JCI The Journal of Clinical Investigation

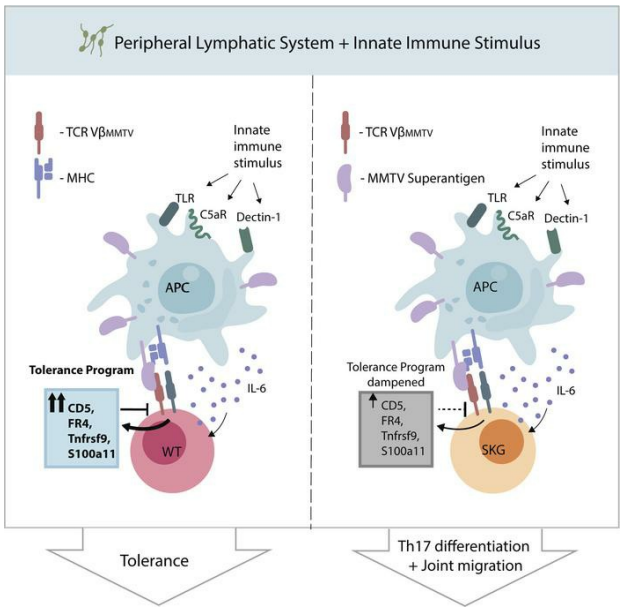
Endogenous antigens shape the transcriptome and TCR repertoire in an autoimmune arthritis model

Elizabeth E. McCarthy, … , Arthur Weiss, Judith F. Ashouri

J Clin Invest. 2024. https://doi.org/10.1172/JCI174647.

Research In-Press Preview Autoimmunity Immunology

Graphical abstract



Endogenous antigens shape the transcriptome and TCR repertoire in an autoimmune arthritis model

One Sentence Summary: Escaping negative selection, self-reactive 'naïve' $CD4^+$ T cells with impaired TCR signaling exhibit an autonomous peripheral tolerance defect, which, coupled with chronic antigen encounter, fosters the development of arthritogenic T cell responses.

Authors: Elizabeth E. McCarthy[1,2,3], Steven Yu[1], Noah Perlmutter[1], Yuka Nakao[1], Ryota Naito[1], Charles Lin[1,4], Vivienne Riekher[1,5], Joe DeRisi[6,7,8], Chun Jimmie Ye[1,3,,7,8]*, Arthur Weiss[1]*, Judith F. Ashouri[1]*

Affiliations:

[1]Rosalind Russell and Ephraim P. Engleman Rheumatology Research Center, Department of Medicine, UCSF; San Francisco, CA, USA.

[2]Department of Epidemiology and Biostatistics, UCSF; San Francisco, CA, USA.

[3]Institute for Human Genetics, UCSF; San Francisco, CA, USA.

[4]Stanford University; Palo Alto, CA, USA.

[5]Friedrich-Alexander University of Erlangen-Nuremburg, Bavaria, Germany.

[6]Department of Biochemistry and Biophysics, UCSF; San Francisco, CA, USA.

[7]Chan Zuckerberg Biohub; San Francisco, CA, USA.

[8]Bakar Computational Health Sciences Institute, UCSF; San Francisco, CA, USA.

*Co-corresponding authors. Email: Judith.Ashouri@ucsf.edu; Arthur.Weiss@ucsf.edu; Jimmie.Ye@ucsf.edu

Arthur Weiss, MD, PhD
513 Parnassus Ave.
Medical Sciences, S-1032C
San Francisco, CA 94143
415-476-1291

Jimmie Ye, PhD
513 Parnassus Ave.
HSW 1202E
San Francisco, CA 94143
415-476-6933

Judith Ashouri-Sinha, MD
513 Parnassus Ave.
Medical Sciences, S-1032A
San Francisco, CA 94143
415-476-7160

Abstract: The development of pathogenic autoreactive $CD4^+$ T cells, particularly in the context of impaired signaling, remains poorly understood. Unraveling how defective signaling pathways contribute to their activation and persistence is crucial for identifying new therapeutic targets. We profiled a highly arthritogenic subset of naïve $CD4^+$ T cells using bulk and single-cell RNA and TCR sequencing from SKG mice, which develop $CD4^+$ T cell mediated autoimmune arthritis driven by a hypomorphic mutation in Zap70—a key TCR signaling kinase. Despite impaired signaling, these cells exhibit heightened expression of T cell activation and cytokine signaling genes, but diminished expression of a subset of tolerogenic markers (*Izumo1r, Tnfrsf9, Cd5, S100a11*) compared to wild-type cells. The arthritogenic cells show an enrichment for TCR variable beta (Vβ) chains targeting superantigens from the endogenous mouse mammary tumor virus (MMTV) but exhibit diminished induction of tolerogenic markers following peripheral antigen encounter, contrasting with the robust induction of negative regulators seen in wild-type cells. In arthritic joints, cells expressing superantigen-reactive Vβs expand alongside detectable MMTV proviruses. Antiretroviral treatment and superantigen-reactive T cell depletion curtail SKG arthritis, suggesting that endogenous retroviruses disrupt peripheral tolerance and promote the activation and differentiation of autoreactive $CD4^+$ T cells into pathogenic effector cells.

INTRODUCTION

The connection between immunodeficiency and autoimmunity has long intrigued clinical immunologists. This dynamic is particularly evident in patients with primary immunodeficiency, where up to a quarter develop autoimmune conditions, such as autoimmune arthritis (1-3). Some autoimmune diseases are believed to arise from the activation of conventional $CD4^+$ T cells that recognize self-antigens (4).

In certain T cell-mediated autoimmune diseases like rheumatoid arthritis (RA), a paradox arises: CD4+ T cells can adopt a pathogenic effector state despite impaired TCR signaling (5-11). How these T cells evade tolerance to drive disease remains unclear. Self-tolerance is maintained through T cell-intrinsic mechanisms during thymic development, primarily through negative selection, and in the periphery via functional unresponsiveness or 'anergy.' Dysregulated T cell signaling is implicated in the pathogenesis of T cell-mediated autoimmune diseases, as seen in both humans and murine models. The SKG mouse model of autoimmune arthritis offers valuable insights into the mechanisms driving these complex interactions.

The SKG mouse, derived from the BALB/c strain, serves as a powerful tool for studying how impaired immunity leads to tolerance defects and autoimmune arthritis. A hypomorphic mutation in *Zap70*, a key tyrosine kinase in TCR signaling, disrupts thymocyte negative selection, allowing self-reactive $CD4^+$ T cells to escape into the periphery (12-15). When exposed to environmental triggers, these dormant arthritogenic $CD4^+$ T cells become activated, causing erosive inflammatory arthritis resembling RA (12, 16) and sharing features with spondyloarthritis (17). SKG $CD4^+$ T cells are required to cause arthritis (12), and even naïve $CD4^+$ ($CD4^+_{naive}$) T cells transferred into immunodeficient hosts trigger disease (13). However, the mechanisms driving their differentiation into pathogenic effector cells, despite impaired TCR signaling, remain unclear.

To investigate this, we used the SKGNur mouse, combining the SKG model with the Nur77-eGFP TCR signaling reporter (13). In this model, green fluorescent protein (GFP) is expressed under the control of *Nr4a1*, which encodes NUR77, a TCR signaling marker. Antigen, but not inflammation, triggers GFP expression, allowing us to track antigen-activated T cells (18, 19) during disease progression. Our previous studies revealed that high Nur77-eGFP expression marks self-reactive $CD4^+_{naive}$ T cells with heightened interleukin 6

(IL-6) sensitivity, making them more arthritogenic (**figure S1A**) (13). Chronic antigen exposure in peripheral SKG $CD4^+$ T cells emerges as a potential catalyst, downregulating suppressor of cytokine signaling 3 (SOCS3)—a key regulator of IL-6 signaling—an observation mirrored in patients with RA (13). We hypothesize that additional immune regulators may be dysregulated in SKG $CD4^+$ T cells, heightening their susceptibility to peripheral tolerance breakdown.

To test this, we analyzed the transcriptome and TCR repertoire of SKG $CD4^+_{naive}$ T cells using bulk and single cell RNA-sequencing. The SKGNur model allowed us to identify and capture arthritogenic (SKGNur GFP^{hi}) cells before disease onset, similar to the pre-clinical phase of autoimmune disease (20). This provided insight into early pathogenic mechanisms and potential targets for tolerance preservation. Within GFP^{hi} $CD4^+_{naive}$ T cells, a subset with elevated *Nr4a1* expression ($T.N4_{Nr4a1}$) exhibited gene programs linked to TCR signaling in response to antigen engagement. However, SKGNur GFP^{hi} $T.N4_{Nr4a1}$ cells exhibited impaired induction of a subset of tolerogenic genes compared to WT counterparts. TCR sequencing revealed a significant enrichment of variable beta (Vβ) chains in SKG T cells that recognize superantigens (Sag) from mouse mammary tumor virus (MMTV), an endogenous retrovirus (ERV) in BALB/c mice. Dysregulated TCR signaling was evident in these Sag-reactive cells, accompanied by a greater propensity for Th17 differentiation. Furthermore, these Sag-reactive $CD4^+$ T cells expanded within arthritic joints of SKG mice, potentially driving arthritis. Treatment with antiretrovirals or depletion of Sag-reactive T cells significantly delayed arthritis onset. These results underscore a breakdown in peripheral tolerance among self-reactive T cells that bypass negative selection. Coupled with chronic antigen stimulation, these circumstances set the stage for autoimmune disease. Our findings also reveal a unique role for ERV Sags in driving pathogenic T cell responses, contributing to disease development.

RESULTS

Arthritogenic SKG CD4$^{+}_{naïve}$ T cells display a gene signature of TCR activation: We previously identified arthritogenic CD4$^{+}_{naive}$ T cells prior to disease onset using Nur77-eGFP expression in SKG mice (13). To understand the transcriptional profile of these SKGNur GFP^{hi} cells, we performed bulk RNA-sequencing on naïve (CD62L^{hi} CD44^{lo}CD25^{-}) CD4^{+} T cells with the highest (GFP^{hi}) and lowest (GFP^{lo}) GFP expression from SKGNur and wild-type (WTNur) mice before arthritis onset (**Figure 1A**, **figure S1B**, **data file S1**). Principal component analysis (PCA) revealed that all four subgroups are transcriptionally distinct (**Figure 1B**). Hierarchical clustering of the 991 differentially expressed genes (DEGs) between groups identified six gene modules (**Figure 1C**, **data file S2**). Gene ontology analysis (21) highlighted functional differences between, and in some cases within, these modules (**figure S1C**). Among the 260 DEGs between WTNur GFP^{hi} and SKGNur GFP^{hi} cells (**figure S1, D** and **E**), SKGNur GFP^{hi} cells showed higher expression of cell cycle genes (*Cdca3, Cdk2nc*, *Mki67*, represented in module 1) and genes linked to cytokine signaling (*Socs1, Tnfsf14, Il2ra, H2-Aa, H2-Ab1* represented in module 6) (**Figure 1, C** and **D**, **figure S1E**, **data file S1**, **S3**). Interestingly, module 6 genes were highly expressed in SKGNur cells regardless of GFP expression, suggesting a distinct regulatory pathway in these cells (**Figure 1C**).

Despite their hypomorphic *Zap70* allele and impaired proximal TCR signaling (12, 13, 15), SKGNur GFP^{hi} cells induce both positive (*Egr1*, *Id3, Icos, Irf4, Tnfrsf9, Tnfrsf4, Myb*) and negative regulators (*Nr4a1, Nr4a3, Cd5, Folr4/Izumo1r, Tigit, Tox, Pdcd1, Lag3, Ctla4, Birc5, Nrp1*) of TCR signaling, primarily found in modules 1 and 2 (**Figure 1C**, **figure S1E, data file S2**). Paradoxically, SKGNur GFP^{hi} CD4$^{+}_{naïve}$ T cells show higher expression of activation and signaling pathways compared to WTNur GFP^{hi} cells (**Figure 1D**, **data file S3**). This likely reflects the chronic endogenous antigen encounter experienced by self-reactive SKG T cells.

SKG CD4$^{+}_{naïve}$ T cells that have most recently encountered antigen demonstrate an enhanced TCR signaling program: The long half-life of eGFP (22) compared to the more dynamic turnover of NUR77/*Nr4a1* protein and transcript (23-25) (**figure S2, A-C**) means GFP^{hi} cells likely consist of mixed populations of more and less recently stimulated cells. To explore the heterogeneity in these subsets, we performed single-cell

RNA and TCR sequencing (scRNA- and scTCR-seq) on GFP^{hi} and GFP^{lo} CD4$^{+}_{naïve}$ T cells from SKGNur and WTNur mice (**Figure 2A**). In our scRNA-seq dataset, we identify nine distinct clusters (**Figure 2, B** and **C**) which recapitulate our bulk RNA-seq gene signatures (**Figure 1**, **data file S4**) and demonstrate pronounced heterogeneity within the GFP^{hi} population (**figure S2, D** and **E**). GFP^{hi} CD4^{+} T cells from both SKGNur and WTNur mice were present across all nine clusters but were enriched in the T.N4$_{Nr4a1}$ cluster, which had the highest expression of *Nr4a1* and eGFP, compared to GFP^{lo} CD4^{+} T cells by a mean of > 4-fold (**Figure 2C**, **figure S2, F** and **G**, **data file S5**). GFP^{hi} T cells were also enriched in the T.N4$_{Izumo1r_Id2}$ and to a lesser extent in the Cyto$_{Nkg7}$ clusters (**figure S2, D-G**).

Given its specificity as a reporter of TCR signaling (22, 26, 27), the high expression of *Nr4a1* in the T.N4$_{Nr4a1}$ cluster signifies that these cells most recently encountered endogenous antigen(s) (28, 29). Indeed, T.N4$_{Nr4a1}$ cells overwhelmingly express genes associated with TCR signal transduction *(*including *Nr4a1, Nr4a3, Egr1-3, Tnfrsf9, Tnfrsf4, Ifr4*, *Cd5*, *Cd69*, **Figure 2D**, **data file S4-5**) which most closely mirror TCR signaling genes found in module 2 of our bulk RNA-seq analysis (**figure S2H**, **data file S2**). Similarly, we found that surface expression of OX-40 (*Tnfrsf4*), CD5, and CD69 was significantly higher in naïve CD4^{+} GFP^{hi} T cells compared to naïve CD4^{+} GFP^{lo} T cells in WTNur and SKGNur mice (**figure S2I**).

Several of the TCR response genes highly expressed in cluster T.N4$_{Nr4a1}$ have been identified as targets of the LAT-PLCγ-HDAC7 pathway (including *Nr4a1, Egr1-3, Irf4*) and correlate with strength of tonic signaling (28, 29). These TCR signaling signatures are more highly expressed in SKGNur GFP^{hi} CD4^{+} T cells in the T.N4$_{Nr4a1}$ cluster compared to those of WT (including genes *Nr4a1, Nr4a3, Relb, Nfkbia, Myb, Lag3*, **Figure 2, E** and **F**, **data file S3**, **S5**), in congruence with our bulk RNA-seq dataset. This suggests that despite their intrinsic signaling defect and dampened inducible signaling (**figure S2J** and reference (13)), the SKGNur GFP^{hi} cells in cluster T.N4$_{Nr4a1}$ have responded more strongly to or encountered more antigen than WTNur GFP^{hi} cells.

