Unveiling the two faces of immune dysregulation in Systemic Sclerosis
A multi-omics approach

Nila Hendrika Servaas

Unveiling the two faces of immune dysregulation in Systemic Sclerosis
A multi-omics approach

Onthulling van de twee gezichten van immuunontregeling bij Systemische Sclerose
Een multi-omics benadering

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de
Universiteit Utrecht
op gezag van de
rector magnificus, prof.dr. H.R.B.M. Kummeling,
ingevolge het besluit van het college voor promoties
in het openbaar te verdedigen op

donderdag 25 november 2021 des middags te 4.15 uur

door

Nila Hendrika Servaas

geboren op 30 juni 1993
te Gouda

THESIS OUTLINE

Chapter 1

General introduction

The immune system: a constant interplay of opposing forces

The ancient Chinese concept of Yin and Yang describes two contrasting forces that exist as inseparable and contradictory opposites[1]. This concept was beautifully described in the Zhuangzi, (Chuang-tzu), Chapter 21: "*When the state of Yin was perfect, all was cold and severe; when the state of Yang was perfect, all was turbulent and agitated. The coldness and severity came forth from Heaven; the turbulence and agitation issued from Earth. The two states communicating together, a harmony ensued and things were produced.*"[2]. The constant interplay of Yin and Yang is thought to underlie many processes in nature, which depend on the balance and stability of various biological systems. In fact, everything in nature exhibits some degree of duality, and the immune system is no exception.

The immune system forms an intricate network of various immune cells (leukocytes), which carry out specific functions in response to endogenous and exogenous stimuli. The immune system has a crucial role in the defense against outside invaders, including viruses and bacteria, but it also plays a vital role in maintaining the body's homeostasis, for example through wound healing[3] and detection of cancer cells[4]. Accordingly, immune cells have dual roles, with some cells being poised to attack (generally referred to as inflammation), while others display more defensive characteristics (generally referred to as tolerance). This duality is critical for the initiation and control of immune responses that support the return to health. Responses that are too weak result in the failure to resolve infections, whereas responses that are too strong or wrongly directed can lead to the development of autoimmune disease. On top of this, the immune system displays another layer of duality, namely the balance between innate and adaptive immune responses[5]. The innate immune system forms the first line of defense against pathogens and comprises non-specific inborn immune mechanisms. The adaptive immune system forms the second line of defense and comprises highly specific immune mechanisms acquired throughout life. Although these two components of the immune system are often studied as separate coexisting entities, the interplay between the innate and adaptive immune system is crucial for the timely initiation and termination of immune responses. Thus, the immune system walks a fine line, constantly being pushed and pulled between an activating and suppressive state, and is controlled by checks and balances of immune cells that support so-called peripheral tolerance[6]. In fact, it has been proposed that immunological tolerance against self-antigens is far from absolute, and specific auto-reactive cells of the adaptive immune system can often be found back in circulation[7, 8]. Under healthy conditions, these auto-reactive cells are reined in by tolerogenic cells of the innate immune system. However, when these cells cross the line towards the activating side, the immune system starts mounting responses to healthy tissue, and attacks the very cells that it is meant to protect. This is when autoimmunity arises.

Systemic sclerosis: the immune system gone awry

Systemic sclerosis (SSc) is an example of an autoimmune disease where the immune system has crossed the line towards hyper activation. SSc is a complex, heterogeneous disease mainly characterized by vascular abnormalities, immune
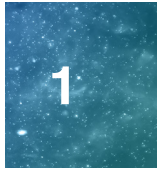
involvement and extensive fibrosis of the skin and internal organs[9]. Based on the extent of skin fibrosis, SSc patients can be broadly classified into two major subsets: limited-cutaneous SSc (lcSSc) and the more severe form, diffuse-cutaneous SSc (dcSSc)[10]. Moreover, patients that fulfill the classification criteria for SSc, but do not present with skin fibrosis are classified as non-cutaneous (ncSSc) patients. Lastly, individuals who suffer from vascular abnormalities, manifesting as Raynaud's phenomenon, in combination with typical SSc features, including vascular abnormalities, or auto-antibodies, are classified as early (eaSSc) patients. These patients have a high probability to develop definite SSc within 10 years[11].

Since skin fibrosis is the cardinal feature of SSc, the disease was historically considered to be a fibrotic disease. In fact, the disease name is derived from the Greek terms *skleros* and *derma*, literally translating to hard skin. However, over the last decades, mounting evidence has pointed towards dysregulation of the immune system as an indispensable factor underlying SSc pathogenesis. One of the first lines of evidence for the participation of the immune system in SSc pathology was the discovery of specific antibodies reactive towards self-antigens (autoantibodies), which are detectable in 90-95% of SSc patients[12, 13]. Moreover, histology-based studies have shown that activated immune cells infiltrate the skin of SSc patients already from very early disease stages, before the onset of fibrosis[14]. It is now broadly accepted that the immune system plays a prominent part in SSc pathogenesis. Whereas fibroblast activation was first thought to be an effect of intrinsic dysregulation of the fibroblasts themselves, it is now becoming more and more clear that immune cell activation and inflammation induce the differentiation and activation of fibroblasts that deposit excessive amounts of extracellular matrix, eventually leading to fibrosis[15]. In turn, activated fibroblasts can perpetuate immune cell dysregulation in SSc through the release of cytokines and direct interactions with immune cells[16], thereby creating a self-sustaining loop between inflammation and fibrosis.

Contribution of different immune cell types to systemic sclerosis pathogenesis

Although immune system dysregulation is now recognized as one of the main culprits of SSc pathogenesis, the molecular mechanisms that underlie this dysregulation are not fully understood. Various types of immune cells display signs of activation and are thought to contribute to the immunopathology in one way or another. These include innate cell types such as myeloid cells (monocytes, macrophages and dendritic cells) as well as adaptive immune cells (B and T lymphocytes). Based on their frequency, molecular phenotype and altered functionality, different roles have been described for different immune cell types in the pathogenesis of SSc.

The innate immune system in systemic sclerosis

The cells of the innate immune system form the first line of defense against invading pathogens. They are crucial cells in initiating immune responses, but at the same time, they also play indispensable roles in the dampening and resolution of immune responses once danger is resolved, and immune cell populations have returned to homeostasis. Inflammation through innate immune responses is mainly triggered

by circulating as well as tissue resident innate immune cells including monocytes, macrophages and dendritic cells (DCs), which can recognize broad molecular patterns using specialized surface receptors.

Toll-like receptors

The functioning of innate immune cells largely relies on the recognition of pathogen associated molecular patterns (PAMPs) and danger associated molecular patterns (DAMPs) by pattern recognition receptors (PRRs). Toll-like receptors (TLRs) comprise one well-studied family of PRRs that are broadly involved in the recognition of viral and bacterial antigens[17], although more such receptor families are now described, including RIG-I–like receptors, Nod-like receptors, and C-type lectin receptors[18]. TLRs are particularly interesting in the context of SSc as, besides exogenous antigens, they also have a well described role in the recognition of various endogenous antigens released upon cellular stress or tissue damage[19]. Indeed, endogenous TLR agonists including TLR4 binding fibronectin (an extracellular matrix component) and TLR7/8/9 binding RNA/DNA complexes, are increased in affected tissue and circulation of SSc patients[20–22]. TLR engagement by these endogenous DAMPs activates innate immune cells and induces the production of inflammatory mediators including the prototypical NF-κB-dependent cytokines tumor necrosis factor (TNF) and interleukin-6 (IL-6), as well as type I interferon (IFN), and the pro-fibrotic molecules transforming growth factor beta (TGF-β) and C-X-C motif ligand 4 (CXCL4). Notably, many of these factors have been described to be increased in the serum and plasma of SSc patients[23–25] and contribute to fibrosis[26, 27]. These data support a driving role of immune cell activation through TLR signaling in the disease pathogenesis. Moreover, distinct TLR ligands can activate distinct signaling pathways in innate immune cells, meaning that they can induce context specific immune responses. Thus, TLRs are not only critical in initiating pro-inflammatory responses but also prime the immune system towards distinct inflammatory states. This is critical in the context of SSc, where TLR signaling does not only contribute to immune system hyper-activation, but also directs the immune system towards a state that facilitates fibrosis.

Monocytes

Monocytes are circulating innate immune cells that constitute around 2-8% of all leukocytes in peripheral blood of healthy individuals. Based on the expression of the cell surface receptors CD14 and CD16, at least three types of monocytes can be distinguished in human circulation[28]. These include classical monocytes (CD14^{++}CD16^{-}), intermediate monocytes (CD14^{++}CD16^{+}) and non-classical monocytes (CD14^{+}CD16^{++}). Monocytes are equipped with various TLRs, chemokine receptors and adhesion molecules that allow them to promptly respond to danger signals and facilitate their extravasation from circulating blood into affected tissues during inflammation. Upon entering the tissue, monocytes can differentiate into mononuclear phagocytes including macrophages or monocyte derived dendritic cells (moDCs), depending on the differentiation signals they receive[29]. Here they can give rise to distinct phenotypes of macrophages which can be distinguished on the basis of different surface markers. These constitute the
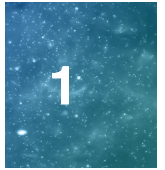
classically activated M1 macrophages, majorly involved in pro-inflammatory responses, or alternatively activated M2 macrophages, majorly involved in anti-inflammatory responses and wound healing[30, 31]. However, monocytes do not merely constitute a reservoir for tissue-resident cells. They also orchestrate immune responses through the secretion of pro-inflammatory cytokines and function as antigen-presenting cells without requiring the differentiation to downstream cell types[32]. Moreover, they are involved in the maintenance of tissue homeostasis through their roles in vascular remodeling[33], tissue repair[34], and the production of anti-inflammatory cytokines[35]. Thus, monocytes are multifaceted immune cells that are involved in many different processes which makes them crucial to consider in autoimmunity related inflammation.
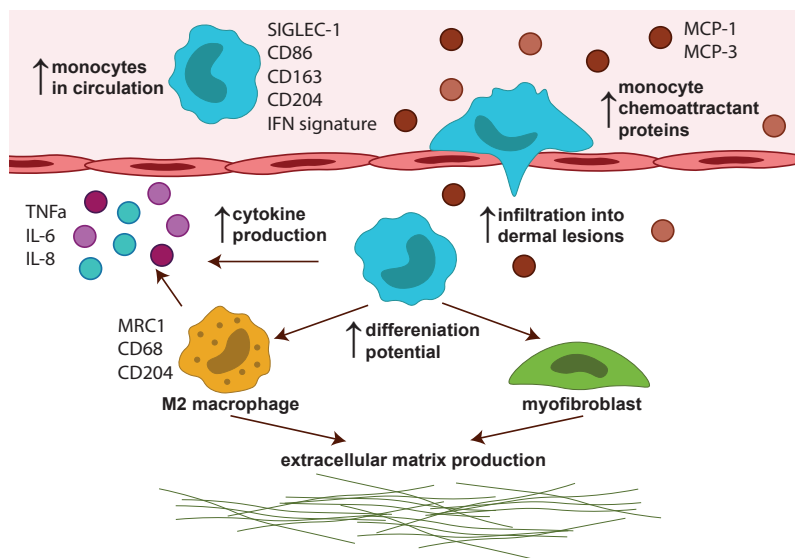
Figure 1. Monocyte dysregulation in the pathogenesis of SSc. Monocytes are increased in circulation of SSc patients and show enhanced expression of activation markers Siglec-1, CD86, CD163, CD204 and the expression of type I IFN genes (IFN signature). Increased monocyte chemoattractant proteins (MCP1, MCP-3) can lead to the increased infiltration of monocytes observed in SSc dermal lesions. Here monocytes produce large amounts of pro-inflammatory and pro-fibrotic cytokines, and they differentiate into M2 macrophages and myofibroblasts that produce extracellular matrix material.

Numerous lines of evidence implicate monocytes in the pathogenesis of SSc (Figure 1). Firstly, monocytes are increased in frequency in the circulation of SSc patients[36, 37], which is already evident at early, pre-fibrotic disease stages[38]. Here, their frequencies correlate with clinical manifestations, including the occurrence of interstitial lung disease (ILD) and the extent of skin fibrosis[38]. Next to a numerical increase of monocytes themselves, monocyte chemoattractant proteins (MCP-1 and MCP-3), which are important mediators of monocyte tissue-directed migration, are also increased in circulating blood and skin of SSc patients[39, 40]. In line with the increase of these chemoattractants, monocytes are amongst the most predominantly infiltrating cells in SSc dermal lesions[37, 41], and increased numbers of macrophages in SSc skin have

also been reported[14, 37, 42]. Moreover, these cells display an activated and pro-fibrotic phenotype evident from the expression of surface markers including activating macrophage mannose receptor 1 (MRC1)[43] and M2 markers CD68 and CD204 on macrophages[44], as well as CD86, CD163, CD204 and Siglec-1 on circulating monocytes[37, 45]. In addition, monocytes of SSc patients are characterized by an increased expression of type I IFN-regulated genes[45, 46] (generally referred to as the type I IFN signature), further highlighting their increased activation status in the disease process. Lastly, SSc monocytes have an increased potential to differentiate into myofibroblasts[47], a special type of fibroblast that is associated with SSc pathophysiology[48]. All this evidence highlights a major role for monocytes in SSc pathogenesis, and positions them at the cross-roads of inflammatory and fibrotic processes in the disease. However, the exact molecular mechanisms that initiate and maintain monocyte dysregulation in SSc are yet to be elucidated.

Dendritic cells

Besides monocytes, DCs have been put forth as important innate immune cells involved in SSc pathogenesis (Figure 2). DCs, named as such by Ralph Steinman and Zanvil Cohn because of their characteristic branching morphology[49], are another type of antigen-presenting cells that are a central part of the immune system as they bridge innate with adaptive immune responses. They are often referred to as the 'sentinels' of the immune system, as their main function is to take up and process antigens and present these to lymph-node-resident antigen-unexperienced T and B cells, thereby initiating adaptive immune responses[50]. Analogous to marker-based monocyte stratification in distinct functional groups, DCs can be further classified into distinct subtypes based on the expression of several surface markers as well as their morphology. Currently, two main groups of DCs have been described: conventional dendritic cells (cDCs, historically referred to as myeloid/mDCs, characterized by a typical branching morphology and high expression of CD11c) and plasmacytoid dendritic cells (pDCs, characterized by a plasma cell-like morphology and high expression of CD123)[51]. Whereas cDCs are most well known for their ability to activate T cells, pDCs are known for their ability to produce large amounts of IFNα. cDCs can be further subdivided into cDC1s and cDC2s based on the surface expression of CD141 or CD1c respectively. To further complicate DC classification, recent single-cell RNA sequencing and flow/mass cytometry studies have led to the identification further subdivision of DCs populations within the pDC, cDC1 and cDC2 compartments[52–55], with some having pro-inflammatory properties while others exhibit a more regulatory phenotype[56]. Thus, DCs are a rather heterogeneous population of bridging immune cells that represent critical regulators of immune responses.

Several studies have been performed to dissect the role of DCs in SSc pathogenesis (Figure 2). pDCs have been extensively investigated for their obvious link to the type I IFN signature that characterizes SSc patients. Indeed SSc pDCs have been shown to spontaneously secrete IFNα[57], and also produce higher levels of IFNα upon TLR stimulation compared to healthy pDCs[25, 57]. Additionally, in line with observations in monocytes, pDCs from SSc patients display a consistent upregulation of IFN responsive
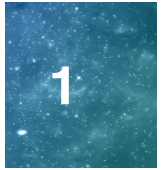
genes[58]. However, as opposed to monocytes, frequencies of pDCs are decreased in SSc circulation[57]. This disease-feature is proposed as a consequence of an increased migration of pDCs towards affected tissues of SSc patients. Indeed multiple studies have shown that these cells are increased in SSc skin and lung tissue[25, 57, 59]. Similar observations have been made for cDCs, which have also been found to be increased in dermal lesions in the skin of SSc patients[60]. This increased migration of DCs has been hypothesized to be a consequence of increased DC chemoattractants including CCL18, CCL19 and CXCL13[61]. In affected tissues, DCs have the potential to regulate other immune cells important in SSc pathogenesis, including fibroblasts and cells implicated in regulation of vascular remodeling during inflammation[62, 63]. SSc cDCs also produce increased levels of pro-fibrotic and pro-inflammatory cytokines including IL-6, IL-8, IL-12 and TNFα upon TLR stimulation[64, 65], further highlighting their role in the regulation of fibrotic and inflammatory processes in SSc. However, further detailed molecular profiling of these cells is needed to fully explore their contribution to disease pathogenesis.
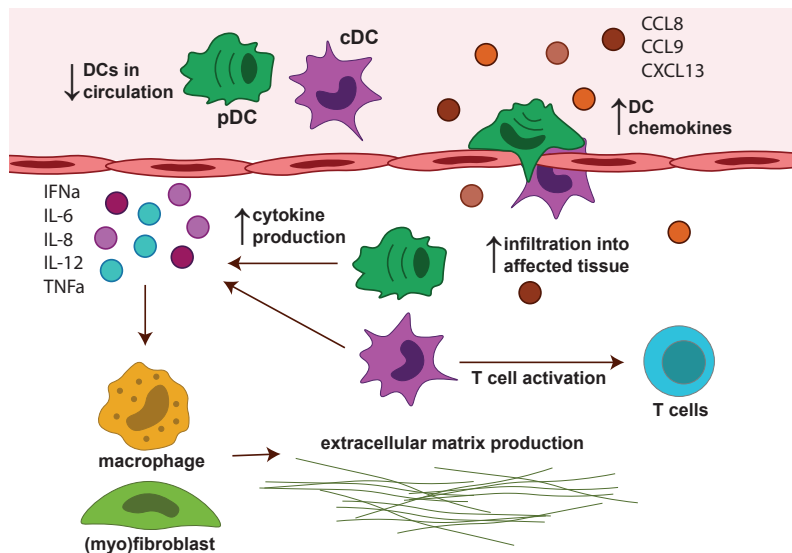
Figure 2. Dendritic cell dysregulation in the pathogenesis of SSc. Dendritic cells (DCs) are decreased in numbers in circulation of SSc patients. Increased chemokines (CCL8, CCL9 and CXCL13) can lead to the increased infiltration of plasmacytoid dendritic cells (pDCs) and classical dendritic cells (cDCs) observed in affected tissue of SSc patients. Here both cell types are capable of producing large amounts of pro-inflammatory and pro-fibrotic cytokines, and thereby stimulate macrophages and (myo)fibroblasts to deposit excessive amounts of extra cellular matrix material. Moreover, activated cDCs are potent T cell activators, putting them at the crossroads of adaptive and innate immune responses in SSc.

The adaptive immune system in systemic sclerosis

Alterations of functioning of the innate immune system in SSc likely contribute to priming of the adaptive immune system towards a state of hyper-activation[66]. Accordingly, the cross-talk between innate and adaptive immune cells perpetuates inflammation over the disease course. The adaptive arm of the immune system mainly

comprises T and B cells that are armed with highly variable and high affinity antigen-binding receptors. In contrast to PRRs expressed on innate immune cells that are antigen non-specific, the antigen-binding receptors of the adaptive immune system are highly specific to recognize highly selective antigens[67]. The high diversity of these antigen-binding receptors arises from the genetic rearrangement of specific DNA segments occurring during the early stages of T and B cell maturation, equipping every T cell with a highly variable T cell receptor (TCR) and every B cell with a highly variable B cell receptor (BCR)[68]. The high variability and specificity of these receptors underlies the breadth and potency of the adaptive immune system.

T cells

T cells are a major immune cell subset of the adaptive immune system. CD8+ T cells have the capacity to execute direct killing of pathogen-infected host cells, which is why they are generally referred to as cytotoxic T cells. CD4+ T cells, at large, have a crucial role in shaping the immune response through the activation of other immune cells, which is why they are generally referred to as helper T (Th) cells. Like many cells of the immune system, Th cells can be further classified into distinct subsets including the major Th1 and Th2 subsets as well as the more recently described Th17 subset[69]. Th1 generally develop in response to viral triggers and produce large amounts of IFNy. Th2 generally develop in response to extracellular parasites and bacteria, and facilitate the initiation of antibody-mediated immune responses through the secretion of IL-4, IL-5 and IL-13. Th17 form a subset of pro-inflammatory T cells that are characterized by the production of large amounts of IL-17, and are implicated in many autoimmune diseases[70]. Lastly, a subset of regulatory T cells (Tregs, characterized by the expression of surface markers CD4 and CD25, and the transcription factor FOXP3) have also been distinguished within the Th compartment. These cells represent a population of T lymphocytes with a suppressive capacity that is crucial for the maintenance of tolerance to auto-antigens[71]. The activation and polarization of all T cells relies on the recognition of antigens presented on the major histocompatibility complex (MHC) of antigen presenting cells (APCs, mostly DCs) through the TCR. Moreover, in order for T cells to become fully activated and clonally expand, a secondary signal is needed which is generated through signaling of costimulatory molecules present on APCs. Lastly, during activation, the polarization of Th subsets is generally induced through the secretion of diverse cocktails of T cell-polarizing cytokines by APCs, which are largely dependent on the inflammatory milieu and the danger signals to be eliminated[72]. This again highlights the importance of the communication between innate and adaptive immune cells in shaping a proper immune response.

T cells appear to have a prominent role in the pathogenesis of SSc. First, next to monocytes and DCs, skin of SSc patients is also characterized by the infiltration of cytotoxic and Th cells[73–75], along with a reduction of regulatory T cells[76]. Notably, these infiltrating T cells are characterized by the expression of CD69[77], an early activation marker[78], suggesting that infiltrating T cells are actively responding to antigens expressed in the skin of SSc patients. In line with these observations, circulating CD4+ and CD8+ T cells in peripheral blood of SSc patients exhibit an enhanced expression of
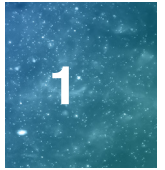
T cell activation markers including IL-2R, HLA-DR, and CD29[79, 80], and secrete large amounts of pro-inflammatory and pro-fibrotic cytokines following stimulation[81, 82]. Adding to the enhanced activation of SSc T cells, Treg function appears to be impaired in SSc patients[83], suggesting that a loss of inhibition also contributes to aberrant T cell responses in the disease. Regarding T cell phenotypes, the Th2 polarizing cytokines IL-4 and IL-13 are increased in circulation[84], and T cells with a predominant Th2 profile have been identified in SSc patients[85–88], leading to the proposal that SSc mainly is a Th2 driven disease. In line with this, Th2 cells are known to directly promote fibrosis through the secretion of pro-fibrotic cytokines IL-4, IL-13 and TGFβ, thereby stimulating fibroblasts as well as macrophages to produce large amounts of extracellular matrix material[89]. One of the most important outstanding questions regarding T cell dysregulation in SSc is whether enhanced T cell activation is driven by the direct recognition and response to specific (auto)antigens. The skin and circulating immune cell pool of SSc patients contain a remarkably oligoclonal TCR repertoire (meaning that an expansion of T cells carrying the same TCR has occurred)[90, 91], providing further evidence for an antigen specific T cell response. However, little is known about the antigen specificity and temporal dynamics of T cells in the disease pathogenesis. Thus, more detailed characterization of the TCR repertoire and the dynamics of T cells responses are required to better understand the role of antigen specific T cell responses in SSc.

