24 Lee, J. Y. *et al.* The transcription factor KLF2 restrains CD4(+) T follicular helper cell differentiation. *Immunity* **42**, 252-264 (2015). https://doi.org:10.1016/j.immuni.2015.01.013

25 Basdeo, S. A. *et al.* Ex-Th17 (Nonclassical Th1) Cells Are Functionally Distinct from Classical Th1 and Th17 Cells and Are Not Constrained by Regulatory T Cells. *J Immunol* **198**, 2249-2259 (2017). https://doi.org:10.4049/jimmunol.1600737

26 Komatsu, N. *et al.* Pathogenic conversion of Foxp3+ T cells into TH17 cells in autoimmune arthritis. *Nat Med* **20**, 62-68 (2014). https://doi.org:10.1038/nm.3432

27 Etori, K. *et al.* Fibroblast growth factor receptor 1 as a potential marker of terminal effector peripheral T helper cells in rheumatoid arthritis patients. *Rheumatology (Oxford)* **62**, 3763-3769 (2023). https://doi.org:10.1093/rheumatology/kead220

28 Zhang, B. *et al.* CD127 imprints functional heterogeneity to diversify monocyte responses in inflammatory diseases. *J Exp Med* **219** (2022). https://doi.org:10.1084/jem.20211191

29 Tang, S. *et al.* Single-cell atlas of human infrapatellar fat pad and synovium implicates APOE signaling in osteoarthritis pathology. *Sci Transl Med* **16**, eadf4590 (2024). https://doi.org:10.1126/scitranslmed.adf4590

30 Yasumizu, Y. *et al.* Single-cell transcriptome landscape of circulating CD4(+) T cell populations in autoimmune diseases. *Cell Genom* **4**, 100473 (2024). https://doi.org:10.1016/j.xgen.2023.100473

31 Tong, L. *et al.* Current understanding of osteoarthritis pathogenesis and relevant new approaches. *Bone Res* **10**, 60 (2022). https://doi.org:10.1038/s41413-022-00226-9

32 Chiffelle, J. *et al.* T-cell repertoire analysis and metrics of diversity and clonality. *Curr Opin Biotechnol* **65**, 284-295 (2020). https://doi.org:10.1016/j.copbio.2020.07.010

33 Dash, P. *et al.* Quantifiable predictive features define epitope-specific T cell receptor repertoires. *Nature* **547**, 89-93 (2017). https://doi.org:10.1038/nature22383

34 Valencia, X. *et al.* TNF downmodulates the function of human CD4+CD25hi T-regulatory cells. *Blood* **108**, 253-261 (2006). https://doi.org:10.1182/blood-2005-11-4567

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

36 Liu, Z. *et al.* Immune homeostasis enforced by co-localized effector and regulatory T cells. *Nature* **528**, 225-230 (2015). https://doi.org:10.1038/nature16169

37 Vento-Tormo, R. *et al.* Single-cell reconstruction of the early maternal-fetal interface in humans. *Nature* **563**, 347-353 (2018). https://doi.org:10.1038/s41586-018-0698-6

38 Logunova, N. N. *et al.* MHC-II alleles shape the CDR3 repertoires of conventional and regulatory naive CD4(+) T cells. *Proc Natl Acad Sci U S A* **117**, 13659-13669 (2020). https://doi.org:10.1073/pnas.2003170117

39 Lu, D. R. *et al.* Dynamic changes in the regulatory T-cell heterogeneity and function by murine IL-2 mutein. *Life Sci Alliance* **3** (2020). https://doi.org:10.26508/lsa.201900520

40 Ko, A. *et al.* TCR Repertoires of Thymic Conventional and Regulatory T Cells: Identification and Characterization of Both Unique and Shared TCR Sequences. *J Immunol* **204**, 858-867 (2020). https://doi.org:10.4049/jimmunol.1901006

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

42 Lagattuta, K. A. *et al.* Repertoire analyses reveal T cell antigen receptor sequence features that influence T cell fate. *Nat Immunol* **23**, 446-457 (2022). https://doi.org:10.1038/s41590-022-01129-x

43 Ooi, J. D. *et al.* Dominant protection from HLA-linked autoimmunity by antigen-specific regulatory T cells. *Nature* **545**, 243-247 (2017). https://doi.org:10.1038/nature22329

44 Yi, J. *et al.* Antigen-specific depletion of CD4(+) T cells by CAR T cells reveals distinct roles of higher- and lower-affinity TCRs during autoimmunity. *Sci Immunol* **7**, eabo0777 (2022). https://doi.org:10.1126/sciimmunol.abo0777

45 Klarenbeek, P. L. *et al.* Inflamed target tissue provides a specific niche for highly expanded T-cell clones in early human autoimmune disease. *Ann Rheum Dis* **71**, 1088-1093 (2012). https://doi.org:10.1136/annrheumdis-2011-200612