SKGNur GFP^{hi} T cells have impaired expression of a subset of tolerogenic genes: We next investigated additional T cell transcriptomic signatures to understand how SKGNur GFP^{hi} CD4$^{+}_{naïve}$ T cells may evade

tolerance and differentiate into pathogenic effector cells. We analyzed the expression of candidate genes linked to tolerance programs (30-36) in cells that recently encountered antigen. We found that GFP^{hi} CD4$^{+}_{naïve}$ T cells, compared to GFP^{lo} cells from both WTNur and SKGNur in cluster T.N4$_{Nr4a1}$, induce genes associated with tolerogenic programs (*Izumo1r, Pdcd1, Lag3, Tox)* and additional 'natural anergy' genes identified by ElTanbouly et al., including *Nfatc1, Hif1a, Icos* (**Figure 2, D** and **G**, **figure S3A**, **data file S5**). This likely indicates the activation of a negative regulatory program in CD4$^{+}_{naïve}$ T cells in response to persistent TCR signaling (13, 37, 38). This process is partially driven by NR4A family members, which have been shown to play inhibitory roles in peripheral T cells (24, 35, 39, 40).

While genes associated with tolerogenic programs and TCR signaling are broadly induced in SKGNur and WTNur GFP^{hi} cells, we found that several of these TCR negative regulators are less efficiently induced in SKGNur GFP^{hi} cells compared to WTNur GFP^{hi} cells in cluster T.N4$_{Nr4a1}$ (*Izumo1r, S100a11,Tnfrsf9, Cd5*) (**Figure 2, D, E**, and **G**, **figure S3A**). The lower expression of *Izumo1r, which* encodes Folate Receptor 4 (FR4), a specific marker of anergic cells, suggests that perhaps SKGNur GFP^{hi} CD4^{+} T cells may sub-optimally induce anergy and/or other tolerogenic programs. In our bulk RNA-seq analysis, we also saw a lower magnitude of *Izumo1r*—also known as *Folr4*—expression in SKGNur GFP^{hi} cells compared to WTNur GFP^{hi} cells (**data file S1**). Therefore, in addition to a known loss in central tolerance, SKG mice likely have an independent defect in mechanisms maintaining peripheral tolerance. This defect is likely derived from their impaired proximal TCR signaling capacity and may explain the reduced frequency of anergic peripheral CD4^{+} T cells we previously reported in SKGNur mice (13).

SKG's hyperresponsiveness to IL-6 is pre-programmed transcriptionally: IL-6 production in SKG mice is indispensable for SKG arthritis development (41, 42). Recognition of major histocompatibility complex/self-peptide complexes stimulate antigen-presenting cells (APCs) to secrete IL-6 (42). We previously found that SKGNur GFP^{hi} T cells were more responsive to IL-6 and more readily produced IL-17 in the most self-reactive T cells, in part due to lower levels of SOCS3 (suppressor of cytokine signaling 3) – a critical negative regulator of IL-6 (13) (**figure S1A**). In our current study, we found that genes associated with IL-6 signaling machinery

and the Th17 pathway were uniquely enriched in SKGNur GFP^{hi} T cells versus WTNur GFP^{hi} T cells (43) in the T.N4$_{Nr4a1}$ cluster (**figure S3B**).

SOCS3 is suppressed in CD4$^{+}_{naïve}$ T cells in response to antigen (44) and in patients with RA (13, 45). Its expression has a strong inverse correlation with murine arthritis severity (46-48) and is one of the genes most strongly suppressed in T.4N$_{Nr4a1}$ cells in our dataset. Therefore, we examined the expression of additional SOCS family members to determine whether this suppression was unique to *Socs3* (**figure S3C**). Of the SOCS family members, *Socs3* was specifically reduced in GFP^{hi} cells versus GFP^{lo} cells within the T.N4$_{Nr4a1}$ cluster (**Figure 2D**, **data file S5**). Moreover, we found a striking inverse correlation between the expression of *Nr4a1* and *Socs3* (**figure S3D**) validating our previous results. This highlights the interdependence between signaling via the TCR and heightened sensitivity to cytokines such as IL-6 (13).

T.N4$_{Nr4a1}$ cells segregate into two distinct TCR signaling subclusters: To more broadly examine genes correlated with *Nr4a1* within our entire dataset, we performed co-expression analysis of highly variable genes (HVGs) in all cells. We identified three gene modules of HVGs that positively correlated with *Nr4a1* (**Figure 3A**). Genes from two of these modules, *Egr* family members (immediate early gene transcription factors) and *Tnfrsf9* (4-1BB – the TCR inducible co-stimulatory receptor), identified distinct subclusters of cells within the T.N4$_{Nr4a1}$ cluster (**Figure 3B**, **figure S4A**). Cell-cycle stages did not fully account for the division between *Egr* family members and *Tnfrsf9* expression (**figure S4B**). We found *Egr2*$^{+}$ cells expressed genes induced early after TCR stimulation (including *Egr1, Egr2, Cd69, Ier2, Egr3, Nfkbid, Junb, Fos, Myc, Cd40lg*), whereas the *Tnfrsf9*$^{+}$ cells expressed genes enriched in pathways induced in response to prolonged TCR signaling (such as *Pou2f2, Myb, Tnfrsf4, Lag3*) (**Figure 3, C** and **D**, **figure S4C**, **data file S3**, **S6**). Moreover, the *Tnfrsf9*$^{+}$ cells more robustly induce a previously identified TCR activation gene module compared to *Egr2*$^{+}$ cells (36) (**figure S4, D** and **E**). These findings suggest T.N4$_{Nr4a1}$ cells segregate into subclusters driven by their TCR signaling kinetics (early versus prolonged stimulation). A subset of tolerogenic genes identified in **Figure 2G** are also induced at lower magnitudes in SKGNur GFP^{hi} *Tnfrsf9*$^{+}$ T cells compared to the corresponding WT subset (**Figure 3E**). This suggests a defect in the expression of tolerogenic gene programs in SKGNur GFP^{hi} CD4^{+} T

cells responding to prolonged TCR stimulation rather than SKGNur simply having a decreased frequency of those cells.

Cell states and trajectories of T.4N$_{Nr4a1}$ cells have a distinct distribution in SKGNur GFP^{hi} subset: We explored if the early versus prolonged TCR signaling states in the T.N4$_{Nr4a1}$ cluster represented endpoints of a trajectory. We identified a continuum of cell states in the T.N4$_{Nr4a1}$ cluster ordered by latent time using RNA velocity (49) (**Figure 4A**). The expression of the *Egr* family peaks in earlier latent time cells, while the expression of *Tnfrsf9* and associated genes peaks in later latent time cells (**Figure 4, B** and **C**).

We used a Gaussian mixture model (50) to deconvolute the distribution of all cells across latent time into four cell states labelled “Stage 1” to “Stage 4” from earlier to later latent time (**Figure 4D**, **figure S4F**). The number of underlying distributions, or stages, were determined using the Bayesian Information Criterion (BIC) and Akaike Information Criterion (AIC) as clustering validity metrics, and the resulting stage labels had an average silhouette score of 0.554 indicating good quality clusters (**figure S4G)**. The RNA velocity vector field (**figure S4H**) and trajectory inference analysis (51) supported a trajectory from Stage 1 to Stage 4 (**Figure 4E**). The expression of *Egr2* and *Nr4a1* peak within cells from Stage 1 while the expression of *Tnfrsf9* peaks within cells from Stage 4 (**Figure 4F**, **data file S7**). The genes overexpressed in Stage 1 and Stage 4 cells are enriched for early or prolonged TCR stimulation pathways, respectively (**figure S4I**). Thus, these cell states seem to be the endpoints of a trajectory of cell states from early to prolonged TCR stimulation.

Cells from SKGNur GFP^{hi} and WTNur GFP^{hi} groups had significantly different distributions across latent time with a higher density at earlier latent time for the SKGNur GFP^{hi} cells, which also had an increased odds of being in Stage 1 versus Stage 4 compared to WTNur GFP^{hi} cells (OR = 1.25, p = 0.02). This difference was not observed between the GFP^{lo} subgroups (**Figure 4, G-H**).

We hypothesized this imbalance may reflect either slower progression of SKG cells through the stages or higher input, due to higher proliferation, into Stage 1 in the SKG pool. There was no significant difference in cell cycle distribution between SKGNur GFP^{hi} and WTNur GFP^{hi} cells within Stage 1 suggesting the SKGNur GFP^{hi} cells in Stage 1 do not proliferate more. While this result favors our slower progression hypothesis, the

two hypotheses are not mutually exclusive. Slower progression of the SKGNur GFP^{hi} cells would suggest that SKG CD4^{+} T cells have a defect in peripheral tolerance induction—a program which is activated as the cells progress through the stages—and could explain the reduced frequency of anergic cells we previously observed in SKG mice (13).

SKGNur GFP^{hi} CD4$^{+}_{naïve}$ T cells demonstrate a biased TCR beta variable (TRBV) gene repertoire: We previously demonstrated that SKGNur GFP^{hi} cells exhibit heightened self-reactivity and can proliferate in response to unknown endogenous antigens (13). This led us to investigate how their TCR repertoire influences their activation in the periphery. With scTCR-seq (**Figure 2A**) we detected paired TCR α (TRA) and TCR β (TRB) genes in 86% of cells (**figure S5A**). We did not find oligoclonal expansion in the naïve T cells (**data file S8**). Instead, we found that SKGNur GFP^{hi} T cells demonstrate a biased TCR β variable (*TRBV*) gene usage, but not TCR variable α (*TRAV*) usage (**Figure 5, A-C**, **figure S5B**). SKGNur GFP^{hi} CD4$^{+}_{naïve}$ T cells had significantly higher usage of *TRBV26* (corresponding to TCR variable beta 3, Vβ3, protein), *TRBV12-1* (Vβ5), *TRBV15* (Vβ12), *TRBV16* (Vβ11), *TRBV3*, and *TRBV29* (Vβ7*)* compared to paired SKGNur GFP^{lo} cells. Each of these *TRBV* genes also had a higher mean frequency in SKGNur GFP^{hi} cells compared to WTNur GFP^{hi} cells (**Figure 5, A**, **C,** and **D**).

Polyclonal Vβ expansion occurs in the presence of superantigen (Sag) in both humans and mice (52, 53). The *TRBV* genes enriched in SKGNur GFP^{hi} T cells mark Vβs that recognize endogenous retroviral (ERV) Sag from mouse mammary tumor virus (MMTV) (**Table S1**) (54-56). We confirmed that our SKG colony harbors all three endogenous MMTV proviruses (*Mtv-6*, *Mtv-8*, *Mtv-9*) known to be present in BALB/c mice (54, 55, 57, 58) (**figure S5C**). Exogenous MMTV infection can stimulate cell proliferation and facilitate infection by increasing the number of cell targets, but Sag expression from endogenous *Mtv* leads to clonal T-cell deletion in the thymus and resistance to infection owing to the absence of these Sag-reactive Vβ expressing T cells (59). However, due to impairment in SKG TCR signaling, thymic clonal T-cell deletion in response to endogenous *Mtv* Sag is incomplete (15) allowing for partial escape of these Sag-reactive T cells into the periphery. In contrast to the *TRBV* genes uniquely enriched in SKGNur GFP^{hi} cells, *TRBV* genes for Vβs that

do not recognize *Mtv* Sags in BALB/c (such as *TRBV19*/Vβ6, *TRBV13-2*/Vβ8, *TRBV31*/Vβ14) are not enriched in SKGNur GFP^{hi} T cells (**Figure 5C**, **figure S5D**). These results show that negative selection is defective in SKG mice and that encounter with endogenous *Mtv* Sag in the periphery further biases the *TRBV* repertoire in SKGNur GFP^{hi} CD4^{+} T cells.

SKGNur GFP^{hi} cells within the T.N4$_{Nr4a1}$ cluster also show enrichment of several Sag-reactive TCRs (*TRBV15* (Vβ12), *TRBV16* (Vβ11), *TRBV29* (Vβ7), **figure S6, A-D**). This enrichment, occurring without TCRα restriction, suggests activation and expansion due to Sag exposure, which contrasts with the typical deletion of Sag-reactive T cells following peripheral *Mtv* Sag encounter (60).

SKGNur GFP^{hi} CD4^{+} T cells are enriched for Vβs driven by MMTV Sag(s): To validate our scTCR-seq results, we assessed TCR Vβ protein levels in SKGNur and WTNur peripheral CD4^{+} T cells before arthritis onset using antibodies against selected Vβs (gating strategy in **figure S7A**). We found that Vβ protein levels mirrored the transcript abundances from our scTCR-seq dataset. Vβ3, Vβ5, Vβ11 (corresponding to *TRBV26*, *-12*, *-16,* respectively) were significantly enriched in SKGNur GFP^{hi} naïve CD4^{+} T cells from lymph nodes (LN) (**Figure 5, E** and **F**) and spleen, whereas non-MMTV Sag targets like Vβ6, Vβ8, Vβ14 (corresponding to *TRBV19*, *-13*, *-31,* respectively) were not enriched in SKGNur GFP^{hi} cells (**figure S7, B** and **C**).

Vβ enrichment in SKGNur GFP^{hi} T cells subset may be driven by Sag encounter in the periphery and even the joints. Indeed, we detected BALB/c specific *Mtv* proviruses in SKG joints (**figure S8A**). Therefore, it is feasible that intra-articular *Mtv* Sag expression could engage and enrich for SKG T cells uniquely expressing these Sag-reactive Vβs (Vβ3, Vβ5, and Vβ11) during arthritis. To investigate this possibility, we induced moderate to severe inflammatory arthritis in SKG mice (**figure S8B**) with zymosan and examined Vβ usage in CD4^{+} T cells harvested from regional joint draining LN (dLN) and arthritic joints compared to CD4^{+} T cells from joint dLN in non-arthritic mice (treated with PBS). We found an increased frequency of MMTV Sag-reactive Vβ3, Vβ5, and Vβ11 in the arthritic joints compared to the periphery (**Figure 6, A** and **B**), but not of the non-Sag-reactive Vβs (control Vβs) (**figure S8C**). Zymosan had limited effect on Sag-reactive and non-Sag-

reactive Vβ frequencies in peripheral naïve or memory $CD4^+$ T cells (**Figure 6B**, **figure S8C**). The Sag-reactive $CD4^+$ T cells had significantly higher Nur77-eGFP mean fluorescence intensity (MFI) compared to $CD4^+$ T cells with control Vβs in SKG arthritic joints (**Figure 6C**), as well as in peripheral naïve $CD4^+$ T cells (**figure S8D**). Elevated Nur77 levels in Sag-reactive T cells remained unaffected by zymosan exposure (**figure S8D**), consistent with our previous findings in human $CD4^+$ T cells (22). Furthermore, we found a significantly higher frequency of Sag-reactive Vβ3, Vβ5, and Vβ11 in SKGNur GFP^{hi} T cells infiltrating the arthritic joints compared to intra-articular GFP^{lo} cells or to GFP^{hi} cells from dLN (**Figure 6, D** and **E**, **figure S8E**). This further enrichment suggests Sag-reactive $CD4^+$ T cells expand after encounter with intra-articular antigen in SKG inflamed joints. This enrichment in the joint was not observed in SKGNur GFP^{hi} T cells expressing control Vβs (**figure S8, F** and **G**).

