

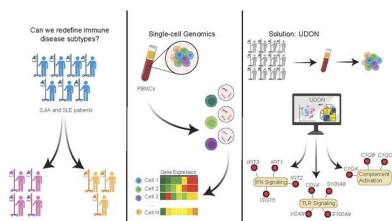
Population-level single-cell genomics reveals conserved gene programs in systemic juvenile idiopathic arthritis

Emely L. Verweyen, ... , Nathan Salomonis, Grant S. Schulert

J Clin Invest. 2023. <https://doi.org/10.1172/JCI166741>.

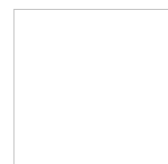
Research In-Press Preview Inflammation

Graphical abstract



Find the latest version:

<https://jci.me/166741/pdf>



Population-level single-cell genomics reveals conserved gene programs in systemic juvenile idiopathic arthritis

Emely L. Verweyen^{*1}, Kairavee Thakkar^{*2,3}, Sanjeev Dhakal¹, Elizabeth Baker¹, Kashish Chetal², Daniel Schnell², Scott Canna⁴, Alexei A. Grom^{1,5}, Nathan Salomonis^{2,5†}, Grant S. Schulert^{1,5,†}

*These authors share first authorship

¹Cincinnati Children's Hospital Medical Center, Division of Rheumatology, Cincinnati, Ohio, USA

²Cincinnati Children's Hospital Medical Center, Division of Biomedical Informatics, Cincinnati, Ohio, USA

³Department of Pharmacology and Systems Physiology, University of Cincinnati College of Medicine, Cincinnati, Ohio, USA

⁴Children's Hospital Philadelphia, Division of Rheumatology, Philadelphia, PA, USA

⁵Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, Ohio, USA

†Corresponding authors:

Nathan Salomonis, Nathan.Salomonis@cchmc.org, Cincinnati Children's Hospital Medical Center, Division of Biomedical Informatics, Cincinnati, Ohio, USA, 513-636-4200 and

Grant Schulert, Grant.Schulert@cchmc.org, Cincinnati Children's Hospital Medical Center, Division of Rheumatology, Cincinnati, Ohio, USA, 513-636-3894

Conflict-of-interest statement:

Dr. Grom reports research support and consulting fees from Novartis, SOBI, and AB2Bio. Dr. Schulert reports consulting fees from Novartis and SOBI and research support from IpiNovyx. Dr. Canna has received consulting fees from SOBI, Apollo, and Simcha. All other authors have declared that no conflict of interest exists.

Abstract

Systemic autoimmune and autoinflammatory diseases are characterized by genetic and cellular heterogeneity. While current single-cell genomics methods provide insights into known disease subtypes, these analysis methods do not readily reveal novel cell-type perturbation programs shared amongst distinct patient subsets. Here, we performed single-cell RNA-Seq of PBMCs of systemic juvenile idiopathic arthritis (SJIA) patients with diverse clinical manifestations, including macrophage activation syndrome (MAS) and lung disease (LD). We introduced two new computational frameworks called UDON and SATAY-UDON which define new patient subtypes based on their underlying disrupted cellular programs as well as associated biomarkers or clinical features. Among twelve independently identified subtypes, this analysis uncovered a novel complement and interferon activation program identified in SJIA-LD monocytes. Extending these analyses to adult and pediatric lupus patients found new but also shared disease programs with SJIA, including interferon and complement activation. Finally, supervised comparison of these programs in a compiled single-cell pan-immune atlas of over 1,000 healthy donors found a handful of normal healthy donors with evidence of early inflammatory activation in subsets of monocytes and platelets, nominating new possible biomarkers for early disease detection. Thus, integrative pan-immune single-cell analysis resolved new conserved gene programs underlying inflammatory disease pathogenesis and associated complications.

Introduction

Diseases of the immune system associated with systemic autoimmunity and autoinflammation can result in significant health burdens, which in many disorders include a high-risk for life-threatening complications. Inflammatory diseases are believed to evolve from complex underlying etiologies, including both genetic and environmental, and include common disorders such as type I diabetes to relatively rare but serious disorders such as Systemic juvenile idiopathic arthritis (SJIA). The immunopathogenesis of SJIA is multifaceted, with features of myeloid activation, autoimmunity, classical autoinflammation, and interferonopathy linked to disease heterogeneity and complications (1). Recently, single-cell RNA-Sequencing (scRNA-Seq) has elucidated underlying gene expression and genetic associations, through independent population-level and focused diseased cohort analyses in disorders such as systemic lupus erythematosus (SLE) (2). Such analyses have led to important insights into gene expression impacts in myeloid and lymphoid cell populations that are associated with common genetic variation, when considering nearly a thousand healthy donors (3).

While single-cell genomics provides the opportunity to understand cellular, molecular and genetic associations among patients and controls, existing approaches are focused on harmonization of cells across patients without considering novel patient subsets associated with distinct gene regulatory programs (4). Such analyses are crucial in systemic inflammatory disorders such as SJIA and SLE, in which patients are characterized by diverse disease states associated with different inflammatory signaling and immune-cell specific impacts, that often remain largely unknown. Indeed, given the large incidence of autoimmune disease within the population (~7%), it is likely that many presumably healthy patients have underlying autoimmune dysfunction that has not yet manifested in diagnosed disease. Hence, new integrative analysis and strategies that leverage large cohorts of controls and patient samples from distinct inflammatory disorders are required to begin understanding the novel common and unique autoimmune cellular programs and their functional relationship to patient phenotypes.

To address these challenges, we developed a new computational framework designed to exploit differences in the gene expression programs of individual patients versus presumed healthy controls, at the level of individual cell-type. This workflow, Unsupervised Discovery Of Novel disease programs or UDON, is designed to discover both known and new disease subtypes associated with coherent gene-expression differences unique to a subset of patients. These subtypes comprise both patient samples and cell-types, associated with distinct gene expression modules. By discerning such patient-cell type subsets, we can readily identify clinical (e.g., histologic) or laboratory (e.g., biomarker) covariates for each cell-type that are associated with different UDON clusters (SATAY-UDON). First, to assess the ability of this approach to resolve new disease subsets, we performed the first in-depth scRNA-Seq of a comprehensive patient cohort of SJIA with clinically diverse etiologies (active disease, inactive disease, SJIA-associated lung disease (SJIA-LD), SJIA-macrophage

activation syndrome (MAS)) and matched pediatric healthy controls. In addition to resolving entirely new cellular and molecular subtypes of SJIA, these analyses identify peripheral biomarkers that predict distinct inflammatory responses including multiple divergent myeloid phenotypes. We find a novel complement and interferon (IFN) activation program enriched in SJIA-LD monocytes and confirmed from the serum of independent patients. To broadly assess systemic inflammatory disease subtypes, we performed a comprehensive pan-immune single-cell survey of de novo subtypes in over 1,000 autoimmune disease (SLE) and healthy donors, to identify both SJIA-specific as well as broadly conserved transcriptional programs. These data provide critical new insights into SJIA pathogenesis, the broader transcriptional landscape present in systemic inflammatory diseases, and the conserved inflammatory programs in subsets of healthy donors.

Results

Peripheral blood mononuclear cell populations defined in children with SJIA, SJIA-MAS, and SJIA-LD

While adult and pediatric forms of SLE have been extensively characterized through scRNA-Seq (2), other severe systemic inflammatory disorders such as SJIA have not. To define the immune landscape of SJIA, we performed scRNA-Seq analysis on PBMCs from a cohort of patients across disease activity and complications. Clinical disease course and treatment response in SJIA is highly variable; in addition, some children experience potentially fatal episodes of macrophage activation syndrome (SJIA-MAS). SJIA-MAS represents a systemic cytokine storm syndrome considered a form of secondary hemophagocytic lymphohistiocytosis characterized by decreased cytolytic function, excessive activation of hemophagocytic macrophages, and expansion of T cells (5, 6). Chronic lung disease (SJIA-LD) is an increasingly recognized pulmonary complication encompassing varying levels of pulmonary alveolar proteinosis (PAP), pulmonary artery hypertension, and fibrosis. 80% of SJIA-LD patients also have a history of MAS, suggesting these complications are pathogenically linked (7, 8). Samples were obtained from 20 children with SJIA (5 inactive SJIA, 7 active SJIA, 7 SJIA-LD (6 individual patients, one patient sampled twice indicated as “A” and “B”) and 2 SJIA-MAS), as well as 5 pediatric healthy controls (**Figure 1A**). All active SJIA and SJIA-MAS patients, and 4/7 SJIA-LD samples, had clinical features of active disease at time of sampling. Treatments included biologic therapy and/or steroids for most patients; some patients with active SJIA were newly diagnosed and sampled prior to initiation of therapy (**Figure 1B, Supplementary Table 1**). Laboratory parameters including serum ferritin, IL-18, CXCL9 and S100 proteins were frequently elevated particularly in active SJIA, SJIA-MAS and SJIA-LD patient samples, while most parameters were normal in inactive SJIA patients (**Supplementary Table 2, Figure 1B**).

We integrated all patient and healthy donor PBMCs (n=234,128 cells), considering possible donor and disease differences, to produce a compendium of 30 candidate cell populations (**Figure 1C, Methods**). These populations were annotated using a well-curated PBMC reference dataset using Azimuth (9) and manual annotation (**Methods, Supplementary Tables 3-4**). These included all major blood constituents, including B cells, T cells, Monocytes, DCs (dendritic cells), NK (Natural killer) cells, Erythrocytes, Platelets, in addition to sub-cell- types with distinct marker genes (**Figure 1D**). No donor-specific cell populations were observed, except certain erythrocyte populations were highly enriched for one patient with MAS, which is a common observation in highly active SJIA-MAS and HLH patients (**Supplementary Figure 1**) (10).

Comparing cell- type frequencies among patients and controls in this integrated compendium nominates somewhat consistent differences among these clinically defined patient subsets (**Figure 1E, Supplementary Figure 1**). We find that active SJIA and SJIA-LD patients have by trend increased NK cells, while Platelets and

1 Platelet Megakaryocyte were lesser in controls and inactive SJIA patients. Additionally, there was a
2 significantly lower proportion of MAIT cells in all patients as compared to controls, while Erythrocytes and
3 double negative T cells (dNTs) increased by trend in the SJIA-MAS patients (**Figure 1E, Supplementary**
4 **Figure 1-2**). Significantly lower proportions ($> 5\%$, t-test p-value < 0.05) of CD4 Naive and CD8 TEM cells were
5 observed in SJIA-LD as well as active and inactive SJIA, respectively, as compared to controls (Supplementary
6 Table 3).

7 **SJIA-MAS patients are distinguished from other SJIA patients by a highly expressed interferon gene** 8 **signature**

9 Distinctive, but often overlapping transcriptional signatures have been identified in SJIA, including interleukin
10 (IL)-1, IL-18 and Toll-like receptors (TLRs), and inflammasome signaling (11–14), while its complications
11 SJIA-MAS and SJIA-LD have been linked to both type I and type II interferon pathway activation (7, 15, 16). In
12 order to define dysregulated gene expression signatures in SJIA, we first identified differentially expressed genes
13 (DEGs) comparing all patients for each clinical subtype versus controls using the software cellHarmony (17).
14 Here, rather than compare individual cells, we pooled all cells from each patient cell-population into cell-type
15 pseudobulks for disease versus healthy control comparisons, to increase rigor. cellHarmony found that the
16 transcriptional landscape of SJIA-LD and SJIA-MAS was more dysregulated than clinically inactive or active
17 SJIA patients. SJIA-MAS MAIT cells were the most dysregulated cell-type with 614 upregulated DEGs
18 (**Supplementary Figure 3**). Comparing all active SJIA, SJIA-MAS, and SJIA-LD patients into a combined
19 disease group versus only the healthy controls, we observed 467 commonly deregulated genes, including 18
20 genes, associated with myeloid populations, previously described as significantly up- or downregulated in a
21 PBMCs bulk gene expression study (13) (**Figure 2A**).

