medRxiv preprint doi: https://doi.org/10.1101/2022.10.20.22281307; this version posted October 25, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

CD4$^+$ T cell subset imbalances in rheumatoid arthritis

An unsupervised clustering algorithm identifies rare T helper cell subset aberrations

Mulhearn B[1,4*], Marshall L[1*], Sutcliffe M[1], Hannes SK[5,6], Fonseka C[5,7], Hussell T[2], Raychaudhuri S[1,5], Barton A[1,3], Viatte S[1,3§]

[1]Versus Arthritis Centre for Genetics and Genomics, Centre for Musculoskeletal Research, Manchester Academic Health Science Centre, The University of Manchester, Manchester, Oxford Road, Manchester, M13 9PT, UK.

[2]Lydia Becker Institute of Immunology and Inflammation, Faculty of Biology, Medicine and Health, The University of Manchester, Manchester, M13 9PL, UK.

[3]NIHR Manchester Musculoskeletal Biomedical Research Centre, Central Manchester NHS Foundation Trust, Manchester Academic Health Sciences Centre, Grafton Street, Manchester, M13 9WL, UK.

[4]Department of Life Sciences, University of Bath, Claverton Down, Bath, BA2 7AY, UK.

[5]Division of Rheumatology, Inflammation, and Immunity, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA.

[6]BioNTech, Cell Therapy Unit (Molecular Immunology), Cambridge, MA, USA.

[7]eGenesis, Cambridge, MA, USA.

*Equal contribution

[§]Corresponding author: sebastien.viatte@manchester.ac.uk

medRxiv preprint doi: https://doi.org/10.1101/2022.10.20.22281307; this version posted October 25, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Abstract

Background: Despite the report of an imbalance between $CD4^+$ T helper (Th) cell subsets in rheumatoid arthritis (RA), patient stratification for precision medicine has been hindered by the discovery of ever more Th cell subsets, as well as contradictory association results.

Objectives: To capture previously reported Th imbalance in RA with deep immunophenotyping techniques; to compare hypothesis-free unsupervised automated clustering with hypothesis-driven conventional biaxial gating and explore if Th cell heterogeneity accounts for conflicting association results.

Methods: Unstimulated and stimulated peripheral blood mononuclear cells from RA patients and controls were immunophenotyped with a 37-marker panel by mass cytometry (chemokine receptors, intra-cellular cytokines, intra-nuclear transcription factors). First, conventional biaxial gating and standard definitions of Th cell subsets were applied to compare subset frequencies between cases and controls. Second, unsupervised clustering was performed with FlowSOM and analysed using mixed-effects modelling of Associations of Single Cells (MASC).

Results: Conventional analytical techniques fail to identify classical Th subset imbalance, while unsupervised automated clustering, by allowing for unusual marker combinations, identified an imbalance between pro- and anti-inflammatory subsets. For example, a pro-inflammatory Th1-like (IL-2^+ T-bet^+) subset and an unconventional but pro-inflammatory IL-17^+ T-bet^+ subset were significantly enriched in RA (odds ratio=5.7, $p=2.2 \times 10^{-3}$; odds ratio=9.7, $p=1.5\times10^{-3}$, respectively). In contrast, a $FoxP3^+$ IL-2^+ HLA-DR^+ Treg subset was reduced in RA (odds ratio=0.1, $p=7.7\times10^{-7}$).

Conclusions: Innovative approaches capture known $CD4^+$ T cell subset imbalances in RA blood.

Keywords: precision medicine, rheumatoid arthritis, $CD4^+$ T cells, mass cytometry, heterogeneity

medRxiv preprint doi: https://doi.org/10.1101/2022.10.20.22281307; this version posted October 25, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Introduction

Current treatment strategies for rheumatoid arthritis (RA) standardise treatments across patient groups, but as a heterogeneous disease, differences will exist in the underlying immune mechanisms. As a consequence, not all patients will respond similarly to the same drug with only 60-70% having a response to any biologic drug [1]. Given that an increase of pro-inflammatory and/or a decrease of anti-inflammatory $CD4^+$ T cell subsets has been reported in some patients, these cell types could represent therapeutic targets and aid patient stratification for precision medicine. However, which $CD4^+$ T cell subset is associated with the disease, and in what proportion of patients, remains unclear.

The study of T cells in RA was revolutionised after the Th1/Th2 paradigm was proposed by Mossman *et al.* in 1986 [2]. This paradigm was described as a dichotomy between type 1 T helper cells (Th1), characterised by interferon-γ (IFN-γ) and tumour necrosis factor (TNF), and type 2 T helper cells (Th2). RA pathogenesis predominantly involves Th1 cytokines, and the Th1 immune response is antagonised by Th2 cytokines (reviewed in [3]). The Th1/Th2 paradigm was modified further when pathogenic IL-17-producing T helper cells (Th17) and regulatory T cells (Tregs) were discovered to play critical roles in initiating and regulating autoimmunity, respectively (reviewed in [4]). Increases in Th1 and Th17 cells which antagonise, and are antagonised by, Th2 and Treg cells has been the focus of much research in autoimmune diseases including RA [5,6]. More recently, interferon-producing Th1 memory cells were found to be associated with RA using novel techniques of analysing high-dimensional single-cell data [7]. A meta-analysis found that circulating Treg cells, as defined by both FOXP3 and CD25, were decreased in RA patients in 9 combined studies [8]. However, other studies found no decrease in Treg frequency or function in RA [9]. To add to the complexity of the paradigm, there are an ever-growing list of T helper cells (Th9 cells [10]; Th22 cells [11]; T follicular helper cells (Tfh) [12]; peripheral helper T cells (Tph) [13]). Critically, it has now been shown that many of these T cell types exhibit plasticity. It was previously thought that the mutually exclusive expression of a master transcription factor determined the fate of T cells, with T-bet, GATA3, RORγt and FoxP3 determining the fate of Th1, Th2, Th17 and Treg cells respectively. However, co-expression of these transcription factors may temporarily alter the effector function of T cell subsets

medRxiv preprint doi: https://doi.org/10.1101/2022.10.20.22281307; this version posted October 25, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

(reviewed in [14]) and many of these T cells might also be transitional and therefore appear only in very low frequencies.

The plasticity of T cell subsets and heterogeneity of RA may also explain why clinical trials targeting specific cytokines, for instance IL-17, have not reached their primary endpoint [15]. Stratified medicine may allow clinicians to identify RA patients who have predominantly IL-17-driven disease to improve the design of such trials, as researchers have successfully shown in psoriatic arthritis [16]. Exploring peripheral blood immunophenotypes by flow cytometry in RA has been shown to mirror findings in the synovial compartment [17] and still represents the most widely used and cost effective technique to enumerate immune cell populations. The advent of mass cytometry (CyTOF) has vastly increased the number of cellular markers which can be detected simultaneously [18] and together with the development of high-throughput automated methods of data analysis [19–23], rare and novel cell types involved in the aetiology of RA have been recently identified [7,13,24,25]. Automated cell clustering algorithms allow for unbiased marker combinations and therefore the hypothesis-free discovery of unanticipated cell subsets, relevant to disease and precision medicine [7,22,26,27]. We postulate that Th cell plasticity and overlap between subsets and definitions explain conflicting associations and lack of reproducibility in small sample sizes. Therefore, we developed a 37 marker T cell mass cytometry panel to encompass most definitions for the most studied $CD4^+$ T cell subsets to date (Th1, Th2, Th17 and Treg).

The aims of this study are first, to confirm the T helper cell subset imbalance previously described in RA; second, to explore whether this imbalance is detectable in a small sample size, if the standard definition of Th subsets is relaxed in favour of Th cell plasticity (unbiased marker combinations) and third, to test whether innovative techniques (mass cytometry and unsupervised clustering algorithm) facilitate the identification of pro- and anti-inflammatory $CD4^+$ T cell subsets over standard techniques (flow cytometry and manual bi-axial gating) in a small sample size.

medRxiv preprint doi: https://doi.org/10.1101/2022.10.20.22281307; this version posted October 25, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Methods

Patient and public involvement

The Versus Arthritis Centre for Genetics and Genomics in Manchester has a Research User Group (RUG) of patients with various rheumatologic conditions, which includes RA. Members of the RUG meet regularly to review the research carried out in the Centre. They highlighted the importance of understanding basic mechanisms of disease from a patient's perspective. The RUG supports research of basic disease mechanisms to identify biomarkers for stratification into treatment response categories. Some of the following comments were made: "the sooner anyone with RA can get on the correct treatment, the better"; "Personalising from onset would be perfect"; "If only a blood test is needed, then even better".

Patient selection and sample collection

A cohort of 10 RA patients stabilised on therapy and 10 healthy volunteers were recruited from the National Repository (North West Ethics committee approval MREC 99/8/84) at the Manchester Royal Infirmary (NIHR portfolio ID 7881). Peripheral Blood Mononuclear Cells (PBMC) were isolated from 18 ml of blood by density gradient centrifugation using Ficoll-Paque plus (GE Healthcare Life Sciences) and cryopreserved at -150°C.

Mass cytometry antibody panel

We adapted a previously published mass cytometry T cell panel [13] to encompass surface (chemokine receptors), intra-cytoplasmic (cytokines) and intra-nuclear markers (transcription factors) used to define Th and Treg subsets (Table 1).

