Title：Age-associated CD4$^+$ T cells with B cell-promoting functions are regulated by ZEB2 in autoimmunity

Authors: Manaka Goto, M.D.[1†], Hideyuki Takahashi, M.D., Ph.D.[1†], Ryochi Yoshida, M.D., Ph.D.[1†], Takahiro Itamiya, M.D., Ph.D.[1, 2], Masahiro Nakano, M.D., Ph.D.[1, 3, 4], Yasuo Nagafuchi, M.D., Ph.D.[1, 2], Hiroaki Harada, M.D., Ph.D.[1], Toshiaki Shimizu, M.D., Ph.D.[1], Meiko Maeda, M.D., Ph.D.[5], Akatsuki Kubota, M.D., Ph.D.[5], Tatsushi Toda, M.D., Ph.D.[5], Hiroaki Hatano, M.D., Ph.D.[1, 3], Yusuke Sugimori, M.D., Ph.D.[1], Kimito Kawahata, M.D., Ph.D.[6], Kazuhiko Yamamoto, M.D., Ph.D.[4], Hirofumi Shoda, M.D., Ph.D.[1], Kazuyoshi Ishigaki, M.D., Ph.D.[3], Mineto Ota, M.D., Ph.D.[1]*, Tomohisa Okamura, M.D., Ph.D.[1, 2]*‡, Keishi Fujio, M.D., Ph.D.[1]*

Affiliations:

[1] Department of Allergy and Rheumatology, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan.

[2] Department of Functional Genomics and Immunological Diseases, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan.

[3] Laboratory for Human Immunogenetics, RIKEN Center for Integrative Medical Sciences, Yokohama, Kanagawa, Japan.

[4] Laboratory for Autoimmune Diseases, RIKEN Center for Integrative Medical Sciences, Yokohama, Kanagawa, Japan.

[5] Department of Neurology, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan.

[6] Department of Rheumatology and Allergology, St. Marianna University School of Medicine, Kawasaki, Kanagawa, Japan

† These authors contributed equally to this work.

‡ Lead contact.

*Corresponding authors. Email:

Tomohisa Okamura, M.D., Ph.D. E-mail: okamura@m.u-tokyo.ac.jp

Keishi Fujio, M.D., Ph.D. E-mail: FUJIOK-INT@h.u-tokyo.ac.jp

Mineto Ota, M.D., Ph.D. E-mail: mineto@stanford.edu

Abstract: Aging is a significant risk factor for autoimmunity, and many autoimmune diseases tend to onset during adulthood. We conducted an extensive analysis of $CD4^+$ T cell subsets from 354 autoimmune disease patients and healthy controls via flow cytometry and bulk RNA sequencing. As a result, we identified a distinct $CXCR3^{mid}CD4^+$ effector memory T cell subset that expands with age, which we designated "age-associated helper T (ThA) cells". ThA cells exhibited both a cytotoxic phenotype and B cell helper functions, and these features were regulated by the transcription factor ZEB2. Consistent with the highly skewed T cell receptor usage of ThA cells, gene expression in ThA cells from systemic lupus erythematosus patients reflected disease activity and was affected by treatment with a calcineurin inhibitor. Moreover, analysis of single-cell RNA sequencing data revealed that ThA cells infiltrate damaged organs in patients with autoimmune diseases. Together, our characterization of ThA cells may facilitate improved understanding of the relationship between aging and autoimmune diseases.

One Sentence Summary: An age-associated $CXCR3^{mid}CD4^+$ T cell subset with a cytotoxic phenotype and B cell helper functions reflects disease activity in lupus.

Main Text:

INTRODUCTION

Autoimmune diseases are induced by breakdown of self-tolerance in the immune system (*1*), with aging representing a major risk factor (*2*, *3*). For example, the incidences of rheumatoid arthritis (RA) and systemic lupus erythematosus (SLE), two of the most common autoimmune diseases, are increased in middle or older ages (*4–7*). Although tremendous efforts have been made in genetic studies of autoimmune diseases, and associations of these diseases with inherited variants have been identified, the reason for their acute onset in adulthood remains unclear. Identifying the specific immune cells involved in aging and autoimmunity and the drugs that target these subsets is likely to aid the development of precision medicine approaches.

T cell cytotoxicity and B cell activation are heavily involved in the pathogenesis of autoimmune diseases (*1*, *8*), with the production of autoantibodies often the defining feature. Autoantibodies are produced by memory B cells or plasma cells, the differentiation of which is supported by $CD4^+$ helper T cells expressing CXCL13 and interleukin (IL)-21 (*9*, *10*). The maturation of B cells generally takes place in the follicles of lymphoid tissues (*11*, *12*). However, the extrafollicular pathway has gained attention due to its role in the differentiation of autoantibody-producing B cells (*13*, *14*). For example, $CXCR5^-PD\text{-}1^{hi}$ peripheral helper T (Tph) cells are involved in the extrafollicular pathway leading to B cell activation (*15*). The function of Tph cells has been extensively investigated in various autoimmune diseases (*16–18*). The main population of B cells that produce autoantibodies in the extrafollicular pathway are likely double negative 2 (DN2), defined as $CD27^-IgD^-CXCR5^-CD11c^+$ (*13*, *14*). DN2 B cells overlap with "age-associated B cells (ABCs)" that increase with age (*19*), perhaps contributing to the age-

related risk for autoimmune diseases. However, the relationship between extrafollicular helper T cells and aging remains unclear.

T cell cytotoxicity is also involved in autoimmune tissue damage. In addition to widely recognized $CD8^+$ cytotoxic T lymphocytes (CTLs) (*20*, *21*), $CD4^+$ CTLs are increased in the blood of autoimmune patients (*22–24*) and infiltrate inflamed organ sites in autoimmune diseases (*25*, *26*). Additionally, the amount of $CD4^+$ CTLs infiltrating organs correlates with disease activity (*27*). Thus, $CD4^+$ CTLs have gained attention for their possible involvement in the pathogenesis of autoimmune diseases. The increased $CD4^+$ CTLs in healthy elderly individuals have also been investigated for their contribution to longevity and biological defense (*28*, *29*). However, the connection between $CD4^+$ CTLs in autoimmune diseases and those in the elderly have not been delineated.

We recently constructed a database termed the "Immune Cell Gene Expression Atlas from the University of Tokyo (ImmuNexUT)", which consists of gene expression and genomic data from a wide range of peripheral blood (PB) immune cells from healthy individuals and patients with autoimmune diseases (*30*). In the present study, we identify age-associated helper T (ThA) cells, a subset of effector memory $CD4^+$ T cells that increases with age and is defined by a $CXCR3^{mid}$ immunophenotype. Notably, ThA cells possess characteristics of both extrafollicular helper T cells and $CD4^+$ CTLs that shape the pathogenesis of autoimmune diseases, suggesting a shared basis for aging and the development of autoimmunity. We also found that the gene expression profile of ThA cells is associated with SLE disease activity and is affected by administration of calcineurin inhibitors. Together, these findings may provide insight into the T cell-driven mechanisms behind age-related induction of autoimmune diseases.

RESULTS

A distinct CXCR3^{mid}CD4^{+} T cell subset is associated with age

Our previously developed ImmuNexUT database contains 28 immune cell subsets derived from healthy controls (HC) and various autoimmune disease patients, including those with RA, SLE, and idiopathic inflammatory myopathy (IIM) (**Fig. 1A, Table S1**) (*30*). Effector memory CD4^{+} T cells (CD4^{+} Tem: CD3^{+}CD4^{+}CCR7^{-}CD45RA^{-}) are strongly involved in immune responses at sites of inflammation (*31*). Within the CD4^{+} Tem population, we noted a phenotype (CXCR3^{mid}CD25^{-}CXCR5^{-}CCR6^{-}CCR4^{-}) that did not match those of any well-described CD4^{+} T cell subset (**Fig. 1B, Table S2, S3**). To evaluate the characteristics of this CXCR3^{mid}CD4^{+} T cell subset, we compared it with eight other well-defined CD4^{+} T subsets: naïve CD4, T helper type (Th) 1, Th2, Th17, Tfh, fraction I naïve regulatory T, fraction II effector regulatory T, and fraction III non-regulatory T cells. In total, 1,562 bulk RNA-seq samples and 354 samples of flow cytometric data from PB mononuclear cells (PBMCs) from HCs and RA, SLE, and IIM patients were analyzed together (**Fig. 1A**). Principal component analysis (PCA) of the total RNA-seq data showed that the CXCR3^{mid}CD4^{+} T cell subset formed a distinct cluster independent of other CD4^{+} T cell subsets and was adjacent to a cluster of Th1 cells (**Fig. 1C**). Intriguingly, among all nine CD4^{+} T cell subsets, only the proportion of the CXCR3^{mid}CD4^{+} T cell subset was positively correlated with age, and this correlation was independent of sex or therapeutic steroid dose (**Fig. 1D, 1E, and fig. S1**). Based on these findings, we termed this CXCR3^{mid}CD4^{+} T cell subset "age-associated helper T (ThA) cells". The proportion of ThA cells was positively correlated with age in HCs and RA, SLE, and IIM patients (**Fig. 1F**).

In order to compare ThA cells and the previously described Tph cells, we examined flow cytometry data from 49 autoimmune disease patients and healthy controls. Almost 10% of ThA

cells overlapped phenotypically with Tph cells (**fig. S2A-C**). Intriguingly, unlike ThA cells (**fig. S1**), the proportion of Tph cells showed no correlation with age (**fig. S2D**), which is consistent with prior findings (*15*). Together, these findings suggest that ThA cells largely represented a distinct population from Tph cells.

ThA cell signature is associated with a cytotoxic phenotype

To evaluate the differences in gene expression among $CD4^+$ T cell subsets, we examined the top specifically expressed genes (SEGs) of each subset in RNA-seq using an algorithm for detecting specific expression focused on entropy (*32*). We found that ThA cells specifically expressed the genes encoding cytotoxic molecules, including *GZMA* and *GZMB*, suggesting that ThA cells may be $CD4^+$ CTLs (**Fig. 2A, Data file S1**). In addition, *ZNF683* was in the top SEGs of ThA cells and encodes Hobit, a transcription factor that is characteristically expressed in $CD4^+$ CTLs (*33*). *ADGRG1*, also in the top ThA SEGs, encodes GPR56, which also regulates cytotoxicity (*34*). ThA cells highly expressed genes related to CTL activity including *GZMA*, *GZMB*, *GZMH,* and *PRF1* (**Fig. 2B**). Other characteristics of $CD4^+$ CTLs (*29*, *35*) such as lack of expression of CD28 and expression of CX3CR1, KLRG1, and NKG7, were also observed in ThA cells both in the HCs and autoimmune disease patients (**Fig. 2C**). Indeed, intracellular staining of HC PBMC revealed that ThA cells expressed high levels of granzyme A, granzyme B, and perforin proteins compared with naïve $CD4^+$ T cells, Th1 cells, Th2 cells, Th17 cells, and Tfh cells, similar to that of $CD8^+$ T cells (**Fig. 2D,E**). Interestingly, the $CXCR3^{mid}$ cell population corresponding to ThA cells markedly exhibited higher levels of granzyme A, granzyme B, and perforin proteins than neighboring $CXCR3^{neg}$ and $CXCR3^{hi}$ cells (**Fig. 2F**), suggesting that $CD4^+$ T cells with moderate expression of CXCR3 have potent CTL activity. Moreover, ThA cells produced high levels of

granzyme A protein after T cell receptor (TCR) stimulation (**fig. S3A**). Cytotoxic granules have been observed as multi-core granules (MCGs) in the intracellular structure of CTLs (*36*). Our transmission electron microscopic observation also revealed the presence of MCGs in ThA cells and $CD8^+$ T cells, but not in naïve $CD4^+$ T cells or Th1 cells (**Fig. 2G**). Indeed, ThA cells exhibited cytotoxic activity toward K562 cells (**fig. S3B**). These results demonstrate that ThA cells are $CD4^+$ CTLs delineated by a $CXCR3^{mid}$ chemokine receptor profile.

ThA cells exert B cell helper functions

Next, we explored the disease-associated phenotypic changes in ThA cells by comparing RA patients and HCs. Age- and sex-matched comparisons showed that the proportion of ThA cells among PB $CD4^+$ T cells was increased in RA patients compared with HCs (**Fig. 3A**). Evaluation of differentially expressed genes (DEGs) in RNA-seq revealed that *CXCL13* and *IL21* were upregulated in RA-derived ThA cells (**Fig. 3B**). CXCL13 is involved in B cell recruitment as a ligand for CXCR5 and is an important effector molecule for the B cell helper function of $CD4^+$ T cells, along with IL-21 (*15*, *18*). Consistent with their gene expression, ThA cells expressed CXCL13 protein upon TCR stimulation (**Fig. 3C, and fig. S3C**). Co-culture of RA-derived PB $CD4^+$ T cell subsets and B cells revealed that PB ThA cells induced antibody production by B cells comparable with PB Tfh cells (**Fig. 3D**). Antibody production was partially suppressed by neutralization of IL-21 and completely suppressed by blocking CD84 (SLAMF5) (**Fig. 3E**), a cell surface protein essential for T-B interaction (*37*). Comparable results were observed with HC ThA cells *in vitro* (**fig. S3D, S3E**). When direct contact between ThA and B cells was prevented by Transwell plates, such B cell helper function by ThA cells disappeared (**Fig. 3F**).

Consequently, direct cell-to-cell contact and subsequent signaling via CD84 and IL-21 are needed for the B cell helper function of ThA cells.

Next, we examined ThA cells in SLE, which is a prototypical autoimmune disease characterized by a variety of autoantibodies and diverse organ lesions in the skin, joints, and kidneys. Additionally, type I interferon (IFN) plays a pivotal role in SLE pathogenesis (*38*). The proportion of ThA cells in PB $CD4^+$ T cells was increased in SLE patients compared with HCs (**Fig. 3G**). *CXCL13* was also one of the top DEGs in ThA cells in SLE patients versus HCs (**Fig. 3H**), and *CXCL13* and *IL21* expression in SLE-derived ThA cells was positively correlated with the SLE disease activity index 2000 (SLEDAI-2K) (*39*) (**Fig. 3I**), suggesting that B cell helper function of ThA cells is involved in the pathogenesis of SLE. In addition, *CXCL13* and *IL21* expression in SLE-derived ThA cells was positively correlated with the IFN-stimulated gene (ISG) score (*40*) (**Fig. 3J**), implying that cytokines, such as type I IFN, are related to the B cell helper function of ThA cells in SLE. Indeed, following anti-CD3 monoclonal antibody (mAb) stimulation, the presence of IFN-α markedly the proportion of ThA cells producing CXCL13, whereas Tfh cells did not show such a response (**Fig. 3K, fig. S3F**).

We performed an analogous analysis of ThA cells derived from IIM patients. IIM is an autoimmune disease characterized by chronic inflammation of skeletal muscle leading to muscle weakness with extensive involvement of skin, joints, and lungs, in which autoantibodies substantially determine the phenotype (*41*). Although there was no increase in the proportion of ThA cells among $CD4^+$ T cells in the PB of IIM patients (**fig. S4A**), the DEGs between IIM-derived ThA cells and HC-derived ThA cells also contained *CXCL13* (**fig. S4B**), suggesting that ThA cells in IIM patients contribute to humoral immunity. When the expression of *CXCL13* and *IL21* of each $CD4^+$ T cell subset was compared between patients with new-onset IIM and the

other IIM patients, *CXCL13* and *IL21* expression of ThA cells was upregulated in patients with new-onset IIM (**fig. S4C**). Thus, the B cell helper function of ThA cells may contribute to the initial pathophysiology of IIM.

ZEB2 and *TBX21* are signature genes of ThA cells

The TCR defines the antigen specificity of T cells, and T cells possessing autoreactive TCRs play a central role in the development of autoimmunity (*42*). TCR repertoire analysis of the nine $CD4^+$ T cell subsets included here (**Fig. 1A**) revealed a marked reduction in TCR diversity (increased Gini index) in ThA cells both in HC and autoimmune diseases (**Fig. 4A**). The TCR repertoire of ThA cells most overlapped with that of Th1 cells (**Fig. 4B**). Th1 cells are helper T cells that activate macrophages and NK cells by producing IFN-γ and are characterized by high, but not mid, expression of CXCR3 on the cell surface (*43*). The overlapping TCR repertoire was consistent with our PCA observations that showed the ThA cell cluster was nearest to the Th1 cell cluster (**Fig. 1C**). Therefore, we evaluated DEGs between ThA cells and Th1 cells. Two transcription factors *ZEB2* and *TBX21* were identified as upregulated genes in ThA cells (**Fig. 4C**). *TBX21* encodes the transcription factor T-bet, the master regulatory gene of Th1 cells (*44*, *45*). Surprisingly, ThA cells expressed about five-times more *TBX21* than Th1 cells (**Fig. 4D**). In contrast, *GATA3, RORC,* and *FOXP3*, which are crucial for Th2, Th17, and Treg cell differentiation, respectively, were not highly expressed in ThA cells. Consistent with the expression of cell surface markers in flow cytometry, gene expression of *CXCR3* in ThA cells was moderate and lower than that in Th1 cells. Of note, IFN-γ production by stimulated ThA cells *in vitro* was comparable with that of Th1 cells (**fig. S5A**).

ZEB2 was also highly expressed in ThA cells among $CD4^+$ T cells (**Fig. 4D**). ZEB2 belongs to the zinc-finger E homeobox-binding family of transcription factors and contributes to the epithelial to mesenchymal transition (*46*). The function of ZEB2 in $CD4^+$ T cells is not well-studied (*47*); however, a recent study showed that ZEB2 expression in mouse Th1 cells transforms them into CTLs (*48*). In our expression quantitative trait locus (eQTL) analysis (*30*), *ZEB2* and *TBX21* were both included in the gene cluster that regulates eQTL effects in memory $CD4^+$ T cells. Expression of this gene cluster increased with age, and the eQTL effects of the regulated genes increased with increasing levels of *ZEB2* and *TBX21*, suggesting that these two transcription factors may be functionally relevant. Using weighted gene correlation network analysis (WGCNA), we identified 4 gene regulatory modules in ThA cells whose expression was significantly and positively correlated with aging (**Fig. 4E**). *ZEB2* and *TBX21* were identified at the center of the gene regulatory network of the age-associated module M22 (**Fig. 4F**). There were no significant differences in the expression of *ZEB2*, *TBX21*, *GZMA*, *CXCL13*, or *IL21* in ThA cells between younger and older individuals (**fig. S6A**).

Next, we conducted single-cell RNA sequencing (scRNA-seq) analysis of PB $CD4^+$ T cells from HCs, which generated 10 clusters (**fig. S5B**). Among them, we identified a ThA cluster with various $CD4^+$ CTL characteristics (expressing granzymes, perforin, *CX3CR1*, *KLRG1*, *NKG7*, *ADGRG1*; lacking expression of *CXCR5*, *FOXP3*, *CD27*, *CD28*) and high expression of *ZEB2* and *TBX21* (**fig. S5C, S5D, Data file S2**). Another $CD4^+$ CTL cluster was observed along with the ThA cluster, likely cytotoxic Th17 (*49–51*), since these cells expressed high levels of *RORC* and *CCR6*. *ADGRG1*, encoding GPR56 and identified in ThA SEGs using RNA-seq (**Fig. 2A**), was also specifically expressed by the ThA cluster in our scRNA-seq analysis (**fig. S5D, Data file S2**). We simultaneously performed cellular indexing of transcriptomes and epitopes by

sequencing (CITE-seq) to evaluate cell surface protein expression using antibody-derived tags. The distributions of CXCR3 protein and mRNA showed strong concordance (**fig. S5E**). Consistent with our flow cytometry data (**Fig. 1B**), CITE-seq showed that the ThA cell cluster also exhibited moderate expression of CXCR3 protein and the effector memory ($CD45RA^{lo}CCR7^{neg}$) phenotype (**fig. S5F**). $CD4^+$ T cells expressing GPR56 were $CXCR3^{mid}$ and mostly belonged to the ThA cell cluster (**fig. S5G**). Flow cytometry of HC PBMC showed expression of GPR56 specifically in the ThA cell fraction of $CXCR3^{mid}$ cells (**fig. S5H**). Thus, data obtained from bulk RNA-seq, scRNA-seq, CITE-seq, and flow cytometry analysis exhibited concordance.

ThA cell numbers increased with age and expressed high levels of genes characteristic of exhausted T cells, such as *PDCD1* (**fig. S6B**), but were not considered to be fully exhausted because they retained cytokine-secreting ability. ThA cells also expressed high levels of *KLRG1*, *KLRK1* (coding NKG2D), and *CD244* (coding 2B4) (**fig. S6B**), which are expressed by senescent T cells, but their proliferative ability upon TCR stimulation was not anergic (**fig. S6C**), and no significant telomere shortening was observed (**fig. S6D**). Thus, ThA cells are unlikely to be senescent or functionally exhausted.

Because ThA cells are age-related T cells that are functionally preserved, they may be involved in the regulation of infectious diseases in the elderly. The potential importance of $CD4^+$ CTLs has been recognized in COVID-19 patients, leading us to further investigate ThA cells in COVID-19 (*52*, *53*). In previously published data from hospitalized COVID-19 patients (*54*), PB memory $CD4^+$ T cells included ThA cells, which expressed *ZEB2*, *GZMA*, and *ADGRG1* but not *CD28* (**fig. S6E**). The late inflammatory response (day 17-23 after symptom onset) represents a turning point in determining the severity of SARS-CoV-2 infection (*54*), at which

the ThA cluster appeared to reach a maximum (**fig. S6F**), suggesting that ThA cells may also be involved in the severity of COVID-19.

ZEB2 regulates the function of ThA cells

Our data suggests roles of the two transcription factors *ZEB2* and *TBX21* in ThA cells, leading us to analyze the function of both genes *in vitro*. When we overexpressed *TBX21* in Jurkat cells, a model human $CD4^+$ T cell line (*55*), *CXCL13* was modestly upregulated, but there were no apparent changes in the expression of *ZEB2* or cytotoxicity-related genes such as *GZMA* and *PRF1* (**Fig. 4G**). When *ZEB2* was overexpressed in the same manner, we observed induction of *GZMA* expression but no increase in *TBX21* or *CXCL13* expression (**Fig. 4H**). Since *ZEB2* or *TBX21* alone could not reproduce ThA cell-associated changes in gene expression, we evaluated the effects when both transcription factors were overexpressed simultaneously. Co-expression of *ZEB2* and *TBX21* upregulated not only the cytotoxicity-related genes *GZMA* and *PRF1*, but also *CXCL13* (**Fig. 4I**). Additionally, when primary human $CD4^+$ T cells were examined, simultaneous ectopic expression of *ZEB2* and *TBX21* upregulated *GZMA* and *CXCL13* expression (**Fig. 4J**).

To further examine the roles of *ZEB2* and *TBX21* in ThA cells, we performed knockout experiments in human ThA cells using the CRISPR/Cas9 system. Expression of *GZMA, PRF1,* and *CXCL13* was downregulated when *ZEB2* was knocked out (**Fig. 4K**). In contrast, knockout of *TBX21* did not result in the downregulation of these genes, suggesting that the *ZEB2* is responsible for maintaining the function of differentiated ThA cells. Interestingly, knockout of *ZEB2* alone reduced *TBX21* expression (**Fig. 4K**), suggesting that *ZEB2* is required to maintain high *TBX21* expression in ThA cells.

Gene expression profiles of ThA cells reflect the disease activity of SLE

We recently reported the relationship between immune cell subsets and SLE therapeutic agents, examining the clinical parameters of SLE with RNA-seq gene expression profiles of immune cell subsets in PBMC (*57*), but did not examine ThA cells. Here, we performed an analogous analysis of SLE and HC using bulk RNA-seq of nine subsets of $CD4^+$ T cells including ThA cells. To quantify the transcriptomic patterns established in SLE, we evaluated the expression of 100 IFN-related genes (IRG) previously identified in a scRNA-seq analysis of SLE PBMC (*58*). In hierarchical clustering, ThA cells had the closest IRG expression profiles to those of Th1 cells (**Fig. 5A**). Cluster 1 (C1) of IRG was highly expressed in HC ThA cells and SLE ThA cells, suggesting that C1 is fundamentally related to ThA cells. The variability of gene expression data for every immune cell subset was greater within SLE than within HC in our previous report (*57*), and this was also the case for ThA cells (**Fig. 5B**), implying that there are strong transcriptomic perturbations in SLE ThA cells as well as other T cell subsets.

Next, we quantified the contribution of clinical parameters (sex, age, immunosuppressant, prednisolone (PSL), and disease activity) to the variations in gene expression within SLE using weighted variance partitioning analysis. SLE disease activity had the largest effect on gene expression variation in SLE ThA cells, which was also the highest among the nine $CD4^+$ T cell subsets (**Fig. 5C**). We then calculated DEGs for each subset, comparing case-controls (reflecting disease status) and between SLE in remission and SLE with high disease activity (reflecting disease activity). The total number of DEGs in ThA cells was the second highest after Th1 cells (**Fig. 5D**). Additionally, the majority of DEGs in ThA cells were associated with disease activity. Thus, gene expression in SLE-derived ThA cells closely reflected the disease activity among

$CD4^+$ T cells. Enrichment analysis of pathways showed a similar pattern in ThA and Th1 cells for both metabolism- and cellular process-related pathways (**Fig. 5E**). Specifically, ThA cells were enriched in oxidative phosphorylation signaling and cell cycling as disease activity signatures. Thus, we anticipate that ThA cells likely proliferate and undergo increased oxidative phosphorylation in association with disease activity in SLE.

In SLE, specific therapeutic agents should be tailored to reduce disease pathophysiology. However, it remains unclear how to optimally target the specific immune cell subsets involved in certain aspects of SLE pathophysiology (*38*). Therefore, we evaluated the contribution of each organ lesion and therapeutic drug to the variations in gene expression using weighted variance partitioning analysis. Gene expression variation in ThA cells was significantly associated with the presence of many organ lesions (**Fig. 5F**). PSL, hydroxychloroquine (HCQ), and tacrolimus (TAC) significantly altered the gene expression of ThA cells (**Fig. 5G**), suggesting that ThA cells might be important targets of these agents, especially TAC, in SLE. In contrast, the use of mycophenolate mofetil (MMF) significantly affected the gene expression of Th1 and Th2 cells.

Serum levels of CXCL13 are correlated with SLE disease activity (*59*), and the expression levels of *CXCL13* in Tfh cells and Tph cells are higher in SLE patients than in HC (*17*), suggesting a role of CXCL13 in the pathogenesis of SLE. Indeed, of all the immune cell subsets from SLE patients in our dataset, ThA, Tfh, and Th1 cells exhibited characteristically high expression of *CXCL13* (**fig. S7A**). To evaluate the contribution of *CXCL13*-expressing $CD4^+$ T cell subsets to humoral immunity, we examined the association between *CXCL13* expression of three subsets and autoantibody titers in SLE, as adjusted for age, sex and drug treatment. Total IgG and anti-Sm antibody titer were significantly associated with *CXCL13* expression in ThA cells (**Fig. 5H, Table S4**). In contrast, anti-RNP antibody titer correlated with *CXCL13*

expression in Tfh cells. Thus, it is possible that the production of anti-RNP antibody depends on the follicular pathway mediated by Tfh cells, whereas the production of anti-Sm antibody (which is more specific for SLE) stems from an extrafollicular pathway mediated by ThA cells.