22 We assessed cellHarmony-derived signatures using gene set enrichment to identify the unique and shared
23 gene programs among these different clinical subsets. Pairwise comparison of all 306 SJIA regulated
24 cellHarmony gene-sets (up- and down-regulated), identified 22 overlapping gene clusters of modules (M) shared
25 in at least 3 signatures, typically associated with different SJIA subtypes in the same or related cell-types (**Figure**
26 **2B, Methods**). The largest module represented shared down-regulated genes in diverse lymphoid cell
27 populations, principally associated with SJIA-LD but including SJIA-MAS and a active SJIA (M22). The other
28 two largest modules represent up-regulated genes shared among lymphoid cell populations in SJIA-LD (M1) or
29 SJIA-MAS (M2), which showed weak but significant association with each other. Other commonly dysregulated
30 cell-type gene programs were found in Monocytic (M3, M4, M14, M21), Plasmablast (M6, M13), B-cell (M7, M10,
31 M16, M19), T/NK cell (M15, M18, M20), platelets (M17), erythroid (M12) and HSCP (M5, M11). Gene-network
32 analysis of these sets highlighted important commonalities and differences. Specifically, *STAT1* and *MYC* were
33 predicted as key transcriptional regulators in SJIA-LD lymphoid populations (M1), while *STAT1* and *IRF1* were

the dominant predicted regulators in SJIA-MAS (M2), suggesting both commonalities and differences in the IFN signature in these patient types (**Figure 2C** and **Supplementary Figure 4**). While much smaller, the down-regulated lymphoid (M16) and B-cell differentiation (M22) modules were also denoted by different predicted core regulators (*CTCF* and *RBL2* vs. *JUN* and *JUND*, respectively) (**Supplementary Figure 4**). Comparison of dysregulated Gene Ontology terms showed M2 (SJIA-MAS) resulted in the broadest group of transcripts corresponding to diverse processes, while the other sets tended to show more specific modulation of processes involved in epigenetic regulation, focal adhesion, proliferation, inflammatory signaling, TNF and mTOR signaling, among others (**Figure 2D**). Hence, these data indicate both shared and unique gene networks, that differentially impinge upon broad chromatin regulators and inflammatory signaling pathways.

The above findings and previous studies suggest that IFNs play a key role in the disease pathogenesis of SJIA-MAS and SJIA-LD. To further determine IFN responses across PBMC, we performed visualization of previously defined IFN modules from whole blood, which finds substantial variation in the cellular source for different IFN mediated genes (**Figure 2E**) (2, 7, 16, 18, 19). Specifically, we observe a shared CD4 IFN⁺-cell specific induction of IFN genes in active SJIA, SJIA-LD, and SJIA-MAS associated with a subset of IFN-targets (most pronounced in module 1.2, which is predominantly driven by IFN β , a type I IFN) and expanding pan-IFN response most strongly in SJIA-MAS spanning the majority of myeloid, lymphoid and B-cell populations (particularly in the IFN γ driven modules 3.4 and 5.12) (**Figure 2E**, **Supplementary Figure 5**). Such impacts were not observed in control and inactive SJIA patients, which had few upregulated IFN related genes.

Transcriptional activation in monocytes and changes in lymphocyte cell frequency separate ongoing disease from inactive SJIA and controls

Monocytes are considered as central pathogenic drivers of SJIA and targeting of monocyte-derived pro-inflammatory cytokines is considered first line therapy. We identified four distinct monocytic populations in our patient compendium, CD14⁺ (classical), CD16⁺ (non-classical), intermediate, and an unclassified population (monocyte undefined) (**Figure 3A-B**, **Supplementary Tables 3-4**). These populations had varying frequency among individual patients, although 6/7 SJIA-LD samples show less intermediate monocytes (**Figure 3C**). These intermediate monocytes presented a mix of CD14 and CD16 monocyte features, with fewer discriminating features. We questioned how previously described alternations in SJIA monocytes mapped across these cellular subpopulations. Supervised comparison to a prior blood monocyte gene signature from SJIA patients with high serum ferritin (20) found three distinct gene modules, segregated by these myeloid populations (**Figure 3D**). Cluster 1 genes were enriched in IL-8 signaling markers, with the highest expression in CD16⁺ monocytes in SJIA-MAS (**Figure 3E**). Cluster 2 was characterized by preferential expression in CD14⁺ monocytes in active disease groups (**Figure 3D**), with enrichment for IL-1 signaling, endogenous TLR signaling, and IFN signaling (**Figure 3E**). Cluster 3 was most dominant in the undefined monocyte population,

enriched in centrosome and mitosis and included *MTOR*, *IL5RA*, and *IL11RA* (**Figure 3D-E**). Hence, we find monocytic subsets that are preferentially enriched, but not restricted to distinct SJIA clinical phenotypes, and associated with multiple signaling and proliferative processes.

We next examined the relative abundance and activation state of lymphocyte populations in SJIA. Here, we find profound differences in the distribution of T cell populations identified across SJIA patients (**Figure 3F, Supplementary Figure 6**). Considering all disease groups (active SJIA, SJIA-LD and SJIA-MAS) relative to inactive SJIA and controls, we find a broad decrease in CD4 Naïve, CD4 TCM, CD8 Mixed cells and MAIT cells (**Figure 3 G-H**). Recent data suggest the cytokine environment in SJIA may alter T cell polarization and represent a new possible avenue of therapy (21). Analysis of T cell polarization marker genes for Th1, Th2 and Th17 finds slightly higher expression of both Th1 (*CCL4*, *IFNG*, *TBX21*) and Th2 (*CCR3*, *CCR6*) markers in active SJIA, SJIA-LD, and MAS (**Supplementary Figure 7**), suggesting a shift from more naïve to more active T cell populations in SJIA. While there is some patient-level variation, we observed no significant differences in relative abundance of the other main cell-types (B cells, NK cells, Platelets and Erythrocytes, DCs and HSCPs) (**Supplementary Figures 8-11**).

Unsupervised discovery with UDON finds complement activation in the monocytes of SJIA-LD and a subset of active SJIA patients

The above findings, including IFN pathway activation, monocyte activation and changes in the T cell compartment, provide important insights into the broad pathogenesis of SJIA. However, given the significant patient-level clinical heterogeneity in SJIA (**Figure 1B**), we reasoned that clinical disease groups such as “active”, “inactive”, and “MAS” may be largely arbitrary, with respect to the underlying biology. To address this limitation, we developed an unsupervised strategy to uncover de novo shared transcriptional programs and patient subtypes that extends to cell-type level. Rather than focus on the individual cells, our approach, called UDON, specifically considers transcriptomic differences for each patient cell-type specific pseudobulk, compared to the combination of controls for that cell-type (**Figure 4A**). This approach collapses gene expression for all cells in a cell-population into a single vector. As this vector is normalized against the average gene expression profile of all healthy matched controls for the same cell-type, only patient-specific disease patterns should emerge. Unsupervised clustering of these control normalized patient pseudobulk differentials is performed in the software ICGS2 to find shared disease-specific gene expression programs that emerge from all cell populations, patients and genes (**Methods**) (22, 23). To inform underlying biology, UDON reports the dominant impacted pathways for each discovered UDON gene cluster (**Figure 4A**). In contrast to other approaches such as covarying neighborhood analysis (CNA), that identify gene modules that covary across samples, UDON identifies pseudobulk clusters and their most discriminate markers (4). When applied to our SJIA cohort, UDON found 12 distinct clusters, which include those enriched in type I and II IFN signaling (U12),

T cell cytotoxicity/IL12 (U4), Erythrocyte Development (U2) and Macrophage polarization (U10). Importantly, we were able to confirm the existence of all UDON clusters by using bulk transcriptomes of two independent SJIA cohorts (201 subjects and 53 controls (13, 24), which demonstrate these signatures in subsets of patients. (**Figure 4B, Table 1, Supplementary Figure 12, Supplementary Tables 5-6**). While UDON clusters were derived from a small number of patients with SJIA-MAS, bulk PBMC transcriptomes from previously reported MAS patients (n=5)(13) further display the same enrichment of type I and II IFN signaling (U12). Finally, using CNA, we identified 7 out of the 12 UDON clusters, based on the comparison of top correlated genes from CNA's reported top 10 principal components (**Supplementary Figure 13, Supplementary Table 7, Methods**), further supporting the validity of UDON clusters. As UDON is a fully unsupervised approach, with no prior imposed gene sets, a potential limitation is that individual UDON clusters can be composed entirely of pseudobulks derived from only one patient. Nonetheless, we observed no UDON clusters derived from a single patient or from only one disease group, and cell- types did not exclusively group together in the same UDON clusters (except U5), indicating clusters are driven rather by gene programs expressed across cell- types (**Figure 4C-F**). To confirm UDON results are stable with different target clustering resolutions or fewer samples, we tested a range of clustering resolutions and a reduced dataset of patients and controls (**Methods**). These analyses demonstrate that UDON is highly consistent when most patients that comprise an UDON cluster are present (**Supplementary Table 8, Supplementary Figure 14**).

Two UDON clusters, U4 (T cell cytotoxicity/IL-12) and U6 (complement induction), were particularly surprising in our analysis which result in distinct inflammatory pathways unique to active and SJIA-LD (**Figure 4G**). Specifically, U6 was intriguing as it was consistently induced in monocytes and pre-DC from all SJIA-LD patients, including complement genes *C1QA-C* along with *IFITM3* and *FLT3*. This cluster was further characterized by induction of surfactant genes *SFTPA1*, *SFTPA2* and *SFTPB*, which is particularly striking given the association of U6 with SJIA-LD pseudobulks (FDR-adjusted p-value < 0.1) and the histologic finding of dysregulated surfactant processing and PAP in such patients (7) (**Table 1, Methods**). Complement activation has not been well described in SJIA pathogenesis; thus we performed an external validation for complement components in SJIA patient serum (**Figure 4H, Supplementary Figure 15**). Using ELISA, we found significantly elevated levels of C9 in SJIA-LD, SJIA-MAS and active SJIA patients compared to controls or inactive SJIA patients. Significantly elevated levels for C5a were also observed for active SJIA patients versus controls and inactive SJIA, while C4 was elevated by trend in SJIA-LD and SJIA-MAS patients compared to controls (**Figure 4H, Supplementary Figure 15**), supporting that enhanced monocytic and macrophage production of complement represents a high-value target for further research.

Identification of new cellular phenotypes associated with de novo SJIA subtypes using SATAY-UDON

While UDON clusters provide intriguing putative insights into cellular and patient heterogeneity with disease, to understand complex phenotypic associations, we require methods to link clinical and diagnostic assay metadata with these predictions. To solve this challenge, we developed an accessory approach called Statistical Association Test for ClinicAI PhenotYpes (SATAY-UDON). SATAY-UDON considers phenotypes (e.g., disease classification, histology) and molecular correlates (e.g., metabolic readouts) together with non-redundant donor and cell-type associations in each UDON cluster using a metadata enrichment protocol (**Figure 5A, Methods**). Applied to our SJIA cohort, SATAY-UDON identified 40 phenotype-to-UDON cluster associations, suggesting patients represented in the UDON clusters share underlying clinical or diagnostic features (**Figure 5B, Supplementary Figure 16**). A subset of these SATAY-UDON associations is also identified in the top 5 expanded cell types in CNA's phenotypic association tests (Supplementary Figure 13C, Methods). Importantly, nearly all these observations were unique to UDON clusters as opposed to clinically defined SJIA subtypes.

SATAY-UDON finds a bias towards patients treated with steroids in UDON cluster U2, which contained genes involved in Erythrocyte differentiation (e.g., *FAM10B*, *FECH*, *BPGM*, and *AHSP*). All U2 were immature Erythrocytes from three active SJIA, three SJIA-LD and one SJIA-MAS sample. In contrast, U1 was associated with Fever, which was also comprised of Erythrocytes (and Platelets) and enriched in genes encoding Hemoglobin subunits as opposed to differentiation (**Supplementary Figure 17, Table 1**).

