Online Supplementary Materials

Integrin and transcriptomic profiles identify a distinctive synovial CD8+ T cell subpopulation in spondyloarthritis

Authors: Zoya Qaiyum, Eric Gracey, Yuchen Yao, Robert D. Inman

Contents

Pgs 2-6: Supplementary methods and references

Pg 7: Supplementary Figure 1: Gating strategy for live and major cell populations

Pg 8: Supplementary Figure 2: Visualization of macrophages in AS SF

Pg 9: Supplementary Figure 3: Heatmap statistics of monocytes and NK cells

Pg 10: Supplementary Figure 4: $CD56^{hi}$ and $CD56^{dim}$ NK cells in blood and synovial fluid exhibit characteristic frequencies and CD16 profiles

Pg 11: Supplementary Figure 5: Extended characterization of CD103+CD8+ T cells in synovial fluid

Pg 12: Supplementary Figure 6: Regression analysis on SFMC and PBMC

Pg 13: Supplementary Figure 7: Extended RNA-seq data of InEx, Non-InEx and naïve cells

Pg 14: Supplementary Figure 8: Protein-protein interaction network from unique and elevated genes

Pg 15: Supplementary Figure 9: Cytokine multiplex assay on unstimulated and stimulated InEx and Non-InEx cells

Pg 16: Supplementary Table 1: Cohort demographics

Pg 17: Supplementary Table 2: CyTOF panel of antibodies

Pg 18: Supplementary Table 3: Antibodies for FACS sorting

Pgs 19-21: Supplementary Figures 1-9 and Tables 1-3 Legends

SUPPLEMENTARY METHODS

Patient Cohorts

We have established a biobank of peripheral blood mononuclear cells (PBMC) and synovial fluid mononuclear cells (SFMC) from AS patients in our longitudinal observation cohort, PBMC and SFMC from rheumatoid arthritis (RA) patients; and PBMC from healthy controls (HC). These were isolated via Ficol-Paque density centrifugation (GE). Our studies have identified young males with AS as having the most pronounced inflammatory signature[1], so we assembled a cohort of such patients fulfilling the modified New York (mNY) criteria for AS. Active joint effusion in AS patients reflects active joint inflammation. HC were age- and sex-matched. Samples from RA patients were selected as disease controls but were not age- or sex-matched as RA is predominantly an older female cohort[2] (Table S1).

Mass Cytometry

PBMCs and SFMCs stored in liquid N2 were thawed and incubated overnight at 37°C in RPMI supplemented with 10% heat-inactivated FCS, penicillin/streptomycin and 50nM β-mercaptoethanol. Subsequently, cells were processed for mass cytometry staining. After blocking target cells with anti-CD16/CD32 Fc receptor blocking solution (BioLegend), the cells were stained with a customized 36-marker antibody panel (Table S2). Since this is the first report assessing trafficking markers in AS extensively using high-dimensional methods, panel design was based on factors such as function, marker expression by cell type and tissue location (primarily gut or joint), and disease implications based on knockout and in vitro studies from RA and IBD literature, in order to be able to extrapolate some of the findings in AS patients. In the final panel, markers were either immune markers, activating/identifying markers, chemokine receptors or integrins. All antibodies were validated by conventional flow cytometry for expression on fresh vs thawed PBMC or SFMC. Where indicated, antibodies were purchased pre-conjugated from Fluidigm or custom conjugated with metal tags using MaxPar Ready© antibodies from BioLegend, according to the manufacturer's protocol. Antibodies were titrated prior to staining and MMO (metal-minus-one) controls used for gating purposed when required. After staining with antibodies, 1μM cisplatin (Fluidigm) was used to identify live cells. DNA intercalation solution (Fluidigm) was added for 1 hr (0.05μM) or overnight (0.078μM) to distinguish cells from debris. Samples were acquired on the CyTOF2 instrument (Fluidigm) at Peter Gilgan Centre for Research and Learning (PGCRL), Hospital for Sick Children, Toronto, Canada. Only matched samples were thawed and stained on the same day and in order to further minimize batch effects, a mix of samples were thawed and stained on the same day (for instance, not all HCs were processed on the same day, rather they were included with AS PBMC or RA PBMC). Where indicated, raw CyTOF files were pre-processed in FlowJo to select live cell populations of interest for unsupervised analysis. Where indicated, populations of interest were concatenated using FlowJo v10 into a single file for all subjects in each disease group. Range of live events, rounded off to nearest thousand, prior to concatenation as follows: AS PBMC= 77,000-127,000; HC PBMC= 75,000-113,000; AS SFMC= 75,000-115,000. Median live events, rounded off to nearest thousand, prior to concatenation as follows: AS PBMC= 112,000; HC PBMC= 103,000; AS SFMC= 101,000. Range of CD3+ events, rounded off to nearest thousand,

prior to concatenation as follows: AS PBMC= 50,000-100,000; HC PBMC= 50,000-84,000; AS SFMC= 9,000-84,000. Median CD3+ events, rounded off to nearest thousand, prior to concatenation as follows: AS PBMC= 78,000; HC PBMC= 71,000; AS SFMC= 62,000. For data analysis, SPADE[3] and viSNE[4] programs were employed on the Cytobank platform (cytobank.org) for unsupervised clustering analysis using concatenated FCS files. All scales were Arcsinh transformed with argument 5. For SPADE (Figures 1A and 2A), 200 target nodes were input with 10% of down-sampled target events. All parameters were selected for clustering, with the exception of the following markers as they had low expression on total live and CD3+ T cells: CD66a, CD122, CX3CR1, CD61, CD41 and CXCR6. The viSNE algorithm (Figure 3A) for mature CD8+ T cells was selected with a maximum of 1,000,000 events, proportional sampling and clustering parameters (CD103, CD49a, CD49b, CD29, CD18, CD161, CD11a, β7, CCR5, CCR6, CCR9, CXCR3, CXCR4, CXCR6) to assess concatenated FCS files of mature CD8+ T cell population files (AS-PBMCs, AS-SFMCs). viSNE clustering parameters were selected based on trafficking marker expression. Other markers such as cell identifiers and low expressing trafficking molecules were excluded. Gating in FlowJo for selection of populations prior to concatenation given in Figures S1.

Mass Cytometry Statistical Analysis

Mass cytometry statistical analysis detailed in Figure Legends and Supplementary Figures. FlowJo v10 and GraphPad Prism v6 were used for all supporting statistical analysis unless indicated in text. All data tested for normality using D'Agostino and Pearson omnibus normality test before selecting appropriate test indicated in the respective Figure Legends (Figures 1B, 2B, 2C, 3D, S3A, S3B, S4A, S4B, S5A, S5B, S5D and S5E). For all plots, a single point represents an individual. Data shown in graphs as mean ± SEM. **** $p<0.0001$, *** $p<0.001$, ** $p<0.01$, * $p<0.05$, n.s. $p>0.05$ (non-significant results not highlighted on graph, unless indicated). To investigate the influence of TNF inhibitor treatment on InEx cell frequency in SF (Figure S6A) and blood (Figure S6B), a regression analysis was performed using IBM SPSS Statistics v22.0.0, with biologic status 'yes' given a value of 1 and 'no' given a value of 0.

Microscopy Imaging

1mL whole blood taken from an AS patient was treated with 20mL red blood cell (RBC) lysis buffer (Qiagen) for 30 min at room temperature. Whole SF from an unmatched AS patient was used as well. 500,000 cells from both samples were used to prepare cytospin slides, which were stained with Wright-Giemsa stain (Sigma-Aldrich) to detect nuclei. The Aperio Scanscope AT2 Whole Slide Scanner was used to digitize slides at 40x magnification, and the Aperio ImageScope v12.3.2.8013 was used for colour correction and capturing snapshots (Figure S2).

FACS Sorting

SFMC from 5 AS patients were prepared from frozen samples as described. The panel used to sort samples on the BD Aria III at the Toronto Western Krembil Discovery Tower (KDT)-UHN Flow Cytometry Facility, Toronto, Canada, is listed in Table S3, with gating strategy in Figure S7A. 50,000-500,000 cells per sample were collected in RPMI with 50% heat-inactivated FCS from the InEx cells, Non-InEx and Naïve CD8+ T cell populations.