46 Al Khabouri, S. *et al.* TCRbeta Sequencing Reveals Spatial and Temporal Evolution of Clonal CD4 T Cell Responses in a Breach of Tolerance Model of Inflammatory Arthritis. *Front Immunol* **12**, 669856 (2021). https://doi.org:10.3389/fimmu.2021.669856

47 Gaublomme, J. T. *et al.* Single-Cell Genomics Unveils Critical Regulators of Th17 Cell Pathogenicity. *Cell* **163**, 1400-1412 (2015). https://doi.org:10.1016/j.cell.2015.11.009

48 Malchow, S. *et al.* Aire Enforces Immune Tolerance by Directing Autoreactive T Cells into the Regulatory T Cell Lineage. *Immunity* **44**, 1102-1113 (2016). https://doi.org:10.1016/j.immuni.2016.02.009

49 Menezes, J. S. *et al.* A public T cell clonotype within a heterogeneous autoreactive repertoire is dominant in driving EAE. *The Journal of Clinical Investigation* **117**, 2176-2185 (2007). https://doi.org:10.1172/JCI28277

50 Zhao, Y. *et al.* Autoimmune susceptibility imposed by public TCRbeta chains. *Sci Rep* **6**, 37543 (2016). https://doi.org:10.1038/srep37543

51 Shugay, M. *et al.* VDJdb: a curated database of T-cell receptor sequences with known antigen specificity. *Nucleic Acids Res* **46**, D419-D427 (2018). https://doi.org:10.1093/nar/gkx760

52 Boyle, L. H., Goodall, J. C. & Gaston, J. S. Major histocompatibility complex class I-restricted alloreactive CD4+ T cells. *Immunology* **112**, 54-63 (2004). https://doi.org:10.1111/j.1365-2567.2004.01857.x

53 Boyle, L. H., Goodall, J. C., Opat, S. S. & Gaston, J. S. The recognition of HLA-B27 by human CD4(+) T lymphocytes. *J Immunol* **167**, 2619-2624 (2001). https://doi.org:10.4049/jimmunol.167.5.2619

54 Singh, N. K. *et al.* A class-mismatched TCR bypasses MHC restriction via an unorthodox but fully functional binding geometry. *Nat Commun* **13**, 7189 (2022). https://doi.org:10.1038/s41467-022-34896-0

55 Kieback, E. *et al.* Thymus-Derived Regulatory T Cells Are Positively Selected on Natural Self-Antigen through Cognate Interactions of High Functional Avidity. *Immunity* **44**, 1114-1126 (2016). https://doi.org:10.1016/j.immuni.2016.04.018

56 Leonard, J. D. *et al.* Identification of Natural Regulatory T Cell Epitopes Reveals Convergence on a Dominant Autoantigen. *Immunity* **47**, 107-117 e108 (2017). https://doi.org:10.1016/j.immuni.2017.06.015

57 Hsieh, C. S. *et al.* Recognition of the peripheral self by naturally arising CD25+ CD4+ T cell receptors. *Immunity* **21**, 267-277 (2004). https://doi.org:10.1016/j.immuni.2004.07.009

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

59 Eggenhuizen, P. J. *et al.* Smith-specific regulatory T cells halt the progression of lupus nephritis. *Nat Commun* **15**, 899 (2024). https://doi.org:10.1038/s41467-024-45056-x

60 Ohkura, N. *et al.* T cell receptor stimulation-induced epigenetic changes and Foxp3 expression are independent and complementary events required for Treg cell development. *Immunity* **37**, 785-799 (2012). https://doi.org:10.1016/j.immuni.2012.09.010

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

62 Stoeckius, M. *et al.* Cell Hashing with barcoded antibodies enables multiplexing and doublet detection for single cell genomics. *Genome Biol* **19**, 224 (2018). https://doi.org:10.1186/s13059-018-1603-1

63 Loza, M., Teraguchi, S., Standley, D. M. & Diez, D. Unbiased integration of single cell transcriptome replicates. *NAR Genom Bioinform* **4**, lqac022 (2022). https://doi.org:10.1093/nargab/lqac022

64 Chen, E. Y. *et al.* Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* **14**, 128 (2013). https://doi.org:10.1186/1471-2105-14-128

65 Street, K. *et al.* Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genomics* **19**, 477 (2018). https://doi.org:10.1186/s12864-018-4772-0

66 Bolotin, D. A. *et al.* MiXCR: software for comprehensive adaptive immunity profiling. *Nat Methods* **12**, 380-381 (2015). https://doi.org:10.1038/nmeth.3364

67 Wilamowski, J. *et al.* InterClone: Store, Search and Cluster Adaptive Immune Receptor Repertoires. *bioRxiv*, 2022.2007.2031.501809 (2022). https://doi.org:10.1101/2022.07.31.501809

Acknowledgements

This work was funded by the Japan Agency for Medical Research and Development (AMED), grant number JP223fa627002 and Platform Project for Supporting Drug Discovery and Life Science Research (Basis for Supporting Innovative Drug Discovery and Life Science Research) under JP21am0101108; and by JSPS KAKENHI Grant Numbers JA23H034980 (to D.S.), 20K16286 (to M.L.C.), and 26860331, 17K15723, 22H02920 (to A.T.).