Additionally, we investigated whether ThA cells infiltrated local organ lesions in SLE using publicly available scRNA-seq data of lupus nephritis specimens generated by the Accelerating Medicines Partnership (AMP) (*60*). Our analysis showed that among $CD4^+$ T cell clusters, $PDCD1^+CXCR5^+$ CT3b (TFH-like cells) expressed *ZEB2*, *TBX21,* and *CXCL13*, confirming the presence of a population similar to ThA cells (**fig. S7B-D**). Thus, it appears that ThA-like cells are also involved in organ lesions of SLE and could represent immune targets of therapy.

ThA cells infiltrate organs targeted in autoimmune diseases

We next assessed the infiltration of ThA cells into inflammatory organ sites in IIM patients using scRNA-seq of lymphocytes in muscles and bronchoalveolar lavage fluid (BALF). Interstitial pneumonia frequently coexists with myopathy in anti-ARS antibody-positive patients (*61*). scRNA-seq of $CD45^+$ lymphocytes obtained from BALF of three anti-ARS antibody-positive IIM patients with interstitial pneumonia (**Table S5**) yielded 25 clusters (**fig. S8A-C**) including six cell types, most of which were $CD4^+$ T cells (**Fig. 6A**). Reclustering of $CD4^+$ T cells resulted in the identification of 11 clusters, most of which harbored $CD4^+$ CTLs expressing granzymes and perforin (**Fig. 6B, 6C**). Notably, reclustered $CD4^+$ T cells did not express NKT cell markers (**fig. S8D**). These cells included several clusters expressing *TBX21*, but only C8 expressed *ZEB2* and was identified as a ThA cluster based on expression of key ThA cell-associated genes (**Fig. 6C**). C8 expressed genes related to cytotoxicity and was the only cluster expressing *CXCL13*, suggesting that CTL activity and B cell helper function might coexist, as in

PB ThA cells. Moreover, only the ThA-like cluster (C8) highly expressed proliferation markers such as *MKI67* and *E2F2*, suggesting *in situ* proliferation (**Fig. 6C**). Gene set enrichment analysis showed that the ThA cluster was enriched for both cell cycle-related genes targeting E2F2 and oxidative phosphorylation-related genes (**Fig. 6D**). Indeed, genes such as *TIMMDC1* (*62*), which is associated with multiple members of the mitochondrial complex I assembly complex, was specifically expressed in the ThA cell cluster (**Fig. 6C**). ThA cell proliferation and increased oxidative phosphorylation were also observed in SLE and were associated with disease activity (**Fig. 5E**). Trajectory analysis inferred a differentiation pathway from naïve $CD4^+$ T cells (C5) to CTL (C0) and then from C4 (potentially "pre-ThA cells") not expressing *CXCL13* to ThA cells (C8) expressing *CXCL13* in lung lesions (**Fig. 6E**). Although the ThA cell cluster (C8) partially expressed *PDCD1*, a marker of Tph cells, other clusters also expressed that marker (**Fig. 6C**), implying that ThA cells may correspond to a subfraction of Tph cells.

In order to further examine the connection between ThA and Tph cells (which are not contained in the ImmuNexUT database), we performed bulk RNA-seq of $CD4^+$ T cell subsets, including Tph cells, from four autoimmune disease patients. As expected, the expression levels of *TBX21*, *ZEB2*, *GZMA*, and *GZMB* were upregulated in ThA cells compared with Tph cells (**fig. S9A**). Although there was no significant difference in the mRNA expression levels of *CD28* and *CXCL13* between ThA and Tph cells (**fig. S9A**), ThA cells exhibited a lower frequency of $CD28^+$ cells compared with Tph cells (**fig. S9B**). Notably, while over 80% of unstimulated ThA cells expressed granzyme B protein, less than 40% of unstimulated Tph cells showed expression of this protein (**fig. S9C**). Moreover, higher levels of CXCL13 and granzyme B protein were also observed in TCR-stimulated ThA cells compared to Tph cells (**fig. S9C**). These findings further

indicate that although there may be some overlap between ThA and Tph cells, ThA cells represent a distinct population.

Next, we performed scRNA-seq of muscle-infiltrating lymphocytes using biopsy specimens from three patients with anti-ARS antibody-positive IIM (**Table S6**). Cells were divided into 14 clusters and 7 cell types were annotated (**Fig. 6F, and fig. S10A, S10B**). Among them, *CXCL13* expression was confined to $CD4^+$ T cells, some of which also expressed *PDCD1* (**Fig. 6G**). Reclustering of $CD4^+$ T cells resulted in clusters of ThA cells, Treg cells, and non-ThA/non-Treg cells (**Fig. 6H**). The Treg cluster exhibited high expression of *FOXP3* and *IL2RA* (encoding CD25) (**Fig. 6I, Data file S3**). Conversely, the ThA cluster expressed both *CXCL13* and *GZMA*. Thus, muscle tissue-infiltrating ThA cells had a gene expression profile comparable with that of ThA cells in BALF and PBMC. Trajectory analysis suggested that cells consecutively differentiated from non-ThA/non-Treg cells into muscle-infiltrating ThA-like cells late in pseudotime (**Fig. 6J**). The gene expression changes along pseudotime showed that *TBX21* was expressed in the early phase, and *CXCL13* and *IL21* were expressed with *ZEB2* later in pseudotime (**Fig. 6K**). Genes related to CTL activity, such as *GZMA* and *PRF1*, showed enhanced expression in the mid to late-phase, but subsequently decreased (**Fig. 6K**). These change in gene expression mostly showed an inverse relationship to those of *MX1*, an IFN-related gene, suggesting a potential involvement of IFN signaling in their reduction. Together, these findings support a model in which CTL- and B cell helper function-related genes are induced by *ZEB2* expression and basal expression of *TBX21* at the site of muscle inflammation. Thus, ThA cells may be involved in the pathogenesis of autoimmune diseases not only in the PB, but also in local organs.

DISCUSSION

Here, using a database that included both healthy individuals and patients with autoimmune diseases, we identified a distinct age-associated $CD4^+$ T-cell population that we have designated as ThA cells. ThA cells have a characteristic chemokine receptor profile, $CXCR3^{mid}CD4^+$ Tem, and a phenotype that does not overlap with other described subsets of $CD4^+$ Th cells. ThA cells have aspects of $CD4^+$ CTL and extrafollicular helper T cells, functions we found to be regulated by ZEB2. Furthermore, the gene expression profiles of SLE-associated ThA cells were strongly associated with disease activity and targeted by drugs such as calcineurin inhibitors, indicating potential as predictive biomarkers. Together, our results provide key insights into the cellular mechanisms behind the association of age with autoimmune diseases, which could be important for developing precision medicine strategies to treat SLE.

Interactions between T and B cells are essential in the humoral immune response, with CXCL13 playing a particularly important role. In our study, *CXCL13* expression in ThA cells was correlated with titers of total IgG and anti-Sm antibodies in patients with SLE. These results suggest that ThA cells play a fundamental role in the homeostasis of humoral immunity, and that the production of anti-Sm antibodies, one of the specific autoantibodies to SLE, depends on an extrafollicular pathway mediated by ThA cells. Recently, the efficacy of MMF and TAC combinations in the treatment of lupus nephritis has received considerable attention (*63*). MMF is a pro-drug of mycophenolic acid that inhibits DNA synthesis, and TAC is a representative calcineurin inhibitor. However, it remains unclear which cells are targeted by combination therapy with TAC and MMF. Here, we found that treatment with TAC and MMF substantially affected gene expression variation in SLE-derived ThA cells and Th1 cells, respectively (**Fig. 5G**). Because ThA cells are much more clonally expanded than Th1 cells and other $CD4^+$ T

cells, we expect they are subject to sustained TCR signaling stimulation. Since calcineurin inhibitors target NFAT, which is activated by TCR signaling, it is reasonable that TAC may selectively target ThA cells. Our findings may explain the mechanism of such effective treatment combinations and guide the treatment selection in lupus nephritis. TAC is also effective for myositis and interstitial pneumonia associated with IIM (*64*), which we found was infiltrated by ThA cells.

The presence of $CD4^+$ CTLs consistent with ThA cells has been frequently observed in older individuals and in autoimmune diseases. $CD4^+$ CTLs are increased in the PB of supercentenarians, and these cells are characterized by lack of CD28 and CD27 expression, KLRG1 expression, and skewing of the TCR repertoire (*28*). In addition, $CD27^-CD28^-$ $CD4^+$ Tem cells are increased in the elderly (*65*). ThA-like $CD4^+$ T cell clusters, which expressed *TBX21*, *GZMB*, and *PDCD1* with low expression of *CD28* and *CD27*, were correlated with aging in a large scale scRNA-seq analysis of PB (*66*). In autoimmune diseases, $CD28^-CD4^+$ CTLs infiltrated the skin of patients with systemic sclerosis (SSc) (*26*), and the muscles of IIM patients (*22*). Similarly, cytotoxic $CD4^+CD28^-CD27^-$ Tem cells were found to clonally proliferate and expressed *ZEB2* in the PB of patients with IgG4-related diseases (*23*, *25*). In Graves' disease with ophthalmopathy, $CD4^+$ CTLs expressing *ZEB2*, *TBX21*, and *GZMA* and with a skewed TCR repertoire were found to be expanded (*67*). Although $CD4^+$ CTLs consistent with ThA cells have been reported in healthy elderly and autoimmune diseases, we found that ThA cells have both cytotoxic potential and B cell helper function. Further, we observed that CXCL13 production by ThA cells was enhanced by external stimulation with type I interferons. Thus, we propose that this physiologically expanded cell type in older individuals can transform into an autoimmune-provoking phenotype under inflammatory conditions.

Here, we demonstrated that transcription factor ZEB2 was essential for the induction of the ThA phenotype in $CD4^+$ T cells. Intriguingly, for induction of the ThA cell phenotype by ZEB2 in Jurkat cells, simultaneous activation of other transcription factors, such as T-bet, was required. However, knockdown of *TBX21* in ThA cells did not affect the expression levels of ThA signature genes, suggesting that other transcription factors might compensate for the function of T-bet in ThA cells. The high expression of *TBX21* in ThA cells and the shared TCR repertoire between ThA and Th1 cells suggest that ThA and Th1 cells have a common origin. Indeed, previous reports have shown that mouse Th1 cells exhibit ZEB2-dependent cytotoxicity with decreasing CXCR3 expression (*48*) and that naïve $CD4^+$ T cells turn into $CD4^+$ CTLs after prolonged culture under Th1 differentiation conditions (*68*). However, to confirm ThA differentiation conditions that will facilitate further functional assessment, additional *in vitro* studies will be needed.

Both ThA and Tph cells are extrafollicular helper T cells in autoimmune diseases, thus, their functions likely partially overlap. In our BALF scRNA-seq study of interstitial pneumonia patients, only the ThA cluster expressed *CXCL13* among the multiple $CD4^+$ T cell clusters expressing *PDCD1*, suggesting that ThA cells may correspond to a subfraction of Tph cells with a particularly strong B cell helper function. Indeed, Tph cells at arthritic sites reportedly express high levels of T-bet (*69*) and $CXCR3^+$ Tph cells in SLE have been designated as Tph1 cells (*70*, *71, 72*). ThA cells share some similarities with Tph and Tph1 cells; however, flow cytometry analysis shows that ThA cells and Tph cells exhibit limited overlap (**Fig. S2C**). Moreover, the majority of ThA cells are $CD28^{neg}$, while Tph cells are $CD28^+$ (*73*). Based on the difference in the distribution in the flow cytometry analysis and distinct CD28 and CXCR3 expression profiles, we surmise that they are different populations.

The current study was limited by the small number of available tissue specimens from IIM patients. Thus, scRNA-seq of tissue-infiltrating lymphocytes did not permit disease stratification by gene expression profiling of inflammatory tissue-infiltrating ThA cells. While the B cell helper function of ThA cells was suggested to be linked to disease, it remains to be seen how enhanced cytotoxicity, which is also common in ThA cells from healthy individuals, affects the human body. To assess the interactions of ThA cells with other immune subsets, detailed scRNA-seq analysis and spatial RNA expression analysis of biopsied tissues from various autoimmune disease cases are needed. In the future, additional analyses of lymphocytes infiltrating inflammatory tissues as well as evaluation of the effects of depleting or suppressing ThA cells will be required to elucidate their functional roles.

The numbers of ThA cells described in this study were increased in healthy elderly individuals and they were also associated with the activity of autoimmune disease. Thus, clarifying the differentiation path from ordinary ThA cells to pathological ThA cells should enhance our understanding of the mechanisms driving autoimmune pathogenesis. Future research on the interaction between ThA cells and other immune cells are likely to reveal how ThA cells change in healthy human aging versus autoimmune disease development, providing a deeper understanding of the regulation of the human immune system.

MATERIALS AND METHODS

Study design

To construct the initial dataset (related to Figs. 1–4) consisting of RNA-seq and flow cytometry data for peripheral blood immune cells, 124 HCs, 44 patients with RA, 120 patients with SLE, and 66 patients with IIM were recruited. For further analysis using SLE patient data (related to

Fig. 5), peripheral blood RNA-seq data from 136 patients with SLE and 89 HCs (the secondary dataset), which overlapped with the initial dataset, were obtained and included in the analysis. For scRNA-seq of tissue samples from patients with IIM (related to Fig. 6), three muscle specimens from muscle biopsies for muscle lesions and three bronchoalveolar lavage fluid (BALF) specimens from bronchoscopy for interstitial pneumonia were obtained from patients with IIM with anti-ARS antibody-positive myositis; two of the three cases in both groups were overlapping cases.

Human samples

For healthy controls (HCs), adults with no underlying disease or family history of autoimmune disease were recruited, and participants who consented to the study were included. The rheumatoid arthritis (RA) cohorts comprised patients attending the University of Tokyo Hospital who met the 2010 American College of Rheumatology/European League Against Rheumatism (ACR/EULAR) classification criteria for RA. The systemic lupus erythematosus (SLE) cohort comprised patients meeting the 1997 ACR classification criteria for SLE. The idiopathic inflammatory myopathy (IIM) cohort comprised patients meeting at least one of the following criteria: Bohan & Peter's criteria (*74*), the European Neuromuscular Center criteria (*75*), Sontheimer criteria (*76*), or Griggs criteria (*77*). When selecting patients for the three diseases, cases with malignancy under treatment or chronic infectious diseases, such as hepatitis B and hepatitis C, were excluded. This study was approved by the Ethics Committees of the University of Tokyo (approval numbers G10095 and G10137). Written informed consent was obtained from all participants.

Sample processing for the initial dataset (related to Figs. 1–4)

Peripheral blood samples obtained from patients were subjected to density gradient centrifugation by Ficoll-Paque PLUS (Cytiva) using a Leucosep tube (Greiner Bio-One) immediately after blood collection, and peripheral blood mononuclear cells (PBMCs) were isolated. Residual erythrocytes were hemolyzed with ACK Lysing Buffer (Thermo Fisher Scientific).

Flow cytometry and cell sorting

PBMCs isolated from the peripheral blood were treated with anti-Fcγ receptor antibodies (Thermo Fisher Scientific), stained with the fluorescent antibody reagent shown in Table S2 (4°C, 20 min), and analyzed by flow cytometry (BD FACSAria Fusion cell sorter; BD Biosciences). Nine $CD4^+$ T cell subsets (naïve CD4, Th1, Th2, Th17, Tfh, fraction I naïve regulatory T, fraction II effector regulatory T, fraction III non-regulatory T, and ThA cells) were sorted. FACS gating was performed (Table S3).

Intracellular staining experiments for cytotoxic proteins were performed using PBMCs. PBMCs were stained for cell surface markers with the fluorescent antibody reagents listed in Table S7 (4°C, 20 min), and dead cells were stained with a LIVE/DEAD Fixable Aqua Dead Cell Stain Kit (Thermo Fisher Scientific). After staining (20°C, 15 min), cells were fixed, permeabilized with a BD Human Th1/Th2/Th17 Phenotyping Kit (BD), and stained for intracellular granzyme A, granzyme B, and perforin proteins (20°C, 30 min). The positivity rate of each of the six subsets of $CD4^+$ T cells (naïve CD4 cells, Th1 cells, Th2 cells, Th17 cells, Tfh cells, and ThA cells) and total $CD8^+$ T cells was evaluated by flow cytometry. To detect intracellular CXCL13,

cells were cultured in 96-well plates coated with 1 μg/mL anti-CD3 antibody (Thermo Fisher Scientific; clone OKT3) for 24 hours. Interferon (IFN)-α 2a (R&D systems) was added at a concentration of 1×10^4 U/mL in RPMI medium under specific conditions. A Protein Transport Inhibitor Cocktail (Thermo Fisher Scientific) was applied during the last 5 hours of culture. Cells were harvested and stained for dead cell detection with the LIVE/DEAD Fixable Aqua Dead Cell Stain Kit (Thermo Fisher Scientific). Subsequently, fixation and permeabilization were performed according to the manufacturer's instructions (Intracellular Fixation & Permeabilization Buffer Set; Thermo Fisher Scientific), and intracellular staining was carried out. In the analysis of cryopreserved cells, the selection of viable cells was performed through staining LIVE/DEAD Fixable Aqua Dead Cell Stain Kit (Thermo Fisher Scientific).

RNA sequencing and quality control (QC) of the expression data

Each $CD4^+$ T cell subset sorted by flow cytometry was subjected to total RNA extraction using a MagMAX-96 Total RNA Isolation Kit (Thermo Fisher Scientific), followed by library preparation using a SMART-seq v4 Ultra Low Input RNA Kit (TaKaRa Bio). The generated libraries were sequenced on a HiSeq 2500 Illumina platform to generate 100 base paired-end reads. Adaptor trimming, read alignment, count quantification, and exclusion of low-quality samples were conducted as previously reported (*57*). Briefly, adaptor sequences were trimmed with Cutadapt (v1.16) (*78*) with default parameters, followed by removal of reads containing low-quality bases (Phred quality score < 20 in > 20% of the bases). Reads were mapped using STAR (v2.5.3) (*79*) with the UCSC (provided in the iGenomes UCSC hg38 bundle; https://support.illumina.com/sequencing/sequencing_software/igenome.html), and read counts were quantified using HTSeq (v0.11.2) (*80*). To detect outlier samples, we calculated pairwise

Spearman's correlations of expression with samples from the same cell type and removed samples with a mean correlation coefficient less than 0.9. Furthermore, to detect potentially swapped samples, we calculated the concordance ratio between the genotype from RNA-seq data and the genotype from WGS data at the heterozygous loci, then removed samples with a concordance ratio less than 0.9.

For analyses using $CD4^+$ T cell RNA-seq data with multiple diseases and HCs, we utilized only phase 2 samples (*30*). Finally, we attained 1,562 RNA-seq samples from 185 donors.

We then filtered out low-expression genes (< 10 counts in > 20% of samples or < 0.5 counts per million [CPM] in > 20% of samples), followed by a trimmed mean of M values (TMM) normalization. The data were then converted to CPM with R (v4.1.0) package edgeR (v3.36.0) (*81*) for each cell type. For each filtering process, the genes described in this paper were retained such that we could compare results between diseases or subsets.

PCA of RNA-seq data

We combined normalized expression using the intersection of genes across all subsets, and then sex and disease (when all the samples were analyzed) were regressed out using a linear model. PCA was performed using the prcomp function in R (v4.1.0).

Identification of specifically expressed genes in RNA-seq

To identify subset-specific expression, we filtered genes with low expression levels in all subsets across HCs. Counts were then normalized and averaged for each subset. Cell type-specific expressed genes were identified using the ROKU function in the TCC package (v1.34.0) (*32*). Although this function identifies both specifically upregulated genes and downregulated genes

based on Shannon entropies, we selected only specifically upregulated genes in each subset for an illustration.

Differential gene expression analysis

To detect differentially expressed genes, we performed quasi-likelihood F-tests implemented in edgeR. In analyses between subsets or between diseases, sex was treated as a covariate.

Coculture of $CD4^+$ T cells and B cells

$CD4^+$ T cell subsets and total B cells from PBMCs were sorted on a FACSAria Fusion (BD). B cells were defined as $CD19^+CD3^-$ cells. For cocultures, 5×10^4 T cells were seeded with 1×10^5 B cells in the presence of anti-CD3/CD28 beads (Thermo Fisher Scientific) at a 1:1 bead/T-cell ratio for 7 days. The culture medium was RPMI 1640 (Thermo Fisher Scientific) supplemented with 10% fetal bovine serum (BioWest), 100 µg/mL streptomycin, 100 U/mL penicillin G, and 0.3 mg/mL L-glutamine (Thermo Fisher Scientific). Recombinant IL-21R His-tag protein (R&D Systems; 50 µg/mL) or anti-CD84 antibodies (BioLegend; 20 µg/mL) were added in blocking experiments. For Transwell cocultures, the B cells were seeded onto 0.4-µm Transwell membranes (Corning) and cultured separately from the T cells at the bottom of the well in 200 µL medium. Concentrations of total IgG were measured with ELISA.

Gene transfection into Jurkat cells

To induce *ZEB2* overexpression, Jurkat cells (ECACC) were nucleofected with 0.5 mg mock ZsGreen plasmid vector (TaKaRa) or *ZEB2* ZsGreen plasmid vector per 2×10^5 cells using the SE Cell Line 4D-Nucleofecton X Kit S (Lonza). After incubation for 72 hours at 37°C in a 5%

CO_2 incubator, ZsGreen-positive or -negative cells were sorted by flow cytometry, and the expression of *TBX21*, *ZEB2*, *GZMA*, *PRF1*, and *CXCL13* was evaluated by reverse-transcription qPCR (RT-qPCR). To induce *TBX21* overexpression, Jurkat cells were nucleofected with 0.5 mg mock DsRed plasmid vector (TaKaRa) or *TBX21* DsRed plasmid vector per 2×10^5 cells, and after 72 hours of culture, DsRed-positive cells were sorted by flow cytometry and evaluated by RT-qPCR. To induce both *ZEB2* and *TBX21* overexpression, nucleofection was performed using a combination of two of the four plasmid vectors (mock ZsGreen, mock DsRed, *ZEB2* ZsGreen, and *TBX21* DsRed), and after 72 hours of culture, cells positive for both ZsGreen and DsRed were sorted by flow cytometry and evaluated by RT-qPCR. The culture medium was RPMI 1640 (Thermo Fisher Scientific) supplemented with 10% fetal bovine serum (BioWest), 100 μg/mL streptomycin, 100 U/mL penicillin G, and 0.3 mg/mL L-glutamine (Thermo Fisher Scientific).

Gene knockout in human ThA cells using the CRISPR/Cas9 system

First, human PB ThA cells were sorted by flow cytometry and divided into four conditions (1×10^5 cells/condition). A green fluorescent protein (GFP)-fused Cas9 protein (26 pmol; Sigma-Aldrich) was incubated with 100 pmol single guide RNA (nontargeting, *ZEB2*, or *TBX21* single guide RNA) for 10 minutes and then nucleofected into ThA cells using the P2 Primary Cell 4D-Nucleofector X Kit S (Lonza). After 5 hours, the cells were cultured for 48 hours in 96-well plates that had been coated overnight with 1 μg/mL anti-human CD3 antibody (Thermo Fisher Scientific). The culture medium was High-Glucose RPMI 1640 (Wako) supplemented with 10% fetal bovine serum (BioWest), 100 μg/mL streptomycin, 100 U/mL penicillin G, and 0.3 mg/mL L-glutamine (Thermo Fisher Scientific). After culture, GFP-positive or -negative cells were

sorted by flow cytometry, and the expression of *TBX21*, *ZEB2*, *GZMA*, *PRF1*, and *CXCL13* was evaluated by RT-qPCR.

Statistical Analysis

The statistical analyses are also described in the figure captions. To calculate the correlation coefficients between the percentage of each $CD4^+$ T cell subset and age, Spearman's rank correlations were used. For comparisons of multiple groups, one-way analysis of variance (ANOVA) with Tukey multiple comparison test was used for the ELISA, intracellular CXCL13 staining, proliferation assay and telomere length data; one-way ANOVA with Bonferroni multiple comparison test was used for the CRISPR/Cas9 assay, intracellular granzyme A/B expression, and perforin expression data; one-way repeated measures ANOVA with Fisher's LSD post hoc test were used for cytotoxicity assay and intracellular CXCL13/granzyme B staining; and multiple test correction using the Benjamini–Hochberg method was used for the other data. For comparisons between two groups, the Wilcoxon rank sum test was used for the proportion of ThA cells among $CD4^+$ T cells in PB, and the two-sided unpaired Student's t test was used for *TBX21* and/or *ZEB2* overexpression in Jurkat cells and for cell surface CD28 positivity. Statistical analyses were conducted by R (v4.1.0), GraphPad Prism 6 or 9.

Supplementary Materials

Materials and Methods

Figs. S1 to S10

Table S1 to S8

Data file S1 to S5

References and Notes

1. Z. Szekanecz, I. B. McInnes, G. Schett, S. Szamosi, S. Benkő, G. Szűcs, Autoinflammation and autoimmunity across rheumatic and musculoskeletal diseases. *Nat. Rev. Rheumatol.* **17**, 585–595 (2021).

2. J. J. Goronzy, C. M. Weyand, Immune aging and autoimmunity. *Cell. Mol. Life Sci.* **69**, 1615–1623 (2012).

3. Y. Wang, Z. Fu, X. Li, Y. Liang, S. Pei, S. Hao, Q. Zhu, T. Yu, Y. Pei, J. Yuan, J. Ye, J. Fu, J. Xu, J. Hong, R. Yang, H. Hou, X. Huang, C. Peng, M. Zheng, Y. Xiao, Cytoplasmic DNA sensing by KU complex in aged CD4+ T cell potentiates T cell activation and aging-related autoimmune inflammation. *Immunity*. **54**, 632-647.e9 (2021).

4. E. Myasoedova, J. Davis, E. L. Matteson, C. S. Crowson, Is the epidemiology of rheumatoid arthritis changing? Results from a population-based incidence study, 1985-2014. *Ann. Rheum. Dis.* **79**, 440–444 (2020).

5. C. M. Weyand, J. J. Goronzy, The immunology of rheumatoid arthritis. *Nat. Immunol.* **22**, 10–18 (2021).

6. F. Rees, M. Doherty, M. Grainge, G. Davenport, P. Lanyon, W. Zhang, The incidence and prevalence of systemic lupus erythematosus in the UK, 1999-2012. *Ann. Rheum. Dis.* **75**, 136–141 (2016).