Omics technologies to further unravel immune dysregulation in systemic sclerosis

As outlined above, it is clear that immune cell dysregulation plays a pivotal role in SSc pathogenesis. However, the molecular processes that drive immune cell activation are still incompletely understood. Therefore, a detailed molecular view of the factors underlying immune system dysregulation in SSc is needed to unravel the interplay between different immune cells in the disease pathogenesis. To accomplish this, vast amounts of data from various molecular layers and from different levels of cellular organization are required. These data can be obtained by applying different omics technologies, which yield detailed information on the genome (genomics), gene regulation (epigenomics), gene expression (transcriptomics), proteins (proteomics) and metabolites (metabolomics) (Figure 3). The use of these omics technologies to study rheumatic autoimmune diseases, including SSc, has rapidly increased over the past 20 years, especially when looking at transcriptomics and epigenomics studies (Figure 4A).

Genomics
DNA sequence alterations
• DNA-sequencing
• GWAS

Epigenomics
Reversible modifications of DNA or chromatin
• DNA methylation
• ChIP-seq
• ATAC-seq

Transcriptomics
Production of functional RNA species
• Micro-array
• RNA-sequencing
• Co-expression

Proteomics
Alterations in protein levels
• Mass spectrometry
• Immunoassays
• Protein arrays

Metabolomics
Monitoring enzymatic/cellular activity
• Mass spectometry
• NMR spectroscopy

Figure 3. Core 'omics' technologies: genomics, epigenomics, transcriptomics, proteomics and metabolomics.

A

B
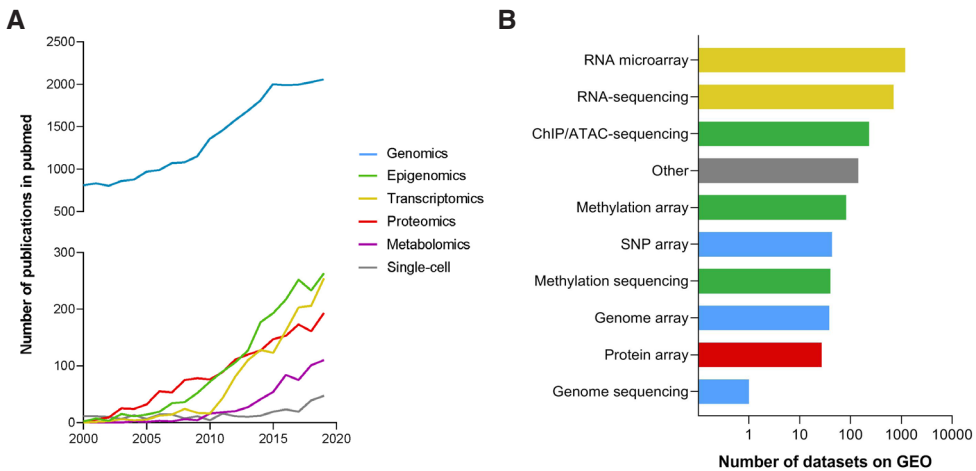
Figure 4. Number of publications in PubMed mentioning rheumatic autoimmune diseases and each omic in its title or abstract since the year 2000. Left: different omics technologies are indicated by different colors. Pubmed queries included: "rheumatology"[MeSH Terms] OR "musculoskeletal diseases"[MeSH Terms] OR "rheumatic diseases"[MeSH Terms] and -ome OR –omic search terms (e.g. genome or genomic), and were filtered for human studies only. Right: number of publicly available datasets available on the NCBI GEO database. Query: "rheumatology"[MeSH Terms] OR "musculoskeletal¬ diseases"[MeSH Terms] OR "rheumatic diseases"[MeSH Terms] (filtered for human). Datasets were categorized according to the type of data as indicated on GEO (indicated on the y-axis, color denotes the omics category as indicated in (A)). Retrieved on 18-02-2021.

Increasingly, omics datasets are being made publicly available, allowing for re-analysis and integration of independent datasets from different patient groups, cell types and tissues (Figure 4B). Following recent technological advances, these omics technologies can now even be applied on the level of single cells, providing a more detailed insight into the contribution of very specific cell populations to disease phenotypes. Application of these techniques to study immune cells in SSc should provide detailed molecular insights to help us better understand factors driving the dysregulation of these cells.

Genomics

Genomics is broadly defined as the systematic study of genetic information (the total DNA, or the genome) of an organism. Genetic mutations in intronic, extronic, and regulatory genomic regions can affect biological phenotypes and play a critical role in disease development. Therefore, assessing the structure and functioning of the genome can offer valuable insights into disease pathogenesis. To this end, a wide array of next generation sequencing (NGS) platforms are currently available, which can be used to gain insight into genomic variation between single individuals or whole patient populations[92]. These sequencing platforms come in different flavors, ranging from short read sequencing for the identification of single-nucleotide polymorphisms (SNPs), to ultra-long sequencing of large genomic regions for the investigation of more complex genomic variation (e.g. structural variants, or repeated regions)[93]. Multiple large genomic studies have already identified genetic variants associated with SSc. Recently, a large genome-wide association study in 9846 Systemic Sclerosis (SSc) patients and

18333 healthy individuals identified 13 new risk loci for SSc[94]. Notably, a large number of the genetic variants identified in this study were linked to regulatory regions. Moreover, specific associations could be linked to specific clinical subtypes of SSc, emphasizing the potential of genetic associations for classification of disease (sub)phenotypes. However, although several studies have identified genomic risk loci for SSc, the concordance rate of SSc in monozygotic twins is quite low (4,2%)[95], suggesting that genomics have a minimal contribution to the development of SSc.

Epigenomics

Since the evidence for genetic dispositions driving SSc is limited, and many genetic risk loci for SSc are linked to regulatory elements, epigenomic modifications are now being considered as imperative factors in SSc pathogenesis. Epigenomic modifications are reversible modifications of DNA or chromatin structures, which affect gene expression without altering the primary DNA sequence. The epigenome is defined as the complete set of epigenomic modifications of a cell, comprising histone modifications, open chromatin regions, DNA methylation and non-coding RNAs. The epigenomic landscape exists in a stable state, but it is also dynamic in the sense that changes can take place in response to environmental triggers, as is for example observed in immune cells in response to infection[96].

A wide range of molecular biology techniques have been developed to study different epigenomic modifications in a high throughput manner. Binding sites of DNA-associated proteins such as transcription factors and histone modifications can be studied using ChIP-seq (chromatin immunoprecipitation combined with parallel DNA sequencing). Chromatin accessibility can be investigated using ATAC-seq (Assay for Transposase-Accessible Chromatin using sequencing), which captures open chromatin sites representing actively transcribed genomic regions[97]. DNA methylation, which typically acts to repress gene transcription[98], can be studied in a high-throughput manner by whole genome bisulfite sequencing (WGBS)[99], or the more recently developed Illumina MethylationEPIC BeadChip assay[100]. These high throughput techniques can be used to identify epigenomic modifications that potentially underlie changes in gene expression in immune cells implicated in SSc.

Transcriptomics

Genomic alterations and changes in epigenomic marks together lead to an effect on gene expression, or in other words, the production of functional RNA. The production of an RNA copy from the DNA strand is referred to as transcription, and the transcriptome represents the sum of all RNA transcripts, ranging from mRNA to noncoding RNA. The transcriptome provides a snapshot of active/quiescent cellular processes, reflecting the effects of genomics and epigenomic alterations on gene expression. Thus, studying the transcriptome is important to understand how altered expression of genetic variants contributes to the development of complex phenotypes such as SSc. The transcriptome can be assessed on a large scale by microarrays or RNA-sequencing (RNA-seq). Microarrays work through the quantification of fluorescently labelled cDNA molecules that can hybridize with known DNA probes[101]. Because microarrays depend on the

hybridization with probes, they require prior knowledge on the genome of interest. Thus, specific genetic variants or unknown transcripts/isoforms cannot be detected using this method. To overcome this, RNA-seq has been developed, which employs next-generation sequencing technology to allow for the detection of unannotated genes and sequence level variations[102]. Moreover, RNA-seq data can be used to study long non-coding RNA transcripts, which are often not captured on microarrays. These long non-coding RNAs are transcripts with a minimal length of 200 nucleotides that do not possess any protein-coding capacity, but instead act as critical regulators of gene expression[103]. Additionally, specific transcriptomic sequencing methods have been developed to study variable TCR sequences at high resolution[104]. Next to studying expression patterns of single genes, transcriptomic data can also be used to construct gene co-expression networks. Gene co-expression networks consist of genes that display similar expression patterns across samples and are therefore thought to be functionally related[105]. Gene co-expression networks can thus be employed to describe the molecular state of a cell type/tissue using transcriptomic profiling on a broader, systems-level.

Proteomics

Although transcriptomics may provide valuable insights into gene-expression alterations, changes in transcript levels may not always correspond to analogous changes in protein levels due to multiple biological (e.g. mRNA stability, ribosome occupancy, protein turnover) as well as technical (e.g. experimental error and noise) factors[106]. Therefore, it is also important to study alterations at protein level to get a complete picture of the downstream effects of transcriptomic variation. The proteome encompasses the complete set of proteins produced in a system, and the large-scale study of these proteins using high throughput techniques is referred to as proteomics[107]. The gold standard high throughput technique in proteomics is mass spectrometry (MS), which can detect proteins using their fundamental biochemical properties such as mass and charge. MS is a complex technology that comes in numerous flavors and, besides the very specific detection and quantification of proteins, can also be used to study protein modifications and protein-protein interactions[108]. Recently, the MS based technique "cytometry by time-of-flight (CyToF)" has been developed to study immune cell proteomic profiles at a single cell level[109]. Because of its possibility to evaluate the expression of a large number of proteins on single cells (much more than traditional flow cytometry approaches), CyToF can be used to explore the immune cell landscape in patients with autoimmune disease in high resolution. These techniques are very helpful to dissect the precise contribution of individual immune cell subsets to disease pathogenesis.

Aims and outline of this thesis

By applying various high-throughput omics approaches, the studies presented in this thesis aim to explore the molecular mechanisms contributing to immune dysregulation in SSc. Following the duality-based characterization of the immune system, this thesis is split up in two distinct yet complementary parts. In the first part, I focus on the innate immune system in SSc, exploring the role of monocytes and DCs in the disease pathogenesis. **Chapter 2** describes the transcriptomic profiling of healthy as well as SSc monocytes

1

to identify long non-coding RNAs involved in the regulation of TLR4 induced type I IFN responses in these cells. The work presented in **chapter 3** further characterizes the roles of long non-coding RNAs in monocytes of SSc patients. Exploiting transcriptomic data of monocytes from SSc and healthy donors, in this work, multiple lncRNAs were identified to have a potential role in the regulation of apoptotic pathways, IFN signaling and cytokine secretion in SSc monocytes. In **chapter 4** I explored histone modifications characterizing the epigenomic landscape of monocytes from SSc patients as well as patients suffering from two other rheumatic autoimmune diseases: systemic lupus erythematosus (SLE) and rheumatoid arthritis (RA). Paired transcriptomic analysis and ChIP-seq allowed for the identification of disease specific and shared epigenomic imprinting underlying monocyte dysregulation in these three autoimmune diseases. Moving from monocytes to DCs, in **chapter 5**, I performed transcriptomic profiling combined with co-expression network analysis and transcription factor ChIP-seq to identify transcriptional regulators of pro-inflammatory responses in cDCs from healthy individuals and SSc patients.

In the second part of the thesis, I explore the role of the adaptive immune system in SSc. In **chapter 6**, I address the question of repertoire dynamics and antigen specificity of T cells in SSc. Herein I present a study of high-throughput sequencing of CD4+ and CD8+ TCR repertoires from blood samples obtained longitudinally from four SSc patients collected over a minimum of two years. To compare the dynamics of T cell responses of SSc to another autoimmune disease that is also characterized by local instead of systemic inflammation, in **chapter 7** the immune architecture and TCR repertoire dynamics of peripheral blood and affected joints of juvenile idiopathic arthritis patients is investigated. This study included T cell phenotyping by CyTOF as well as TCR sequencing of effector T cells and Tregs, with the aim to get detailed insights into the T cell populations involved in local autoimmune responses.

Finally, all the findings and insights from the studies presented in this thesis are summarized and discussed in **chapter 8**.
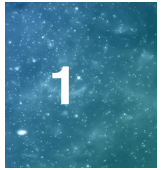
REFERENCES

1. W. Chan. A Source Book in Chinese Philosophy. *Princeton University Press*, 1963.
2. James Legge. The Writings of Chuang Tzu. *Oxford University Press*, 1891.
3. J. Larouche, S. Sheoran, K. Maruyama, et al. Immune Regulation of Skin Wound Healing: Mechanisms and Novel Therapeutic Targets. *Adv. Wound Care*, vol. 7, no. 7, pp. 209–231, Jul. 2018.
4. M. L. Disis. Immune Regulation of Cancer. *J. Clin. Oncol.*, vol. 28, no. 29, pp. 4531–4538, Oct. 2010.
5. K. Hoebe, E. Janssen, and B. Beutler, The interface between innate and adaptive immunity. *Nat. Immunol.*, vol. 5, no. 10, pp. 971–974, Oct. 2004.
6. D. L. Mueller. Mechanisms maintaining peripheral tolerance. *Nat. Immunol.*, vol. 11, no. 1, pp. 21–27, Jan. 2010.
7. G. Y. Liu, P. J. Fairchild, R. M. Smith, et al. Low avidity recognition of self-antigen by T cells permits escape from central tolerance. *Immunity*, vol. 3, no. 4, pp. 407–415, Oct. 1995.
8. C. Bouneaud, P. Kourilsky, and P. Bousso. Impact of Negative Selection on the T Cell Repertoire Reactive to a Self-Peptide. *Immunity*, vol. 13, no. 6, pp. 829–840, Dec. 2000.
9. A. Gabrielli, E. V. Avvedimento, and T. Krieg. Scleroderma. *N. Engl. J. Med.*, vol. 360, no. 19, pp. 1989–2003, May 2009.
10. F. van den Hoogen, D. Khanna, J. Fransen, et al. 2013 classification criteria for systemic sclerosis: an American college of rheumatology/European league against rheumatism collaborative initiative. *Ann. Rheum. Dis.*, vol. 72, no. 11, pp. 1747–1755, Nov. 2013.
11. E. C. LeRoy and T. A. Medsger. Criteria for the classification of early systemic sclerosis. *J. Rheumatol.,* vol. 28, no. 7, pp. 1573–1576, Jul. 2001
12. R. Mierau, P. Moinzadeh, G. Riemekasten et al. Frequency of disease-associated and other nuclear autoantibodies in patients of the German network for systemic scleroderma: correlation with characteristic clinical features. *Arthritis Res. Ther.,* vol. 13, no. 5, p. R172, 2011.
13. V. D. Steen. Autoantibodies in Systemic Sclerosis. *Semin. Arthritis Rheum.*, vol. 35, no. 1, pp. 35–42, Aug. 2005.
14. R. J. Prescott, A. J. Freemont, C. J. P. Jones, et al. Sequential dermal microvascular and perivascular changes in the development of scleroderma. *J. Pathol.*, vol. 166, no. 3, pp. 255–263, Mar. 1992.
15. D. J. Abraham, T. Krieg, J. Distler, et al. Overview of pathogenesis of systemic sclerosis. *Rheumatology (Oxford).*, vol. 48 Suppl 3, no. suppl_3, pp. iii3--7, Jun. 2009.
16. J. C. Worrell and S. O'Reilly. Bi-directional communication: Conversations between fibroblasts and immune cells in systemic sclerosis. *J. Autoimmun.*, vol. 113, p. 102526, Sep. 2020.
17. T. Kawasaki and T. Kawai. Toll-Like Receptor Signaling Pathways. *Front. Immunol.*, vol. 5, Sep. 2014.
18. Y. Kumagai and S. Akira. Identification and functions of pattern-recognition receptors. *J. Allergy Clin. Immunol.*, vol. 125, no. 5, pp. 985–992, May 2010.
19. L. Yu, L. Wang, and S. Chen. Endogenous toll-like receptor ligands and their biological significance. *J. Cell. Mol. Med.*, vol. 14, no. 11, pp. 2592–2603, Nov. 2010.
20. R. Lande, E. Y. Lee, R. Palazzo, et al. CXCL4 assembles DNA into liquid crystalline complexes to amplify TLR9-mediated interferon-α production in systemic sclerosis. *Nat. Commun.*, vol. 10, no. 1, p. 1731, Dec. 2019.
21. S. Bhattacharyya, Z. Tamaki, W. Wang, et al. FibronectinEDA Promotes Chronic Cutaneous Fibrosis Through Toll-Like Receptor Signaling. *Sci. Transl. Med.*, vol. 6, no. 232, pp. 232ra50-232ra50, Apr. 2014.
22. D. Kim, A. Peck, D. Santer, et al. Induction of interferon-α by scleroderma sera containing autoantibodies to topoisomerase I: Association of higher interferon-α activity with lung fibrosis. *Arthritis Rheum.*, vol. 58, no. 7, pp. 2163–2173, Jul. 2008.
23. P. Gourh, F.C. Arnett, S. Assassi, et al. Plasma cytokine profiles in systemic sclerosis: associations with autoantibody subsets and clinical manifestations. *Arthritis Res. Ther.*, vol. 11, no. 5, p. R147, 2009.
24. H. L. Zhu, Q. DU, W. L. Chen, et al. Altered serum cytokine expression profile in systemic sclerosis and its regulatory mechanisms. *Beijing Da Xue Xue Bao.*, 2019.
25. L. Van Bon, A.J. Affandi, J. Broen, et al. Proteome-wide analysis and CXCL4 as a biomarker in systemic sclerosis. *N. Engl. J. Med.*, vol. 370, no. 5, pp. 433–443, 2014.
26. H. Higley, K. Persichitte, S. Chu, et al. Immunocytochemical Localization and Serologic Detection of Transforming Growth Factor β1. *Arthritis Rheum.*, vol. 37, no. 2, pp. 278–288, Feb. 1994.
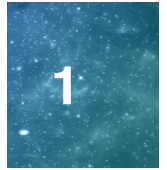
27. M. van der Kroef, T. Carvalheiro, M. Rossato, et al. CXCL4 triggers monocytes and macrophages to produce PDGF-BB, culminating in fibroblast activation: Implications for systemic sclerosis. *J. Autoimmun.,* vol. 111, p. 102444, Jul. 2020.

28. L. Ziegler-Heitbrock, P. Ancuta, S. Crowe, et al. Nomenclature of monocytes and dendritic cells in blood. *Blood*, vol. 116, no. 16, pp. e74–e80, Oct. 2010.

29. F. Tacke and G. J. Randolph. Migratory fate and differentiation of blood monocyte subsets. *Immunobiology*, vol. 211, no. 6–8, pp. 609–618, Sep. 2006.

30. C. D. Mills, K. Kincaid, J. M. Alt, et al. M-1/M-2 Macrophages and the Th1/Th2 Paradigm. *J. Immunol.*, vol. 164, no. 12, pp. 6166–6173, Jun. 2000.

31. P. Krzyszczyk, R. Schloss, A. Palmer, et al. The Role of Macrophages in Acute and Chronic Wound Healing and Interventions to Promote Pro-wound Healing Phenotypes. *Front. Physiol.*, vol. 9, May 2018.

32. C. Jakubzick, E.L. Gautier, S.L. Gibbings, et al. Minimal Differentiation of Classical Monocytes as They Survey Steady-State Tissues and Transport Antigen to Lymph Nodes. *Immunity*, vol. 39, no. 3, pp. 599–610, Sep. 2013.

33. C. E. Bergmann, I.E. Hoefer, B. Meder, et al. Arteriogenesis depends on circulating monocytes and macrophage accumulation and is severely depressed in op/op mice. J*. Leukoc. Biol.*, vol. 80, no. 1, pp. 59–65, May 2006.

34. M. E. Ogle, C. E. Segar, S. Sridhar, et al. Monocytes and macrophages in tissue repair: Implications for immunoregenerative biomaterial design. *Exp. Biol. Med.*, vol. 241, no. 10, pp. 1084–1097, May 2016.

35. G. Varga and D. Foell. Anti-inflammatory monocytes—interplay of innate and adaptive immunity. *Mol. Cell. Pediatr.*, vol. 5, no. 1, p. 5, Dec. 2018.

36. S. K. Mathai, M. Gulati, X. Peng, et al. Circulating monocytes from systemic sclerosis patients with interstitial lung disease show an enhanced profibrotic phenotype. *Lab. Investig.*, vol. 90, no. 6, pp. 812–823, Jun. 201.

37. N. Higashi-Kuwata, M. Jinnin, T. Makino, et al. Characterization of monocyte/macrophage subsets in the skin and peripheral blood derived from patients with systemic sclerosis. *Arthritis Res. Ther.*, vol. 12, no. 4, p. R128, 2010.

38. M. Kroef, L.L. van den Hooge, J.S. Mertens et al. Cytometry by time of flight identifies distinct signatures in patients with systemic sclerosis, systemic lupus erythematosus and Sjögrens syndrome. *Eur. J. Immunol.*, vol. 50, no. 1, pp. 119–129, Jan. 2020.

39. M. Hasegawa, S. Sato, and K. Takehara. Augmented production of chemokines (monocyte chemotactic protein-1 (MCP-1), macrophage inflammatory protein-1α (MIP-1α) and MIP-1β) in patients with systemic sclerosis: MCP-1 and MIP-1α may be involved in the development of pulmonary fibrosis. *Clin. Exp. Immunol.*, vol. 117, no. 1, pp. 159–165, Jul. 1999.

40. O. Distler, T. Pap, O. Kowal-Bielecka, et al. Overexpression of monocyte chemoattractant protein 1 in systemic sclerosis: Role of platelet-derived growth factor and effects on monocyte chemotaxis and collagen synthesis. *Arthritis Rheum.*, vol. 44, no. 11, pp. 2665–2678, Nov. 2001.

41. B. M. Kräling, G. G. Maul, and S. A. Jimenez. Mononuclear cellular infiltrates in clinically involved skin from patients with systemic sclerosis of recent onset predominantly consist of monocytes/macrophages. *Pathobiology*, vol. 63, no. 1, pp. 48–56, 1995.

42. O. Ishikawa and H. Ishikawa. Macrophage infiltration in the skin of patients with systemic sclerosis. *J. Rheumatol.*, vol. 19, no. 8, pp. 1202–6, Aug. 1992

43. R. B. Christmann, E. Hayes, S. Pendergrass, et al. Interferon and alternative activation of monocyte/macrophages in systemic sclerosis-associated pulmonary arterial hypertension. *Arthritis Rheum.*, vol. 63, no. 6, pp. 1718–1728, Jun. 2011.

44. N. Higashi-Kuwata, T. Makino, Y. Inoue, et al. Alternatively activated macrophages (M2 macrophages) in the skin of patient with localized scleroderma. *Exp. Dermatol.*, vol. 18, no. 8, pp. 727–729, 2009.

45. M. R. York, T. Nagai, A. J. Mangini, et al. A macrophage marker, siglec-1, is increased on circulating monocytes in patients with systemic sclerosis and induced by type i interferons and toll-like receptor agonists. *Arthritis Rheum.*, vol. 56, no. 3, pp. 1010–1020, 2007.

46. Z. Brkic, L. van Bon, M. Cossu, et al. The interferon type I signature is present in systemic sclerosis before overt fibrosis and might contribute to its pathogenesis through high BAFF gene expression and high collagen synthesis. *Ann. Rheum. Dis.*, vol. 75, no. 8, pp. 1567–1573, Aug. 2016.

47. N. Binai, S. O'Reilly, B. Griffiths, et al. Differentiation potential of CD14 + monocytes into myofibroblasts in patients with systemic sclerosis. *PLoS One*, vol. 7, no. 3, pp. 1–7, 2012.