Author contributions

D.S. and M.L.C. conceived the study. M.L.C., S.T. and A.T. designed experiments. M.L.C., E.L., and A.T performed experiments. M.L.C., M.L.L., D.D., Z.X., performed data analysis. M.L., A.T. and D.S. interpreted data and wrote the paper. S.T. and Z.X. designed custom software for TCRs analysis. D.M. and S.S. supervised the project. All the authors read and approved the final manuscript.

Competing interests

A.T. reported grants from Shionogi & Co., Ltd. outside the submitted work. No other disclosures were reported.

Correspondence and requests for materials should be addressed to D.M. or M.L.C

Figure 1

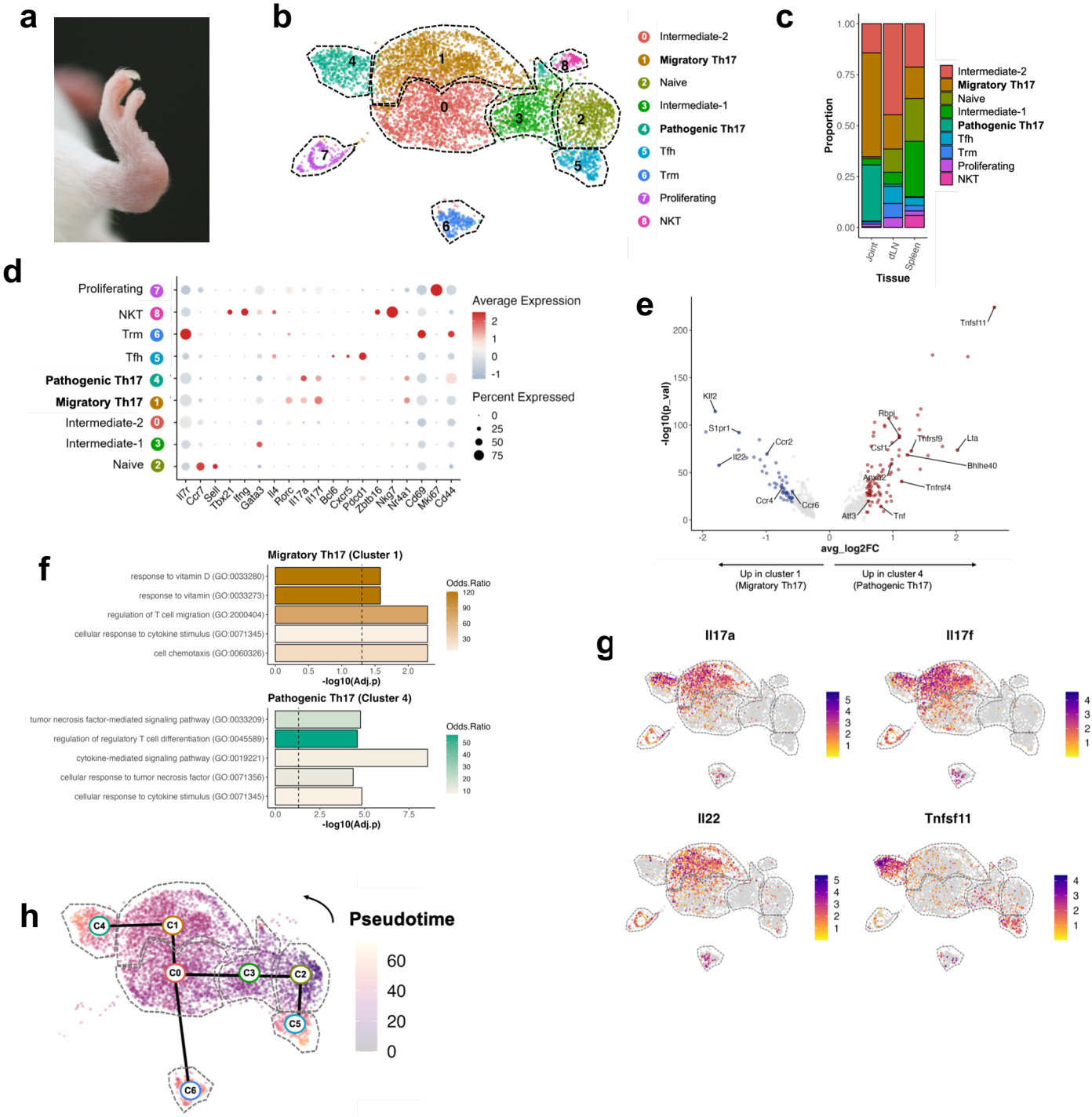

Figure 1. Two transcriptionally different Th17 cell populations exist in inflamed joints of arthritic ZAC mice. a) Lateral view of a hindpaw of a 23-week old arthritic ZAC mouse. **b)** UMAP plot of Tconv cells obtained from joints, draining lymph node (dLN) and spleen of arthritic ZAC mice colored by cluster. Unsupervised clustering was applied by the Louvain algorithm and clusters were annotated as described in the text. **c)** Proportion of annotated Tconv cell populations in joints, dLN and spleen of arthritic ZAC mice. **d)** Relative expression of canonical Tconv markers across gene expression clusters. Circle color and intensity represent the average expression whereas circle size represents the percentage of cells from each cluster expressing the gene. **e)** Volcano plot of the differentially expressed genes (DEG) among cluster 1 (migratory Th17) and cluster 4 (pathogenic Th17). **f)** Top 5 gene Ontology (GO) Terms enriched in DEGs from cluster 1 (migratory Th17) vs cluster 4 (pathogenic Th17). The color intensity of bars 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 *Il17a*, *Il17f*, *Il22* or *Tnfsf11*. **h)** UMAP plot of Tconv cells from tissues of arthritic ZAC mice, with cells colored according to their inferred pseudotime values; cluster centers are connected according to trajectory prediction by Slingshot's pipeline.