In UDON cluster U4, we observe a strong association with high Absolute Neutrophil Count (ANC) and higher C-reactive protein (CRP), well accepted markers for high underlying disease activity in SJIA (**Figure 5B, Supplementary Figure 17**). U4 consists primarily of cytotoxic T cells and CD4 and CD8 Mixed cells from three different active SJIA samples and 5 SJIA-LD samples. U4 marker genes are enriched in IL-12 mediated signaling and predicted cytotoxicity associated genes (e.g., *KLRF1*, *PRF1*, *CCL4*) (**Supplementary Figure 17, Table 1**).

Intriguingly, UDON cluster U12 showed associations of both Fever (with CD16 Monocytes, preDCs) and CXCL9 secretion (several myeloid and lymphoid populations), comprising all MAS patients plus one defined as active SJIA but noted to have subclinical MAS, further validating the clinical connection between these patients. U12 was principally associated with type I and II IFN signaling, with IFN γ as the central driver of CXCL9 secretion (16). A more focused transcriptional analysis of CD16+ monocyte pseudobulks found three different CD16+ monocyte transcriptional phenotypes, associated with distinct inflammatory and RNA-binding pathways, which subdivided SJIA-LD, active and MAS patients into novel subsets. We identified an expanded U12 cluster of MAS, active and SJIA-LD patients, as well as a U10 cluster of CD16+ monocyte pseudobulks from active and SJIA-LD patients, both of which were distinct from other CD16+ monocyte pseudobulks. The U10 cluster, defined by marker genes *IL1R2* and *CD163*, was associated with elevated S100A12 serum levels. The U12 cluster was associated with elevated CXCL9 serum levels and upstream of both type I and II IFN pathway genes, we

observed upregulation of a network of key inflammatory transcription factors previously implicated in IFN regulation (*STAT1*, *IRF1*) (**Figure 5C-D**). This further supports a central role of CD16 monocytes in driving the IFN response in SJIA-MAS.

Finally, UDON cluster U7 cluster, enriched in TLR signaling among Platelets and Megakaryocytes was found to be associated with elevated serum levels of S100A8/A9 and S100A12. Notably, *S100A8* and *S100A9* mRNAs were the principal markers of this population. Unsupervised analysis of these Platelet pseudobulks finds an expanded U7 cluster comprised of three active and one MAS patient, enriched in genes involved in the induction of apoptosis (e.g., *NOTCH2*, *TNFSF10*) and NF-kappaB signaling (e.g., *IRAK1*, *TRAF3*) (**Figure 5E-F**). Given emerging data that S100 proteins signal through TLR (25), these findings could indicate a novel mechanism where platelet-mediated activation via S100 proteins drives inflammation in SJIA. Hence, these findings support the notion that heterogenous immunological diseases with variable clinical features and biomarkers, may stem from identifiable cell-type specific disease subtypes.

UDON clusters represent broadly conserved transcriptional programs across a pan-immune landscape

To determine if these disrupted signaling networks are unique to SJIA or shared across other systemic inflammatory disorders, we performed a comprehensive pan-immune survey of de novo subtypes, leveraging existing single-cell profiles from 41 autoimmune disease (SLE) and 982 healthy donors (OneK1K cohort) (2, 3). We first performed UDON on a previously reported cohort of 33 patients with childhood-onset systemic lupus erythematosus (cSLE), as well as 8 adult SLE (aSLE) patients with matched controls (2). Using our 30 PBMC cell populations as a common reference, we produced pseudobulk folds for each child and adult relative to their matched controls. To assess the disease significance, we also produced pseudobulk folds for all healthy controls, relative to their collective average (**Figure 6A, Methods**). When applied to these SLE pseudobulk differentials, UDON found 21 stable clusters, after considering a range of possible resolutions (**Figure 6B, Methods**). While most of the UDON clusters were primarily composed of pediatric cases, only three (U5, U16, U19) were unique to cSLE, and no clusters were unique to a single patient (**Supplementary Figure 18**). Notably, 8 of the SLE-UDON clusters mapped to at least one SJIA-UDON cluster, with most of the remaining mapping to SJIA cell and/or subtype signature (e.g., CD16 monocytes in SJIA-LD) (**Supplementary Figure 18, Table 2**). Further analysis of SLE UDON cluster U3 showed that this CD14/CD16 monocyte dominated cluster corresponded to our Complement-associated SJIA-LD enriched cluster U6 and type I and II IFN-associated SJIA-MAS enriched U12, based on gene-set enrichment (**Figure 6C, Supplementary Figure 18**).

To examine the associations of UDON clusters with markers of SLE disease activity, we used SATAY-UDON with the associated SLE clinical metadata (2), which replicated earlier findings as well as proposed new disease associations (**Figure 6D, Supplementary Figure 19**, **Supplementary Table 9**). For these analyses, we employed both the Cochran–Mantel–Haenszel (CMH) procedure to account for covariate association differences

among adult and pediatric patients in addition to the standard Fisher Exact test, applied separately to each age group (Methods). Prior work with this cohort demonstrated a strong IFN signature across several cell-types, and associated abundance of those clusters with higher cSLE disease activity (2, 19). In support of this we find that SLE UDON clusters that map to IFN signaling (SLE-U3 and SLE-U11) (**Supplementary Figure 18**) in CD16 Monocytes and B Memory cells are associated with high systemic lupus erythematosus disease activity index (SLEDAI) scores, as well as in NK cells associated with kidney involvement (adjusted $p < 0.1$, CMH). More age-specific associations of IFN signaling UDON clusters, U3 and U11, were observed with dsDNA levels, serum sub-score of the SLEDAI, and pyuria (**Figure 6D**).

Strikingly, we also find a previously unrecognized association of higher complement component C4 levels in children with SLE-U13, which maps to IL-6 mediated signaling events, and this association was driven by a subset of lymphoid cell-types including NK, CD8 Mixed, and regulatory T cells. C4 was similarly associated with IL-5 mediated signaling SLE-U5 cluster, driven by monocytic cell-types (adjusted $p < 0.1$ for CMH test) (**Figure 6D**). These findings are notable as they nominate gene pathways that may be relevant in lupus patients without hypocomplementemia. We furthermore identify associations of specific treatments such as hydroxychloroquine with SLE-U18 (mapping to the clotting cascade), and other age-group specific associations of SLE-U3 and SLE-U11 with rash, erythrocyte sedimentation rate (ESR) and arthritis in children (**Supplementary Figure 19, Supplementary Table 9**).

Finally, to determine the overall pan-immune landscape of these inflammatory diseases we joint embedded SLE pseudobulk differentials and those derived from 982 normal donors from the OneK1K cohort into a SJIA-centric UMAP via UMAP projection. Considering all 30 cell-types we show broad alignment of controls from all three cohorts with inactive SJIA, as well as alignment of SJIA-MAS with aSLE (**Figure 6E-F, Methods**). Projecting labels from SLE to SJIA pseudobulk differentials, we found that the monocytes of a subset of cSLE patients phenocopy the SJIA macrophage activation (SJIA-U12) (**Figure 6G-H**). While the OneK1K cohort is comprised of only presumably healthy controls, it was previously discovered that known and novel autoimmunity quantitative trait loci associated with distinct autoimmunity implicated regulators in specific cell-types (3). Given these results, we projected all healthy control pseudobulks differentials into both SJIA and SLE UDON clusters. Inspection of these results found that multiple SLE UDON subtypes were assigned to a small number of healthy controls (**Supplementary Figure 18C**), including intermediate-monocytes of complement/interferon-associated SLE-U3 and fibrin clotting-associated platelets from SLE-U18 (**Figure 6I Supplementary Figure 19C**). These intriguing findings support the hypothesis that single-cell genomics can identify emerging conserved autoimmune programs in health and disease, as well dominate novel diagnostic gene-regulatory programs (**Figure 6J**).

Discussion

In this study we discover subtypes of systemic inflammatory disease that selectively associate with clinical features and biomarkers. To overcome existing analytical limitations, we defined a new computational approach to define subtypes that redraw established clinical classifications through transcriptomics at the level of individual cell-types. These data suggest the existence of conserved, pre-specified gene programs within the same or similar cell-types in distinct system inflammatory disorders. Our focused analysis of SJIA reveals distinct impacted inflammatory pathways that resolve different stages of disease, including active disease, inactive disease and MAS. Although prior work has identified potential pathogenic programs in SJIA including monocyte, neutrophil, T and B cell activation (12, 20, 26–29), the underlying causes and contributors to this heterogeneity remain unknown. Here, our initial analysis revealed activation of IFN-related genes across multiple cell-types particularly in SJIA-MAS, distinct transcriptional changes in monocyte populations, and differences in the T cell compartment. However, given the marked patient-level heterogeneity, we developed new methods called UDON and SATAY-UDON to identify and describe novel disease-associated transcriptional programs implicating new potential drivers of SJIA pathogenesis including IFN activation, multiple distinct monocyte phenotypes, and platelet activation. Critically we found that many of these transcriptional programs are broadly conserved across the immune landscape and represent previously hidden drivers of inflammatory diseases.

By identifying novel patient subclusters at individual cell-population level, UDON can overcome an inherent limitation of existing supervised comparison methods. A principal aim of single-cell genomics analysis is the ability to not only resolve cell-populations, but also complex underlying disease programs. UDON exploits well-established integration and single-cell clustering approaches to define patient-specific differences that underlie hidden disease programs. While the number of donors maybe limited within this cohort, we orthogonally confirmed the presence of UDON signatures using independent bulk transcriptomic cohorts of SJIA, CNA, and experimental validations. SATAY-UDON then further illustrates how these previously hidden programs are associated with clinical measures. Together, UDON and SATAY-UDON predictions provide new insights such as the predominance of IFN-driven activation in SJIA-MAS, and a monocytic-driven complement and interferon phenotype in SJIA-LD.

Critically, we found that UDON clusters represent not just features of SJIA pathogenesis, but broadly conserved transcriptional programs present across inflammatory disease states. We identified homologous UDON clusters present in both adult and childhood-onset SLE, with UDON clusters associated with distinct clinical and serological markers of lupus disease activity. Projecting transcriptomic data from over 1000 healthy donors and SLE patients across UDON clusters, we demonstrated that healthy individuals largely diverge from those with

inflammatory disorders. However, a minority of healthy individuals have UDON clusters that co-segregate with those from SJIA and SLE patients. It is tempting to speculate that such individuals have subclinical disease or underlying genetic predisposition to autoimmunity.

An important example is a new UDON-SJIA cluster dominated by IFN signaling (U12) present in patients with clinical evidence of overt and subclinical MAS, expressed by CD16+ and intermediate monocytes, preDCs, and other lymphocytic cell types. SATAY-UDON analysis associated this UDON cluster with elevated levels of CXCL9, an IFN γ -induced chemokine and specific MAS biomarker (16). These findings align with our demonstration that the most distinct IFN signature was expressed by monocytic and CD4 IFN+ cells from SJIA-MAS patients, with a lesser degree of elevated expression of these genes for monocytes from active SJIA and SJIA-LD. Prior work showing that monocytes from SJIA patients and untreated MAS patients or patients with secondary hemophagocytic lymphohistiocytosis (sHLH) are hyperresponsive to IFN γ further supports the hypothesis that this IFN signature driving MAS is derived from monocytes (20, 30, 31).

Our analysis also revealed that distinct monocyte transcriptional programs exist across the SJIA disease spectrum, including a myeloid polarization cluster (U10), expressed by active SJIA and SJIA-LD patients, and associated with elevated S100A12 levels. The transcriptional profiles of CD16+ monocytes in this UDON cluster showed marked differences from those of other patients, including high *CD163* expression, a myeloid differentiation marker which is responsible for binding and engulfing hemoglobin:haptoglobin complexes. Intriguingly, elevated *CD163* expression has also previously been identified as a marker for patients who fail anti-IL-1 therapy, further highlighting how different monocyte phenotypes link to disease biology (14).