RNA Sequencing

RNA was immediately isolated from the FACS sorted InEx, Non-InEx and naïve cells using TRIzol reagent (Life Technologies) following the manufacturer's protocol. RNA integrity was assessed by Agilent Bioanalyzer to determine sample RIN and concentration was assessed by the Qubit RNA HS assay kit (Life Technologies). All samples had RIN values>8. cDNA libraries were created from 2ng of RNA using the SMARTer Stranded Total RNA-seq Kit (Takara Bio USA Inc.). Final cDNA library sizing was verified on the Agilent high sensitivity DNA kit while library concentration was quantified by qPCR using the KAPA SYBR FAST qPCR kit (Kapa Biosystems). Libraries were normalized to 10nM, diluted to 2nM, denatured using 0.2N NaOH, and diluted again to 1.7pM before loading onto the NextSeq 500 system. The samples were sequenced as a pair-end 75 cycle sequencing run to achieve a minimum of ~40M reads per sample. The raw sequence data, in the form of FASTQ files, was aligned to the human genome (UCSC hg38 and GTF definition file) using the HISAT2 v2.1.0. Accessory programs for the alignment stage included samtools 1.3.1. Transcripts were assembled and gene expressions were estimated with StringTie 1.3.3 using UCSC GTF file and GTF-only mode. StringTie output results were loaded into R 3.4.1/ballgown 2.8.4 for gene/transcript differential expression analysis. All RNA-seq processing (subsequent to sorting) was conducted at the Princess Margaret Genomics Centre (PMGC), Toronto, Canada.

ClustVis[5] and Venny[6] algorithms were used to generate differential gene expression graphs (Figures 4A, 4B and S7B) and principal component analysis (PCA) graph (Figure S7C). Differential gene expression graphs were plotted with significance level $p<0.05$ and fold change (FC) greater than or less than 2. For the PCA plot, unit variance scaling was applied to the row values and SVD with imputation used to calculate principal components. X and Y axes show principal component 1 and principal component 2 that explain 28.9% and 14.8% of the total variance, respectively. Uncorrected p values were used for our analyses, due to our low sample numbers. Volcano plots generated using GraphPad Prism 5 (Figure S7D). Pathway analysis (Figures 4C and S7E) was conducted using the ClueGo application from Cytoscape program ($p<0.05$), all ontologies selected and plotted using GraphPad Prism v6.

Pathway enrichment analysis (Figure 4D) was performed using the pathDIP[7] software. Gene lists were entered, all pathway sources selected and only the literature curated (core) pathway memberships option selected. From the output, q values (FDR: BH-method) were used to plot the results in GraphPad Prism v6.

NanoString

SFMC from 6 AS patients were prepared from frozen samples as described. InEx and Non-InEx cells were FACS sorted at the KDT flow facility as above and stimulated in X-VIVO15 serum-free media (Lonza) with anti-CD2/CD3/CD28 T cell activation beads (Miltenyi) and 500U/mL IL-2 (GenScript) for 72 hr at 37° C. Subsequently, RNA was isolated using TRIzol reagent (Life Technologies), according to the manufacturer's protocol, and processed for hybridization for 12-30 hr at 65°C using the nCounter mRNA assay for the nCounter Immunology panel v2 (Human) (NanoString Technologies) at the Princess Margaret Genomics Centre (PMGC). Post-hybridization was performed using nCounter Prep Station and Digital Analyzer was used for data collection. Raw RNA transcript counts were assessed for quality control and normalized using the nSolver v4 software. Normalized RNA transcripts were converted into log2 values for statistical analysis (Figure 5).

Protein-Protein Interactions

Differentially expressed genes among InEx, Non-InEx and naïve cells were entered into the Integrated Interactions Database (IID)[8] v2018-11 program, and only interactions among query proteins retrieval strategy was selected. Resulting output files were input into the NAViGaTOR[9] 3.0 software. In the 1,124 gene network analysis (Figure 6), only those proteins with degree of interaction ≥ 15 were selected. Enriched pathways obtained from the pathDIP program were highlighted. In Figure S8A and S8B, only those proteins with degree of interaction ≥ 1 were selected and enriched pathway highlighted (Figure S8B).

Cytokine Multiplex Assay

SFMC from 5 AS patients were prepared from frozen samples as described. InEx and Non-InEx cells were FACS sorted at the KDT flow facility as above and stimulated in X-VIVO15 serum-free media (Lonza) with anti-CD2/CD3/CD28 T cell activation beads (Miltenyi) and 500U/mL IL-2 (GenScript) or left unstimulated in X-VIVO15 serum-free media (Lonza) for 72 hr at 37° C. Supernatants were collected and stored at -80° C. Once all samples were collected, all supernatants were used in the LEGENDplex™ Human CD8/NK Panel (BioLegend) assay, according to manufacturer's protocol. Stimulated and unstimulated supernatants were run at a dilution of 1:4 and 1:2, respectively. Acquisition of samples was performed on the FACS Canto II machine at the KDT flow facility. FCS files were entered into the LEGENDplex v8.0 analysis software and steps outlined by the manufacturer's protocol for analysis was followed. Resulting concentrations (pg/mL) of select cytokines were plotted using GraphPad Prism v6 (Figure S9).

References

1 Gracey E, Yao Y, Green B, *et al*. Sexual dimorphism in the T_H17 signature of ankylosing spondylitis. *Arthritis Rheumatol* 2016;68(3):679-689.

2 Cannon CP, Curtis SP, Bolognese JA, *et al*. Clinical trial design and patient demographics of the Multinational Etoricoxib and Diclofenac Arthritis Long-term (MEDAL) study program: cardiovascular outcomes with etoricoxib versus diclofenac in patients with osteoarthritis and rheumatoid arthritis. *Am Heart J* 2006;152(2):237-245.
3 Qiu P, *et al*. Extracting a cellular hierarchy from high-dimensional cytometry data with SPADE. *Nat Biotechnol* 2011;29(10):886-891.
4 Amir ED, *et al*. viSNE enables visualization of high-dimensional single-cell data and reveals phenotypic heterogeneity of leukemia. *Nat Biotechnol* 2013;31(6):545-552.
5 Metsalu T, Vilo J. ClustVis: a web tool for visualizing clustering of multivariate data using Principal Component Analysis and heatmap. *Nucleic Acids Res* 2015;43(W1):W566-W570.
6 Oliveros JC. Venny. An interactive tool for comparing lists with Venn's diagrams. (2007-2015). http://bioinfogp.cnb.csic.es/tools/venny/index.html
7 Rahmati S, Abovsky M, Pastrello C, Jurisica I. pathDIP: An annotated resource for known and predicted human gene-pathway associations and pathway enrichment analysis. *Nucl Acids Res* 2017;45(D1):D419-D426.
8 Kotlyar M, Pastrello C, Malik Z, Jurisica I. IID 2018 update: context-specific physical protein-protein interactions in human, model organisms and domesticated species. *Nucleic Acids Res* 2019;47(D1):D581-589.
9 Brown KR, Otasek D, Ali M, McGuffin MJ, Xie W, Devani B, Toch IL, Jurisica I. NAViGaTOR: Network Analysis, Visualization and Graphing Toronto. *Bioinformatics* 2009;25(24):3327-3329.