Figure 2

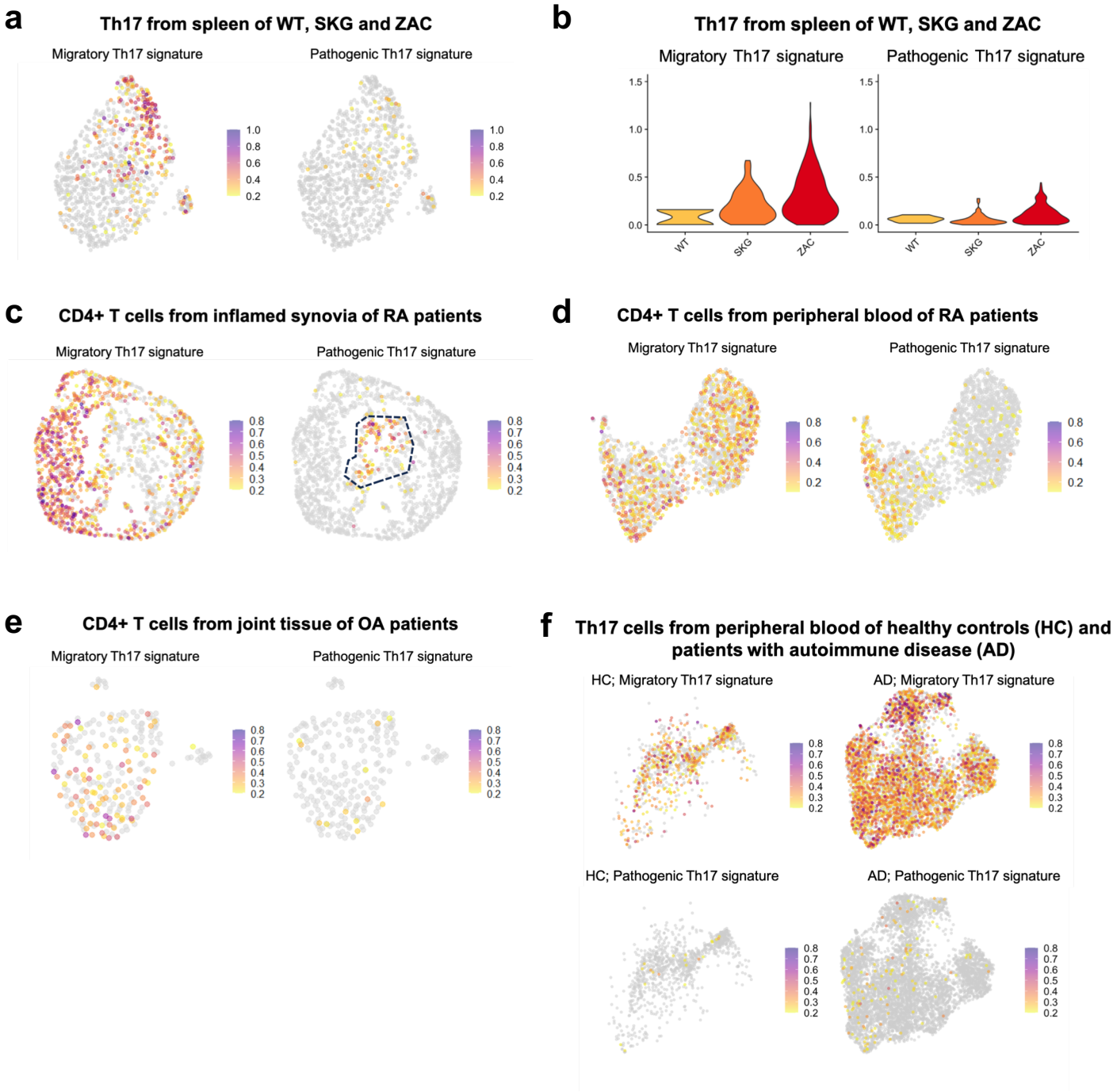

Figure 2. The pathogenic Th17 gene signature is uniquely detected in CD4+ T cells from inflamed synovia of rheumatoid arthritis (RA) patients. Gene signatures of migratory Th17 and pathogenic Th17 cells were obtained and used for calculation of module scores in public scRNAseq datasets of **a)** Th17 cells obtained from spleens of WT, SKG and ZAC mice (n=2 each), **c)** CD4+ T cells from inflamed synovia of RA patients (n=3)[27], **d)** CD4+ T cells from peripheral blood of RA patients (n=1)[28], **e)** CD4+ T cells from synovia and infrapatellar fat pad (IPFP) tissue of osteoarthritis (OA) patients (n=3)[29], **f)** Circulating Th17 cells from healthy donors (n=3), and patients with autoimmune diseases (AD) other than RA: myasthenia gravis (MG) (n=3), multiple sclerosis (MS) (n=4), systemic lupus erythematosus (SLE) (n=3)[30]; **b)** Violin plot of module scores for the gene signatures of migratory Th17 (left) and pathogenic Th17 (right) cells in Th17 cells from spleens of WT, SKG and arthritic ZAC mice.