7. P. Y. Leong, J. Y. Huang, J. Y. Chiou, Y. C. Bai, J. C. C. Wei, The prevalence and incidence of systemic lupus erythematosus in Taiwan: a nationwide population-based study. *Sci. Rep.* **11**, 5631 (2021).

8. U. Khan, H. Ghazanfar, *T Lymphocytes and Autoimmunity.* **341**, 125-168 (2018).

9. J. M. Anaya, Common mechanisms of autoimmune diseases (the autoimmune tautology). *Autoimmun. Rev.* **11**, 781–784 (2012).

10. J. E. Craft, Follicular helper T cells in immunity and systemic autoimmunity. *Nat. Rev. Rheumatol.* **8**, 337–347 (2012).

11. C. G. Vinuesa, M. A. Linterman, D. Yu, I. C. M. Maclennan, Follicular Helper T Cells. *Annu. Rev. Immunol.* **34**, 335–368 (2016).

12. W. Ise, T. Kurosaki, Plasma cell generation during T-cell-dependent immune responses. *Int. Immunol.* **33**, 797–801 (2021).

13. S. A. Jenks, K. S. Cashman, M. C. Woodruff, F. E. H. Lee, I. Sanz, Extrafollicular responses in humans and SLE. *Immunol. Rev.* **288**, 136–148 (2019).

14. S. A. Jenks, K. S. Cashman, E. Zumaquero, U. M. Marigorta, A. V. Patel, X. Wang, D. Tomar, M. C. Woodruff, Z. Simon, R. Bugrovsky, E. L. Blalock, C. D. Scharer, C. M. Tipton, C. Wei, S. S. Lim, M. Petri, T. B. Niewold, J. H. Anolik, G. Gibson, F. E. H. Lee, J. M. Boss, F. E. Lund, I. Sanz, Distinct Effector B Cells Induced by Unregulated Toll-like Receptor 7 Contribute to Pathogenic Responses in Systemic Lupus Erythematosus. *Immunity*. **49**, 725-739.e6 (2018).

15. D. A. Rao, M. F. Gurish, J. L. Marshall, K. Slowikowski, C. Y. Fonseka, Y. Liu, L. T. Donlin, L. A. Henderson, K. Wei, F. Mizoguchi, N. C. Teslovich, M. E. Weinblatt, E. M. Massarotti, J. S. Coblyn, S. M. Helfgott, Y. C. Lee, D. J. Todd, V. P. Bykerk, S. M. Goodman, A. B. Pernis, L. B. Ivashkiv, E. W. Karlson, P. A. Nigrovic, A. Filer, C. D. Buckley, J. A. Lederer, S. Raychaudhuri, M. B. Brenner, Pathologically expanded

peripheral T helper cell subset drives B cells in rheumatoid arthritis. *Nature*. **542**, 110–114 (2017).

16. F. Zhang, K. Wei, K. Slowikowski, C. Y. Fonseka, D. A. Rao, S. Kelly, S. M. Goodman, D. Tabechian, L. B. Hughes, K. Salomon-Escoto, G. F. M. Watts, A. H. Jonsson, J. Rangel-Moreno, N. Meednu, C. Rozo, W. Apruzzese, T. M. Eisenhaure, D. J. Lieb, D. L. Boyle, A. M. Mandelin, J. Albrecht, S. L. Bridges, C. D. Buckley, J. H. Buckner, J. Dolan, J. M. Guthridge, M. Gutierrez-Arcelus, L. B. Ivashkiv, E. A. James, J. A. James, J. Keegan, Y. C. Lee, M. J. McGeachy, M. A. McNamara, J. R. Mears, F. Mizoguchi, J. P. Nguyen, A. Noma, D. E. Orange, M. Rohani-Pichavant, C. Ritchlin, W. H. Robinson, A. Seshadri, D. Sutherby, J. Seifert, J. D. Turner, P. J. Utz, B. F. Boyce, E. DiCarlo, E. M. Gravallese, P. K. Gregersen, L. Moreland, G. S. Firestein, N. Hacohen, C. Nusbaum, J. A. Lederer, H. Perlman, C. Pitzalis, A. Filer, V. M. Holers, V. P. Bykerk, L. T. Donlin, J. H. Anolik, M. B. Brenner, S. Raychaudhuri, Defining inflammatory cell states in rheumatoid arthritis joint synovial tissues by integrating single-cell transcriptomics and mass cytometry. *Nat. Immunol.* **20**, 928–942 (2019).

17. A. V. Bocharnikov, J. Keegan, V. S. Wacleche, Y. Cao, C. Y. Fonseka, G. Wang, E. S. Muise, K. X. Zhang, A. Arazi, G. Keras, Z. J. Li, Y. Qu, M. F. Gurish, M. Petri, J. P. Buyon, C. Putterman, D. Wofsy, J. A. James, J. M. Guthridge, B. Diamond, J. H. Anolik, M. F. Mackey, S. E. Alves, P. A. Nigrovic, K. H. Costenbader, M. B. Brenner, J. A. Lederer, D. A. Rao, PD-1hiCXCR5- T peripheral helper cells promote B cell responses in lupus via MAF and IL-21. *JCI Insight*. **4**, e130062 (2019).

18. Y. Gao, G. Dunlap, M. Elahee, D. A. Rao, Patterns of T-Cell Phenotypes in Rheumatic Diseases From Single-Cell Studies of Tissue. *ACR Open Rheumatol.* **3**, 601–613 (2021).

19. A. V. Rubtsov, K. Rubtsova, A. Fischer, R. T. Meehan, J. Z. Gillis, J. W. Kappler, P. Marrack, Toll-like receptor 7 (TLR7)-driven accumulation of a novel CD11c+ B-cell population is important for the development of autoimmunity. *Blood*. **118**, 1305–1315 (2011).

20. M. Reina-Campos, N. E. Scharping, A. W. Goldrath, CD8+ T cell metabolism in infection and cancer. *Nat. Rev. Immunol.* **21**, 718–738 (2021).

21. P. Golstein, G. M. Griffiths, An early history of T cell-mediated cytotoxicity. *Nat. Rev. Immunol.* **18**, 527–535 (2018).

22. A. E. R. Fasth, M. Dastmalchi, A. Rahbar, S. Salomonsson, J. M. Pandya, E. Lindroos, I. Nennesmo, K.-J. Malmberg, C. Söderberg-Nauclér, C. Trollmo, I. E. Lundberg, V. Malmström, T Cell Infiltrates in the Muscles of Patients with Dermatomyositis and Polymyositis Are Dominated by CD28 null T Cells . *J. Immunol.* **183**, 4792–4799 (2009).

23. H. Mattoo, V. S. Mahajan, T. Maehara, V. Deshpande, E. Della-Torre, Z. S. Wallace, M. Kulikova, J. M. Drijvers, J. Daccache, M. N. Carruthers, F. V. Castelino, J. R. Stone, J. H. Stone, S. Pillai, Clonal expansion of CD4+ cytotoxic T lymphocytes in patients with IgG4-related disease. *J. Allergy Clin. Immunol.* **138**, 825–838 (2016).

24. X. Hong, S. Meng, D. Tang, T. Wang, L. Ding, H. Yu, H. Li, D. Liu, Y. Dai, M. Yang, Single-Cell RNA Sequencing Reveals the Expansion of Cytotoxic CD4+ T Lymphocytes and a Landscape of Immune Cells in Primary Sjögren's Syndrome. *Front. Immunol.* **11**, 594658 (2021).

25. C. A. Perugino, N. Kaneko, T. Maehara, H. Mattoo, J. Kers, H. Allard-Chamard, V. S. Mahajan, H. Liu, E. Della-Torre, S. J. H. Murphy, M. Ghebremichael, Z. S. Wallace, M.

B. Bolster, L. M. Harvey, G. Mylvaganam, Y. Tuncay, L. Liang, S. B. Montesi, X. Zhang, A. Tinju, K. Mochizuki, R. Munemura, M. Sakamoto, M. Moriyama, S. Nakamura, N. Yosef, J. H. Stone, S. Pillai, CD4+ and CD8+ cytotoxic T lymphocytes may induce mesenchymal cell apoptosis in IgG4-related disease. *J. Allergy Clin. Immunol.* **147**, 368-382 (2021).

26. T. Maehara, N. Kaneko, C. A. Perugino, H. Mattoo, J. Kers, H. Allard-Chamard, V. S. Mahajan, H. Liu, S. J. H. Murphy, M. Ghebremichael, D. Fox, A. S. Payne, R. Lafyatis, J. H. Stone, D. Khanna, S. Pillai, Cytotoxic CD4+ T lymphocytes may induce endothelial cell apoptosis in systemic sclerosis. *J. Clin. Invest.* **130**, 2451–2464 (2020).

27. T. Maehara, H. Mattoo, M. Ohta, V. S. Mahajan, M. Moriyama, M. Yamauchi, J. Drijvers, S. Nakamura, J. H. Stone, S. S. Pillai, Lesional CD4+ IFN-γ+ cytotoxic T lymphocytes in IgG4-related dacryoadenitis and sialoadenitis. *Ann. Rheum. Dis.* **76**, 377–385 (2017).

28. K. Hashimoto, T. Kouno, T. Ikawa, N. Hayatsu, Y. Miyajima, H. Yabukami, T. Terooatea, T. Sasaki, T. Suzuki, M. Valentine, G. Pascarella, Y. Okazaki, H. Suzuki, J. W. Shin, A. Minoda, I. Taniuchi, H. Okano, Y. Arai, N. Hirose, P. Carninci, Single-cell transcriptomics reveals expansion of cytotoxic CD4 T cells in supercentenarians. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 24242–24251 (2019).

29. W. Xu, A. Larbi, Markers of T cell senescence in humans. *Int. J. Mol. Sci.* **18**, 1742 (2017).

30. M. Ota, Y. Nagafuchi, H. Hatano, K. Ishigaki, C. Terao, Y. Takeshima, H. Yanaoka, S. Kobayashi, M. Okubo, H. Shirai, Y. Sugimori, J. Maeda, M. Nakano, S. Yamada, R. Yoshida, H. Tsuchiya, Y. Tsuchida, S. Akizuki, H. Yoshifuji, K. Ohmura, T. Mimori, K.

Yoshida, D. Kurosaka, M. Okada, K. Setoguchi, H. Kaneko, N. Ban, N. Yabuki, K. Matsuki, H. Mutoh, S. Oyama, M. Okazaki, H. Tsunoda, Y. Iwasaki, S. Sumitomo, H. Shoda, Y. Kochi, Y. Okada, K. Yamamoto, T. Okamura, K. Fujio, Dynamic landscape of immune cell-specific gene regulation in immune-mediated diseases. *Cell*. **184**, 3006-3021.e17 (2021).

31. S. N. Mueller, T. Gebhardt, F. R. Carbone, W. R. Heath, Memory T Cell Subsets, Migration Patterns, and Tissue Residence. *Annu. Rev. Immunol.* **31**, 137–161 (2013).

32. J. Sun, T. Nishiyama, K. Shimizu, K. Kadota, TCC: An R package for comparing tag count data with robust normalization strategies. *BMC Bioinformatics*. **14**, 219 (2013).

33. A. E. Oja, F. A. Vieira Braga, E. B. M. Remmerswaal, N. A. M. Kragten, K. M. L. Hertoghs, J. Zuo, P. A. Moss, R. A. W. van Lier, K. P. J. M. van Gisbergen, P. Hombrink, The transcription factor hobit identifies human cytotoxic CD4+ T cells. *Front. Immunol.* **8**, 325 (2017).

34. G. W. Chang, C. C. Hsiao, Y. M. Peng, F. A. Vieira Braga, N. A. M. Kragten, E. B. M. Remmerswaal, M. D. B. van de Garde, R. Straussberg, G. M. König, E. Kostenis, V. Knäuper, L. Meyaard, R. A. W. van Lier, K. P. J. M. van Gisbergen, H. H. Lin, J. Hamann, The Adhesion G Protein-Coupled Receptor GPR56/ADGRG1 Is an Inhibitory Receptor on Human NK Cells. *Cell Rep.* **15**, 1757–1770 (2016).

35. M. Cenerenti, M. Saillard, P. Romero, C. Jandus, The Era of Cytotoxic CD4 T Cells. *Front. Immunol.* **13**, 867189 (2022).

36. H.-F. Chang, C. Schirra, M. Ninov, U. Hahn, K. Ravichandran, E. Krause, U. Becherer, Š. Bálint, M. Harkiolaki, H. Urlaub, S. Valitutti, C. T. Baldari, M. L. Dustin, R. Jahn, J.

Rettig, Identification of distinct cytotoxic granules as the origin of supramolecular attack particles in T lymphocytes. *Nat. Commun.* **13**, 1029 (2022).

37. M. Cuenca, J. Sintes, Á. Lányi, P. Engel, CD84 cell surface signaling molecule: An emerging biomarker and target for cancer and autoimmune disorders. *Clin. Immunol.* **204**, 43–49 (2019).

38. G. C. Tsokos, M. S. Lo, P. C. Reis, K. E. Sullivan, New insights into the immunopathogenesis of systemic lupus erythematosus. *Nat. Rev. Rheumatol.* **12**, 716–730 (2016).

39. Z. Touma, M. B. Urowitz, D. D. Gladman, Systemic Lupus Erythematosus Disease Activity Index 2000 Responder Index-50 Website. *J. Rheumatol.* **40**, 733 (2013).

40. R. Y. O. Hisada, M. Kato, E. R. I. Sugawara, M. Kanda, Y. Fujieda, K. Oku, T. Bohgaki, O. Amengual, T. Horita, S. Yasuda, T. Atsumi, Circulating plasmablasts contribute to antiphospholipid antibody production, associated with type I interferon upregulation. *J. Thromb. Haemost.* **17**, 1134–1143 (2019).

41. I. E. Lundberg, M. De Visser, V. P. Werth, Classification of myositis. *Nat. Rev. Rheumatol.* **14**, 269–278 (2018).

42. Y. Takeuchi, K. Hirota, S. Sakaguchi, Impaired T cell receptor signaling and development of T cell–mediated autoimmune arthritis. *Immunol. Rev.* **294**, 164–176 (2020).

43. R. Saigusa, H. Winkels, K. Ley, T cell subsets and functions in atherosclerosis. *Nat. Rev. Cardiol.* **17**, 387–401 (2020).

44. J. Zhu, H. Yamane, W. E. Paul, Differentiation of Effector CD4 T Cell Populations. *Annu. Rev. Immunol.* **28**, 445–489 (2010).

45. A. Kallies, K. L. Good-Jacobson, Transcription Factor T-bet Orchestrates Lineage Development and Function in the Immune System. *Trends Immunol.* **38**, 287–297 (2017).

46. M. Fardi, M. Alivand, B. Baradaran, M. Farshdousti Hagh, S. Solali, The crucial role of ZEB2: From development to epithelial-to-mesenchymal transition and cancer complexity. *J. Cell. Physiol.* **234**, 14783–14799 (2019).

47. C. L. Scott, K. D. Omilusik, ZEBs: Novel Players in Immune Cell Development and Function. *Trends Immunol.* **40**, 431–446 (2019).

48. P. D. Krueger, M. F. Goldberg, S.-W. Hong, K. C. Osum, R. A. Langlois, D. I. Kotov, T. Dileepan, M. K. Jenkins, Two sequential activation modules control the differentiation of protective T helper-1 (Th1) cells. *Immunity*. **54**, 687-701.e4 (2021).

49. H. Jamann, Q.-L. Cui, H. L. Desu, F. Pernin, O. Tastet, A. Halaweh, N. Farzam-kia, V. H. Mamane, O. Ouédraogo, A. Cleret-Buhot, A. Daigneault, R. Balthazard, W. Klement, F. Lemaître, N. Arbour, J. Antel, J. A. Stratton, C. Larochelle, Contact-Dependent Granzyme B-Mediated Cytotoxicity of Th17-Polarized Cells Toward Human Oligodendrocytes. *Front. Immunol.* **13**, 850616 (2022).

50. L. Hou, K. Yuki, CCR6 and CXCR6 Identify the Th17 Cells With Cytotoxicity in Experimental Autoimmune Encephalomyelitis. *Front. Immunol.* **13**, 819224 (2022).

51. Y. Mikami, R. L. Philips, G. Sciumè, F. Petermann, F. Meylan, H. Nagashima, C. Yao, F. P. Davis, S. R. Brooks, H. W. Sun, H. Takahashi, A. C. Poholek, H. Y. Shih, B. Afzali, S. A. Muljo, M. Hafner, Y. Kanno, J. J. O'Shea, MicroRNA-221 and -222 modulate intestinal inflammatory Th17 cell response as negative feedback regulators downstream of interleukin-23. *Immunity*. **54**, 514-525.e6 (2021).

52. P. Georg, R. Astaburuaga-García, L. Bonaguro, S. Brumhard, L. Michalick, L. J. Lippert, T. Kostevc, C. Gäbel, M. Schneider, M. Streitz, V. Demichev, I. Gemünd, M. Barone, P. Tober-Lau, E. T. Helbig, D. Hillus, L. Petrov, J. Stein, H.-P. Dey, D. Paclik, C. Iwert, M. Mülleder, S. K. Aulakh, S. Djudjaj, R. D. Bülow, H. E. Mei, A. R. Schulz, A. Thiel, S. Hippenstiel, A.-E. Saliba, R. Eils, I. Lehmann, M. A. Mall, S. Stricker, J. Röhmel, V. M. Corman, D. Beule, E. Wyler, M. Landthaler, B. Obermayer, S. von Stillfried, P. Boor, M. Demir, H. Wesselmann, N. Suttorp, A. Uhrig, H. Müller-Redetzky, J. Nattermann, W. M. Kuebler, C. Meisel, M. Ralser, J. L. Schultze, A. C. Aschenbrenner, C. Thibeault, F. Kurth, L. E. Sander, N. Blüthgen, B. Sawitzki, Complement activation induces excessive T cell cytotoxicity in severe COVID-19. *Cell*, **185**, 493-512.e25 (2022).

53. B. J. Meckiff, C. Ramírez-Suástegui, V. Fajardo, S. J. Chee, A. Kusnadi, H. Simon, S. Eschweiler, A. Grifoni, E. Pelosi, D. Weiskopf, A. Sette, F. Ay, G. Seumois, C. H. Ottensmeier, P. Vijayanand, Imbalance of Regulatory and Cytotoxic SARS-CoV-2-Reactive CD4+ T Cells in COVID-19. *Cell*. **183**, 1340-1353.e16 (2020).

54. C. Liu, A. J. Martins, W. W. Lau, N. Rachmaninoff, J. Chen, L. Imberti, D. Mostaghimi, D. L. Fink, P. D. Burbelo, K. Dobbs, O. M. Delmonte, N. Bansal, L. Failla, A. Sottini, E. Quiros-Roldan, K. L. Han, B. A. Sellers, F. Cheung, R. Sparks, T. W. Chun, S. Moir, M. S. Lionakis, M. S. Abers, R. Apps, M. Bosticardo, P. Milanez-Almeida, M. P. Mulè, E. Shaw, Y. Zhang, F. Castelli, M. L. Muiesan, G. Tomasoni, F. Scolari, A. Tucci, C. Rossi, H. C. Su, D. B. Kuhns, J. I. Cohen, L. D. Notarangelo, J. S. Tsang, Time-resolved systems immunology reveals a late juncture linked to fatal COVID-19. *Cell*. **184**, 1836-1857.e22 (2021).

55. H. Wang, Y. Wang, X. Jiang, Z. Wang, B. Zhong, Y. Fang, The molecular mechanism of curcumol on inducing cell growth arrest and apoptosis in Jurkat cells, a model of CD4+ T cells. *Int. Immunopharmacol.* **21**, 375–382 (2014).

56. T. L. W. Muskardin, T. B. Niewold, Type I interferon in rheumatic diseases. *Nat. Rev. Rheumatol.* **14**, 214–228 (2018).

57. M. Nakano, M. Ota, Y. Takeshima, Y. Iwasaki, H. Hatano, Y. Nagafuchi, T. Itamiya, J. Maeda, R. Yoshida, S. Yamada, A. Nishiwaki, H. Takahashi, H. Takahashi, Y. Akutsu, T. Kusuda, H. Suetsugu, L. Liu, K. Kim, X. Yin, S.-Y. Bang, Y. Cui, H.-S. Lee, H. Shoda, X. Zhang, S.-C. Bae, C. Terao, K. Yamamoto, T. Okamura, K. Ishigaki, K. Fujio, Distinct transcriptome architectures underlying lupus establishment and exacerbation. *Cell*. **185**, 3375-3389.e21 (2022).

58. D. Nehar-Belaid, S. Hong, R. Marches, G. Chen, M. Bolisetty, J. Baisch, L. Walters, M. Punaro, R. J. Rossi, C. H. Chung, R. P. Huynh, P. Singh, W. F. Flynn, J. A. Tabanor-Gayle, N. Kuchipudi, A. Mejias, M. A. Collet, A. L. Lucido, K. Palucka, P. Robson, S. Lakshminarayanan, O. Ramilo, T. Wright, V. Pascual, J. F. Banchereau, Mapping systemic lupus erythematosus heterogeneity at the single-cell level. *Nat. Immunol.* **21**, 1094–1106 (2020).

59. L. Schiffer, K. Worthmann, H. Haller, M. Schiffer, CXCL13 as a new biomarker of systemic lupus erythematosus and lupus nephritis - from bench to bedside? *Clin. Exp. Immunol.* **179**, 85–89 (2015).

60. A. Arazi, D. A. Rao, C. C. Berthier, A. Davidson, Y. Liu, P. J. Hoover, A. Chicoine, T. M. Eisenhaure, A. H. Jonsson, S. Li, D. J. Lieb, F. Zhang, K. Slowikowski, E. P. Browne, A.

Noma, D. Sutherby, S. Steelman, D. E. Smilek, P. Tosta, W. Apruzzese, E. Massarotti, M. Dall'Era, M. Park, D. L. Kamen, R. A. Furie, F. Payan-Schober, W. F. Pendergraft, E. A. McInnis, J. P. Buyon, M. A. Petri, C. Putterman, K. C. Kalunian, E. S. Woodle, J. A. Lederer, D. A. Hildeman, C. Nusbaum, S. Raychaudhuri, M. Kretzler, J. H. Anolik, M. B. Brenner, D. Wofsy, N. Hacohen, B. Diamond, D. Waguespack, S. M. Connery, M. A. McMahon, W. J. McCune, R. B. Kado, R. Hsu, M. A. Cunningham, P. J. Utz, M. Pichavant, H. T. Maecker, R. Gupta, J. A. James, J. M. Guthridge, C. Fonseka, E. Der, R. Clancy, T. Tuschl, H. Suryawanshi, A. Fava, D. H. Goldman, The immune cell landscape in kidneys of patients with lupus nephritis. *Nat. Immunol.* **20**, 902–914 (2019).

61. M. Mahler, F. W. Miller, M. J. Fritzler, Idiopathic inflammatory myopathies and the anti-synthetase syndrome: A comprehensive review. *Autoimmun. Rev.* **13**, 367–371 (2014).

62. S. Guerrero-Castillo, F. Baertling, D. Kownatzki, H. J. Wessels, S. Arnold, U. Brandt, L. Nijtmans, The Assembly Pathway of Mitochondrial Respiratory Chain Complex I. *Cell Metab.* **25**, 128–139 (2017).

63. Z. Liu, H. Zhang, Z. Liu, C. Xing, P. Fu, Z. Ni, J. Chen, H. Lin, F. Liu, Y. He, Y. He, L. Miao, N. Chen, Y. Li, Y. Gu, W. Shi, W. Hu, Z. Liu, H. Bao, C. Zeng, M. Zhou, Multitarget therapy for induction treatment of lupus nephritis: A randomized trial. *Ann. Intern. Med.* **162**, 18–26 (2015).

64. I. E. Lundberg, M. Fujimoto, J. Vencovsky, R. Aggarwal, M. Holmqvist, L. Christopher-Stine, A. L. Mammen, F. W. Miller, Idiopathic inflammatory myopathies. *Nat. Rev. Dis. Prim.* **7**, 86 (2021).

65. S. Koch, A. Larbi, E. Derhovanessian, D. Özcelik, E. Naumova, G. Pawelec, Multiparameter flow cytometric analysis of CD4 and CD8 T cell subsets in young and old people. *Immun. Ageing*. **5**, 6 (2008).

66. A. Nathan, J. I. Beynor, Y. Baglaenko, S. Suliman, K. Ishigaki, S. Asgari, C. C. Huang, Y. Luo, Z. Zhang, K. Lopez, C. S. Lindestam Arlehamn, J. D. Ernst, J. Jimenez, R. I. Calderón, L. Lecca, I. Van Rhijn, D. B. Moody, M. B. Murray, S. Raychaudhuri, Multimodally profiling memory T cells from a tuberculosis cohort identifies cell state associations with demographics, environment and disease. *Nat. Immunol.* **22**, 781–793 (2021).

67. Y. Wang, Z. Chen, T. Wang, H. Guo, Y. Liu, N. Dang, S. Hu, L. Wu, C. Zhang, K. Ye, B. Shi, A novel CD4+ CTL subtype characterized by chemotaxis and inflammation is involved in the pathogenesis of Graves' orbitopathy. *Cell. Mol. Immunol.* **18**, 735–745 (2021).

68. Y. Serroukh, C. Gu-Trantien, B. Hooshiar Kashani, M. Defrance, T. P. Vu Manh, A. Azouz, A. Detavernier, A. Hoyois, J. Das, M. Bizet, E. Pollet, T. Tabbuso, E. Calonne, K. Van Gisbergen, M. Dalod, F. Fuks, S. Goriely, A. Marchant, The transcription factors Runx3 and ThPOK cross-regulate acquisition of cytotoxic function by human Th1 lymphocytes. *Elife*. **7**, e30496 (2018).

69. J. Fischer, J. Dirks, J. Klaussner, G. Haase, A. Holl-Wieden, C. Hofmann, S. Hackenberg, H. Girschick, H. Morbach, Effect of Clonally Expanded PD-1highCXCR5–CD4+ Peripheral T Helper Cells on B Cell Differentiation in the Joints of Patients With Antinuclear Antibody–Positive Juvenile Idiopathic Arthritis. *Arthritis Rheumatol.* **74**, 150–162 (2022).

70. S. Caielli, D. T. Veiga, P. Balasubramanian, S. Athale, B. Domic, E. Murat, R. Banchereau, Z. Xu, M. Chandra, C. H. Chung, L. Walters, J. Baisch, T. Wright, M. Punaro, L. Nassi, K. Stewart, J. Fuller, D. Ucar, H. Ueno, J. Zhou, J. Banchereau, V. Pascual, A CD4+ T cell population expanded in lupus blood provides B cell help through interleukin-10 and succinate. *Nat. Med.* **25**, 75–81 (2019).