Sag-reactive SKG CD4 T cells exhibit impaired tolerance induction, dampened signaling, and a propensity to differentiate into Th17 cells: We next compared the transcriptional profiles of *Mtv* Sag-reactive TRBVs with non-Sag-reactive TRBVs, focusing on the impact of peripheral Sag encounter. WT GFP^{hi} cells with Sag-reactive TRBVs (WT GFP^{hi} $TRBV_{enriched}$) compared to WT GFP^{hi} cells with non-Sag-reactive TRBVs (WT GFP^{hi} $TRBV_{non\text{-}enriched}$) have significantly higher expression of genes (*Tnfrsf9*, *Cd200*, *Ikzf2*, *Myb*) which are part of the *Tnfrsf9* subcluster/chronic activation signature described in **Figure 3**. In contrast, we found that SKG GFP^{hi} cells with *Mtv* Sag-reactive TRBVs ($TRBV_{enriched}$) compared to SKG GFP^{hi} cells with non-Sag-reactive TRBVs (SKG GFP^{hi} $TRBV_{non\text{-}enriched}$) had either no significant difference in expression or diminished induction, compared to WT, of TCR response genes associated with tolerance and negative regulation of TCR signaling (*Izumo1r, Cd5, Tox, Cd200, Tnfrsf9*) (**Figure 7, A** and **B**, **data file S9**). Additionally, we found that WT GFP^{hi} $TRBV_{enriched}$ cells from cluster $T.4N_{Nr4a1}$ were significantly more likely than the WT GFP^{hi} $TRBV_{non\text{-}enriched}$ cells to be part of the *Tnfrsf9* subcluster rather than the *Egr2* subcluster (OR = 1.5, p =0.04) while there was no significant difference in distribution for those cells from SKG GFP^{hi} mice (**figure S9A**). We then compared the transcriptional programs of WT and SKG GFP^{hi} $TRBV_{enriched}$ cells within the $T.4N_{Nr4a1}$ cluster, indicative of recent antigen encounter. WT GFP^{hi} $TRBV_{enriched}$ cells expressed higher

levels of negative regulators of TCR signaling such as *Izumo1r, Tnfrsf9,* and *Cd5* compared to their SKG counterparts, which instead showed higher expression of *Foxo1* and *Il6r* (**Figure 7C, data file S9**). *Mtv* Sags typically induce anergy in Sag-reactive T cells that escape thymic and peripheral deletion (61, 62). In WT mice, most of these cells are eliminated, and those that persist effectively induce a transcriptional tolerance program (**Figure 7, A** and **C**). However, in SKG mice, these cells not only expand and persist (**Figure 6**) but also fail to activate a robust tolerance program despite ongoing peripheral Sag antigen encounter (**Figure 7, B** and **C, data file S9**).

We next examined the surface expression of CD5, FR4, LAG3, CD73 and additional negative regulators of TCR signaling induced after antigen encounter. To capture a larger number of cells, we analyzed these markers in the total Sag-reactive population of naïve $CD4^+$ cells from SKGNur and WTNur mice. Thus, we are no longer focusing on the GFP^{hi} subsets from our transcriptional analysis. Yet, we still found that Sag-reactive (Sag $V\beta^+$) naïve CD4 T cells in both WT and SKG mice expressed higher levels of several inhibitory surface markers compared to non-Sag-reactive (Sag $V\beta^-$) T cells (**Figure 7D**, **figure S9B**). This likely represents compensatory tolerance mechanism to constrain self-reactive T cells, similar to published data on Nur77-GFP^{hi} T cells (13, 63). In results that support our transcriptional analysis, we saw significantly lower surface expression of CD5 and a trend towards lower levels of FR4/*Izumo1r* in SKG Sag-reactive T cells compared to WT (**Figure 7D**, **figure S9B**). Moreover, WT Sag-reactive T cells exhibit a reduced response to TCR stimulation compared to WT non-Sag-reactive T cells, likely due to chronic antigen exposure (**Figure 7E**). This dampening effect is less pronounced in SKG T cells, possibly due to not only their impaired proximal signaling defect, but also impaired tolerance mechanisms, which results in minimal differences in signaling between Sag-reactive and non-Sag-reactive SKG T cells (**Figure 7E**). We then assessed the capacity of Sag-reactive T cells from SKG mice to differentiate into IL-17-producing CD4 T cells under both pathogenic and non-pathogenic Th17 conditions. Th17 cells can be categorized as either pathogenic, associated with inflammation and autoimmune disease, or non-pathogenic, associated with tissue-homeostasis based on the cytokines present in their microenvironment (64, 65). Pathogenic Th17 cells are driven by signals from TGF-β3 and IL-6, or a combination of IL-6, IL-23, and IL-1β, while non-pathogenic Th17 cells are induced by TGF-β1

and IL-6 (65). We observed a higher frequency of IL-17$^+$ CD4$^+$ T cells among SKG Sag-reactive (Sag Vβ$^+$) cells compared to non-Sag-reactive (Sag Vβ$^-$) cells in both pathogenic and non-pathogenic conditions (**Figure 7F**).

Antiretroviral therapy ameliorates SKG arthritis: We tested whether inhibition of retroviral elements could curtail SKG arthritis and impede Sag-reactive T cell activation and expansion. Mice were treated with Truvada, a combination of antiretroviral reverse transcriptase inhibitors emtricitabine and tenofovir, or vehicle control prior to and during arthritis development (**Figure 8A**). MMTV reverse transcriptase is sensitive to these compounds which prematurely terminate nascent cDNA synthesis during reverse transcription and have been reported to decrease viral protein expression (66, 67) and improve mouse models of inflammatory colitis and autoimmune biliary disease (66, 68). Truvada significantly reduced arthritis severity and delayed disease onset in SKG mice (**Figure 8, B-D**). These data support the idea that ERVs may contribute to arthritis activity in SKG mice. We cannot exclude the possibility that antiretrovirals are inhibiting recombined live MMTV as a result of 'endogenous resurrection'—a phenomenon described in immune compromised mice (69). However, the absence of mammary tumors in aged SKG mice from our colony (**figure S9, C** and **D**) suggests that this is unlikely (70).

Sag-reactive SKG CD4 T cells are arthritogenic: We tested whether Sag-reactive CD4 T cells, which comprise ~4% of the SKG CD4 T cell population, initiate or augment disease. We adoptively transferred SKG CD4$^+$CD25$^-$ T cells (Vβ Total group) or SKG CD4$^+$ T cells depleted of most MMTV Sag-reactive Vβ specific cells using available Vβ antibodies (Vβ depleted) into SCID recipients (sorting strategy in **figure S9E**). Recipient mice that received SKG CD4 T cells depleted of most Sag-reactive Vβs exhibited a significant delay in arthritis onset and a trend for less severe disease (**Figure 8, E-H**, **figure S9, F** and **G**). Furthermore, we observed > 5-fold expansion of Sag-reactive Vβs in recipient mice that received Vβ Total cells (**figure S9H**). Together, these results suggest that a portion of the arthritis pathogenicity is contained within the Sag-reactive T cells.

DISCUSSION

This study explores the gene expression and TCR repertoire in arthritogenic SKGNur GFP^{hi} CD4$^{+}_{naïve}$ T cells before arthritis onset. Our sequencing analyses revealed that, after antigen exposure and before arthritis onset, these cells exhibit 1) heightened expression of T cell activation genes (*Nr4a1, Egr1-3, Irf4*) despite impaired proximal TCR signaling capacity, 2) diminished expression of key tolerogenic markers compared to wild-type (WT) cells.

Our findings provide evidence for a peripheral tolerance breach in arthritogenic SKG CD4^{+} T cells, independent of their central tolerance failure. Normally, TCR activation induces negative regulators that maintain peripheral tolerance and prevent immunopathology (71-74). However, in arthritogenic SKGNur GFP^{hi} CD4^{+} T cells, chronic exposure to endogenous antigens leads to incomplete induction of tolerance genes following TCR signaling, unlike the more complete induction of negative regulators observed in WT cells. This compromised signaling in SKG mice fails to fully establish a protective anergy state upon full antigen encounter as shown in **Figures 2** and **7**. Additionally, RNA velocity analysis adds a temporal perspective and reveals that these arthritogenic SKG cells are predominantly found in early TCR signaling states (Stage 1) and fail to advance to stages associated with extended and more effective TCR signaling (Stage 4).

Our study reveals that arthritogenic SKG T cells have an enrichment of Sag-reactive T cells. These cells, which escape negative selection in the thymus and evade peripheral deletion by endogenous *Mtv* Sag encounter (60), show a significant TCR Vβ bias in the peripheral naïve SKG repertoire. This bias suggests substantial peripheral Sag engagement. Through tracking these Sag-reactive T cells by their Vβs, we observe that these cells in WT mice exhibit increased expression of tolerance genes and inhibitory receptors, indicating substantial peripheral Sag engagement. Notably, WT Sag-reactive T cells respond more poorly than non-Sag reactive T cells to ex vivo restimulation. In contrast, SKG Sag-reactive T cells exhibit reduced expression of inhibitory genes and a subset of inhibitory receptors (on the protein level) and are more prone to differentiate into IL-17-producing cells.

The peripheral Vβ bias extends to SKG arthritic joints, indicating expansion due to intra-articular *Mtv* Sag encounter. While alternative joint-specific endogenous antigen responses cannot be ruled out, our data strongly suggest that the expansion of Sag-reactive T cells is closely associated with intra-articular *Mtv* Sag interactions, as evidenced by significantly higher levels of Nur77-eGFP in Sag-reactive T cells infiltrating the joint. The potential role of *Mtv* Sag in arthritis development is underscored by our findings that ERV reverse transcriptase inhibitors (such as Truvada) significantly mitigate SKG arthritis, likely by disrupting the impact of ERVs on innate and adaptive immune responses (67, 75, 76) and possibly influencing Sag-reactive T cell activation. Furthermore, depleting Sag-reactive T cells curtails SKG arthritis, highlighting the pivotal role of endogenous Sag in disease progression. Our results also emphasize the crucial role of chronic antigen encounter in breaking peripheral tolerance in T cells with impaired signaling, leading to their activation and differentiation into pathogenic effector cells.

Our proposed model (**Figure 9**) synthesizes our current and previous findings (12, 13, 15, 77) to explain the complex interplay between compromised TCR signaling, inefficient negative selection, chronic antigen encounter, and altered peripheral tolerance in autoimmunity. This interplay leads to a breakdown in tolerance, characterized by reduced induction of negative immune regulators and fewer anergic cells (13). The compromised TCR signaling threshold in Tregs, coupled with changes in their repertoire, exacerbates this loss of tolerance (14). Molecules like IL-6 may act as costimulatory signals, enhancing T cell survival and lowering their threshold for activation and differentiation. The biased self-reactive TCR Vβ repertoire in SKGNur GFP^{hi} CD4^{+} T cells, along with their activated state, primes these cells to respond to innate immune stimuli, potentially initiating or propagating disease (**Figure 9**). Our findings show that some of the pathogenicity is contained within Sag-reactive T cells and that antiretroviral therapy significantly reduces SKG arthritis development. Future studies will directly investigate the role of Sag-reactive Vβs, *Mtv* Sags, and ERVs in SKG arthritis, and their relevance to human autoimmune disorders (76, 78-81), including RA (78-80, 82, 83) and other autoimmune arthritides (84-88), where specific Vβ-expressing T cells expand and persist in the synovial microenvironment.

MATERIALS AND METHODS

Mice. BALB/c and C57BL/6J mice were purchased from Jackson laboratory, and BALB/cNur77-eGFP and SKGNur77-eGFP mice were bred in our facility (University of California, San Francisco) as previously described (13). All mice were housed and bred in specific pathogen-free conditions in the Animal Barrier Facility at UCSF according to the University Animal Care Committee and NIH guidelines.

Sex as a biological variable. Female mice were used in sequencing studies and arthritis experiments due to higher disease penetrance and severity (12). Both sexes were used in signaling studies, sex was not controlled as a biological variable in those experiments.

Murine synovial tissue preparation. See supplementary methods.

Antibodies and reagents. See supplementary methods.

Surface and intracellular staining. See supplementary methods.

Nuclei isolation and NFAT nuclear staining. See supplementary methods. The nuclei isolation protocol was adapted from supplementary methods from reference (89) and methods from (90).

In vivo treatments. Power analyses was performed based on preliminary data to calculate number of mice needed in each group to reach a power of 0.8 and detect a 50% difference between groups with standard deviation of 30% and type I error = 0.05. Adoptive transfer experiments were performed as described (12, 13). Negatively selected $CD4^+$ T cells from 8–12-week-old female SKG mice were sorted on $CD4^+CD25^-$ markers ± depleted of the following Sag-reactive Vβs: Vβ3, Vβ5, Vβ11, Vβ12 (using Vβ antibodies conjugated to PE) and 4.5e5 cells were adoptively transferred by tail vein injection into 8-week-old female SCID recipients. To reduce potential for cage effect, female mice were randomized prior to study start to different cages and groups controlling for litter, cage, and age. Bedding was also mixed between all cages to reduce effect of potential baseline differences in the microbiome 1 week prior to study start date. Mice were housed 2-3 mice/ cage to minimize cage effect after study induction. Scorer was blinded to treatment condition in adoptive transfer studies.

Th17 differentiation. See supplementary methods.

Flow cytometry and cell sorting. Cells were stained with antibodies of the indicated specificities and analyzed on a BD LSR Fortessa and BD LSR Fortessa DUAL (cytometer 2, figure S8B) flow cytometers (BD Biosciences). Flow cytometry plots and analyses were performed using FlowJo v.10.8.0-v.10.10 (Tree Star). Cells were sorted to >95% purity using a MoFlo XDP (Beckman Coulter).

PCR and RT-PCR. See supplementary methods.

Bulk RNA sequencing. Negatively selected $CD4^+$ T cells from the lymph node were sorted for $CD62L^{hi}CD44^{lo}CD25^-$ and the 10% highest (GFP^{hi}) or lowest (GFP^{lo}) expressing T cells. Cells were washed, pelleted and immediately flash frozen using dry ice in ethyl alcohol. Samples were processed for bulk RNA-sequencing by Q2 solutions using the TruSeq Stranded mRNA kit (Illumina: RS-122-2103) for library preparation. The resulting libraries pool into three batches and sequenced on an Illumina HiSeq 2500 sequencer over three lanes.

Alignment and initial processing of bulk RNA sequencing data. The raw fastq files were clipped and filtered using fastq-mcf v.1.04.636 to remove low quality reads and bases, homopolymers, and adapter sequences. The filtered reads were aligned using the STAR v.2.4 (91) with the default settings to the mm10 transcriptome and the resulting bam files were converted to count matrices for each sample with RSEM v.1.2.14. Genes with less than 10 counts across all the samples were filtered out. Raw counts were normalized and transformed by the variance stabilizing transformation (VST) function from DESeq2 v.1.22.2 (92).

PCA analysis. The VST normalized features were used for principal component analysis with the function plotPCA from DESeq2.