Figure 3

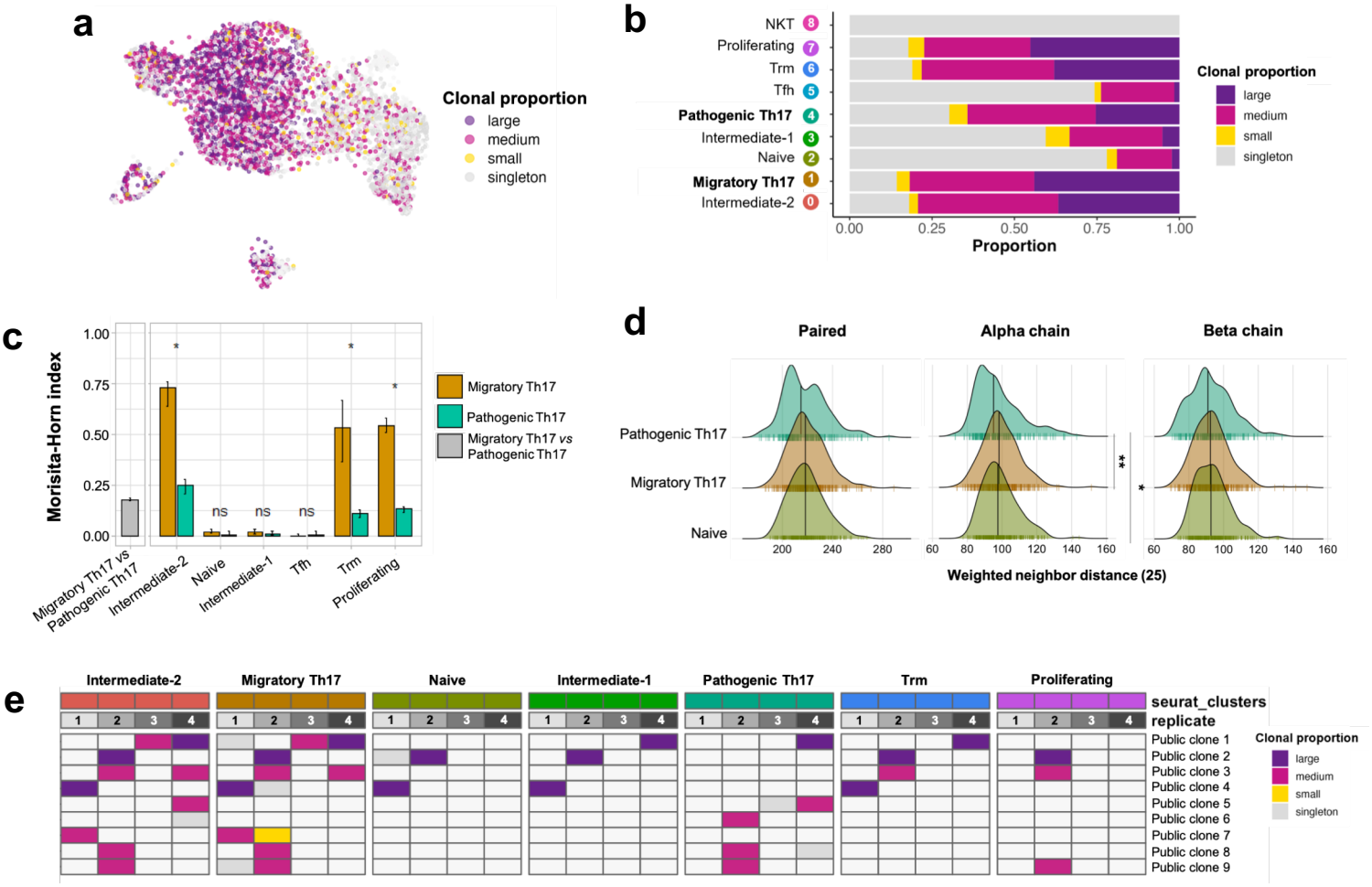

Figure 3. The TCR repertoire of pathogenic Th17 cells is unique and convergent, and contain public clones. a) UMAP plot of Tconv cells from tissues of arthritic ZAC mice, cells are colored according to clonal proportion classification. **b)** Distribution of clonal proportion classes across ZAC Tconv gene expression clusters. **c)** Morisita-Horn index of the paired CDR3 amino acid similitude between the migratory and pathogenic Th17 repertoires (gray bar); and between both Th17 populations vs the other Tconv phenotypes in ZAC tissues (brown and green bars). **d)** Distribution of neighbor distances to the nearest 25th percentile (methods) in the repertoires of Naïve, migratory Th17 and pathogenic Th17 cells. Paired (left) and single alpha (middle) and beta (right) TCR chain distributions are shown. **e)** Public clones and their clonal proportion classification and distribution across Tconv phenotypes in each mouse replicate. *: $p<0.05$, **$p<0.01$; Wilcoxon Test.