UDON intriguingly discovered a distinct, monocytic driven complement and interferon program (U6), which was present in SJIA-LD patients and several active SJIA patients. This monocyte program is particularly notable as it also shows high levels of surfactant protein expression; dysregulated surfactant metabolism in the lungs is a key histologic feature of SJIA-LD (7). SJIA-LD shares characteristics with pulmonary alveolar proteinosis (PAP), which is defined by intra-alveolar accumulation of surfactant proteins due to impaired clearance by alveolar macrophages (32). Surfactant expression in circulating monocytes, as seen here, may reflect a specific subset of blood monocytes primed to migrate into lung tissue, where they subsequently differentiate into alveolar macrophages (33). This transcriptional program was also found in monocytes of cSLE patients, and associated with markers of severe disease activity.

While prior work has found increased expression of some complement pathway genes or proteins in JIA (34–39), to our knowledge, no study has investigated the involvement of complement in SJIA-LD. The detected increase of several serum complement proteins in SJIA patients observed here may partially derive from hepatocytes in the liver, which are considered the predominant source of complement components (40).

However, recent scRNA-Seq studies have intriguingly highlighted a distinct monocyte/macrophage subset expressing *C1QA-C* in patients with pediatric SLE, Behçet's and Kawasaki disease, adult Rheumatoid Arthritis, and bacterial infections (2, 41–44). In addition, a human cross-tissue scRNA-Seq study characterized a specialized lung alveolar macrophage subset strongly expressing *C1QA-C* (45), and our recent work examining lung tissue in a mouse model of MAS found a similar MAS-specific macrophage cell cluster enriched for complement activation (46). Clinically, low C3/C4, possibly reflecting complement consumption, has been described in 2 SJIA-MAS patients and 1 adult-onset Still's disease-MAS (47). Furthermore, a recent study of 23 patients with refractory HLH found 70% simultaneously present with complement mediated thrombotic microangiopathy (TMA). These authors hypothesized that the high levels of IFN γ in HLH activates complement, which then causes endothelial injury and damage in TMA (48). Together with the findings of Zheng and colleagues, who demonstrate that the C1Q-high inflammatory monocyte phenotype present in Behçet's disease patients is in vitro inducible with IFN γ (44), this highlights the potential interactions of complement activation and IFN signaling in driving pathogenesis in the SJIA clinical spectrum, and the conserved role of this monocytic program across the pan-immune landscape.

SATAY-UDON revealed an unexpected association of increased serum levels of the alarmin proteins S100A8/A9 and S100A12 with Platelet Megakaryocytes in UDON cluster U7, which was enriched for genes involved in TLR signaling. S100A8/A9 has been previously shown to signal through both RAGE and TLR4 to amplify inflammatory responses, and is likely the key driver of the TLR signaling pathways detected in U7 (25). Platelet S100A8/A9 levels are also increased in patients with SLE and peripheral artery disease and are thought to promote thrombosis and cardiovascular disease (25, 49, 50). Such elevated Platelet and Platelet Megakaryocyte frequencies are detected in active SJIA and MAS (58), which supports the novel hypothesis that in SJIA, activated platelets may contribute to inflammation by releasing S100A8/A9 in the microenvironment to drive inflammation and thrombosis (50).

We also find a strong association with well-described clinical markers of high disease activity in SJIA – both laboratory parameters (CRP, ANC) and active arthritis – and cytolytic T cell populations in the IL-12 signaling/cytotoxicity cluster U4. Indeed, most cytotoxic T cell pseudobulks were associated with U4, except for those from the SJIA-MAS patients which rather clustered with the IFN-driven U12. More broadly, our analysis of the T cell landscape supports a shift from more naïve to more active populations in SJIA. Recent work identifying a common HLA-DRB1*15 haplotype in many SJIA-LD patients (51) has suggested that this clinical subtype could have distinct patterns of T cell activation (21). Indeed, our analyses found that lymphocyte population in SJIA-LD have a striking transcriptional signature that has some similarity but is distinct from that seen in SJIA-MAS. However, we saw similar patterns of Th1/Th2 markers across the active disease populations. We also observed no changes in Th17 polarization as has been previously reported (26). Intriguingly while the MAIT cells was significantly lesser in the disease group, these cells presented a strongly dysregulated gene expression

profile particularly in SJIA-MAS patients, but also in active SJIA and SJIA-LD. Finally, SJIA-MAS patients had significantly more dNTs, a cell- type which was also expanded in pediatric lupus and is proposed to represent an end-stage T cell subset particularly efficient in cytokine secretion and cytotoxicity (52). While these clinical associations suggest novel hypotheses underlying the transcriptional programs, precise validation approaches are required in specific cell- populations in patients.

In conclusion, UDON and SATAY-UDON offer a novel exploratory computational strategy to discover distinct transcriptional programs in large clinically heterogeneous patient cohorts. Through this approach, we find a previously unexplored shared monocytic complement and interferon gene program in SJIA-LD, which is also present in lupus and associated with markers of high disease activity. We also discover a role for platelets in driving SJIA inflammation and evidence for distinct monocyte transcriptional phenotypes present across inflammatory disorders including lupus, SJIA, and MAS. Importantly, this method highlights heterogeneous clinical phenotypes and serum measurements that underlie these novel subtypes, suggesting transcriptional programs result in separable and durable disease responses. Together, this approach can identify diverse transcriptional programs found across cell- types in severe systemic autoimmune and autoinflammatory disorders. We anticipate new opportunities to improve and expand these computational approaches in the future, including improved means to distinguish cell-type specific heterogeneous diseases impacts and automatically accounting for potential batch effects in various stages of the analyses.

Methods

Experimental design

This was a cohort study of children with SJIA, SJIA-MAS, and SJIA-LD. Clinical and laboratory features were obtained from the electronic medical record using a standardized case report form. All patients were diagnosed with SJIA based on the International League of Associations for Rheumatology (ILAR) criteria (53); however, for newly diagnosed patients, samples were obtained and treatment initiated with a disease duration <6 months, using the operational definition of SJIA (54). Inactive disease was defined based on the Wallace criteria (55); conversely patients were considered to have active SJIA if they had any active clinical features including arthritis, rash, fever, adenopathy, hepatosplenomegaly, or elevated c-reactive protein or erythrocyte sedimentation rate. Patients were diagnosed with MAS based on diagnosis of the treating physician. Patients were considered to have SJIA-LD if they had both clinical and radiographic features of lung disease based on definition of probable or definite SJIA-LD (1).

Sample collection

1 Fresh whole blood was collected in CPT tubes by venipuncture, mixed by inverting the tube 8-10 times and then
2 centrifuged at 1600g, 20 min at room temperature. The cell pellet containing isolated peripheral blood
3 mononuclear cells (PBMCs) was washed twice with PBS, resuspended at approximately $2-4 \times 10^6$ PBMCs/ml in
4 freezing media (90% FCS/10% DMSO) and stored in liquid nitrogen until further processing.

6 **Complement protein analysis**

7 Serum levels of complement proteins were analyzed either by Luminex (R&D Systems, Inc., Minneapolis MN,
8 USA) for C5a, C1q, C4 and MBL, by C9 ELISA (Abcam, Cambridge, United Kingdom) or TCC ELISA
9 (Mybiosource, San Diego, CA, USA). Statistical analysis was performed with GraphPad Prism 9.3.1. Except
10 where noted, results of statistical tests were significant when $p < 0.05$.

12 **PBMC isolation and Single-cell RNA-Sequencing**

13 Frozen PBMCs were thawed and washed twice with PBS. Dead cells were removed with dead cell removal kit
14 following the manufacturer's instructions (Miltenyi Biotec, Bergisch Gladbach, Germany)), and then resuspended
15 at 700-1,200 cells/ μ l in PBS (total count 12,800 cells). Samples were then sequenced with 10x Genomics
16 (Pleasanton, CA, USA), version Chromium NextGEM (Chemistry: 3'v3 Assay)

18 **Single-cell RNA-Seq Analysis**

19 Details on read alignment, cell clustering, differential gene expression analyses, and supervised clustering are
20 provided in the supplementary methods section.

22 **UDON**

23 *Intuition.* UDON is an unsupervised approach to detect novel gene regulatory programs that are common or
24 different among heterogeneous disease samples based on single-cell RNA-Seq. We assume that prior defined
25 clinically annotated disease subtypes are imprecise, with the goal of redefining them at a pseudobulk fold level.
26 UDON leverages the common concept of a pseudobulk profile, which is the average gene expression of a set of
27 cells (usually from a cell-type or a cluster) from a donor, patient, or sample. For larger single-cell disease
28 cohorts, individual patients can have varying cell-type frequencies. Combining all cells together without
29 explicitly considering which cells are associated with which patients, can lead to disease observations driven
30 by a single patient; however, pseudobulks provide an effective way around this. The central premise of this
31 approach is that once cell-types are defined from a cohort (outside of this program), case and control
32 cohort designs can be exploited to identify common gene expression responses between sample-level gene
33 expression changes in individual cell populations, as opposed to single-cells. Here, populations-level effects
34 within a sample are defined by first calculating pseudobulks. Specifically, for the controls within a cohort,
35 an aggregate pseudobulk for a cell-population is calculated across all controls. If different controls exist for
36 different patients, based on age, batch or other factors (referred to here as batch), batch-specific pseudobulks

are computed. UDON computes fold change between patient-specific pseudobulks compared to the aggregate controls for each cell-type. In special scenarios, such as the analysis of the OneK1K cohort, pseudobulk fold changes are also computed for each control donor and cell-type. The fold change is calculated between the sample-specific pseudobulks against the aggregate control pseudobulks. We cluster pseudobulk fold change instead of pseudobulks, as the latter alone would only identify distinct cell-type clusters, as opposed to common disease-specific transcriptional differences. Using the pseudobulk fold changes from samples with disease, the software finds clusters of pseudobulks with common gene expression changes by applying sparse non-negative matrix factorization (NMF)-based clustering (ICGS2). ICGS2 is an unsupervised method to define populations from bulk or scRNA-Seq in AltAnalyze and defines clusters using iterative guide-gene focused sample/cell clustering to find coherent correlated gene modules. UDON is thus fully unsupervised and independent of prior gene sets and patient subtypes. An UDON cluster can be composed entirely of normalized pseudobulks from one cell-type or many. It is a naive subtype identification approach which is only dependent on the initial cell-type definitions and the selection of explicit control samples. This program can identify patient populations with or without a pre-imposed k -parameter (desired number of clusters), but only reports NMF-defined clusters with unique marker gene expression. For ICGS2, we used the default program options with the additional parameters ρ set to 0.3, $markerPearsonCutoff$ set to 0.2, k set to a range of tested values (10-30), and $FoldDiff$ set to 2. A final k resolution of 15 was selected in SJIA and 25 for SLE, as the number of predicted clusters beyond these numbers failed to substantially increase the number of “stable” clusters. For each cluster, unique marker genes are determined using the MarkerFinder module of AltAnalyze. Table 1 displays the top 10 unique ranked marker genes using this MarkerFinder statistic.

Assumptions. We make the following assumptions prior to applying this algorithm:

1. A sufficiently sized cohort of patients is required to identify shared gene expression programs among two or more patients
2. A relatively homogenous set of controls is needed.
3. Gene expression datasets have undergone quality control (including batch effect removal) and have been log-scaled.
4. Cells have a cell-type annotation assigned manually or from an external program.
5. Phenotypically distinct cells comprise a reasonable subset of those in at least one cell-type from a patient.
6. (If applying SATAY-UDON) There is at least one quantifiable clinical measurement for at least two disease samples included in the UDON analyses and the clinical measurement is/transformed into a binary variable.

Defining sample-level gene expression change between diseased and control samples. Consider a cohort of d number of disease samples and h number of healthy samples. Let D and H be the gene expression matrices for disease and healthy samples respectively such that D_i is a $m \times n_i$ matrix of gene expression, where m is

the number of genes and n_i is the number of cells for disease sample $i \in [1, d]$ and H_i is a $m \times n_i$ matrix of gene expression, $i \in [1, h]$,

$$\begin{aligned} D &= [D_1, D_2, \dots, D_d] \\ H &= [H_1, H_2, \dots, H_h] \end{aligned}$$

Let $j \in [1, c]$ where j is a cell-type belonging to c cell-types assigned to the cells of the samples. Then, for a disease sample $i \in [1, d]$,

$$\tilde{P}_{ij} = \sum_{r=1}^{n_{ij}} \frac{D_{ijr}}{n_{ij}}$$

where \tilde{P}_{ij} is a m dimensional vector representing the sample-level disease cell-type pseudobulk from n_{ij} cells to capture sample-specific gene expression programs.