Target	Epitope location	Marker significance	Label	Clone
CD8a	Extracellular	Cytotoxic T cell	141Pr	RPA-T8
HLA-DR	Extracellular	Activation marker	143Nd	L243
IL-2	Intracellular	T cell cytokine	144Nd	MQ1-17H12

medRxiv preprint doi: https://doi.org/10.1101/2022.10.20.22281307; this version posted October 25, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

CD4	Extracellular	T helper cell	145Nd	RPA-T4
CD127 (IL-7Ra)	Extracellular	Cytokine receptor	147Sm	A019D5
CD278/ICOS	Extracellular	Checkpoint protein	148Nd	C398.4A
CD200	Extracellular	Myeloid cell regulator	149Sm	OX-104
IL-22	Intracellular	Th2 cytokine	150Nd	22URTI
IL-5	Intracellular	Th2 cytokine	151Eu	TRFK5
TNF	Extracellular	Inflammatory cytokine	152Sm	Mab11
CD62L (L-selectin)	Extracellular	Adhesion molecule	153Eu	DREG-56
CD38	Extracellular	Lymphocyte activator	154Sm	HIT2
CD279 (PD-1)	Extracellular	T cell suppressor	155Gd	EH12.2H7
CD134 (OX40)	Extracellular	Co-stimulatory molecule	156Gd	ACT35
IFNγ	Intracellular	Th1 cytokine	158Gd	B27
FoxP3	Intracellular	Treg marker	159Tb	259D/C7
Tbet	Intracellular	Th1 transcription factor	160Gd	4B10
CD197 (CCR7)	Extracellular	Chemokine receptor involved in lymph node homing	161Dy	G043H7
CD152 (CTLA-4)	Extracellular	Checkpoint protein	162Dy	14D3
CD183 (CXCR3)	Extracellular	Th1 chemokine receptor	163Dy	G025H7
IL-17A	Intracellular	Th17 cytokine	164Dy	N49-653
CD45RO	Extracellular	Memory T cell marker	165Ho	UCHL1
IL-10	Intracellular	Regulatory cytokine	166Er	JES3-9D7
Gata3	Intracellular	Th2 transcription factor	167Er	TWAJ
ROR gamma	Intracellular	Th17 transcription factor	168Er	600214
CD25 (IL-2R)	Extracellular	IL-2 receptor	169Tm	2A3
CD28	Extracellular	Costimulatory molecule	170Er	CD28.2
CD185 (CXCR5)	Extracellular	T follicular helper cell	171Yb	RF8B2

medRxiv preprint doi: https://doi.org/10.1101/2022.10.20.22281307; this version posted October 25, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

		chemokine receptor		
IL-21	Intracellular	Regulatory cytokine	172Yb	3A3-N2
IL-4	Intracellular	Th2 cytokine	173Yb	MP4-25D2
CD154 (CD40L)	Extracellular	Activated T cell marker	174Yb	24-31
Perforin	Intracellular	Cytotoxic T cell effector protein	175Lu	B-D48
CD196 (CCR6)	Extracellular	Th17 chemokine receptor	176Yb	11A9
Iridium – CellID		DNA binder to identify cell events	191Ir	
Iridium – CellID		DNA binder to identify cell events	193Ir	
Cisplatin		DNA binder to assess cell viability	194Pt	
CD3	Extracellular	Pan-T cell marker	209Bi	UCHT1

Table 1. The antigen target, heavy metal and clone of each conjugated antibody used in the CyTOF T cell panel. Target location indicates if the antibody was used in the extracellular or intracellular antibody cocktail. All antibodies were purchased from Fluidigm.

T cell enrichment and stimulation

Thawed PBMCs were rested at 37°C for one hour prior to enrichment of $CD3^+$ T cells by positive selection using magnetic cell separation (MACS) with CD3 MicroBeads (Miltenyi Biotec). T cell receptor stimulation was achieved with Dynabeads Human T-Activator CD3/CD28 beads (Fisher Scientific UK Ltd) at a concentration of 1 bead per 2 cells, in the presence of brefeldin A and monensin (both Fisher Scientific UK Ltd) and incubated for 4.5 hours at 37 °C.

Mass cytometry staining protocol

$CD3^+$ T cells were incubated with cisplatin prior to incubation with an extracellular antibody cocktail (detailed protocol in Supplementary Methods). Subsequently, cells were fixed,

medRxiv preprint doi: https://doi.org/10.1101/2022.10.20.22281307; this version posted October 25, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

permeabilised and stained for intracellular antigens for 30 minutes on ice. A solution of Intercalator-Ir was added to each well. Samples were run on Helios mass cytometers by the Longwood Medical Area CyTOF Core Facility at the Dana-Farber Cancer Institute, Boston, USA.

Data pre-processing

The Nolan Normalizer MATLAB plugin [28] was used to normalise the signal in each channel to the signals from the Maxpar Calibration beads.

Traditional gating analysis

Traditional biaxial gating was carried out using FlowJo V.10.8 (BD Biosciences) to identify well-described $CD4^+$ Th and Treg subsets using standard definitions [29] (example gating strategy in supplementary Fig. 1). The percentages of positive cells are represented as a proportion of the total T cell population ($CD3^+$). Geometric mean intensity is used to quantify cytokine expression on an individual cell level. The Mann-Whitney *U* test was used to compare RA and Healthy Controls (HC). All p-values are unadjusted (tests are not independent).

Automated clustering workflow

The CyTOF clustering workflow from Nowicka *et al*. [20] was modified to include an extended quality control approach and a different statistical framework for association testing (see below). All plots were produced with *ggplot2* (v3.3.5) [30], unless stated otherwise.

Quality control steps

Normalised cytometric data was manually inspected using FlowJo V.10.8 (BD Biosciences) to ensure that at least 10 cell events were identified in a cluster in at least 3 samples. IL-5 was excluded due to the absence of any positive cell clusters (supplementary figure 2). Furthermore, only samples with at least 1000 live single T cells were included in subsequent analyses. Cell events with extreme expression were excluded per individual marker. After the removal of unsuccessful markers and extreme events, analysis was performed in R (v4.1.0) on data transformed with the inverse hyperbolic sine (arcsinh) function (cofactor = 5). Finally, to

medRxiv preprint doi: https://doi.org/10.1101/2022.10.20.22281307; this version posted October 25, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

identify outlying samples and potential batch effects, multidimensional scaling (MDS) and principal component analysis (PCA) plots were produced from median expression of panel markers in each sample.

Clustering with FlowSOM/ConsensusClusterPlus

The dataset was downsampled to an equal number of randomly selected cells (n) from each sample, where n was equal to the number of cells in the smallest sample. The automated clustering steps were performed with *FlowSOM* (v2.0.0) [19] and *ConsensusClusterPlus* (v1.56.0) [23] using agglomerative hierarchical consensus clustering. Cells were clustered into 20 populations using Euclidean distance and average linkage metrics (supplementary figure 3). Heatmaps of median marker expression for each cluster were generated with *pheatmap* (v1.0.12) [31]. For visualisation, these values were scaled between 0 and 1, with the 0.01 and 0.99 quantiles for each marker as the lower and upper boundary, respectively.

Cluster visualisation with t-SNE

The similarity of single cells in two-dimensional space was visualised with the dimensionality reduction technique t-stochastic neighbour embedding (t-SNE), implemented with *Rtsne* (v0.15) [32]. Clusters were annotated manually by their surface phenotype.

Statistical analysis

To identify clusters with differential representation in RA samples compared to healthy controls, we used mixed-effects modelling of associations of single cells (MASC) [7] for each cluster, including age and sex of donors as fixed-effects covariates, sample ID as a random-effect covariate and the case-control status as the contrast term. Clusters that are significantly enriched or depleted in RA are defined as those with a P value ≤ 0.05 after Bonferroni correction (n = number of clusters tested).

medRxiv preprint doi: https://doi.org/10.1101/2022.10.20.22281307; this version posted October 25, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Data sharing

All mass cytometry data is freely available at https://flowrepository.org/ (repository ID FR-FCM-Z5RM) [46].

medRxiv preprint doi: https://doi.org/10.1101/2022.10.20.22281307; this version posted October 25, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Results

Study design and subject characteristics

PBMCs from 10 RA patients and 10 healthy controls (Table 2) were left unstimulated or stimulated with anti-CD3/CD28 beads (*n* = 40 samples) prior to deep immunophenotyping by mass cytometry.

	Cases (*n* = 10)	**Controls (*n* = 10)**	***P* value**
Age (mean ± SD)	61 ± 14	46 ± 8	0.02
Female	7	6	1
RF-positive	7	—	—
ACPA-positive	6	—	—
DAS28 (mean ± SD)	3.4 ± 2.2	—	—
Glucocorticoids	2	0	—
Methotrexate	5	0	—
Other DMARD*	3	0	—
Biologics	1	0	—

*Sulfasalazine or hydroxychloroquine

Table 2. Subject characteristics of cases (RA) and healthy controls. *P* values calculated for age and gender by Mann-Whitney U and Fisher's exact test, respectively. RF: rheumatoid factor. ACPA: anti-citrullinated protein antibody. DAS: disease activity score. DMARD: disease-modifying anti-rheumatic drug.

Manual gating with traditional statistical analysis fails to identify large differences between blood from patients with RA and healthy controls.

Sequential biaxial gating of CyTOF data was used to identify commonly described $CD4^+$ and $CD8^+$ T cell subsets as defined by expression of surface markers, transcription factors and/or intracellular cytokines. Conventional statistical analysis with Mann-Whitney *U* found no significant differences in the proportions of Th $CD4^+$ subsets between RA and HC (figure 1); only

medRxiv preprint doi: https://doi.org/10.1101/2022.10.20.22281307; this version posted October 25, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Treg (defined as $CD4^+$ $FoxP3^+$) were markedly decreased in RA. No $CD8^+$ T cell subsets were found to be differentially abundant in RA (figure 2).

CyTOF automated clustering pipeline quality control steps

Low quality cells with extreme expression values were excluded after individual appraisal of expression distributions for each marker. This step removed 0.34% (4,146/1,230,364) of cells from the full dataset, with between 0.09% and 1.4% of cells removed from each sample. Diagnostic MDS and PCA plots highlight notable separation of RA and HC in the second dimension, indicating preliminary differences in marker expression between the two groups (figure 3A). On the same PCA (figure 3B), samples largely cluster by age group (figure 3C) similarly to previous reports [49], suggesting that the structure in the data is attributable to participant age rather than technical batch effects. Therefore, we included age as a co-variate for statistical association testing.

Automated clustering identifies well-described T cell subsets

Automated clustering was performed on downsampled quality-controlled CyTOF data; analyses of unstimulated and stimulated T cell datasets were performed separately. Within the unstimulated dataset, we identified 20 distinct T cell subsets, including 11 $CD4^+$ and nine $CD8^+$ (figure 4A). Amongst the $CD4^+$ populations are a $FoxP3^+$ subset (subset 11; a Treg subset), two IL-17^+ subsets (subsets 16 and 20; Th17 subsets), a Th1-like subset (subset 19) and a $CD38^+$ subset (subset 18).

Although analysis was performed independently, clusters identified in the stimulated dataset were broadly similar to those of the unstimulated dataset (figure 4B). As expected, t-SNE plots coloured by expression level of several markers highlight an increase in the number of cells expressing TNF upon stimulation (figures 4B and 4C).

MASC identifies multiple T cell clusters with differential expression between RA and healthy samples

medRxiv preprint doi: https://doi.org/10.1101/2022.10.20.22281307; this version posted October 25, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

MASC was used to identify clusters with differential abundance between RA samples and HC, adjusting for the effects of age, gender, and donor. Six unstimulated cell subsets and 8 stimulated subsets were found to be differentially abundant with MASC (figure 5). The majority were $CD4^+$ and, of these, three populations were present in both unstimulated and stimulated states, including enrichment of Th1 and Th17 subsets ($CD4^+$ T-bet$^+$ IL-17$^+$ and $CD4^+$ T-bet$^+$ IL-2$^+$ cell clusters) and depletion of a Treg subset ($CD4^+$ FoxP3$^+$ IL-2$^+$) in RA (figure 6). Of note, stimulation further enriched the $CD4^+$ T-bet$^+$ IL-2$^+$ cluster in RA samples relative to healthy controls ($P = 2.2 \times 10^{-3}$, OR = 5.7, figure 7). Interestingly, a Th2 subset ($CD4^+$ T_{CM} GATA3$^+$) was decreased in RA (although it did not reach statistical significance).