Bulk RNA sequencing differential expression. Differential gene expression for the bulk RNA sequencing samples was performed with the raw counts from the filtered gene list for the indicated samples as the inputs. The analysis was run using a negative binomial model with multiple testing correction with Benjamini-Hochberg

implemented via the DESeq function which includes an internal normalization from DESeq2. For differential gene expression between samples within the same genotype, mouse identity was included as a covariate.

Functional enrichment analysis. See supplementary methods.

Gene set enrichment analysis. See supplementary methods.

Single-cell RNA and TCR sequencing. Negatively selected $CD4^+$ T cells from the lymph node and spleen were sorted for $CD62L^{hi}CD44^{lo}CD25^-$ and the 10% highest (GFP^{hi}) or lowest (GFP^{lo}) expressing T cells. Droplet-based paired single-cell RNA and TCR sequencing was performed using the 10x single-cell 5'+V(D)J v.1 kit per manufacturer's instructions. The resulting cDNA libraries were sequenced on four lanes of an Illumina Novaseq 6000 sequencer to yield gene expression (GEX) and T cell receptor (TCR) fastqs.

Alignment and initial processing of single cell sequencing data. The raw fastq files were aligned using CellRanger v3.0.1 and 3.0.2 software with the default settings to the mm10 transcriptome with the addition of the sequence for the eGFP transcript and the vdj GRCm38 v 3.1.0 reference for the GEX and TCR fastqs, respectively.

eGFP transcript sequence. See supplementary methods.

Cell type classification and clustering. See supplementary methods.

Single-cell differential expression analysis. Single-cell differential expression was performed using the log-normalized gene counts with the rank_genes_groups function from scanpy with the Wilcoxon rank-sum method and multiple testing correction with Benjamini-Hochberg. Additionally, the adjusted p values that were equal to 0 were updated to the minimum representable positive normalized float (2.2250738585072014e-308).

Cell cycle phase assignment and module scoring. See Supplementary methods.

RNA velocity analysis. See supplementary methods.

TCR analysis. Cells with <=2 TRA chains and <=1 TRB chains were used in the TCR clonotype analyses (52). Cells with two TRA chains were removed for the TRBV and TRAV analyses since the highest frequency for any dual TRA was 0.09% in any one sample (~1 cell). This removed 10,598 cells or 13.6% of all cells which is consistent with the expected dual TRA frequency. TRBV and TRAV genes which were not present in at least two mice from the same subgroup (SKGNur GFP^{hi}, WTNur GFP^{hi}, SKGNur GFP^{lo}, and WTNur GFP^{lo}) were removed from the downstream TRBV and TRAV analyses.

Statistics. Flow cytometry data were analyzed by comparison of means using paired or unpaired 2-tailed Student's *t* tests using Prism v.9.2.0 or v.9.3.1 for Mac (GraphPad Software). Unpaired t-test with Welch's correction was used to calculate differences in arthritis scores and log-rank Mantel-Cox test used to calculate differences in Kaplan Meier Survival.

Significant differences in the TRBV frequencies between subgroups was determined by exact permutation test for unpaired and paired samples (for N > 5 paired samples) (93) or paired t-test with Benjamini-Hochberg correction (for N <= 5 paired samples) using scipy v.1.4.1 and statsmodels v.0.11.1.

The latent time distributions from different subgroups were compared using the Kolmogorov-Smirnov test. The cell cycle distributions between subgroups within stage 1 were compared using Pearson's chi-squared test.

Significant difference in GFP mean fluorescence intensity (MFI) for cells assigned TRBVs in the enriched or not-enriched groups was determined with a linear mixed effect model eGFP MFI ~ TRBV group (enriched or not-enriched) + mouse id (for paired data) with a random intercept for each TRBV protein followed by Benjamini-Hochberg correction. Significant difference in MFI for surface protein markers between groups for Sag-reactive and Sag non-reactive cells from WT GFP^{hi} and SKG GFP^{hi} were determined by linear mixed effect model MFI ~ subgroup for unpaired samples (e.g., WT Sag-reactive versus SKG Sag-reactive) and MFI ~ sag type + mouse for paired samples (e.g., WT Sag-reactive versus WT Sag non-reactive) with a random intercept for the flow cytometer machine, and p values were adjusted using Benjamini-Hochberg correction for multiple testing. Samples were collected on different cytometers due to cytometer 1 malfunction. Data in all

figures represent mean ± SEM unless otherwise indicated. Differences were considered significant at $P < 0.05$: *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$, and ****$P < 0.0001$.

Odds ratio was calculated using a conditional maximum likelihood (CML) estimator with scipy v.1.6.1.

Study approval. All animal experiments were approved by the UCSF Institutional Animal Care and Use Committee (IRB AN192722).

Data and Materials Availability. Bulk RNA sequencing data and single-cell RNA and TCR sequencing data discussed in this publication have been deposited in NCBI's Gene Expression Omnibus and are accessible through GEO Series accession number GSE185577 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE185577). All other data are available in the main text or as supplementary data. Values for all data points in the graphs are reported in the Supporting Data Values file. Code for analysis is available at: https://github.com/yelabucsf/SKG_rheum

Supplementary Materials

Supplementary Methods

Figure S1 to S9

Table S1

Data files S1 to S9

Acknowledgments: We thank Lewis Lanier and Julie Zikherman for valuable comments that improved the manuscript. We thank Z. Wang for cell sorting, A. Roque for animal husbandry, Chan Zuckerberg Biohub for single-cell sequencing, and SciStories, LLC for their assistance in creating **Figure 9**.

Funding:

NIH grants K08AR072144 (JFA), R01AI165706 (JFA), T32GM007618 (EM), F30CA257291 (EM), R01HG011239 (CJY), R01AI136972 (CJY), U01HG012192 (CJY), R37AI114575 (AW)
Howard Hughes Medical Institute (AW)
Chan Zuckerberg Biohub (JD, CJY)
Chan Zuckerberg Initiative (CJY)
Parker Institute for Cancer Immunotherapy (CJY)
Nora Eccles Treadwell Foundation (AW, JFA)
UCSF Center for Rheumatic Diseases (JFA)

Rosalind Russell Medical Research Fou=ndation Bechtel Award (JFA)
Arthritis National Research Foundation (JFA)
Rheumatology Research Foundation K Supplement Award (JFA)
UCSF Research Evaluation Allocation Committee (JFA)

Author Contributions:

Conceptualization: J.F.A., A.W., C.J.Y., J.D.
Methodology: J.F.A, C.J.Y, A.W., E.E.M., S.Y., Y.N.
Investigation: J.F.A, E.E.M., S.Y., C.L., N.P., Y.N., R.N., V.R.
Formal Analysis: J.F.A, E.E.M., C.J.Y.
Visualization: J.F.A, E.E.M., S.Y., C.L., N.P., Y.N., R.N.
Funding acquisition: J.F.A., A.W., C.J.Y., J.D.
Project administration: J.F.A., A.W., C.J.Y., J.D., E.E.M.
Resources: J.F.A., A.W., C.J.Y., J.D.
Supervision: J.F.A., A.W., C.J.Y.
Writing – original draft: J.F.A, E.E.M.
Writing – review & editing: J.F.A., E.E.M., A.W., C.J.Y.

Competing Interests: C.J.Y. is a Scientific Advisory Board member for and hold equity in Related Sciences and ImmunAI, a consultant for and hold equity in Maze Therapeutics, and a consultant for TReX Bio. C.J.Y. has received research support from Chan Zuckerberg Initiative, Chan Zuckerberg Biohub, and Genentech. E.E.M. is a consultant for Alixia and Gate Bioscience.

References

1. Arkwright PD, Abinun M, and Cant AJ. Autoimmunity in human primary immunodeficiency diseases. *Blood.* 2002;99(8):2694-702.
2. Etzioni A. Immune deficiency and autoimmunity. *Autoimmunity reviews.* 2003;2(6):364-9.
3. Fischer A, Provot J, Jais JP, Alcais A, and Mahlaoui N. Autoimmune and inflammatory manifestations occur frequently in patients with primary immunodeficiencies. *The Journal of allergy and clinical immunology.* 2017;140(5):1388-93.e8.
4. Cope AP. T cells in rheumatoid arthritis. *Arthritis research & therapy.* 2008;10 Suppl 1(S1).
5. Schmidt D, Goronzy JJ, and Weyand CM. CD4+ CD7- CD28- T cells are expanded in rheumatoid arthritis and are characterized by autoreactivity. *The Journal of clinical investigation.* 1996;97(9):2027-37.
6. Thomas R, McIlraith M, Davis LS, and Lipsky PE. Rheumatoid synovium is enriched in CD45RBdim mature memory T cells that are potent helpers for B cell differentiation. *Arthritis and rheumatism.* 1992;35(12):1455-65.
7. Weyand CM, Fujii H, Shao L, and Goronzy JJ. Rejuvenating the immune system in rheumatoid arthritis. *Nature reviews Rheumatology.* 2009;5(10):583-8.
8. Zhang Z, Gorman CL, Vermi AC, Monaco C, Foey A, Owen S, et al. TCRzetadim lymphocytes define populations of circulating effector cells that migrate to inflamed tissues. *Blood.* 2007;109(10):4328-35.
9. Gringhuis SI, Papendrecht-van der Voort EA, Leow A, Nivine Levarht EW, Breedveld FC, and Verweij CL. Effect of redox balance alterations on cellular localization of LAT and downstream T-cell receptor signaling pathways. *Molecular and cellular biology.* 2002;22(2):400-11.
10. Maurice MM, Lankester AC, Bezemer AC, Geertsma MF, Tak PP, Breedveld FC, et al. Defective TCR-mediated signaling in synovial T cells in rheumatoid arthritis. *Journal of immunology (Baltimore, Md : 1950).* 1997;159(6):2973-8.
11. Romagnoli P, Strahan D, Pelosi M, Cantagrel A, and van Meerwijk JP. A potential role for protein tyrosine kinase p56(lck) in rheumatoid arthritis synovial fluid T lymphocyte hyporesponsiveness. *International immunology.* 2001;13(3):305-12.
12. Sakaguchi N, Takahashi T, Hata H, Nomura T, Tagami T, Yamazaki S, et al. Altered thymic T-cell selection due to a mutation of the ZAP-70 gene causes autoimmune arthritis in mice. *Nature.* 2003;426(6965):454-60.
13. Ashouri JF, Hsu L-Y, Yu S, Rychkov D, Chen Y, Cheng DA, et al. Reporters of TCR signaling identify arthritogenic T cells in murine and human autoimmune arthritis. *Proceedings of the National Academy of Sciences.* 2019;116(37):18517-27.
14. Tanaka S, Maeda S, Hashimoto M, Fujimori C, Ito Y, Teradaira S, et al. Graded attenuation of TCR signaling elicits distinct autoimmune diseases by altering thymic T cell selection and regulatory T cell function. *Journal of immunology (Baltimore, Md : 1950).* 2010;185(4):2295-305.
15. Hsu LY, Tan YX, Xiao Z, Malissen M, and Weiss A. A hypomorphic allele of ZAP-70 reveals a distinct thymic threshold for autoimmune disease versus autoimmune reactivity. *J Exp Med.* 2009;206(11):2527-41.

16. Yoshitomi H, Sakaguchi N, Kobayashi K, Brown GD, Tagami T, Sakihama T, et al. A role for fungal {beta}-glucans and their receptor Dectin-1 in the induction of autoimmune arthritis in genetically susceptible mice. *J Exp Med.* 2005;201(6):949-60.
17. Ruutu M, Thomas G, Steck R, Degli-Esposti MA, Zinkernagel MS, Alexander K, et al. β-glucan triggers spondylarthritis and Crohn's disease-like ileitis in SKG mice. *Arthritis and rheumatism.* 2012;64(7):2211-22.
18. Evans RM. The steroid and thyroid hormone receptor superfamily. *Science (New York, NY).* 1988;240(4854):889-95.
19. Olefsky JM. Nuclear receptor minireview series. *The Journal of biological chemistry.* 2001;276(40):36863-4.
20. Deane KD, and Holers VM. The Natural History of Rheumatoid Arthritis. *Clin Ther.* 2019;41(7):1256-69.
21. Raudvere U, Kolberg L, Kuzmin I, Arak T, Adler P, Peterson H, et al. g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic acids research.* 2019;47(W1):W191-w8.
22. Ashouri JF, and Weiss A. Endogenous Nur77 Is a Specific Indicator of Antigen Receptor Signaling in Human T and B Cells. *Journal of immunology (Baltimore, Md : 1950).* 2017;198(2):657-68.
23. Huang B, Pei HZ, Chang HW, and Baek SH. The E3 ubiquitin ligase Trim13 regulates Nur77 stability via casein kinase 2alpha. *Scientific reports.* 2018;8(1):13895.
24. Tan C, Hiwa R, Mueller JL, Vykunta V, Hibiya K, Noviski M, et al. NR4A nuclear receptors restrain B cell responses to antigen when second signals are absent or limiting. *Nature immunology.* 2020;21(10):1267-79.
25. Zhang L, Xie F, Zhang J, Dijke PT, and Zhou F. SUMO-triggered ubiquitination of NR4A1 controls macrophage cell death. *Cell Death Differ.* 2017;24(9):1530-9.
26. Jennings E, Elliot TAE, Thawait N, Kanabar S, Yam-Puc JC, Ono M, et al. Nr4a1 and Nr4a3 Reporter Mice Are Differentially Sensitive to T Cell Receptor Signal Strength and Duration. *Cell reports.* 2020;33(5):108328.
27. Moran AE, Holzapfel KL, Xing Y, Cunningham NR, Maltzman JS, Punt J, et al. T cell receptor signal strength in Treg and iNKT cell development demonstrated by a novel fluorescent reporter mouse. *J Exp Med.* 2011;208(6):1279-89.
28. Myers DR, Lau T, Markegard E, Lim HW, Kasler H, Zhu M, et al. Tonic LAT-HDAC7 Signals Sustain Nur77 and Irf4 Expression to Tune Naive CD4 T Cells. *Cell Rep.* 2017;19(8):1558-71.
29. Myers DR, Zikherman J, and Roose JP. Tonic Signals: Why Do Lymphocytes Bother? *Trends in immunology.* 2017;38(11):844-57.
30. Kalekar LA, Schmiel SE, Nandiwada SL, Lam WY, Barsness LO, Zhang N, et al. CD4(+) T cell anergy prevents autoimmunity and generates regulatory T cell precursors. *Nature immunology.* 2016;17(3):304-14.
31. Mueller DL. E3 ubiquitin ligases as T cell anergy factors. *Nature immunology.* 2004;5(9):883-90.
32. Zheng Y, Zha Y, Driessens G, Locke F, and Gajewski TF. Transcriptional regulator early growth response gene 2 (Egr2) is required for T cell anergy in vitro and in vivo. *J Exp Med.* 2012;209(12):2157-63.

33. Crawford A, Angelosanto JM, Kao C, Doering TA, Odorizzi PM, Barnett BE, et al. Molecular and transcriptional basis of CD4(+) T cell dysfunction during chronic infection. *Immunity.* 2014;40(2):289-302.

34. Trefzer A, Kadam P, Wang SH, Pennavaria S, Lober B, Akcabozan B, et al. Dynamic adoption of anergy by antigen-exhausted CD4(+) T cells. *Cell reports.* 2021;34(6):108748.