The control samples, however, are assumed to be homogenous in UDON and therefore, the cell-type pseudobulk P_j (a m dimensional vector) is computed as an aggregate value across all controls. It is defined as:

$$P_j = \frac{\sum_{i=1}^h \sum_{r=1}^{n_{ij}} H_{ijr}}{\sum_{i=1}^h n_{ij}}$$

The fold change between the disease sample i and the controls, for all $j \in [1, c]$, is simply defined as:

$$\hat{P}_{ij} = \tilde{P}_{ij} - P_j$$

where \hat{P}_{ij} is a $m \times c$ matrix.

Lastly, prior to applying the sparse non-negative matrix factorization clustering algorithm to the pseudobulk profiles, we ensure that \hat{P}_{ij} has non-negative entries, a required condition for the algorithm. Now,

$$\hat{P}_{ij} = \hat{P}_{ij} - M_{ijg}$$

where M_{ijg} is the minimum value of \hat{P}_{ij} for gene $g \in [1, m]$.

Applying sparse non-negative matrix factorization (NMF) for clustering pseudobulk profiles.

Let $[\hat{P}_{ij}]$ be the matrix be the collection of \hat{P}_{ij} vectors for all $i \in [1, d]$ and $j \in [1, c]$. We provide $[\hat{P}_{ij}]$ as the input for ICGS2, which applies the sparse non-negative matrix factorization (NMF) algorithm to cluster the pseudobulk profiles (the columns of $[\hat{P}_{ij}]$). UDON provides an optional cluster resolution parameter k for NMF analysis that allows users to explore broader or more granular clusters.

For the downstream analyses presented in the paper, we have considered UDON clusters derived with k set to 15. The resolution of 15 was selected as stable (consistent with higher resolution results) after considering $k=10, 15, 20$ or 30 , which reported 10, 12, 14 and 17 final UDON clusters, respectively. Dominant impacted

pathways (Pathway Commons by default) are reported for each UDON cluster, which represent common or heterogenous groups of patients and cell-types.

Statistical analyses in SATAY-UDON

To discover underlying clinical or phenotypic associations from UDON clusters, SATAY-UDON provides a statistical phenotype enrichment protocol, which considers non-redundant donor and cell-type associations per UDON cluster. Our null hypothesis is that there is no association between a sample's clinical condition and the sample's pseudobulk profile in a gene response program (as indicated by an UDON cluster). We test this hypothesis for positive enrichment by performing a one-sided Fisher's Exact test on all pseudobulk profiles for a given cell-type in an UDON cluster. Fisher's exact test requires categorical data, and thus, the continuous clinical variables are transformed into a binary variable based on clinical expert-set thresholds. This option is not provided in SATAY-UDON and must be determined by the user outside of the program. For a given cell-type and clinical measure, in a 2 by 2 contingency table, let the number of pseudobulks associated with samples with the clinical condition in and not in UDON cluster U be Q and r respectively. Similarly, let the number of pseudobulks associated with samples that do not have the clinical condition in and not in UDON cluster U be S and t respectively. Then, SATAY-UDON applies Fisher's exact test on the described contingency table.

By default, four or more samples are required for an association to take place between an UDON cluster and clinical covariate, indicated by Q value in the contingency table. For a cell-type, an association between a UDON cluster and a clinical variable is considered positive and visualized if the one-sided p-value < 0.1 for the above-mentioned Fisher's exact test; however, SATAY-UDON provides the confidence level (p-value) as a user provided parameters for users. We report the false discovery rate (FDR) adjusted p-value by applying the Benjamini-Hochberg adjustment to the p-values from each clinical variable and visualizing the associations with an FDR adjusted $p < 0.1$.

Cochran-Mantel-Haenszel test for confounding variables.

To protect SATAY-UDON associations against confounding variables (such as age, sex, etc.), we employ the Cochran-Mantel-Haenszel (CMH) procedure to produce stratified estimates of association between UDON clusters and clinical covariates. The CMH test stratifies the samples by the categories in the confounding variable (for example, male and female if sex is a confounding variable) and considers a series of 2 by 2 contingency tables of the binary predictors for each stratum. In SATAY-UDON, for each stratum, like the Fisher's Exact Test, the i th 2 by 2 contingency table, where $i \in [1, c]$ and is the stratum in a list of c categories of a confounding variable, has the following values defined in same way as mentioned above: Q_i, r_i indicate the number of pseudobulks associated with samples with the clinical condition of interest in and not in UDON cluster,

respectively, and S_i, t_i indicate the number of pseudobulks associated with samples without the clinical condition of interest in and not in UDON cluster, respectively.

By default, two or more samples are required in each stratum for an association to take place between an UDON cluster and clinical covariate, indicated by Q_i value in the contingency table . CMH calculates a p-value and an odd ratio that represents a weighted association between a UDON cluster and a covariate across the strata. We report the false discovery rate (FDR) adjusted p-value by applying the Benjamini-Hochberg adjustment to the p-values from each clinical variable and visualizing the associations with an adjusted $p < 0.1$. For the sLE dataset, we applied the CMH test, considering age as the confounding variable, to identify the associations between a UDON cluster and clinical covariate that are protected against adult and pediatric data imbalances. At least two samples are required in each age group of samples (also referred as stratum) for the binary predictors, UDON cluster and clinical covariate, to be considered for the CMH test. For determining age group-specific associations, we perform SATAY-UDON using the Fisher's Exact Test on only age group-specific samples and report the raw and FDR-adjusted p-values. We note that other study designs may necessitate distinct testing procedures and batch effects correction beyond the parameters described here.

Covarying Neighborhood Analysis (CNA)

Details on CNA are provided in the supplementary methods.

External bulk and scRNA-seq dataset analyses

Details on external SJIA and SJIA-MAS dataset analyses are provided in the supplementary methods.

Study approval

This study was approved by the CCHMC institutional review board (IRB# 2018-2408) and written informed consent was obtained from each parent or guardian. Child assent was obtained where appropriate.

Code and Data Availability

The UDON workflow is composed of independent python and R modules. Scripts for pre-processing, data normalization and clinical covariate association analyses (SATAY-UDON) are available on Github (<https://github.com/kairaveet/udon-sjia-sle>). Unsupervised iterative clustering via guide-gene selection are accomplished through the existing python2 module ICGS2 in AltAnalyze (<http://www.altanalyze.org> and <https://github.com/nsalomonis/altanalyze>). Excel file indicating supporting data values for figures is provided in addition to the supplemental tables. The processed PBMC SJIA scRNA-Seq and associated metadata have

been deposited in the Gene Expression Omnibus (GEO) database (GSE207633). Raw sequencing data is being made available in dbGAP, associated with this GEO study.

Acknowledgements

We thank the patients and families for participating in this study, and the clinicians who referred patients for evaluation. We thank the CCHMC Gene Expression Core for assistance in performing scRNA-seq, and the Pediatric Rheumatology Tissue Repository (PRTR) for sample storage and processing. This work was partially supported by an Academic Research Clinical (ARC) Award from the Cincinnati Children's Research Foundation (GS and AG), the Systemic JIA Foundation, the Jellen Family Foundation, a K-supplement Award from the Rheumatology Research Foundation (GS), National Heart Lung and Blood Institute (N.S., U24HL148865), the Chan Zuckerberg Pediatric Network for the Human Cell Atlas (N.S.) and the Cincinnati Children's Pediatric Cell Atlas Center (N.S.). Dr. Schulert was supported by NIH K08-AR072075. Dr. Verweyen was supported by the DFG (German Research Foundation, project 448863690).

Author contributions

Wrote the manuscript: EV, GSS, NS, KT. Acquired data: SD, EB, KC, DS. Analyzed data: KT, EV, NS. Designed the study: GSS, NS, AAG. Provided samples: SC. Order of first authorship: EV wrote the initial draft of the manuscript.

References:

1. Erkens R, Esteban Y, Towe C, Schulert G, Vastert S. Pathogenesis and Treatment of Refractory Disease Courses in Systemic Juvenile Idiopathic Arthritis: Refractory Arthritis, Recurrent Macrophage Activation Syndrome and Chronic Lung Disease. [Internet]. *Rheum Dis Clin North Am* 2021;47(4):585–606.
2. Nehar-Belaid D et al. Mapping systemic lupus erythematosus heterogeneity at the single-cell level. *Nat Immunol* 2020;21(9):1094–1106.
3. Yazar S et al. Single-cell eQTL mapping identifies cell type-specific genetic control of autoimmune disease. [Internet]. *Science* 2022;376(6589):eabf3041.
4. Reshef YA et al. Co-varying neighborhood analysis identifies cell populations associated with phenotypes of interest from single-cell transcriptomics. [Internet]. *Nat Biotechnol* 2022;40(3):355–363.
5. Crayne CB, Albeituni S, Nichols KE, Cron RQ. The Immunology of Macrophage Activation Syndrome [Internet]. *Front Immunol* 2019;10:119.
6. Schulert GS, Grom AA. Pathogenesis of macrophage activation syndrome and potential for cytokine-directed therapies. [Internet]. *Annu Rev Med* 2015;66:145–59.
7. Schulert GS et al. Systemic Juvenile Idiopathic Arthritis–Associated Lung Disease: Characterization and Risk Factors. *Arthritis & Rheumatology* 2019;71(11):1943–1954.
8. Saper VE et al. Emergent high fatality lung disease in systemic juvenile arthritis [Internet]. *Ann Rheum Dis* 2019;78(12):1722–1731.
9. Hao Y et al. Integrated analysis of multimodal single-cell data. *Cell* 2021;184(13):3573–3587.e29.
10. Hinze CH et al. Immature cell populations and an erythropoiesis gene-expression signature in systemic juvenile idiopathic arthritis: implications for pathogenesis. [Internet]. *Arthritis Res Ther* 2010;12(3):R123.