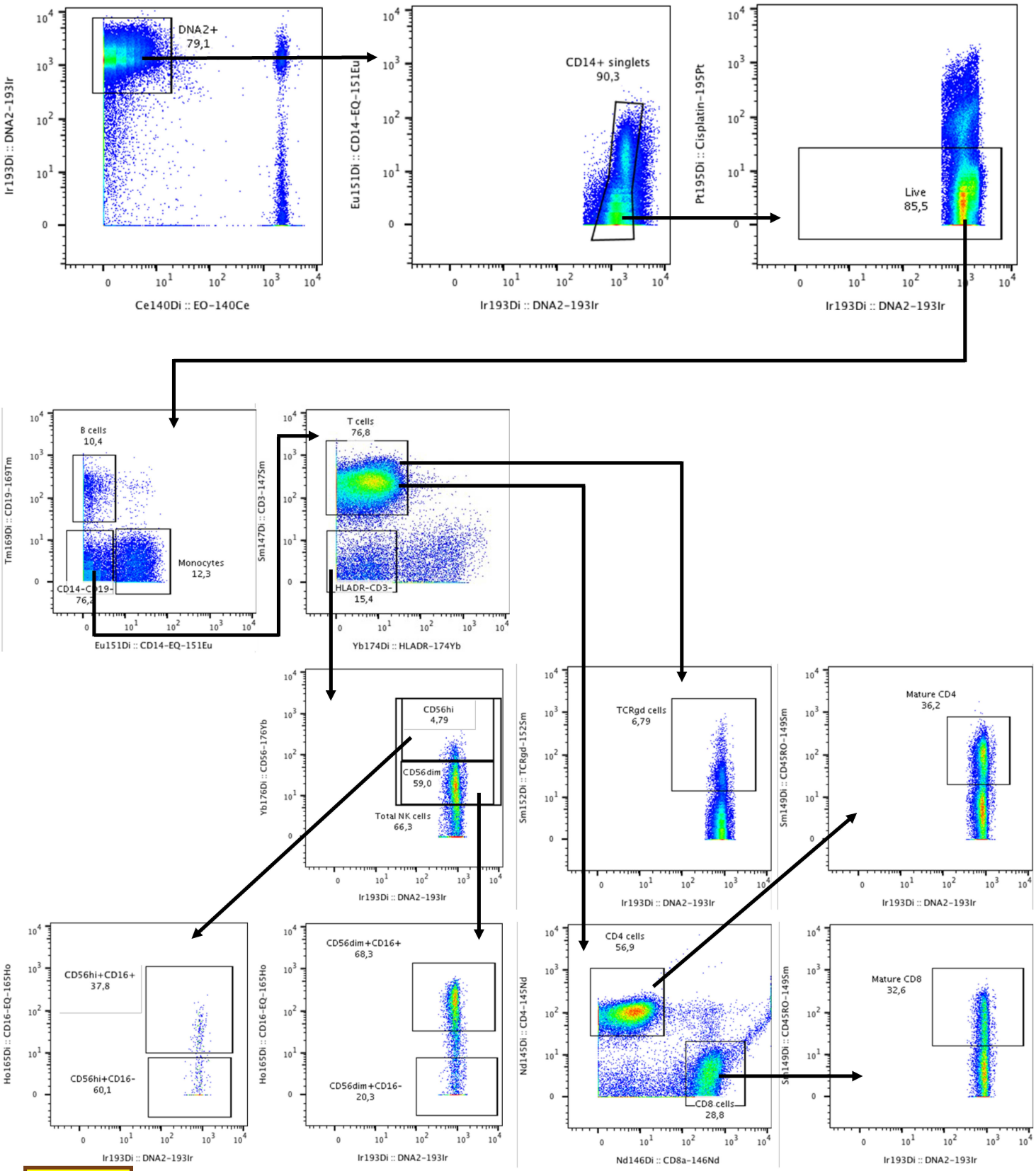
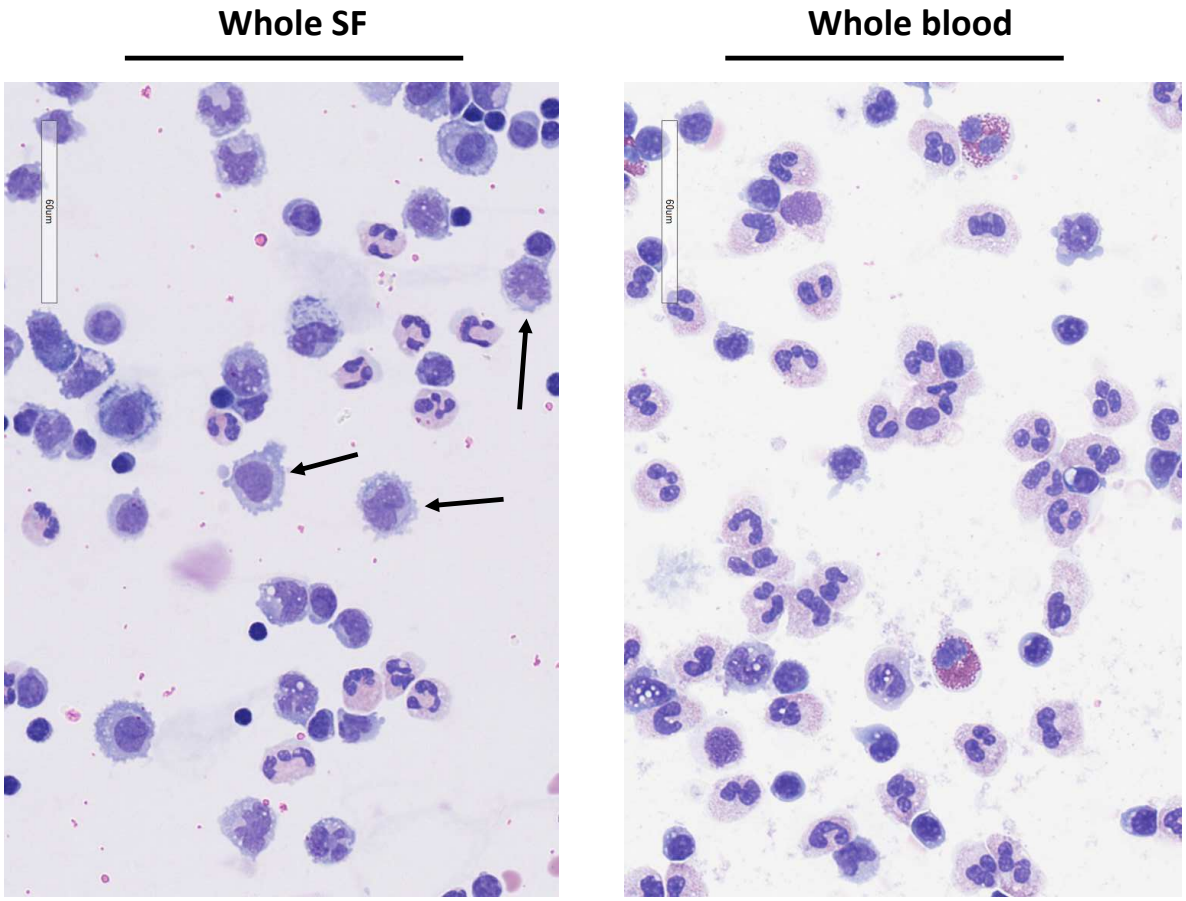

DNA2+
79,1
Ir193Di :: DNA2-193Ir
Ce140Di :: EQ-140Ce
CD14+ singlets
90,3
Eu151Di :: CD14-EQ-151Eu
Ir193Di :: DNA2-193Ir
Live
85,5
Pt195Di :: Cisplatin-195Pt
Ir193Di :: DNA2-193Ir
B cells
10,4
Monocytes
12,3
CD14-CD19-
76,2
Tm169Di :: CD19-169Tm
Eu151Di :: CD14-EQ-151Eu
T cells
76,8
HLADR-CD3-
15,4
Sm147Di :: CD3-147Sm
Yb174Di :: HLADR-174Yb
CD56hi
4,79
CD56dim
59,0
Total NK cells
66,3
Yb176Di :: CD56-176Yb
Ir193Di :: DNA2-193Ir
TCRgd cells
6,79
Sm152Di :: TCRgd-152Sm
Ir193Di :: DNA2-193Ir
Mature CD4
36,2
Sm149Di :: CD45RO-149Sm
Ir193Di :: DNA2-193Ir
CD56hi+CD16+
37,8
CD56hi+CD16-
60,1
Ho165Di :: CD16-EQ-165Ho
Ir193Di :: DNA2-193Ir
CD56dim+CD16+
68,3
CD56dim+CD16-
20,3
Ho165Di :: CD16-EQ-165Ho
Ir193Di :: DNA2-193Ir
CD4 cells
56,9
CD8 cells
28,8
Nd145Di :: CD4-145Nd
Nd146Di :: CD8a-146Nd
Mature CD8
32,6
Sm149Di :: CD45RO-149Sm
Ir193Di :: DNA2-193Ir

Figure S1

Figure S2



Qaiyum Z, *et al. Ann Rheum Dis* 2019; 0:1566–1575. doi: 10.1136/annrheumdis-2019-215349