35. Chen J, Lopez-Moyado IF, Seo H, Lio CJ, Hempleman LJ, Sekiya T, et al. NR4A transcription factors limit CAR T cell function in solid tumours. *Nature.* 2019.

36. ElTanbouly MA, Zhao Y, Nowak E, Li J, Schaafsma E, Le Mercier I, et al. VISTA is a checkpoint regulator for naïve T cell quiescence and peripheral tolerance. *Science (New York, NY).* 2020;367(6475).

37. Zinzow-Kramer WM, Weiss A, and Au-Yeung BB. Adaptation by naïve CD4(+) T cells to self-antigen-dependent TCR signaling induces functional heterogeneity and tolerance. *Proc Natl Acad Sci U S A.* 2019;116(30):15160-9.

38. Nguyen TTT, Wang ZE, Shen L, Schroeder A, Eckalbar W, and Weiss A. Cbl-b deficiency prevents functional but not phenotypic T cell anergy. *J Exp Med.* 2021;218(7).

39. Liebmann M, Hucke S, Koch K, Eschborn M, Ghelman J, Chasan AI, et al. Nur77 serves as a molecular brake of the metabolic switch during T cell activation to restrict autoimmunity. *Proc Natl Acad Sci U S A.* 2018;115(34):E8017-e26.

40. Liu X, Wang Y, Lu H, Li J, Yan X, Xiao M, et al. Genome-wide analysis identifies NR4A1 as a key mediator of T cell dysfunction. *Nature.* 2019;567(7749):525-9.

41. Hata H, Sakaguchi N, Yoshitomi H, Iwakura Y, Sekikawa K, Azuma Y, et al. Distinct contribution of IL-6, TNF-alpha, IL-1, and IL-10 to T cell-mediated spontaneous autoimmune arthritis in mice. *The Journal of clinical investigation.* 2004;114(4):582-8.

42. Hirota K, Hashimoto M, Yoshitomi H, Tanaka S, Nomura T, Yamaguchi T, et al. T cell self-reactivity forms a cytokine milieu for spontaneous development of IL-17+ Th cells that cause autoimmune arthritis. *J Exp Med.* 2007;204(1):41-7.

43. Ciofani M, Madar A, Galan C, Sellars M, Mace K, Pauli F, et al. A validated regulatory network for Th17 cell specification. *Cell.* 2012;151(2):289-303.

44. Yu CR, Mahdi RM, Ebong S, Vistica BP, Gery I, and Egwuagu CE. Suppressor of cytokine signaling 3 regulates proliferation and activation of T-helper cells. *The Journal of biological chemistry.* 2003;278(32):29752-9.

45. Ye H, Zhang J, Wang J, Gao Y, Du Y, Li C, et al. CD4 T-cell transcriptome analysis reveals aberrant regulation of STAT3 and Wnt signaling pathways in rheumatoid arthritis: evidence from a case-control study. *Arthritis research & therapy.* 2015;17:76.

46. Atsumi T, Ishihara K, Kamimura D, Ikushima H, Ohtani T, Hirota S, et al. A point mutation of Tyr-759 in interleukin 6 family cytokine receptor subunit gp130 causes autoimmune arthritis. *J Exp Med.* 2002;196(7):979-90.

47. Shouda T, Yoshida T, Hanada T, Wakioka T, Oishi M, Miyoshi K, et al. Induction of the cytokine signal regulator SOCS3/CIS3 as a therapeutic strategy for treating inflammatory arthritis. *The Journal of clinical investigation.* 2001;108(12):1781-8.

48. Wong PK, Egan PJ, Croker BA, O'Donnell K, Sims NA, Drake S, et al. SOCS-3 negatively regulates innate and adaptive immune mechanisms in acute IL-1-dependent inflammatory arthritis. *The Journal of clinical investigation.* 2006;116(6):1571-81.

49. Bergen V, Lange M, Peidli S, Wolf FA, and Theis FJ. Generalizing RNA velocity to transient cell states through dynamical modeling. *Nature biotechnology.* 2020;38(12):1408-14.

50. Reynolds D. In: Li SZ, and Jain A eds. *Encyclopedia of Biometrics*. Boston, MA: Springer US; 2009:659-63.

51. Wolf FA, Hamey FK, Plass M, Solana J, Dahlin JS, Göttgens B, et al. PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. *Genome biology.* 2019;20(1):59.

52. Abdelnour A, Bremell T, Holmdahl R, and Tarkowski A. Clonal expansion of T lymphocytes causes arthritis and mortality in mice infected with toxic shock syndrome toxin-1-producing staphylococci. *European journal of immunology.* 1994;24(5):1161-6.

53. Kappler J, Kotzin B, Herron L, Gelfand EW, Bigler RD, Boylston A, et al. V beta-specific stimulation of human T cells by staphylococcal toxins. *Science (New York, NY).* 1989;244(4906):811-3.

54. Hodes RJ, and Abe R. Mouse endogenous superantigens: Ms and Mls-like determinants encoded by mouse retroviruses. *Current protocols in immunology.* 2001;Appendix 1:Appendix 1F.

55. Matsutani T, Ohmori T, Ogata M, Soga H, Yoshioka T, Suzuki R, et al. Alteration of T-cell receptor repertoires during thymic T-cell development. *Scandinavian journal of immunology.* 2006;64(1):53-60.

56. Czarneski J, Rassa JC, and Ross SR. Mouse mammary tumor virus and the immune system. *Immunol Res.* 2003;27(2-3):469-80.

57. Acha-Orbea H, Shakhov AN, Scarpellino L, Kolb E, Müller V, Vessaz-Shaw A, et al. Clonal deletion of V beta 14-bearing T cells in mice transgenic for mammary tumour virus. *Nature.* 1991;350(6315):207-11.

58. Choi Y, Kappler JW, and Marrack P. A superantigen encoded in the open reading frame of the 3' long terminal repeat of mouse mammary tumour virus. *Nature.* 1991;350(6315):203-7.

59. Frankel WN, Rudy C, Coffin JM, and Huber BT. Linkage of Mls genes to endogenous mammary tumour viruses of inbred mice. *Nature.* 1991;349(6309):526-8.

60. Barnett A, Mustafa F, Wrona TJ, Lozano M, and Dudley JP. Expression of mouse mammary tumor virus superantigen mRNA in the thymus correlates with kinetics of self-reactive T-cell loss. *Journal of virology.* 1999;73(8):6634-45.

61. Scherer MT, Ignatowicz L, Pullen A, Kappler J, and Marrack P. The use of mammary tumor virus (Mtv)-negative and single-Mtv mice to evaluate the effects of endogenous viral superantigens on the T cell repertoire. *J Exp Med.* 1995;182(5):1493-504.

62. Acha-Orbea H, Shakhov AN, and Finke D. Immune response to MMTV infection. *Front Biosci.* 2007;12:1594-609.

63. Zinzow-Kramer WM, Kolawole EM, Eggert J, Evavold BD, Scharer CD, and Au-Yeung BB. Strong Basal/Tonic TCR Signals Are Associated with Negative Regulation of Naive CD4(+) T Cells. *Immunohorizons.* 2022;6(9):671-83.

64. Omenetti S, Bussi C, Metidji A, Iseppon A, Lee S, Tolaini M, et al. The Intestine Harbors Functionally Distinct Homeostatic Tissue-Resident and Inflammatory Th17 Cells. *Immunity.* 2019;51(1):77-89.e6.

65. Peters A, Lee Y, and Kuchroo VK. The many faces of Th17 cells. *Current opinion in immunology.* 2011;23(6):702-6.

66. Sharon D, Chen M, Zhang G, Girgis S, Sis B, Graham D, et al. Impact of combination antiretroviral therapy in the NOD.c3c4 mouse model of autoimmune biliary disease. *Liver Int.* 2015;35(4):1442-50.

67. Lima-Junior DS, Krishnamurthy SR, Bouladoux N, Collins N, Han SJ, Chen EY, et al. Endogenous retroviruses promote homeostatic and inflammatory responses to the microbiota. *Cell.* 2021;184(14):3794-811.e19.

68. Armstrong H, Rahbari M, Park H, Sharon D, Thiesen A, Hotte N, et al. Mouse mammary tumor virus is implicated in severity of colitis and dysbiosis in the IL-10(-/-) mouse model of inflammatory bowel disease. *Microbiome.* 2023;11(1):39.

69. Young GR, Eksmond U, Salcedo R, Alexopoulou L, Stoye JP, and Kassiotis G. Resurrection of endogenous retroviruses in antibody-deficient mice. *Nature.* 2012;491(7426):774-8.

70. Bhadra S, Lozano MM, Payne SM, and Dudley JP. Endogenous MMTV proviruses induce susceptibility to both viral and bacterial pathogens. *PLoS Pathog.* 2006;2(12):e128.

71. Bour-Jordan H, Esensten JH, Martinez-Llordella M, Penaranda C, Stumpf M, and Bluestone JA. Intrinsic and extrinsic control of peripheral T-cell tolerance by costimulatory molecules of the CD28/ B7 family. *Immunological reviews.* 2011;241(1):180-205.

72. Chen L, and Flies DB. Molecular mechanisms of T cell co-stimulation and co-inhibition. *Nature reviews Immunology.* 2013;13(4):227-42.

73. Fife BT, and Bluestone JA. Control of peripheral T-cell tolerance and autoimmunity via the CTLA-4 and PD-1 pathways. *Immunological reviews.* 2008;224:166-82.

74. Lee J, Su EW, Zhu C, Hainline S, Phuah J, Moroco JA, et al. Phosphotyrosine-dependent coupling of Tim-3 to T-cell receptor signaling pathways. *Molecular and cellular biology.* 2011;31(19):3963-74.

75. Ross SR. Mouse mammary tumor virus and the immune system. *Adv Pharmacol.* 1997;39:21-46.

76. Tugnet N, Rylance P, Roden D, Trela M, and Nelson P. Human Endogenous Retroviruses (HERVs) and Autoimmune Rheumatic Disease: Is There a Link? *Open Rheumatol J.* 2013;7:13-21.

77. Ito Y, Hashimoto M, Hirota K, Ohkura N, Morikawa H, Nishikawa H, et al. Detection of T cell responses to a ubiquitous cellular protein in autoimmune disease. *Science (New York, NY).* 2014;346(6207):363-8.

78. Howell MD, Diveley JP, Lundeen KA, Esty A, Winters ST, Carlo DJ, et al. Limited T-cell receptor beta-chain heterogeneity among interleukin 2 receptor-positive synovial T cells suggests a role for superantigen in rheumatoid arthritis. *Proc Natl Acad Sci U S A.* 1991;88(23):10921-5.

79. Paliard X, West SG, Lafferty JA, Clements JR, Kappler JW, Marrack P, et al. Evidence for the effects of a superantigen in rheumatoid arthritis. *Science (New York, NY).* 1991;253(5017):325-9.

80. Jenkins RN, Nikaein A, Zimmermann A, Meek K, and Lipsky PE. T cell receptor V beta gene bias in rheumatoid arthritis. *The Journal of clinical investigation.* 1993;92(6):2688-701.

81. Joo YB, Lim YH, Kim KJ, Park KS, and Park YJ. Respiratory viral infections and the risk of rheumatoid arthritis. *Arthritis research & therapy.* 2019;21(1):199.

82. Turcinov S, Af Klint E, Van Schoubroeck B, Kouwenhoven A, Mia S, Chemin K, et al. Diversity and Clonality of T Cell Receptor Repertoire and Antigen Specificities in Small Joints of Early Rheumatoid Arthritis. *Arthritis & rheumatology (Hoboken, NJ).* 2023;75(5):673-84.

83. Zhou J, Kong C, Yu J, Dong H, Jin C, and Song Q. Skewness of TCR Vβ of peripheral blood and synovial fluid of patients with rheumatoid arthritis. *J Immunoassay Immunochem.* 2014;35(2):207-19.

84. Bhardwaj N, Hodtsev AS, Nisanian A, Kabak S, Friedman SM, Cole BC, et al. Human T-cell responses to Mycoplasma arthritidis-derived superantigen. *Infection and immunity.* 1994;62(1):135-44.

85. Britanova OV, Lupyr KR, Staroverov DB, Shagina IA, Aleksandrov AA, Ustyugov YY, et al. Targeted depletion of TRBV9(+) T cells as immunotherapy in a patient with ankylosing spondylitis. *Nature medicine.* 2023;29(11):2731-6.

86. Faham M, Carlton V, Moorhead M, Zheng J, Klinger M, Pepin F, et al. Discovery of T Cell Receptor β Motifs Specific to HLA-B27-Positive Ankylosing Spondylitis by Deep Repertoire Sequence Analysis. *Arthritis & rheumatology (Hoboken, NJ).* 2017;69(4):774-84.

87. Grom AA, Thompson SD, Luyrink L, Passo M, Choi E, and Glass DN. Dominant T-cell-receptor beta chain variable region V beta 14+ clones in juvenile rheumatoid arthritis. *Proc Natl Acad Sci U S A.* 1993;90(23):11104-8.

88. Komech EA, Pogorelyy MV, Egorov ES, Britanova OV, Rebrikov DV, Bochkova AG, et al. CD8+ T cells with characteristic T cell receptor beta motif are detected in blood and expanded in synovial fluid of ankylosing spondylitis patients. *Rheumatology (Oxford, England).* 2018;57(6):1097-104.

89. Gallagher MP, Conley JM, and Berg LJ. Peptide Antigen Concentration Modulates Digital NFAT1 Activation in Primary Mouse Naive CD8(+) T Cells as Measured by Flow Cytometry of Isolated Cell Nuclei. *Immunohorizons.* 2018;2(7):208-15.

90. Brooks JF, Riggs J, Mueller JL, Mathenge R, Wholey WY, Meyer AR, et al. Molecular basis for potent B cell responses to antigen displayed on particles of viral size. *Nature immunology.* 2023;24(10):1762-77.

91. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics (Oxford, England).* 2013;29(1):15-21.

92. Love MI, Huber W, and Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology.* 2014;15(12):550.

93. Thompson WL, White GC, and Gowan C. In: Thompson WL, White GC, and Gowan C eds. *Monitoring Vertebrate Populations*. Academic Press; 1998:145-69.