71. A. Makiyama, A. Chiba, D. Noto, G. Murayama, K. Yamaji, N. Tamura, S. Miyake, Expanded circulating peripheral helper T cells in systemic lupus erythematosus: Association with disease activity and B cell differentiation. *Rheumatol. (United Kingdom)*. **58**, 1861–1869 (2019).

72. S. C. Eisenbarth, D. Baumjohann, J. Craft, N. Fazilleau, C. S. Ma, S. G. Tangye, C. G. Vinuesa, M. A. Linterman, CD4+ T cells that help B cells – a proposal for uniform nomenclature. *Trends Immunol.* **42**, 658–669 (2021).

73. L. Yong, W. Chunyan, Y. Yan, L. Wanyu, J. Huifan, Z. Pingwei, J. Yanfang, Expanded circulating peripheral helper T cells in primary biliary cholangitis: Tph cells in PBC. *Mol. Immunol.* **131**, 44–50 (2021).

74. A. Bohan, J. B. Peter., Polymyositis and dermatomyositis (first of two parts). *N Engl J Med.* **292**, 344–347 (1975).

75. J. E. Hoogendijk, A. A. Amato, B. R. Lecky, E. H. Choy, I. E. Lundberg, M. R. Rose, J. Vencovsky, M. De Visser, R. A. Hughes, 119th ENMC international workshop: Trial design in adult idiopathic inflammatory myopathies, with the exception of inclusion body myositis, 10-12 October 2003, Naarden, The Netherlands. *Neuromuscul. Disord.* **14**, 337–345 (2004).

76. R. D. Sontheimer, Would a new name hasten the acceptance of amyopathic dermatomyositis (dermatomyositis siné myositis) as a distinctive subset within the idiopathic inflammatory dermatomyopathies spectrum of clinical illness? *J. Am. Acad. Dermatol.* **46**, 626–636 (2002).

77. K. Sivakumar, M. C. Dalakas, Inclusion body myositis and myopathies. *Curr. Opin. Neurol.* **10**, 413–420 (1997).

78. M. Martin, Cutadapt Removes Adapter Sequences From High-Throughput Sequencing Reads. *EMBnet.journal*. **17**, 10–12 (2011).

79. A. Dobin, C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, T. R. Gingeras, STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*. **29**, 15–21 (2013).

80. S. Anders, P. T. Pyl, W. Huber, HTSeq-A Python framework to work with high-throughput sequencing data. *Bioinformatics*. **31**, 166–169 (2015).

81. M. D. Robinson, D. J. McCarthy, G. K. Smyth, edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. **26**, 139–140 (2009).

82. D. E. Ho, G. King, E. A. Stuart, K. Imai, MatchIt : Nonparametric Preprocessing for Parametric Casusal Inference. *J. Stat. Softw.* **42**, 1–28 (2011).

83. Y. M. El-Sherbiny, A. Psarras, M. Y. M. Yusof, E. M. A. Hensor, R. Tooze, G. Doody, A. A. A. Mohamed, D. McGonagle, M. Wittmann, P. Emery, E. M. Vital, A novel two-score system for interferon status segregates autoimmune diseases and correlates with clinical features. *Sci. Rep.* **8**, 5793 (2018).

84. Y. Nagafuchi, M. Ota, H. Hatano, M. Inoue, S. Kobayashi, M. Okubo, Y. Sugimori, M. Nakano, S. Yamada, R. Yoshida, Y. Tsuchida, Y. Iwasaki, H. Shoda, Y. Okada, K. Yamamoto, K. Ishigaki, T. Okamura, K. Fujio, Control of naive and effector CD4 T cell receptor repertoires by rheumatoid-arthritis-risk HLA alleles. *J. Autoimmun.* **133**, 102907 (2022).

85. D. A. Bolotin, S. Poslavsky, I. Mitrophanov, M. Shugay, I. Z. Mamedov, E. V. Putintseva, D. M. Chudakov, MiXCR: Software for comprehensive adaptive immunity profiling. *Nat. Methods*. **12**, 380–381 (2015).

86. P. Langfelder, S. Horvath, WGCNA: An R package for weighted correlation network analysis. *BMC Bioinformatics*. **9**, 559 (2008).

87. T. Stuart, A. Butler, P. Hoffman, C. Hafemeister, E. Papalexi, W. M. Mauck, Y. Hao, M. Stoeckius, P. Smibert, R. Satija, Comprehensive Integration of Single-Cell Data. *Cell*. **177**, 1888-1902.e21 (2019).

88. C. Trapnell, D. Cacchiarelli, J. Grimsby, P. Pokharel, S. Li, M. Morse, N. J. Lennon, K. J. Livak, T. S. Mikkelsen, J. L. Rinn, The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* **32**, 381–386 (2014).

89. N. Borcherding, A. Vishwakarma, A. P. Voigt, A. Bellizzi, J. Kaplan, K. Nepple, A. K. Salem, R. W. Jenkins, Y. Zakharia, W. Zhang, Mapping the immune environment in clear cell renal carcinoma by single-cell genomics. *Commun. Biol.* **4**, 122 (2021).

90. G. E. Hoffman, E. E. Schadt, variancePartition: Interpreting drivers of variation in complex gene expression studies. *BMC Bioinformatics*. **17**, 483 (2016).

91. G. Yu, L. G. Wang, Y. Han, Q. Y. He, ClusterProfiler: An R package for comparing biological themes among gene clusters. *Omi. A J. Integr. Biol.* **16**, 284–287 (2012).

92. A. Liberzon, C. Birger, H. Thorvaldsdóttir, M. Ghandi, J. P. Mesirov, P. Tamayo, The Molecular Signatures Database Hallmark Gene Set Collection. *Cell Syst.* **1**, 417–425 (2015).

93. M. Kanehisa, S. Goto, KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).

Acknowledgments: We thank all participants and collaborators in the University of Tokyo and Chugai Pharmaceutical for the contribution to sample processing. We appreciate T. Otsuki, M. Matsumoto, H. Hamada, and J. Izumisawa, our experimental assistants, for their help in obtaining the data for this study. The super-computing resource was provided by Human Genome Center, Institute of Medical Sciences, The University of Tokyo. Current institutional affiliations: M.O., Department of Genetics, Stanford University, Stanford, CA, USA., Gladstone-UCSF Institute of Genomic Immunology, San Francisco, CA, USA; Y.N., Department of Rheumatology, Leiden University Medical Center, Leiden, The Netherlands.

Funding: Supported by The Center of Innovation Program from Japan Science and Technology Agency (JST) (JPMJCE1304 to KF), The Ministry of Education, Culture, Sports, and the Japan Agency for Medical Research and Development (AMED) (JP21tm0424221 to KF), The Ministry of Education, Culture, Sports, and the Japan Agency for Medical Research and Development (AMED) (JP21zf0127004 to KF), The Ministry of Education, Culture, Sports, and the Japan Agency for Medical Research and Development (AMED) (JP22ek0410074 to KF), The Ministry

of Education, Culture, Sports, and the Japan Agency for Medical Research and Development (AMED) (JP223fa627001 to KF), The Ministry of Education, Culture, Sports, and the Japan Agency for Medical Research and Development (AMED) (JP233fa627001 to KF), The Ministry of Education, Culture, Sports, and the Japan Agency for Medical Research and Development (AMED) (JP23gm1810005 to KF, TO and KI), JSPS Grant-in-Aid for Scientific Research (B) (JP19H03697, JP22H03110 to TO), Chugai Pharmaceutical Co., Ltd., Tokyo, Japan.

Author contributions: MG, HT, RY, TO, MO, KF designed study; MG, HT, RY performed in vitro experiments; MG, HT, RY, TI, MO, MN, YN, KI, TO, KF analyzed data; MN, KI designed the model of Fig. 5; MG, HT, RY, HHatano, YS, TS, HHarada, MM, AK, TT obtained human samples and clinical data; KK, KY, HS conducted critical reading of the manuscript; HT, MG, TO, KF wrote the manuscript with critical inputs from TI, MN; HT edited figures.

Competing interests: TO, TI, and YN belong to the Social Cooperation Program, Department of Functional Genomics and Immunological Diseases, supported by Chugai Pharmaceutical. K.F. receives speaking fees, consulting honoraria, and research support from Chugai Pharmaceutical. All other authors declare no competing interests.

Data and materials availability:

The RNA-seq data used in this study have been deposited at the National Bioscience Database Center (NBDC) Human Database (https://humandbs.biosciencedbc.jp/en/) and are available under accession numbers JGAS000626, JGAS000627, JGAS000648. Tabulated data underlying the figures is available in Data file S5. All other data needed to support the conclusions of the paper are present in the paper or the Supplementary Materials.

Main figures:

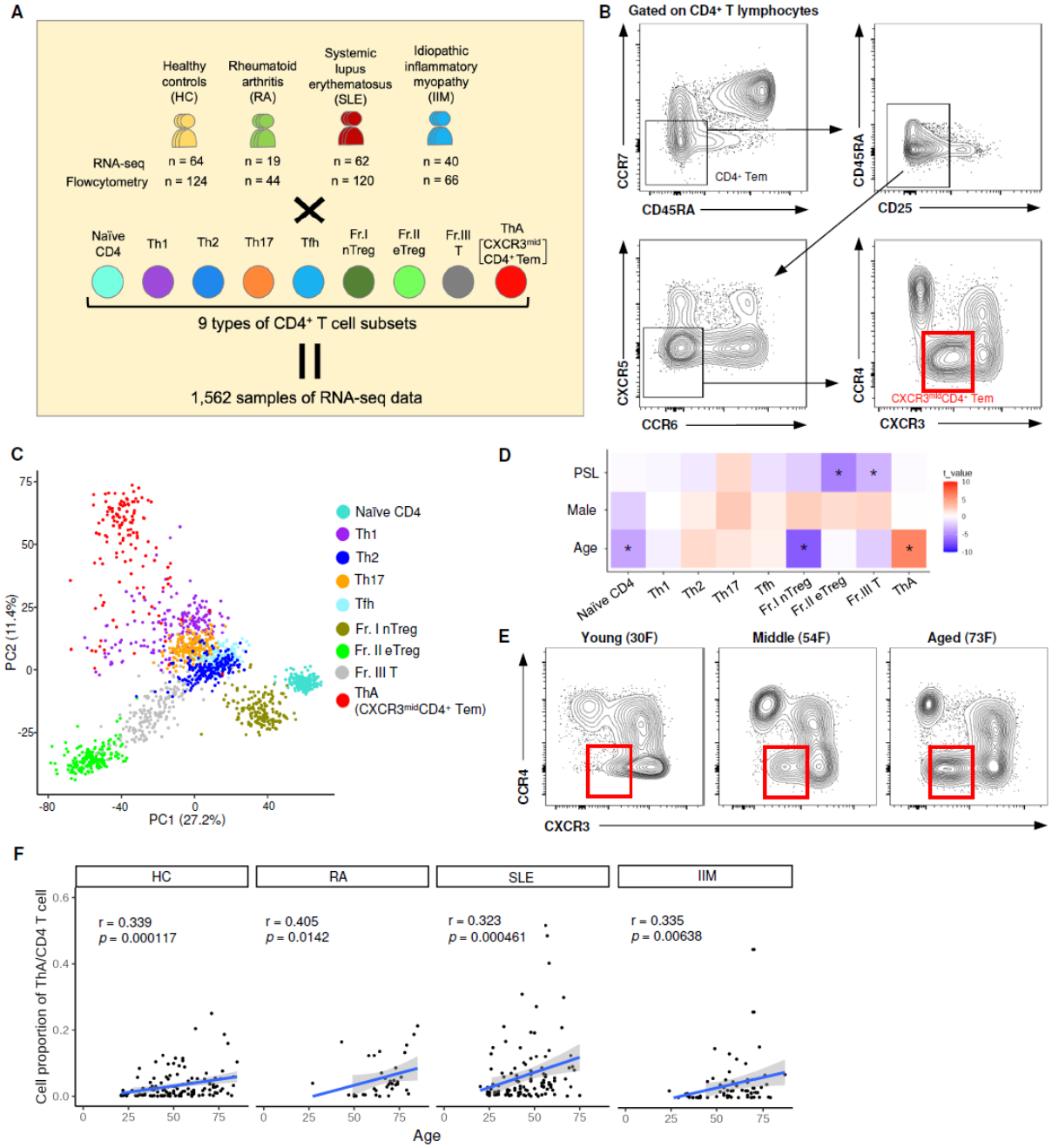

Fig. 1. Characterization of age-associated CXCR3^{mid}CD4^{+} effector memory T cells. (A) Overview of the initial dataset of this study. Nine CD4^{+} T cell subsets were collected from PB of HC and three autoimmune diseases. Flow cytometric data and 1,562 RNA sequencing datasets were generated from the initial dataset. **(B)** Flow cytometric hierarchical gating strategy for ThA

cells ($CXCR3^{mid}CD4^{+}$ Tem). Representative data from 69 years old HC female is shown. **(C)** Principal component analysis (PCA) of RNA-seq data from the initial dataset. **(D)** Heatmap of *t* values for the coefficients obtained by multivariate analysis of the associations between the proportion of each $CD4^{+}$ T cell subset among total $CD4^{+}$ T cells detected by flow cytometry and the clinical data, i.e., prednisolone (PSL) use, sex, and age. $*p < 0.05$. **(E)** Representative flow cytometric data from healthy controls at different ages. The gating strategy is the same as Fig. 1B. Gates indicate ThA cells. **(F)** The proportion of ThA cells among total $CD4^{+}$ T cells, detected by flow cytometry from the initial dataset, plotted against age. The regression line is shown. The correlation coefficients and *p* values were determined using Spearman's rank correlations.

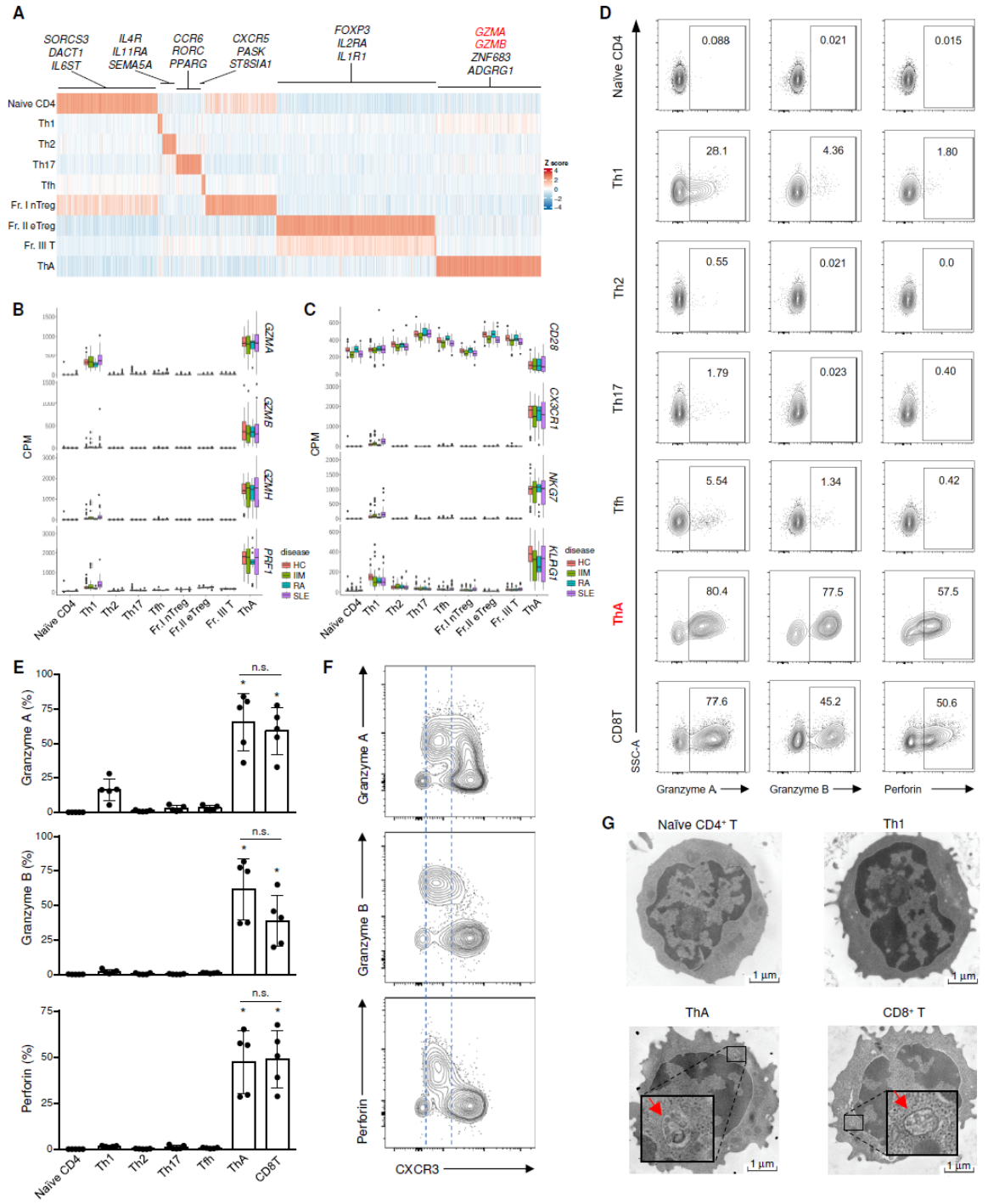

Fig. 2. The ThA cell signature is associated with a cytotoxic phenotype. (A) Cell-type-specific expressed genes (SEG) across each cell subset were extracted from the initial dataset. Column-wise Z scores of normalized counts are plotted. Representative genes are annotated on

the top. Full lists of SEG are provided in **Data file S1**. **(B, C)** Expression of genes directly involved in cytotoxic activity **(B)** and other $CD4^+$ CTL signature genes **(C)** in RNA-seq data from the initial dataset. CPM: counts per million. **(D, E)** Intracellular staining of proteins involved in cytotoxic activity. Representative plots (**D**: HC) and a graph summarizing positive rates (**E**: HC, n = 5 biological replicates, mean ± SD) are shown. *****, Significantly different from naïve $CD4^+$ T cells (adjusted $p < 0.05$, one-way ANOVA with Bonferroni multiple comparison test). n.s.: not significant. **(F)** Representative flow cytometric analysis of surface CXCR3 vs. intracellular granzymes and perforin in $CD25^-CXCR5^-CCR6^-CCR4^-$ effector memory ($CD45RA^-CCR7^-$) $CD4^+$ T cells. **(G)** Transmission electron microscopy images of T cells. Arrows indicate multi-core granules (MCGs), which represent cytotoxic granules in T cells. Representative data from three HC are shown.

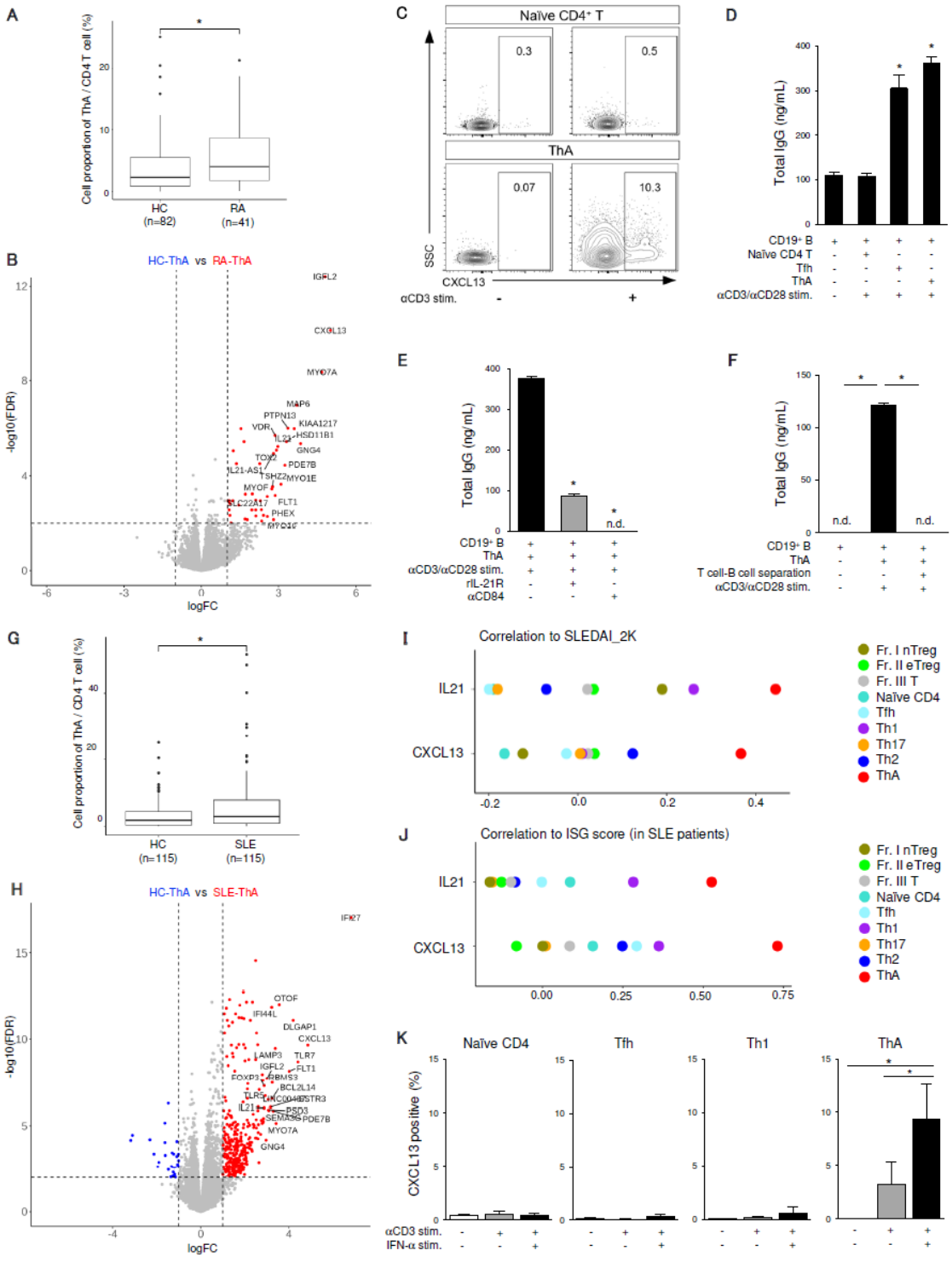

Fig. 3. ThA cells exert B cell helper functions. (A) Comparison of the proportion of ThA cells between age/sex matched HC (n = 82) and RA cases (n = 41). *, $p < 0.05$, Wilcoxon rank sum test.

(B) Volcano plot of DEGs in RNA-seq between HC-ThA and RA-ThA cells. **(C)** CXCL13 expression of $CD4^+$ T cell subsets in RA patients. The data are representative of three experiments. **(D)** Total IgG in supernatant of each $CD4^+$ T cell subset co-cultured with B cells obtained from RA patients, quantified by ELISA (mean ± SD, n = 3). *****, $p < 0.05$, compared with B cell alone culture condition. **(E)** Total IgG in supernatants during co-culture of ThA and B cells obtained from RA patients with or without IL-21 neutralization or CD84 blocking, quantified by ELISA (mean ± SD, n = 4). *****, $p < 0.05$, compared with co-culture of ThA and B cells under αCD3/αCD28 mAb stimulation. **(F)** Total IgG in supernatants when ThA and B cells obtained from HC were co-cultured under separate or mixed conditions on Transwell plates, quantified by ELISA (mean ± SD, n = 3). n.d.: not detected. *****, $p < 0.05$. **(G)** Comparison of the proportion of ThA cells between age/sex matched HC and SLE cases (n = 115 each). *****, $p < 0.05$, Wilcoxon rank sum test. **(H)** Volcano plots of DEGs in RNA-seq between HC-ThA and SLE-ThA cells. **(I and J)** Pearson correlation coefficients between gene expression and SLEDAI-2K (I) or ISG score (J) in each $CD4^+$ T cell subset of SLE patients. **(K)** Intracellular $CXCL13^+$ cells for each HC $CD4^+$ T cell subset upon αCD3 mAb and/or IFN-α stimulation (mean ± SD, n = 4 biological replicates). *****, $p < 0.05$, one-way ANOVA with Tukey multiple comparison tests. **(D–F)** The data are representative of three experiments (three to four technical replicates of three biological replicates). Statistical significance was calculated by one-way ANOVA with Tukey multiple comparison tests.