48. A. van Caam, M. Vonk, F. van den Hoogen, et al. Unraveling SSc Pathophysiology; The Myofibroblast. *Front. Immunol.*, vol. 9, Nov. 2018.
49. R. M. Steinman and Z. A. Cohn. Identification of a novel cell type in peripheral lymphoid organs of mice. I. Morphology, quantitation, tissue distribution. J*. Exp. Med.*, vol. 137, no. 5, pp. 1142–1162, May 1973.
50. J. Banchereau and R. M. Steinman. Dendritic cells and the control of immunity. *Nature*, vol. 392, no. 6673, pp. 245–252, Mar. 1998.
51. C. Macri, E. S. Pang, T. Patton, et al. Dendritic cell subsets. *Semin. Cell Dev. Biol.*, vol. 84, pp. 11–21, Dec. 2018.
52. A.-C. Villani, R. Satija, G. Reynolds, et al. Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. *Science* (80-.)., vol. 356, no. 6335, p. eaah4573, Apr. 2017.
53. C.-A. Dutertre, E. Becht, S.E. Irac, et al. Single-Cell Analysis of Human Mononuclear Phagocytes Reveals Subset-Defining Markers and Identifies Circulating Inflammatory Dendritic Cells. *Immunity*, vol. 51, no. 3, pp. 573-589.e8, Sep. 2019.
54. C. C. Brown, H. Gudjonson, Y. Pritykin et al. Transcriptional Basis of Mouse and Human Dendritic Cell Heterogeneity. *Cell*, vol. 179, no. 4, pp. 846-863.e24, Oct. 2019.
55. S. G. Alculumbre, V. Saint-André, J. Di Domizio, et al. Diversification of human plasmacytoid predendritic cells in response to a single stimulus. *Nat. Immunol.,* vol. 19, no. 1, pp. 63–75, Jan. 2018.
56. J. Villar and E. Segura. Decoding the Heterogeneity of Human Dendritic Cell Subsets. *Trends Immunol.*, vol. 41, no. 12, pp. 1062–1071, Dec. 2020.
57. M. D. A. Kioon, C. Tripodo, D. Fernandez, et al. Plasmacytoid dendritic cells promote systemic sclerosis with a key role for TLR8. *Sci. Transl. Med.*, vol. 10, no. 423, pp. 1–14, 2018.
58. E. Chouri, M. Wang, M.R. Hillen, et al. Implication of miR-126 and miR-139-5p in Plasmacytoid Dendritic Cell Dysregulation in Systemic Sclerosis. *J. Clin. Med.*, vol. 10, no. 3, p. 491, Jan. 2021.
59. S. Kafaja, I. Valera, A.A. Divekar, et al. pDCs in lung and skin fibrosis in a bleomycin-induced model and patients with systemic sclerosis. *JCI Insight*, vol. 3, no. 9, May 2018.
60. S. Mokuda, T. Miyazaki, Y. Ubara, et al. CD1a+ survivin+ dendritic cell infiltration in dermal lesions of systemic sclerosis. *Arthritis Res. Ther.*, vol. 17, no. 1, p. 275, Dec. 2015.
61. A. L. Mathes, R.B. Christmann, G. Stifano, et al. Global chemokine expression in systemic sclerosis (SSc): CCL19 expression correlates with vascular inflammation in SSc skin. *Ann. Rheum. Dis.*, vol. 73, no. 10, pp. 1864–1872, Oct. 2014.
62. T.-C. Tzeng, S. Chyou, S. Tian, et al. CD11c hi Dendritic Cells Regulate the Re-establishment of Vascular Quiescence and Stabilization after Immune Stimulation of Lymph Nodes. *J. Immunol.*, vol. 184, no. 8, pp. 4247–4257, Apr. 2010.
63. L. Ding, T. Liu, Z. Wu, et al. Bone Marrow CD11c + Cell–Derived Amphiregulin Promotes Pulmonary Fibrosis. *J. Immunol.*, vol. 197, no. 1, pp. 303–312, Jul. 2016.
64. L. van Bon, C. Popa, R. Huijbens, et al. Distinct evolution of TLR-mediated dendritic cell cytokine secretion in patients with limited and diffuse cutaneous systemic sclerosis. *Ann. Rheum. Dis.*, vol. 69, no. 8, pp. 1539–1547, Aug. 2010.
65. T. Carvalheiro, S. Horta, J.A.G. van Roon et al. Increased frequencies of circulating CXCL10-, CXCL8- and CCL4-producing monocytes and Siglec-3-expressing myeloid dendritic cells in systemic sclerosis patients. *Inflamm. Res.*, vol. 67, no. 2, pp. 169–177, Feb. 2018.
66. P. Laurent, V. Sisirak, E. Lazaro, et al. Innate Immunity in Systemic Sclerosis Fibrosis: Recent Advances. *Front. Immunol.*, vol. 9, Jul. 2018.
67. J. Charles A Janeway, P. Travers, et al. Antigen Recognition by B-cell and T-cell Receptors. *Immunobiology: The Immune System in Health and Disease,* 2001.
68. C. Brack, M. Hirama, R. Lenhard-Schuller, et al. A complete immunoglobulin gene is created by somatic recombination. *Cell*, vol. 15, no. 1, pp. 1–14, Sep. 1978.
69. S. L. Swain, K. K. McKinstry, and T. M. Strutt. Expanding roles for CD4+ T cells in immunity to viruses. *Nat. Rev. Immunol.*, vol. 12, no. 2, pp. 136–148, Feb. 2012.
70. K. Yasuda, Y. Takeuchi, and K. Hirota. The pathogenicity of Th17 cells in autoimmune diseases. *Semin. Immunopathol.*, vol. 41, no. 3, pp. 283–297, May 2019.
71. S. Z. Josefowicz, L.-F. Lu, and A. Y. Rudensky. Regulatory T Cells: Mechanisms of Differentiation and Function. *Annu. Rev. Immunol.*, vol. 30, no. 1, pp. 531–564, Apr. 2012.
72. M. L. Kapsenberg. Dendritic-cell control of pathogen-driven T-cell polarization. *Nat. Rev. Immunol.*,
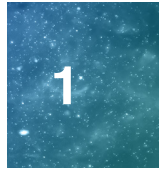
vol. 3, no. 12, pp. 984–993, Dec. 2003.

73. R. Fleischmajer, J. S. Perlish, and J. R. T. Reeves. Cellular infiltrates in scleroderma skin. *Arthritis Rheum.*, vol. 20, no. 4, pp. 975–984, May 1977.

74. A. D. Roumm, T. L. Whiteside, T. A. Medsger, et al. Lymphocytes in the skin of patients with progressive systemic sclerosis. *Arthritis Rheum.*, vol. 27, no. 6, pp. 645–653, Jun. 1984.

75. T. Maehara, N. Kaneko, C.A. Perugino, et al. Cytotoxic CD4+ T lymphocytes may induce endothelial cell apoptosis in systemic sclerosis. *J. Clin. Invest.*, vol. 130, no. 5, pp. 2451–2464, Apr. 2020.

76. S. Klein, C.C. Kretz, V. Ruland, et al. Reduction of regulatory T cells in skin lesions but not in peripheral blood of patients with systemic scleroderma. *Ann. Rheum. Dis.*, vol. 70, no. 8, pp. 1475–1481, Aug. 2011.

77. A. Kalogerou, E. Gelou, S. Mountantonakis, et al. Early T cell activation in the skin from patients with systemic sclerosis. *Ann. Rheum. Dis.*, vol. 64, no. 8, pp. 1233–1235, Aug. 2005.

78. S. F. Ziegler, F. Ramsdell, and M. R. Alderson. The activation antigen CD69. *Stem Cells*, vol. 12, no. 5, pp. 456–465, 1994.

79. A. Kahan, A. Kahan, F. Picard, et al. Abnormalities of T lymphocyte subsets in systemic sclerosis demonstrated with anti-CD45RA and anti-CD29 monoclonal antibodies. *Ann. Rheum. Dis.,* vol. 50, no. 6, pp. 354–358, Jun. 1991.

80. L. Parolin Ercole, M. Malvezzi, A. C. Boaretti, et al. Analysis of lymphocyte subpopulations in systemic sclerosis. *J. Investig. Allergol. Clin. Immunol.,* 2003.

81. P. Fuschiotti, A. T. Larregina, J. Ho, et al. Interleukin-13-producing CD8+ T cells mediate dermal fibrosis in patients with systemic sclerosis. *Arthritis Rheum.*, vol. 65, no. 1, pp. 236–246, Jan. 2013.

82. P. Fuschiotti. T cells and cytokines in systemic sclerosis. *Curr. Opin. Rheumatol.*, vol. 30, no. 6, pp. 594–599, Nov. 2018.

83. T. R. D. J. Radstake, L. van Bon, J. Broen, et al. Increased Frequency and Compromised Function of T Regulatory Cells in Systemic Sclerosis (SSc) Is Related to a Diminished CD69 and TGFβ Expression. *PLoS One*, vol. 4, no. 6, p. e5981, Jun. 2009.

84. M. Hasegawa, M. Fujimoto, K. Kikuchi, et al. Elevated serum levels of interleukin 4 (IL-4), IL-10, and IL-13 in patients with systemic sclerosis. *J. Rheumatol.*, vol. 24, no. 2, pp. 328–332, 1997.

85. F. Meloni, N. Solari, L. Cavagna, et al. Frequency of Th1, Th2 and Th17 producing T lymphocytes in bronchoalveolar lavage of patients with systemic sclerosis. *Clin. Exp. Rheumatol.*, vol. 27, no. 5, pp. 765–72, 2009.

86. C. Mavalia, C. Scaletti, P. Romagnani, et al. Type 2 helper T-cell predominance and high CD30 expression in systemic sclerosis. *Am. J. Pathol.*, vol. 151, no. 6, pp. 1751–8, Dec. 1997.

87. M. E. Truchetet, N. C. Brembilla, E. Montanari, et al. Increased frequency of circulating Th22 in addition to Th17 and Th2 lymphocytes in systemic sclerosis: Association with interstitial lung disease. *Arthritis Res. Ther.*, vol. 13, no. 5, p. R166, Oct. 2011.

88. J. Tang, L. Lei, J. Pan, et al. Higher levels of serum interleukin-35 are associated with the severity of pulmonary fibrosis and Th2 responses in patients with systemic sclerosis. *Rheumatol. Int.*, vol. 38, no. 8, pp. 1511–1519, Aug. 2018.

89. L. Barron and T. A. Wynn. Fibrosis is regulated by Th2 and Th17 responses and by dynamic interactions between fibroblasts and macrophages. *Am. J. Physiol. Liver Physiol.*, vol. 300, no. 5, pp. G723–G728, May 2011.

90. L. I. Sakkas, B. Xu, C. M. Artlett, et al. Oligoclonal T Cell Expansion in the Skin of Patients with Systemic Sclerosis. *J. Immunol.*, vol. 168, no. 7, pp. 3649–3659, 2002.

91. A. Kreuter, S. Höxtermann, C. Tigges, et al. Clonal T-cell populations are frequent in the skin and blood of patients with systemic sclerosis. *Br. J. Dermatol.*, vol. 161, no. 4, pp. 785–790, Oct. 2009.

92. A. M. Kanzi, J.E. San, B. Chimukangara, et al. Next Generation Sequencing and Bioinformatics Analysis of Family Genetic Inheritance. *Front. Genet.*, vol. 11, no. October, pp. 1–18, 2020.

93. T. Xiao and W. Zhou. The third generation sequencing: the advanced approach to genetic diseases. *Transl. Pediatr.*, vol. 9, no. 2, pp. 163–173, Apr. 2020.

94. E. López-Isac, M. Acosta-Herrera, M. Kerick, et al. GWAS for systemic sclerosis identifies multiple risk loci and highlights fibrotic and vasculopathy pathways. *Nat. Commun.*, vol. 10, no. 1, p. 4955, Dec. 2019.

95. C. Feghali-Bostwick, T. A. Medsger, and T. M. Wright. Analysis of systemic sclerosis in twins reveals low concordance for disease and high concordance for the presence of antinuclear antibodies. *Arthritis Rheum.,* vol. 48, no. 7, pp. 1956–1963, Jul. 2003.

96. Q. Zhang and X. Cao. Epigenetic regulation of the innate immune response to infection. *Nat. Rev.*

Immunol., vol. 19, no. 7, pp. 417–432, Jul. 2019.

97. J. D. Buenrostro, P. G. Giresi, L. C. Zaba, et al. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods*, vol. 10, no. 12, pp. 1213–1218, 2013.

98. P. L. Jones, G.J. Veenstra, P.A. Wade, et al. Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. *Nat. Genet.*, vol. 19, no. 2, pp. 187–191, Jun. 1998.

99. M. Frommer, L.E. McDonald, D.S. Millar, et al. A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proc. Natl. Acad. Sci.*, vol. 89, no. 5, pp. 1827–1831, Mar. 1992.

100. R. Pidsley, E. Zotenko, T.J. Peters et al. Critical evaluation of the Illumina MethylationEPIC BeadChip microarray for whole-genome DNA methylation profiling. *Genome Biol.*, vol. 17, no. 1, p. 208, Dec. 2016.

101. M. Schena, D. Shalon, R. W. Davis, et al. Quantitative Monitoring of Gene Expression Patterns with a Complementary DNA Microarray. *Science* (80-.)., vol. 270, no. 5235, pp. 467–470, Oct. 1995.

102. F. Ozsolak and P. M. Milos. RNA sequencing: advances, challenges and opportunities. *Nat. Rev. Genet.,* vol. 12, no. 2, pp. 87–98, Feb. 2011.

103. L. Statello, C. J. Guo, L. L. Chen, et al. Gene regulation by long non-coding RNAs and its biological functions. *Nat. Rev. Mol. Cell Biol.*, vol. 22, no. February, 2020.

104. E. Rosati, C. M. Dowds, E. Liaskou, et al. Overview of methodologies for T-cell receptor repertoire analysis. *BMC Biotechnol.*, vol. 17, no. 1, p. 61, Dec. 2017.

105. B. Zhang and S. Horvath. A General Framework for Weighted Gene Co-Expression Network Analysis. *Stat. Appl. Genet. Mol. Biol.*, vol. 4, no. 1, Jan. 2005.

106. T. Maier, M. Güell, and L. Serrano. Correlation of mRNA and protein in complex biological samples. *FEBS Lett.*, vol. 583, no. 24, pp. 3966–3973, Dec. 2009.

107. W. P. Blackstock and M. P. Weir. Proteomics: quantitative and physical mapping of cellular proteins. *Trends Biotechnol.*, vol. 17, no. 3, pp. 121–127, Mar. 1999.

108. Z. Zhang, S. Wu, D. L. Stenoien, et al. High-Throughput Proteomics. *Annu. Rev. Anal. Chem.*, vol. 7, no. 1, pp. 427–454, Jun. 2014.

109. D. R. Bandura, V.I. Baranov, O.I. Ornatsky, et al. Mass Cytometry: Technique for Real Time Single Cell Multitarget Immunoassay Based on Inductively Coupled Plasma Time-of-Flight Mass Spectrometry. *Anal. Chem.*, vol. 81, no. 16, pp. 6813–6822, Aug. 2009.

1
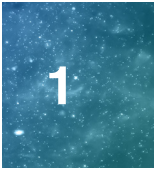
Part 1

Innate immunity in Systemic Sclerosis

Chapter 2

2

The Long Non-coding RNA NRIR Drives IFN-Response in Monocytes: Implication for Systemic Sclerosis

Authors:

B. Mariotti[1], N.H. Servaas[2,3], M. Rossato[2,3,4], N. Tamassia[1], M.A. Cassatella[1], M. Cossu[2,3], L. Beretta[5], M. van der Kroef[2,3], T.R.D.J. Radstake[2,3†], F. Bazzoni[1†]

[†] These authors contributed equally

Affiliations:

1. General Pathology Section, Department of Medicine, University of Verona, Verona, Italy.
2. Department of Rheumatology and Clinical Immunology, University Medical Center Utrecht, Utrecht, Netherlands.
3. Laboratory of Translational Immunology, University Medical Center Utrecht, Utrecht, Netherlands.
4. Department of Biotechnology, University of Verona, Verona, Italy.
5. Scleroderma Unit, Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico di Milano, Referral Center for Systemic Autoimmune Diseases, Milan, Italy.

ABSTRACT

TLR4 activation initiates a signaling cascade leading to the production of type I IFNs and of the downstream IFN-stimulated genes (ISGs). Recently, a number of IFN-induced long non-coding RNAs (lncRNAs) that feed-back regulate the IFN response have been identified. Dysregulation of this process, collectively known as the "Interferon (IFN) Response," represents a common molecular basis in the development of autoimmune and autoinflammatory disorders. Concurrently, alteration of lncRNA profile has been described in several type I IFN-driven autoimmune diseases. In particular, both TLR activation and the upregulation of ISGs in peripheral blood mononuclear cells have been identified as possible contributors to the pathogenesis of systemic sclerosis (SSc), a connective tissue disease characterized by vascular abnormalities, immune activation, and fibrosis. However, hitherto, a potential link between specific lncRNA and the presence of a type I IFN signature remains unclear in SSc. In this study, we identified, by RNA sequencing, a group of lncRNAs related to the IFN and anti-viral response consistently modulated in a type I IFN-dependent manner in human monocytes in response to TLR4 activation by LPS. Remarkably, these lncRNAs were concurrently upregulated in a total of 46 SSc patients in different stages of their disease as compared to 18 healthy controls enrolled in this study. Among these lncRNAs, Negative Regulator of the IFN Response (NRIR) was found significantly upregulated in vivo in SSc monocytes, strongly correlating with the IFN score of SSc patients. Weighted Gene Co-expression Network Analysis showed that NRIR-specific modules, identified in the two datasets, were enriched in "type I IFN" and "viral response" biological processes. Protein coding genes common to the two distinct NRIR modules were selected as putative NRIR target genes. Fifteen in silico-predicted NRIR target genes were experimentally validated in NRIR-silenced monocytes. Remarkably, induction of CXCL10 and CXCL11, two IFN-related chemokines associated with SSc pathogenesis, was reduced in NRIR-knockdown monocytes, while their plasmatic level was increased in SSc patients. Collectively, our data show that NRIR affects the expression of ISGs and that dysregulation of NRIR in SSc monocytes may account, at least in part, for the type I IFN signature present in SSc patients.

INTRODUCTION

Toll-like receptor 4 (TLR4) is a member of the pattern recognition receptors (PRR) family, which detects conserved structures found in a broad range of pathogens and triggers innate immune responses. TLR4 signals through two major pathways: (i) the MyD88-dependent pathway, that elicits the release of pro-inflammatory cytokines, such as TNF-α, IL-6, and IL-12p40; (ii) the TRIF (Toll/IL-1R domain-containing adaptor- inducing IFN-beta)-dependent pathway, that contributes to pro-inflammatory cytokine responses and, most importantly, induces type I IFN responses, particularly IFN-β[1]. IFNs confer their activity by regulating networks of interferon-stimulated genes (ISGs), a process that requires de novo transcription and translation of both IFN and downstream ISGs[2]. Other than being activated by different exogenous pathogen-associated molecular patterns (PAMPs), the IFN pathway is activated also by TLR4 ligation of endogenous danger-associated molecular patters (DAMPs) released upon cell damage or stress[3, 4]. Thus, TLR4-mediated activation of innate immunity plays a key role not only in host defense against pathogens but also in numerous autoimmune diseases, including systemic sclerosis (SSc)[5]. Indeed, endogenous ligand-induced TLR4 activation has been recognized as a key player driving the persistent fibrotic response in SSc[5–7]. Different endogenous TLR4 ligands, including fibronectin extra domain A (FnEDA) and S100A8/A9, are indeed increased in the circulation of SSc patients and have been correlated with fibrotic-related clinical complications[8, 9]. Moreover, activation of TLR4 response leads to transforming growth factor-β production, a crucial mediator for fibrosis development in SSc[10].

Likewise, production of type I interferon is closely linked to TLR4-mediated innate immune signaling in SSc[11–13]. In fact, several lines of evidence suggest that both the IFN network and monocytes are implicated in SSc immune-pathogenesis. First, the development of SSc has been reported in patients undergoing IFN treatment[14] and IFN-α injections worsen SSc-related clinical features[15]. Most importantly, increased expression of type I IFN-regulated genes, known as "type I IFN signature," is a hallmark of SSc, and type I IFN signature is present both in the fibrotic skin and in peripheral blood cells[11, 13], as well as in monocytes of SSc patients from the earliest phases of the disease, even before the skin fibrosis is evident[16]. Moreover, in the fibrotic subsets of SSc patients we identified an increase in non-classical monocytes spontaneously producing the IFN-responsive CXCL10[17], a chemokine associated with faster progression rate from pre-fibrotic SSc to worse disease stages[18].

The IFN pathway downstream TLR4 activation has been focus of intense investigation and a number of known protein-mediated mechanisms that mediate the complex signaling pathways and gene expression programs involved in the interferon response have been identified[2]. Recent studies point at long non-coding RNAs (lncRNAs) as a novel class of IFN pathway regulatory molecules[19]. LncRNAs are RNA transcripts longer than 200 nucleotides, characterized by lacking protein coding capability, but able to regulate gene expression both at the transcriptional and post-transcriptional levels[20]. Existing data indicate that lncRNAs are critically involved in various biological and immunological processes[21], including several pathways related to innate immunity[22–29]. However, with respect to the IFN response, while IFN-induced
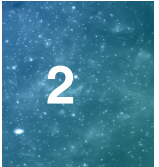
changes in the expression of protein-coding RNAs and their functional outcome have been well-documented, our knowledge of the impact of IFNs on lncRNA genes is highly incomplete. Moreover, the involvement of lncRNAs in diseases such as SSc, where both TLR4 and type I IFN concur to disease pathogenesis, is still unexplored.

This study aims to investigate the profile and the role of lncRNAs in the IFN response initiated by TLR4 activation of primary human monocytes and their implication in the immune dysregulation present in SSc patients.

MATERIALS AND METHODS

Patients

Patients affected by systemic sclerosis (SSc) and sex- and age-matched healthy controls (HC) were obtained from the University Medical Center Utrecht (UMCU), The Netherlands, and the Scleroderma Unit of Fondazione IRCCS Policlinico of Milan, Italy. Patients fulfilling the ACR/EULAR 2013 criteria[30] were classified in relation to the extent of skin fibrosis as limited cutaneous (lcSSc) or diffuse cutaneous SSc (dcSSc) [31]; patients satisfying the classification criteria without skin fibrosis were referred to as non-cutaneous SSc (ncSSc). Additionally, early SSc (eaSSc) subjects were defined as patients presenting with Raynaud's phenomenon and SSc-specific autoantibodies and/ or typical nailfold videocapillaroscopy abnormalities[32]. Three separate cohorts, herein named "definite SSc" cohort, "non-fibrotic SSc" cohort, "SSc cohort 3," were recruited for the current study. Demographics and clinical characteristics of the three cohorts are reported in Tables 1–3. All patients and healthy donors signed an informed consent form approved by the local institutional review boards prior to participation in the study. Samples and clinical information were made de-identified immediately after collection.