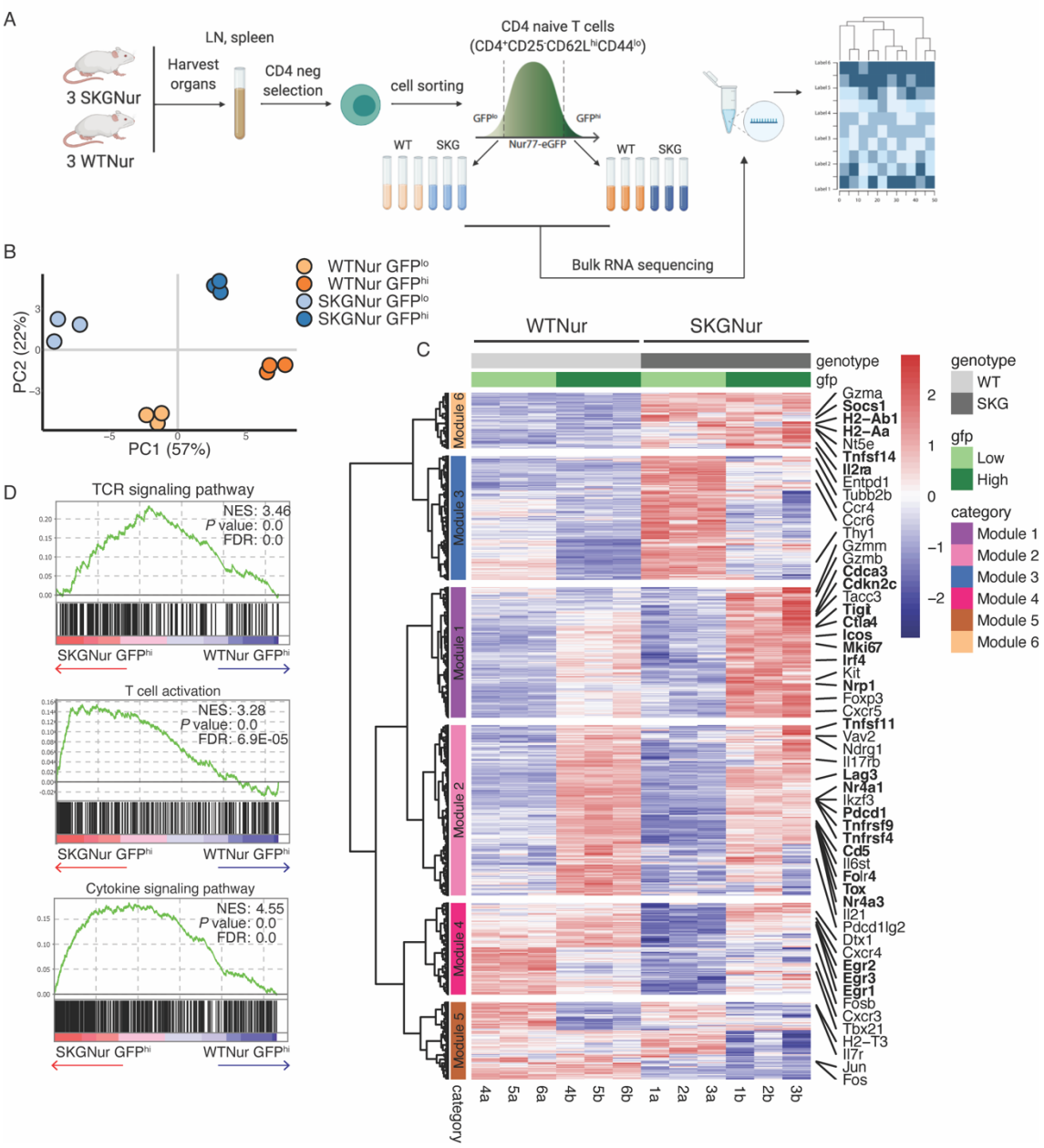


Figure 1. Pre-arthritic naïve SKG T cells demonstrate enhanced T cell activation. (**A**) Experimental schematic of bulk RNA-seq analysis. (**B**) Principal component analysis (PCA) based on transcriptomic data from bulk RNA-seq data shows distribution of SKGNur and WTNur GFP^{lo} and GFP^{hi} CD4 naïve T cell subsets as shown in (A) (*n*=3 per subgroup). (**C**) Heatmap showing expression of 991 significantly differentially expressed genes (DEGs, |log2(FC)| > 1, adjusted P value < 0.05) from pairwise comparisons for all samples grouped by subgroup. Hierarchical clustering was used to group DEGs into 6 modules (indicated by dendrogram and row annotation color bar on left). (**D**) Enrichment plots of TCR signaling and cytokine pathways from GSEA analysis of all GO:BP pathways for ranked genes from SKGNur GFP^{hi} and WTNur GFP^{hi} differential expression analysis. FDR, false discovery rate. NES, normalized enrichment score.

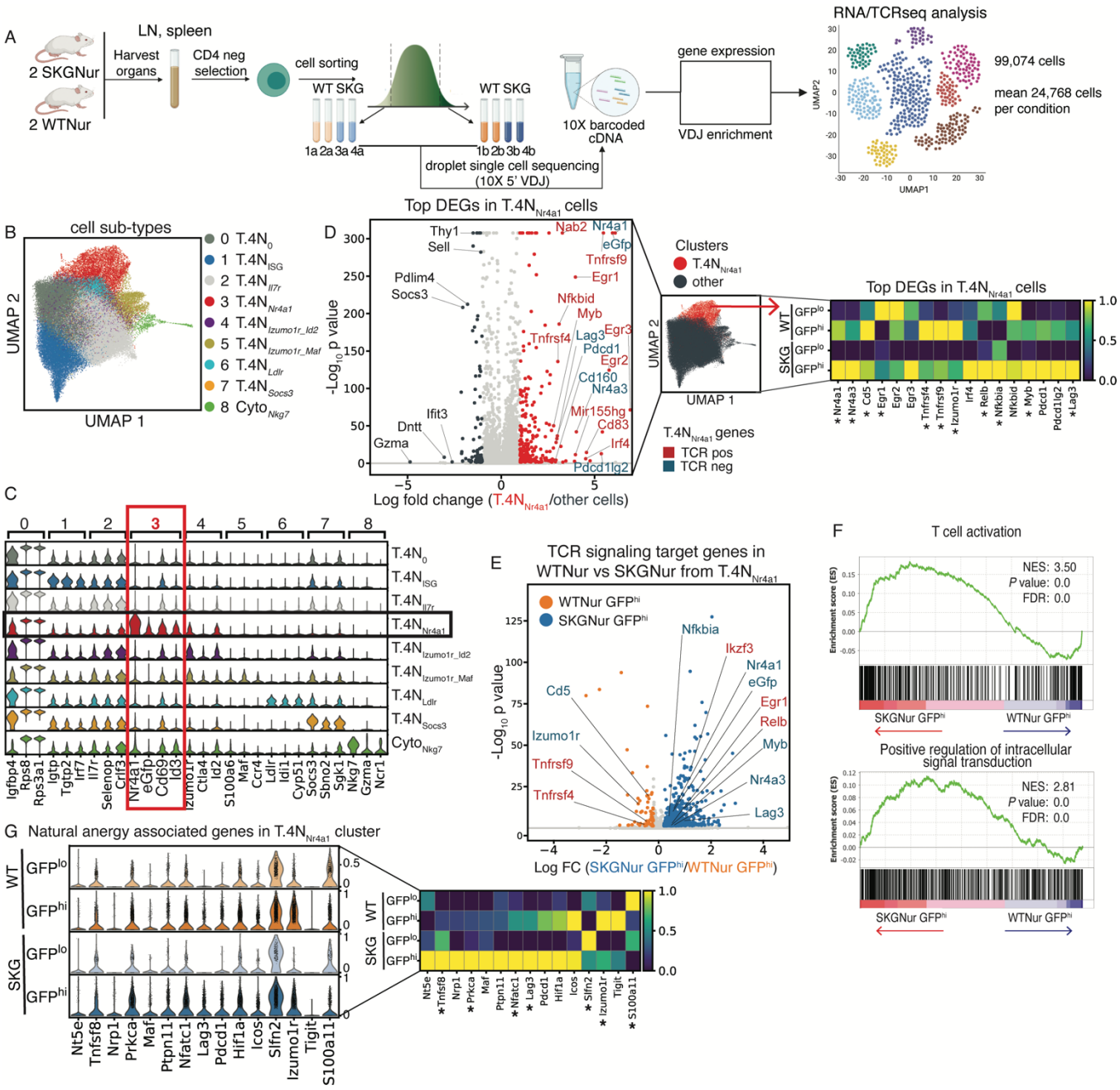


Figure 2. Single-cell RNA sequencing reveals heterogeneity in naïve CD4 T cells, highlighting a subset defined by TCR signaling genes. (**A**) Experimental design of paired scRNA- and TCR-seq of sorted GFP^{hi} and GFP^{lo} naïve CD4 T cells. (**B**) UMAP of 99,074 naïve T cells from 8 samples in (**A**) colored by merged clusters. (**C**) Violin plots of log-normalized expression of marker genes for each cluster. Black box highlights T.4N$_{Nr4a1}$ cluster, red box highlights genes uniquely expressed in T.4N$_{Nr4a1}$ cluster. (**D**) Volcano plot of DEGs in the T.4N$_{Nr4a1}$ cluster versus all other cells. Dots are colored by significant overexpression (|log2(FC)| > 1, adj. P < 0.05) in the T.4N$_{Nr4a1}$ cluster (red), other cells (dark grey), or no significant difference (light grey). Labelled genes are colored by role in TCR signaling regulation: positive (red) or negative (blue). Heatmap shows average expression of the labelled genes by subgroup normalized by standard scale for each gene. (**E**) Volcano plot of DEGs from SKGNur versus WTNur GFP^{hi} in T.4N$_{Nr4a1}$ cluster. Dots are colored as significantly overexpressed (|log2(FC)| > 0.2, adj. P < 0.05) in WTNur (orange) or SKGNur (blue) GFP^{hi} or not significantly different between groups (grey). Labelled genes involved in TCR signaling are colored as in **(D)**. (**F**)

Enrichment plots of TCR activation and signaling pathways from GSEA analysis of GO:BP pathways for ranked DEGs of SKGNur versus WTNur GFP^{hi} cells from T.4N$_{Nr4a1}$ cluster. FDR, false discovery rate. NES, normalized enrichment score. (**G**) Violin plots show expression of 'natural anergy' genes in WTNur and SKGNur GFP^{lo} and GFP^{hi} CD4 naïve cells from the T.4N$_{Nr4a1}$ cluster. Heatmap displays average gene expression by subgroup. Both are normalized by standard scale. Asterisks (*) in heatmap (**E** and **G**) indicate significant differential gene expression between SKGNur and WTNur GFP^{hi} cells.

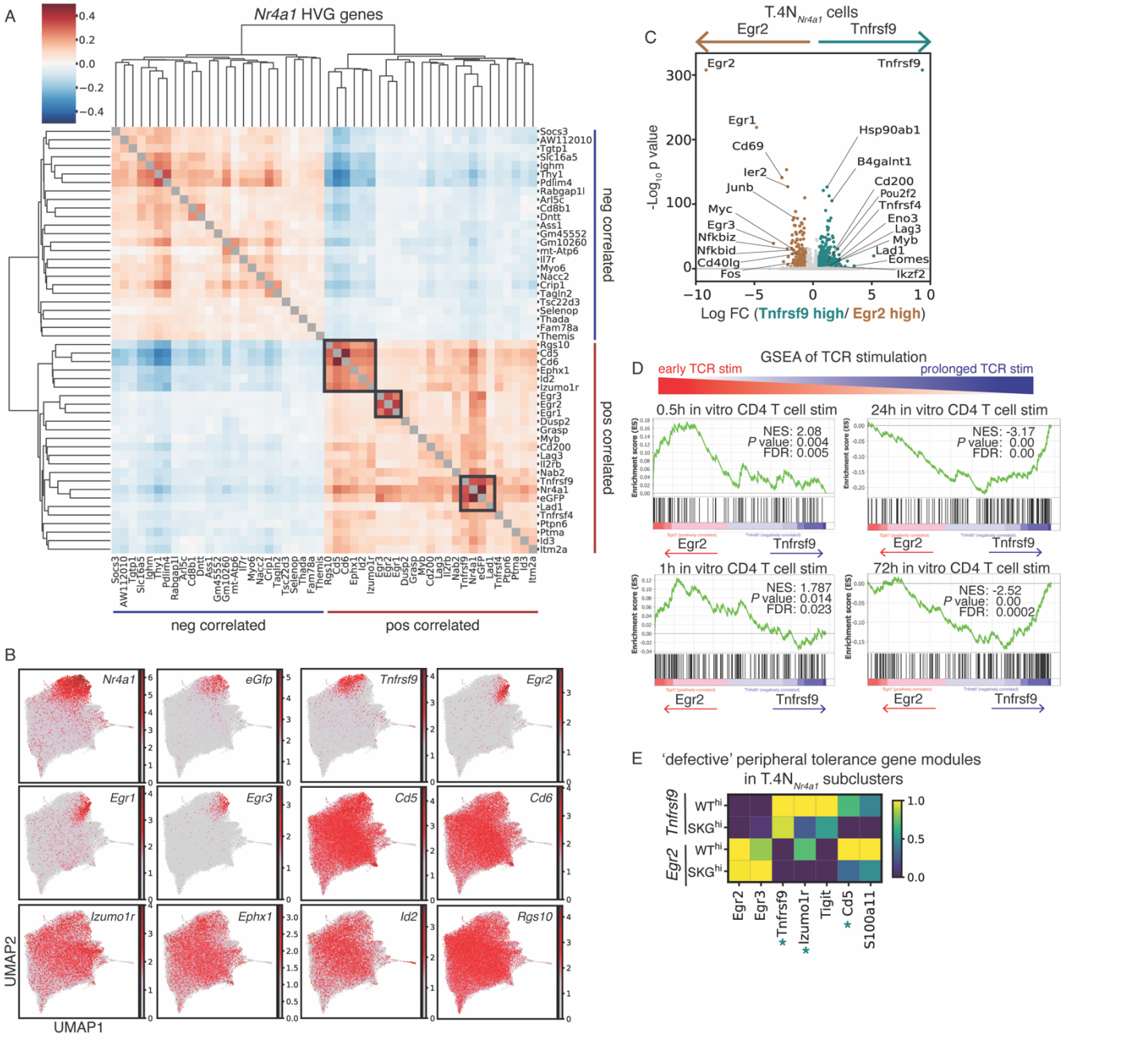


Figure 3. T.N4*Nr4a1* cells segregate into two distinct TCR signaling modules that segregate acute from chronic antigen-activated T cells. (**A**) Correlation matrix shows hierarchical clustering of Spearman's correlation of top 25 highly variable genes (HVG) that positively and negatively correlate with *Nr4a1* expression across all cells (SKG and WT). Diagonal grey boxes represent correlation of 1. Dark grey boxes mark distinct gene modules from genes that positively correlate with *Nr4a1* expression. (**B**) UMAP plots show the expression levels in all cells of indicated marker genes positively correlating with *Nr4a1* as identified in (**A**). Scale represents the log-transformed normalized counts of genes. (**C**) Volcano plot shows DEGs for SKG and WT cells in T.4N$_{Nr4a1}$ cluster that expressed (log normalized expression > 1) *Egr2* or *Tnfrsf9* with dots colored by significant overexpression (|log2(fold-change)| > 0.5, adj. P < 0.05) in *Egr2* (tan) or *Tnfrsf9* (teal) expressing cells. (**D**) Enrichment plots from GSEA analysis of study GSE17974 pathways of time course *in vitro* activation of CD4+ T cells with αCD3 + CD28 for ranked genes from differential gene expression analysis of cells in

T.4N$_{Nr4a1}$ cluster that express *Egr2* versus *Tnfrsf9*. FDR, false discovery rate. NES, normalized enrichment score. (**E**) Heatmap of average expression of peripheral tolerance defect signature genes from WTNur and SKGNur GFP^{hi} cells expressing *Egr2* or Tnfrsf9 in T.4N$_{Nr4a1}$ cluster normalized by standard scale for each gene. Teal asterisk (*) next to genes in heatmap marks significant differential expression between SKGNur GFP^{hi} and WTNur GFP^{hi} cells in the *Tnfrsf9* subcluster.