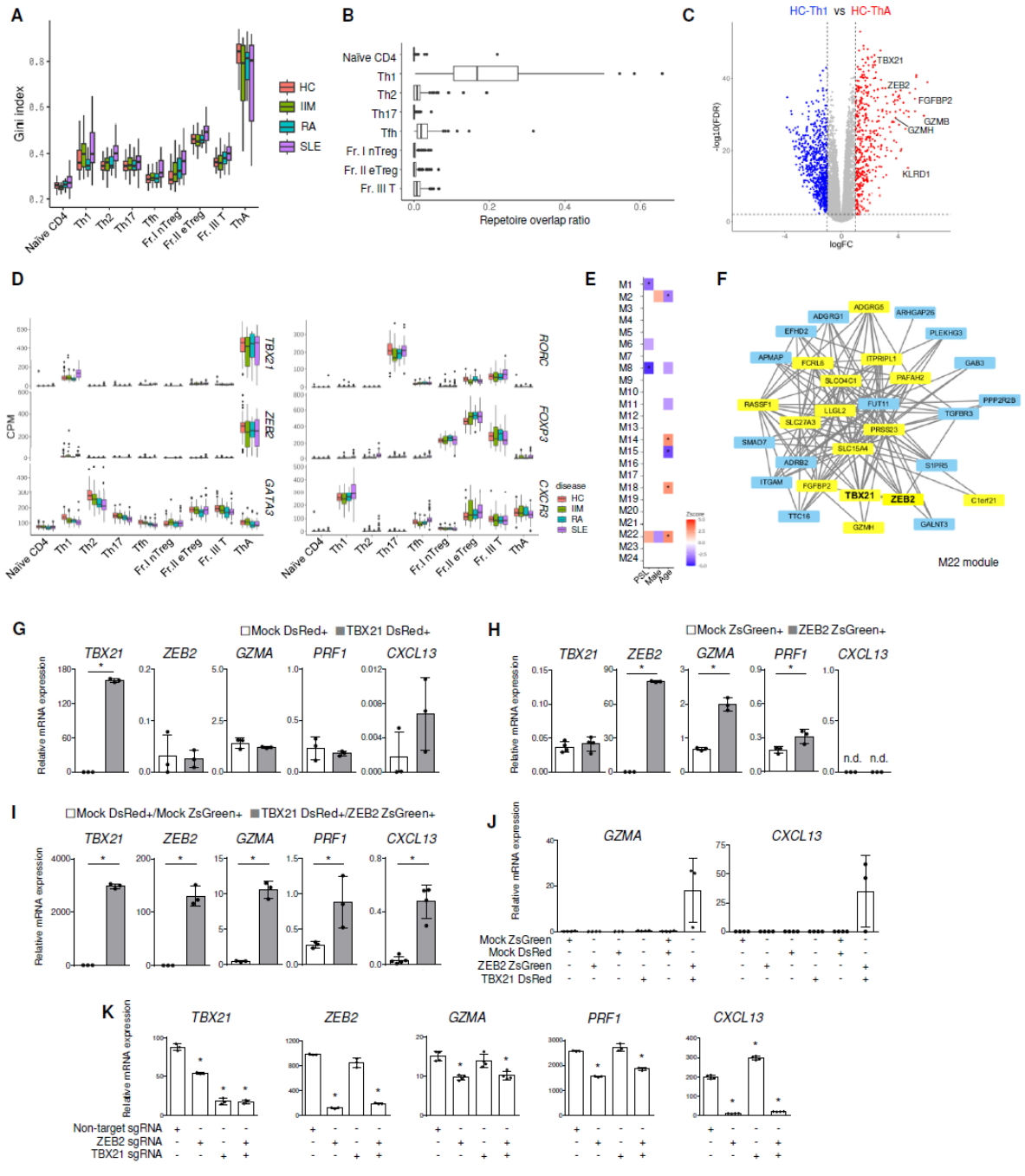

Fig. 4. ZEB2 regulates the function of ThA cells. (A, B) T cell receptor (TCR) repertoire analysis of initial dataset. Gini skewing indices of TCRβ clones in each T cell subset **(A)**, and overlap ratio of TCR repertoire with ThA cells **(B)**. **(C)** Volcano plots of RNA-seq data depicting DEGs between HC-Th1 and HC-ThA cells. **(D)** Gene expression levels of known master regulatory genes of Th

cells and *CXCR3* in RNA-seq data from the initial dataset. **(E, F)** Weighted gene coexpression network analysis of ThA cell RNA-seq data from the initial dataset. Characteristics of each gene network module **(E)**, and the top 30 hub genes that comprise its aging-related M22 module **(F)**. Top 15 hub genes are highlighted yellow. **(G, H, I)** Gene expression in Jurkat cells quantified by qRT-PCR with induction of *TBX21* or mock vector **(G)** (n = 3), *ZEB2* or mock vector **(H)** (n = 3 or 4), and both *TBX21* and *ZEB2* or mock vectors **(I)** (n = 3 to 5) (mean ± SD). *****, $p < 0.05$, two-sided unpaired *t*-test. **(J)** Gene expression quantified by qRT-PCR of human primary $CD4^+$ T cells with induction of *TBX21* and/or *ZEB2* or mock vectors (mean ± SD, n = 4). **(K)** Gene expression quantified by qRT-PCR of HC ThA cells after *TBX21* and/or *ZEB2* gene knock-out via the CRISPR/Cas9 system (mean ± SD, n = 3 to 4). *****, adjusted $p < 0.05$, compared with the non-target single guide_RNA condition, one-way ANOVA with Bonferroni multiple comparison test. **(G–K)** The data are representative of three experiments (three to five technical replicates of three biological replicates).

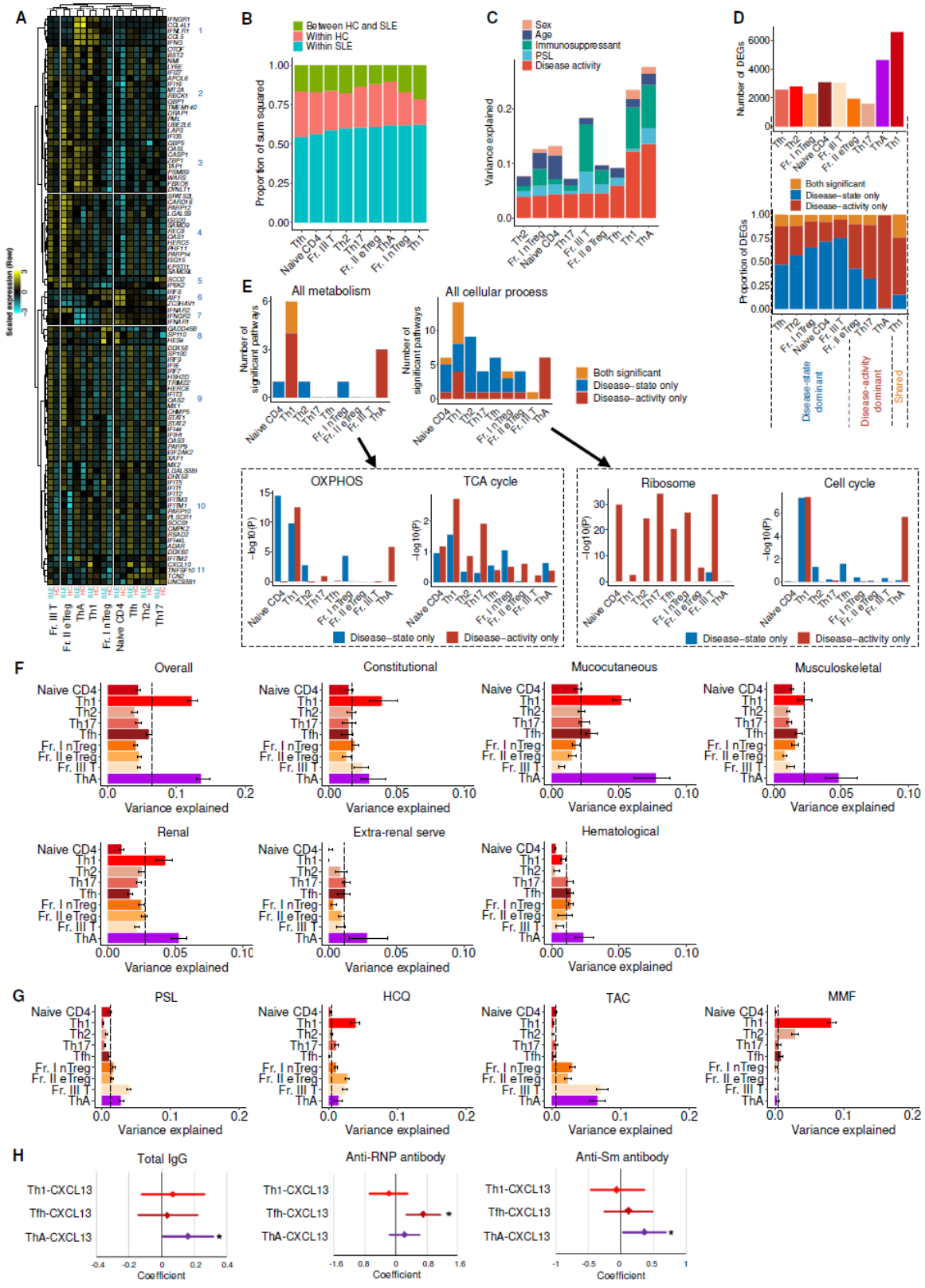
A
Scaled expression (Row)
Fr. III T
Fr. II eTreg
ThA
Th1
Fr. I nTreg
Naive CD4
Th
Th2
Th17
B
Between HC and SLE
Within HC
Within SLE
Proportion of sum squared
C
Sex
Age
Immunosuppressant
PSL
Disease activity
Variance explained
D
Number of DEGs
Both significant
Disease−state only
Disease−activity only
Proportion of DEGs
Disease-state dominant
Disease-activity dominant
Shared
E
All metabolism
All cellular process
Number of significant pathways
Both significant
Disease−state only
Disease−activity only
OXPHOS
TCA cycle
Ribosome
Cell cycle
−log10(P)
Disease−state only
Disease−activity only
F
Overall
Constitutional
Mucocutaneous
Musculoskeletal
Renal
Extra-renal serve
Hematological
Variance explained
G
PSL
HCQ
TAC
MMF
Variance explained
H
Total IgG
Anti-RNP antibody
Anti-Sm antibody
Th1-CXCL13
Tfh-CXCL13
ThA-CXCL13
Coefficient

Fig. 5. Gene expression profiles of ThA cells reflect lupus disease activity. (A) Heatmap depicting the mean expression of 100 IRGs across all $CD4^+$ T cell subsets and diseases. Genes and cell subsets were hierarchically clustered based on the IRG expression levels of our dataset. **(B)** Proportion of sum-squared deviations within HC, SLE, and between HC and SLE data in each cell subset. **(C)** Proportion of variance explained by indicated clinical parameters within SLE data in each cell subset. **(D)** Number of disease-state/-activity signature DEGs (top) and the proportion of DEG types in each cell subset (bottom). Cell subsets were separated into three groups. **(E)** Number of significant enrichments of metabolism- and cellular process-related pathways (top) and the enrichment of representative pathways for each signature (bottom). *p* values in one-sided Fisher's exact test. **(F, G)** Proportion of variance within-SLE data in each cell subset, explained by the overall disease activity and representative organ activities **(F)**, and the use of key therapeutic agents **(G)**. Error bars and dashed vertical lines indicate 95% confidence intervals from jackknife resampling and the median values across nine cell types, respectively. **(H)** Forest plot showing the estimated correlation between autoantibody titers and *CXCL13* expression in RNA-seq data of each T cell subset in SLE patients, after adjusting for age, sex, and therapeutic agents (see also **Table S4**). Point represents the median value and line spans 95% confidence interval. *, Adjusted $p < 0.05$, multiple linear regression analysis. **(A-H)** The secondary dataset was used for the analysis.

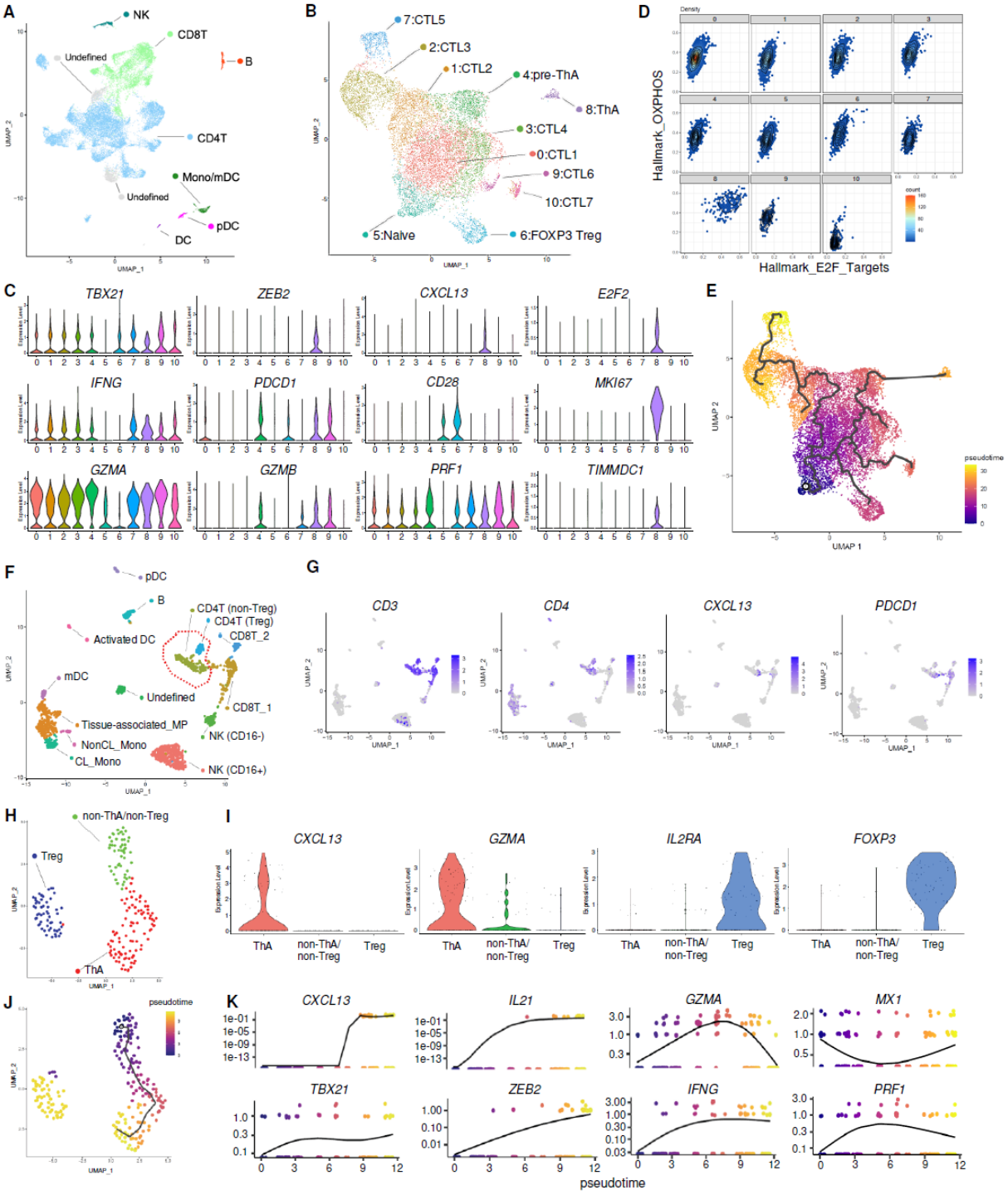

Fig. 6. ThA cells infiltrate the target organs of autoimmune disease. (A) UMAP of scRNA-seq data for CD45$^+$ cells from BALF of three patients with anti-ARS antibody-positive IIM. **(B)** UMAP reclustering analysis of CD4$^+$ T cells from CD45$^+$ BALF cells in **A**. **(C)** Violin plots

depicting expression of key genes of ThA cells across cell clusters in **B**. **(D)** Gene Set Enrichment Analysis (GSEA) of proliferative capacity (Hallmark_E2F_TARGETS) and oxidative phosphorylation (Hallmark_OXPHOS) of each $CD4^+$ T cell cluster in **B**. **(E)** Pseudotime values were calculated and plotted across cell clusters in **B** (deep blue to yellow for low to high pseudotime values, respectively). **(F, G)** UMAP **(F)** and feature plot **(G)** of scRNA-seq data for $CD45^+$ cells from muscle biopsy specimens of three anti-ARS antibody-positive IIM patients. **(H)** UMAP reclustering analysis of $CD4^+$ T cells from $CD45^+$ muscle-infiltrating cells in **F**. **(I)** Violin plot depicting expression of *CXCL13, GZMA, IL2RA,* and *FOXP3* across cell clusters in **H**. **(J)** Pseudotime values were calculated and plotted across cell clusters in **H** (deep blue to yellow for low to high pseudotime values, respectively). **(K)** Gene expression changes along the pseudotime values in **J**.

Supplementary Materials for

Title：Age-associated CD4$^+$ T cells with B cell-promoting function regulated by ZEB2 in autoimmune diseases

Manaka Goto, Hideyuki Takahashi, Ryochi Yoshida *et al.*

Corresponding author: Keishi Fujio, Tomohisa Okamura, Mineto Ota

This file includes:

Supplementary Methods

Figs. S1 to S10

Tables S1 to S8

Other Supplementary Materials for this manuscript includes following:

Data file S1 to S5

MDAR Reproducibility Checklist

Supplementary Materials

Materials and Methods

Analysis of the association between cell proportions and clinical data

In each subset, cell proportion was regressed according to rank-transformed age, rank-transformed prednisolone (PSL) dosage, and sex using a linear model, with disease treated as a covariate. *T*-values of the regression coefficient were compared across subsets and clinical traits (Fig. 1D).

Analysis of the association between clinical data and transcriptome data

We combined normalized expression using the intersection of the genes from each subset, and sex and disease (when all the samples were analyzed) were then regressed out using a linear model. After clinical data were joined with the transcriptome data, their relationships were evaluated by taking Spearman's correlation *r* in each combination of subset and gene. *p* values were adjusted with the Benjamini and Hochberg method to correct for multiple testing.

Propensity score matching

To compare cell proportions of ThA cells among $CD4^+$ T cells with controlled baseline characteristics between HCs and targeted diseases (RA, SLE, or IIM), we matched HCs to patients with the targeted disease based on propensity scores calculated with a logistic regression of sex and age using greedy nearest neighbor matching without replacement by MatchIt package (v4.5.3) (*82*). Matching ratio was set to 2:1 in RA, and 1:1 in SLE and IIM.

Calculation of the IFN-stimulated gene (ISG) score

In each sample, sex was regressed out from normalized data, and then we calculated ISG scores using genes and loading factors listed in the IFN score A proposed by El-Sherbiny et al. (*83*).

T-cell receptor (TCR) repertoire analysis

We utilized TCR data attained from our previous report (*84*). Briefly, RNA-seq data were aligned with the MiXCR (v 3.0.11) (*85*) using the "analyze shotgun" command with default parameters. We used these TCR data, the Gini index, and sequence motif of clones analyzed with immunarch (v0.5.5, https://immunarch.com/).

Weighted gene co-expression network analysis

We performed weighted gene co-expression network analysis to normalized data for ThA cells with the "signed network" option (*86*). The soft threshold power with the adjacency matrix was set to 12 to generate gene modules, and gene modules whose first eigenvectors were highly correlated (Pearson's correlation > 0.9) were then merged.

In silico analyses including Tph cells

In in-silico analyses including Tph cells, we utilized samples from 4 donors (2 RA donors and 2 IIM donors), which were collected separately from samples in the main analysis. Sample normalization was conducted in the same way as described above. Expression differences between Tph and ThA cells were investigated for some genes without batch correction due to the small sample size.

Enzyme-linked immunosorbent assay (ELISA)

The concentration of total IgG in the culture supernatant was measured with commercial kits from Bethyl Laboratories, Mabtech, or Thermo Fisher Scientific. For granzyme A, IFN-γ, and CXCL13 detection, Duoset kits (R&D Systems) were used. Each ELISA was performed as described in the manufacturer's protocols.

Electron microscopy of human T cells

Naïve $CD4^+$ T cells, Th1 cells, ThA cells, and $CD8^+$ T cells (2×10^5 cells each) were sorted by flow cytometry. The sorted cells were prefixed with half Karnovsky's solution (a mixture of 2.5% glutaraldehyde and 2% paraformaldehyde) at Kamakura Techno Science and then agar-embedded with 2% agar solution. Samples were then postfixed with 1% osmium tetroxide solution, dehydrated with ethanol and propylene oxide, and embedded in resin. Ultrathin sections (60–80 nm thickness) were prepared from the embedded blocks, electrons were stained (double stained with uranium acetate and lead citrate), and the internal cell structure was observed by transmission electron microscopy (Hitachi HT7700).

Cytotoxicity assay

Naïve $CD4^+$ T cells, ThA cells and effector memory $CD8^+$ T cells ($CD3^+CD4^-CD8^+$ $CCR7^-$ $CD45RA^-$ cells) were sorted and co-cultured with K562 cells at an effector-to-target (E:T) ratio of 10:1 for 4 hours. The culture medium was RPMI 1640 without phenol red (Thermo Fisher Scientific) containing 10% fetal bovine serum (Thermo Fisher Scientific), 100 ng/mL anti-CD3 antibody (Thermo Fisher Scientific; clone OKT3, functional grade) and 500 ng/mL recombinant human IFN-γ (R&D Systems). The LDH release was measured in the supernatant using CytoTox

96 Non-Radioactive Cytotoxicity Assay (Promega). The percentage of specific cytotoxicity was calculated using the following formula according to the manufacturer's instructions.

Specific Lysis (%) = 100 * (Experimental – Effector Spontaneous – Target Spontaneous) / (Target Maximum – Target Spontaneous)

The LDH release from the following conditions was used in the calculation. Experimental: co-culture of effector cells and K562 cells, Effector Spontaneous: individual culture of effector cells, Target Spontaneous: individual culture of K562 cells, Target Maximum: 0.8% Triton X-100 treated K562 cells.

Intracellular staining

Sorted Naïve $CD4^+$ T cells, Tfh cells, and ThA cells were cultured in RPMI 1640 (Thermo Fisher Scientific) supplemented with 10% fetal bovine serum (BioWest), 100 μg/mL streptomycin, 100 U/mL penicillin G, and 0.3 mg/mL L-glutamine (Thermo Fisher Scientific). Under the stimulation conditions, anti-CD3/CD28 beads (Thermo Fisher Scientific) at a 1:1 ratio of cells to beads and/or 1×10^4 U/mL of IFN-α 2a (R&D systems) were added to the medium. To evaluate CXCL13 and Granzyme B expression, Naïve $CD4^+$ T cells, Tph cells, and ThA cells were sorted and cultured with plate-coated anti-CD3 antibody (1 μg/mL, Thermo Fisher Scientific; clone OKT3) for 24 hours. Cells were treated with a Protein Transport Inhibitor Cocktail (Thermo Fisher Scientific) during the final 5 hours of culture and subsequently assessed by dead cell staining and intracellular staining with Intracellular Fixation & Permeabilization Buffer Set (Thermo Fisher Scientific) as described in the methods section of the main text.

T-cell proliferation assay

$CD4^+$ T cell subsets were sorted and labeled with 0.25 μM 5- or 6-(N-succinimidyloxycarbonyl) fluorescein 3′,6′-diacetate (CFSE; Dojindo). Each subset was incubated for 5 days in the presence or absence of anti-CD3/CD28 beads (Thermo Fisher Scientific) in RPMI 1640 (Thermo Fisher Scientific) supplemented with 10% fetal bovine serum (BioWest), 100 μg/mL streptomycin, 100 U/mL penicillin G, and 0.3 mg/mL L-glutamine (Thermo Fisher Scientific). T-cell proliferation was measured by CFSE (Dojindo) dilution.

Telomere length quantification

An Absolute Human Telomere Length Quantification qPCR Assay Kit (ScienCell) was used to measure telomere length. Genomic DNA was extracted from sorted $CD4^+$ T cell subsets using a QIAamp DNA Mini kit (Qiagen). Quantitative polymerase chain reaction (qPCR) was performed to amplify the telomere sequence and single-copy reference on chromosome 17 for each sample. Absolute telomere length was calculated by comparing the quantification cycle number of the telomere and single-copy reference between the target samples and the reference genomic DNA sample from the kit.

Gene transfection into human primary $CD4^+$ T cells

Human primary $CD4^+$ T cells were extracted from PBMCs with MojoSort Human CD4 T Cell Isolation Kit (BioLegend). To induce *ZEB2* overexpression, human primary $CD4^+$ T cells were nucleofected with 0.5 mg mock ZsGreen plasmid vector (TaKaRa) or *ZEB2* ZsGreen plasmid vector per 2×10^6 cells using the P3 Primary Cell 4D-Nucleofecton X Kit S (Lonza). After incubation for 48 hours at 37°C in a 5% CO_2 incubator, ZsGreen-positive or -negative cells were sorted by flow cytometry, and the expression of *TBX21*, *ZEB2*, *GZMA*, *PRF1*, and *CXCL13* was evaluated by reverse-transcription qPCR (RT-qPCR). To induce *TBX21* overexpression, human

primary $CD4^+$ T cells were nucleofected with 0.5 mg mock DsRed plasmid vector (TaKaRa) or *TBX21* DsRed plasmid vector per 2×10^6 cells, and after 48 hours of culture, DsRed-positive cells were sorted by flow cytometry and evaluated by RT-qPCR. To induce both *ZEB2* and *TBX21* overexpression, nucleofection was performed using a combination of two of the four plasmid vectors (mock ZsGreen, mock DsRed, *ZEB2* ZsGreen, and *TBX21* DsRed), and after 48 hours of culture, cells positive for both ZsGreen and DsRed were sorted by flow cytometry and evaluated by RT-qPCR. The cells were cultured for 48 hours in 96-well plates that had been coated overnight with 1 μg/mL anti-human CD3 antibody (Thermo Fisher Scientific) and 2 μg/mL anti-human CD28 antibody (BD). The culture medium was Iscove's Modified Dulbecco's Medium (IMDM) (Thermo Fisher Scientific) supplemented with 10% fetal bovine serum (BioWest), 100 μg/mL streptomycin, 100 U/mL penicillin G, and 0.3 mg/mL L-glutamine (Thermo Fisher Scientific).

RT-qPCR

Total RNA was extracted from sorted cells using a MagMAX-96 Total RNA Isolation Kit (Thermo Fisher Scientific), and complementary DNA was then synthesized using random primers (Thermo Fisher Scientific) and Superscript IV (Thermo Fisher Scientific) according to the manufacturer’s protocols. RT-qPCR was performed using the CFX Connect Real-Time PCR detection system (Bio-Rad Laboratories) and the intercalation method with SYBR Green (Qiagen). Relative mRNA abundance was determined by normalization to glyceraldehyde-3-phosphate dehydrogenase expression. The primer sequences used in this study are shown in table S8.

scRNA-seq of peripheral blood $CD4^+$ T cells from HCs

$CD4^+$ T cells from PBMCs obtained from HCs were sorted by flow cytometry, and libraries were prepared using a Chromium Next GEM Single Cell 5′ Reagent Kit (v2 Dual Index; 10× Chromium). Sequencing was performed on a NovaSeq 6000 system (Illumina). The generated sequencing data were mapped with Cell Ranger (10× Genomics) using GRCh38 (Genome Reference Consortium) as the reference genome, and integrated analysis was then performed with Seurat (v3.2.3) (*87*). Data were adopted from cells with detected gene counts between 500 and 3,500, and data from cells with mitochondrial gene content greater than 5% were excluded. The number of dimensions was set based on the PCA and Elbowplot results, and dimensionality compression was performed using the unsupervised machine learning algorithm uniform manifold approximation and projection (UMAP). In this scRNA-seq, CITE-seq technology was simultaneously applied to evaluate cell surface protein expression using antibody-derived tags (ADTs). TotalSeq-C Custom Human Panel (BioLegend) shown in data file S4 was used for CITE-seq.

Collection and processing of tissue specimens from patients with IIM

Muscle specimens were transferred into a cryopreservation solution Cryostor CS10 (STEMCELL Technology), slow-frozen, and thawed for analysis at a later date. After thawing, the samples were shredded to 0.5–1 mm in Advanced DMEM (Thermo Fisher Scientific) containing 0.25 mg/mL Liberase TL (Roche) and 0.5 mg/mL DNase I (Roche). The shredded samples were heated in a 37°C water bath for 5 min. The treated samples were diluted in RPMI 1640 (Thermo Fisher Scientific) supplemented with 10% fetal bovine serum (BioWest), 100 μg/mL streptomycin, 100 U/mL penicillin G, and 0.3 mg/mL L-glutamine (Thermo Fisher Scientific) to inactivate Liberase TL, filtered through a 30-mm cell strainer.

BALF specimens were filtered through a 40-mm cell strainer, hemolyzed with ACK Lysing Buffer (Thermo Fisher Scientific), slowly frozen in cryopreservation solution containing DMSO, and later thawed for analysis.