Patient group (n)	HC (9)	ncSSc (7)	lcSSc (11)	dcSSc (7)
Age (yr.)	52 (30–64)	45 (26–63)	59 (45–70)	58 (34–72)
Female (n, %)	5 (56%)	6 (86%)	8 (73%)	3 (43%)
ANA (n pos, %)	–	6 (86%)	10 (91%)	7 (100%)
ACA (n pos, %)	–	3 (43%)	6* (55%)	1 (14%)
Scl70 (n pos, %)	–	2 (29%)	2* (18%)	4 (57%)
mRSS	–	0	6 (0–12)	14* (5–36)
ILD	–	1 (14%)	2 (18%)	5 (71%)
Disease Duration (yr.)	–	4 (1–12)	9 (1–19)	10 (2–27)

Table 1. Demographics and clinical characteristics of the donors included in the definite SSc cohort. Values reported indicate the number (n) of patients and the median for each parameter [Interquartile Range (IQR)], if not otherwise indicated. ACA, anticentromere antibodies; ANA, antinuclear antibodies; dcSSc, diffuse cutaneous SSc; HC, healthy controls; ILD, Interstitial Lung disease; lcSSc, limited cutaneous SSc; mRSS, modified Rodman Skin score; ncSSc, non-cutaneous SSc; pos, positivity; Scl70, anti-topoisomerase antibodies; Yr., years. *1 patient unknown.
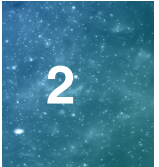
Patient group (n)	HC (9)	eaSSc (11)	ncSSc (10)
Age (yr.)	38 (28–49)	57 (40–77)	52 (25–70)
Female (n, %)	9 (100%)	11 (100%)	10 (100%)
ANA (n pos, %)	–	10 (91%)	10 (100%)
ACA (n pos, %)	–	7 (64%)	8 (80%)
Scl70 (n pos, %)	–	2 (18%)	1 (10%)
mRSS	–	0	0
ILD	–	0	0
Disease Duration (yr.)	–	–	Unknown

2

Table 2. Demographics and clinical characteristics of the donors included in the non-fibrotic SSc cohort. Values reported indicate the number (n) of patients and the median for each parameter [Interquartile Range (IQR)], if not otherwise indicated. ACA, anticentromere antibodies; ANA, antinuclear antibodies; eaSSc, early SSc; HC, healthy controls; ILD, Interstitial Lung disease; mRSS, modified Rodman Skin score; ncSSc, non-cutaneous SSc; pos, positivity; Scl70, anti-topoisomerase antibodies; Yr., years

Patient group (n)	HC (21)	eaSSc (15)	ncSSc (27)	lcSSc (23)	dcSSc (19)
Age (yr.)	52 (35–82)	62 (25–81)	59 (29–80)	60 (41–80)	52 (27–80)
Female (n, %)	19 (90%)	15 (100%)	27 (100%)	22 (96%)	15 (79%)
ANA (n pos, %)	–	15 (100%)	26 (96%)	22 (96%)	16 (84%)
ACA (n pos, %)	–	12 (80%)	20 (74%)	12 (52%)	0 (0%)
Scl70 (n pos, %)	–	2 (13%)	1 (4%)	9 (39%)	11 (58%)
mRSS	–	0	0	4 (0–8)	12 (2–29)
ILD	–	0	2 (7%)	7 (30%)	14 (74%)
Disease Duration (yr.)	–	N.A.	10* (0–29)	16** (1–38)	10 (1–25)

Table 3. Demographics and clinical characteristics of the donors included in the SSc cohort 3. Values reported indicate the number (n) of patients and the median for each parameter [Interquartile Range (IQR)], if not otherwise indicated. ACA, anticentromere antibodies; ANA, antinuclear antibodies; dcSSc, diffuse cutaneous SSc; eaSSc, early SSc; HC, healthy controls; ILD, Interstitial Lung disease; lcSSc, limited cutaneous SSc; mRSS, modified Rodman Skin score; N.A., not assessed; ncSSc, non-cutaneous SSc; pos, positivity; Scl70, anti-topoisomerase antibodies; Yr., years. *2 patients unknown. **3 patients unknown.

Cell Purification and Culture

Human CD14+ monocytes and neutrophils (PMNs) were purified from heparinised whole blood of SSc patients and matched HC or from buffy coats of healthy donors after centrifugation over Ficoll-Paque gradient. Briefly, CD14+ monocytes were purified from PBMCs using the anti-CD14 microbeads (Miltenyi Biotec), on the autoMACs Pro Separator (Miltenyi Biotec) according to manufacturer's protocol. Purity of monocyte preparations was usually >98%. PMNs were recovered after dextran sedimentation and hypotonic lysis of erythrocytes followed by EasySep neutrophil enrichment kit (StemCell Technologies, Vancouver, Canada)[33]. Purity of neutrophils preparations was usually 99.7 ± 0.2%.

Monocytes (3×10^6 cells/ml) and PMNs (5×10^6 cells/ml) were cultured in RPMI 1640 (Gibco) supplemented with 10% FCS (<0.5 EU/ml; Sigma-Aldrich) and 2 mM Glu in the presence or absence of 100 ng/ml ultra-pure lipopolysaccharide (LPS, from E. coli strain O111:B4, InvivoGen, San Diego, CA, USA), 5 μM R848 (Invivogen), 1,000 U/ml IFNα CRI003B, Cell Sciences), 100 ng/ml palmitoyl-3-cysteine-serine-lysine-4 (Pam3CSK4, Invivogen), 50 μg/ml polynosinic:polycytidylic acids [poly(I:C), Invivogen], as indicated. In selected experiments, CD14+ monocytes were incubated for 30 min with 5 μg/ml Brefeldin A (BFA, Sigma-Aldrich) or 5 μg/ml αIFNAR (PBL InterferonSource, Piscataway, NJ, USA) or its isotype control antibody (mouse IgG2a), before cell stimulation.

Human Monocyte Transfection

Freshly purified monocytes (8×10^6) were transfected with 200 pmol NRIR-specific Silencer Select siRNA or Silencer Select negative control #2 (both from Ambion, Thermo Scientific), using the Human Monocyte Nucleofector Kit and the AMAXA Nucleofector II device (both from Lonza), according to the manufacturer's protocol. Once transfected, cells were plated in recovery medium [50% RPMI 1640 + 10% FCS + 2 mM Glu, and 50% IMDM (Lonza) + 10% FCS + 2 mM Glu], at 3×10^6 cells/ml overnight. The next day, medium was changed to RPMI 1640 + 10% FCS + 2 mM Glu, and cells were stimulated as indicated. NRIR specific Silencer Select siRNA sequence[34] is reported in Table S1.

Extraction of Total RNA

Total RNA was purified with the RNeasy Mini Kit (Qiagen), according to the manufacturer's instructions. DNAse treatment (RNAse Free DNase I set, Qiagen) on column was performed. RNA quantification, purity and integrity were assessed at the Nanodrop 2000 spectrophotometer (Thermo Scientific) and by capillary electrophoresis on an Agilent Bioanalyzer (Agilent Technologies), respectively. Purified RNA was used for sequencing analysis or RT-qPCR, as described below.

RNA Sequencing Analysis

RNA sequencing data of peripheral blood monocytes purified from SSc, together with sex- and age-matched healthy controls (HC) enrolled in the "definite SSc" cohort, were obtained from the University Medical Center Utrecht (UMCU), The Netherlands[35].

RNA sequencing libraries were generated from total RNA extracted from CD14+ monocytes of SSc patients and matched HC enrolled in the "definite SSc" and "non-fibrotic SSc" cohorts, or from RNA pools of three different donors of freshly isolated and LPS-treated monocytes. RNA-seq library preparation was accomplished using the TruSeq RNA Sample Prep Kit v2 (Illumina Inc., San Diego, CA, USA). Libraries were sequenced on a HiSeq 2000 system (Illumina) using pair-end sequencing reads (2 × 90 bp for SSc and matched HC libraries and 2 × 51 bp for resting and LPS-treated monocytes libraries); a minimum of 20 million fragments per sample were analyzed. After quality filtering according to the Illumina pipeline, reads were firstly aligned to the human transcriptome annotated in Ensembl 77 (Homo sapiens gene model annotation) and secondly converted to genomic mapping using as reference the human reference

genome GrCh38 (Genome Reference Consortium Human build 38) by means of TopHat (v 2.0.14)[36]. On average, 23,969,150 (concordant pair alignment rate: 91.84%), 24,404,133 (concordant pair alignment rate: 89.90%), and 43,071,006 (concordant pair alignment rate: 92.67%) paired-reads of the "definite SSc," "non-fibrotic SSc" and LPS-treated-monocytes dataset, respectively, mapped to the reference genome.

Differential expression analysis was performed using the generalized linear model (GLM) implemented in DESeq2 (v 1.6.3) on the summed exon reads count per gene estimated using HTSeq-count (v 0.6.1p1)[37, 38]. Differentially expressed genes were identified from the comparison of each single SSc group and matched HC. Significance was tested using the Wald test. Genes with a log2(FC) value ≥0.58 or ≤-0.58 and a p ≤ 0.05, were considered significantly modulated. Differentially expressed genes in LPS-treated monocytes were identified using the Likelihood Ratio Test (LRT). Raw p-values from differential expression analyses were adjusted to control the false discovery rate (FDR) using the Benjamini–Hochberg method. Genes with adjusted p ≤ 0.05 were considered significantly modulated by LPS. Gene expression levels were expressed as variance stabilized data (vsd) or FPKM, calculated according to DESeq2 instructions. Gene type were associated according to the Ensembl 77 annotation. All genes not belonging to the gene type protein coding and pseudogene and with a transcript length of at least 200 bp were considered as lncRNAs. Raw and processed sequencing data are available from Gene Expression Omnibus under the following accession numbers: GSE123532 and GSE124075.
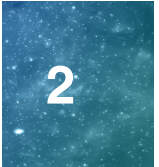
Gene Expression Data of PBMC From SLE Patients and Relative Healthy Controls

Gene expression profiles of PBMC purified from systemic lupus erythematosus (SLE) and relative healthy donors (HC) were downloaded from Gene Expression Omnibus Database (GEO number: GSE122459). Gene expression levels and differential expression analysis were retrieved from the dataset present in the GEO database.

GO-Term and Pathway Enrichment Analysis

Protein coding genes (PCGs) were subjected to Gene Ontology (GO) and pathway enrichment analysis using ToppFun (https://toppgene.cchmc.org/enrichment.jsp[39]). p-value was calculated according to the probability density function and corrected for the False Discovery Rate (FDR) according to Benjamini-Hochberg method. Pathways and GO- terms associated to biological processes (BP) with a FDR ≤ 0.05 were considered significantly enriched.

Weighted Gene Co-expression Network Analysis

Co-expression networks were generated using WGCNA R-package[40]. Signed weighted adjacency matrix of connection strengths was constructed using the soft-threshold approach with a scale-independent topological power β=18 for LPS-treated and freshly isolated monocytes and β=13 for the definite SSc data. Genes were aggregated into modules by hierarchical clustering and refined by the dynamic tree cut algorithm. Biological function of each module was evaluated by pathway and BP GO-terms enrichment analysis using ToppFun[39]. All terms enriched with a FDR ≤ 0.05,

were considered. Redundancy of significantly enriched BP GO-terms was solved by means of REVIGO[41] using the simRel score to assess similarity between two GO-terms[42]. NRIR-specific modules were visualized using Cytoscape v3.2.1[43].

Gene Expression Analysis by Real-Time PCR

RNA samples were reverse transcribed using 5 ng/μl random primers, 1 U/μl RNase inhibitor (RNAse Out, Invitrogen) and 5 U/μl reverse transcriptase (SuperScript III, Invitrogen), according to manufacturer's instruction. NRIR expression was quantified in duplicates by RT-qPCR from 9 ng RNA-equivalent cDNA in the presence of SYBR Select Master Mix (ThermoFisher Scientific, Applied Biosystems) and 400 nM specific primers (Table S1), on the ViiATM 7 Real-Time PCR System (ThermoFisher Scientific, Applied Biosystems) using the standard protocol. PCG expression was quantified in duplicates by RT-qPCR from 9 ng RNA-equivalent cDNA in the presence of Fast SYBR Green Master mix (ThermoFisher Scientific, Applied Biosystems) and 200 nM of specific primer pairs (Table S1), on the ViiATM 7 Real-Time PCR System (ThermoFisher Scientific, Applied Biosystems). Primers were designed using the Oligo Explorer software (http://www.genelink.com/tools/gl-downloads.asp), for only fifty-six out seventy-nine NRIR putative target genes was possible to design specific primer pairs. Data were analyzed with LinReg PCR 7.0 (http:/LinRegPCR.nl) and Q-Gene software (http://www.gene-quantification.de/download.html). Gene expression was calculated as mean normalized expression (MNE [44]) units after normalization over the stably expressed RPL32 or ACTIN B.

Multiplex Immunoassay

CXCL10, CXCL11, and CCL8 concentrations in cell-free supernatants and/or plasma from SSc patients and matched HC enrolled in the "SSc cohort 3" were measured using an in-house developed and validated (ISO9001 certified) multiplex immunoassay (Laboratory of Translational Immunology, University Medical Center Utrecht) based on Luminex technology (xMAP, Luminex Austin TX USA). The assay was performed as previously described [45]. Aspecific heterophilic immunoglobulins were pre-absorbed from all plasma samples with heteroblock (Omega Biologicals, Bozeman MT, USA). All samples were measured with the Biorad FlexMAP3D (Biorad laboratories, Hercules USA) in combination with the xPONENT software (v 4.2, Luminex). Data were analyzed by a 5-parametric curve fitting using the Bio-Plex Manager software (v 6.1.1, Biorad).

Statistical Analysis

Data are expressed as mean ± SEM unless otherwise indicated. Statistical evaluation was determined using the Mann Whitney test or the two-way analysis of variance (ANOVA), followed by Bonferroni post-test, with α set to 0.05. Correlation analysis were performed using the rcorr() function in R using the non-parametric Spearman method. Correlation with $p < 0.05$ were considered significant.

RESULTS

Identification of LPS-Modulated lncRNAs in Primary Human Monocytes

To identify lncRNAs potentially involved in the responses of peripheral human monocytes downstream TLR4 activation, CD14$^+$ monocytes purified from buffy coats of healthy donors were cultured in the presence or absence of LPS (100 ng/ml) for 1.5 h or 4 h, and subsequently subjected to RNA sequencing. 1,812 transcripts annotated as lncRNAs in Ensemble (Figure 1A) were identified as significantly (p-adj ≤ 0.05) modulated in response to LPS. Specifically, 1278 lncRNAs (i.e., 70.53%) were up-regulated, while 534 lncRNAs (i.e., 29.47%) were down-regulated (Figure 1B). Moreover, K-means clustering arranged the LPS-modulated lncRNAs in three main groups according to their kinetic of expression (Figure 1C): (i) lncRNAs rapidly and consistently modulated by LPS within 1.5 h, representing the majority (52.32%) of LPS-modulated lncRNAs (early group, Figure 1D); (ii) lncRNAs modulated by LPS within 1.5 h in a transient manner (22.57%) (early and transient group, Figure 1E); (iii) lncRNAs modulated by LPS at 4 h (25.11%) (late group, Figure 1F).

### Identification of Type I IFN Signature-Associated lncRNAs

LncRNAs possibly involved in the regulation of type I IFN pathway activated downstream TLR4 were identified using the strategy depicted in Figure 2. Specifically, 3,248 PCGs up- regulated in response to LPS (FPKM > 2) were retrieved and subjected to GO term enrichment analysis. 469 LPS-induced PCGs associated to significantly enriched IFN-response and anti-viral response-related GO-terms were then subjected to correlation analysis with the 1,812 LPS-modulated lncRNAs. Finally, based on the knowledge that lncRNAs can regulate the transcription of PCGs located in cis[46], only the lncRNAs localized in cis (± 150 Kb) to correlated PCGs were retrieved (n = 99) (Figure 2 and Table S2). This group of lncRNAs (n = 99) will be referred from now on as the "IFN/viral" lncRNAs.

To verify whether the selected "IFN/viral" lncRNAs were effectively related to the IFN signature in an *in vivo* setting where the IFN pathway is known to play a pathogenetic role, the expression level of the 99 selected lncRNAs was then retrieved and analyzed from the transcriptomic profile of monocytes purified from the "definite SSc" [35] and "non-fibrotic SSc" cohorts of patients and matched healthy donors (Tables 1, 2). The patient cohorts included individuals presenting with different SSc phenotypes according to clinical features and the extent of skin fibrosis, i.e., early SSc (eaSSc, n = 11), non-cutaneous SSc (ncSSc, n = 17), limited cutaneous SSc (lcSSc, n = 11), diffuse cutaneous SSc (dcSSc, n = 7).
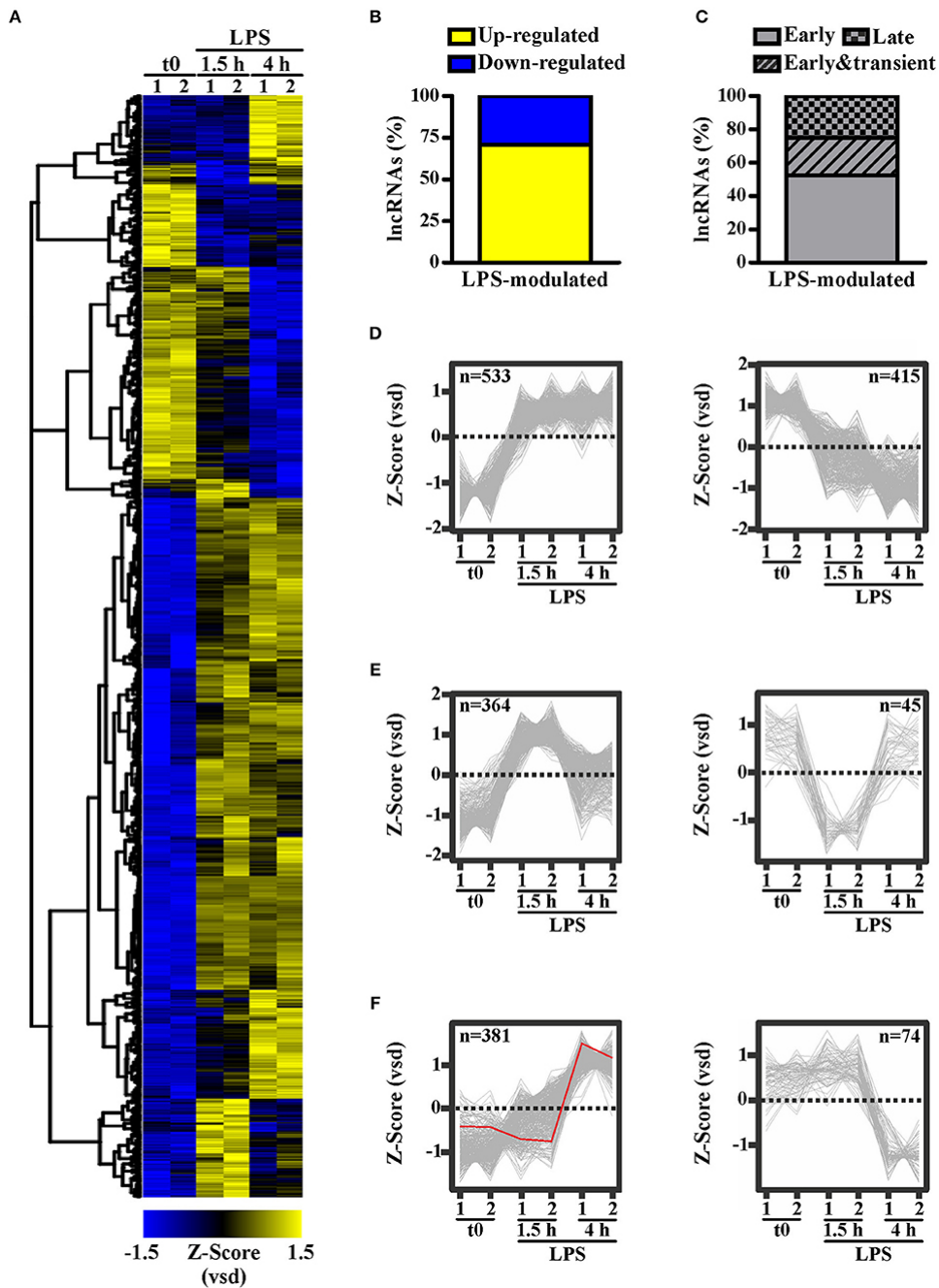
Figure 1. LPS modulates the expression of long noncoding transcripts in human monocytes. CD14+ monocytes were cultured for 1.5 or 4 h with LPS (100 ng/ml) or left untreated (t0). Two pools of three donors for each condition were used to create polyA library for RNA-seq. Sequencing data were analyzed as described in Materials and Methods. The expression levels of the LPS-modulated (adjusted p < 0.05) lncRNAs (A) are shown as row mean-centered z-Score of the variance stabilized data (vsd). (B) The percentage of up- and down-regulated lncRNAs modulated by LPS. (C) The percentage *(continued)*

of early, early & transient and late lncRNAs modulated by LPS. K-means clustering analysis was applied on the significantly modulated lncRNAs. Early modulated (D), early and transiently modulated (E) as well as late modulated (F) lncRNAs are shown. The expression of each lncRNA belonging to the three groups is shown. LncRNAs up regulated and down regulated by LPS are shown separately. NRIR expression is highlighted in red. LncRNA expression is depicted as row mean-centered z-Score of the variance stabilized data (vsd), number of lncRNA belonging to each KMC group is shown.

2
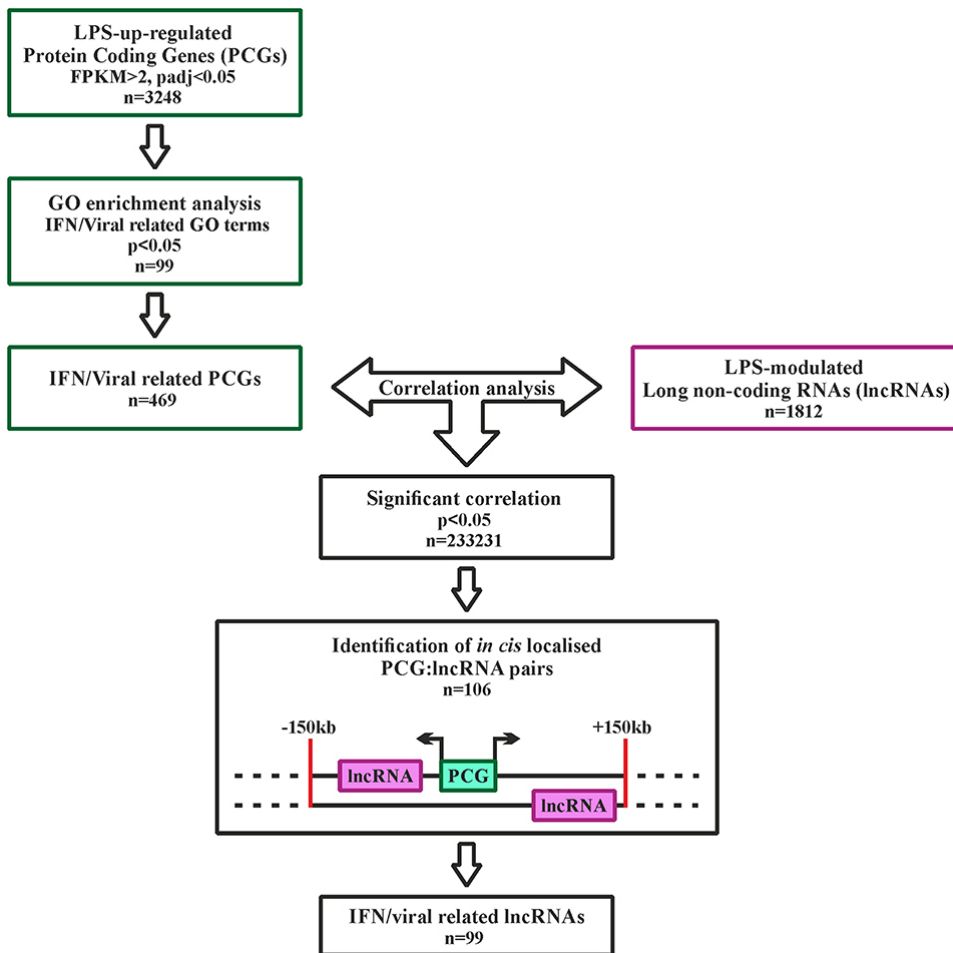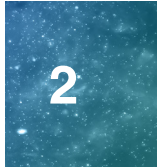
Figure 2. Analysis pipeline to identify IFN/viral-related lncRNAs, modulated by LPS in monocytes. Green squares represent the selection of IFN/viral related protein coding genes, while the purple square represents the selected lncRNAs modulated by LPS. Black squares represent the workflow for integration of protein coding genes and lncRNAs by correlation analysis.