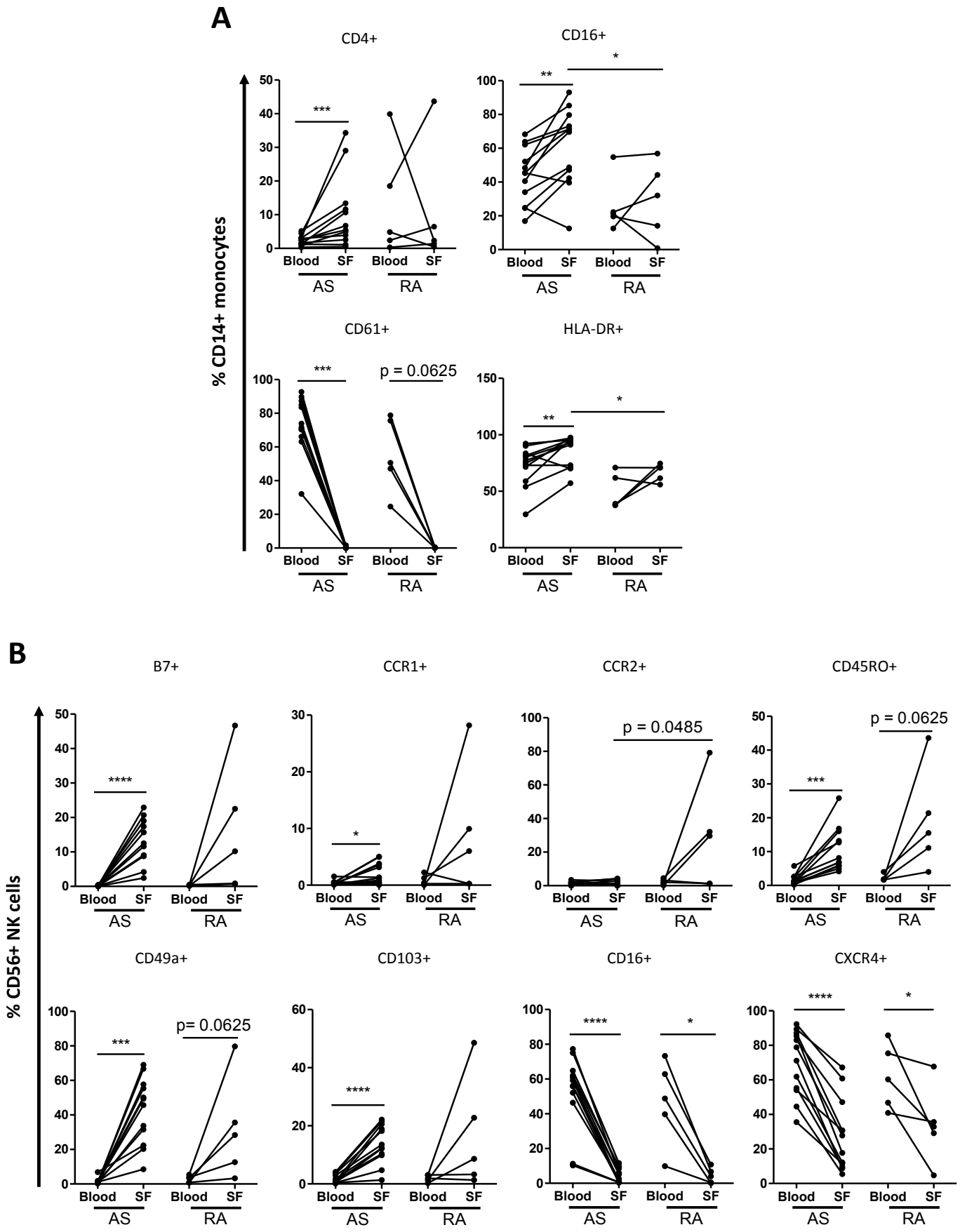

A
CD4+
CD16+
CD61+
HLA-DR+
% CD14+ monocytes
Blood
SF
AS
RA
p = 0.0625
B
B7+
CCR1+
CCR2+
CD45RO+
CD49a+
CD103+
CD16+
CXCR4+
% CD56+ NK cells
p = 0.0485
p= 0.0625

Figure S3

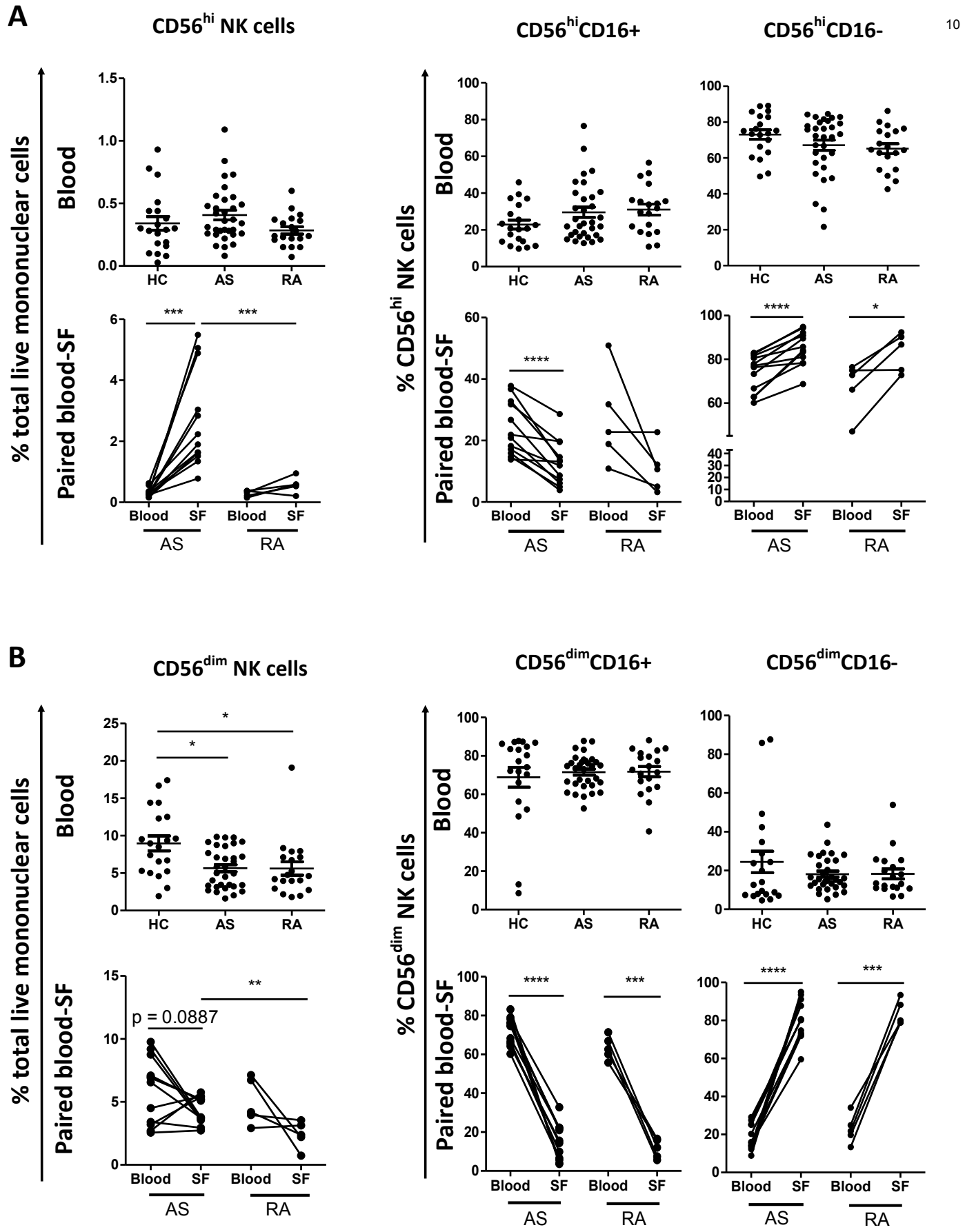

A
CD56hi NK cells
CD56hiCD16+
CD56hiCD16-
% total live mononuclear cells
Blood
Paired blood-SF
% CD56hi NK cells
HC
AS
RA
Blood
SF
AS
RA
***
****
*
B
CD56dim NK cells
CD56dimCD16+
CD56dimCD16-
% CD56dim NK cells
p = 0.0887
**