After the above treatments, each sample was treated with anti-Fcγ receptor antibody (Thermo Fisher Scientific), stained with anti-CD45-BUV395 antibody (BD) at 4°C for 20 minutes and a LIVE/DEAD Fixable Aqua Dead Cell Stain Kit (Thermo Fisher Scientific) for 15 minutes at 20°C, and then sorted for CD45-positive live cells using flow cytometry (BD FACSAria Fusion cell sorter; BD Biosciences).

scRNA-seq of tissue specimens from patients with IIM

Libraries were generated from CD45-positive viable immune cells sorted from muscle and BALF specimens from patients with IIM using a Chromium Next GEM Single Cell 5′ Reagent Kit (v2 Dual Index; 10× Chromium). Sequencing was performed on a NovaSeq 6000 system (Illumina). The generated sequencing data were mapped with Cell Ranger (10× Genomics) using GRCh38 (Genome Reference Consortium) as the reference genome, and integrated analysis was then performed with Seurat (v3.2.3) (R package). Data were adopted from cells with detected gene counts between 1,000 or 1,500 (BALF or muscle, respectively) and 5,500, and data from cells with mitochondrial gene content greater than 5% were excluded. Data were dimensionally compressed using UMAP. Pseudotime trajectory analysis for each cell was estimated using Monocle3 (v1.0.0) (R package) (*88*). Gene set enrichment analysis (GSEA) was performed with Escape (v1.2.0) (R package) (*89*). As for the Human Gene Set in GSEA, proliferative capacity (HALLMARK_E2F_TARGETS) and oxidative phosphorylation (HALLMARK_OXPHOS) were evaluated. In these scRNA-seq data, CITE-seq technology was simultaneously applied to

evaluate cell surface protein expression using ADTs. TotalSeq-C Human TBNK Cocktail (BioLegend) shown in data file S4 was used for CITE-seq.

Analysis from publicly available scRNA-seq data

Using the publicly available scRNA-seq data browser, we analyzed $CD4^+$ T cells in the peripheral blood of patients with coronavirus disease-2019 (COVID-19) and $CD4^+$ T cells infiltrating the renal lesions of patients with SLE. Memory $CD4^+$ T-cell scRNA-seq data were extracted from publicly available data of blood from patients with COVID-19 (*54*). Feature plots of *ZEB2*, *CD28*, *TBX21*, *ADGRG1*, *PDCD1*, and *GZMA* and cell frequency according to days after onset were analyzed with a cellxgene browser (https://cellxgene.cziscience.com/e/21d3e683-80a4-4d9b-bc89-ebb2df513dde.cxg/). Gene expression in $CD4^+$ T cells was evaluated using publicly available scRNA-seq data for lupus nephritis specimens provided by the Accelerating Medicines Partnership (AMP) (*60*). Data were analyzed using the following website: https://singlecell.broadinstitute.org/single_cell/study/SCP279/.

Expression of IFN-related genes (IRGs) in patients with SLE and HCs (related to Fig. 5)

In this analysis, we used the same pipeline as in our previous study (*57*). We evaluated the expression of 100 IRGs reported in a recent PBMC scRNA-seq study (*58*) across case-control data and nine cell types from the $CD4^+$ T-cell lineage in the ImmuNexUT data (Fig. 5A). Expression levels were scaled for each gene across the nine cell types. Cell types were arranged based on the hierarchical clustering of IRGs and cell types using IRG expression. For the

hierarchical clustering, Euclidean distances of the scaled expression levels were used with Ward's method.

Association of transcriptome data with SLE disease status (related to Fig. 5)

In this analysis, we used the same pipeline as in our previous study (*57*). To quantify transcriptome variations within HCs, within patients with SLE, and between HCs and patients with SLE, we conducted PCA, calculated the proportion of sum squared deviations for each PC score, and then inferred the average value of the proportion weighted by the eigenvalue of each PC (Fig. 5B). We used PC1–7 for the analysis, which showed significant associations with clinical parameters in a previous study (*57*). To calculate the explained variance of each clinical parameter for whole transcriptome variation within SLE, we fitted each PC score to the linear mixed models in R package variancePartition (v1.20.0) (*90*) with the following regression models and then inferred the average value of the explained variance weighted by the eigenvalue of each PC for each cell type (Fig. 5C, F, G).

$$\text{Fig. 5C: } y = \beta \cdot x\,[0,1,2,3] + \sum_{k=1,2,3} \gamma \cdot I_k[0,1] + \delta \cdot PSL[mg] + \varepsilon \cdot Age[yrs] + \epsilon \cdot Sex[0,1] + \theta$$

For overall disease activity, we defined four categories: i) inactive as SLEDAI-2K (*39*) = 0, ii) low disease activity (LDA) as $1 \leq$ SLEDAI-2K ≤ 4, iii) moderate disease activity (MDA) as $5 \leq$ SLEDAI-2K ≤ 8, and iv) high disease activity (HDA) as SLEDAI-2K ≥ 9. Here, x represents disease activity (inactive, LDA, MDA, and HDA) and $I_k(k = 1,2,3)$ represents each immunosuppressant (mycophenolate mofetil [MMF], hydroxychloroquine [HCQ], and tacrolimus [TAC]) as covariates.

$$\text{Fig. 5F: } y = \sum_{j=1\ldots7} \beta \cdot x_j[0,1] + \sum_{k=1,2,3} \gamma \cdot I_k[0,1] + \delta \cdot PSL[mg] + \varepsilon \cdot Age[yrs] + \epsilon \cdot Sex[0,1] + \theta$$

Here, $x_j (j = 1 \ldots 7)$ represents each organ/domain activity (constitutional, mucocutaneous, musculoskeletal, renal, extrarenal severe, hematological, and serological).

$$\text{Fig. 5G: } y = \sum_{k=1,2,3} \gamma \cdot I_k[0,1] + \beta \cdot x[0,1,2,3] + \delta \cdot PSL[mg] + \varepsilon \cdot Age[yrs] + \epsilon \cdot Sex[0,1] + \theta$$

To verify whether the inferred explained variance was not biased by outlier samples, we estimated the standard errors (SEs) of the explained variance using the jackknife resampling method (Fig. 5F, G). When we had *n* samples for one cell type, we recalculated the explained variance *n* times by excluding each one of the samples. We then evaluated the distribution of *n* explained variance, quantified the SE for each clinical parameter, and compared the explained variance in each cell type against the median explained variance across the nine cell types.

Differential gene expression analysis for patients with SLE and HCs (related to Fig. 5)

In this analysis, we used the same pipeline as in our previous study (*57*). To detect differentially expressed genes (DEGs) in each cell type, we fitted the TMM-normalized counts to the generalized linear models (GLMs) with negative binomial distribution using edgeR (v3.32.1) (*81*) (Fig. 5D). We defined (1) "disease-state signature genes" as significant DEGs between patients with inactive SLE and HCs and (2) "disease-activity signature genes" as significant DEGs between patients with HDA and inactive SLE for each cell type. The equations in the GLM were as follows:

(1) For disease-state (x: patients with inactive SLE versus HCs),

$$y = \beta \cdot x[0,1] + \varepsilon \cdot Age[yrs] + \epsilon \cdot Sex[0,1] + \vartheta \cdot Batch + \theta$$

(2) For disease-activity (x: patients with HDA versus inactive SLE),

$$y = \beta \cdot x[0,1] + \sum_{k=1,2,3} \gamma \cdot I_k[0,1] + \delta \cdot PSL[mg] + \varepsilon \cdot Age[yrs] + \epsilon \cdot Sex[0,1] + \vartheta \cdot Batch + \theta$$

Here, $I_k(k = 1,2,3)$ represents each immunosuppressant (MMF, HCQ, TAC) as covariates. Statistical significance was set at a false discovery rate (FDR) of less than 0.05 with the Benjamini-Hochberg method. In Fig. 5D, bar plots show (top) the number of the union of disease-state/activity signature genes and (bottom) the proportion of DEG types in each cell type. To examine the specificity of these signatures, we calculated the Jaccard similarity index in each cell type to quantify the shared genes between signatures; we considered that a gene was shared when it was included in both signatures with a concordant sign. Based on the proportion of DEGs and the Jaccard index, we found three different transcriptome perturbation patterns: disease-state dominant pattern, disease-activity dominant pattern, and shared pattern (*57*).

Pathway enrichment analysis for SLE-relevant signatures (related to Fig. 5)

In this analysis, we used the same pipeline as in our previous study (*57*). To test pathway enrichment in disease-state and activity signature genes, we performed over-representation analyses with one-sided Fisher's exact test in R package clusterProfiler (v.3.18.1) (*91*) (Fig. 5E). Statistical significance was set at an FDR of less than 0.05 with the Benjamini-Hochberg method. For pathway data sets, we used the MsigDB hallmark gene set collection (50 annotations) (*92*) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway (548 annotations) (*93*). To capture the cell-type-specific biology linked to disease-state and activity signatures genes, we set the union of both signature genes in all cell types as the background gene sets. To overview the biology unique to each signature, we first classified all significant KEGG pathways into 33 subcategories and then 4 main categories ("Immune disease system," "Non-immune disease system," "Metabolism," and "Cellular process") according to the definitions on the developer's website (https://www.genome.jp/kegg/pathway.html) and our previous study

(*57*). In Fig. 5E, we summarized the number of enrichments for the “Metabolism” and “Cellular process” categories.

Multiple linear regression analysis of antibody titers in patients with SLE

To examine the association between autoantibody titers in SLE and *CXCL13* expression in three $CD4^+$ T cell subsets (Th1, Tfh, and ThA), we conducted multiple linear regression analysis. Age, sex, and drug treatment, including treatment with PSL, TAC, CyA, HCQ, and MMF, were adjusted as covariates. All variables were analyzed after normalization. GraphPad Prism version 9 software was used for this analysis.

Supplementary Figures

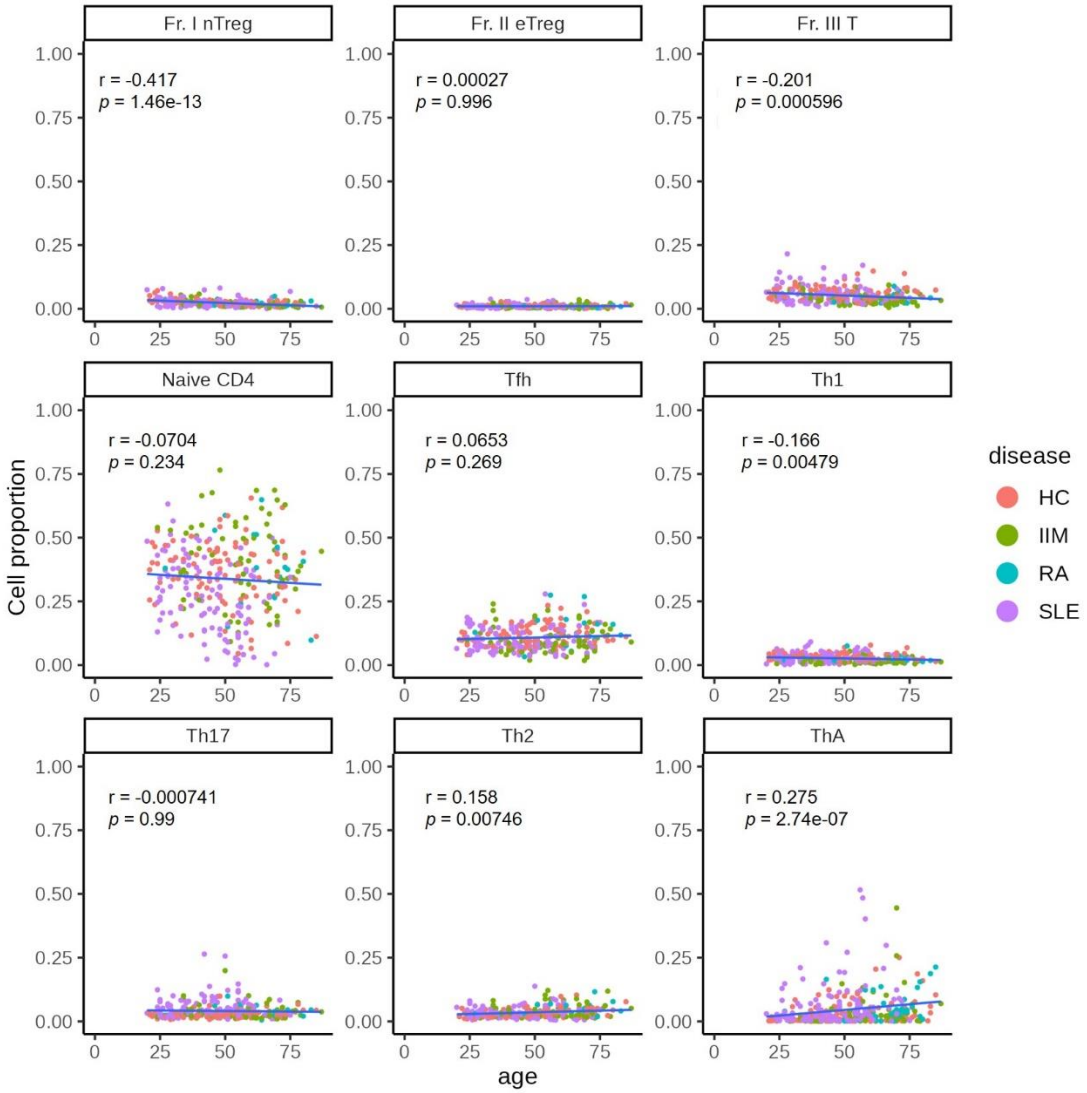

Figure S1. Distribution of age and the percentage of each of nine $CD4^+$ T cell subsets in flow cytometry. The data are presented for all flow cytometry data (HC = 124, RA = 44, SLE = 120, IIM = 66) included in the initial data set (**Fig. 1A**). The percentage of each cell subset was calculated as the percentage in the total $CD4^+$ T cells. The correlation coefficients and *p* values were determined using Spearman's rank correlations.

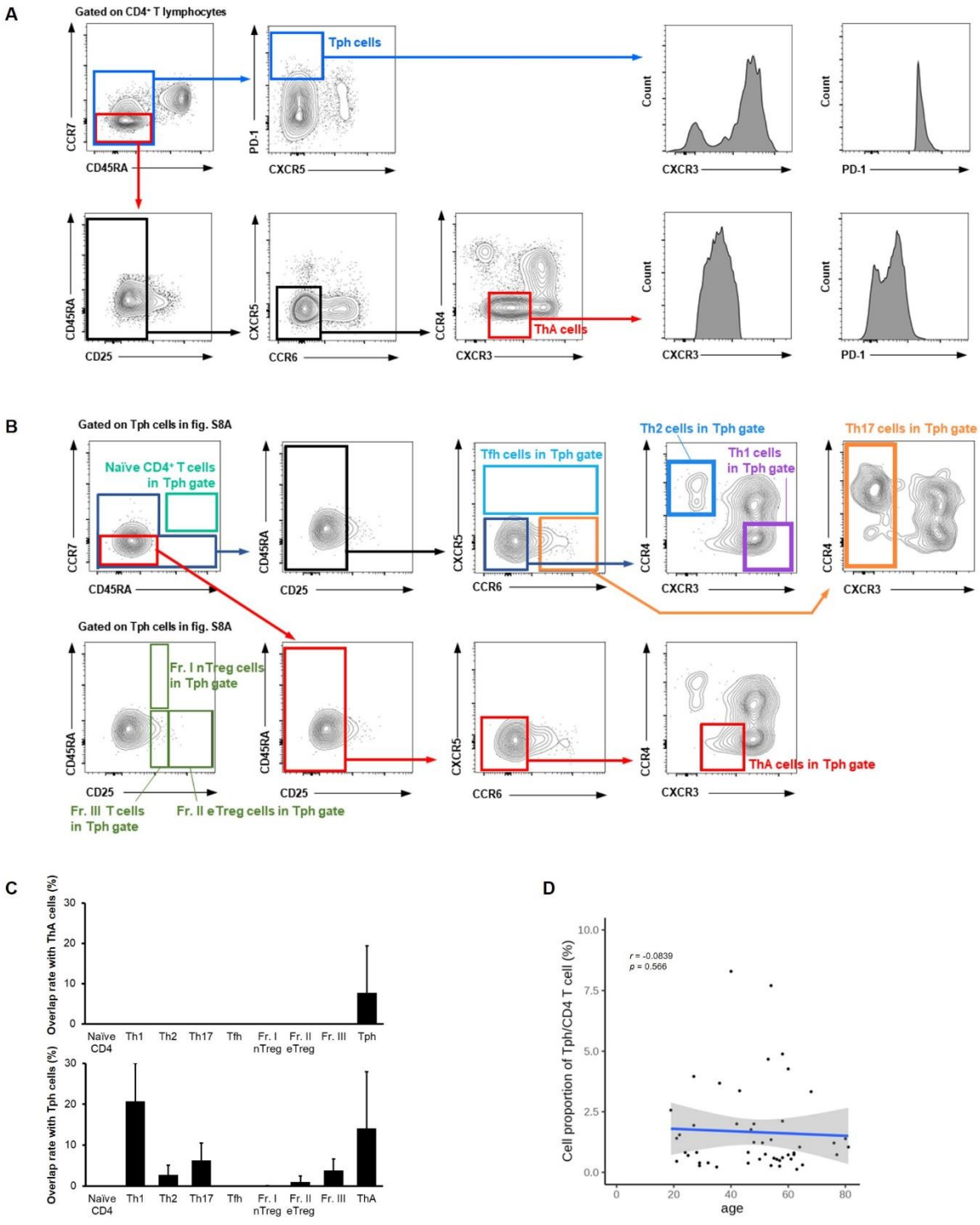

Figure S2. Examination of overlap between Tph cells and other CD4⁺ T cells. **(A)** Flow cytometric hierarchical gating strategy for Tph and ThA cells. Representative data from 54-years-old HC female is shown. **(B)** Flow cytometric hierarchical gating strategy for each

indicated T cell subsets gated on Tph cells fig. S2A. **(C)** Rate of overlap among $CD4^+$ T cell subsets from healthy controls and autoimmune disease patients according to flow cytometric analysis (mean ± SD, n = 49, HC: 22, SLE: 11, RA: 16). The gating strategy is as shown in **A** and **B**. **(D)** Distributions of age and the Tph cell proportion according to the flow cytometry data shown in **C**. The proportion of each cell subset was calculated as the percentage of the total $CD4^+$ T cells. The correlation coefficients and *p* values were determined using Spearman's rank correlations.

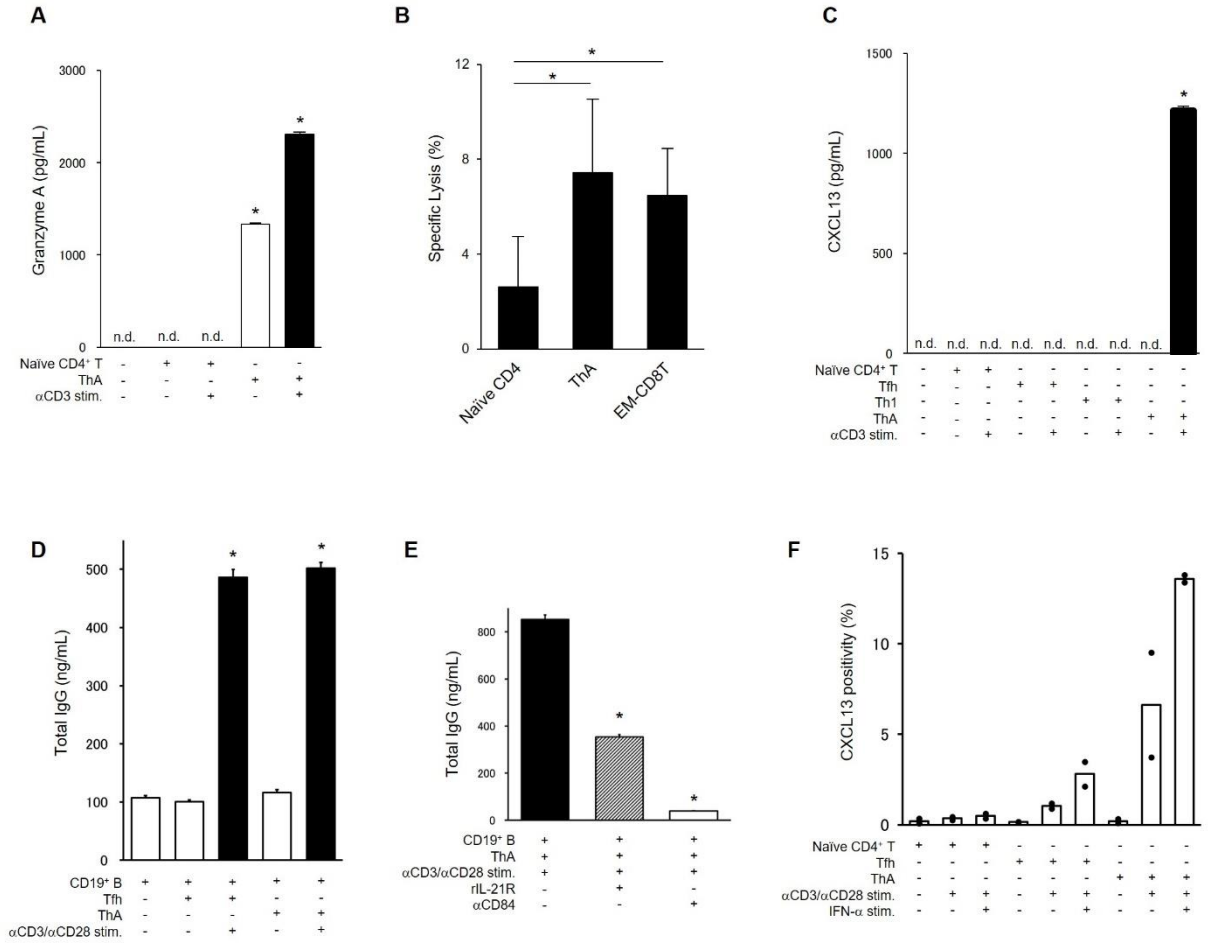

Figure S3. Functional analysis of ThA cells (related to Figures 2 and 3). (A) Granzyme A concentration in culture supernatant of naïve $CD4^+$ T cells and ThA cells obtained from HC, quantified by ELISA (mean ± SD, n = 4). *, $p < 0.05$, compared with naïve $CD4^+$ T cell culture conditions with αCD3 mAb stimulation. **(B)** Cytotoxic effect of αCD3 mAb-stimulated naïve $CD4^+$ T cells, ThA cells, or effector memory $CD8^+$ T cells from healthy donors on K562 target cells. Cytotoxic responses were evaluated by the lactate dehydrogenase (LDH) assay at an effector-to-target (E:T) ratio of 10:1 (n = 3 biological replicates). *$p < 0.05$ by one-way repeated measures ANOVA with Fisher's LSD post hoc test. **(C)** CXCL13 concentration in culture supernatant of $CD4^+$ T cell subsets obtained from RA patients, quantified by ELISA (mean ±

SD, n = 4). *, $p < 0.05$, compared with naïve $CD4^+$ T cell culture conditions with αCD3 mAb stimulation. **(D)** Total IgG concentration in culture supernatant of $CD4^+$ T cell subset co-cultured with B cells obtained from HC, quantified by ELISA (mean ± SD, n = 4). *, $p < 0.05$, compared with the B cell alone culture condition. **(E)** Total IgG concentration in culture supernatant during co-culture of ThA and B cells obtained from HC with or without IL-21 neutralization or CD84 blocking, quantified by ELISA (mean ± SD, n = 4). *, $p < 0.05$, compared with the co-culture condition of ThA and B cells under αCD3/αCD28 mAb stimulation. **(F)** Intracellular CXCL13-positivity rates for each $CD4^+$ T cell subset obtained from RA patients after αCD3 mAb/αCD28 mAb and/or IFN-α stimulation (n = 2 biological replicates). **(A, C–E)** Data are representative of three experiments (four technical replicates of three biological replicates). Statistical significance was calculated by one-way ANOVA with Tukey multiple comparison tests.

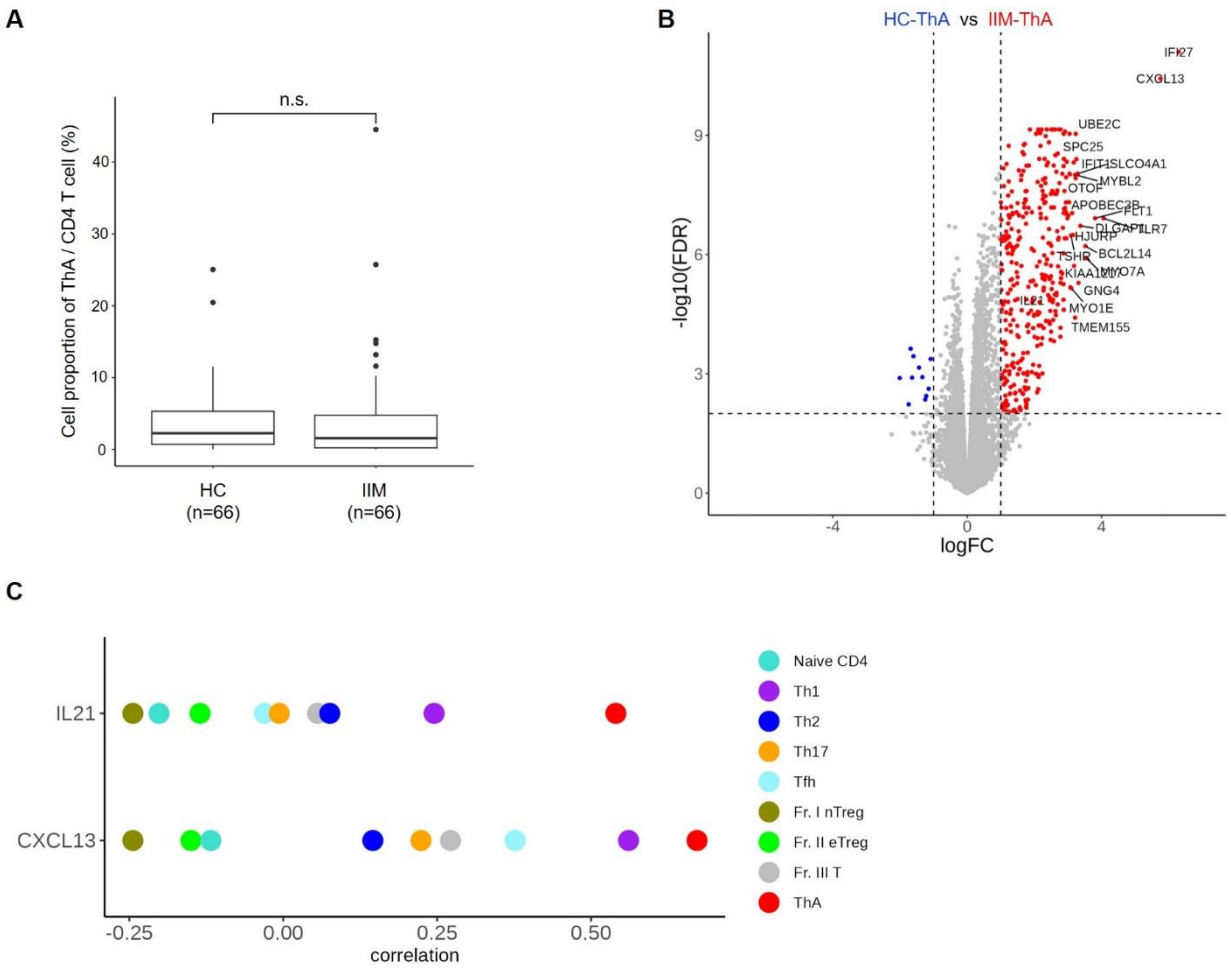

Figure S4. Comparison of IIM ThA cells and HC ThA cells in the initial data set. (A) Comparison of the proportion of ThA cells in PB $CD4^+$ T cells between age/sex matched HC and IIM cases (n = 66 each). *, $p < 0.05$, Wilcoxon rank sum test. n.s.: not significant. **(B)** Volcano plot showing DEGs between HC-ThA and IIM-ThA cells. **(C)** Pearson correlation coefficients between *IL21* and *CXCL13* gene expression and disease state in IIM patients (n = 40). Correlation coefficients were calculated with cases evaluated at onset as 1 and those evaluated at other times as 0.