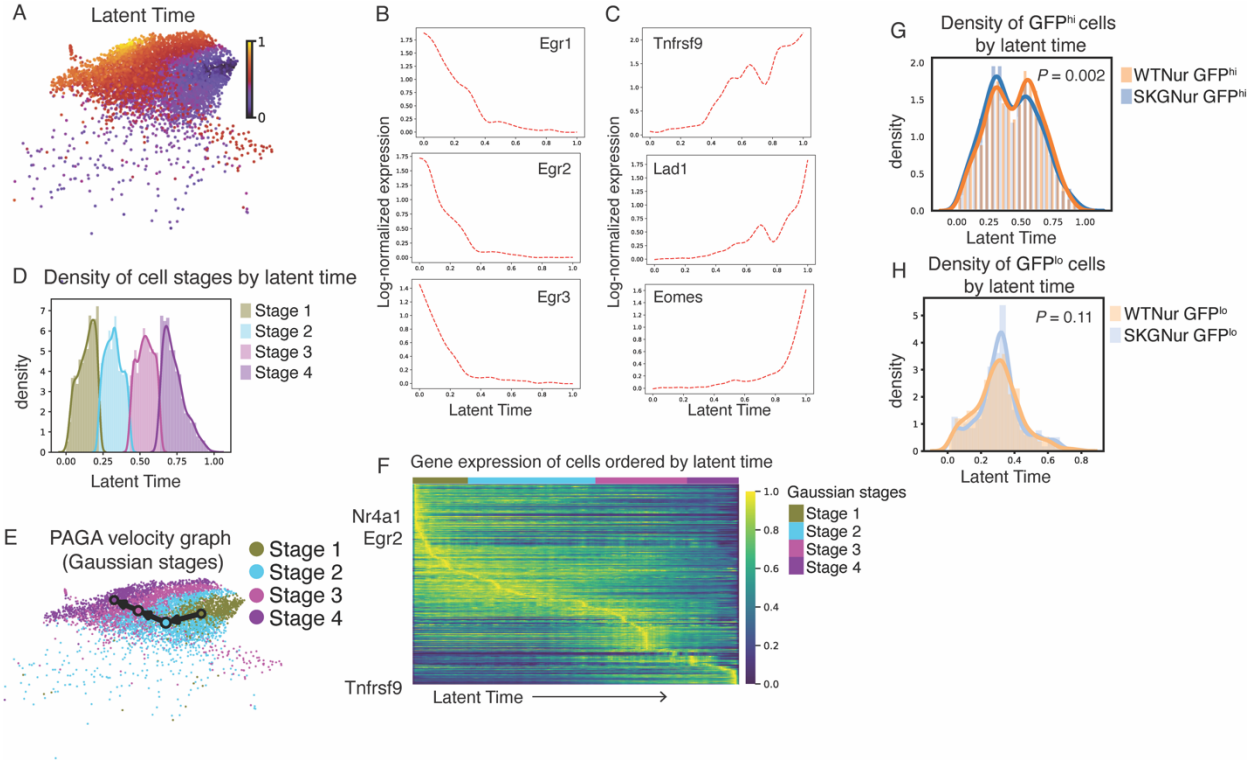


Figure 4. Trajectory analysis of T.4N$_{Nr4a1}$ cells orthogonally uncovers acute versus chronic antigen-activated T cell states with a distinct distribution in the SKGNur GFP^{hi} subset. (**A**) UMAP of cells from T.4N$_{Nr4a1}$ cluster colored by latent time. (**B-C**) Smoothed gene expression from cells in T.4N$_{Nr4a1}$ cluster of selected genes with highest expression earlier (**B**) or later (**C**) along latent time axis. (**D**) Probability densities of latent time distribution of cells from T.4N$_{Nr4a1}$ cluster assigned to 4 distinct clusters (labelled “Stage 1” – “Stage 4”) by a Gaussian mixture model. (**E**) Predicted transitions from PAGA algorithm between cells from stages indicated in **(D)**. (**F**) Heatmap of single cell standard scale normalized expression of genes ordered top to bottom by peak expression at earlier to later latent time. Chosen genes are the top 300 highest confidence genes used in modeling of latent time. Column annotation bar indicates stage assignment of the cell in each column. (**G-H**) Probability densities of latent time distribution for GFP^{hi} **(G)** and GFP^{lo} **(H)** cells from WTNur and SKGNur mice with P values (Kolmogorov-Smirnov test).

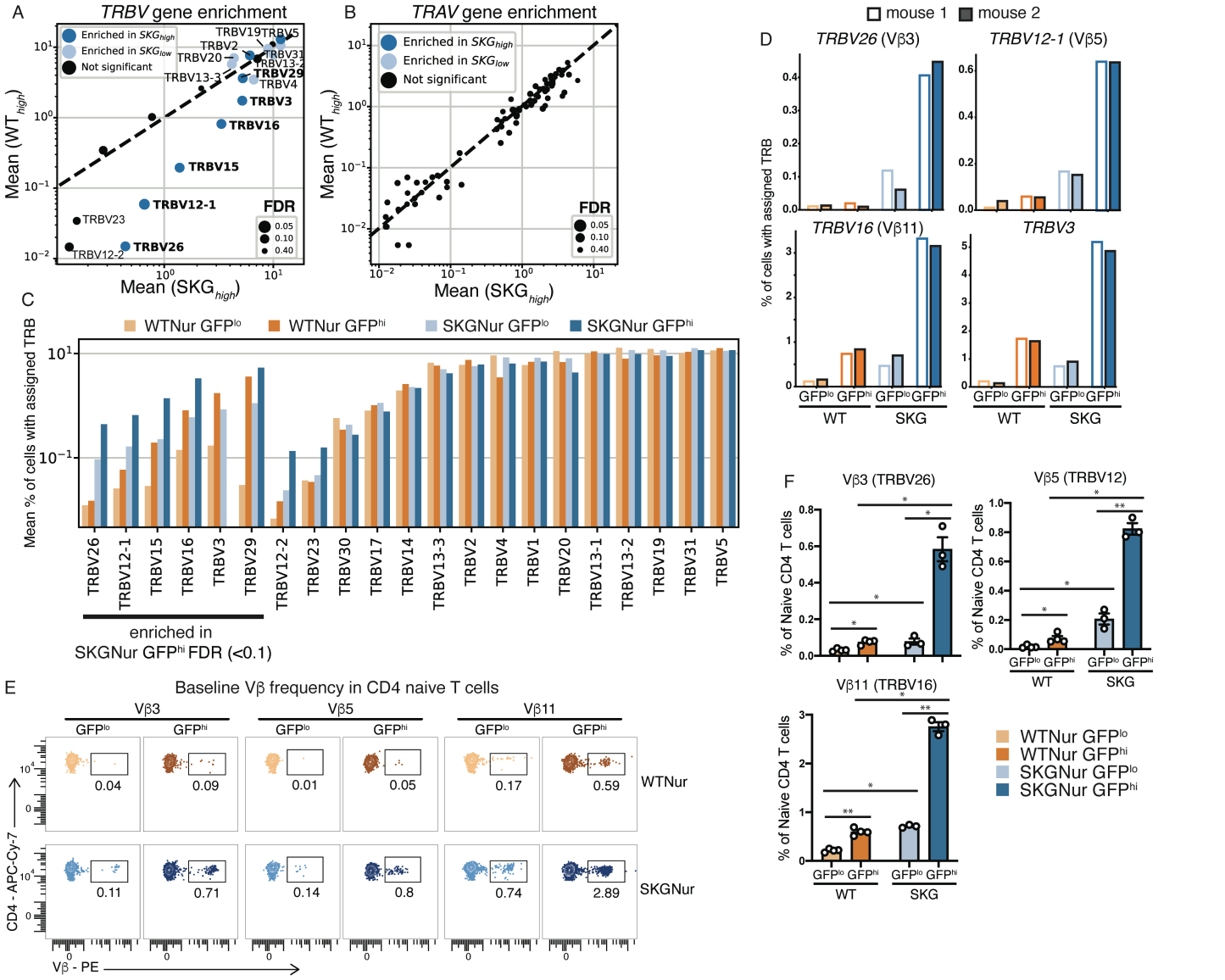


Figure 5. SKG CD4 T cells harbor a biased TCR variable beta gene repertoire. (**A-B**) Scatterplot of mean frequency of cells expressing each TRBV (**A**) or TRAV (**B**) gene for the SKGNur GFP^{hi} samples versus the WTNur GFP^{hi} samples. Dots for each TRBV and TRAV genes are sized according to the FDR from a one-sided paired t-test comparing frequency in SKGNur GFP^{hi} versus SKGNur GFP^{lo}. Dots are colored as either significantly enriched (FDR < 0.1) in SKGNur GFP^{hi} (dark blue), significantly enriched in SKGNur GFP^{lo} (light blue), or not significantly enriched in either subgroup (black). Dots for significant TRBV genes are labelled with the TRBV gene name. Labels for TRBV genes that were significantly enriched in SKGNur GFP^{hi} and were also more highly expressed in SKGNur GFP^{hi} samples versus WTNur GFP^{hi} samples are bolded. (**C**) Bar plot of mean value of cells expressing each TRBV gene as a percentage of all cells in each sample with an assigned TRBV. Bars are colored according to subgroup and ordered with the TRBV genes enriched in SKGNur GFP^{hi} from **(A)** followed by the other TRBV genes ordered by increasing overall frequency. (**D**) Bar plots of frequency of cells expressing indicated TRBV genes significantly enriched in SKGNur GFP^{hi} for the two replicate mice in each subgroup. (**E-F**) Representative FACS plots (**E**) of naïve peripheral CD4 T cells with indicated TCR Vβ protein usage determined by flow cytometry in GFP^{lo} and GFP^{hi} T cells from lymph nodes (LN) of WTNur and SKGNur mice prior to arthritis induction and quantified in (**F**) where bar graphs depict mean frequency (± SEM), n = 3-4 mice per group, experiment repeated at least 3 times. Significance indicated by asterisk for FDR (paired t-test with BH correction) or P value (exact permutation test) < 0.05 (*), < 0.01 (**).

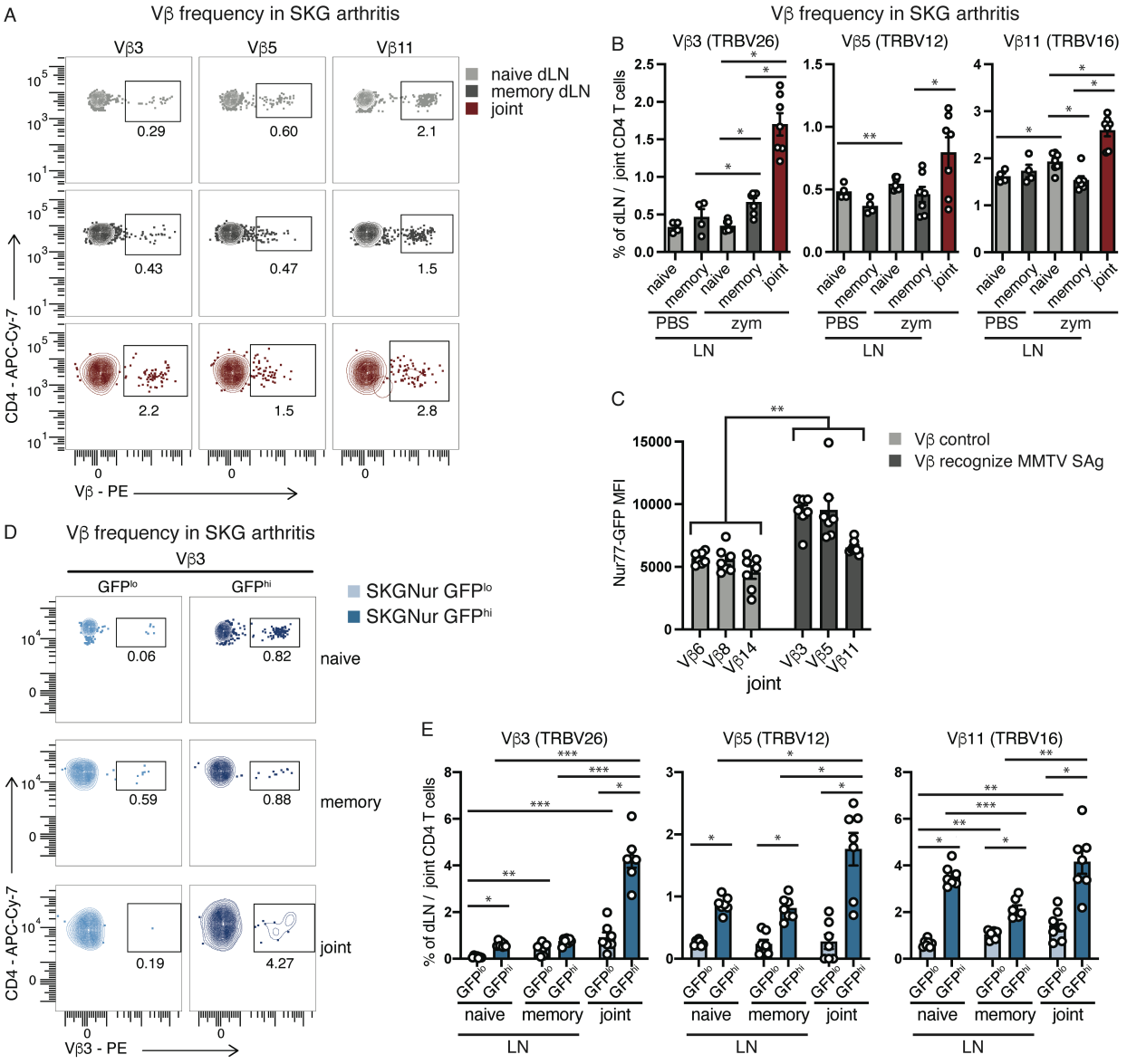


Figure 6. Arthritogenic CD4 T cells are enriched for TCR Vβs likely driven by endogenous superantigen(s). (**A-B**) Representative FACS plots (**A**) of peripheral naïve or memory, or joint CD4 T cells with indicated TCR Vβ protein usage determined by flow cytometry in CD4 T cells from draining LN or joints of SKGNur mice 2.5 weeks after arthritis induction with zymosan or vehicle treatment with PBS (as seen in **figure S8B**) and quantified in (**B**) where bar graphs depict mean frequency (± SEM). (**C**) Bar graphs of GFP mean fluorescence intensity (MFI ± SEM) of CD4 T cells bearing indicated Vβs from arthritic joints of SKG mice, n = 7 mice pooled from 2 experiments (also reported in **figure S8D**). (**D-E**) Representative FACS plots of (**D**) peripheral naïve or memory, or joint CD4 T cells with indicated TCR Vβ protein usage determined by flow cytometry in GFP^{lo} (light blue) and GFP^{hi} (dark blue) T cells from LN or joints of SKGNur mice 2.5 weeks after arthritis induction with zymosan and quantified in (**E**) where bar graphs depict mean frequency (± SEM), n = 7 mice per group pooled from 2 experiments. Significance indicated by asterisk [< 0.05 (*), < 0.01 (**), or < 0.001 (***)] for FDR (paired t-test with BH correction) or *P* value (exact permutation test) (**B** and **E**) or FDR (linear mixed effect model with BH correction) (**C**).