Figure S4

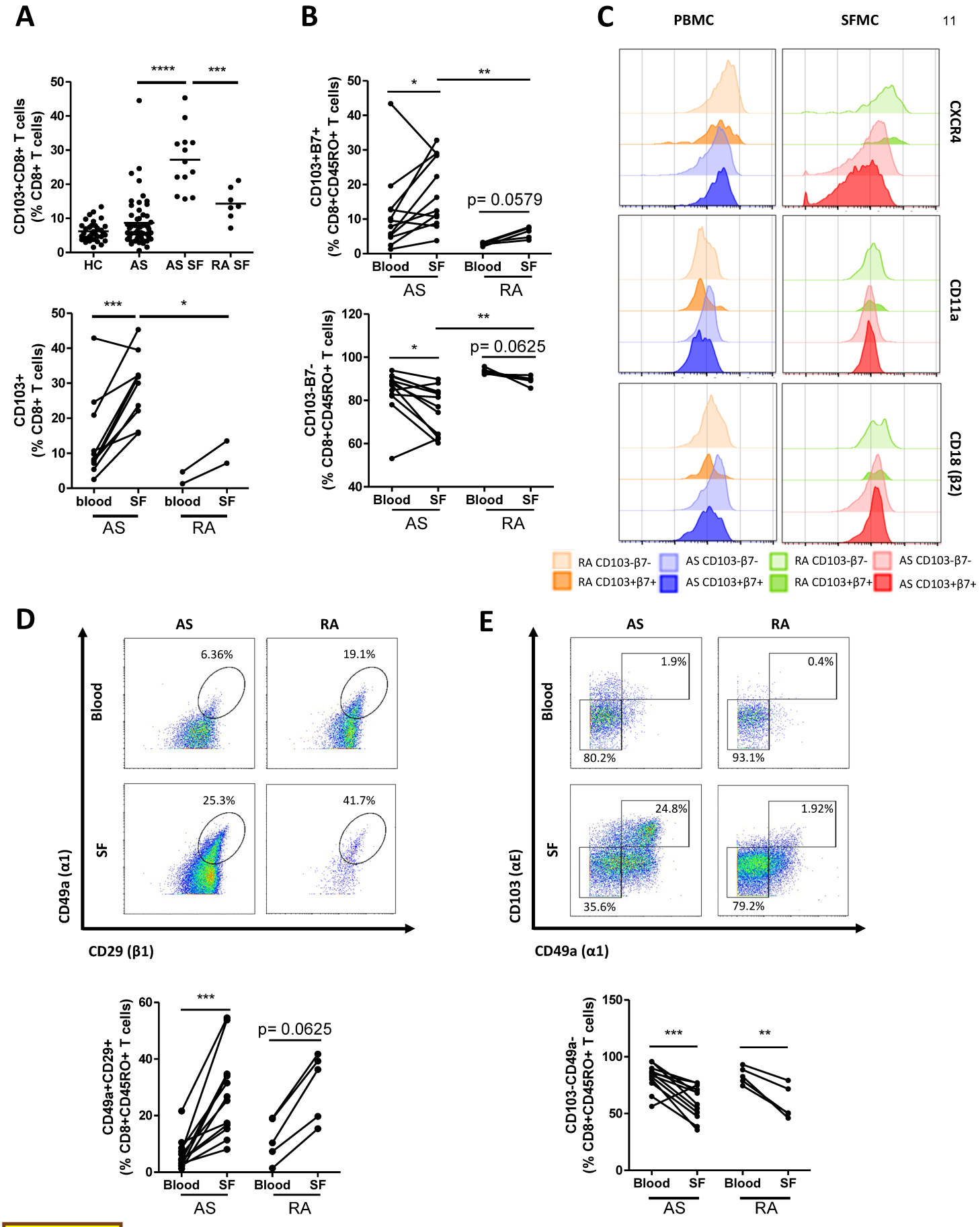

A
CD103+CD8+ T cells (% CD8+ T cells)
HC
AS
AS SF
RA SF
CD103+ (% CD8+ T cells)
blood
SF
AS
RA
B
CD103+B7+ (% CD8+CD45RO+ T cells)
p= 0.0579
CD103-B7- (% CD8+CD45RO+ T cells)
p= 0.0625
Blood
SF
C
PBMC
SFMC
CXCR4
CD11a
CD18 (β2)
RA CD103-β7-
RA CD103+β7+
AS CD103-β7-
AS CD103+β7+
D
AS
RA
Blood
SF
6.36%
19.1%
25.3%
41.7%
CD49a (α1)
CD29 (β1)
CD49a+CD29+ (% CD8+CD45RO+ T cells)
p= 0.0625
E
1.9%
0.4%
80.2%
93.1%
24.8%
1.92%
35.6%
79.2%
CD103 (αE)
CD49a (α1)
CD103-CD49a- (% CD8+CD45RO+ T cells)

Figure S5

A

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.250[a]	.062	.000	5.78134	.062	.999	1	15	.333

a. Predictors: (Constant), Biologics

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	33.392	1	33.392	.999	.333[b]
	Residual	501.359	15	33.424		
	Total	534.751	16			

a. Dependent Variable: lnEx

b. Predictors: (Constant), Biologics

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	7.392	1.828		4.043	.001	3.495	11.289
	Biologics	-2.848	2.849	-.250	-1.000	.333	-8.920	3.225

a. Dependent Variable: lnEx

B

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.217[a]	.047	.027	.608	.047	2.366	1	48	.131

a. Predictors: (Constant), Biologics

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	.873	1	.873	2.366	.131[b]
	Residual	17.720	48	.369		
	Total	18.594	49			

a. Dependent Variable: lnEx

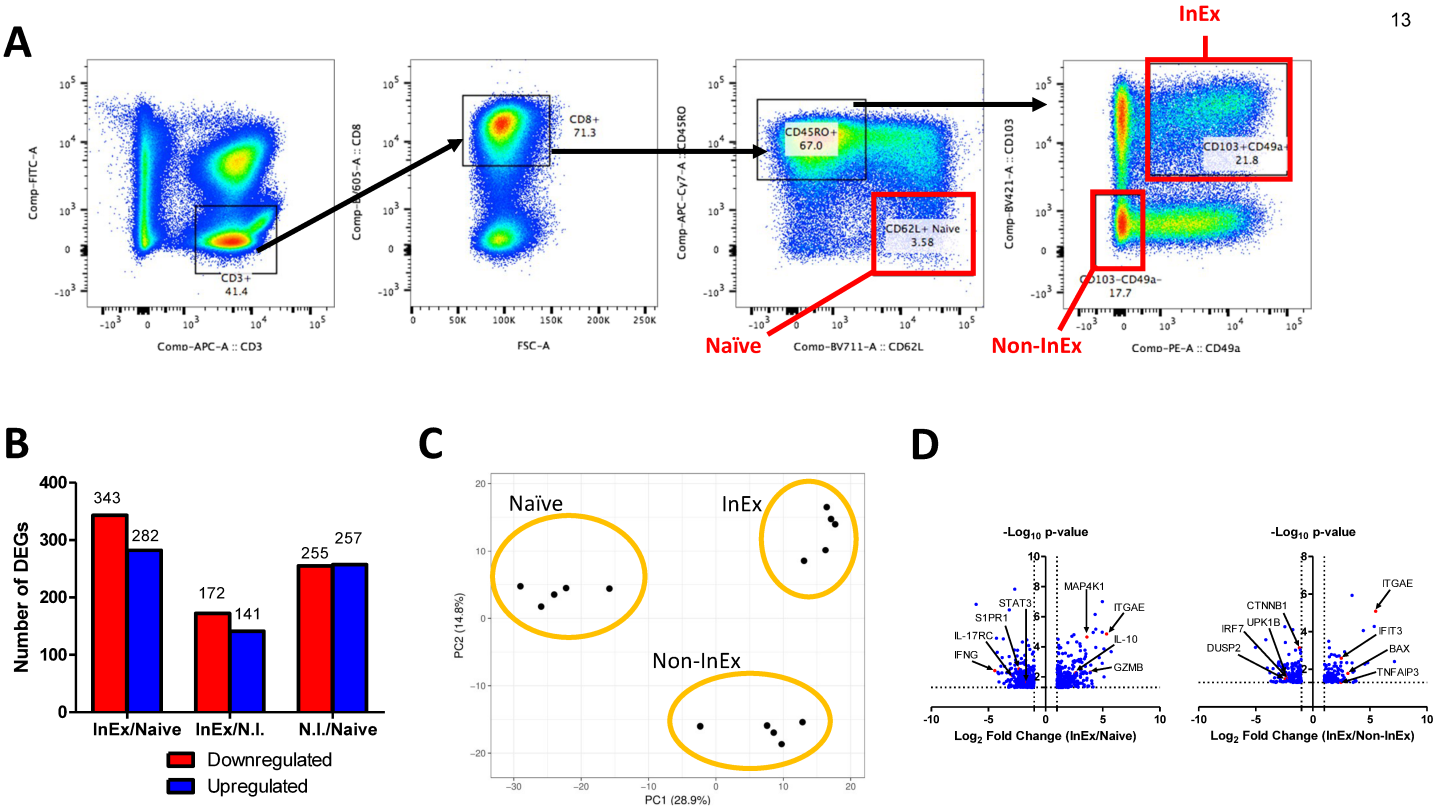
b. Predictors: (Constant), Biologics

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	.681	.122		5.603	.000	.437	.925
	Biologics	.264	.172	.217	1.538	.131	-.081	.610

a. Dependent Variable: lnEx

Figure S6


A
InEx
Comp-FITC-A
Comp-APC-A :: CD3
CD3+
41.4
Comp-BV605-A :: CD8
FSC-A
CD8+
71.3
Comp-APC-Cy7-A :: CD45RO
CD45RO+
67.0
CD62L+ Naive
3.58
Naïve
Comp-BV711-A :: CD62L
Comp-BV421-A :: CD103
CD103+CD49a+
21.8
CD103-CD49a-
17.7
Non-InEx
Comp-PE-A :: CD49a
B
Number of DEGs
343
282
172
141
255
257
InEx/Naive
InEx/N.I.
N.I./Naive
Downregulated
Upregulated
C
Naïve
InEx
Non-InEx
PC2 (14.8%)
PC1 (28.9%)
D
-Log10 p-value
MAP4K1
ITGAE
STAT3
S1PR1
IL-17RC
IL-10
IFNG
GZMB
Log2 Fold Change (InEx/Naive)
-Log10 p-value
ITGAE
CTNNB1
UPK1B
IRF7
IFIT3
DUSP2
BAX
TNFAIP3
Log2 Fold Change (InEx/Non-InEx)