Notably, figures 6 and 7 highlight the heterogeneity in subset proportions between individual patients. For example, the abundance of the $CD4^+$ TNF$^+$ cells range from < 1% to > 8% of cells in stimulated RA samples (figure 7, subset 14). Naïve T cells were increased in frequency in controls compared to cases, but this difference disappeared after statistical adjustment for age implemented in MASC, demonstrating that the correction for the age-driven stratification observed in the PCA plots was important in terms of statistical outcome.

Cell cluster phenotype, abundances, odds ratios, and *P* values for all cell clusters generated by the pipeline are presented in table 3.

medRxiv preprint doi: https://doi.org/10.1101/2022.10.20.22281307; this version posted October 25, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Subset	Number	% total RA cells	% total HC cells	OR (RA vs HC)	*P* value
A: Shared (Unstimulated, stimulated)					
CD4$^+$ T$_{CM}$ T-bet$^+$ IL-17$^+$ (combined Th1 and Th17 phenotype)	20	2.6	0.4	12.2 (5.3 – 28.1)	**1.5 x 10^{-4}**
	16	1.4	0.2	9.7 (3.8 – 24.5)	**1.5 x 10^{-3}**
CD4$^+$ T$_{CM}$ FoxP3$^+$ IL-2$^+$ (Treg)	11	0.5	1.8	0.3 (0.2 – 0.5)	**4.6 x 10^{-3}**
	15 + 17	0.4 / 0.09	2.0 / 0.7	0.1 (0.1 – 0.2) / 0.03 (0.01 – 0.09)	**7.7 x 10^{-7} / 1.9 x 10^{-7}**
CD4$^+$ T$_{EM}$ CCR6$^+$ (Th17)	7	1.6	0.6	4.4 (2.2 – 8.7)	**5.1 x 10^{-3}**
	7	4.8	1.1	9.4 (4.6 – 19.5)	**7.9 x 10^{-5}**
CD4$^+$ T$_{CM}$ T-bet$^+$ IL-2$^+$ (Th1)	19	0.8	0.4	3.5 (1.6 – 7.4)	6.8 x 10^{-2}
	13	0.8	0.2	5.7 (2.6 – 12.3)	**2.2 x 10^{-3}**
CD8$^+$ T$_{EM}$	3	15.9	10.5	1.7 (1.1 – 2.7)	3.4 x 10^{-1}
	1	6.8	9.8	0.6 (0.4 – 0.9)	5.4 x 10^{-1}
CD4$^+$ T$_{CM}$	2	22.8	14.0	1.5 (0.9 – 2.7)	1
	12	22.5	20.6	0.9 (0.5 – 1.5)	1
CD8$^+$ T$_{EM}$ T-bet$^+$ Perforin$^+$	8 + 9	12.2 / 3.7	9.0 / 4.8	1.0 (0.5 – 1.9) / 0.8 (0.4 – 1.5)	1
	2	12.7	9.9	0.9 (0.5 – 1.9)	1
CD4$^+$ T$_{EM}$ TNFα$^+$ (Th1)	13	1.0	0.9	1.1 (0.6 – 1.9)	1
	14	3.8	1.9	1.9 (0.8 – 4.7)	1
CD4$^+$ Naïve	15	7.0	15.4	0.6 (0.3 – 1.1)	1
	10	11.8	17.4	0.8 (0.4 – 1.4)	1
B: Unstimulated					
CD4$^+$ T$_{CM}$ IL-17$^+$ (Th17)	16	0.9	0.2	5.7 (3.1 – 10.3)	**4.0 x 10^{-4}**
CD8$^+$ T$_{EM}$ TNFα$^+$	14	1.5	0.6	3.7 (2.0 – 7.1)	**1.4 x 10^{-2}**

medRxiv preprint doi: https://doi.org/10.1101/2022.10.20.22281307; this version posted October 25, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

CD8$^+$ T$_{CM}$ FoxP3$^+$ IL-2$^+$	12	0.5	0.5	0.2 (0.1 – 0.5)	**2.5 x 10^{-2}**
CD8$^+$ T$_{CM}$ PD-1$^+$	4	2.4	6.2	0.4 (0.2 – 0.7)	6.5 x 10^{-2}
CD8$^+$ T$_{EM}$ CCR6$^+$ Perforin$^+$	17	0.8	0.3	3.5 (1.6 – 7.6)	7.3 x 10^{-2}
CD8$^+$ T$_{CM}$ Perforin$^+$	6	0.9	0.2	3.8 (1.6 – 8.7)	9.4 x 10^{-2}
CD4$^+$ T$_{CM}$ GATA3$^+$	1	15.0	20.3	0.6 (0.3 – 1.2)	1
CD4$^+$ T$_{CM}$ Perforin$^+$	5	6.9	7.9	0.7 (0.4 – 1.2)	1
CD8$^+$ Naïve	10	2.2	5.3	0.8 (0.4 – 1.5)	1
CD4$^+$ T$_{CM}$ CD38$^+$	18	0.8	0.7	1.6 (1.0 – 2.7)	1
C: Stimulated					
CD8$^+$ T$_{EM}$ CD38$^+$ T-bet$^+$	4	1.0	0.4	6.3 (2.3 – 17.0)	**1.8 x 10^{-2}**
CD8$^+$ T$_{EM}$ CCR6$^+$ T-bet$^+$	5	0.9	0.4	3.8 (1.8 – 7.9)	**3.8 x 10^{-2}**
CD4$^+$ T$_{CM}$ TNFα$^+$	19	2.2	0.8	4.0 (1.9 – 8.5)	**4.1 x 10^{-2}**
CD4$^+$ T$_{CM}$ CD38$^+$ TNFα$^+$	18	3.9	9.0	0.3 (0.1 – 0.8)	4.1 x 10^{-1}
CD4$^+$ T$_{CM}$ T-bet$^+$ Perforin$^+$	8	0.2	0.6	0.3 (0.1 – 0.8)	4.1 x 10^{-1}
CD8$^+$ T$_{CM}$	3	11.7	10.2	1.4 (0.9 – 2.0)	1
CD8$^+$ T$_{EM}$ T-bet$^+$ Perforin$^+$	6	3.3	4.5	0.9 (0.4 – 2.0)	1
CD4$^+$ T$_{EM}$	9	9.1	7.6	1.1 (0.5 – 2.4)	1
CD8$^+$ T$_{EM}$ IFNγ$^+$ TNFα$^+$	11	2.3	2.0	0.9 (0.4 – 2.3)	1
CD4$^+$ T$_{CM}$ CCR6$^+$	20	0.5	0.8	0.7 (0.4 – 1.5)	1

Table 3. MASC results for clusters present in both unstimulated and stimulated samples (A), unstimulated samples alone (B) and stimulated samples alone (C). For shared cell subsets, the first row indicates statistics for unstimulated subsets and the second for stimulated subsets. Elements separated by "/" indicates statistics for two identical subsets found in the unstimulated and stimulated datasets. The odds ratio is shown with a 95% confidence interval. Differentially abundant subsets are highlighted by a bold *P* value (α = 0.05). T$_{CM}$: T central memory. T$_{EM}$: T effector memory.

medRxiv preprint doi: https://doi.org/10.1101/2022.10.20.22281307; this version posted October 25, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Agnostically defined cell clusters can be identified using conventional techniques

To provide internal validation of the automated pipeline's significant clusters, manual gating and standard statistical analysis of agnostically defined clusters was performed. Based on the phenotypes presented in figure 4 and the level of expression of each marker, gates were drawn by eye around all positive cells if the cluster had low expression of a marker on the heatmap, but only around high positive cells if the marker had high expression.

Using these definitions of Th and Treg subsets and a standard statistical approach, manual gating was able to identify a clear imbalance between pro-inflammatory Th subsets and anti-inflammatory subsets in RA. For example, out of the unstimulated dataset, manual gating was able to identify that subset 11 ($CD4^+$ Tcm $FoxP3^+$ IL-2^+), a Treg subset, is decreased in RA (P = 0.0004). The same population could also be identified in the stimulated dataset (subset 17, $CD4^+$ Tcm $FoxP3^+$ IL-2^+, P = 0.0002). Pro-inflammatory Th cells were also clearly identified and statistically significantly increased in RA ($CD4^+$ Tcm $Tbet^+$ IL-17^+, subset 20, P = 0.0172; and subset 16, P = 0.0022; figure 8). Therefore, innovative approaches easily capture known $CD4^+$ T cell subset imbalances in RA blood, even in a small sample size, by allowing for unconventional marker combinations.

medRxiv preprint doi: https://doi.org/10.1101/2022.10.20.22281307; this version posted October 25, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Discussion

We show the superiority of innovative immunophenotyping and automated analytical techniques over hypothesis-driven conventional biaxial gating (based on canonical biomarkers) to identify a Th1, Th17/Th2, Treg imbalance in RA peripheral blood in small sample sizes. Our strategy allows for unusual combinations of T cell markers, for instance T-bet and IL-17, capturing the plasticity of T helper cell populations, a more difficult task to achieve with standard gating strategies.

Th17 cells are known to be phenotypically unstable ([33]). In our study, T-bet$^+$ IL-17$^+$ T cells are increased in RA blood. These cells have been termed Th1-like Th17 cells in the literature, co-produce IL-17 and IFNγ, and are known to contribute to inflammation in the context of autoimmunity [33–35]. Interestingly, IFNγ production in Th1-like Th17 cells has been shown to be repressed by miR-146a [35] and loss of function polymorphisms in the microRNA-146a (miR-146a) gene have been associated with an increased risk of developing RA and lupus in both European and Asian genetic association studies [36,37]; taken together, these data provide some mechanistic evidence that Th1-like Th17 cells are involved in the pathogenesis of RA. The methods outlined in this paper easily detect an increase of this subset in RA.

We also found a significant decrease in the number of IL-2$^+$ FoxP3 T cells (Tregs) in RA samples compared to controls. It is well-established that IL-2 is critical in maintaining Treg function [38] and that some Treg produce IL-2 [39]. However, the relative role of paracrine versus autocrine IL-2 secretion from nearby effector T cells [40] versus IL-2$^+$ FoxP3 T cells [39] is unclear. Our data suggests that IL-2-producing Tregs rather than all Treg are playing an important role in RA. Again, our methods of automated analysis were able to easily identify this non-canonical cell type, without any previous hypothesis on its association.

Due to our small sample size, our association results could theoretically represent false positive due to a sampling bias or caused by multiple testing. However, this is very unlikely, as first we confirm a plausible and previously reported imbalance of Th/Treg subsets in RA; second the mechanistic evidence supports our findings, and finally all statistical tests have been stringently corrected for multiple testing. These elements reinforce the credibility of our main conclusion:

medRxiv preprint doi: https://doi.org/10.1101/2022.10.20.22281307; this version posted October 25, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

the superiority of hypothesis-free clustering of many markers in small sample sizes over standard hypothesis-driven bi-axial gating.