11. Quartier P et al. A multicentre, randomised, double-blind, placebo-controlled trial with the interleukin-1 receptor antagonist anakinra in patients with systemic-onset juvenile idiopathic arthritis (ANAJIS trial). *Ann Rheum Dis* 2011;70(5):747–754.
12. Brown RA et al. Neutrophils From Children With Systemic Juvenile Idiopathic Arthritis Exhibit Persistent Proinflammatory Activation Despite Long-Standing Clinically Inactive Disease [Internet]. *Front Immunol* 2018;9:2995.
13. Fall N et al. Gene expression profiling of peripheral blood from patients with untreated new-onset systemic juvenile idiopathic arthritis reveals molecular heterogeneity that may predict macrophage activation syndrome [Internet]. *Arthritis Rheum* 2007;56(11):3793–3804.
14. Verweyen EL, Pickering A, Grom AA, Schultert GS. Distinct Gene Expression Signatures Characterize Strong Clinical Responders Versus Nonresponders to Canakinumab in Children With Systemic Juvenile Idiopathic Arthritis. *Arthritis and Rheumatology* 2021;73(7):1334–1340.
15. Rice GI et al. Assessment of Type I Interferon Signaling in Pediatric Inflammatory Disease. *J Clin Immunol* 2017;37(2):123–132.
16. Bracaglia C et al. Elevated circulating levels of interferon-gamma and interferon-gamma induced chemokines characterize patients with macrophage activation syndrome complicating systemic juvenile idiopathic arthritis (in press) [Internet]. *Ann Rheum Dis* 2016;76(1):1–7.
17. DePasquale EAK et al. cellHarmony: cell-level matching and holistic comparison of single-cell transcriptomes [Internet]. *Nucleic Acids Res* 2019;47(21):e138–e138.
18. Chiche L et al. Modular transcriptional repertoire analyses of adults with systemic lupus erythematosus reveal distinct type i and type ii interferon signatures. *Arthritis and Rheumatology* 2014;66(6):1583–1595.
19. Banchereau R et al. Personalized Immunomonitoring Uncovers Molecular Networks that Stratify Lupus Patients. *Cell* 2016;165(3):551–565.
20. GS S et al. Monocyte and bone marrow macrophage transcriptional phenotypes in systemic juvenile idiopathic arthritis reveal TRIM8 as a mediator of IFN- γ hyper-responsiveness and risk for macrophage activation syndrome [Internet]. *Ann Rheum Dis* [published online ahead of print: 2020]; doi:10.1136/ANNRHEUMDIS-2020-217470
21. Binstadt BA, Nigrovic PA. The Conundrum of Lung Disease and Drug Hypersensitivity-like Reactions in Systemic Juvenile Idiopathic Arthritis. *Arthritis & Rheumatology* 2022;0(0):1–10.
22. Venkatasubramanian M, Chetal K, Schnell DJ, Atluri G, Salomonis N. Resolving single-cell heterogeneity from hundreds of thousands of cells through sequential hybrid clustering and NMF. [Internet]. *Bioinformatics* 2020;36(12):3773–3780.
23. Emig D et al. AltAnalyze and DomainGraph: analyzing and visualizing exon expression data [Internet]. *Nucleic Acids Res* 2010;38(suppl_2):W755–W762.
24. Brachat AH et al. Early changes in gene expression and inflammatory proteins in systemic juvenile idiopathic arthritis patients on canakinumab therapy [Internet]. *Arthritis Res Ther* 2017;19(1):13.
25. Lood C et al. Platelet-Derived S100A8/A9 and Cardiovascular Disease in Systemic Lupus Erythematosus. *Arthritis and Rheumatology* 2016;68(8):1970–1980.
26. Henderson LA et al. Th17 reprogramming of T cells in systemic juvenile idiopathic arthritis. *JCI Insight* 2020;5(6):1–17.
27. Kessel C et al. Proinflammatory Cytokine Environments Can Drive Interleukin-17 Overexpression by γ/δ T Cells in Systemic Juvenile Idiopathic Arthritis. [Internet]. *Arthritis Rheumatol* 2017;69(7):1480–1494.
28. Do T et al. MicroRNA networks associated with active systemic juvenile idiopathic arthritis regulate CD163 expression and anti-inflammatory functions in macrophages through two distinct mechanisms [Internet]. *J Leukoc Biol* 2018;103(1):71–85.
29. Imbrechts M et al. Insufficient IL-10 Production as a Mechanism Underlying the Pathogenesis of Systemic Juvenile Idiopathic Arthritis. *The Journal of Immunology* 2018;201(9):2654–2663.
30. Macaubas C et al. Altered signaling in systemic juvenile idiopathic arthritis monocytes. [Internet]. *Clin Immunol* 2016;163:66–74.

31. Pascarella A et al. Monocytes From Patients With Macrophage Activation Syndrome and Secondary Hemophagocytic Lymphohistiocytosis Are Hyperresponsive to Interferon Gamma. *Front Immunol* 2021;12(March):1–9.
32. Brasch FE et al. Surfactant proteins in pulmonary alveolar proteinosis in adults [Internet]. *Eur Respir J* 2004;24(3):426–435.
33. Evren E et al. Distinct developmental pathways from blood monocytes generate human lung macrophage diversity [Internet]. *Immunity* 2021;54(2):259-275.e7.
34. Barnes MG et al. Subtype-specific peripheral blood gene expression profiles in recent-onset juvenile idiopathic arthritis [Internet]. *Arthritis Rheum* 2009;60(7):2102–2112.
35. Nirmala N et al. Gene-expression analysis of adult-onset Still’s disease and systemic juvenile idiopathic arthritis is consistent with a continuum of a single disease entity. [Internet]. *Pediatr Rheumatol Online J* 2015;13(1):50.
36. Aggarwal A, Bhardwaj A, Alam S, Misra R. Evidence for activation of the alternate complement pathway in patients with juvenile rheumatoid arthritis. *Rheumatology* 2000;39(2):189–192.
37. Brunner J et al. Analysis of the classical, alternative, and mannose binding lectin pathway of the complement system in the pathogenesis of oligoarticular juvenile idiopathic arthritis. *Rheumatol Int* 2012;32(6):1815–1818.
38. Gilliam BE, Reed MR, Chauhan AK, Dehlendorf AB, Moore TL. Significance of complement components C1q and C4 bound to circulating immune complexes in juvenile idiopathic arthritis: Support for classical complement pathway activation. *Clin Exp Rheumatol* 2011;29(6):1049–1056.
39. Miller JJ, Olds LC, Silverman ED, Milgrom H, Curd JG. Different patterns of C3 and C4 activation in the varied types of juvenile arthritis. *Pediatr Res* 1986;20(12):1332–1337.
40. Zhou Z, Xu MJ, Gao B. Hepatocytes: a key cell type for innate immunity [Internet]. *Cell Mol Immunol* 2016;13(3):301–315.
41. Lei H, Wang C, Wang Y, Wang C. Single-cell RNA-Seq revealed profound immune alteration in the peripheral blood of patients with bacterial infection. *International Journal of Infectious Diseases* 2021;103:527–535.
42. Wang Z et al. Single-cell RNA sequencing of peripheral blood mononuclear cells from acute Kawasaki disease patients. *Nat Commun* 2021;12(1):1–10.
43. Zhang F et al. Defining inflammatory cell states in rheumatoid arthritis joint synovial tissues by integrating single-cell transcriptomics and mass cytometry. *Nat Immunol* 2019;20(7):139–148.
44. Zheng W et al. Single-cell analyses highlight the proinflammatory contribution of C1q-high monocytes to Behcet’s disease. *Pnas* 2022;119(26):1–12.
45. Mulder K et al. Cross-tissue single-cell landscape of human monocytes and macrophages in health and disease. *Immunity* 2021;54(8):1883-1900.e5.
46. Gao DK et al. IFN- γ is essential for alveolar macrophage driven pulmonary inflammation in macrophage activation syndrome. *JCI Insight* [published online ahead of print: 2021]; doi:10.1172/jci.insight.147593
47. Gorelik M, Torok KS, Kietz DA, Hirsch R. Hypocomplementemia associated with macrophage activation syndrome in systemic juvenile idiopathic arthritis and adult onset still’s disease: 3 Cases. *Journal of Rheumatology* 2011;38(2):396–397.
48. Gloude NJ et al. Thinking Beyond HLH: Clinical Features of Patients with Concurrent Presentation of Hemophagocytic Lymphohistiocytosis and Thrombotic Microangiopathy. *J Clin Immunol* 2020;40(5):699–707.
49. Wang Y et al. Platelet-derived S100 family member myeloid-related protein-14 regulates thrombosis. *Journal of Clinical Investigation* 2014;124(5):2160–2171.
50. Dann R et al. Platelet-Derived MRP-14 Induces Monocyte Activation in Patients With Symptomatic Peripheral Artery Disease. *J Am Coll Cardiol* 2018;71(1):53–65.
51. Saper VE et al. Severe delayed hypersensitivity reactions to IL-1 and IL-6 inhibitors link to common HLA-DRB1*15 alleles. [Internet]. *Ann Rheum Dis* 2022;81(3):406–415.

- 1 52. Tarbox JA et al. Elevated Double Negative T Cells in Pediatric Autoimmunity. *J Clin Immunol*
2 2014;34(5):594–599.
- 3 53. Petty RE et al. International League of Associations for Rheumatology classification of juvenile idiopathic
4 arthritis: second revision, Edmonton, 2001. [Internet]. *J Rheumatol* 2004;31(2):390–2.
- 5 54. DeWitt EM et al. Consensus Treatment Plans for New-Onset Systemic Juvenile Idiopathic Arthritis.
6 *Arthritis and Rheumatology* 2012;64(7):1001–1010.
- 7 55. Wallace CA, Giannini EH, Huang B, Itert L, Ruperto N. American College of Rheumatology provisional
8 criteria for defining clinical inactive disease in select categories of juvenile idiopathic arthritis. [Internet].
9 *Arthritis Care Res (Hoboken)* 2011;63(7):929–36.
- 10 56. Stuart T et al. Comprehensive Integration of Single-Cell Data. *Cell* 2019;177(7):1888-1902.e21.
- 11 57. Zambon AC et al. GO-Elite: A flexible solution for pathway and ontology over-representation.
12 *Bioinformatics* 2012;28(16):2209–2210.
- 13 58. Olsson A et al. Single-cell analysis of mixed-lineage states leading to a binary cell fate choice [Internet].
14 *Nature* 2016;537(7622):698–702.
- 15
16

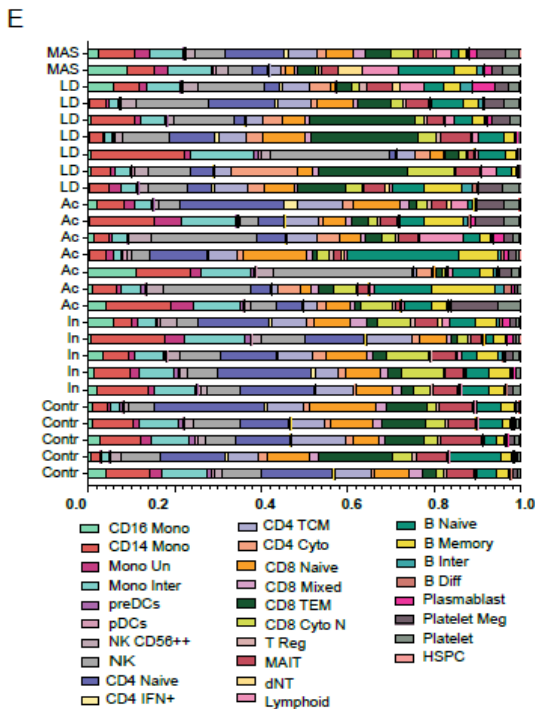


Figure 1. Peripheral blood mononuclear cells vary in composition by pediatric SJIA clinical subtype. A) Study overview illustrating 26 SJIA patients and controls, for which clinical features were collected and PBMCs were analyzed by single cell RNA Sequencing. B) Binary plot depicting treatments, laboratory parameters (protein biomarkers), and systemic features (clinical associations) of each patient at time of sample collection. C) Integrated UMAP of 234,128 single cells and 30 annotated cell populations from SJIA samples and controls. Cluster identity specified on the basis of Azimuth and literature associations. D) Dot plot of average population gene expression for prior defined cell- type marker genes. Dot size indicates the percentage of cells expressing the gene and color intensity indicates the mean expression. E) Bar plot indicating cell frequency of each cell- type (Erythrocytes excluded) per sample in the cohort shown in panel A. In=Inactive SJIA, Ac= active SJIA, LD= SJIA-LD, MAS= SJIA-MAS.