Figure 4

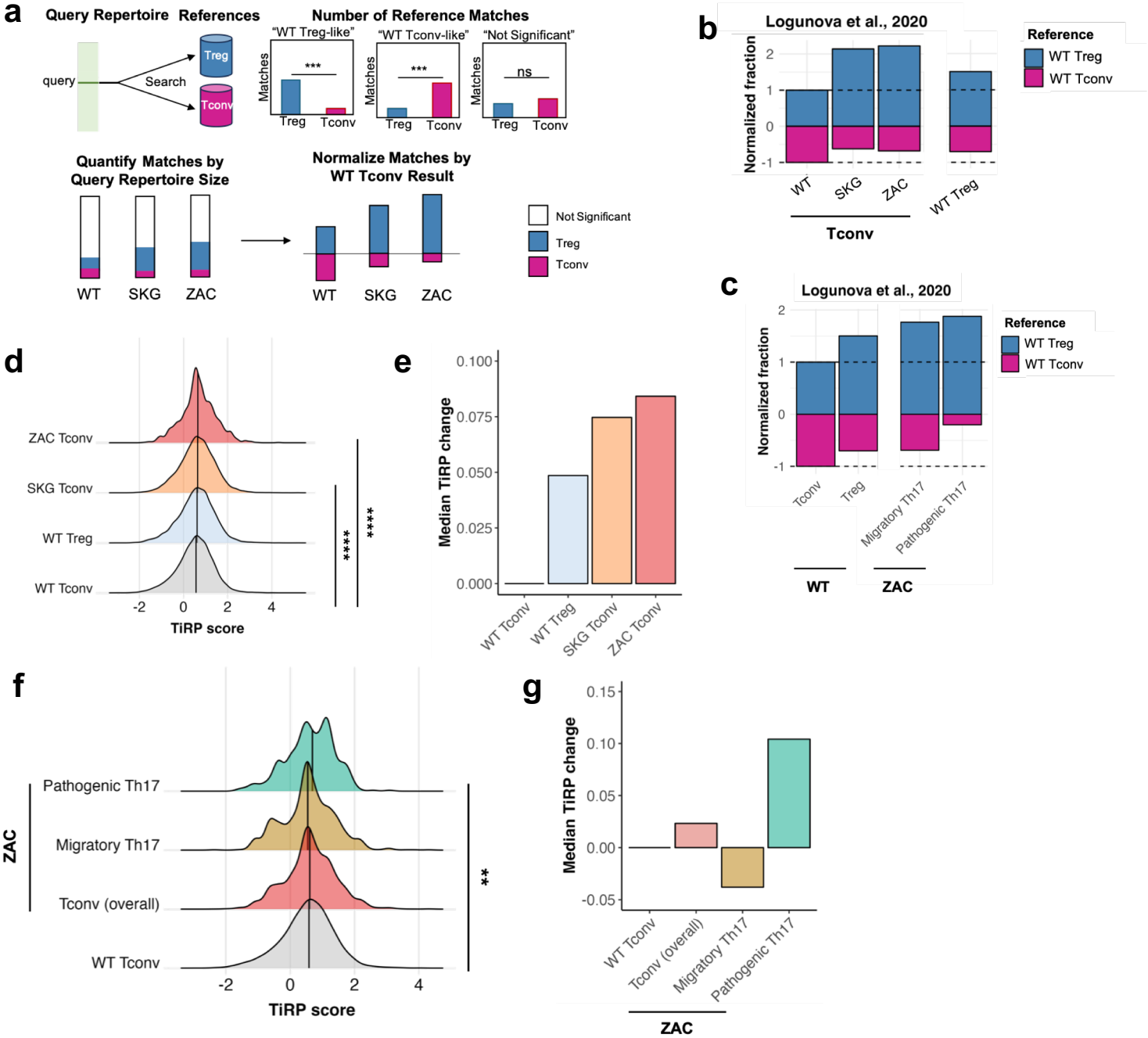

Figure 4. TCR repertoires of ZAP-70 mutant Tconv cells, especially of pathogenic Th17 cells, are more WT Treg-like than WT Tconv-like. a) Description of the method employed for the assessment of repertoire similitude to reference WT Treg and WT Tconv repertoires. Each of our experimentally determined query TCRs were used to search CDR matches (methods) in a reference WT Tconv or WT Treg dataset[38-41]**.** We next computed the number of reference WT Tconv and reference WT Treg matches and deemed the query TCR "WT Treg-like", "WT Tconv-like" or "not significant" using a hypergeometric p-value cutoff of 0.001. The total number of matches was calculated in terms of the query repertoire size and results were normalized to the matches in the query WT Tconvs. **(b-c)** Normalized fraction of query WT Tconv and WT Treg as compared to SKG and ZAC Tconv repertoires **(b)** or ZAC migratory and pathogenic Th17 repertoires **(c)** matching with reference WT Tconv (magenta section, negative values) or WT Treg repertoires (blue section, positive values) obtained from a representative reference