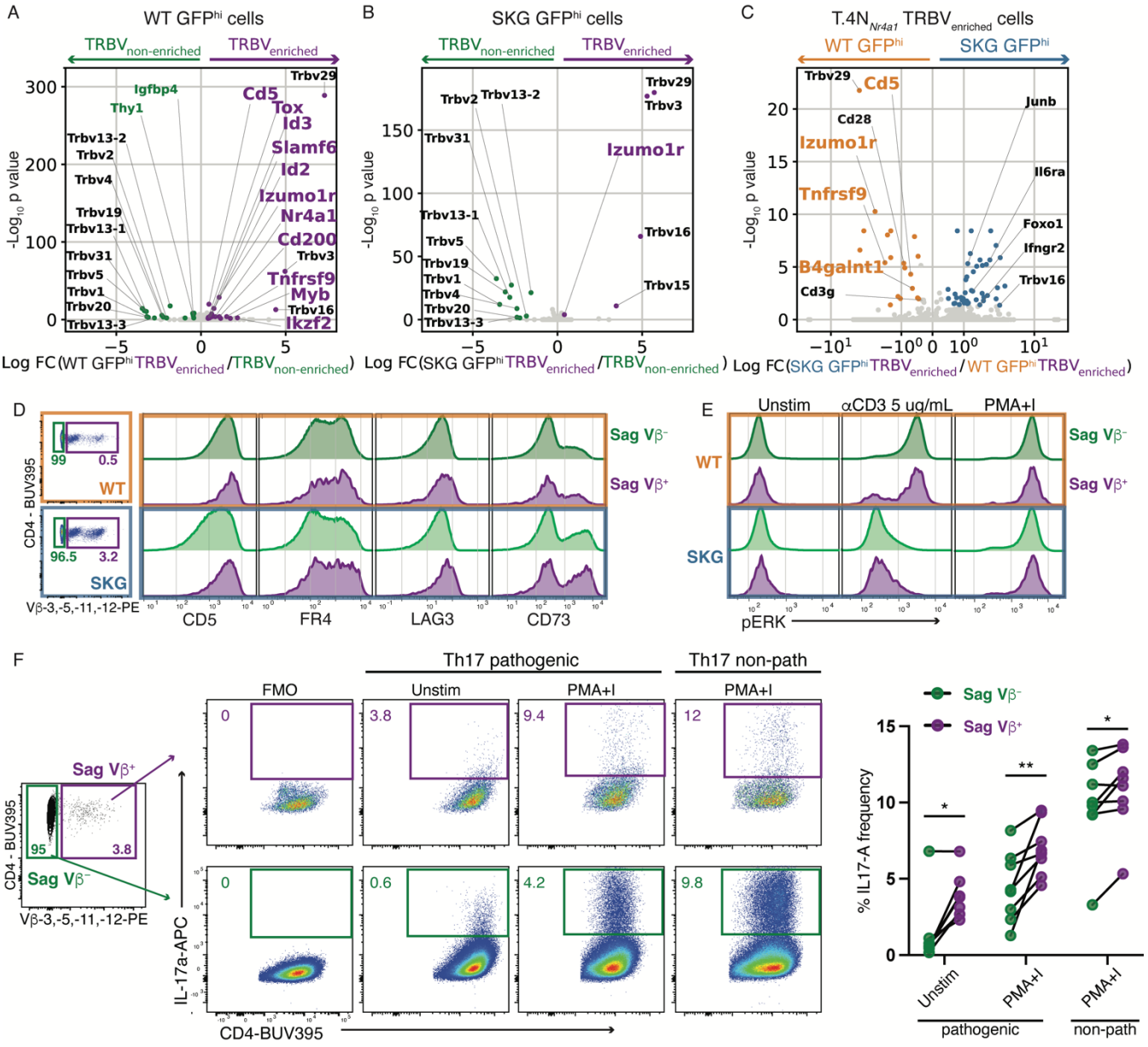


Figure 7. Sag-reactive SKG CD4 T cells show impaired tolerance, defective signaling, and a Th17 differentiation bias. (**A-B**) Volcano plots of DEGs from WT (**A**) or SKG (**B**) GFP^{hi} cells comparing Sag-reactive TRBV$_{enriched}$ vs non-Sag reactive TRBV$_{non-enriched}$ with significant overexpression (|log2(fold-change)| >0.4, adj P <0.05) in TRBV$_{enriched}$ (purple) or TRBV$_{non-enriched}$ (green). (**C**) Volcano plots of DEGs from T.4N$_{Nr4a1}$ cluster TRBV$_{enriched}$ cells from WT versus SKG GFP^{hi} mice with significant overexpression (|log2(fold-change)| >0.4, adj P <0.05) in TRBV$_{enriched}$ (purple) or TRBV$_{non-enriched}$ (green). (**D**) Left panel: representative FACS plots of the gating strategy for peripheral CD4^{+}CD25^{-} naïve Sag-reactive (Vβ3^{+}Vβ5^{+}Vβ11^{+}Vβ12^{+}) and Sag-negative (Vβ3^{-}Vβ5^{-}Vβ11^{-}Vβ12^{-}) T cells from LNs of WT or SKG mice. Right panel: histograms display surface marker expression in unstimulated cell subsets, quantified in **figure S9B**. Data represent 6 mice per genotype from 4 independent experiments. (**E**) Histograms display phospho-ERK (p-ERK) levels in Sag-reactive and Sag non-reactive CD4^{+}CD25^{-} T cells, gated on naïve markers (CD62L^{hi}CD44^{lo}) , from WT and SKG mice after TCR crosslinking for 2 min with α-CD3ε or stimulation with PMA. Data represent at least 4 mice per group from 3 independent experiments. (**F**) Left: FACS plots show IL-17^{+} cell frequencies in Sag-reactive and Sag non-reactive CD4 T cells after restimulation with PMA and ionomycin or vehicle control. Naive CD4^{+}CD25^{-} T cells

from SKG mouse LNs were cultured for 4 days in pathogenic or non-pathogenic Th17 conditions. Right: Quantification of mean IL-17$^+$ cell frequencies in Sag-reactive and Sag-negative CD4 T cell subsets (± SEM, n = 8 independent biological replicates per condition, each dot represents 2 pooled mice, experiment repeated 3 times). Significance indicated by asterisk [< 0.05 (*) or < 0.01 (**)] for *P* value from paired t-tests.

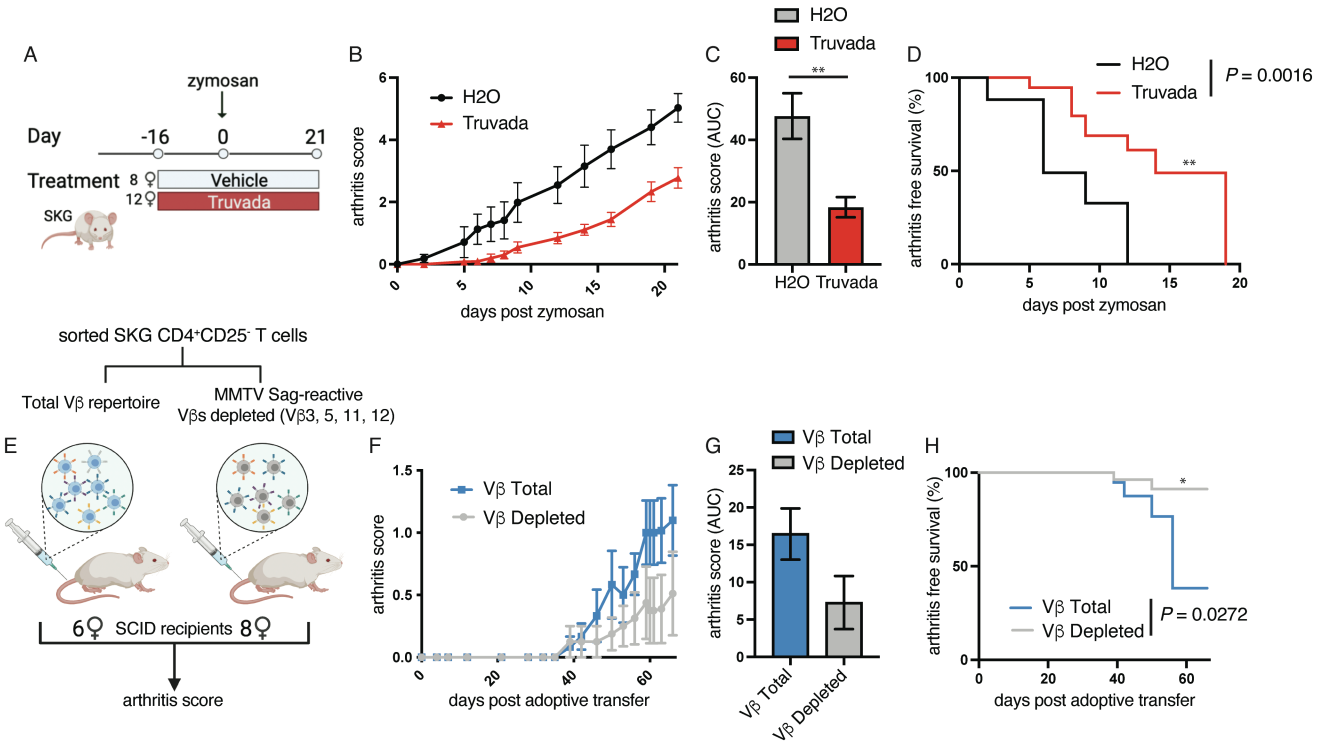


Figure 8. Arthritis pathogenicity partially localized to Sag-reactive SKG T cells. (**A**) Experimental set-up: SKG mice were treated with Truvada (n = 12) or vehicle control (n = 8) at day -16 prior to arthritis induction with intraperitoneal (i.p.) zymosan on day 0. (**B-C**) Arthritis score in SKG mice post zymosan injection in (**B**) and plotted as area under the curve (AUC) in (**C**). (**D**) Arthritis free survival plotted as Kaplan-Meier curve from (**A-B**), representative of 2 independent experiments. (**E**) Sorted SKG $CD4^{+}CD25^{-}$ T cells of the indicated Vβ T cell populations were adoptively transferred into SCID recipients and monitored for arthritis development. (**F-H**) Arthritis score in SCID mice after adoptive transfer in (**F**) and plotted as AUC (P = 0.08) in (**G**) and probability of arthritis free survival in (**H**), n = 6-8 mice per group, representative of 2 independent experiments**.** Significance indicated by asterisk [< 0.05 (*), < 0.01 (**), or < 0.001 (***)] for *P* value from unpaired t-test with Welch's correction (**C** and **G**), or log-rank Mantel-Cox test (**D** and **H**).

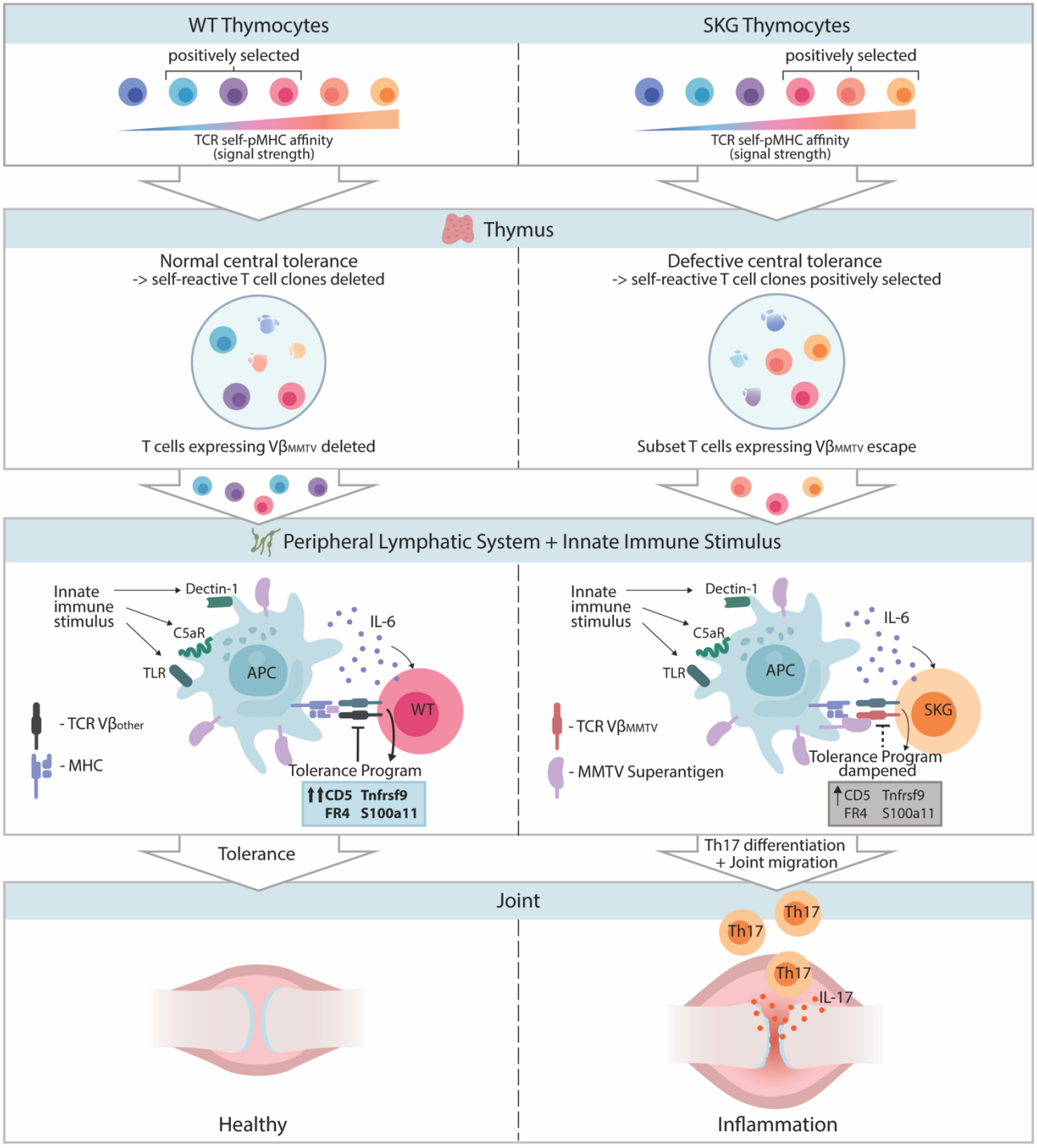


Figure 9. Sag-reactive SKG T cells evade central and peripheral tolerance and contribute to SKG arthritis. Impaired TCR signaling in SKG mice leads to a more self-reactive repertoire and escape of autoreactive, along with MMTV Sag-reactive, CD4 T cells into the periphery. Chronic encounter with peripheral antigens and innate immune stimuli activates these T cells (identified as GFP^{hi} cells) via their TCR. Due to impaired TCR signal transduction, SKG mice show reduced induction of TCR negative regulators and fail to fully establish a protective anergy state upon antigen encounter. Consequently, in the setting of certain environmental cues (e.g., IL-6 signaling), SKG T cells encountering endogenous antigens differentiate into pathogenic IL-17 producing effector T cells that cause erosive arthritis.