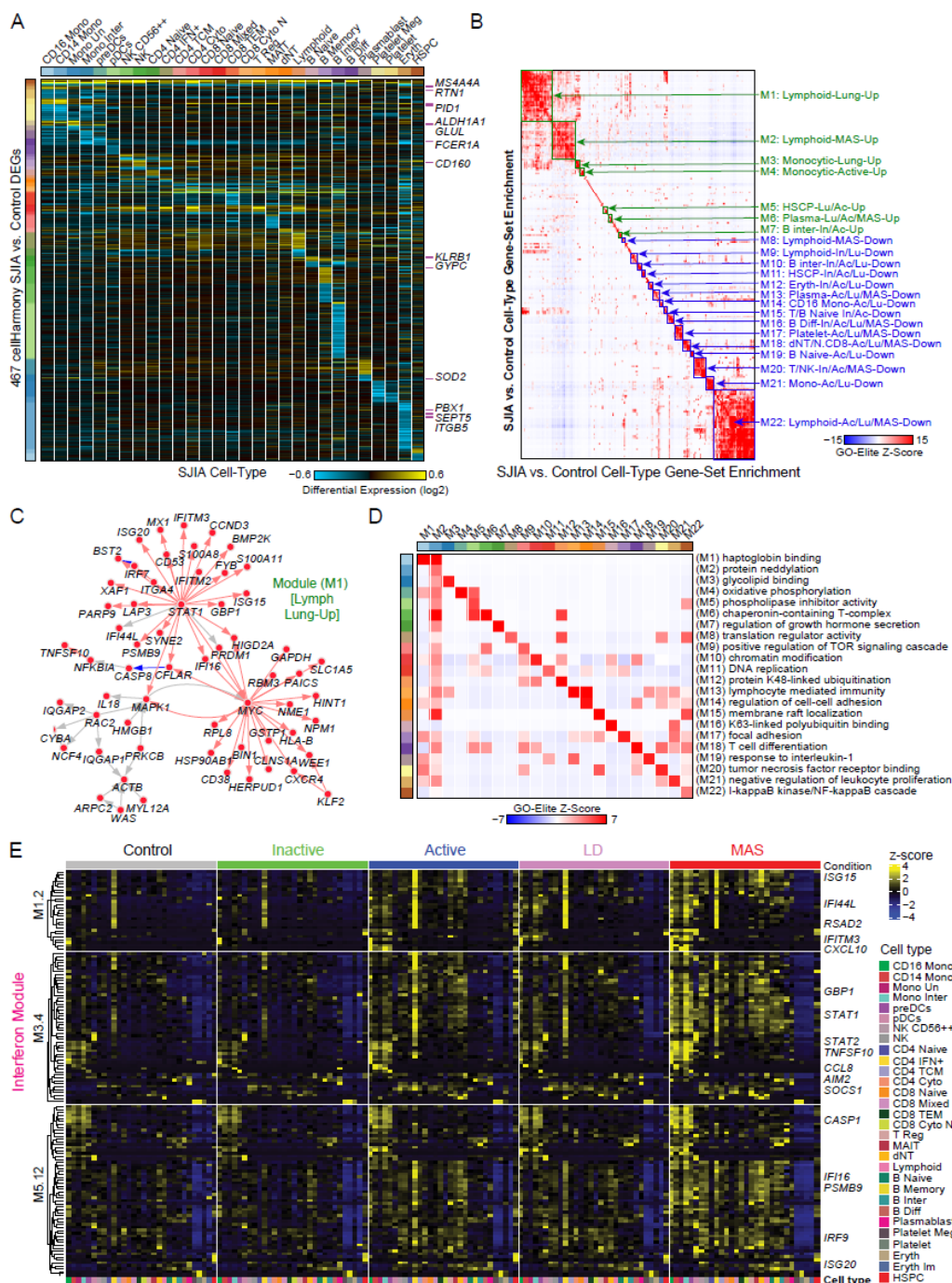


Figure 2. SJIA-MAS patients are distinguished from other SJIA patients by a distinct IFN gene signature. A) cellHarmony differential fold-change heatmap comparing SJIA patient cell pseudo-bulks in Disease (Active, LD, MAS) versus controls. Each column is the mean fold difference for a cell-population and each row a gene (fold>1.2 and empirical Bayes moderated t-test $p < 0.005$, unadjusted). B) Identification of SJIA impacted gene-sets (modules), defined from all constituent cellHarmony SJIA subtype and cell-type comparisons. Each module represents multiple up- or down-regulated patient versus control signatures with mutual gene-set enrichments (GO-Elite). The source signatures include aggregate disease and specific SJIA subtypes versus controls. Module annotations (right) denote the major associated cell-types and subtypes signatures present. C) Transcription factor (TF) and gene interaction networks for shared genes in module M1 from panel B. Red nodes = up-regulation and Blue nodes = down-regulation. Red arrows indicate annotated TF-target interactions in GO-Elite (TRRUST, Pazar, Amadeus). D) Module specific example Gene Ontology terms associated with each of the shared genes for each Module in panel B. E) Heatmap of scaled (z-score) expression values of IFN induced gene modules M1.2, M3.4 and M5.12 as described in Banchereau et al., 2016, across all PBMC cell clusters and clinical groups.

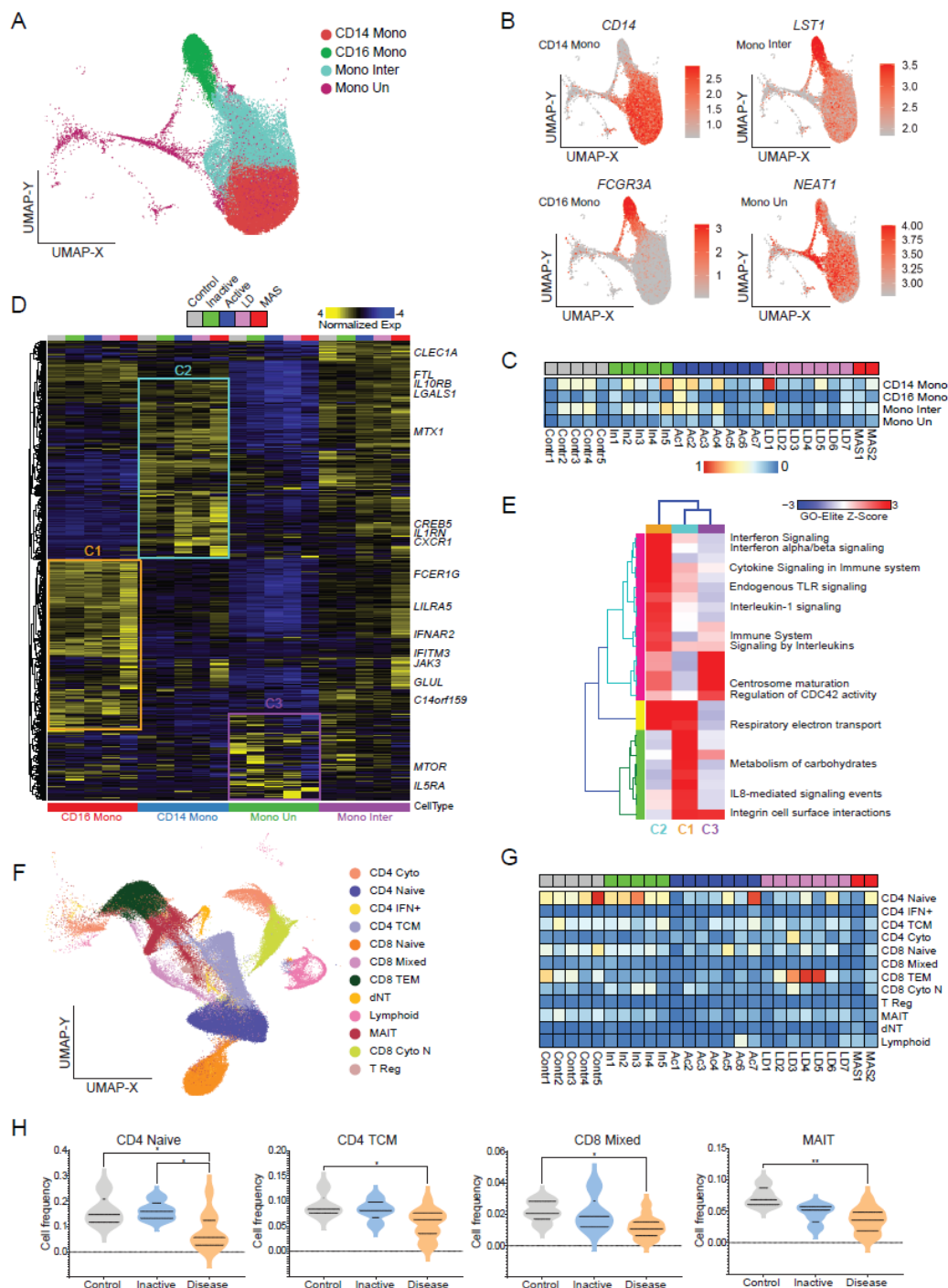


Figure 3. Transcriptional activation in monocytes and changes in lymphocyte cell frequency separate ongoing disease from inactive SJIA and controls. A) UMAP representing all 4 monocyte cell populations identified by scRNA-seq. B) Feature plots indicating the expression levels of selected marker genes of monocyte populations. C) Matrix representing the cell frequency per individual SJIA patient or control for the 4 monocyte cell populations. D) Gene expression heatmap of previously determined monocyte signatures of high ferritin SJIA patients as described in Schulert et al, 2020, across the 4 monocyte populations. Supervised clustering defined clusters C1, C2 and C3 shown in D), and GO-Elite analysis of associated cluster pathways (Pathway Commons) is shown in E). F) UMAP representing all 12 T-cell populations identified by scRNA-seq. G) Matrix representing cell frequency per individual SJIA patient or control. H) Violin plots depicting significant differences in cell frequency in the 4 T-cell populations between controls (n=5), Inactive SJIA (n=5) or Disease (combined data from Active SJIA, SJIA-LD and SJIA-MAS) (n=16). Bars indicate significant differences calculated by one-way ANOVA (*= adj. p-value ≤ 0.05 , **= adj. p-value ≤ 0.005).