The hallmark of precision medicine and the primary aim of stratified medicine initiatives in RA [41] consists of stratifying patients into treatment response categories. Our study shows large heterogeneity across patients with some with a predominantly Th1 profile, others a Th17 profile, or a Th1-like Th17 profile, while others are mainly characterised by a low IL-2^{+} Treg profile. As an example, stratifying patients by their IL-2^{+} Treg titres may dictate response to low-dose IL-2 therapy, as has now been successfully trialled in lupus [42] and RA [43]. Furthermore, a Japanese group found that stratified medicine for psoriatic arthritis patients increased response rates significantly [16]. Immunophenotyped patients received ustekinumab if Th1-dominant and secukinumab if either Th17-dominant or with a Th1/Th17-hybrid phenotype resulting in clinically significant improvement in response rates [16]. The same group have good one year follow-up data [44], suggesting that the immunophenotype could be the main predictor of drug response in precision medicine. Stratifying RA patients by their Th1/Th17 phenotype, or by the Th1-like Th17 phenotype highlighted in this study, may suggest that despite the fact that previous trials of secukinumab in RA failed to reach their primary endpoint, it could be a viable treatment option in RA patient subgroups; further work is required to confirm this hypothesis. [45].

We have shown that non-biased automated analysis of large immune datasets is successful in identifying Th cell imbalance in RA with implications for precision medicine. We now plan to expand on these findings by testing their role in patient stratification for treatment response studies.

medRxiv preprint doi: https://doi.org/10.1101/2022.10.20.22281307; this version posted October 25, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Key messages

What is already known about this subject?

- Large cytometric datasets are difficult to analyse manually due to the multidimensional nature of the data.
- Manual data analysis may introduce significant bias and be underpowered to interrogate multidimensional data.

What does this study add?

- This study has shown that an unbiased automated clustering algorithm can successfully interrogate large cytometric datasets, finding differential expression of 2 rare T cell populations in RA patients (decreased IL-2^{+} Treg cells; increased Th1-like Th17 cells).

How might this impact on clinical practice?

- Understanding the underlying immunopathology of autoimmune disease is a prerequisite to finding new drug targets for treatment.
- As our knowledge of T cell heterogeneity expands, discovering rare T cell populations will directly feed into the concept of precision medicine, to: 'treat arthritis, right first time'.

Competing interests

None of the authors have or had competing interests for the duration of this study or during manuscript preparation.

Contributorship

SR, RS and AB conceived the project, guided it through its stages of development, ensured progress, and contributed to the final manuscript. BM ran the laboratory experiments with help from SKH. MS pre-processed the raw mass cytometry data. BM and LM ran the data analysis

medRxiv preprint doi: https://doi.org/10.1101/2022.10.20.22281307; this version posted October 25, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

and drafted the manuscript. CF supported the automated pipeline analysis and provided R code. TH and SR provided immunology expertise. All authors contributed and proof read the final version of the manuscript.

Acknowledgments

We would like to thank Sarah Ashton (senior study coordinator) for her invaluable support, all the staff at the University of Manchester Versus Arthritis Centre for Genetics and Genomics and at the Kellgren Centre for Rheumatology, Central Manchester University NHS Foundation Trust, UK. Most importantly, we are grateful and thankful for all the patients and volunteers who took part in this study.

Funding and grant awards

This research was supported by Versus Arthritis (grant 21754), the NIHR Manchester Biomedical Research Centre, and the Wellcome Trust Institutional Strategic Support Fund (WT ISSF) in Precision Medicine and Single Cell Research. AB is an NIHR Senior Investigator. BM is an NIHR-funded Academic Clinical Fellow.

medRxiv preprint doi: https://doi.org/10.1101/2022.10.20.22281307; this version posted October 25, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

References

1 Hyrich KL, Watson KD, Silman AJ, *et al.* Predictors of response to anti-TNF-α therapy among patients with rheumatoid arthritis: Results from the British Society for Rheumatology Biologics Register. *Rheumatology* 2006;**45**. doi:10.1093/rheumatology/kel149

2 Mosmann TR, Cherwinski H, Bond MW, *et al.* Two types of murine helper T cell clone. I. Definition according to profiles of lymphokine activities and secreted proteins. *J Immunol* 1986;**136**.

3 Schulze-Koops H, Kalden JR. The balance of Th1/Th2 cytokines in rheumatoid arthritis. *Best Pract Res Clin Rheumatol* 2001;**15**:677–91. doi:10.1053/berh.2001.0187

4 Afzali B, Lombardi G, Lechler RI, *et al.* The role of T helper 17 (Th17) and regulatory T cells (Treg) in human organ transplantation and autoimmune disease. *Clin Exp Immunol* 2007;**148**:32–46. doi:10.1111/j.1365-2249.2007.03356.x

5 Niu Q, Cai B, Huang Z, *et al.* Disturbed Th17/Treg balance in patients with rheumatoid arthritis. *Rheumatol Int* 2012;**32**:2731–6. doi:10.1007/s00296-011-1984-x

6 van Hamburg JP, Asmawidjaja PS, Davelaar N, *et al.* Th17 cells, but not Th1 cells, from patients with early rheumatoid arthritis are potent inducers of matrix metalloproteinases and proinflammatory cytokines upon synovial fibroblast interaction, including autocrine interleukin-17A production. *Arthritis Rheum* 2011;**63**:73–83. doi:10.1002/art.30093

7 Fonseka CY, Rao DA, Teslovich NC, *et al.* Mixed-effects association of single cells identifies an expanded effector CD4+ T cell subset in rheumatoid arthritis. *Sci Transl Med* 2018;**10**. doi:10.1126/scitranslmed.aaq0305

8 Morita T, Shima Y, Wing JB, *et al.* The Proportion of Regulatory T Cells in Patients with Rheumatoid Arthritis: A Meta-Analysis. *PLoS One* 2016;**11**:e0162306. doi:10.1371/journal.pone.0162306

9 Walter GJ, Fleskens V, Frederiksen KS, *et al.* Phenotypic, Functional, and Gene Expression Profiling of Peripheral CD45RA+ and CD45RO+ CD4+CD25+CD127 ^{low} Treg Cells in Patients With Chronic Rheumatoid Arthritis. *Arthritis Rheumatol* 2016;**68**:103–16. doi:10.1002/art.39408

10 Ciccia F, Guggino G, Rizzo A, *et al.* Potential involvement of IL-9 and Th9 cells in the pathogenesis of rheumatoid arthritis. *Rheumatol (United Kingdom)* 2015;**54**. doi:10.1093/rheumatology/kev252

11 Miyazaki Y, Nakayamada S, Kubo S, *et al.* Th22 Cells Promote Osteoclast Differentiation via Production of IL-22 in Rheumatoid Arthritis. *Front Immunol* 2018;**9**. doi:10.3389/fimmu.2018.02901

12 Deng J, Fan C, Gao X, *et al.* Signal transducer and activator of transcription 3 hyperactivation associates with follicular helper T cell differentiation and disease activity in rheumatoid arthritis. *Front Immunol* 2018;**9**. doi:10.3389/fimmu.2018.01226

medRxiv preprint doi: https://doi.org/10.1101/2022.10.20.22281307; this version posted October 25, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

13 Rao DA, Gurish MF, Marshall JL, *et al.* Pathologically expanded peripheral T helper cell subset drives B cells in rheumatoid arthritis. *Nature* 2017;**542**:110–4. doi:10.1038/nature20810

14 Mirlekar B. Co-expression of master transcription factors determines CD4+ T cell plasticity and functions in auto-inflammatory diseases. Immunol. Lett. 2020;**222**. doi:10.1016/j.imlet.2020.03.007

15 Genovese MC, Durez P, Richards HB, *et al.* Efficacy and safety of secukinumab in patients with rheumatoid arthritis: a phase II, dose-finding, double-blind, randomised, placebo controlled study. *Ann Rheum Dis* 2013;**72**:863–9. doi:10.1136/annrheumdis-2012-201601

16 Miyagawa I, Nakayamada S, Nakano K, *et al.* Precision medicine using different biological DMARDs based on characteristic phenotypes of peripheral T helper cells in psoriatic arthritis. *Rheumatol (United Kingdom)* 2019;**58**:336–44. doi:10.1093/rheumatology/key069

17 Ishigaki K, Shoda H, Kochi Y, *et al.* Quantitative and qualitative characterization of expanded CD4 + T cell clones in rheumatoid arthritis patients. *Sci Rep* 2015;**5**:12937. doi:10.1038/srep12937

18 Bendall SC, Simonds EF, Qiu P, *et al.* Single-Cell Mass Cytometry of Differential Immune and Drug Responses Across a Human Hematopoietic Continuum. *Science (80-)* 2011;**332**:687–96. doi:10.1126/science.1198704

19 Van Gassen S, Callebaut B, Van Helden MJ, *et al.* FlowSOM: Using self-organizing maps for visualization and interpretation of cytometry data. *Cytom Part A* 2015;**87**. doi:10.1002/cyto.a.22625

20 Robinson MD, Nowicka M, Krieg C, *et al.* CyTOF workflow: Differential discovery in high-throughput high-dimensional cytometry datasets. *F1000Research* 2017;**6**. doi:10.12688/f1000research.11622.2

21 Qiu P, Simonds EF, Bendall SC, *et al.* Extracting a cellular hierarchy from high-dimensional cytometry data with SPADE. *Nat Biotechnol* 2011;**29**:886–91. doi:10.1038/nbt.1991

22 Bruggner R V, Bodenmiller B, Dill DL, *et al.* Automated identification of stratifying signatures in cellular subpopulations. *Proc Natl Acad Sci U S A* 2014;**111**:E2770-7. doi:10.1073/pnas.1408792111

23 Wilkerson MD, Hayes DN. ConsensusClusterPlus: A class discovery tool with confidence assessments and item tracking. *Bioinformatics* 2010;**26**. doi:10.1093/bioinformatics/btq170

24 Zhang F, Wei K, Slowikowski K, *et al.* Defining inflammatory cell states in rheumatoid arthritis joint synovial tissues by integrating single-cell transcriptomics and mass cytometry. *Nat Immunol* 2019;**20**:928–42. doi:10.1038/s41590-019-0378-1

25 Stahl EA, Raychaudhuri S, Remmers EF, *et al.* Genome-wide association study meta-analysis identifies seven new rheumatoid arthritis risk loci. *Nat Genet* 2010;**42**:508–14.

medRxiv preprint doi: https://doi.org/10.1101/2022.10.20.22281307; this version posted October 25, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

doi:10.1038/ng.582

26 Korsunsky I, Millard N, Fan J, *et al.* Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat Methods* 2019;**16**:1289–96. doi:10.1038/s41592-019-0619-0