dataset[38]. **d)** Distribution of TCR-intrinsic regulatory potential (TiRP) scores for Tconv repertoires of WT, SKG and ZAC mice and the WT Treg repertoire. **e)** Median TiRP score change of WT Treg, and SKG and ZAC Tconv repertoires relative to the WT Tconv repertoire. **f)** Distribution of TiRP scores for WT Tconv, and overall Tconv, migratory Th17, and pathogenicTh17 repertoires of ZAC mice. **g)** Median TiRP score change of overall Tconv, migratory Th17, and pathogenicTh17 repertoires of ZAC mice relative to WT Tconv. **p<0.01, ****: p<0.0001; Wilcoxon Test

Figure 5

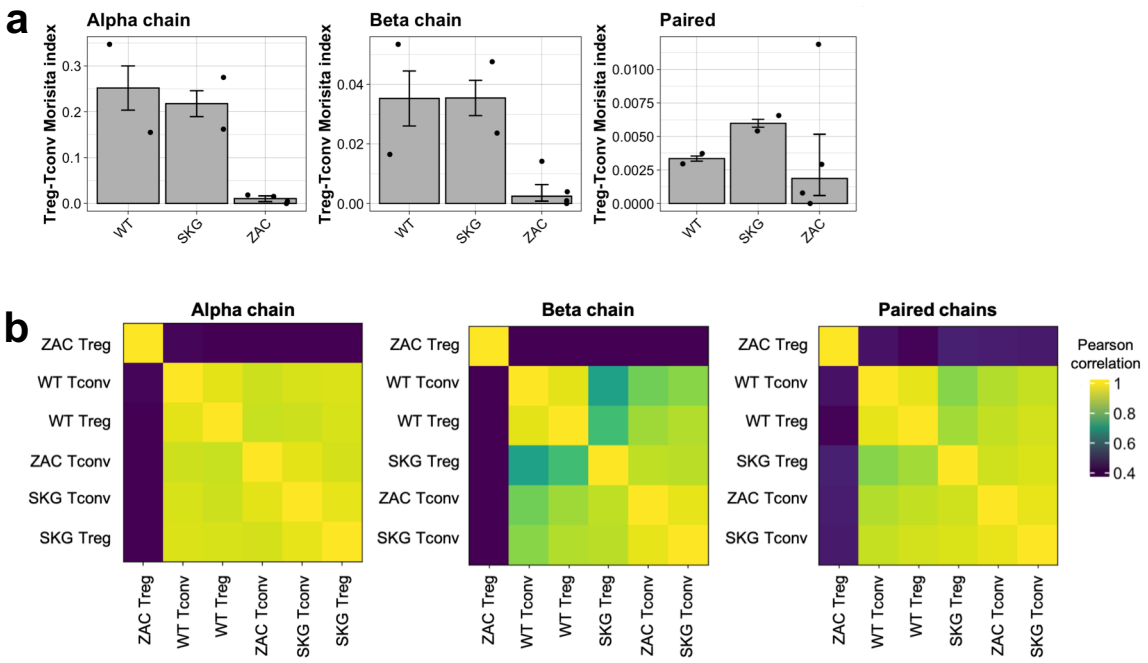

Figure 5. Intra-Treg-Tconv similitude in WT, SKG and ZAC repertoires. a) Morisita-Horn index calculated at the exact CDR3 amino acid level between the Tconv and Treg compartments of WT, SKG and ZAC mice. Left panel shows the analysis for single alpha chains, middle panel shows single beta chains and right panel depicts paired chains. **b)** Heatmap of weighted neighbor distance correlations between WT, SKG and ZAC Treg and Tconv repertoires as calculated by TCRdist (methods), higher correlation scores indicate closer repertoires.