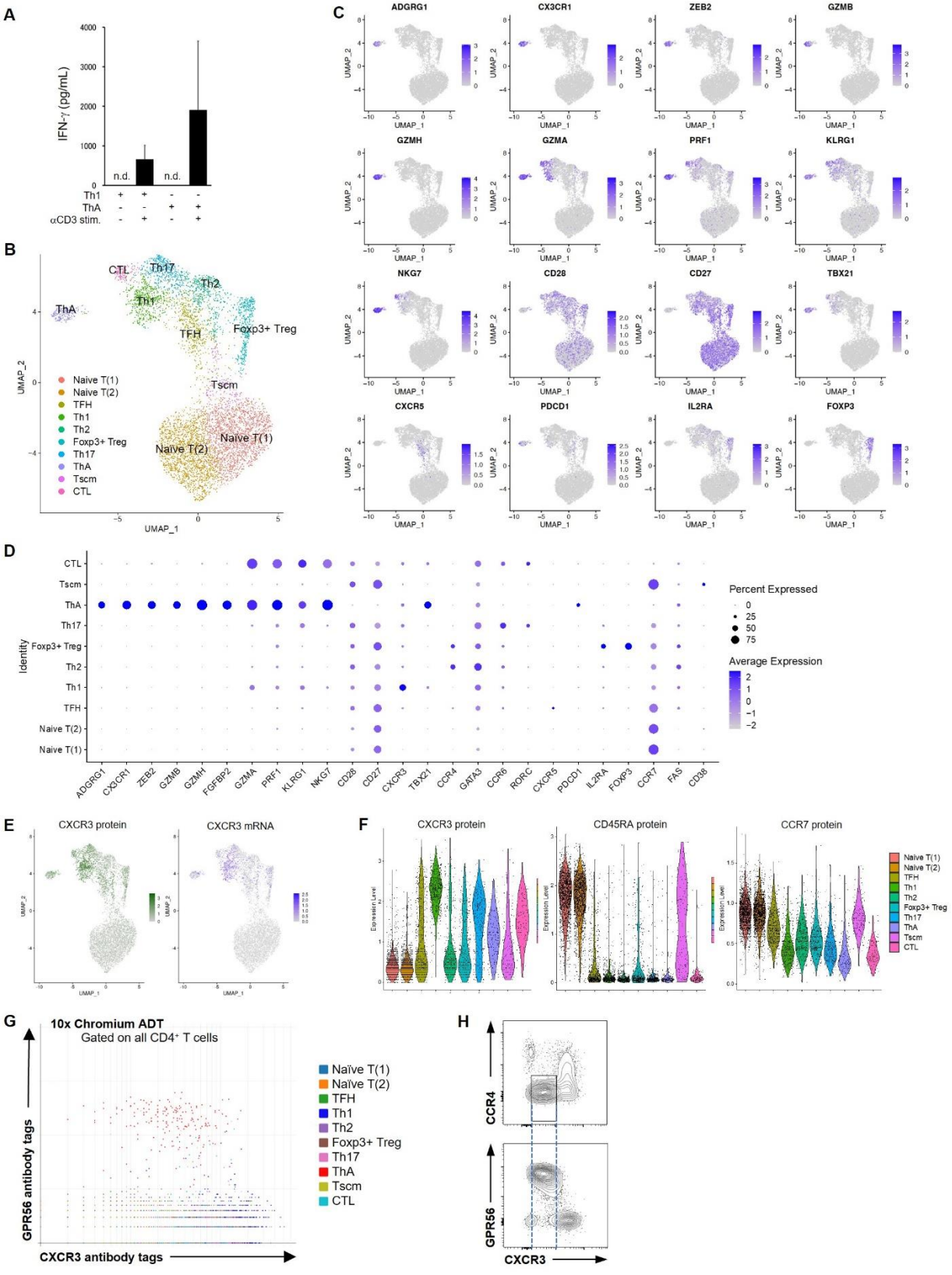
A
IFN-γ (pg/mL)
n.d.
n.d.
Th1
ThA
αCD3 stim.
B
CTL
Th17
Th2
Th1
ThA
TFH
Foxp3+ Treg
Tscm
Naive T(1)
Naive T(2)
UMAP_1
UMAP_2
C
ADGRG1
CX3CR1
ZEB2
GZMB
GZMH
GZMA
PRF1
KLRG1
NKG7
CD28
CD27
TBX21
CXCR5
PDCD1
IL2RA
FOXP3
D
Identity
Percent Expressed
Average Expression
ADGRG1
CX3CR1
ZEB2
GZMB
GZMH
FGFBP2
GZMA
PRF1
KLRG1
NKG7
CD28
CD27
CXCR3
TBX21
CCR4
GATA3
CCR6
RORC
CXCR5
PDCD1
IL2RA
FOXP3
CCR7
FAS
CD38
E
CXCR3 protein
CXCR3 mRNA
F
CXCR3 protein
CD45RA protein
CCR7 protein
Expression Level
G
10x Chromium ADT
Gated on all CD4⁺ T cells
GPR56 antibody tags
CXCR3 antibody tags
Naïve T(1)
Naïve T(2)
TFH
Th1
Th2
Foxp3+ Treg
Th17
ThA
Tscm
CTL
H
CCR4
GPR56
CXCR3

Figure S5. Characterization of ThA cells in the peripheral blood of healthy controls. (A) Interferon-γ concentration in culture supernatant of Th1 and ThA cells obtained from HC, quantified by ELISA (mean ± SD, n = 3 biological replicates). No significant difference, one-way ANOVA. **(B, C)** UMAP **(B)** and feature plots **(C)** of scRNA-seq data from PB $CD4^+$ T cells obtained from a HC. **(D)** Dot plots showing gene expression for each of the ten $CD4^+$ T-cell clusters in **B**. **(E)** Feature plots of CXCR3 protein (left) and mRNA (right) expression according to the scRNA-seq analysis of PB $CD4^+$ T cells shown in **B**. CXCR3 protein expression levels were evaluated using antibody-derived tags (ADTs) via CITE-seq analysis of the scRNA-seq data. **(F)** Cell surface CXCR3, CD45RA, and CCR7 protein expression evaluated using ADTs via CITE-seq analysis of the specimens used in **B**. **(G)** Cell surface CXCR3 and GPR56 protein expression evaluated using ADTs via CITE-seq analysis of the scRNA-seq data from PB $CD4^+$ T cells shown in **B**. **(H)** Representative flow cytometry data showing cell surface CXCR3 and GPR56 expression from HC. The gating strategy is the same as **Fig. 1B**. Rectangles indicate ThA cells.

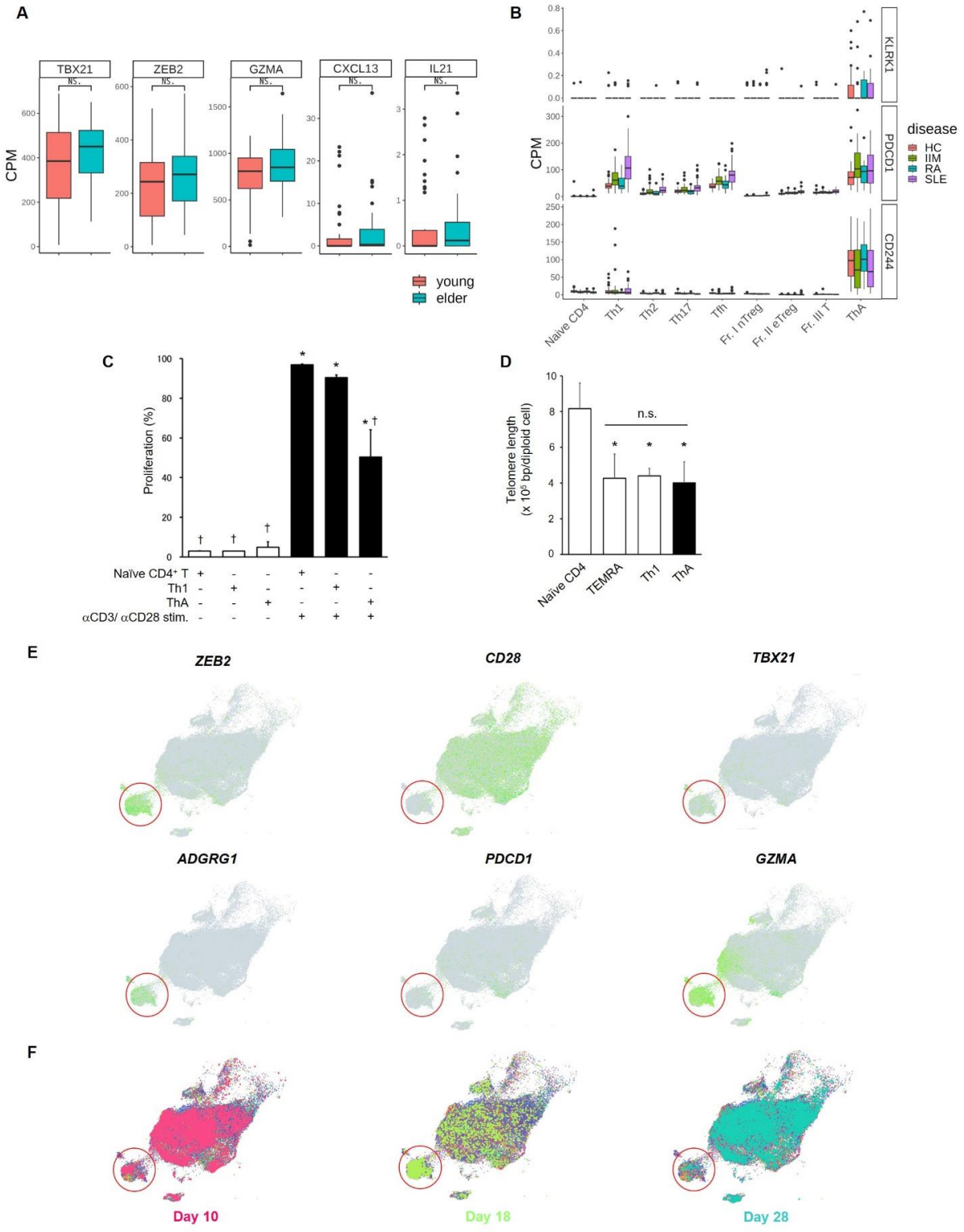

Figure S6. ThA cells in T cell senescence and COVID-19 infection. (A) Expression of the indicated genes in ThA cells according to RNA-seq data from the initial data set (**Fig. 1A**).

Younger and older age groups were defined as under 45 (n = 40) and over 60 (n = 46) years of age, respectively. CPM: counts per million. *, $p < 0.05$, Student's *t*-test. **(B)** Expression of genes related to T cell senescence and exhaustion in bulk RNA-seq data from the initial data set (**Fig. 1A**). **(C)** Percentage of proliferation of three $CD4^+$ T cell subsets evaluated with a CFSE labeling assay after five days of culture with or without αCD3/αCD28 mAb stimulation (mean ± SD, n = 3 biological replicates). *, $p < 0.05$, compared to Naïve $CD4^+$ T without αCD3/αCD28 stim.; †, $p < 0.05$, compared to Naïve $CD4^+$ T with αCD3/αCD28 stim.; one-way ANOVA with Tukey multiple comparison tests. **(D)** Telomere length of genomic DNA of four $CD4^+$ T cell subsets quantified by Absolute Human Telomere Length Quantification qPCR Assay Kit (ScienCell). Cells were obtained from three HCs (mean ± SD, n = 3). *, $p < 0.05$, compared with naïve $CD4^+$ T cells, one-way ANOVA with Tukey multiple comparison tests. n.s., not significant. Data are representative of three experiments (three technical replicates of three biological replicates). **(E, F)** Memory $CD4^+$ T cell scRNA-seq data extracted from publicly available COVID-19-infected patients' blood (*Cell.* 2021;184(7):1836-1857.e22.). Data were analyzed using the cellxgene browser (https://cellxgene.cziscience.com/e/21d3e683-80a4-4d9b-bc89-ebb2df513dde.cxg/). Feature plot of *ZEB2*, *CD28*, *TBX21*, *ADGRG1*, *PDCD1*, *GZMA* **(D)** and cell frequency according to days after onset (day 10: red, day 18: pale green, day 28: green) **(E)** are shown.

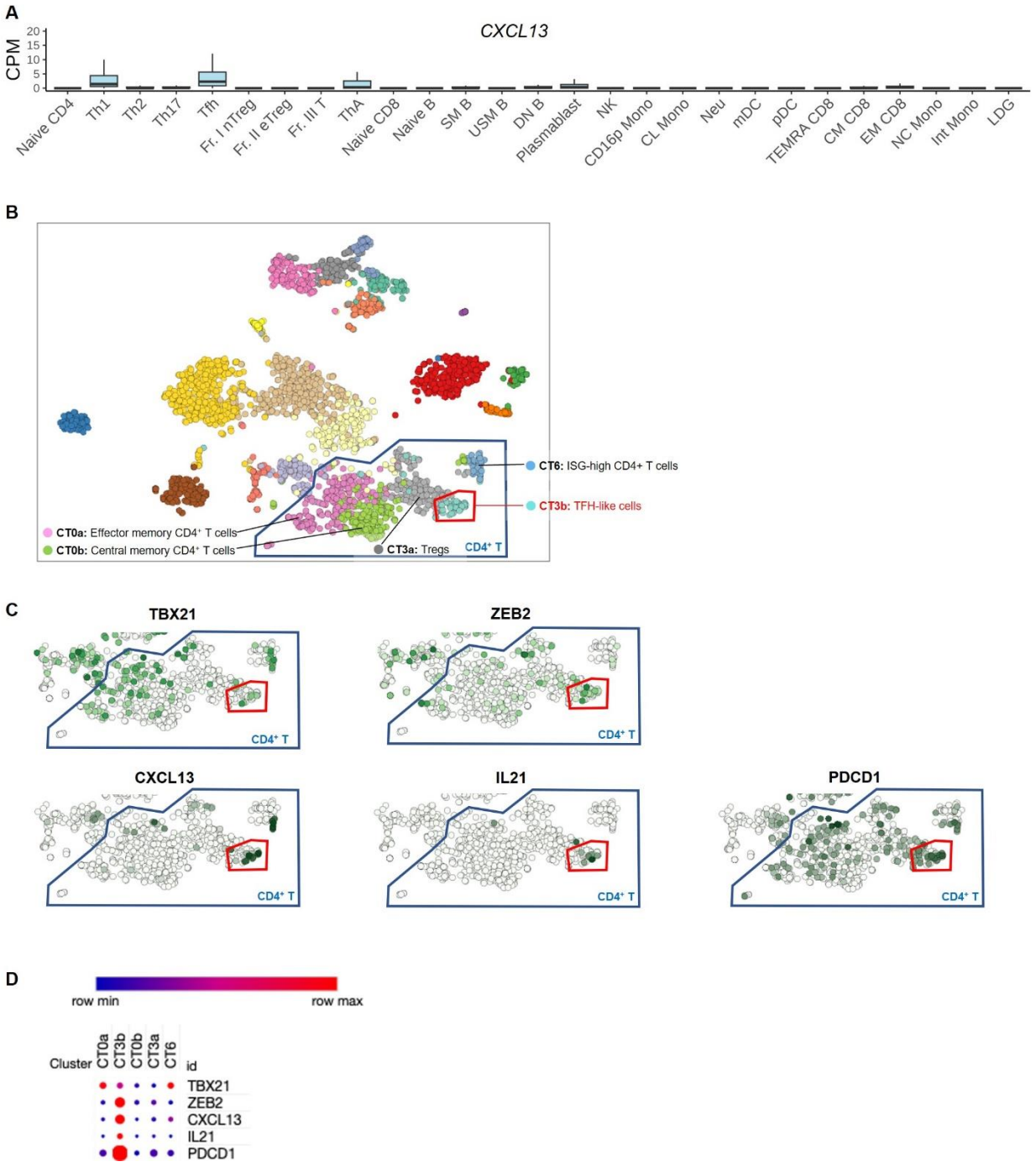

Figure S7. ThA cells infiltrate local organ lesions in SLE. (A) *CXCL13* gene expression for 27 immune cell subsets in 136 SLE patients (*Cell.* 2022;185(18):3375-3389), newly including ThA cells. **(B, C, D)** Gene expression of $CD4^+$ T cells from publicly available scRNA-seq data of

lupus nephritis specimens provided by the Accelerating Medicines Partnership (AMP). Annotations of each $CD4^+$ T cell cluster **(B)** and feature plot showing gene expression **(C)**. Dot plot summarizing gene expression in each $CD4^+$ T cell cluster **(D)**. **B, C, D** were analyzed by the website https://singlecell.broadinstitute.org/single_cell/study/SCP279/.

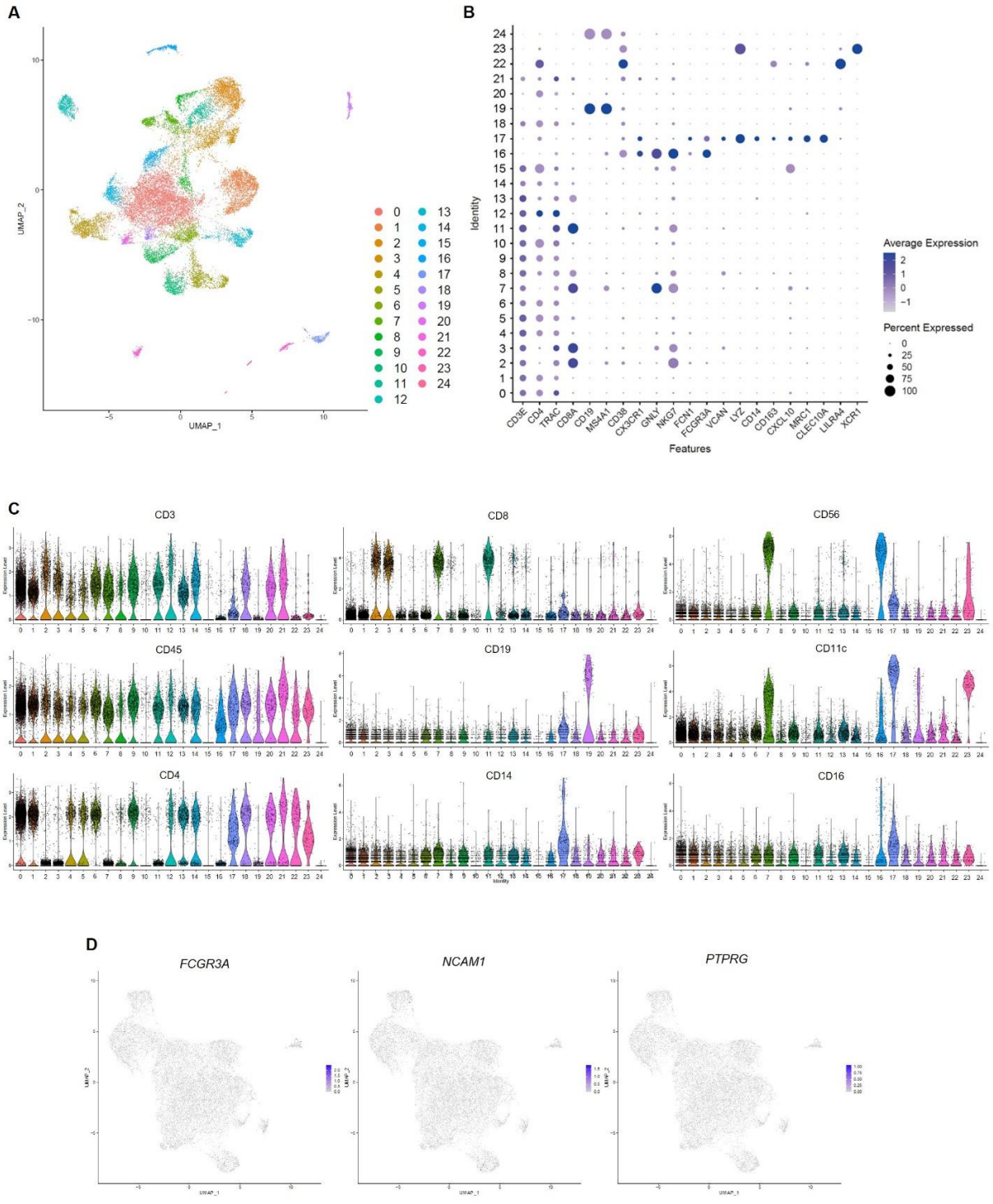

Figure S8. Supplementary data for scRNA-seq of BALF cells and (related to Figure 6).

(A) UMAP of scRNA-seq data for CD45^{+} cells from BALF of three patients with anti-ARS antibody-positive IIM related to **Fig. 6A**. Twenty-five clusters were identified. **(B)** Dot plots

showing gene expression of cell subset signature genes across cell clusters in **A**. **(C)** Cell surface protein expression evaluated with antibody-derived tags (ADTs) using CITE-seq technology on specimens in **A**. **(D)** UMAP plots of the scRNA-seq data from PB $CD4^+$ T cells shown in **Fig. 6B**.

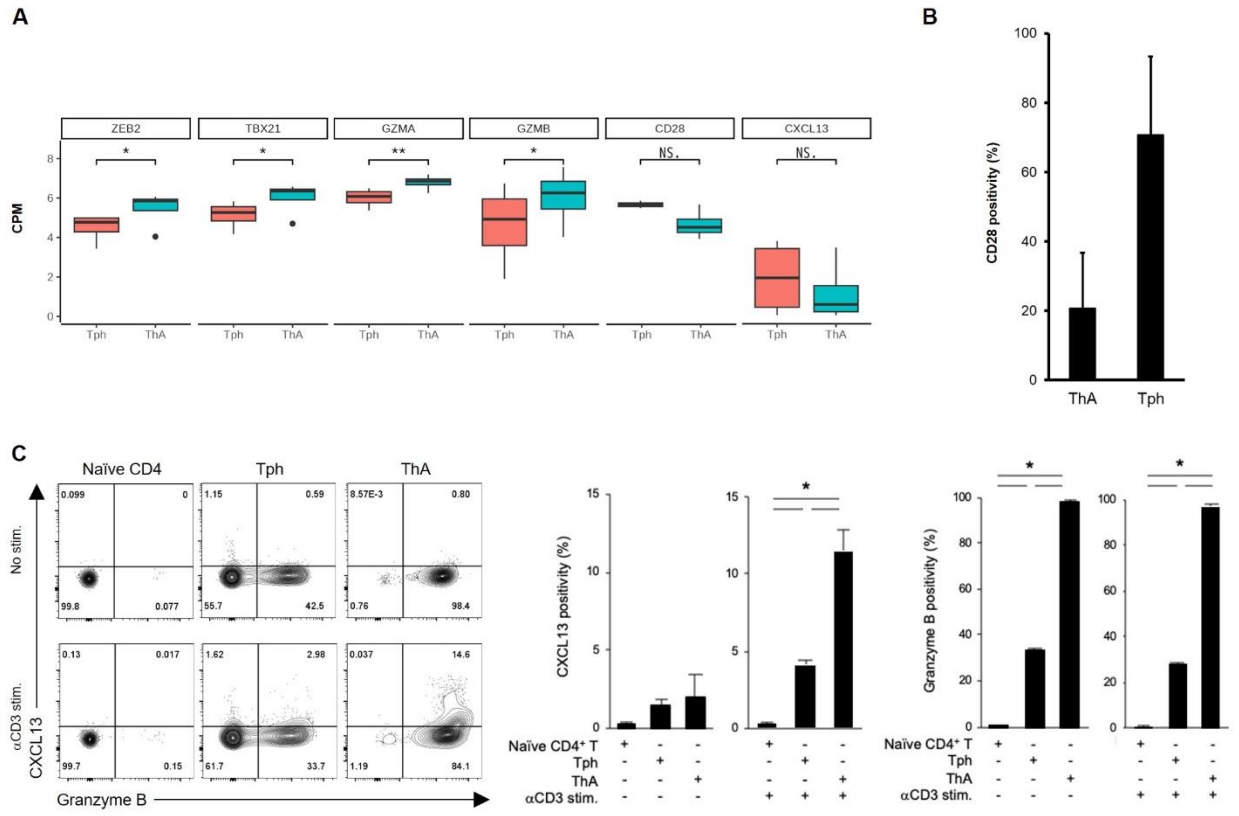

Figure S9. Comparison between ThA and Tph cells (A) Expression of the indicated genes in ThA and Tph cells based on the RNA-seq data from autoimmune disease patients (n = 4: 2 for IIM and 2 for RA). CPM: counts per million. *, $p < 0.05$, two-sided unpaired *t*-test. **(B)** Cell surface CD28-positivity in PB ThA and Tph cells in RA patients (mean ± SD, n = 4 biological replicates). CD28 positivity of ThA cells ($CXCR3^{mid}CD25^{-}CXCR5^{-}CCR6^{-}$ $CD4^{+}$ Tem) and Tph cells ($CXCR5^{-}PD\text{-}1^{hi}$ $CD4^{+}$ T) was calculated from PBMC flow cytometry data. *, $p < 0.05$, two-sided unpaired *t*-test. **(C)** Representative flow cytometric data showing the expression of intracellular CXCL13 and granzyme B proteins in the indicated PB $CD4^{+}$ T cell subsets from HC individuals after αCD3 mAb stimulation. Bar graphs indicate the mean ± SD CXCL13 and granzyme B expression levels in ThA or Tph cells (n = 3 biological replicates). *$p < 0.05$ by one-way repeated measures ANOVA with Fisher's LSD post hoc test.