Four out of 99 lncRNAs, namely NRIR, PSMB8-AS1, RP5-1091N2.9, and RP11-24F11.2, were expressed at significantly higher levels in at least two groups of SSc patients as compared to their respective healthy donors in the "definite SSc" cohort (Figure 3A), whereas only NRIR was significantly up-regulated in ncSSc and showed a trend in eaSSc (FC = 1.30, p = 0.104) in the "non-fibrotic" cohort (Figure 3B). Remarkably, only the expression of NRIR significantly correlated in both cohorts with the patients' IFN score (Figures 3C, D), calculated on the basis of the expression of IFI27, IFI44L, IFIT1, IFIT2, IFIT3, and SERPING1[16].
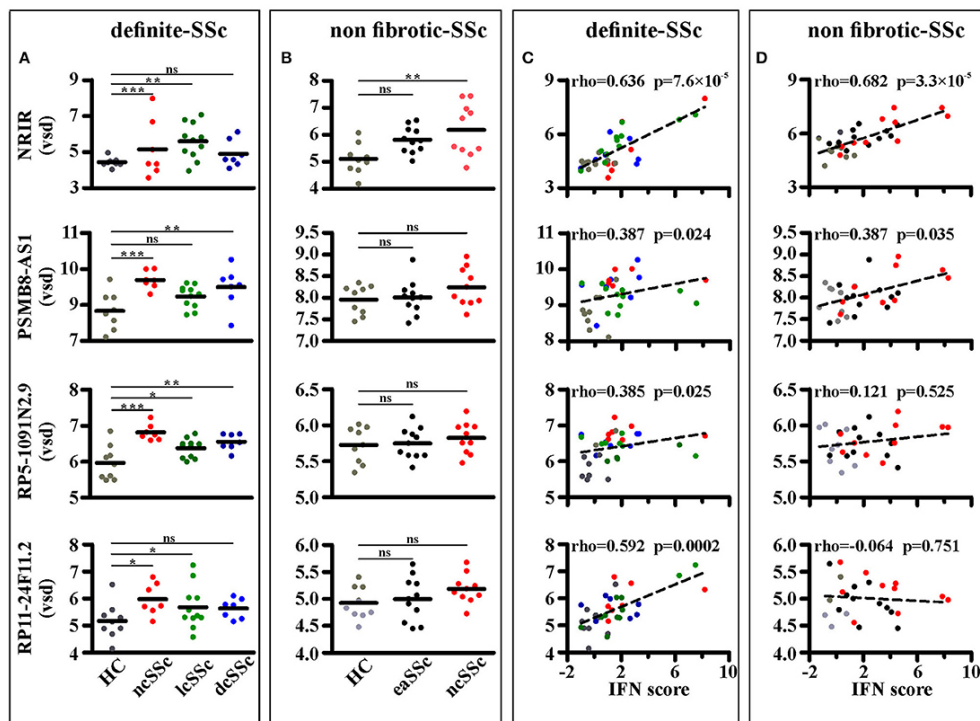
Figure 3. NRIR expression is increased in monocytes from SSc patients and correlates with the IFN-score. RNA sequencing data of CD14+ monocytes from SSc patients and matched healthy controls (HC) from both the definite SSc and non-fibrotic SSc cohorts were analyzed as described in Materials and Methods. NRIR, PSMB8-AS1, RP5-1091N2.9, and RP11-24F11.2 expression were considered. LncRNAs expression in HC and patients with established Systemic Sclerosis (ncSSc, lcSSc, and dcSSc, definite-SSc cohort) (A) and in patients with early stages of SSc (eaSSc and ncSSc, non-fibrotic SSc cohort) (B) is shown. *p < 0.05, **p < 0.01, ***p < 0.001, ns, not significant, by Wald test. (C) Correlation of NRIR expression with the IFN-score of HC (gray), ncSSc (red), lcSSc (green), and dcSSc (blue) patients is depicted. (D) Correlation of NRIR expression with the IFN-score of HC (gray), eaSSc (black) and ncSSc (red) patients is shown. Spearman's Rho and p-value are reported. NRIR expression levels are expressed as vsd, IFN Score was calculated according to Brkic *et al.*[16].

IFNα was demonstrated to be central to the pathogenesis also of other systemic autoimmune diseases, with Systemic lupus erythematosus (SLE) being the prototype one. To verify whether NRIR is effectively related to the IFN signature in an in vivo setting in IFN-related diseases other than SSc, we retrieved from the Gene Expression Omnibus

database RNA-seq data from PBMCs of SLE patients and matched healthy controls (GSE122459). Seventeen out of ninety-nine lncRNAs were commonly modulated in LPS-treated CD14$^+$ transcriptome and SLE PBMCs compared to healthy controls (Figure S1A), and only three lncRNAs, namely NRIR, PSMB8-AS1 and RP5-1091N2.9, were modulated in all the three datasets (i.e., LPS-treated CD14$^+$ monocytes, SSc CD14$^+$ monocytes and PBMCs from SLE patients) (Figure S1B). Remarkably, NRIR was the only one lncRNA significantly up-regulated in all the three datasets and the lncRNA most differentially expressed (log2FC = 1.90, p = 3.83 × 10–8) in PBMCs from SLE patients as compared to healthy controls (Figure S1B).

2

Collectively, data from three different biological datasets (i.e., transcriptome of monocytes activated in vitro by LPS, transcriptome of circulating monocytes from SSc patients and transcriptome of PBMCs from SLE patients) converged in identifying NRIR as belonging to the IFN signature. Therefore, we focused our study on the pathways underlying NRIR upregulation as well as on the role of this lncRNA in the type I IFN signature.

NRIR Is a Type I IFN Dependent lncRNA

Consistent with KMC analysis of RNA-seq data that classified NRIR as a "late" transcript (Figure 1F, red line), kinetic analysis confirmed that NRIR expression is slowly induced by LPS stimulation in monocytes, being detectable after 4 h and steadily increasing over 16 h (Figure 4A). In addition, monocyte activation with agonists of TLR3 [polyinosinic:polycytidylic acid, poly(I:C)] and TLR7/8 (Resiquimod, R848), both known to promote type I IFN production, resulted in up-regulation of NRIR (Figure 4B). Conversely, a synthetic lipoprotein agonist of TLR2 (Pam3CSK4), unable to induce type I IFN transcription and secretion[47], was ineffective (Figure 4B). Consistent with this observation, treatment of monocytes with brefeldin A or with IFNα receptor (αIFNAR) blocking antibodies before LPS stimulation completely abolished NRIR induction by LPS (Figure 4C), indicating that endogenously produced type I IFNs is responsible for the upregulation of NRIR. Additionally, NRIR expression is significantly induced by IFNα but not by LPS, in human polymorphonuclear neutrophils (PMNs), that do not activate the IFN pathway downstream TLR4 (Figure 4D)[48]. Taken together, these data demonstrate that type I IFN production is necessary and sufficient to increase NRIR expression in response to LPS.
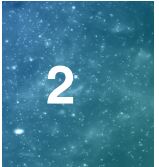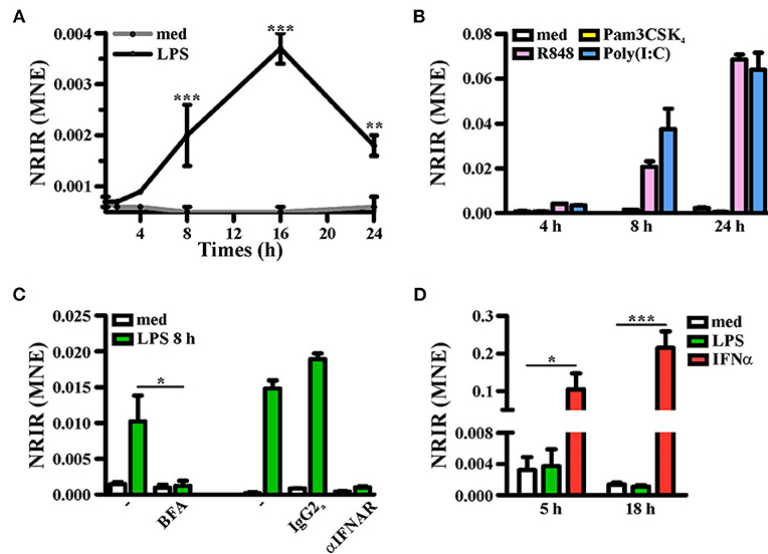
Figure 4. Induction of NRIR expression is IFN-dependent. (A) CD14+ monocytes were cultured for the indicated time point in presence of LPS (100 ng/ml, black line) or left untreated (gray line). NRIR expression levels were analyzed by RT-qPCR and expressed as mean normalized expression (MNE). Results are shown as mean ± SEM of three experiments. **$p < 0.01$, ***$p < 0.001$ by two-way ANOVA. (B) CD14+ monocytes were stimulated with Pam3CSK4 (100 ng/ml), poly(I:C) (50 μg/ml), R848 (5 μM) or left untreated for the indicated time points. NRIR expression levels were analyzed by RT-qPCR and expressed as MNE. One experiment representative of two performed is shown. (C) CD14+ monocytes were stimulated with LPS or left untreated for 8 h in presence or absence of brefeldin A (BFA, left) or αIFNAR or the control IgG2a antibody (right). NRIR expression levels were analyzed by RT-qPCR and expressed as MNE. For BFA experiments results are shown as mean ± SEM of three experiments, *$p < 0.05$ by two-way ANOVA, while for αIFNAR experiments one experiment representative of two performed is shown. (D) Human neutrophils were stimulated with LPS (100 ng/ml), IFNα (1,000 U/ml) or left untreated for 5 and 18 h. NRIR expression levels were analyzed by RT-qPCR and expressed as MNE. Results are shown as mean ± SEM of three experiments. *$p < 0.05$, ***$p < 0.001$ by two-way ANOVA.

The Type I IFN-Dependent NRIR Plays a Role in the Expression of Several ISGs

Identification of pathways likely associated to NRIR function was conducted by weighted gene co-expression analysis (WGCNA). Two specific co-expression networks were created, one composed of 13 modules in the transcriptome of LPS-treated monocytes and the second one composed of 26 modules in the "definite SSc" cohort. The NRIR-related module was identified in both LPS-treated monocytes (blue module) and SSc monocytes (cyan module) co-expression networks. The blue module contained 2060 PCGs and 548 ncRNAs (Figure S2), while the cyan module was composed of 116 PCGs and 8 ncRNAs (Figure S3).

GO-term and pathway enrichment analysis of the PCGs of each module underlined that biological processes related to "response to type I IFN," "response to virus," and "immune system process" (Figure 5) and related pathways (Tables S3, S4) were significantly enriched in both modules. Comparative analysis of the two modules identified 83 common transcripts: specifically, 79 PCGs and 4 ncRNAs (Figure 6A), the majority (63.3%) of which were associated to IFN, antiviral and immune response (Figure 6B). The 79 common PCGs were selected as putative NRIR target genes.

2
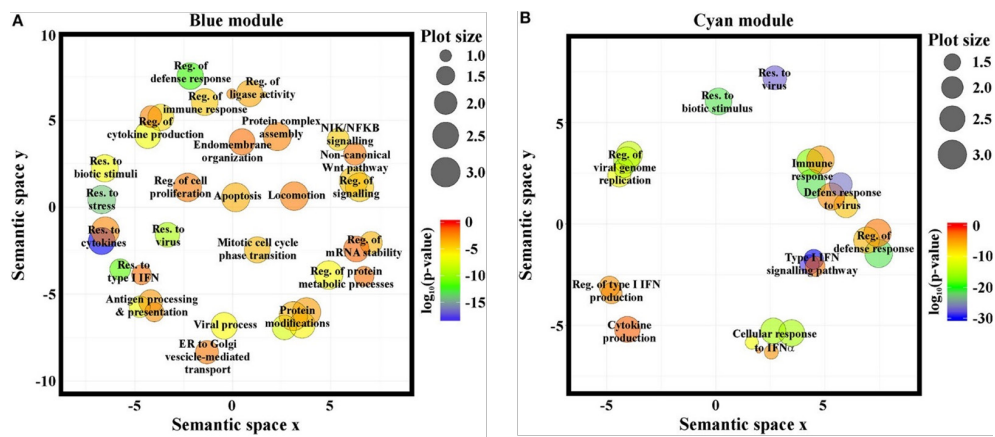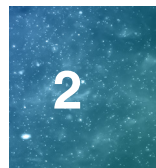
Figure 5. NRIR is implicated in biological processes related to immune response and the IFN/ antiviral response. GO-term enrichment analysis was performed to identify biological processes enriched in the blue- (A) or the cyan-module (B). Significantly enriched GO terms are represented as circles according to their semantic similarities. Circle size represents term specificity (bigger, general terms; smaller, specific terms), while circle color represents the log10 (p-value FDR B&H) of the enrichment.

To investigate the role of NRIR in the regulation of IFN and anti-viral response secondary to TLR4 activation, we analyzed the expression of 56 PCGs, that were co-expressed with NRIR and common to the both blue and cyan modules (Figure 6), in NRIR-silenced monocytes. Monocyte transfection with NRIR siRNA led to an average reduction of 60.83 ± 4.81 and 55.47 ± 4.83% of the constitutive and LPS induced NRIR expression, respectively (Figure 7A). Under these conditions, the induction of fifteen PCGs by LPS was significantly impaired as compared to cells transfected with a scramble siRNA (Figures 7B–P). Precisely, decreased induction of CXCL10, CXCL11, APOBEC3A, MX1, USP18 mRNA was observed 4 h after LPS stimulation and remained reduced at 8 h as well; decreased induction of CCL8, EPSTI1, DDX58, IFI44, IFIH1, IFIT2, and OAS2 was observed at shorter time point (4 h); whereas the ability of LPS to upregulate the expression of IFITM3, ISG15 and OAS3 could be detected only at later time point (8 h) (Figures 7B–P). The induction of the remaining forty-one PCGs was unaffected by NRIR knock- down (Figures S4, S5), Strikingly, all genes modulated by NRIR silencing were also significantly upregulated in at least one group of SSc monocytes as compared to cells isolated from healthy donors (Figure S3).
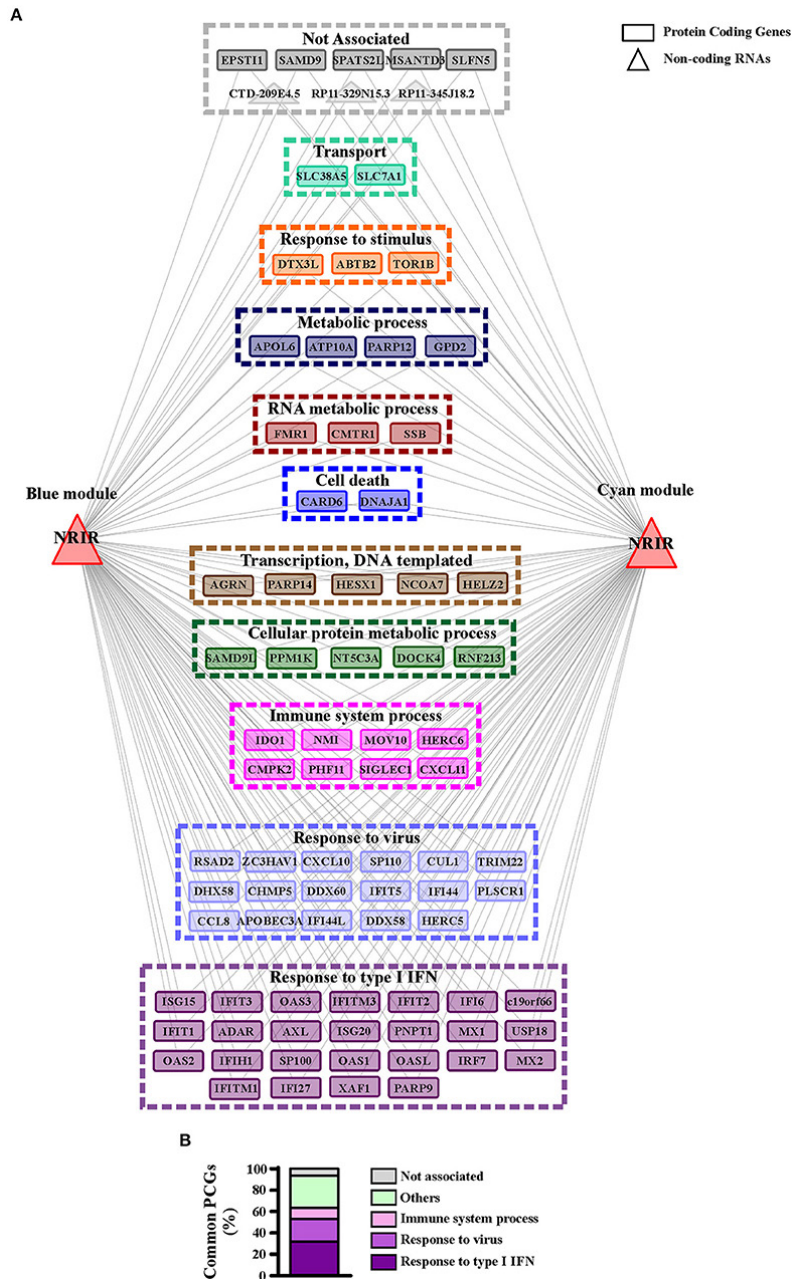
Figure 6. PCGs common to the blue and cyan modules are mainly involved in the immune and IFN/antiviral response. (A) Representation of the transcripts common to blue and cyan-module. The seventy-nine protein coding genes and the four ncRNAs are represented as rectangles and triangles, respectively. Transcripts are grouped according to their associated biological process related GO-terms. Different colors highlight different group of GO-terms, the most general GO term, summarizing each group, is reported. Genes associated to any GO-term are signed as not associated and depicted in gray. (B) Protein coding genes found in both modules are associated to their GO terms. Percent of common protein coding genes associated to different GO terms is shown.

2
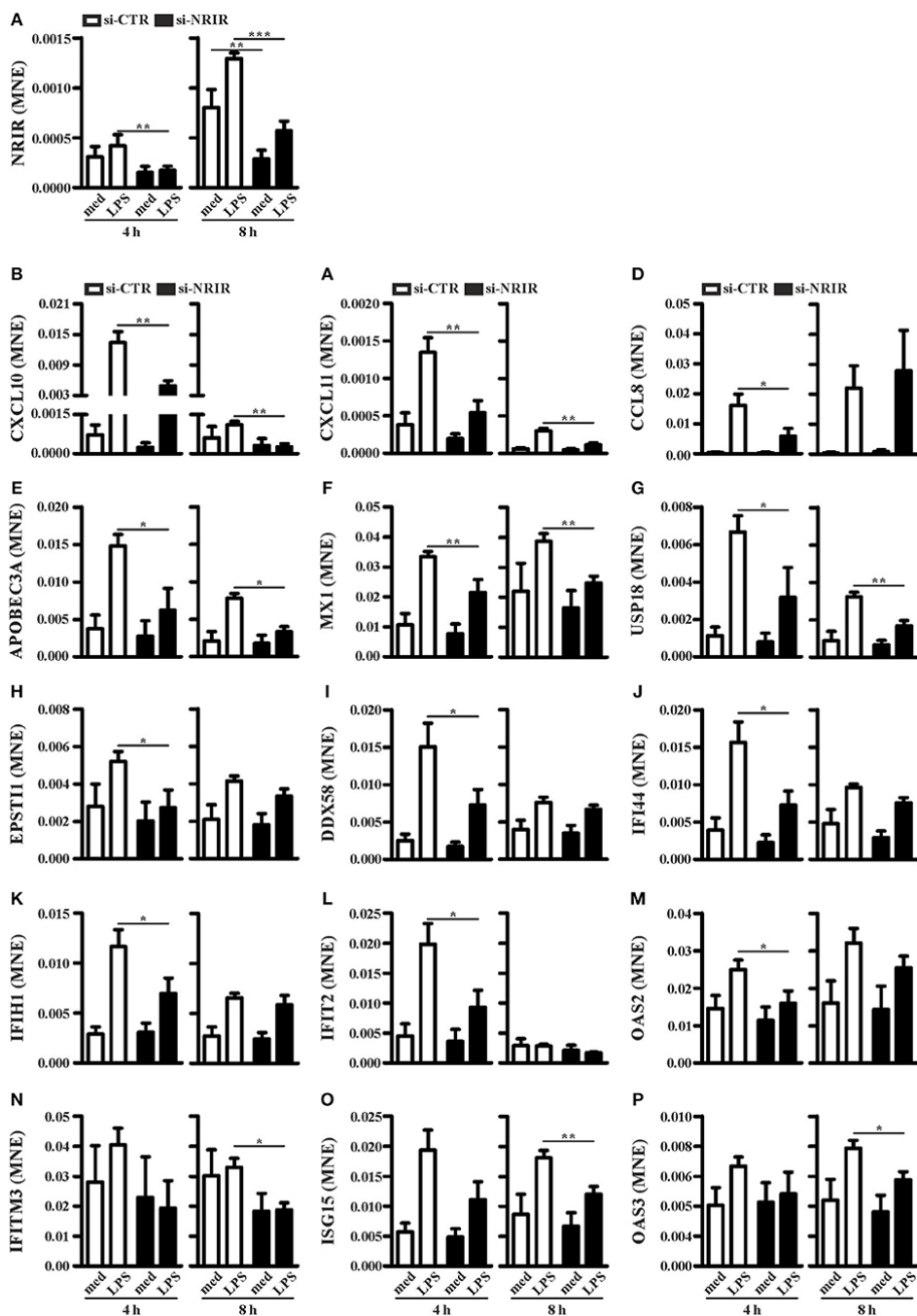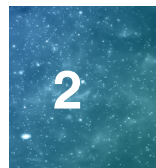
Figure 7. NRIR regulates fifteen of its co-expressed genes. CD14+ monocytes were transfected with si-NRIR or si-CTR and 18 h later were stimulated with LPS for 4 or 8 h or left untreated. The expression of NRIR (A) and its co-expressed genes (B–P) was analyzed by RT-qPCR and expressed as MNE. Results are shown as mean ± SEM of at least three different experiments. *p < 0.05, **p < 0.01, ***p < 0.001 by two-way ANOVA.

Among the IFN-responsive genes, CXCL10, CXCL11 and CCL8 have been shown to be implicated in SSc pathogenesis and/or to correlate with the degree of skin fibrosis[18, 49–51]. Analysis of CXCL10, CXCL11 and CCL8 protein level in cell-free supernatants of LPS-stimulated monocytes showed a significant reduction of CXCL10 (mean reduction: 62.48 ± 8.94%, n = 7) and CCL8 (mean reduction: 56.13 ± 7.37%, n = 7) production in response to LPS (Figures 8A,B), while CXCL11 was below the detection levels (not shown). Noticeably, plasma level of CXCL10 and CXCL11 in the SSc subjects enrolled in this study was significantly higher as compared to their healthy counterparts (Figures 8C,D).