27 Kang JB, Nathan A, Weinand K, *et al.* Efficient and precise single-cell reference atlas mapping with Symphony. *Nat Commun* 2021;**12**. doi:10.1038/s41467-021-25957-x

28 Finck R, Simonds EF, Jager A, *et al.* Normalization of mass cytometry data with bead standards. *Cytometry A* 2013;**83 A**:483–94. doi:10.1002/cyto.a.22271

29 Maecker HT, McCoy JP, Nussenblatt R. Standardizing immunophenotyping for the Human Immunology Project. *Nat Rev Immunol* 2012;**12**:191–200. doi:10.1038/nri3158

30 Wickham H. *ggplot2 -Positioning Elegant Graphics for Data Analysis*. 2016.

31 Kolde R. Package 'pheatmap': Pretty heatmaps. Version 1.0.12. 2019.

32 Van Der Maaten LJP, Hinton GE. Visualizing high-dimensional data using t-sne. *J Mach Learn Res* 2008.

33 Kamali AN, Noorbakhsh SM, Hamedifar H, *et al.* A role for Th1-like Th17 cells in the pathogenesis of inflammatory and autoimmune disorders. *Mol Immunol* 2019;**105**:107–15. doi:10.1016/J.MOLIMM.2018.11.015

34 Cohen CJ, Crome SQ, MacDonald KG, *et al.* Human Th1 and Th17 Cells Exhibit Epigenetic Stability at Signature Cytokine and Transcription Factor Loci. *J Immunol* 2011;**187**:5615–26. doi:10.4049/JIMMUNOL.1101058

35 Schmolka N, Papotto PH, Romero PV, *et al.* MicroRNA-146a controls functional plasticity in T cells by targeting NOD1. *Sci Immunol* 2018;**3**:1392. doi:10.1126/SCIIMMUNOL.AAO1392/SUPPL_FILE/AAO1392_TABLE_S2.ZIP

36 Zhang LL, Wu XX, Wang XF, *et al.* Genetic variant in microRNA-146a gene is associated with risk of rheumatoid arthritis. *Ann Med* 2021;**53**:824–9. doi:10.1080/07853890.2021.1933163

37 Löfgren SE, Frostegård J, Truedsson L, *et al.* Genetic association of miRNA-146a with systemic lupus erythematosus in Europeans through decreased expression of the gene. *Genes Immun* 2012;**13**:268–74. doi:10.1038/GENE.2011.84

38 de la Rosa M, Rutz S, Dorninger H, *et al.* Interleukin-2 is essential for CD4+CD25+ regulatory T cell function. *Eur J Immunol* 2004;**34**:2480–8. doi:10.1002/EJI.200425274

39 Chawla AS, Khalsa JK, Dhar A, *et al.* A role for cell-autocrine interleukin-2 in regulatory T-cell homeostasis. *Immunology* 2020;**160**:295–309. doi:10.1111/IMM.13194

40 Furtado GC, De Curotto Lafaille MA, Kutchukhidze N, *et al.* Interleukin 2 signaling is required for CD4(+) regulatory T cell function. *J Exp Med* 2002;**196**:851–7. doi:10.1084/JEM.20020190

41 Barton A, Pitzalis C. Stratified medicine in rheumatoid arthritis-the MATURA programme. *Rheumatology (Oxford)* 2017;**56**:1247–50. doi:10.1093/rheumatology/kew369

medRxiv preprint doi: https://doi.org/10.1101/2022.10.20.22281307; this version posted October 25, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

42 Humrich JY, Cacoub P, Rosenzwajg M, *et al.* Low-dose interleukin-2 therapy in active systemic lupus erythematosus (LUPIL-2): a multicentre, double-blind, randomised and placebo-controlled phase II trial. *Ann Rheum Dis* 2022;:annrheumdis-2022-222501. doi:10.1136/ARD-2022-222501

43 Zhang X, Miao M, Zhang R, *et al.* Efficacy and safety of low-dose interleukin-2 in combination with methotrexate in patients with active rheumatoid arthritis: a randomized, double-blind, placebo-controlled phase 2 trial. *Signal Transduct Target Ther* 2022;**7**. doi:10.1038/S41392-022-00887-2

44 Miyagawa I, Nakayamada S, Ueno M, *et al.* Precision medicine based on the phenotypic differences in peripheral T helper cells in patients with psoriatic arthritis: One year follow-up outcomes. *Front Med* 2022;**9**:2229. doi:10.3389/FMED.2022.934937/BIBTEX

45 Blanco FJ, Möricke R, Dokoupilova E, *et al.* Secukinumab in Active Rheumatoid Arthritis: A Phase III Randomized, Double-Blind, Active Comparator- and Placebo-Controlled Study. *Arthritis Rheumatol* 2017;**69**:1144–53. doi:10.1002/art.40070

46 Spidlen J, Breuer K, Brinkman R. Preparing a Minimum Information about a Flow Cytometry Experiment (MIFlowCyt) Compliant Manuscript Using the International Society for Advancement of Cytometry (ISAC) FCS File Repository (FlowRepository.org). *Curr Protoc Cytom* 2012;**61**:10.18.1-10.18.26. doi:10.1002/0471142956.CY1018S61

medRxiv preprint doi: https://doi.org/10.1101/2022.10.20.22281307; this version posted October 25, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Figure legends

Figure 1. Abundances of Th and Treg $CD4^+$ T cell subsets identified by manual gating of CyTOF data shown with mean, 95% confidence interval and p-value (shown above the pairwise comparisons). Points represent RA and HC samples. Mann Whitney *U* P values are shown above the pairwise comparisons. Graphs with cytokine expression are taken from the stimulated T cell dataset and those with surface markers and transcription factors are taken from the unstimulated dataset. Box A represents markers associated with Th1 cells, box B Treg cells, box C represents Th17-related markers, and box D Th2-related markers.

Figure 2. Abundances of $CD8^+$ T cell subsets identified by manual gating of CyTOF data shown with mean and 95% confidence interval. Points represent RA and HC samples. Mann Whitney *U* P values are shown above the pairwise comparisons. Graphs with cytokine expression are taken from the stimulated T cell dataset and those with surface markers and transcription factors are taken from the unstimulated dataset.

Figure 3. A) Multidimensional scaling (MDS) plot based on median, arcsinh-transformed marker expression in unstimulated (U) and stimulated (S) T cell samples from rheumatoid arthritis (RA) patients and healthy controls (HC). **B)** PCA plot of the same data coloured by case-control status. **C)** PCA plot of the same data coloured by age group of participants.

Figure 4. A) Heatmap of median marker expression (arcsinh, 0-1 transformed) for 20 T cell clusters identified in unstimulated cell samples by automated clustering. Each sample was downsampled to 2746 cells (n = 10 RA, 10 HC). The dendrogram represents hierarchical clustering using Euclidean distance and average linkage. **B)** As A), for stimulated samples (n = 10 RA, 10 HC). Each sample was downsampled to 2224 cells. **C)** t-SNE plots of the same unstimulated cells, coloured by expression level of a selection of markers. **D)** As C), for stimulated cells.

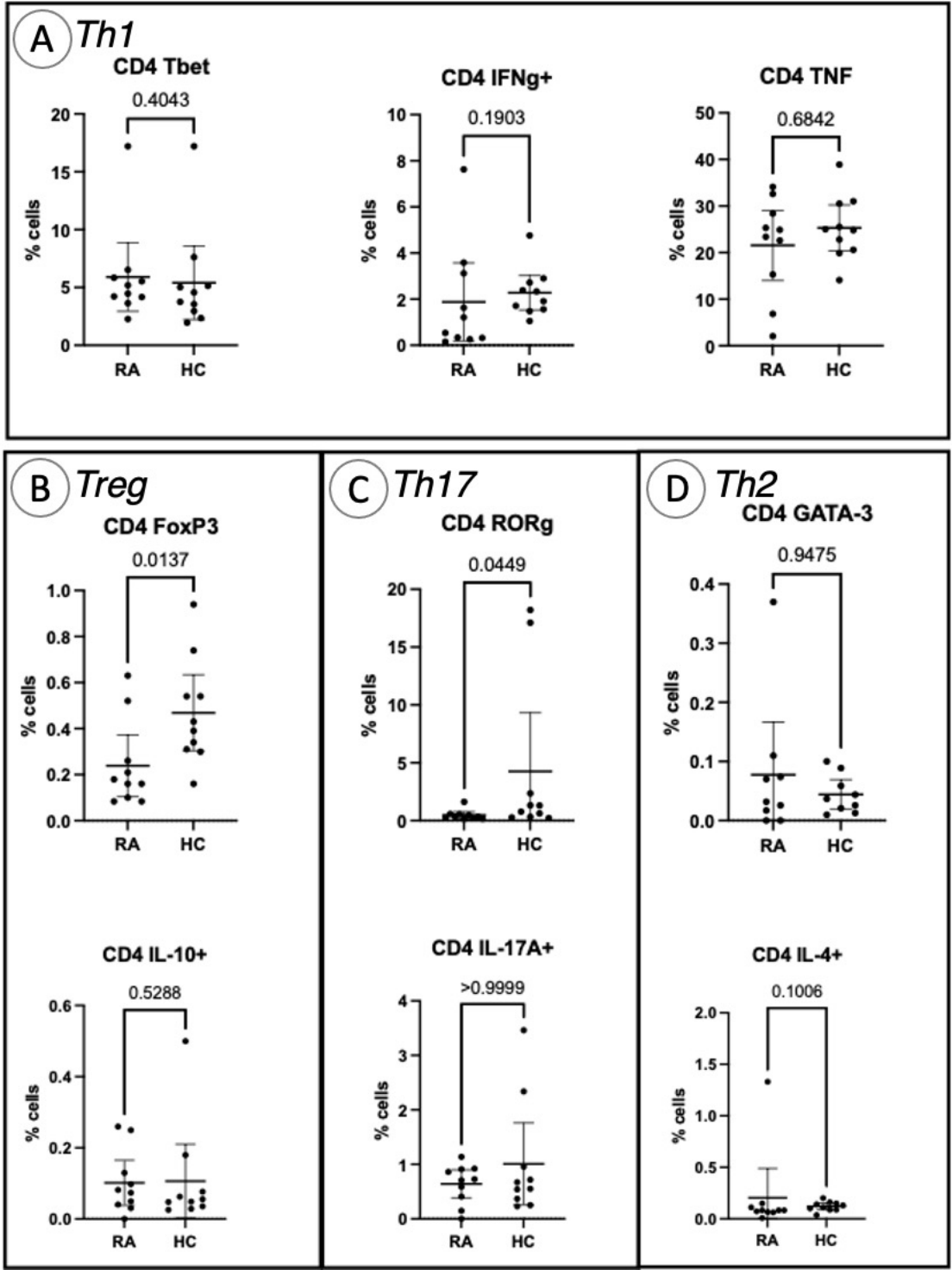
Figure 5. A) tSNE plots of unstimulated cells faceted by case-control status and coloured by cluster assignment. Samples were downsampled to 2746 cells each (n = 20). **B)** tSNE plot of unstimulated cells with conditions combined. Clusters with differential abundance in RA vs healthy controls (HC) according to MASC are highlighted ($P \leq 0.05$). **C)** As A), for stimulated cells. Samples were downsampled to 2224 cells each (n = 20). **D)** As B), for unstimulated cells. Note that unstimulated and stimulated samples were clustered independently.