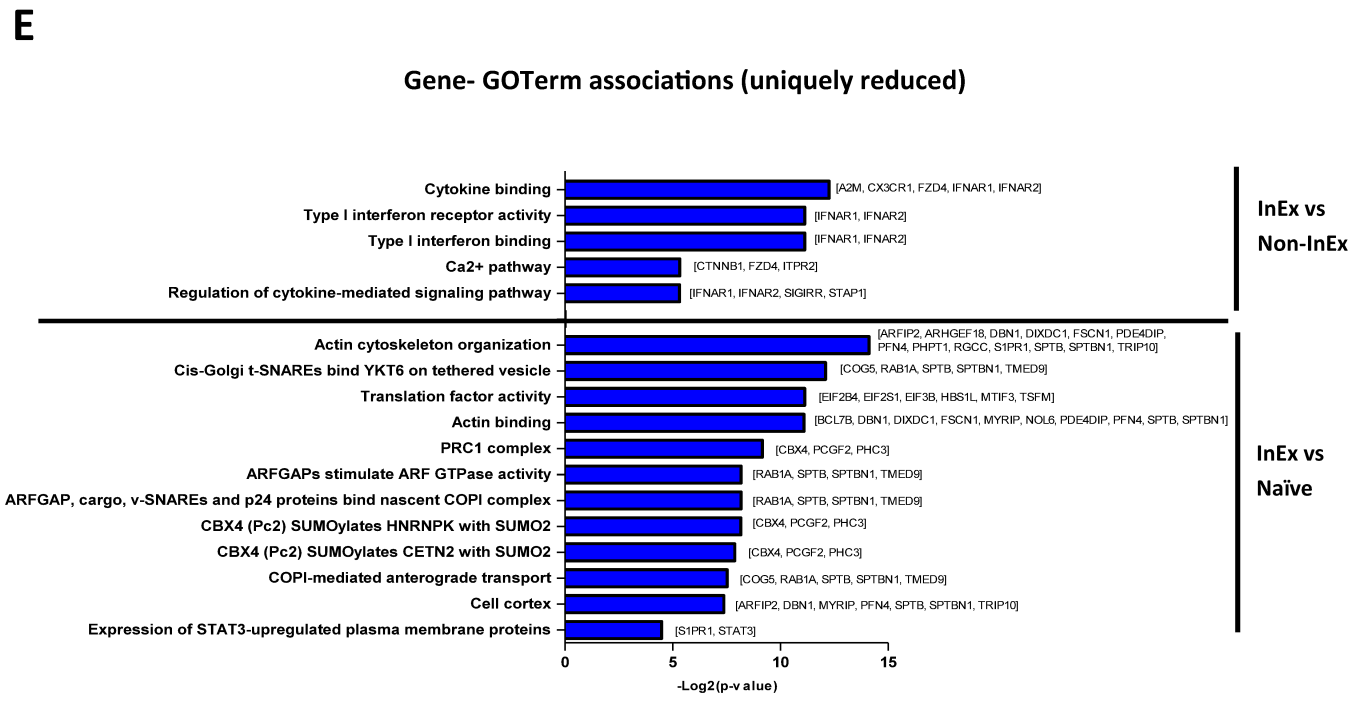

E
Gene- GOTerm associations (uniquely reduced)
Cytokine binding [A2M, CX3CR1, FZD4, IFNAR1, IFNAR2]
Type I interferon receptor activity [IFNAR1, IFNAR2]
Type I interferon binding [IFNAR1, IFNAR2]
Ca2+ pathway [CTNNB1, FZD4, ITPR2]
Regulation of cytokine-mediated signaling pathway [IFNAR1, IFNAR2, SIGIRR, STAP1]
InEx vs Non-InEx
Actin cytoskeleton organization [ARFIP2, ARHGEF18, DBN1, DIXDC1, FSCN1, PDE4DIP, PFN4, PHPT1, RGCC, S1PR1, SPTB, SPTBN1, TRIP10]
Cis-Golgi t-SNAREs bind YKT6 on tethered vesicle [COG5, RAB1A, SPTB, SPTBN1, TMED9]
Translation factor activity [EIF2B4, EIF2S1, EIF3B, HBS1L, MTIF3, TSFM]
Actin binding [BCL7B, DBN1, DIXDC1, FSCN1, MYRIP, NOL6, PDE4DIP, PFN4, SPTB, SPTBN1]
PRC1 complex [CBX4, PCGF2, PHC3]
ARFGAPs stimulate ARF GTPase activity [RAB1A, SPTB, SPTBN1, TMED9]
ARFGAP, cargo, v-SNAREs and p24 proteins bind nascent COPI complex [RAB1A, SPTB, SPTBN1, TMED9]
CBX4 (Pc2) SUMOylates HNRNPK with SUMO2 [CBX4, PCGF2, PHC3]
CBX4 (Pc2) SUMOylates CETN2 with SUMO2 [CBX4, PCGF2, PHC3]
COPI-mediated anterograde transport [COG5, RAB1A, SPTB, SPTBN1, TMED9]
Cell cortex [ARFIP2, DBN1, MYRIP, PFN4, SPTB, SPTBN1, TRIP10]
Expression of STAT3-upregulated plasma membrane proteins [S1PR1, STAT3]
InEx vs Naïve
0
5
10
15
-Log2(p-value)

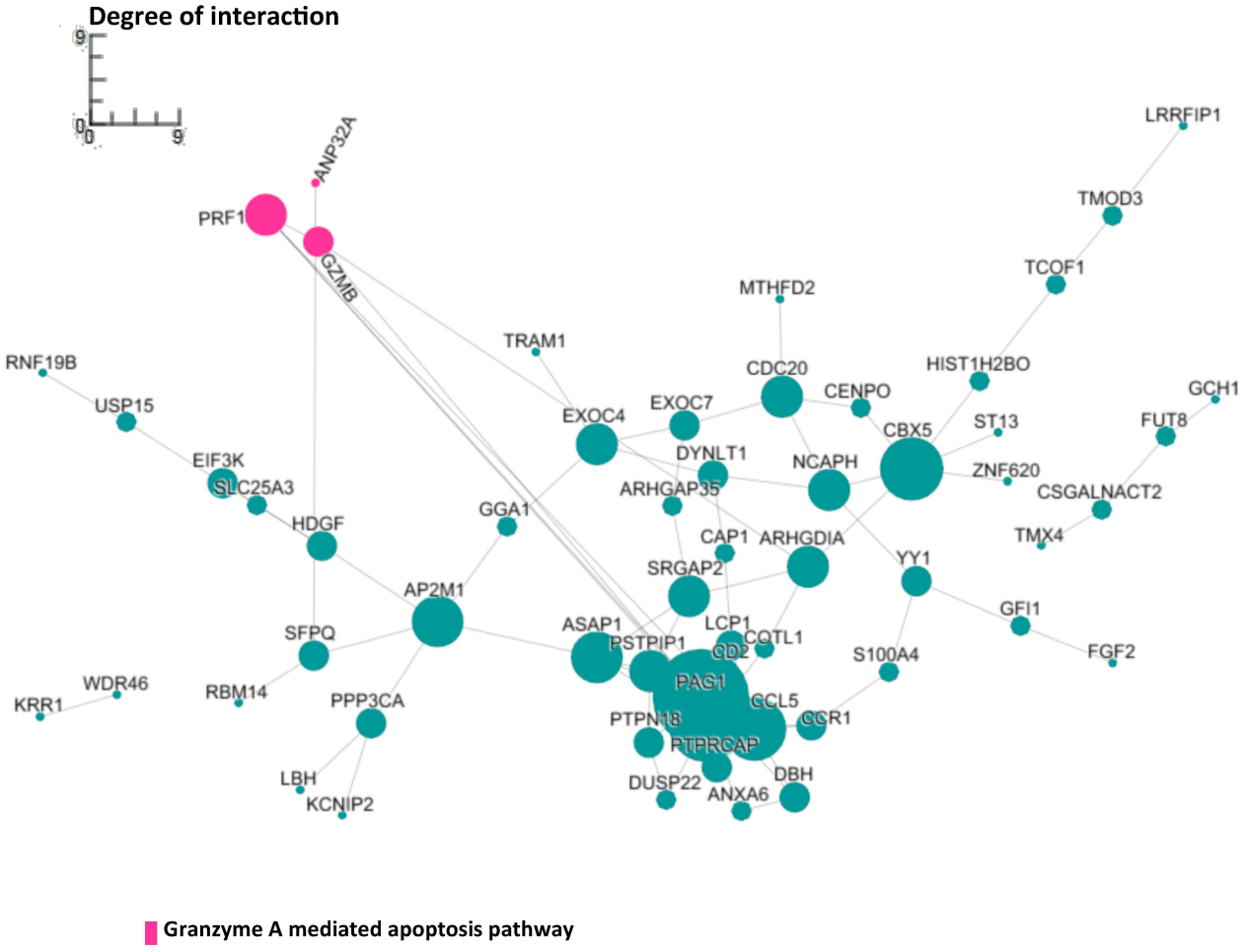
Figure S7

A

Protein-protein interaction network (InEx vs Non-InEx cells)

B

Protein-protein interaction network (InEx vs Naïve cells)