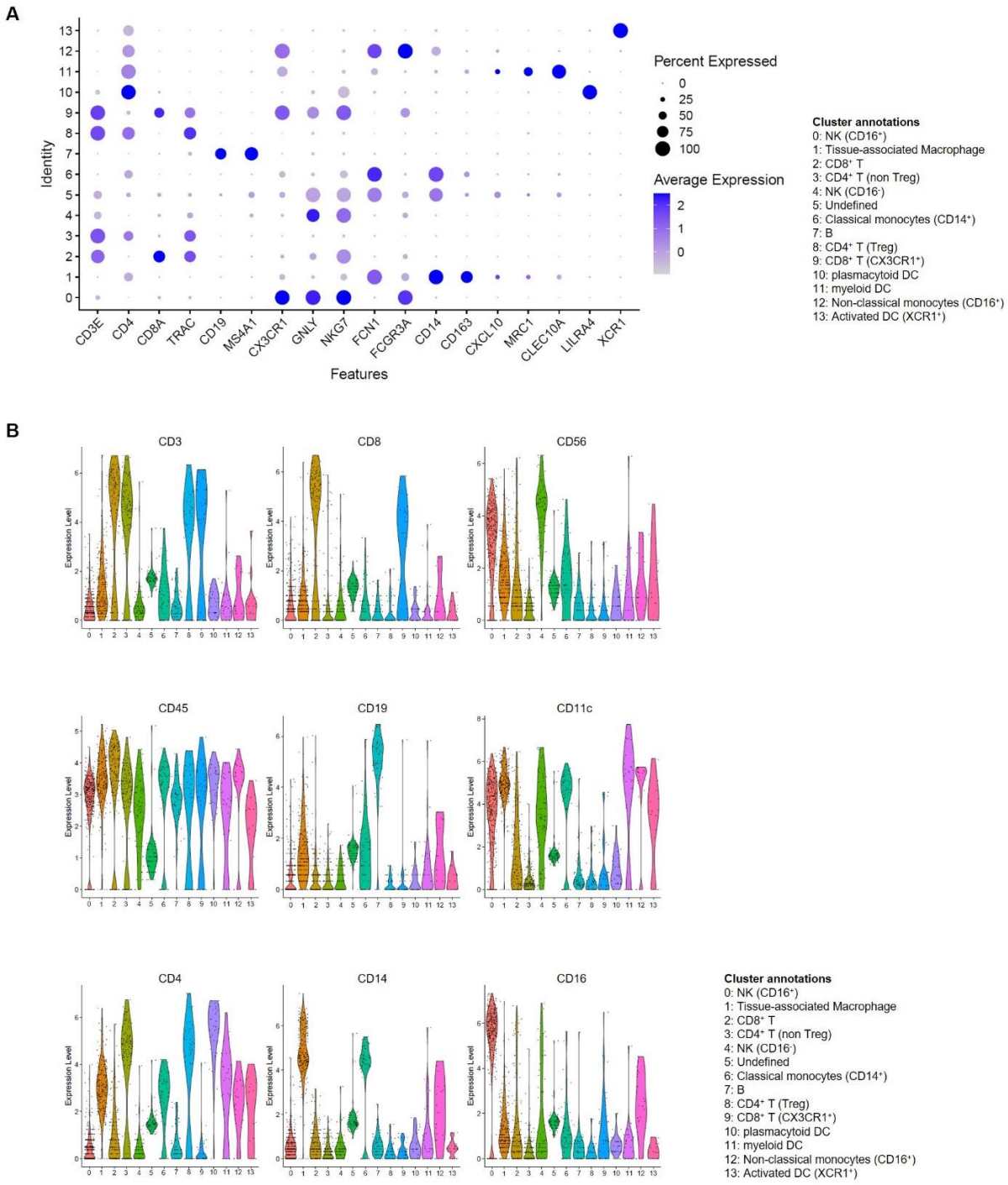

Figure S10. Supplementary data for scRNA-seq of muscle-infiltrating cells from anti-ARS antibody-positive IIM patients (related to Figure 6). (A) Dot plots showing gene expression

of cell subset signature genes across cell clusters in **Fig. 6F**. **(B)** Cell surface protein expression evaluated with ADTs using CITE-seq technology using specimens in **Fig. 6F**.

Supplementary Tables

Table S1. Clinical features of initial data set, Related to Figure 1 - 4

In right three columns, values are median, and inter quartile ranges are in parentheses. PSL, prednisolone.

【RNA-seq (N = 185)】

Diagnosis	Total (n)	Female (n (%))	PSL (mg)	Age (year)	Disease duration (year)
Healthy control	64	48 (75.0)	-	48 (37.8-58)	-
Rheumatoid arthritis	19	18 (94.7)	2.5 (0-5)	64 (55.5-72)	10.5 (0.9-22.5)
Systemic lupus erythematosus	62	57 (91.9)	5 (3-8)	47 (37.5-57.5)	12 (4-20.8)
Idiopathic inflammatory myopathy	40	30 (75.0)	2 (0-6)	58 (52-70)	3.5 (0-12.5)

【Flow cytometry (N = 354)】

Diagnosis	Total (n)	Female (n (%))	PSL (mg)	Age (year)	Disease duration (year)
Healthy control	124	88 (71.0)	-	51 (40-64.5)	-
Rheumatoid arthritis	44	38 (86.4)	2 (0-4)	69 (60.2-73.2)	15 (3.6-26)
Systemic lupus erythematosus	120	105 (87.5)	5 (2.5-8)	44 (34-55)	14 (7-21)
Idiopathic inflammatory myopathy	66	48 (72.7)	5 (0-8)	57 (45-68.5)	3 (0-11.8)

Table S2. Fluorescent-labeled flow-cytometry antibodies,

Used to construct initial data set in this study, Related to Figure 1 - 4

marker	conjugate	clones	source	identifier	RRID
CD25	FITC	M-A251	BD	#560990	AB_2033954
CXCR5	PerCP/Cy5.5	RF8B2	BD	#562781	AB_2313576
CD3	PE/Cy7	UCHT1	BioLegend	#300419	AB_439780
CCR6	APC	11A9	BD	#560619	AB_1727439
CD45RA	APC/Cy7	HI100	BioLegend	#304128	AB_10708880
CXCR3	BV421	1C6	BD	#562558	AB_2737653
CD4	BV510	RPA-T4	BD	#740161	AB_2739914
CCR4	BV605	L291H4	BioLegend	#359418	AB_2562483
CCR7	BV711	G043H7	BioLegend	#353228	AB_2563865
CD8	BUV805	RPA-T8	BD	#749366	AB_2873737

Used to analyze cell proportion of Tph and other CD4$^+$ T subsets,

Related to Figure S2A, S2B, S2C, and S2D

marker	conjugate	clones	source	identifier	RRID
CD25	FITC	M-A251	BD	#560990	AB_2033954
CXCR5	PerCP/Cy5.5	RF8B2	BD	#562781	AB_2313576
CD3	PE/Cy7	UCHT1	BioLegend	#300419	AB_439780
CCR6	APC	11A9	BD	#560619	AB_1727439
CD45RA	APC/Cy7	HI100	BioLegend	#304128	AB_10708880
CXCR3	BV421	1C6	BD	#562558	AB_2737653
CCR4	BV605	L291H4	BioLegend	#359418	AB_2562483
PD-1	BV650	EH12.2H7	BioLegend	#329950	AB_2566362
CCR7	BV711	G043H7	BioLegend	#353228	AB_2563865
CD8	BUV805	RPA-T8	BD	#749366	AB_2873737
CD4	BUV661	SK3	BD	#612962	AB_2739452

Table S3. Immune cell definition in our cohort

Subset name	Abbreviation	Definition
Naïve $CD4^+$ T cells	Naïve CD4	$CD3^+/CD4^+CD8^-/CCR7^+CD45RA^+$
T helper 1 cells	Th1	$CD3^+/CD4^+CD8^-$/non-naïve $CD4^+/CD25^-/CXCR5^-CCR6^-/CXCR3^+CCR4^-$
T helper 2 cells	Th2	$CD3^+/CD4^+CD8^-$/non-naïve $CD4^+/CD25^-/CXCR5^-CCR6^-/CXCR3^-CCR4^+$
T helper 17 cells	Th17	$CD3^+/CD4^+CD8^-$/non-naïve $CD4^+/CD25^-/CXCR5^-CCR6^+/CXCR3^-$
T follicular helper cells	Tfh	$CD3^+/CD4^+CD8^-$/non-naïve $CD4^+/CD25^-/CXCR5^+$
Fraction I naïve regulatory T cells	Fr.I nTreg	$CD3^+/CD4^+CD8^-/CD25^+CD45RA^+$
Fraction II effector regulatory T cells	Fr.II eTreg	$CD3^+/CD4^+CD8^-/CD25^{++}CD45RA^-$
Fraction III non-regulatory T cells	Fr.III T	$CD3^+/CD4^+CD8^-/CD25^+CD45RA^-$
Age-associated helper T cells	ThA	$CD3^+/CD4^+CD8^-/CCR7^-CD45RA^-/CD25^-/CXCR5^-CCR6^-/CXCR3^{mid}$ $CCR4^-$

Non-naïve $CD4^+$ refers to $CD4^+$ T cells other than naïve ($CCR7^+CD45RA^+$).

Table S4. Multiple linear regression analysis of antibody titer in SLE patients,

Related to Figure 5

PSL; prednisolone, TAC; tacrolimus, CyA; cyclosporin A, HCQ; hydroxychloroquine, MMF; mycophenolate mofetil

【Total IgG】

Explanatory variables	Estimated coefficients	95% CI	|*t*|	Adjusted *P*-value
Clinical features				
Sex	0.1211	0.02844 to 0.2138	2.606	0.0112*
Age	-0.0001083	-0.1383 to 0.1381	0.001563	0.9988
Treatment drugs				
PSL	-0.2064	-0.3618 to -0.05096	2.648	0.0100*
TAC	-0.08172	-0.2104 to 0.04694	1.266	0.2095
CyA	0.1311	-0.1502 to 0.4124	0.9292	0.3559
HCQ	-0.04145	-0.1164 to 0.03352	1.102	0.2740
MMF	-0.09173	-0.2400 to 0.05658	1.233	0.2215
Gene expression				
Th1-CXCL13	0.06865	-0.1232 to 0.2605	0.7135	0.4778
Tfh-CXCL13	0.03683	-0.1452 to 0.2188	0.4035	0.6878
ThA-CXCL13	0.1613	0.008326 to 0.3143	2.102	0.0391*

【Anti-RNP antibody titer】

Explanatory variables	Estimated coefficients	95% CI	|*t*|	Adjusted *P*-value
Clinical features				
Sex	0.06960	-0.1482 to 0.2874	0.6375	0.5259
Age	-0.2098	-0.5490 to 0.1293	1.234	0.2213
Treatment drugs				
PSL	-0.02557	-0.3931 to 0.3420	0.1387	0.8901
TAC	0.07678	-0.2262 to 0.3797	0.5056	0.6148
CyA	-0.3303	-1.100 to 0.4398	0.8556	0.3952
HCQ	0.05889	-0.1159 to 0.2337	0.6721	0.5038
MMF	-0.1466	-0.4986 to 0.2054	0.8308	0.4090
Gene expression				
Th1-CXCL13	-0.1811	-0.6690 to 0.3069	0.7403	0.4616
Tfh-CXCL13	0.7193	0.2856 to 1.153	3.309	0.0015*
ThA-CXCL13	0.2316	-0.1489 to 0.6122	1.214	0.2288

【Anti-Sm antibody titer】

Explanatory variables	Estimated coefficients	95% CI	|*t*|	Adjusted *P*-value
Clinical features				
Sex	0.1224	-0.06315 to 0.3079	1.315	0.1927
Age	-0.5116	-0.7973 to -0.2258	3.569	0.0006*
Treatment drugs				
PSL	0.04077	-0.2740 to 0.3556	0.2582	0.7970
TAC	-0.3465	-0.6061 to -0.08699	2.662	0.0096*
CyA	-0.06637	-0.6307 to 0.4980	0.2345	0.8153
HCQ	-0.1173	-0.2625 to 0.02789	1.611	0.1116
MMF	0.04603	-0.2553 to 0.3473	0.3046	0.7615
Gene expression				
Th1-CXCL13	-0.04852	-0.4494 to 0.3524	0.2413	0.8100
Tfh-CXCL13	0.1200	-0.2474 to 0.4874	0.6512	0.5170
ThA-CXCL13	0.3702	0.05197 to 0.6885	2.320	0.0232*

Table S5. Clinical information on 3 BALF lymphocyte scRNA-seq cases,

Related to Figure 6

	Pt. 1	**Pt. 2**	**Pt. 3**
Age (years)	56	49	72
Sex	Male	Female	Female
Myositis-specific antibodies	anti-ARS	anti-ARS	anti-ARS
Duration of disease (years)	0.2	1.8	7
Complications of autoimmune disease	-	-	Sjogren syndrome
Distribution of involvement	arthritis, lung, skin, muscle	lung, skin, muscle	arthritis, lung, skin
Patterns of interstitial pneumonia	NSIP	Probable UIP	NSIP
Myositis	+	+	-
Immunosuppressive treatment at testing	none	none	none

Table S6. Clinical information on 3 muscle infiltrating lymphocytes scRNA-seq cases, Related to Figure 6

	Pt. 1	Pt. 2	Pt. 3
Age (years)	56	49	41
Sex	Male	Female	Female
Myositis-specific antibodies	anti-ARS	anti-ARS	anti-ARS
Duration of disease (years)	0.2	1.8	1.0
Complications of autoimmune disease	-	-	Sjogren syndrome
Distribution of involvement	arthritis, lung, skin, muscle	lung, skin, muscle	skin, muscle
Creatine phosphokinase (U/L)	289	530	12,768
Immunosuppressive treatment at testing	none	none	none

Table S7. Reagents and Resources

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
anti-human CCR4 BV605 (clone L291H4)	BioLegend	Cat#359418; RRID: AB_2562483
anti-human CCR6 APC (clone 11A9)	BD	Cat#560619; RRID: AB_1727439
anti-human CCR6 BV650 (clone 11A9)	BD	Cat#563922; RRID: AB_2738488
anti-human CCR6 BUV395 (clone 11A9)	BD	Cat#743356; RRID: AB_2741448
anti-human CCR7 BV711 (clone G043H7)	BioLegend	Cat#353228; RRID: AB_2563865
anti-human CD19 APC (clone HIB19)	BioLegend	Cat#302212; RRID: AB_314242
anti-human CD19 BUV805 (clone HIB19)	BD	Cat#742007; RRID: AB_2871305
anti-human CD19 PE (clone HIB19)	BioLegend	Cat#302208; RRID: AB_314238
anti-human CD25 FITC (clone M-A251)	BD	Cat#560990; RRID: AB_2033954
anti-human CD25 PE (clone M-A251)	BD	Cat#555432; RRID: AB_395826
anti-human CD27 BUV395 (clone M-T271)	BD	Cat#740291; RRID: AB_2740030
anti-human CD279 (PD-1) BV650 (clone EH12.2H7)	BioLegend	Cat#329950; RRID: AB_2566362
anti-human CD28 BUV805 (clone CD28.2)	BD	Cat#742037; RRID: AB_2871331
anti-human CD28 (clone CD28.2)	BD	Cat#556620; RRID: AB_396492
anti-human CD3 (clone OKT3)	Thermo Fisher Scientific	Cat#14-0037-82; RRID:AB_467057
anti-human CD3 (clone OKT3)	Thermo Fisher Scientific	Cat#16-0037-85; RRID: AB_468854
anti-human CD3 BUV395 (clone UCHT1)	BD	Cat#563546; RRID: AB_2744387
anti-human CD3 PE/Cy7 (clone UCHT1)	BioLegend	Cat#300419; RRID: AB_439780
anti-human CD38 BV650 (clone HIT2)	BD	Cat#740574; RRID: AB_2740275
anti-human CD4 BUV661 (clone SK3)	BD	Cat#612962; RRID: AB_2739452
anti-human CD4 BV510 (clone RPA-T4)	BD	Cat#740161; RRID: AB_2739914
anti-human CD45 BUV395 (clone HI30)	BD	Cat#563792; RRID: AB_2869519
anti-human CD45RA APC/Cy7 (clone HI100)	BioLegend	Cat#304128; RRID: AB_10708880

anti-human CD8 BUV805 (clone RPA-T8)	BD	Cat#749366; RRID: AB_2873737
anti-human CD8 BV510 (clone RPA-T8)	BD	Cat#563256; RRID: AB_2738101
anti-human CD8 PE (clone RPA-T8)	BioLegend	Cat#301008; RRID: AB_314126
anti-human CD84	BioLegend	Cat#CD84.1.21
anti-human CXCL13 APC (clone 53610)	Thermo Fisher Scientific	Cat#MA5-23629; RRID: AB_2610225
anti-human CXCL13 PE (clone 53610)	R&D Systems	Cat# IC801P
anti-human CXCR3 BV421 (clone 1C6/CXCR3)	BD	Cat#562558; RRID: AB_2737653
anti-human CXCR3 PE (clone 1C6/CXCR3)	BD	Cat#557185; RRID: AB_396596
anti-human CXCR5 PerCP/Cy5.5 (clone RF8B2)	BD	Cat#562781; RRID: AB_2313576
anti-human GPR56 PE (clone CG4)	BioLegend	Cat#358204; RRID: AB_2562084
anti-human GPR56 PE/Cy7 (clone CG4)	BioLegend	Cat#358206; RRID: AB_2562090
anti-human Granzyme A PE/Cy7 (clone CB9)	BioLegend	Cat#507222; RRID: AB_2721668
anti-human Granzyme B PE (clone GB11)	BD	Cat#561142; RRID: AB_10561690
anti-human Granzyme B Alexa Fluor 647 (clone GB11)	BioLegend	Cat#515406; RRID: AB_2566333
anti-human Perforin PE (clone B-D48)	BioLegend	Cat#353303; RRID: AB_10915476
Fc Receptor Binding Inhibitor Polyclonal Antibody, eBioscience	Thermo Fisher Scientific	Cat#14-9161-73
TotalSeq-C Custom Human Panel (204 antibodies)	BioLegend	Cat#900000115
TotalSeq-C Human TBNK Cocktail	BioLegend	Cat#399903
Chemicals, peptides, and recombinant proteins		
Ficoll-Paque PLUS	Cytiva	Cat# 17144003
ACK Lysing Buffer	Thermo Fisher Scientific	Cat#A1049201
LIVE/DEAD Fixable Aqua Dead Cell Stain Kit, for 405 nm excitation	Thermo Fisher Scientific	Cat#L34957
Recombinant Human IFN-alpha A (alpha 2a) Protein	R&D Systems	Cat#11100-1
Recombinant Human IFN-gamma Protein	R&D Systems	Cat#285-IF-100
eBioscience Protein Transport Inhibitor Cocktail (500X)	Thermo Fisher Scientific	Cat#00-4980-93
Dynabeads Human T-Activator CD3/CD28	Thermo Fisher Scientific	Cat#DB11131
RPMI 1640 Medium	Thermo Fisher Scientific	Cat#11875119
RPMI 1640 Medium, no phenol red	Thermo Fisher Scientific	Cat#11835030
Iscove's Modified Dulbecco's Medium (IMDM)	Thermo Fisher Scientific	Cat#12440053
Fetal Bovine Serum	BioWest	Cat#S1820
HyClone Fetal Bovine Serum	Thermo Fisher Scientific	Cat# Sh30910.03
Penicillin-Streptomycin-Glutamine (100X)	Thermo Fisher Scientific	Cat#10378016
Recombinant IL-21R His-tag protein	R&D Systems	Cat#9249-R2-100

5- or 6- (N-Succinimidyloxycarbonyl) fluorescein 3',6'-diacetate	Dojindo	Cat#C375
Cas9-GFP Protein	SIGMA-Aldrich	Cat#CAS9GFPPRO-250UG
High-Glucose RPMI 1640 medium	Wako	Cat#187-02705
SuperScript IV Reverse Transcriptase	Thermo Fisher Scientific	Cat#18090200
QuantiTect SYBR® Green PCR Kits	QIAGEN	Cat#204143
Cryostor CS10	STEMCELL Technology	Cat#7930
Advanced DMEM	Thermo Fisher Scientific	Cat#12491015
Liberase TL	Roche	Cat#447530
DNase I	Roche	Cat#10104159001
Critical commercial assays		
BD Human Th1/Th2/Th17 Phenotyping Kit	BD	Cat#560751
eBioscience Intracellular Fixation & Permeabilization Buffer Set	Thermo Fisher Scientific	Cat#88-8824-00
MagMAX-96 Total RNA Isolation Kit	Thermo Fisher Scientific	Cat#AM1830
SMART-seq v4 Ultra Low Input RNA Kit for Sequencing	Takara Bio	Cat#634894
Human IgG ELISA Quantitation Set	Bethyl Laboratories	Cat#E80-104
Human IgG ELISA BASIC kit (ALP)	Mabtech	Cat#3850-1AD-6
IgG (Total) Human Uncoated ELISA Kit with Plates	Thermo Fisher Scientific	Cat#88-50550-22
Human Granzyme A DuoSet ELISA	R&D Systems	Cat#DY2905-05
Human IFN-gamma DuoSet ELISA	R&D Systems	Cat#DY285B
Human CXCL13/BLC/BCA-1 DuoSet ELISA	R&D Systems	Cat#DY801
DuoSet ELISA Ancillary Reagent Kit 2	R&D Systems	Cat#DY008
CytoTox 96 Non-Radioactive Cytotoxicity Assay	Promega	Cat#G1780
Absolute Human Telomere Length Quantification qPCR Assay Kit	ScienCell	Cat#8918
QIAamp DNA Mini kit	QIAGEN	Cat#51304
SE Cell Line 4D-Nucleofecton X Kit S	Lonza	Cat#V4XC-1032
P2 Primary Cell 4D-Nucleofector X Kit S	Lonza	Cat#V4XP-2032
P3 Primary Cell 4D-Nucleofector X Kit S	Lonza	Cat#V4XP-3032
MojoSort Human CD4 T Cell Isolation Kit	BioLegend	Cat#480009
Chromium Next GEM Single Cell 5' Reagent Kits v2 Dual Index	10X Chromium	Cat#1000266
Deposited data		
RNA-seq	This study	NBDC: JGAS000626, JGAS000627, JGAS000648
ImmuNexUT (Immune Cell Gene Expression Atlas from the University of Tokyo)	Cell. 2021;184(11):3006-3021.e17.	https://www.immunexut.org/
Human reference genome NCBI, build 38, GRCh38	Genome Reference Consortium	https://www.ncbi.nlm.nih.gov/grc/human
Time-resolved Systems Immunology Reveals a Late Juncture Linked to Fatal COVID-19	Liu et al., Cell. 2021;184(7):1836-1857.e22.	https://cellxgene.cziscience.com/collections/ed9185e3-5b82-40c7-9824-b2141590c7f0.

The Immune Cell Landscape in Kidneys of Lupus Nephritis Patients	Arazi et al., Nat Immunol. 2019;20(7):902-914.	https://singlecell.broadinstitute.org/single_cell/study/SCP279/
Experimental models: Cell lines		
Jurkat cells	ECACC	Cat# 88042803
K562 cells	ATCC	Cat# CCL-243
Oligonucleotides		
Random Primers	Thermo Fisher Scientific	Cat#48190011
Software and algorithms		
cutadapt (v1.16)	Martin, EMBnetjournal. 2011; 17(1):10-12.	https://cutadapt.readthedocs.io/en/stable/index.html
STAR (v2.5.3)	Dobin et al., Bioinformatics. 2013; 29(1):15–21.	https://github.com/alexdobin/STAR
HTSeq (v0.11.2)	Anders et al., Bioinformatics. 2015;31(2):166-9.	https://github.com/htseq/htseq
R (v4.1.0)	Ihaka, R., and R. Gentleman. 1996. R: a language for data analysis and graphics. J. Comp. Graph. Stat. 5:299-314	https://www.r-project.org/
edgeR (v3.32.1 or v3.36.0, depending on the time of the analysis)	Robinson et al., Bioinformatics. 2009;26(1):139–40.	https://bioconductor.org/packages/release/bioc/html/edgeR.html
TCC (v1.34.0)	Sun et al., BMC Bioinformatics. 2013;14(1).	https://www.bioconductor.org/packages/release/bioc/html/TCC.html
MatchIt (v4.5.3)	Ho et al., J Stat Softw [Internet]. 2011;42(8):1–28.	https://github.com/kosukeimai/MatchIt
MiXCR (v3.0.11)	Bolotin et al., Nat Methods. 2015;12(5):380–1.	https://github.com/milaboratory/mixcr
immunarch (v0.5.5)	ImmunoMind Team, 2019	https://immunarch.com/
Seurat (v3.2.3)	Stuart et al,, Cell 2019; 177, 1888-1902.e1821.	https://satijalab.org/seurat/
Monocle3 (v1.0.0)	Trapnell et al., Nat. Biotechnol. 2014; 32, 381-386.	https://cole-trapnell-lab.github.io/monocle3/
Escape (v1.2.0)	Borcherding et al., Commun Biol. 2021; 4(1):122.	https://github.com/ncborcherding/escape
variancePartition (v1.20.0)	Hoffman et al., BMC Bioinformatics. 2016;17(1):17–22.	https://bioconductor.org/packages/release/bioc/html/variancePartition.html
clusterProfiler (v3.18.1)	Yu et al., Omi A J Integr Biol. 2012;16(5):284–7.	https://bioconductor.org/packages/release/bioc/html/clusterProfiler.html

FlowJo	FlowJo, LLC	https://www.flowjo.com/
10x Cell Ranger	10x Genomics	https://support.10xgenomics.com/single-cell-gene-expression/software/pipelines/latest/what-is-cell-ranger/
GraphPad Prism version 9	GraphPad Software	https://www.graphpad.com/
Other		
ZsGreen plasmid vector	TaKaRa	Cat#632478
DsRed plasmid vector	TaKaRa	Cat#632540

Table S8. Primer sequences used for qRT-PCR, Related to STAR Methods

Gene	Direction	Primer sequence
ZEB2 †	forward	TCTCGCCCGAGTGAAGCCTT
ZEB2 †	reverse	GGGAGAATTGCTTGATGGAGC
ZEB2 *	forward	AAATTCCAGGTCCCCATCCC
ZEB2 *	reverse	GTGGAGTTCTGCTATAGATGGTG
TBX21 †	forward	GATGCGCCAGGAAGTTTCAT
TBX21 †	reverse	GCACAATCATCTGGGTCACATT
TBX21 *	forward	TCCTACCCAACCAGTATCCTG
TBX21 *	reverse	TCATGCTGACTGCTCGAAAC
GZMA	forward	TTTCTGGCATCCTCTCTCTCA
GZMA	reverse	GGGTCATAGCATGGATAGGG
PRF1	forward	GTGGAGTGCC GCTTCTACAGTT
PRF1	reverse	TGCCGTAGTTGGAGATAAGCCT
CXCL13	forward	GAGGCAGATGGAACTTGAGC
CXCL13	reverse	CTGGGGATCTTCGAATGCTA

† used in the gene transfection experiments

* used in the gene knockout experiments using the CRISPR/Cas9 system in human ThA cells