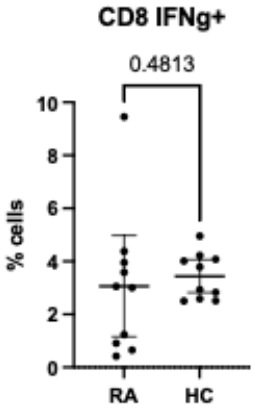
Figure 6. Boxplots of the proportions of 20 T cell clusters identified in unstimulated RA and healthy control (HC) samples (n = 10 RA, 10 HC). The P value obtained from MASC is shown for each cluster, and a green outline indicates differential abundance between RA and HC ($P \leq 0.05$).

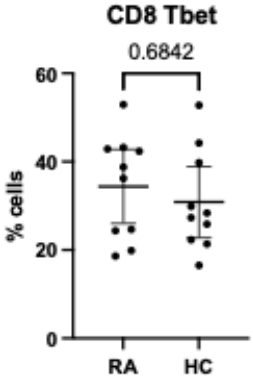
Figure 7. Boxplots of the proportions of 20 T cell clusters identified in stimulated RA and healthy control (HC) samples (n = 10 RA, 10 HC). The P value obtained from MASC is shown for each cluster, and a green outline indicates differential abundance between RA and HC ($P \leq 0.05$).

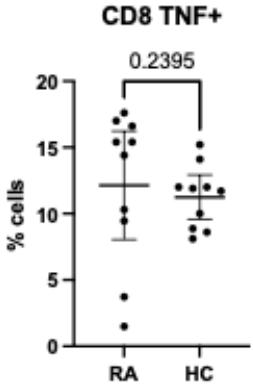
medRxiv preprint doi: https://doi.org/10.1101/2022.10.20.22281307; this version posted October 25, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

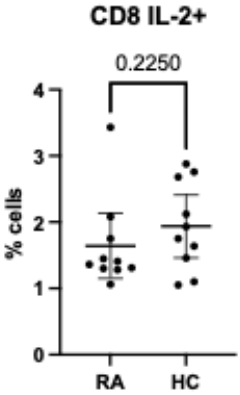
Figure 8. Clusters 11 and 17 represent the abundance of manually gated subsets previously identified by the CyTOF automated clustering pipeline, shown here with mean and 95% confidence intervals. Clusters 20 and 16 represent the expression of CCR7 on Th1-like IL-17^{+} cells. Points represent RA and HC samples. Mann Whitney *U P* values are shown above the pairwise comparisons. TEM: T effector memory cell. TCM: T central memory cell.

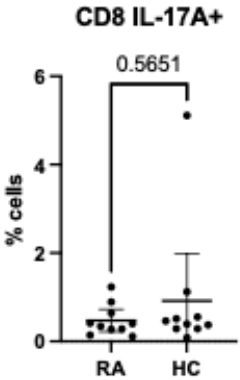

A
Th1
CD4 Tbet
0.4043
% cells
RA
HC
CD4 IFNg+
0.1903
CD4 TNF
0.6842
B
Treg
CD4 FoxP3
0.0137
CD4 IL-10+
0.5288
C
Th17
CD4 RORg
0.0449
CD4 IL-17A+
>0.9999
D
Th2
CD4 GATA-3
0.9475
CD4 IL-4+
0.1006