Granzyme A mediated apoptosis pathway

Figure S8

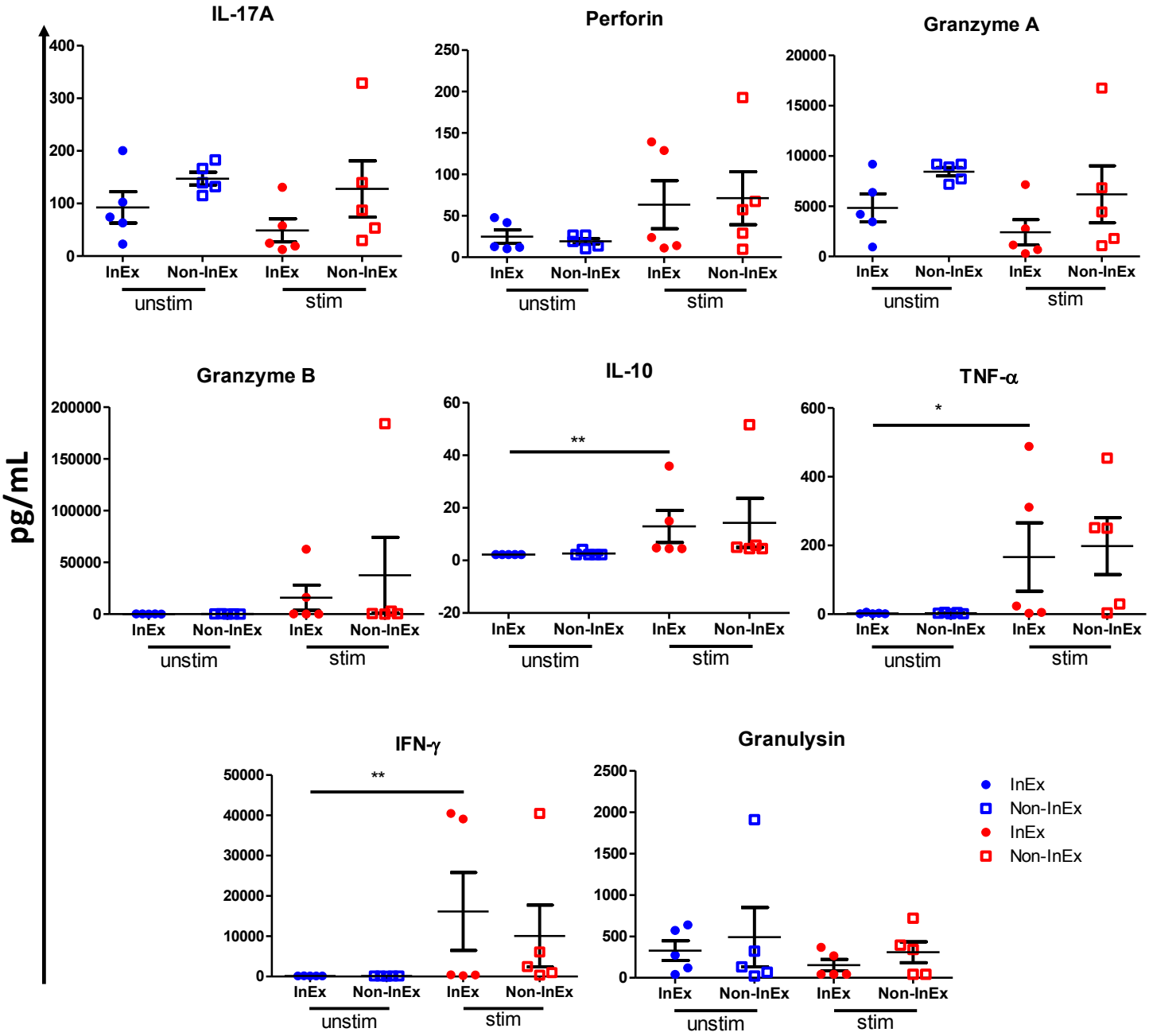

IL-17A
Perforin
Granzyme A
Granzyme B
IL-10
TNF-α
IFN-γ
Granulysin
pg/mL
InEx
Non-InEx
unstim
stim
*
**
InEx
Non-InEx
InEx
Non-InEx

Figure S9

	HC	AS	AS-SF	RA	RA-SF
Sex (M/F)	20/-	29/3	9/3	7/12	2/3
Age (yr ± SD)	25 ± 5	30 ± 9	34 ± 12	57 ± 12	64 ± 15
HLA-B27+ status	-	81%	92%	-	-
Mean age of backpain (yr ± SD)	-	16 ± 6	17 ± 8	-	-
Mean age at diagnosis (yr ± SD)	-	18 ± 8	19 ± 10	-	-
Mean ESR (mm/h ± SD)	-	13 ± 14	19 ± 17	22 ± 19	33 ± 21
Mean CRP (mg/L ± SD)	-	13 ± 19	17 ± 15	10 ± 13	20 ± 20
BASDAI (mean ± SD)	-	3 ± 3	4 ± 3	-	-
Extra-articular manifestations					
Skin (psoriasis)	-	13%	17%	-	-
Gastrointestinal	-	6%	0%		
Biologics					
anti-TNF	-	47%	33%	56%	60%
anti-IL-17	-	0%	0%	0%	0%
anti-Integrin	-	0%	0%	0%	0%
NSAIDs	-	75%	83%	42%	40%

Table S1

				Manufacturer	
	Antibody Target-Metal Isotope	**Clone**	**Dilution Factor**	**Fluidigm**	**BioLegend (custom conjugation)**
Immune markers	CD16- 165Ho	3G8	1:50	✓	
	CD14- 151Eu	M5E2	1:50	✓	
	CD19- 169Tm	HIB19	1:100	✓	
	CD3- 147Sm	UCHT1	1:100		✓
	CD4- 145Nd	RPA-T4	1:50	✓	
	CD8α- 146Nd	RPA-T8	1:100	✓	
	TCRγδ- 152Sm	11F2	1:50	✓	
	CD25- 171Yb	BC96	1:50		✓
	CD56- 176Yb	HCD56	1:50	✓	
	HLADR- 174Yb	L243	1:100	✓	
Activating/identifying markers	CD122- 170Er	Tu27	1:50	✓	
	CD11a- 142Nd	HI111	1:50	✓	
	CD11b- 168Er	ICRF44	1:100		✓
	CD7- 143Nd	CD7-6B7	1:100		✓
	CD11c- 167Er	Bu15	1:100		✓
	CD161- 159Tb	HP-3G10	1:50	✓	
	CD45RO- 149Sm	UCHL1	1:100		✓
	CD62L- 164Dy	DREG-56	1:200		✓
Chemokine receptors	CCR2- 153Eu	K036C2	1:100		✓
	CCR1- 175Lu	5F10B29	1:50		✓
	CCR5- 144Nd	NP-6G4	1:50	✓	
	CCR6- 141Pr	G034E3	1:50	✓	
	CCR9- 154Sm	L053E8	1:50		✓
	CXCR3- 163Dy	G025H7	1:100	✓	
	CXCR4- 173Yb	12G5	1:200	✓	
	CXCR6- 172Yb	K041E5	1:50		✓
	CX3CR1- 161Dy	2A9-1	1:50		✓
Integrins	CD49a(α1)- 150Nd	TS2/7	1:50		✓
	CD18(β2)- 158Gd	TS1/18	1:200		✓
	CD49b(α2)- 166Er	P1E6-C5	1:50		✓
	CD103(αE)- 162Dy	Ber-ACT8	1:100		✓
	CD29(β1)- 156Gd	TS2/16	1:50	✓	
	CD66a- 160Gd	ASL-32	1:50		✓
	β7- 148Nd	FIB504	1:50		✓
	CD41(α2b)- 89Y	HIP8	1:50	✓	
	CD61- 209Bi	VI-PL2	1:50	✓	

Table S2

			Manufacturer	
Antibodiy Target– Fluorochrome	**Clone**	**Concentration (µg/mL)**	**BioLegend**	**BD Biosciences**
CD4– FITC	RPA-T4	10	✓	
CD11c– FITC	Bu15	10	✓	
CD14– FITC	M5E2	10	✓	
CD16– FITC	3G8	6	✓	
CD20– FITC	2H7	2.5	✓	
TCRγδ– FITC	B1	15	✓	
CD3– APC	UCHT1	0.4	✓	
CD8– BV605	RPA-T8	0.625	✓	
CD45RO– APC/Cy7	UCHL1	10	✓	
CD62L– BV711	SK11	2.5		✓
CD103– BV421	Ber-ACT8	1.75	✓	
CD49a– PE	TS2/7	5	✓	

Table S3

SUPPLEMENTARY FIGURE LEGENDS

Supplementary Figure 1 (Figure S1). Gating strategy for live and major cell populations. Gating strategy displayed for live cells. FCS files acquired from CyTOF2 were subject to DNA2 vs. normalization bead (EQ-140Ce) discrimination to distinguish between cells vs. beads. Following that, DNA2+ cells underwent doublet discrimination using CD14 as CD14+ monocytes are larger cells compared to lymphocytes and would have been missed on a traditional DNA2 vs. DNA1 plot to identify single cells. Lack of cisplatin staining further identified live cells. Then, various immune cells were gated on live cells for identification of major cell populations.