Collectively, these data substantiate the role of NRIR in the expression of several interferon-responsive genes upregulated by LPS in vitro or constitutively increased in circulating monocytes from SSc patients.
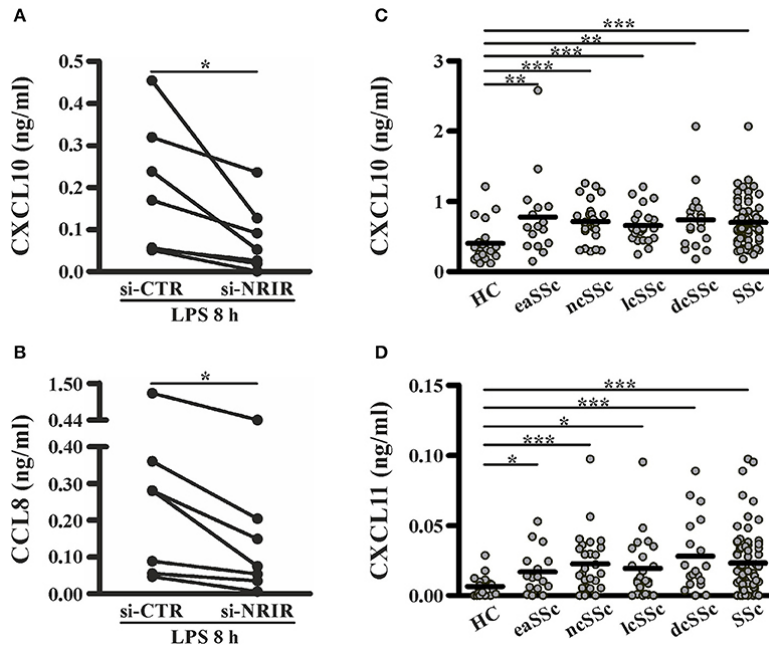
Figure 8. NRIR regulated proteins CXCL10 and CXCL11 are elevated in plasma of SSc patients. CD14+ monocytes from seven different donors were transfected with si-NRIR or si-CTR and 18 h later were stimulated with LPS for 8 h. Cell-free supernatants were collected, and the release of CXCL10 (A) and CCL8 (B) was measured by the multiplex immunoassay. *p < 0.05 by Wilcoxon matched-pairs signed rank test. CXCL10 (C) and CXCL11 (D) level in plasma from SSc patients and matched HC was measured by the multiplex immunoassay. eaSSc, early SSc; ncSSc, non-cutaneous SSc; lcSSc, limited-cutaneous SSc; dcSSc, diffuse-cutaneous SSc. *p < 0.05, **p < 0.01, ***p < 0.001 by Mann Whitney test.

DISCUSSION

The aim of this study was to investigate the potential role of lncRNAs in the type I IFN pathway elicited in human monocytes by TLR4 activation and to explore their functional role in vivo, in the IFN signature displayed by SSc monocytes. Several studies have shown that lncRNAs are involved in numerous aspects of the innate and adaptive immune responses[22], and, more recently, a critical role for a small group of lncRNAs in the regulation of the IFN response has been reported[19]. Likewise, evidence clearly supports the involvement of lncRNAs in the pathogenesis of autoimmune and inflammatory diseases[25, 31], where the physiologic response of immune cells is dysregulated. However, no lncRNA has been associated to the immune dysregulation present in SSc yet. Characterization of the role of lncRNAs in the regulation of monocytes IFN response to TLR4 activating agents is an important aspect to understand both the physiologic response and the disease biology of SSc arising from alteration of physiologic pathways. In fact, the link between monocytes, TLR4 activation and the downstream IFN response with SSc pathogenesis is supported by several observations: (i) circulating monocytes have been indicated as one prominent leukocyte subset playing a role in the pathogenesis of SSc[52–55]; (ii) circulating SSc monocytes are characterized by an increased type I IFN signature[11, 12, 16] (iii) TLR activation may represent the connection between immune activation in SSc and tissue fibrosis[7, 10, 52, 56].

The lncRNA landscape of LPS-activated human monocytes, characterized by RNA sequencing, identified 1,278 annotated lncRNAs as upregulated and 534 as downregulated. Modulated lncRNAs were further clustered according to their kinetic of expression into early, early and transient and late. Correlation with the expression of PCGs enriched in the IFN- and anti-viral response related GO-terms allowed us to retrieve lncRNAs likely comprised into the type I IFN pathway. Moreover, as some lncRNAs have been described to regulate the expression of neighboring genes[46], lncRNAs that may have functional relevance in the expression of LPS-induced mRNAs related to the IFN/anti-viral response were retrieved on the basis of their localization in cis to their respective correlated PCGs.

To validate the relevance of these "IFN/viral" lncRNAs in an in vivo setting where the IFN response constitute a major hallmark, we examined the expression level of each of the 99 lncRNAs in monocytes from two distinct cohorts of SSc patients as compared to the relative healthy control groups. The cohorts comprised patients with the full spectrum of SSc phenotypes, from pre-clinical eaSSc, to definite groups either presenting with (lcSSc and dcSSc) or without (ncSSc) skin fibrosis. Most importantly in both cohorts a remarkable IFN signature had been identified in previous studies[16, 35]. Remarkably, monocytes from lcSSc and ncSSc patients showed consistently higher levels of NRIR expression, that correlated significantly with the IFN signature in both cohorts analyzed, strikingly confirming the implication of NRIR in the IFN response also in a pathological condition. Consistently, it must be noted that NRIR had the highest expression levels in patients with ncSSc, that is the SSc subset presenting with the strongest IFN-signature[16]. In addition, it is intriguing to observe that NRIR shows a trend of upregulation also in the eaSSc group, characterized by higher levels of ISGs as well. Considering that most patients with eaSSc are prompt to progress toward definite

SSc[57, 58], one could speculate a potential implication of NRIR in the IFN signature intertwined with SSc progression. Remarkably, NRIR was the lncRNA most differentially expressed in PBMC from SLE patients as compared to healthy controls, thus further supporting that dysregulation of the IFN-dependent NRIR lncRNA represents a hallmark of different IFN-driven pathologies.

Identification of NRIR-related pathways was conducted according to the "guilt-by-association" method[59], that remains the only approach allowing to characterize lncRNAs based on the function of their co-expressed PCGs. NRIR was found in two distinct co-expression modules, retrieved from WGCNA analysis of the transcriptome of monocyte activated in vitro by LPS or isolated from SSc patients. The majority (63%) of the PCGs common to both modules was included in "response to type I IFN," "response to virus," and "immune system process" biological processes, thus strengthening the likelihood that NRIR plays a role in these processes. Experimental validation of the *in silico* analysis demonstrated that NRIR is a type I IFN-responsive gene, induced in monocytes upon activation of only those TLRs that can trigger type I IFN production (i.e., TLR4, TLR3 and TLR7/8). This is further supported by the demonstration that inhibition of LPS-induced release of soluble mediators, and specifically blockade of type I IFN receptor abolished the ability of LPS to upregulate NRIR. Moreover, monocyte activation with agonists of TLR2 (unable to induce type I IFN transcription and secretion) or neutrophil activation of TLR4 (that does not mobilize the TRIF-IFN pathway[48]) failed to upregulate NRIR expression.

Consistently with the NRIR role suggested by the WGCNA approach, data shows that NRIR-silencing mainly reduces the LPS-induced expression of type I IFN target genes, including, among the others, CXCL10, MX1, IFITM3, and ISG15. Moreover, measurements of CXCL10 and CCL8 secretion further endorsed the role of NRIR as a positive regulator of a subset of LPS-induced IFN-dependent genes.

The inhibition of ISGs upon NRIR-silencing is in sharp contrast with recent reports showing that NRIR acts as a negative regulator of specific ISGs (CMPK2, CXCL10, IFIT3, IFITM1, ISG15, Viperin, and IFITM3) in hepatocytes[34] or epithelial cells[60]. Overall, our findings strengthen the role of NRIR as a regulator of the IFN response, but they strongly point out that NRIR function is highly cell-type or stimulus specific. Such behavior is not uncommon among lncRNAs implicated in the regulation of immune response; one example is IL7-AS, that was described either as a positive regulator of IL-6 expression in IL-1β-activate epithelial cells[61] or as negative regulator in LPS-stimulated monocytes/macrophages as well as in IL-1β activated chondrocytes[62].

It must be underlined that all the ISGs inhibited by NRIR silencing are also upregulated in SSc monocytes, that display concomitantly a prominent IFN signature as well as NRIR upregulation. These observations strengthen the relevance of the NRIR-ISGs axis in both physiological as well as pathological conditions. Among the ISGs inhibited upon NRIR silencing, numerous genes have been frequently linked to SSc. Increased levels of CXCL10, CXCL11, IFI44, and MX1 correlate with the severity of different clinical features in SSc patients[63, 64]. Higher MX1 expression was associated with ischemic ulcers and reduced forced vital capacity[64, 65]. The extent of skin fibrosis measured by the modified Rodman Skin Score (mRSS) correlates with the expression

of IFI44[63]. Most importantly, increased levels of circulating CXCL10 and CXCL11, both NRIR targets, highly correlate with the type I IFN signature as well as with a more severe clinical phenotype, with lung and kidney involvement[11, 63, 66]. In fact, serum level of CXCL10 and CXCL11 has been recently proposed as biomarker for the identification of early and non-fibrotic subset of SSc[18]. Conversely, inhibition of type I IFN signature in SSc patients with anifrolumab, that blocks IFN receptor signaling, leads to lower levels of CXCL10 expression and fibrosis-related transcripts[67].

      Collectively, herein we demonstrate that the IFN-dependent lncRNA NRIR is a positive regulator of the LPS-induced IFN response in human monocytes and highlight, for the first time, that aberrant expression of NRIR can be involved in the dysregulation of immune system intertwined with SSc development.

2
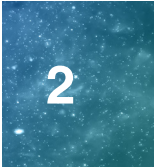
REFERENCES

1. K. Takeda, S. Akira. TLR signaling pathways. *Semin Immunol.* vol. 16, no.1, pp. 3–9, Feb. 2004.
2. L.B. Ivashkiv, L.T. Donlin. Regulation of type i interferon responses. *Nat Rev Immunol.* vol. 14, no. 1, pp. 36–49 Jan. 2014.
3. M. Farrugia, B. Baron. The role of toll-like receptors in autoimmune diseases through failure of the self-recognition mechanism. *Int J Inflam.* vol. 2017, pp. 8391230. May 2017.
4. L. Duffy, S. O'reilly. Toll-like receptors in the pathogenesis of autoimmune diseases: recent and emerging translational developments. *ImmunoTargets Ther.* vol. 5, pp. 69–80. Aug. 2016.
5. S. Bhattacharyya, J. Varga. Emerging roles of innate immune signaling and toll-like receptors in fibrosis and systemic sclerosis. *Curr Rheumatol Rep.* vol. 17, no. 1, pp. 474. Jan. 2015.
6. M. Ciechomska, R. Cant, J. Finnigan, *et al.* Role of toll- like receptors in systemic sclerosis. *Expert Rev Mol Med.* vol.15, pp. e9. Aug. 2013
7. R. Lafyatis, A. Farina. New insights into the mechanisms of innate immune receptor signalling in fibrosis. *Open Rheumatol J.* vol. 6, pp. 72–79. Jun. 2012.
8. S. Bhattacharyya, Z. Tamaki, W. Wang, *et al.* FibronectinEDA promotes chronic cutaneous fibrosis through toll-like receptor signaling. *Sci Transl Med.* vol.6, no. 232, pp. 232ra50. Apr. 2014.
9. L. Van Bon, M. Cossu, A. Loof, *et al.* Proteomic analysis of plasma identifies the Toll-like receptor agonists S100A8/A9 as a novel possible marker for systemic sclerosis phenotype. *Ann Rheum Dis.* vol. 73, no. 8, pp. 1585–1589. Aug. 2014.
10. S. Bhattacharyya, K. Kelley, D.S. Melichian, *et al.* Toll- like receptor 4 signaling augments transforming growth factor-β responses: a novel mechanism for maintaining and amplifying fibrosis in scleroderma. *Am J Pathol.* vol. 182, no. 1, pp.192–205. Jan. 2013.
11. M-L. Eloranta, K. Franck-Larsson, T. Lovgren, *et al.* Type I interferon system activation and association with disease manifestations in systemic sclerosis. *Ann Rheum Dis.* vol.69, no. 7, pp.1396–1402. Jul. 2010.
12. M.R. York, T. Nagai, A.J. Mangini, *et al.* A macrophage marker, siglec-1, is increased on circulating monocytes in patients with systemic sclerosis and induced by type I interferons and toll-like receptor agonists. *Arthritis Rheum.* vol.56, no. 3, pp.1010–1020. Mar. 2007.
13. F.K. Tan, X. Zhou, M.D. Mayes, *et al.* Signatures of differentially regulated interferon gene expression and vasculotrophism in the peripheral blood cells of systemic sclerosis patients. *Rheumatology.* vol.45, no.6, pp. 694–702. Jan. 2006.
14. R. Solans, J.A. Bosch, I. Esteban, *et al.* Systemic sclerosis developing in association with the use of interferon alpha therapy for chronic viral hepatitis. *Clin Exp Rheumatol.* vol. 22, no. 5, pp. 625–628. Oct 2004.
15. C.M. Black, A.J. Silman, A.I. Herrick, *et al.* Interferon-alpha does not improve outcome at one year in patients with diffuse cutaneous scleroderma: results of a randomized, double-blind, placebo-controlled trial. *Arthritis Rheum.* vol. 42, no. 2, pp. 299–305. Feb. 1999.
16. Z. Brkic, L. van Bon, M. Cossu, *et al.* The interferon type I signature is present in systemic sclerosis before overt fibrosis and might contribute to its pathogenesis through high BAFF gene expression and high collagen synthesis. *Ann Rheum Dis.* vol. 75, no. 8, pp. 1567–1573. Aug. 2016.
17. T. Carvalheiro, S. Horta, J.A.G. van Roon, *et al.* Increased frequencies of circulating CXCL10-, CXCL8- and CCL4-producing monocytes and Siglec-3-expressing myeloid dendritic cells in systemic sclerosis patients. *Inflamm Res.* vol.67, no. 2, pp. 169–177. Feb. 2018.
18. M. Cossu, L. van Bon, C. Preti, *et al.* Earliest phase of systemic sclerosis typified by increased levels of inflammatory proteins in the serum. *Arthritis Rheumatol.* vol. 69, no. 12, pp. 2359–2369. Dec. 2017.
19. S. Valadkhan, L.M. Plasek. Long non-coding RNA-mediated regulation of the interferon response: a new perspective on a familiar theme. *Pathog Immun.* vol. 3, no. 1, pp. 126–148. Aug. 2018.
20. I.M. Dykes, C. Emanueli. Transcriptional and post-transcriptional gene regulation by long non-coding RNA. *Genom Proteom Bioinformat.* vol. 15, no. 3, pp. 177–186. Jun. 2017.
21. M-T. Melissari, P. Grote. Roles for long non-coding RNAs in physiology and disease. *Pflugers Arch.* vol. 468, no. 6, pp. 945–958. Mar. 2016.
22. K.A. Fitzgerald, D.R. Caffrey. Long noncoding RNAs in innate and adaptive immunity. *Curr Opin Immunol.* vol. 26, pp. 140–146. Feb. 2014.
23. J.A. Heward, M.A. Lindsay. Long non-coding RNAs in the regulation of the immune response. *Trends Immunol.* vol. 35, no. 9, pp. 408–419. Sep. 2014.

24. K. Imamura, N. Akimitsu. Long non-coding rnas involved in immune responses. *Front Immunol*. vol. 5, pp. 573. Nov. 2014.

25. J.L. Marques-Rocha, M. Samblas, F.I. Milagro, *et al*. Noncoding RNAs, cytokines, and inflammation-related diseases. *FASEB J*. vol. 29, no. 9, pp. 3595–3611. Sep. 2015.

26. A.T. Satpathy, H.Y. Chang. Long noncoding RNA in hematopoiesis and immunity. *Immunity*. vol. 42, no. 5, pp. 792–804. May 2015.

27. K.R. Sigdel, A. Cheng, Y. Wang, *et al*. The emerging functions of long noncoding rna in immune cells: autoimmune diseases. *J Immunol Res*. vol. 2015, pp. 848790. May 2015.

28. A. Stachurska, M.M. Zorro, M.R. van der Sijde, *et al*. Small and long regulatory RNAs in the immune system and immune diseases. *Front Immunol*. vol. 5, pp. 513. Oct. 2014.

29. A.D. Yu, Z. Wang, K.V. Morris. Long noncoding RNAs: A potent source of regulation in immunity and disease. *Immunol Cell Biol*. vol. 93, no. 3, pp. 277–283. Mar. 2015.

30. F. van den Hoogen, D. Khanna, J. Fransen, *et al*. 2013 Classification criteria for systemic sclerosis: an ACR- EULAR collaborative initiative. *Ann Rheum Dis*. vol. 72, no. 11, pp. 1747–1755. Nov. 2013.

31. E.C. LeRoy, C. Black, R. Fleischmajer, *et al*. Scleroderma (systemic sclerosis): classification, subsets and pathogenesis. *J Rheumatol*. vol. 15, no. 2, pp. 202–205. Feb. 1988.

32. E.C. LeRoy, J. Medsger. Criteria for the classification of early systemic sclerosis. *J Rheumatol*. vol. 28, no. 7, pp. 1573–1576. Jul. 2001.

33. F. Calzetti, N. Tamassia, F. Arruda-Silva, *et al*. The importance of being "pure" neutrophils. *J Allergy Clin Immunol*. vol. 139, no. 1, pp. 352–55.e6. Jan. 2017.

34. H. Kambara, F. Niazi, L. Kostadinova, *et al*. Negative regulation of the interferon response by an interferon- induced long non-coding RNA. *Nucleic Acids Res*. vol. 42, no. 16, pp. 10668–10681. Aug. 2014.

35. M. v/d Kroef, M. Castellucci, M. Mokry, *et al*. Histone modifications underlie monocyte dysregulation in Systemic Sclerosis patients, underlining the treatment potential of epigenetic targeting. *Ann Rheum Dis*. vol. 78, no.4, pp. 529–538. Apr. 2019

36. C. Trapnell, L. Pachter, S.L. Salzberg. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics*. vol. 25, no. 9, pp. 1105–1111. May 2009.

37. M.I. Love, W. Huber, S. Anders. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. vol. 15, pp. 550. Dec. 2014.

38. S. Anders, P.T. Pyl, W. Huber. HTSeq-A Python framework to work with high-throughput sequencing data. *Bioinformatics*. vol. 31, no. 2, pp. 166–169. Sep. 2015.

39. J. Chen, E.E. Bardes, B.J. Aronow, *et al*. ToppGene suite for gene list enrichment analysis and candidate gene prioritization. *Nucleic Acids Res*. vol. 37, pp. W305–W311. Jul. 2009.

40. P. Langfelder, S. Horvath. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*. vol. 9, pp. 559. Dec. 2008.

41. F. Supek, M. Bošnjak, N. Škunca, *et al*. REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS ONE*. vol. 6, pp. e21800. Jul. 2011.

42. A. Schlicker, F.S. Domingues, J. Rahnenführer, *et al*. A new measure for functional similarity of gene products based on gene ontology. *BMC Bioinformatics*. vol. 7, pp. 302. Jun. 2006.

43. P. Shannon, A. Markiel, O. Ozier, *et al*. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*. vol. 13, no. 11, pp. 2498–2404. Nov. 2003.

44. P.Y. Muller, H. Janovjak, A.R. Miserez, *et al*. Processing of gene expression data generated by quantitative real-time RT-PCR. *Biotechniques*. vol. 32, no. 6, pp. 1372–1379. Jun. 2002.

45. W. de Jager, B. Prakken, G.T. Rijkers. Cytokine multiplex immunoassay: methodology and (clinical) applications. *Methods Mol Biol*. vol. 514, pp. 119–133. 2009.

46. S. Guil, M. Esteller. Cis-acting noncoding RNAs: friends and foes. *Nat Struct Mol Biol*. vol. 19, no. 11, pp. 1068–1075. Nov. 2012.

47. V. Toshchakov, B.W. Jones, P-Y. Perera, *et al*. TLR4, but not TLR2, mediates IFN-beta-induced STAT1alpha/beta- dependent gene expression in macrophages. *Nat Immunol*. vol. 3, no. 4, pp. 392–398. Apr. 2002.

48. N. Tamassia, V. Le Moigne, F. Calzetti, *et al*. The MyD88-independent pathway is not mobilized in human neutrophils stimulated via TLR4. *J Immunol*. vol. 178, no. 11, pp. 7344–7356. Jun. 2007

49. X. Liu, M.D. Mayes, F.K. Tan, *et al*. Correlation of interferon-inducible chemokine plasma levels with disease severity in systemic sclerosis. *Arthritis Rheum*. vol. 65, no. 1, pp. 226–235. Jan. 2013.

50. C. Crescioli, C. Corinaldesi, V. Riccieri, *et al*. Association of circulating CXCL10 and CXCL11 with systemic sclerosis. *Ann Rheum Dis*. vol. 77, no. 12, pp. 1845–1846. Dec. 2018.

2
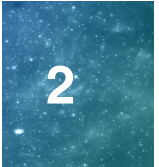
51. J.U. Lee, H.S. Cheong, E.Y. Shim, *et al.* Gene profile of fibroblasts identify relation of CCL8 with idiopathic pulmonary fibrosis. *Respir Res*. vol. 18, no. 1, pp. 3. Jan. 2017.

52. M. Ciechomska, R. Cant, J. Finnigan, *et al.* Role of toll- like receptors in systemic sclerosis. *Expert Rev Mol Med*. vol. 15, pp. e9. Aug. 2013.

53. S. O'Reilly. Innate immunity in systemic sclerosis pathogenesis. *Clin Sci*. vol. 126, no. 5, pp. 329–337. Mar. 2014.

54. N. Binai, S. O'Reilly, B. Griffiths, et al. Differentiation potential of CD14+ monocytes into myofibroblasts in patients with systemic sclerosis. *PLoS ONE*. vol. 7, no. 3, pp. e33508. Mar. 2012.

55. A. Lescoat, V. Lecureur, M. Roussel, *et al.* CD16-positive circulating monocytes and fibrotic manifestations of systemic sclerosis. *Clin Rheumatol*. vol. 36, no. 7, pp. 1649–1654. Jul. 2017.

56. S. Bhattacharyya, W. Wang, W. Qin, *et al.* TLR4- dependent fibroblast activation drives persistent organ fibrosis in skin and lung. *JCI Insight*. vol. 3, no. 13, pp. e98850. Jul. 2018.

57. B. Vigone, A. Santaniello, M. Marchini, *et al.* Role of class II human leucocyte antigens in the progression from early to definite systemic sclerosis. *Rheumatol*. vol. 54, no. 4, pp. 707–11. Apr. 2015.

58. M. Koenig, F. Joyal, M.J. Fritzler, *et al.* Autoantibodies and microvascular damage are independent predictive factors for the progression of Raynaud's phenomenon to systemic sclerosis: a twenty-year prospective study of 586 patients, with validation of proposed criteria for early systemic sclerosis. *Arthritis Rheum*. vol. 58, no. 12, pp. 3902–3912. Dec. 2008.