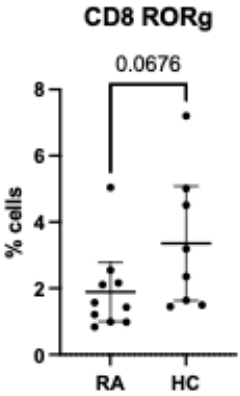

CD8 IFNg+
0.4813
% cells
0
2
4
6
8
10
RA
HC

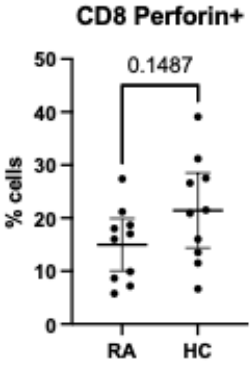

CD8 Tbet
0.6842
% cells
0
20
40
60
RA
HC

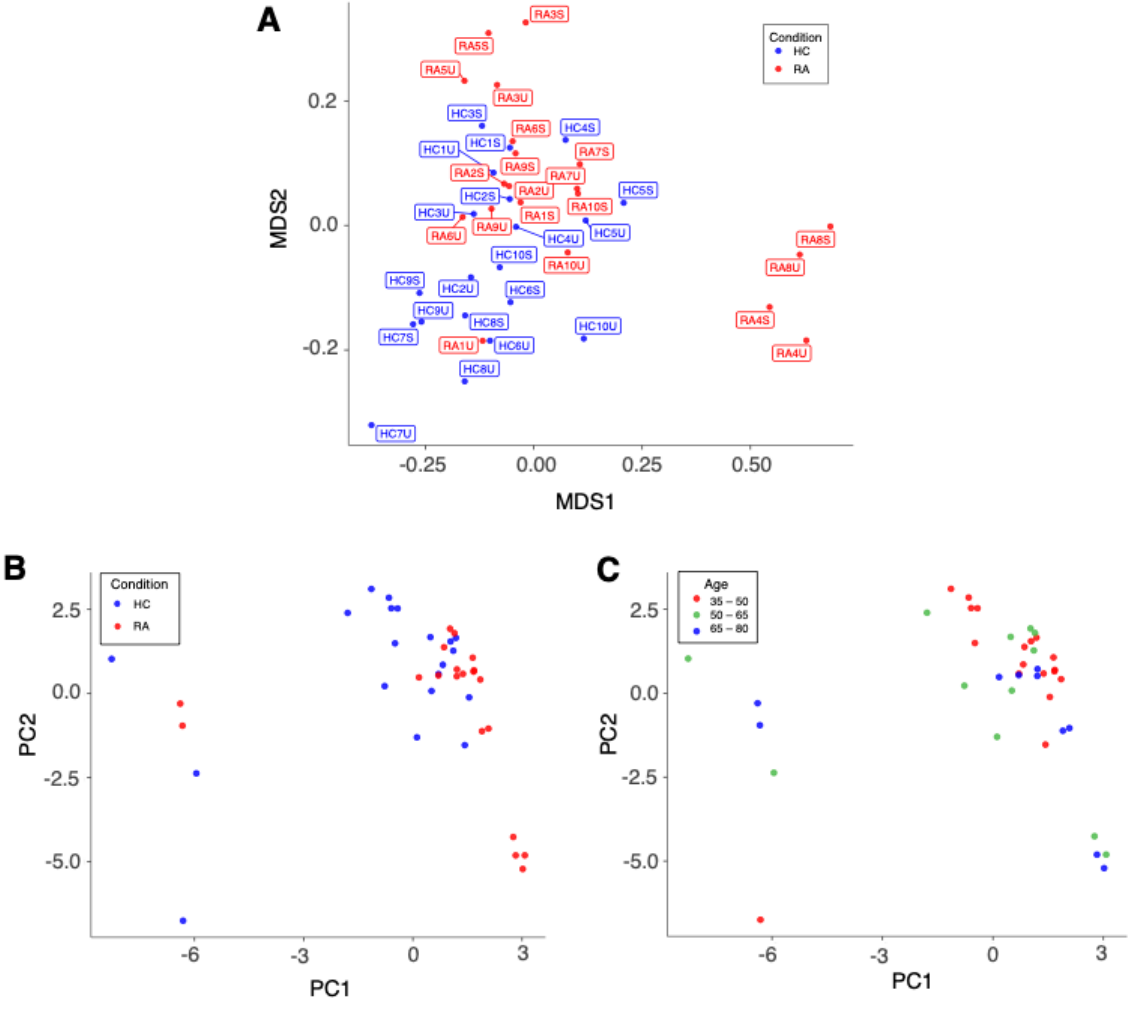

CD8 TNF+
0.2395
% cells
0
5
10
15
20
RA
HC

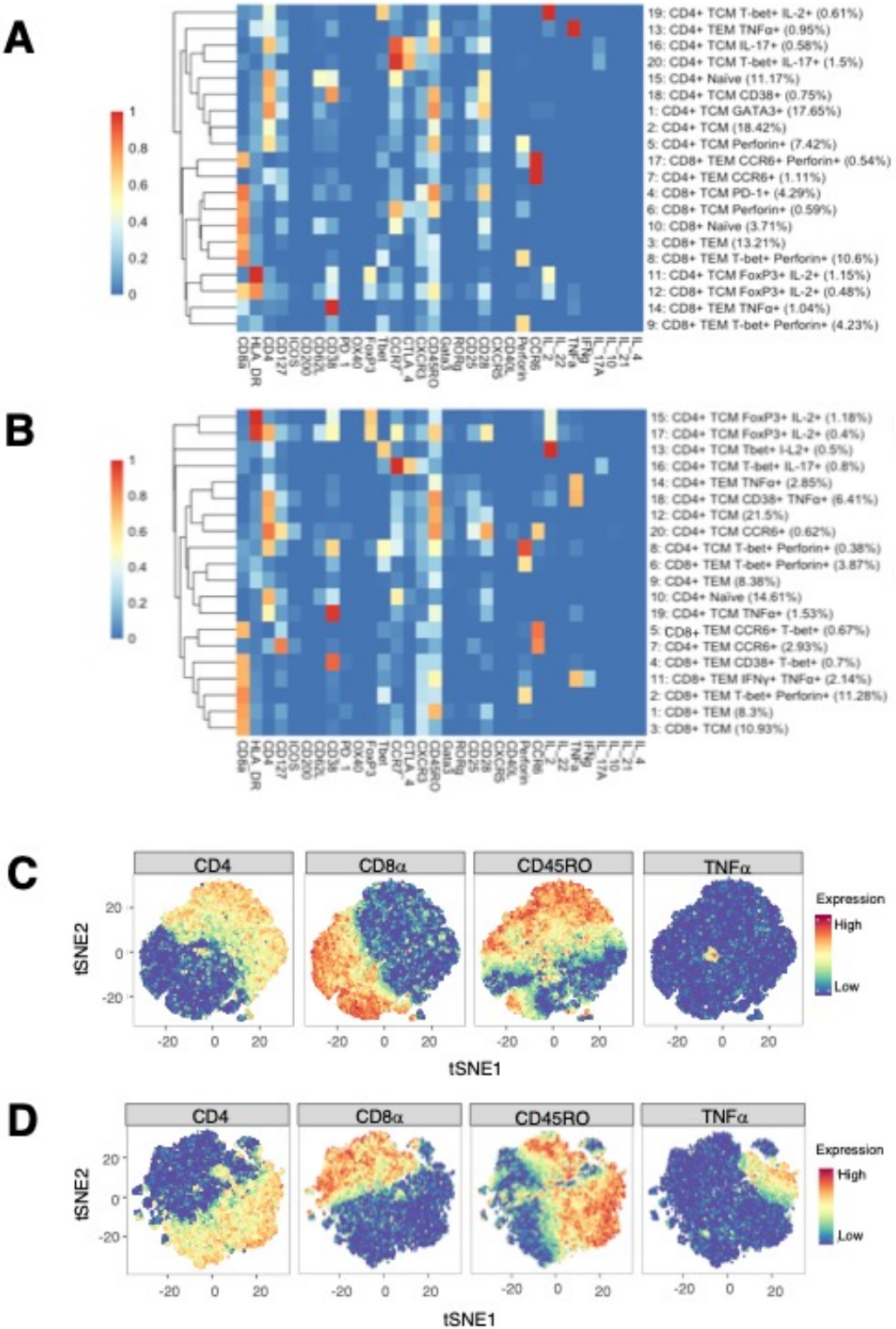

CD8 IL-2+
0.2250
% cells
0
1
2
3
4
RA
HC

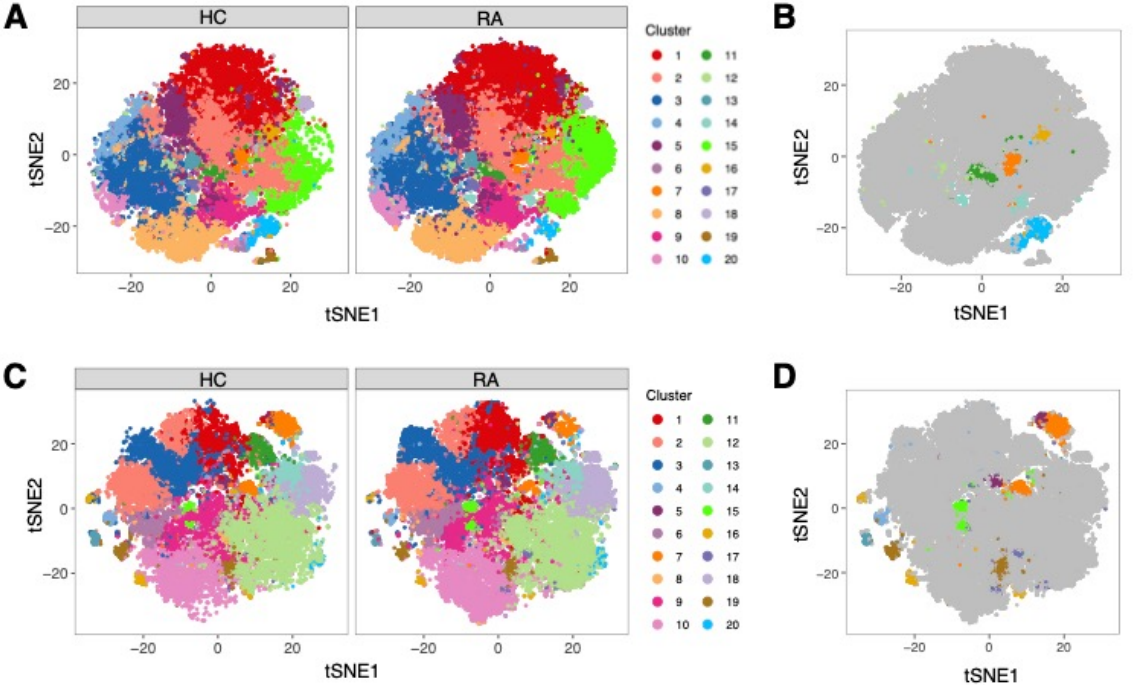

CD8 IL-17A+
0.5651
% cells
0
2
4
6
RA
HC

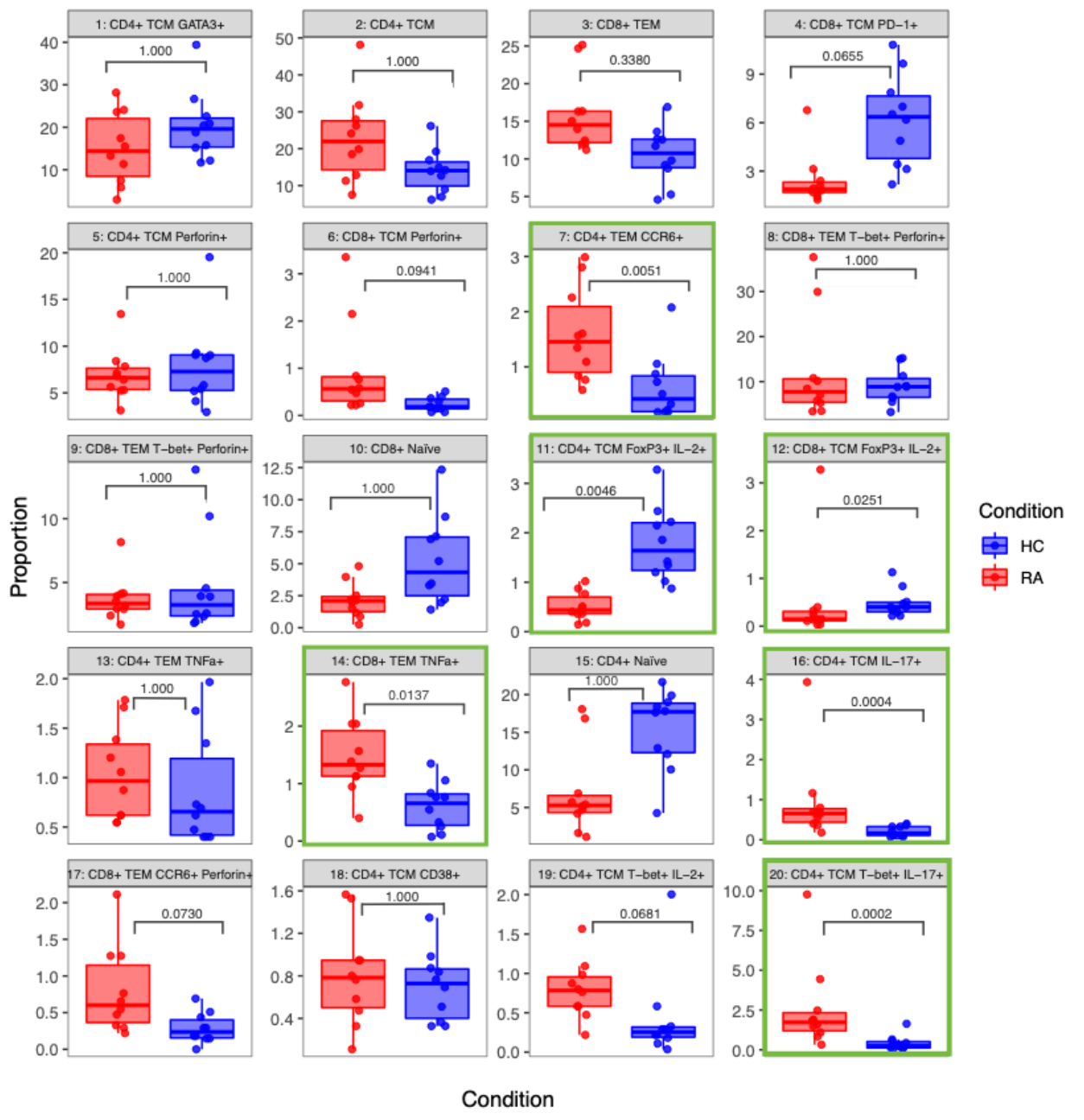

CD8 RORg
0.0676
% cells
0
2
4
6
8
RA
HC

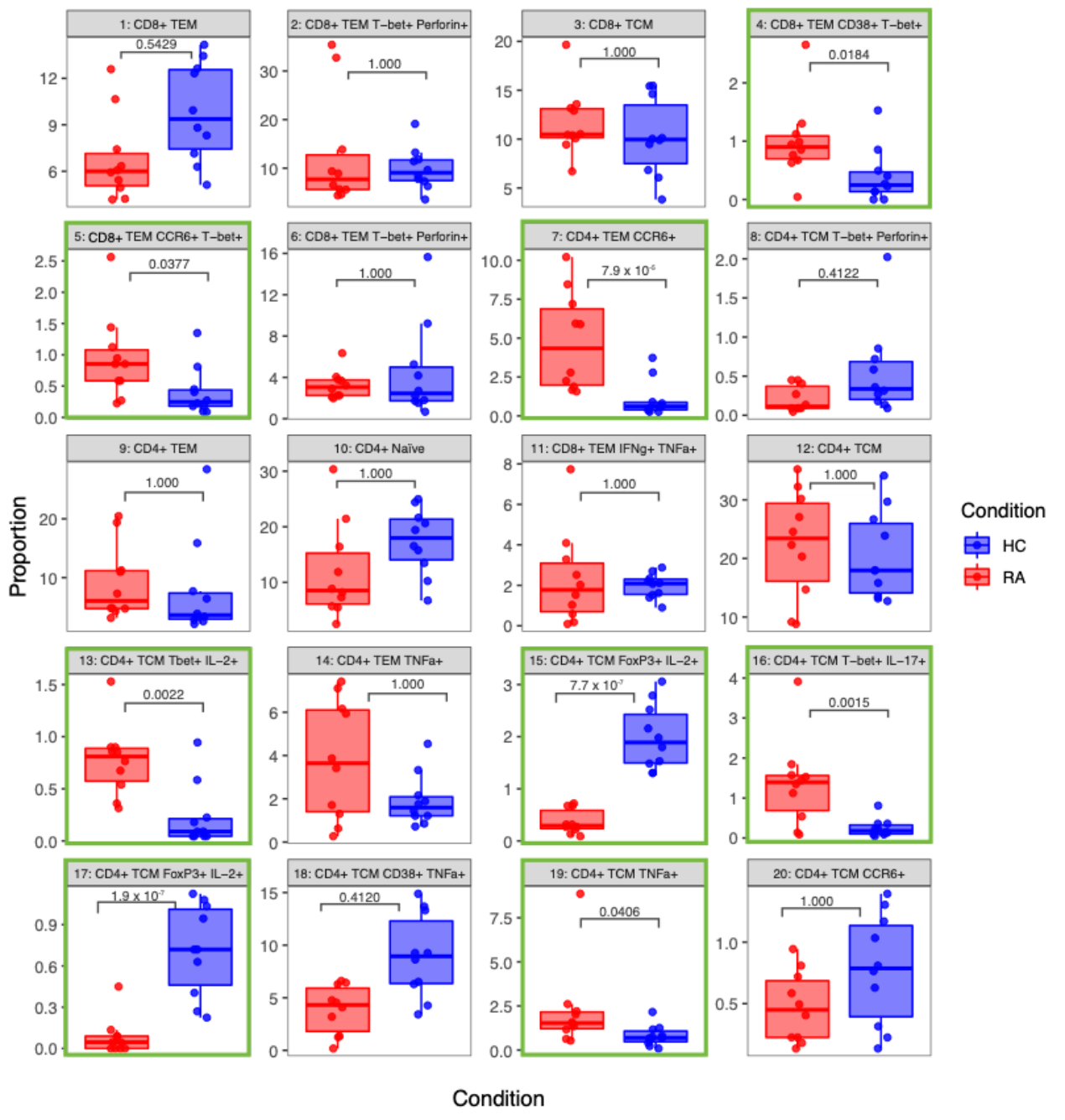

CD8 Perforin+
0.1487
% cells
0
10
20
30
40
50
RA
HC

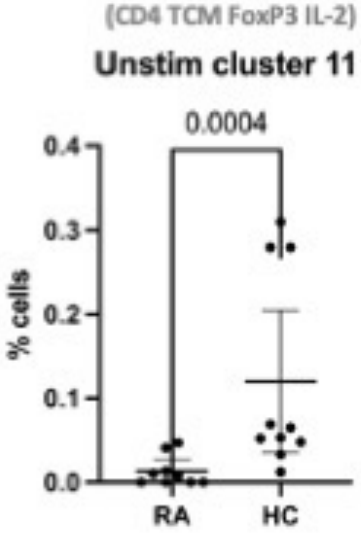

A
Condition
HC
RA
MDS2
MDS1
0.2
0.0
-0.2
-0.25
0.00
0.25
0.50
RA3S
RA5S
RA5U
RA3U
HC3S
RA6S
HC4S
HC1S
HC1U
RA7S
RA9S
RA2S
RA7U
HC5S
HC2S
RA2U
RA10S
HC3U
RA1S
RA9U
HC5U
RA6U
HC4U
RA8S
HC10S
RA10U
RA8U
HC9S
HC2U
HC6S
HC9U
HC8S
HC10U
RA4S
HC7S
RA1U
HC6U
RA4U
HC8U
HC7U
B
Condition
HC
RA
PC2
PC1
2.5
0.0
-2.5
-5.0
-6
-3
0
3
C
Age
35 – 50
50 – 65
65 – 80
PC2
PC1
2.5
0.0
-2.5
-5.0
-6
-3
0
3

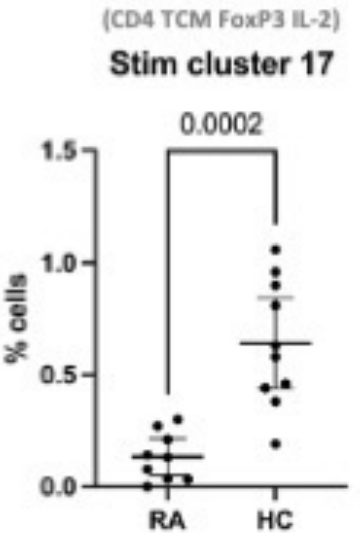

A
19: CD4+ TCM T-bet+ IL-2+ (0.61%)
13: CD4+ TEM TNFα+ (0.95%)
16: CD4+ TCM IL-17+ (0.58%)
20: CD4+ TCM T-bet+ IL-17+ (1.5%)
15: CD4+ Naive (11.17%)
18: CD4+ TCM CD38+ (0.75%)
1: CD4+ TCM GATA3+ (17.65%)
2: CD4+ TCM (18.42%)
5: CD4+ TCM Perforin+ (7.42%)
17: CD8+ TEM CCR6+ Perforin+ (0.54%)
7: CD4+ TEM CCR6+ (1.11%)
4: CD8+ TCM PD-1+ (4.29%)
6: CD8+ TCM Perforin+ (0.59%)
10: CD8+ Naive (3.71%)
3: CD8+ TEM (13.21%)
8: CD8+ TEM T-bet+ Perforin+ (10.6%)
11: CD4+ TCM FoxP3+ IL-2+ (1.15%)
12: CD8+ TCM FoxP3+ IL-2+ (0.48%)
14: CD8+ TEM TNFα+ (1.04%)
9: CD8+ TEM T-bet+ Perforin+ (4.23%)
CD8a HLA_DR CD4 CD127 ICOS CD200 CD62L CD38 PD_1 OX40 FoxP3 Tbet CCR7 CTLA_4 CXCR3 CD45RO Gata3 RORγt CD25 CD28 CXCR5 CD40L Perforin CCR6 IL_2 IL_22 TNFα IFNγ IL_17A IL_10 IL_21 IL_4
B
15: CD4+ TCM FoxP3+ IL-2+ (1.18%)
17: CD4+ TCM FoxP3+ IL-2+ (0.4%)
13: CD4+ TCM Tbet+ I-L2+ (0.5%)
16: CD4+ TCM T-bet+ IL-17+ (0.8%)
14: CD4+ TEM TNFα+ (2.85%)
18: CD4+ TCM CD38+ TNFα+ (6.41%)
12: CD4+ TCM (21.5%)
20: CD4+ TCM CCR6+ (0.62%)
8: CD4+ TCM T-bet+ Perforin+ (0.38%)