Figure 6

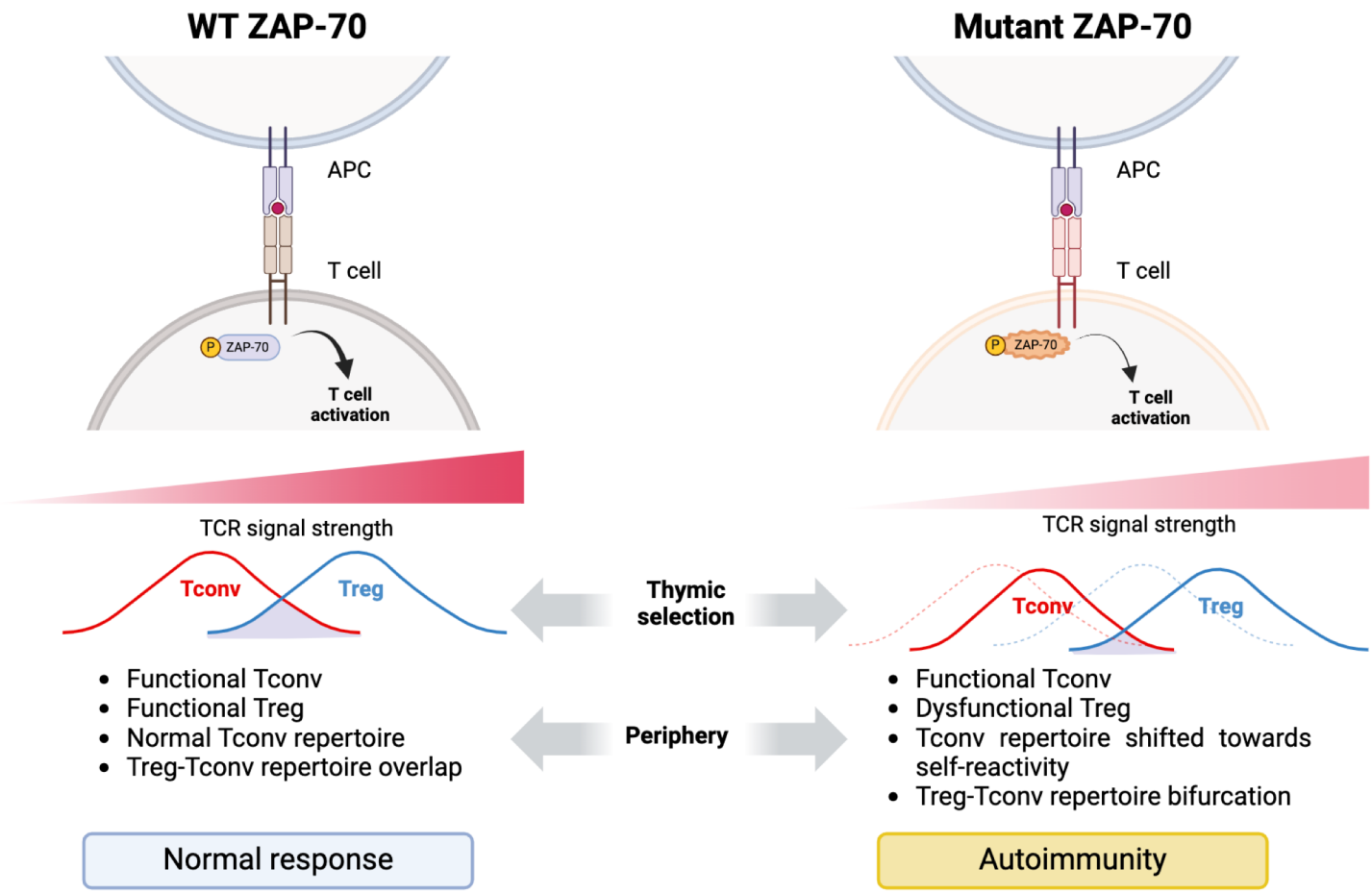

Figure 6. Consequences of weakened TCR signaling in the shaping of Treg and Tconv repertoires and autoimmunity. The SKG (W163C) and ZAC (H165A) mutations in the ZAP-70 gene cause a reduction in its affinity for phosphorylated CD3z within a range that results in the development of autoimmune disease. The consequences of a weakened TCR signaling in the development of autoimmune disease are multilayered. On one hand, cells with self-reactive TCRs are aberrantly selected into the Tconv compartment instead of the Treg compartment (repertoire shift). This defective selection also deprives the Treg repertoires of specific receptors potentially critical for the proper suppression of autoimmune responses and this is evidenced by a reduced similarity between Treg and Tconv repertoires within an individual (repertoire bifurcation). In addition, Treg effector activity upon TCR weakening is significantly disrupted, rendering Tregs insufficient for the suppression of self-reactive Tconvs. These events together, favor the activation of self-reactive T cells, their differentiation into pathogenic phenotypes and their accumulation in target tissues. Created with BioRender.com