59. M. Guttman, I. Amit, G. Garber, *et al.* Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature*. vol. 458, no. 7235, pp. 223–227. Mar. 2009.

60. Z. Xu-Yang, B. Pei-Yu, Y. Chuan-Tao, *et al.* Interferon-induced transmembrane protein 3 inhibits hantaan virus infection, and its single nucleotide polymorphism rs12252 influences the severity of hemorrhagic fever with renal syndrome. *Front Immunol*. vol. 7, pp. 535. Jan. 2017.

61. B.T. Roux, J.A. Heward, L.E. Donnelly, *et al.* Catalog of differentially expressed long non-coding RNA following activation of human and mouse innate immune response. *Front Immunol*. vol. 8, pp. 1038. Aug. 2017.

62. M.J. Pearson, A.M. Philp, J.A. Heward, *et al.* Long intergenic noncoding rnas mediate the human chondrocyte inflammatory response and are differentially expressed in osteoarthritis cartilage. *Arthritis Rheumatol*. vol. 68, no. 4. pp. 845–56. Apr. 2016

63. G. Farina, D. Lafyatis, R. Lemaire, *et al.* A four-gene biomarker predicts skin disease in patients with diffuse cutaneous systemic sclerosis. *Arthritis Rheum*. vol. 62, no. 2, pp. 580–588. Feb. 2010.

64. P. Airò, C. Ghidini, C. Zanotti, *et al.* Upregulation of myxovirus-resistance Protein A: a possible marker of Type I interferon induction in systemic sclerosis. *J Rheumatol*. vol. 35, no. 11, pp. 2192–2200. Oct. 2008.

65. R.B. Christmann, E. Hayes, S. Pendergrass, *et al.* Interferon and alternative activation of monocyte/macrophages in systemic sclerosis-associated pulmonary arterial hypertension. *Arthritis Rheum*. vol. 63, no. 6, pp. 1718–1728. Jun. 2011.

66. A. Antonelli, C. Ferri, P. Fallahi, *et al.* CXCL10 (α) and CCL2 (β) chemokines in systemic sclerosis - a longitudinal study. *Rheumatology*. vol. 47, no. 1, pp. 45–49. Jan. 2008.

67. X. Guo, B.W. Higgs, A-C. Bay-Jensen, *et al.* Suppression of T cell activation and collagen accumulation by an anti-type I interferon receptor monoclonal antibody in adult patients with systemic sclerosis. *J Invest Dermatol*. vol. 135, no. 10, pp. 2402–2409. Oct. 2015.

SUPPLEMENTARY INFORMATION

The Supplementary Tables for this article can be found online at: *https://www.frontiersin.org/articles/10.3389/fimmu.2019.00100/full#supplementary-material*.
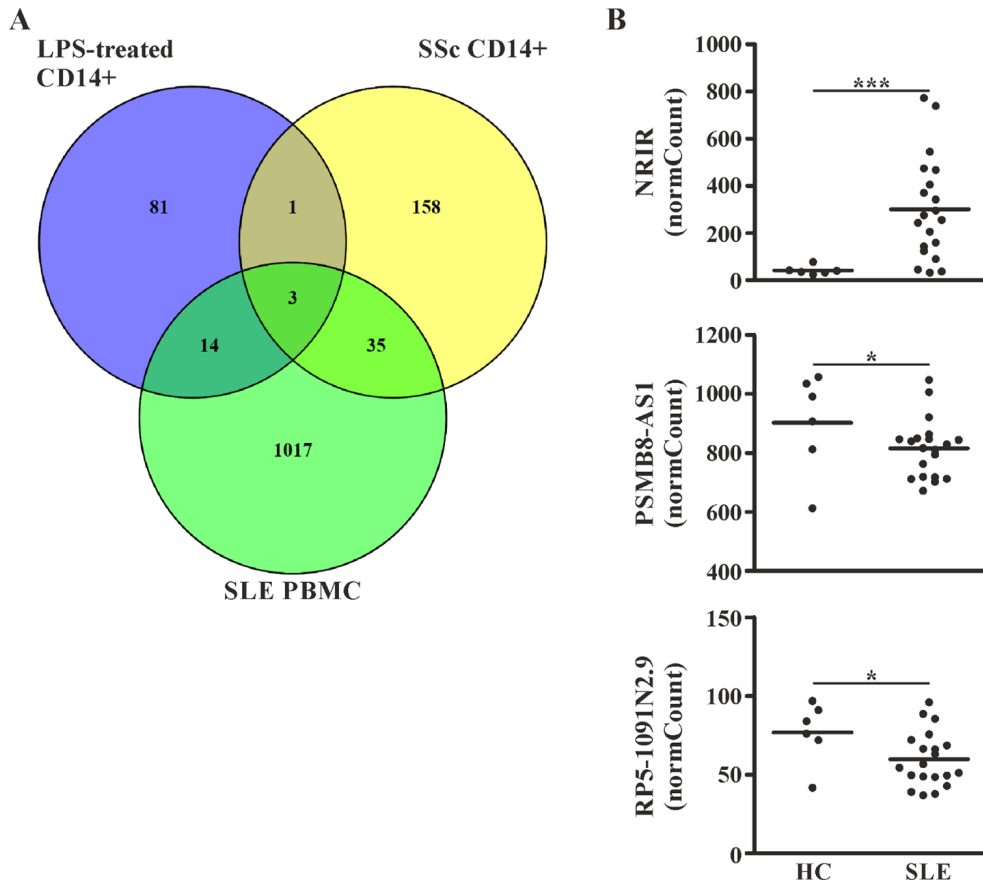
Figure S1. NRIR expression is increased in PBMC from SLE. RNA-seq data of LPS-treated CD14+ monocytes, CD14+ monocytes from SSc and matched healthy controls (HC) and PBMC from SLE patients and relative HC were analyzed as described in Materials and Methods. A) Venn diagram representing the IFN/viral-related lncRNAs modulated by LPS (blue), lncRNAs modulated in at least two SSc groups as compared to HC (yellow) and lncRNAs differentially expressed in PBMC from SLE patients as compared to relative HC (green). The number of specific or common lncRNAs is reported. B) Expression levels of NRIR, PSMB8-AS1 and RP5-1091N2.9 in PBMC from SLE patients as compared to HC is shown. Expression levels are reported as normalized count (normCount) according to DESeq2 normalization. $* p < 0.05$, $*** p < 0.001$ by Wald test.
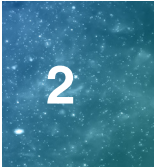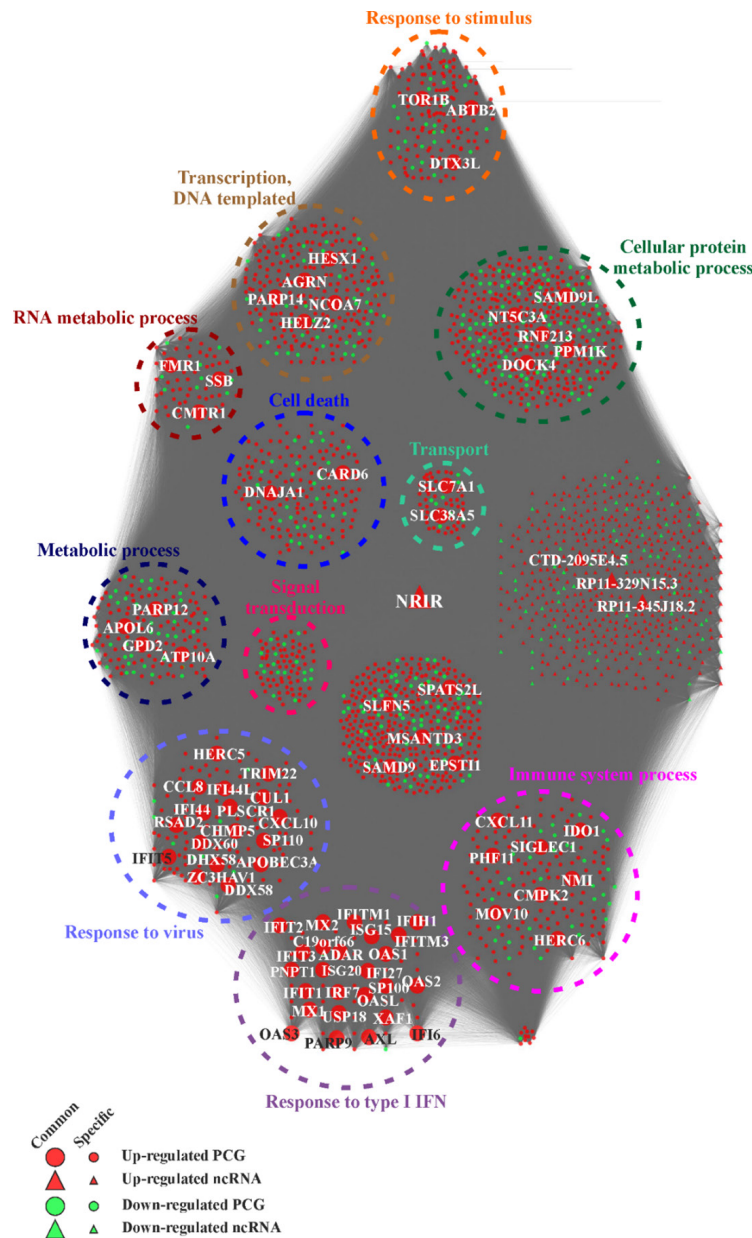
Figure S2. Gene co-expression network of the blue-module. Gene co-expression network was built starting from the LPS-treated and untreated monocytes transcriptome as described in Materials and Methods. The NRIR-associated module (blue-module) is shown. The 2060 protein coding and the 548 non-coding transcripts are represented as circles and triangles, respectively. Red nodes represent LPS-upregulated transcripts while green nodes represent down-regulated transcripts. Nodes' size indicates the overlap between blue-module and the NRIR-associated module identified in the SSc network, where the small, unlabeled nodes represent transcripts specific of the blue-module. Transcripts are grouped according to their associated biological process related GO-terms; the most general GO term, summarizing each group is reported.
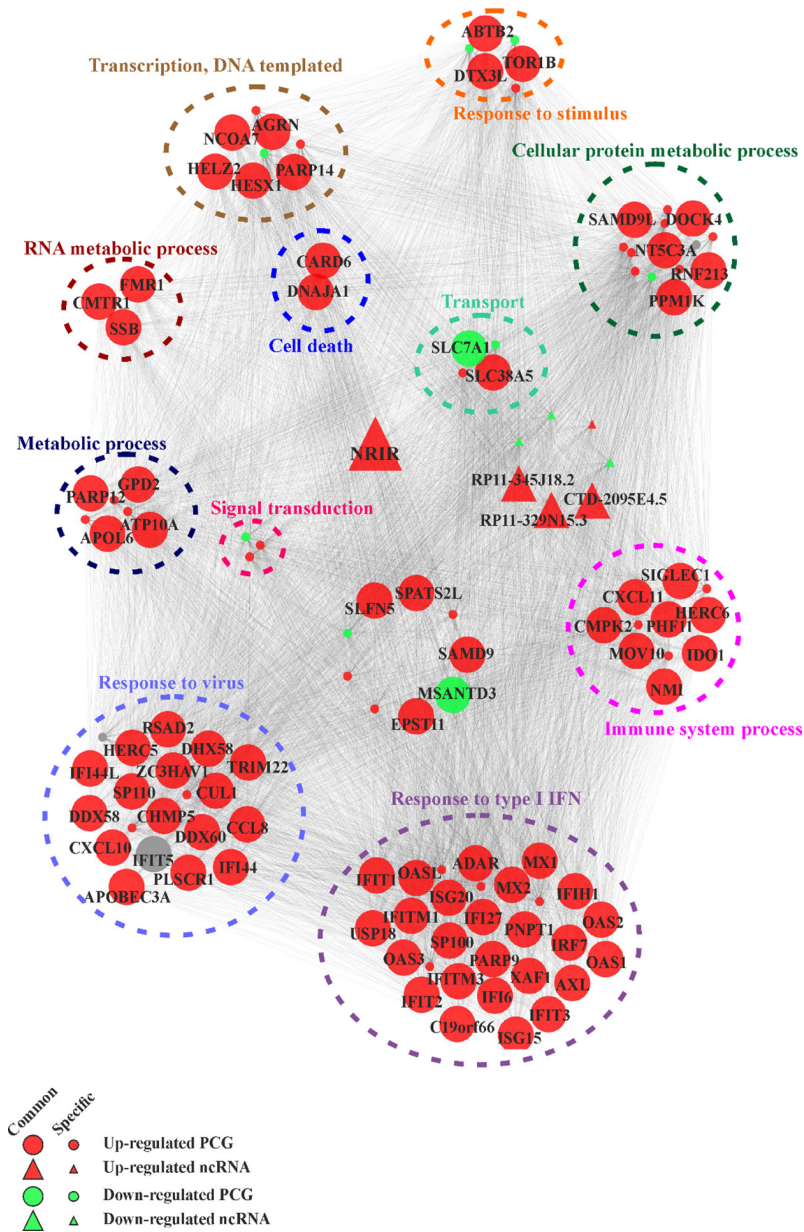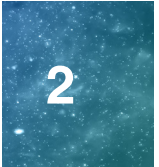
Figure S3. Gene co-expression network of the cyan-module. Gene co-expression network was built starting from the SSc monocytes transcriptome as described in Materials and Methods. The NRIR-associated module (cyan-module) is shown. The 116 protein coding and the 8 non-coding transcripts are represented as circles and triangles, respectively. Red nodes represent transcripts more expressed in at least one SSc subset compared to HC while, green nodes represent less expressed transcripts. Nodes' size indicates the overlap between cyan-module and the NRIR-associated module identified in the LPS-treated monocytes network, where the small, unlabeled nodes represent transcripts specific of the cyan-module. Transcripts are grouped according to their associated biological process related GO-terms; the most general GO term, summarizing each group is reported.
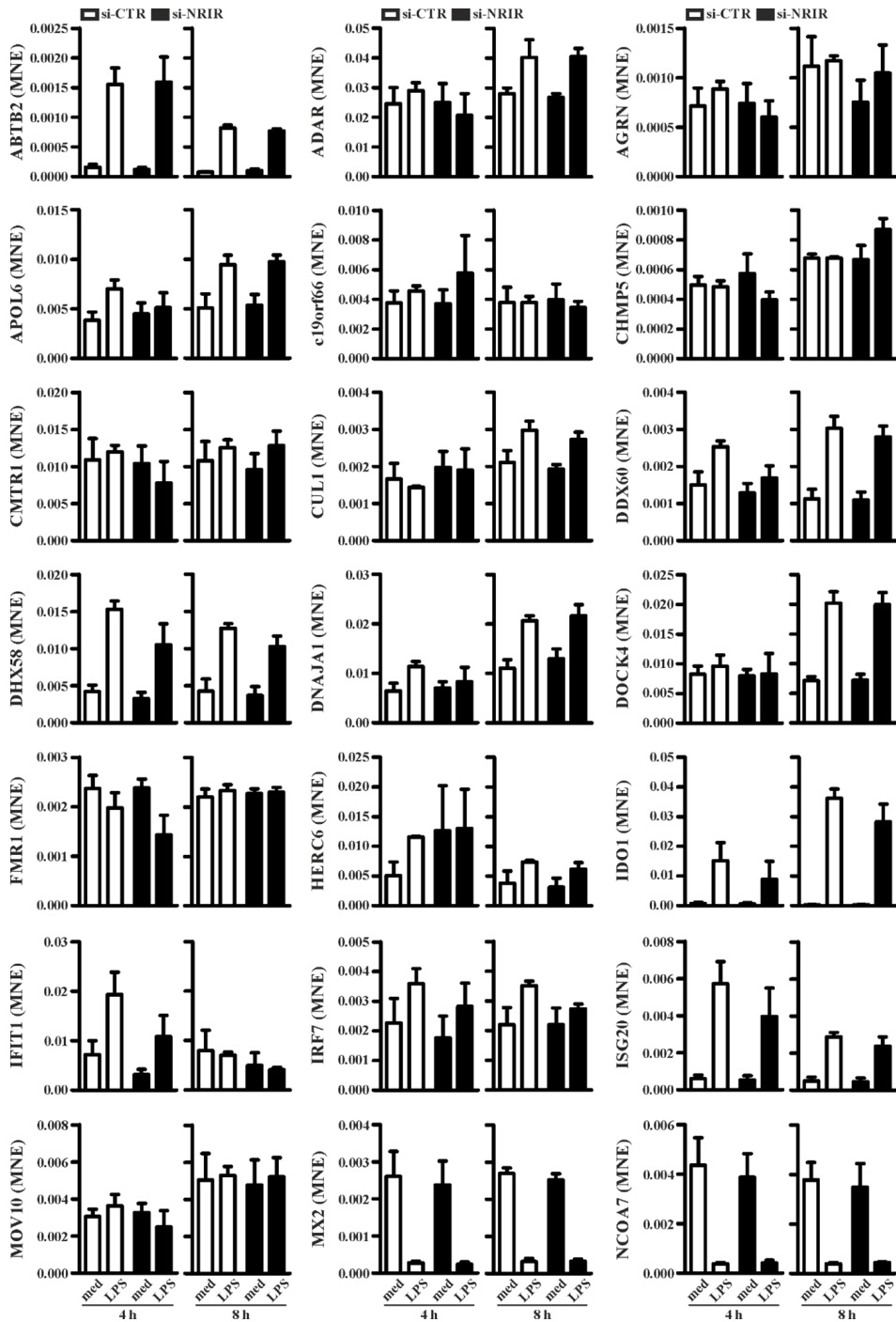
Figure S4. Effect of NRIR silencing on the expression of its putative target genes. CD14+ monocytes were transfected with si-NRIR or si-CTR and 18h later were stimulated with LPS for 4h or 8h or left untreated. The expression of putative NRIR-target genes was analyzed by RT-qPCR and expressed as MNE. Results are shown as mean ± SEM of at least three different experiments.
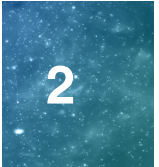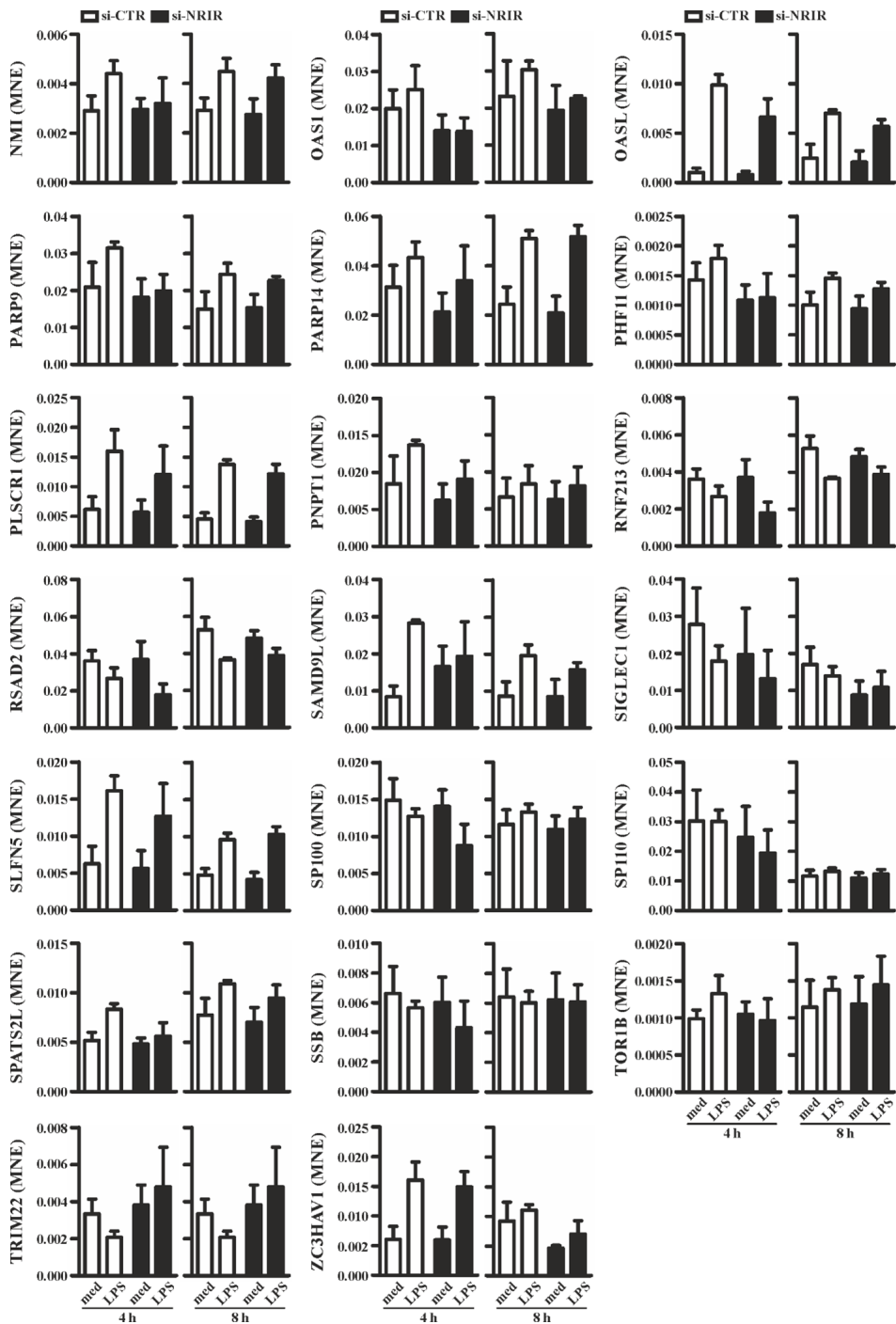
Figure S5. Effect of NRIR silencing on the expression of its putative target genes. CD14+ monocytes were transfected with si-NRIR or si-CTR and 18h later were stimulated with LPS for 4h or 8h or left untreated. The expression of putative NRIR-target genes was analyzed by RT-qPCR and expressed as MNE. Results are shown as mean ± SEM of at least three different experiments.

Chapter 3

Characterization of Long Non-Coding RNAs in Systemic Sclerosis Monocytes: A Potential Role for PSMB8-AS1 in Altered Cytokine Secretion

3

Authors:
N.H. Servaas[1,2*], B. Mariotti[3*], M. van der Kroef[1,2], C.G.K. Wichers[1,2], A. Pandit[1,2], F. Bazzoni[3], T.R.D.J. Radstake[1,2], M. Rossato[4]

*These authors contributed equally

Affiliations:
1. Department of Rheumatology and Clinical Immunology, University Medical Center Utrecht, Utrecht, Netherlands.
2. Laboratory of Translational Immunology, University Medical Center Utrecht, Utrecht, Netherlands.
3. General Pathology Section, Department of Medicine, University of Verona, Verona, Italy.
4. Department of Biotechnology, University of Verona, Verona, Italy.