6: CD8+ TEM T-bet+ Perforin+ (3.87%)
9: CD4+ TEM (8.38%)
10: CD4+ Naive (14.61%)
19: CD4+ TCM TNFα+ (1.53%)
5: CD8+ TEM CCR6+ T-bet+ (0.67%)
7: CD4+ TEM CCR6+ (2.93%)
4: CD8+ TEM CD38+ T-bet+ (0.7%)
11: CD8+ TEM IFNγ+ TNFα+ (2.14%)
2: CD8+ TEM T-bet+ Perforin+ (11.28%)
1: CD8+ TEM (8.3%)
3: CD8+ TCM (10.93%)
C
CD4
CD8α
CD45RO
TNFα
Expression
High
Low
tSNE2
tSNE1
D
CD4
CD8α
CD45RO
TNFα
Expression
High
Low
tSNE2
tSNE1

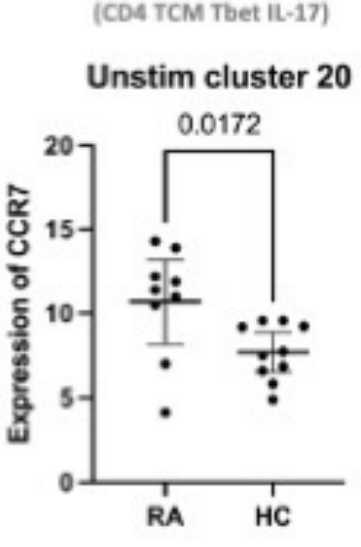

A
HC
RA
Cluster
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
tSNE2
tSNE1
B
C
D

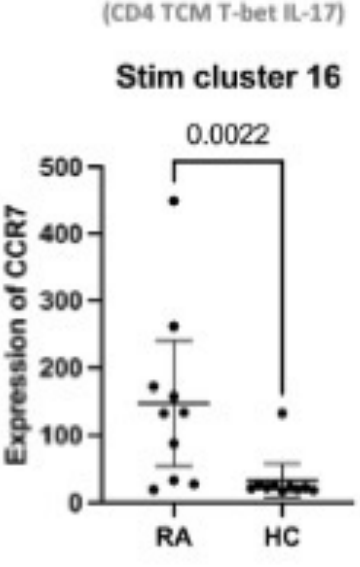

1: CD4+ TCM GATA3+
1.000
2: CD4+ TCM
1.000
3: CD8+ TEM
0.3380
4: CD8+ TCM PD−1+
0.0655
5: CD4+ TCM Perforin+
1.000
6: CD8+ TCM Perforin+
0.0941
7: CD4+ TEM CCR6+
0.0051
8: CD8+ TEM T−bet+ Perforin+
1.000
9: CD8+ TEM T−bet+ Perforin+
1.000
10: CD8+ Naïve
1.000
11: CD4+ TCM FoxP3+ IL−2+
0.0046
12: CD8+ TCM FoxP3+ IL−2+
0.0251
13: CD4+ TEM TNFa+
1.000
14: CD8+ TEM TNFa+
0.0137
15: CD4+ Naïve
1.000
16: CD4+ TCM IL−17+
0.0004
17: CD8+ TEM CCR6+ Perforin+
0.0730
18: CD4+ TCM CD38+
1.000
19: CD4+ TCM T−bet+ IL−2+
0.0681
20: CD4+ TCM T−bet+ IL−17+
0.0002
Proportion
Condition
HC
RA
Condition


1: CD8+ TEM
0.5429
2: CD8+ TEM T−bet+ Perforin+
1.000
3: CD8+ TCM
1.000
4: CD8+ TEM CD38+ T−bet+
0.0184
5: CD8+ TEM CCR6+ T−bet+
0.0377
6: CD8+ TEM T−bet+ Perforin+
1.000
7: CD4+ TEM CCR6+
7.9 x 10⁻⁶
8: CD4+ TCM T−bet+ Perforin+
0.4122
9: CD4+ TEM
1.000
10: CD4+ Naïve
1.000
11: CD8+ TEM IFNg+ TNFa+
1.000
12: CD4+ TCM
1.000
13: CD4+ TCM Tbet+ IL−2+
0.0022
14: CD4+ TEM TNFa+
1.000
15: CD4+ TCM FoxP3+ IL−2+
7.7 x 10⁻⁷
16: CD4+ TCM T−bet+ IL−17+
0.0015
17: CD4+ TCM FoxP3+ IL−2+
1.9 x 10⁻⁷
18: CD4+ TCM CD38+ TNFa+
0.4120
19: CD4+ TCM TNFa+
0.0406
20: CD4+ TCM CCR6+
1.000
Proportion
Condition
HC
RA
Condition


(CD4 TCM FoxP3 IL-2)
Unstim cluster 11
0.0004
% cells
0.4
0.3
0.2
0.1
0.0
RA
HC


(CD4 TCM FoxP3 IL-2)
Stim cluster 17
0.0002
% cells
1.5
1.0
0.5
0.0
RA
HC


(CD4 TCM Tbet IL-17)
Unstim cluster 20
0.0172
Expression of CCR7
20
15
10
5
0
RA
HC


(CD4 TCM T-bet IL-17)
Stim cluster 16
0.0022
Expression of CCR7
500
400
300
200
100
0
RA
HC