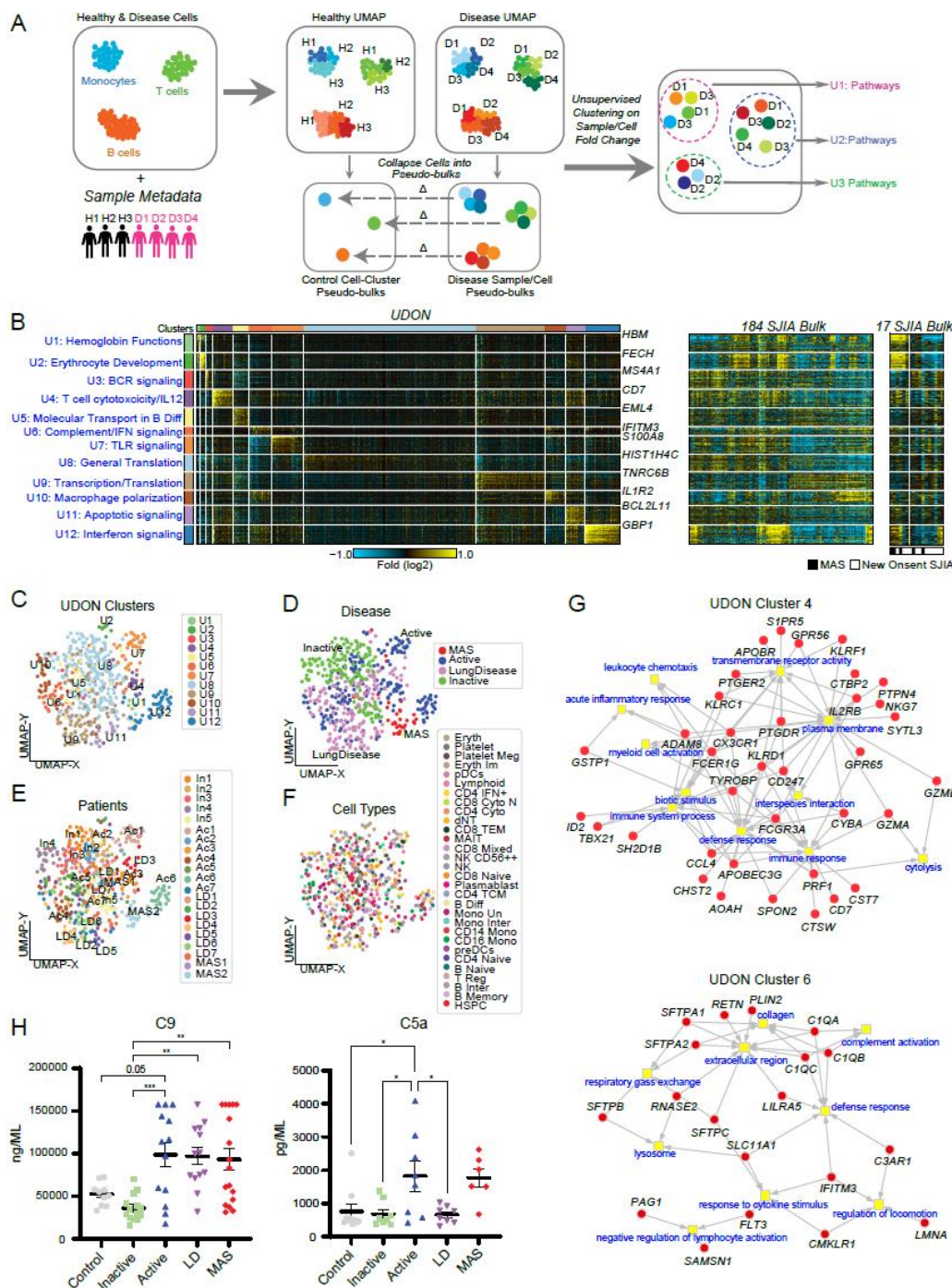


Figure 4. UDON analysis defines new SJIA disease subtypes including complement activation in monocytes in SJIA-LD patients. A) Overview of the UDON analysis pipeline, an unsupervised clustering method applied to control normalized patient pseudo-bulks to define disease subtypes. B) SJIA UDON patient-cell subtypes (UDON clusters 1-12), defined by the top cluster marker genes and top enriched pathways (PathwayCommons), are shown in the left heatmap. Confirmation of UDON SJIA subtypes from independent large bulk transcriptomics cohorts. Confirmation of UDON SJIA subtypes from two independent large bulk SJIA PBMC transcriptomics cohorts are shown to the right of the UDON heatmap, with matching genes, normalized to within cohort controls. C-F) UMAP visualization of control normalized patient pseudobulks for UDON clusters (C), clinical subtypes (D), individual patients (E) and cell-populations (F). G) Gene-to-GO term associations (GO-Elite) for UDON cluster 4 and cluster 6. H) Serum protein expression of complement component C9 and C5a of healthy controls (n=10) and SJIA patients (n=57) by ELISA. Error bars indicate mean \pm standard deviation. Significant differences calculated by one-way ANOVA (*= adj. p-value \leq 0.05, **= adj. p-value \leq 0.005, ***= adj. p-value \leq 0.001).



32

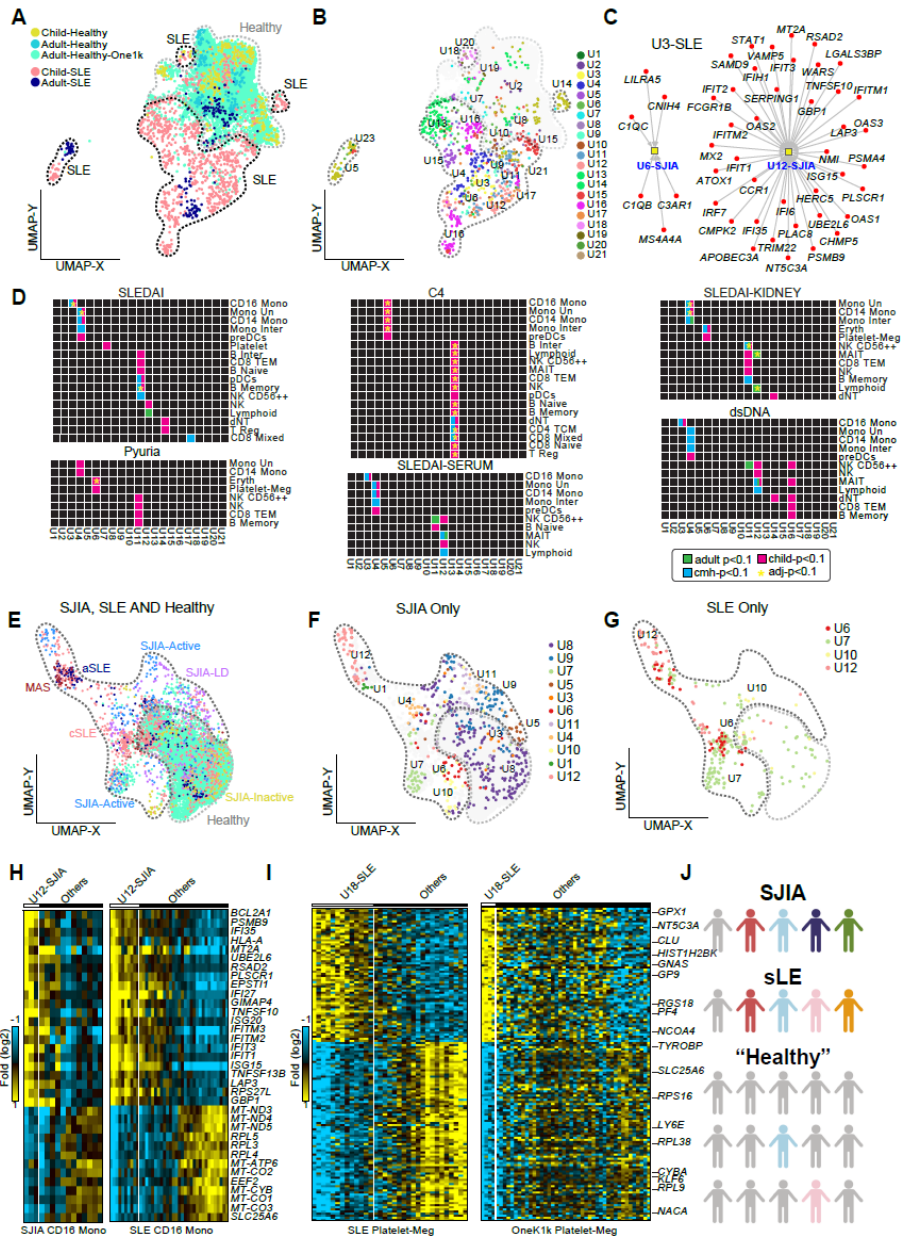


Figure 6. Udon analysis identifies broad subtypes and inflammatory programs in a pan-immune atlas. A) Joint UMAP of SLE cohort (adult, child, controls) with healthy PBMC samples in the OneK1K cohort (pseudobulk folds). The drawn boundaries in the UMAP designate control enriched populations (grey) versus SLE (black). B) Visualization of only SLE samples from panel A, colored according to the UDON cluster number. C) Marker genes in SLE-UDON cluster 3 (n=60) that overlap with enriched SJIA-UDON clusters 6 (complement activation) and 12 (interferon signaling). D) Age-independent and age-specific SATAY-UDON results for selected SLE clinical parameters (blue = $p < 0.10$ Cochran-Mantel-Haenszel test, green = $p < 0.10$ adult-specific, pink = $p < 0.10$ child-specific one-sided Fisher's Exact Test, yellow star = FDR-adjusted $p < 0.10$). E) Joint UMAP of SJIA, OneK1K and SLE samples (each dot is a normalized pseudobulk) for SJIA UDON marker genes, colored by the known clinical subtype of each sample. The drawn boundaries in the UMAP designate control-enriched populations (grey) versus systemic inflammatory disease (black). F) Visualization of only SJIA samples on the same UMAP as panel E, colored according to UDON cluster number from Fig. 4C. G) Same UMAP as panel E but highlighting the SLE samples, colored by projected SJIA-UDON cluster labels from panel F onto SLE samples. H) Heatmap of common differentially expressed genes (fold>1.2 and empirical Bayes t-test $p < 0.05$, two sided), for SJIA UDON cluster 12 CD16 monocyte pseudobulk folds versus other CD16 monocytes in the SJIA cohort and SLE cohort. I) Heatmap of common differentially expressed genes (fold>1.2 and empirical Bayes t-test $p < 0.05$, two sided), for SLE UDON cluster 18 platelet pseudobulk folds versus other platelets in the SLE and OneK1K cohort. J) Proposed model for disease heterogeneity among systemic inflammatory disease patients and presumably healthy donors in the populations.

1 Table 1. UDON cluster summary.

UDON cluster	Cluster name	Patient group(s)	No. Pseudo-bulks	Cell type(s)	No. enriched genes	Top 10 enriched genes (by Pearson rho)
U1	Hemoglobin Functions	2 Active, 2 MAS	5	Erythrocytes/Platelets	90	HBA1, HBM, HBA2, HBB, SLC25A39, IFI27, ADIPOR1, ALAS2, TSTA3, HBD
U2	Erythrocyte Development	3 Active, 3 LD, 1 MAS	7	Eryth Im	168	DCAF12, FECH, FAM210B, MKRN1, SLC25A37, STRADB, BNIP3L, BPGM, RPIA, BLVRB
U3	BCR Signaling	2 Inactive, 2 Active, 3 LD	9	mainly pDCs	75	MS4A1, CD79A, VPREB3, RALGPS2, CD79B, PLEKHF2, FCRLA, BANK1, SWAP70, LY86
U4	T Cell Cytotoxicity/IL12	3 Active, 5 LD	25	cytotoxic T cells, CD4/ CD8 Mixed cells	359	CD7, CTSW, TTC38, CLIC3, KLRF1, PRF1, CST7, CCL4, AOA, SPON2
U5	Molecular Transport in B Diff	5 Inactive, 6 Active, 7 LD, 2 MAS	20	B Diff	533	EML4, LRCH1, ORMDL1, DENND5B, ZBTB25, GOLGB1, NKTR, WEE1, BMP2K, COBLL1
U6	Complement/IFN signaling	3 Active, 7 LD	29	Monocytic	35	IFITM3, FLT3, C1QB, C1QA, C1QC, RETN, PLIN2, MS4A4A, SLC11A1, CNIH4
U7	TLR Signaling	1 Inactive, 3 Active, 1 LD, 1 MAS	40	T cells, mixed	118	S100A8, S100A9, LYZ, VCAN, S100A12, MND, FCN1, LILRB2, CD14, PLBD1
U8	General Translation	all samples	216	mixed	253	HIST1H4C, MZT2A, NBEAL1, EEF1B2, GLTSCR2, TOMM20, TOMM7, EIF3D, MZT2B, EIF3E
U9	Transcription/Translation	4 Inactive, 5 Active, 6 LD, 2 MAS	86	mainly HSPC, T Reg/CD8 TEM	795	TNRC6B, PIK3IP1, ZFP36L2, KIAA1551, CELF2, CABIN1, APBA2, TSC22D3, GPATCH8, ANKRD44
U10	Macrophage Polarization	1 Inactive, 2 Active, 3 LD	27	mainly monocytic	56	IL1R2, CD163, SAP30, JDP2, ARL4A, RNF144B, PHC2, FKBP5, IRS2, KLF9
U11	Apoptotic Signaling	4 Active, 3 LD, 2 MAS	24	mainly MAIT, Erythrocytes and T Reg, other T cells	728	BCL2L11, RNF167, ATP2B4, PYHIN1, SIT1, PPP1CA, MT1E, SAMD3, SLC9A3R1, MXD4
U12	IFN Signaling	1 Active, 2 MAS	44	T cells/Monocytic, DC	742	GBP1, GBP4, STAT1, APOL6, IFIT3, ISG15, EPST11, DTX3L, UBE2L6, GBP5

2
3

1 Table 2. SJIA and sLE UDON cluster comparison summary.

SJIA UDON cluster	cSLE UDON cluster	Cluster name	# cSLE (#cSLE+aSLE) pseudobulks	cSLE cell types	SJIA cell types
U4	U8	T Cell Cytotoxicity / IL-12	17 (20)	Mainly Lymphoid	Cytotoxic T cells, CD4/CD8 Mixed cells
U6	U3	Complement / IFN Signaling	30 (33)	Monocytic	Monocytic
U7	U6	TLR Signaling	19 (34)	CD4 IFN+ cells, Erythrocytes and Platelets	T cells, mixed
U8	U14	General Translation	86 (105)	Mixed (B/T cells, Monocytic)	Mixed
U9	U16	Transcription / Translation	108 (108) cSLE only	Mixed T cells	HSPC, T Reg, CD8 TEM
U10	U4	Macrophage polarization	47 (53)	Monocytic	Mainly monocytic
U11	U12	Apoptotic Signaling	57 (63)	NK and MAIT cells, mixed T cells	Mainly MAIT, erythrocytes, T reg, other T cells
U12	U3	IFN Signaling	30 (33)	Monocytic	T cells, monocytic, DC
U12	U11	IFN Signaling	53 (67)	Mixed, plurality pDC	T cells, monocytic, DC

2
3
4
5