Supplementary Figure 2 (Figure S2). Visualization of macrophages in AS SF. Whole blood from AS patient treated with RBC lysis buffer, and whole SF from unmatched patient used to prepare cytospin slides stained with Wright-Giemsa stain. Slides were digitized at 40x magnification and Aperio ImageScope v12.3.2.8013 software used to generate snapshots. Arrows indicate macrophages in whole SF.

Supplementary Figure 3 (Figure S3). Heatmap statistics of monocytes and NK cells. A) CD4, CD16, CD61 and HLA-DR expression on CD14+ monocytes from paired blood-SF samples assessed. CD4+CD14+, CD61+CD14+ and HLA-DR+CD14+ paired samples analyzed by Wilcoxon matched pairs test. CD16+CD14+ paired samples analyzed by paired T test. AS SF and RA SF analyzed by Mann-Whitney test. **** $p<0.0001$, *** $p<0.001$, ** $p<0.01$, * $p<0.05$, n.s. $p>0.05$ (non-significant results not highlighted on graphs, unless indicated)**.** B) β7, CCR1, CCR2, CD45RO, CD49a, CD103, CD16 and CXCR4 expression on CD56+ NK cells from paired blood-SF samples assessed. β7+CD56+, CCR2+CD56+, CD103+CD56+, CD16+CD56+ and CXCR4+CD56+ paired samples analyzed by paired T test. CCR1+CD56+, CD45RO+CD56+ and CD49a+CD56+ paired samples analyzed by Wilcoxon matched pairs test. AS SF and RA SF analyzed by Mann-Whitney test. **** $p<0.0001$, *** $p<0.001$, ** $p<0.01$, * $p<0.05$, n.s. $p>0.05$ (non-significant results not highlighted on graphs, unless indicated)**.**

Supplementary Figure 4 (Figure S4). $CD56^{hi}$ and $CD56^{dim}$ NK cells in blood and synovial fluid exhibit characteristic frequencies and CD16 profiles. NK cells were divided into two major subgroups, $CD56^{hi}$ and $CD56^{dim}$ NK cells, and CD16 expression was analyzed. A) $CD56^{hi}$ NK cells blood data analyzed by Kruskal Wallis test with Dunns post-test, and paired samples analyzed by paired T test. $CD56^{hi}$CD16+ and $CD56^{hi}$CD16- NK cells blood data analyzed by Kruskal Wallis test with Dunns post-test, and paired samples analyzed by paired T test. AS-SF and RA-SF analyzed by Mann-Whitney test. **** $p<0.0001$, *** $p<0.001$, ** $p<0.01$, * $p<0.05$, n.s. $p>0.05$ (non-significant results not highlighted on graphs, unless indicated). B) $CD56^{dim}$ NK cells blood data analyzed by Kruskal Wallis test with Dunns post-test, and paired samples analyzed by paired T test. $CD56^{dim}$CD16+ and $CD56^{dim}$CD16- NK cells blood data analyzed by

Kruskal Wallis test with Dunns post-test, and paired samples analyzed by paired T test. AS-SF and RA-SF analyzed by Mann-Whitney test. **** p<0.0001, *** p<0.001, ** p<0.01, * p<0.05, n.s. p>0.05 (non-significant results not highlighted on graphs, unless indicated).

Supplementary Figure 5 (Figure S5). Extended characterization of CD103+CD8+ T cells in synovial fluid. A) CD103+ expression on CD8+ T cells from an independent cohort validated using flow cytometry. HC and AS PBMC, AS PBMC and AS SFMC, and AS SFMC and RA SFMC analyzed by Mann Whitney test instead of Kruskal Wallis test (with Dunns post-test) to minimize effect of non-comparative groups. Paired blood-SF data from AS and RA patients also shown and analyzed by paired T test. **** p<0.0001, *** p<0.001, ** p<0.01, * p<0.05 (non-significant results not highlighted on graph, unless indicated). B) Graphs depicting CD103+β7+ and CD103-β7- expression on mature CD8 T cells from paired samples shown. CD103+β7+ and CD103-β7- paired data analyzed by Wilcoxon matched pairs test. AS SF and RA SF analyzed by Mann-Whitney test. **** p<0.0001, *** p<0.001, ** p<0.01, * p<0.05, n.s.>0.05. C) Representative plots of CD103+β7+ or CD103-β7- cells from mature CD8+ T cells on the basis of their CXCR4, CD11a and CD18 expression in AS or RA PBMC and SFMC. D) Plots from representative paired PBMC and SFMC from AS and RA patients shown, on the basis of CD49a and CD29 expression patterns gated on mature CD8+ T cells. CD49a+CD29+ expression on mature CD8 T cells from paired samples graphed. CD49a+CD29+ paired samples analyzed by Wilcoxon matched pairs test. AS SF and RA SF analyzed by Mann-Whitney test. *** p<0.001 (non-significant results not highlighted on graph, unless indicated). E) Plots from representative paired PBMC and SFMC from AS and RA patients shown, on the basis of CD103 and CD49a expression patterns gated on mature CD8 T cells. Graph depicting CD103-CD49a- cells in paired samples, analyzed by paired T test. AS SF and RA SF analyzed by Mann-Whitney test. *** p<0.001, ** p<0.01.

Supplementary Figure 6 (Figure S6). Regression analysis on SFMC and PBMC. Regression analysis was performed on A) SFMC and B) PBMC using all AS or RA patient samples using IBM SPSS Statistics v22.0.0 platform, to test influence of TNF inhibitory therapy and diagnosis on InEx cell frequency. Dependent variable was InEx cell frequency and independent variables were biologic status. Displayed in this figure is the resulting output data, specifically the model summary, F ratio and unstandardized coefficients.

Supplementary Figure 7 (Figure S7). Extended RNA-seq data of InEx, Non-InEx and naïve cells. A) Gating strategy of InEx, Non-InEx and naïve cells for FACS sorting. Prior to RNA-seq analysis, liquid N_2 frozen SFMCs were thawed, incubated overnight, and stained to FACS sort the InEx, Non-InEx and naïve cell populations (highlighted in red boxes) on the basis of their indicated markers. Subsequent to sorting, RNA from each population was collected and then subject to quality check, library preparation and bulk RNA-sequencing. B) Differentially expressed genes (DEGs) in each comparison group displayed, depicting number of genes that

were either upregulated or downregulated in InEx vs. naïve, InEx vs. Non-InEx and Non-InEx vs. naïve. C) Based on the $\log_{10}$ (FPKM+1) values of upregulated and downregulated genes that were subject to fold change greater than or less than 2 and $p<0.05$ cut off, PCA plots were generated. D) After filtering using aforementioned cut offs, DEGs were converted into volcano plots to E) Pathway analysis of downregulated genes unique to InEx vs Non-InEx and InEx vs naïve. 197 genes were differentially regulated in InEx vs Non-InEx cells and 353 genes differentially regulated in InEx vs naïve cells. Out of these, 117 genes and 203 genes were uniquely downregulated, respectively, and were fed into Cytoscape and ClueGo program selected. All pathways/ontologies were selected, with network specificity 'medium' and $p<0.05$. Resulting data was converted into a graph to display pathway analysis of downregulated genes.

Supplementary Figure 8 (Figure S8). Protein-protein interaction network from unique and elevated genes. A) 80 unique and elevated genes in InEx vs Non-InEx cells, and B) 150 unique and elevated genes in InEx vs naïve cells were submitted to the IID platform, from which corresponding protein-protein interactions (PPIs) and degree of interaction for each gene was obtained. These were then input into the NAViGaTOR program, and only those proteins with degree of interactions ≥ 1 were selected for ease of visualization. The size of the node indicates degree of interaction with other proteins amongst the query proteins. Enriched pathways from pathDIP program highlighted in colour.

Supplementary Figure 9 (Figure S9). Cytokine multiplex assay on unstimulated and stimulated InEx and Non-InEx cells. FACS sorted InEx and Non-InEx cells from AS SFMCs (n=5) were stimulated with anti-CD2/CD3/CD28 T cell activation beads and 500U/mL IL-2, or unstimulated for 72hrs/37°C. Supernatants isolated and processed using the LEGENDplex Human CD8/NK panel assay and analyzed using manufacturer's analysis software. Concentrations (pg/mL) of select cytokines plotted as seen in figure.

Supplementary Table 1 (Table S1). Cohort demographics. AS and RA patient and healthy control clinical profiles. Where indicated, values given as mean ± SD. Where applicable, "-" indicates data not collected.

Supplementary Table 2 (Table S2). CyTOF panel of antibodies. Metal isotope-tagged antibodies with their respective clones and dilution factor listed. Antibodies were either purchased directly from Fluidigm or purchased from BioLegend and custom conjugated with Fluidigm kit as indicated.

Supplementary Table 3 (Table S3). Antibodies for FACS sorting. List of 12 antibodies with their respective clones and concentrations (μg/mL) used to FACS sort InEx, Non-InEx and naïve cell populations. CD62L-BV711 was purchased from BD Biosciences, while the rest of the antibodies were purchased from BioLegend as indicated.