ABSTRACT

Systemic sclerosis (SSc) is a chronic autoimmune disease mainly affecting the connective tissue. In SSc patients, monocytes are increased in circulation, infiltrate affected tissues, and show a pro-inflammatory activation status, including the so-called interferon (IFN) signature. We previously demonstrated that the dysregulation of the IFN response in SSc monocytes is sustained by altered epigenetic factors as well as by upregulation of the long non-coding RNA (lncRNA) NRIR. Considering the enormously diverse molecular functions of lncRNAs in immune regulation, the present study investigated the genome-wide profile of lncRNAs in SSc monocytes, with the aim to further unravel their possible role in monocyte dysregulation and disease pathogenesis. Transcriptomic data from two independent cohorts of SSc patients identified 886 lncRNAs with an altered expression in SSc monocytes. Differentially expressed lncRNAs were correlated with neighboring protein coding genes implicated in the regulation of IFN responses and apoptotic signaling in SSc monocytes. In parallel, gene co-expression network analysis identified the lncRNA PSMB8-AS1 as a top-ranking hub gene in co-expression modules implicated in cell activation and response to viral and external stimuli. Functional characterization of PSMB8-AS1 in monocytes demonstrated that this lncRNA is involved in the secretion of IL-6 and TNFα, two pivotal pro-inflammatory cytokines altered in the circulation of SSc patients and associated with fibrosis and disease severity. Collectively, our data showed that lncRNAs are linked to monocyte dysregulation in SSc, and highlight their potential contribution to disease pathogenesis.

INTRODUCTION

Systemic sclerosis (SSc) is a chronic autoimmune disease with a highly heterogeneous clinical phenotype[1]. The disease is characterized by three main hallmarks: vascular abnormalities, immune system dysregulation, and fibrosis. Based on the extent of skin fibrosis and the presence of vascular and immunological abnormalities, SSc patients can be divided into four subsets: early SSc (eaSSc), non-cutaneous SSc (ncSSc), limited-cutaneous SSc (lcSSc), and diffuse-cutaneous SSc (dcSSc)[2, 3]. Vascular abnormalities characterize the pre-clinical stage of SSc, and Raynaud's phenomenon (RP) occurs in 90–98% of patients with SSc, often preceding the disease onset by years[4]. However, the exact immunopathogenic mechanisms leading to the onset and contributing to the progression of SSc remain to be elucidated. Vascular injury and endothelial cell activation appear to be the earliest events in SSc pathogenesis[5]. This vascular damage is hypothesized to lead to the recruitment and activation of various immune cell types including lymphocytes, dendritic cells, and monocytes, which secrete various pro-inflammatory cytokines and growth factors such as IL-6, IL-8, IL-13, TNFα, TGFβ, and MCP-1[6]. The resulting mix of inflammatory mediators induces the differentiation of resident epithelium, endothelium, and fibroblasts into myofibroblasts that deposit excessive amounts of extracellular matrix, leading to fibrosis and permanent tissue scarring[6].

Several lines of evidence implicate monocytes as an important cell type in SSc pathogenesis. Monocytes are among the predominant infiltrating mononuclear cells in SSc skin lesions[7–9], suggesting that these cells are involved in the fibrotic processes underlying the disease. In addition, the population of circulating monocytes is increased in the peripheral blood of SSc patients, and their frequency is correlated with the extent of skin fibrosis and the occurrence of interstitial lung disease (ILD)[10, 11]. Besides their increased frequencies, SSc monocytes also display signs of enhanced activation, evident from an increased expression of interferon (IFN) responsive genes (referred to as the type I IFN signature)[12, 13], and an enhanced production of pro-inflammatory and pro-fibrotic mediators[14–16]. Together, this evidence suggests a critical role for monocytes in the pathogenesis of SSc, linking immune aberrances and fibrosis.

Epigenetic[17, 18] and miRNA-associated[13] alterations have been proposed as potential contributors to monocyte dysregulation in SSc. Next to these, long non-coding RNAs (lncRNAs) have recently gained widespread attention as critical biological regulators of gene expression in immune cells including monocytes[19] and have been linked to SSc pathogenesis[20–24]. lncRNAs are broadly defined as RNA transcripts longer than 200 nucleotides that lack protein coding capacity. They are involved in virtually all levels of gene expression regulation through a variety of biological mechanisms[25]. Based on the genomic localization relative to their targets, lncRNAs can be categorized as cis- or trans-acting, regulating the expression of neighboring or distal protein coding genes[26]. Additionally, the subcellular localization of lncRNAs also underlies their function[27]. Chromatin-associated and nuclear lncRNAs are often involved in the regulation of transcriptional processes, for example, through chromatin remodeling or the recruitment of transcription factors[28], while cytoplasmic lncRNAs most frequently act on post-transcriptional levels, for example, through miRNA sponging, the regulation

3
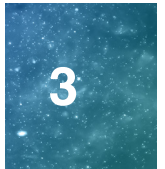
of mRNA translation, or the alteration of protein activity[29].

We recently demonstrated that the lncRNA NRIR plays an important role in the regulation of type I IFN responses in monocytes[20]. NRIR is upregulated in SSc monocytes and promotes IFN-related pathways, thereby contributing to the type I IFN signature observed in these cells[20]. Because of the broad molecular functions of lncRNAs and their involvement in immune system regulation, we hypothesized that more lncRNAs may be implicated in the altered molecular processes characterizing monocytes of SSc patients. Exploiting transcriptomic data of monocytes obtained from SSc patients and matched healthy controls, we identified multiple lncRNAs potentially involved in the regulation of apoptotic pathways and IFN signaling in SSc monocytes. In addition, combining in silico and in vitro approaches, we identified the lncRNA PSMB8-AS1 as a potential regulator of cytokine release in SSc monocytes.

RESULTS

The Expression of lncRNAs Is Altered in SSc Monocytes and Is Correlated with Neighboring Protein Coding Genes

The genome-wide expression profile of lncRNAs in healthy and SSc monocytes was initially assessed in the "Definite SSc cohort", comprising ncSSc (n = 7), lcSSc (n = 11) and dcSSc (n = 7) patients, and matched healthy controls (HC, n = 9, Table 1). A total of 886 lncRNAs were found to be differentially expressed in at least one group of SSc patients versus healthy controls (log2(FC) > 0.58 or < 0.58, and p-value < 0.05) (Figure S1A, Table S1). Of these, 22 lncRNAs were commonly altered in all SSc subsets (Figure S1B).

Since lncRNAs often regulate the transcription of neighboring protein coding genes (PCGs)[30], in cis correlation analysis was performed to identify putative target genes of differentially expressed lncRNAs in the Definite cohort. To this end, the expression levels of differentially expressed lncRNAs were correlated with PCGs located 5 kb upstream or downstream of each lncRNA gene (Figure 1A). Out of 886 differentially expressed, 278 lncRNAs were significantly correlated with PCGs localized in cis (Spearman's rho > 0.4 or < 0.4, and p-value 0.05), allowing for the identification of 332 lncRNA-PCG pairs. Functional enrichment analysis of the correlated PCGs identified pathways associated with IFN response, negative regulation of apoptosis, and inflammatory cell apoptotic processes (Figure 1B), indicating that lncRNAs altered in SSc monocytes potentially regulate genes involved in these pathways.

In order to substantiate these results and identify putative lncRNA-PCG pairs involved already at the early stages of SSc development, we repeated the cis correlation analysis in an additional cohort comprising ncSSc patients and SSc patients at the early disease stage (eaSSc), as well as individuals with Raynaud's Phenomenon (RP) ("Non-cutaneous cohort", Table 1). This analysis highlighted that 143 out of 332 correlated lncRNA-PCG pairs identified in the Definite cohort were reproduced in the Non-cutaneous cohort (Spearman's rho >0.4 or < 0.4, and p-value 0.05, Figure 1C). GO-term enrichment analysis of the PCGs identified in the replicated pairs showed a significant enrichment

	HC (9)	-	-	ncSSc (7)	lcSSc (11)	dcSSc (7)
Definite Cohort	HC (9)	-	-	ncSSc (7)	lcSSc (11)	dcSSc (7)
Non-Cutaneous cohort	HC (9)	RP (9)	eaSSc (11)	ncSSc (10)	-	-
Replication cohort	HC (8)	-	eaSSc (5)	ncSSc (6)	lcSSc (10)	dcSSc (6)
Age (yr.)	52 (30–64)	-	-	45 (26–63)	59 (45–70)	58 (34–72)
	38 (28–49)	47(22–70)	57 (40–77)	52 (25–70)	-	-
	57 (31–64)	-	47 (22–61)	41 (36–55)	58 (38–69)	56 (53–72)
Female/Male, n	5/4	-	-	6/1	8/3	3/4
	9/0	9/0	11/0	10/0	-	-
	7/1	-	4/1	4/2	8/2	4/2
ANA, n (% pos.)	-	-	-	6 (86%)	10 (91%)	7 (100%)
	-	3 (33%)	10 (91%)	10 (100%)	-	-
	-	-	4 (80%)	6 (100%)	8 * (80%)	6 (100%)
ACA, n (% pos.)	-	-	-	3 (43%)	6 * (55%)	1 (14%)
	-	0 (0%)	7 (64%)	8 (80%)	-	-
	-	-	1 (20%)	1 (17%)	4 * (40%)	1 (17%)
Scl70, n (% pos.)	-	-	-	2 (29%)	2* (18%)	4 (57%)
	-	0 (0%)	2 (18%)	1 (10%)	-	-
	-	-	1 (20%)	2 (33%)	2 * (20%)	4 (67%)
ILD, n (% pos.)	-	-	-	1 (14%)	2 (18%)	5 (71%)
	-	0 (0%)	0 (0%)	0 (0%)	-	-
	-	-	1 (20%)	2 (33%)	3 * (30%)	3 (50%)
mRSS	-	-	-	0	6 (0–12)	14* (5–36)
	-	0	0	0	-	-
	-	-	0	0	4* (2–14)	13 (4–23)
Tel., n (%)	-	-	-	3 * (43%)	4 (36%)	4 (57%)
	-	0 (0%)	1 (9%)	4 (40%)	-	-
	-	-	1 * (20%)	3 (50%)	6 * (60%)	2 * (33%)
NVC early, n (%)	-	-	-	2 * (29%)	2 ** (18%)	1 ***(14%)
	-	0 (0%)	9 (82%)	5 (50%)	-	-
	-	-	3 * (60%)	4 (66%)	3 ** (30%)	3 (50%)
NVC late/active, n (%)	-	-	-	4 * (57%)	3 ** (27%)	2 ***(28%)
	-	0 (0%)	0 (0%)	5 (50%)	-	-
	-	-	1 * (20%)	2 (33%)	1 ** (10%)	3 (50%)
Steroids, n (%)	-	-	-	0 (0%)	0 (9%)	2 (28%)
	-	0 (0%)	0 (0%)	0 (0%)	-	-
	-	-	1 * (20%)	0 (0%)	1 * (10%)	1 (17%)
Immunosup., n (%)	-	-	-	0 (0%)	1 (9%)	3 (43%)
	-	0 (0%)	1 (9%)	0 (0%)	-	-
	-	-	1 (20%)	2 (33%)	1 * (10%)	4 (66%)

3

Table 1. Demographics and clinical features of subjects enrolled in the study. Values reported indicate the number (n) of patients and the median for each parameter (Interquartile Range (IQR)). ACA, anticentromere antibodies; ANA, antinuclear antibodies; dcSSc, diffuse cutaneous SSc; eaSSc, early SSc; HC, healthy controls; ILD, Interstitial Lung Disease; Immunosupp., immunosuppressive therapy; lcSSc, limited cutaneous SSc; mRSS, modified Rodnan Skin Score; ncSSc, non-cutaneous SSc; NVC, nailfold videocapillaroscopy; pos, positivity; RP, Raynaud's Phenomenon; Scl70, anti-topoisomerase antibodies; Tel., telangiectasia; yr., years. * = 1 patient, ** = 6 patients, *** = 3 patients unknown.
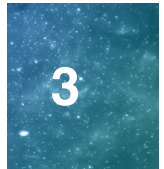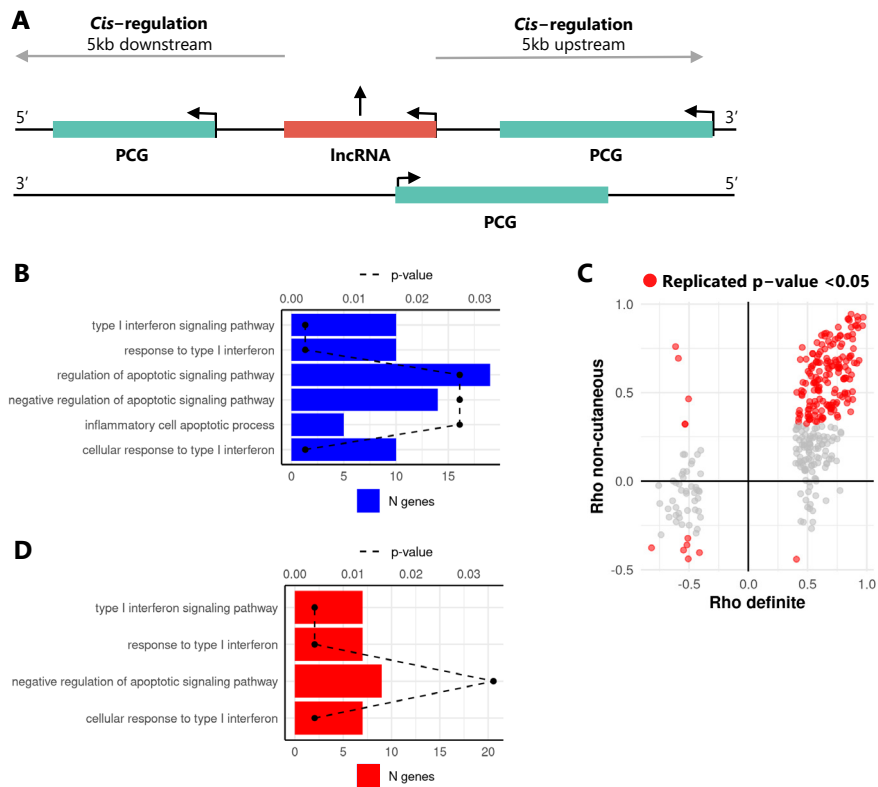
Figure 1. Differentially expressed lncRNAs are correlated with cis localized protein coding genes relevant for SSc pathogenesis. (A) Schematic overview of the cis correlation approach to identify neighboring long non-coding RNAs (lncRNA, red) and protein coding genes (PCGs, green). Arrows indicate the direction of transcription. (B) GO-term enrichment analysis results of PCGs identified in the cis correlation analysis in the Definite SSc cohort. GO terms for significantly enriched biological processes are given on the y-axis. Bars depict the number of genes identified within the enriched pathway (N genes, bottom x-axis), dashed line indicates B&H corrected p-value of the enrichment (p-value, top x-axis). (C) Correlation coefficients (Spearman's Rho) between lncRNAs and PCGs in the Definite (x-axis) and Non-cutaneous (y-axis) cohorts. Each dot represents a lncRNA-PCG pair. Grey dots represent pairs significantly correlated in the Definite cohort only, while red dots represent pairs significantly correlated in both cohorts (Spearman's rho >0.4 or <–0.4, and p-value ≤ 0.05). (D) GO-term enrichment analysis results of protein coding genes from the in cis correlation analysis replicated the Non-cutaneous cohort.

for GO terms related to type I IFN and apoptosis (Figure 1D), suggesting that lncRNAs altered during SSc development are implicated in these processes. The 15 lncRNA-PCG pairs that were annotated in these biological pathways are given in Table 2.

Weighted Gene Co-Expression Network Analysis Identifies Clusters of Tightly Correlated RNAs Associated to SSc Clinical Features and Relevant Biological Processes

Next to cis regulatory lncRNAs, trans-acting lncRNAs are also emerging as important regulators of gene expression, especially at the post-transcriptional level[31]. To explore the regulatory potential of trans-acting lncRNAs in SSc monocytes, genome-

GO-Term	PCG	lncRNA	Definite cohort			Non-cutaneous cohort		
			BM	R	p	BM	R	p
Type I interferon response (GO:0071357, GO:0060337, GO:0034340)	PSMB8	PSMB8-AS1	645.23	0.67	0.000	262.01	0.68	0.000
	OAS1	RP1-71H24.6	46.71	0.50	0.003	17.74	0.62	0.000
	IRF2	RP11-326I11.3	75.09	0.61	0.000	40.19	0.73	0.000
	CACTIN	CACTIN-AS1	6.68	0.60	0.000	3.25	0.62	0.000
	IFITM3	RP11-326C3.11	14.01	0.58	0.000	9.65	0.68	0.000
	IFI6	RP11-288L9.4	16.93	0.41	0.017	7.28	0.50	0.001
	MX1	AP001610.5	7.46	0.59	0.000	8.20	0.72	0.000
	PAM16	RP11-295D4.3	30.27	0.88	0.000	40.74	0.80	0.000
Negative regulation of apoptotic signaling pathway (GO:2001234)	FAS	RP11-399O19.9	12.26	0.56	0.001	15.18	0.56	0.000
	BIRC6	AL133243.2	50.33	0.45	0.008	87.28	0.51	0.001
	THBS1	CTD-2033D15.2	40.78	0.76	0.000	76.70	0.87	0.000
	SGMS1	RP11-521C22.2	33.86	0.60	0.000	54.93	0.53	0.001
	CCAR2	RP11-582J16.5	34.78	0.59	0.000	64.17	0.67	0.000
	AATF	CTC-268N12.3	0.58	−0.41	0.015	0.56	−0.40	0.011
	TNAIP3	RP11-356I2.4	66.03	0.84	0.000	96.56	0.50	0.001
	IFI6	RP11-288L9.4	16.93	0.41	0.017	7.28	0.50	0.001

Table 2. Replicated cis correlating lncRNA-PCG pairs are annotated in biological processes relevant for SSc. PCG, protein coding gene; lncRNA, long non-coding RNA; BM, base mean expression level; R, Spearman's rank correlation coefficient; p, p-value.

wide co-expression network analysis was performed in parallel to the cis correlation analysis. Weighted gene co-expression network analysis (WGCNA) generated 18 distinct co-expression modules of highly correlated genes in the Definite cohort (Table S2). Correlation between module eigengenes (MEs) and clinical traits for SSc identified 10 co-expression modules that were significantly correlated with clinical parameters associated with SSc (Figure 2A, Pearson correlation, p-value < 0.05). Comparison of the global ME expression across SSc patients and healthy controls showed that the honeydew1, brown4, darkturquoise, yellowgreen, darkorange2, and paleviolet modules were lower in SSc patients, while the ME expression of the darkgreen module was higher in SSc patients (Figure 2B). Since the remaining violet, white, and blue modules did not show a distinct ME expression pattern in SSc patients, they were not considered for subsequent analysis (Figure 2B). Next, GO-term enrichment analysis was performed to annotate the seven modules with a distinct ME expression pattern in SSc patients. The darkgreen module was linked to vesicle transport and the regulation of autophagy, while the darkorange and yellowgreen modules were linked to IkappaB-NF-kappaB signaling, myeloid cell differentiation, and protein modifications (Figure 2C). No significant enrichment was identified for the remaining modules that were therefore not considered for further investigation (Figure 2D).

Identification of the lncRNA PSMB8-AS1 as a Reproducible Hub Gene Relevant for SSc and Monocyte Biology

The co-expression network analysis was repeated in the Non-cutaneous cohort, and the extent of overlap between the modules from the Definite and Non-cutaneous cohorts was assessed to identify reproducible co-expression modules (Figure S2, Table S3). Thirteen modules from the Non-cutaneous cohort showed a significant overlap with the 3 selected modules from the Definite cohort (i.e., darkgreen, darkorange, and
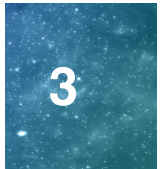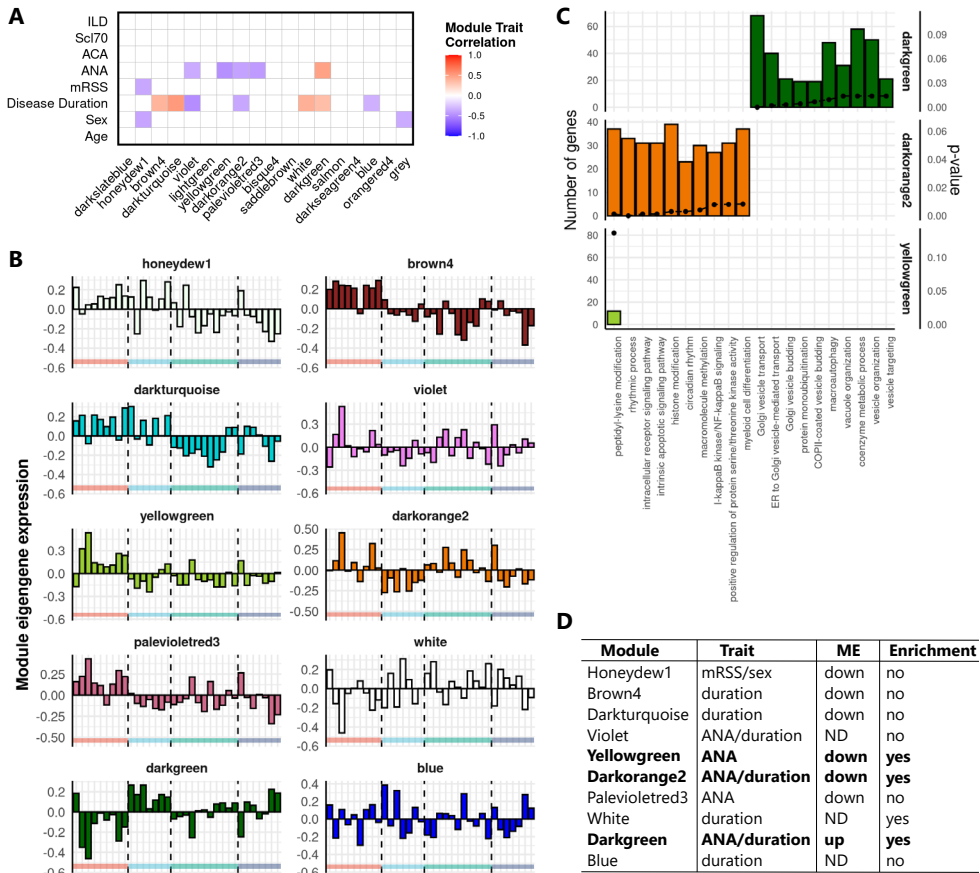
Figure 2. Identification and functional annotation of co-expression modules correlated with clinical traits in SSc. (A) Correlation of module eigengenes (MEs) to SSc clinical traits. Rows indicate the clinical traits, and columns indicate modules identified in the Definite cohort. Cells of significant correlations (Pearson, p-value < 0.05) are color-coded by the degree and direction of the correlation (red = positive; blue = negative). Abbreviations: mRSS, modified Rodnan Skin Score; ANA, antinuclear antibodies; ACA, anticentromere antibodies; Scl70, anti-topoisomerase I antibodies; ILD, interstitial lung disease. (B) ME expression (first principal component, y-axis) of modules significantly correlated with clinical traits. Bars represent individual donors grouped according to their disease subset represented by the colors on the x-axis (red = HC, light blue = ncSSc, green = lcSSc, dark blue = dcSSc). (C) GO-term enrichment analysis of selected modules. Top 10 enriched terms for each module are shown (B&H corrected p-value < 0.05). Bars depict the number of module genes associated to enriched GO terms, and dots represent the p-value for the enrichment. (D) Table indicating the characteristics for selected modules (Column 1), considering correlations to clinical traits (Column 2), distinct ME expression pattern versus heathy controls (Column 3, up = higher ME in SSc, down = lower ME in SSc, and ND = ME not distinct from healthy subjects), and functional enrichment (Column 4). Modules that were selected for subsequent analysis are highlighted in bold.

yellowgreen, Fisher's exact test, p-value < 0.05, Figure 3A), demonstrating that clinically relevant modules are reproducible across different cohorts of SSc monocytes, including pre-clinical SSc stages.
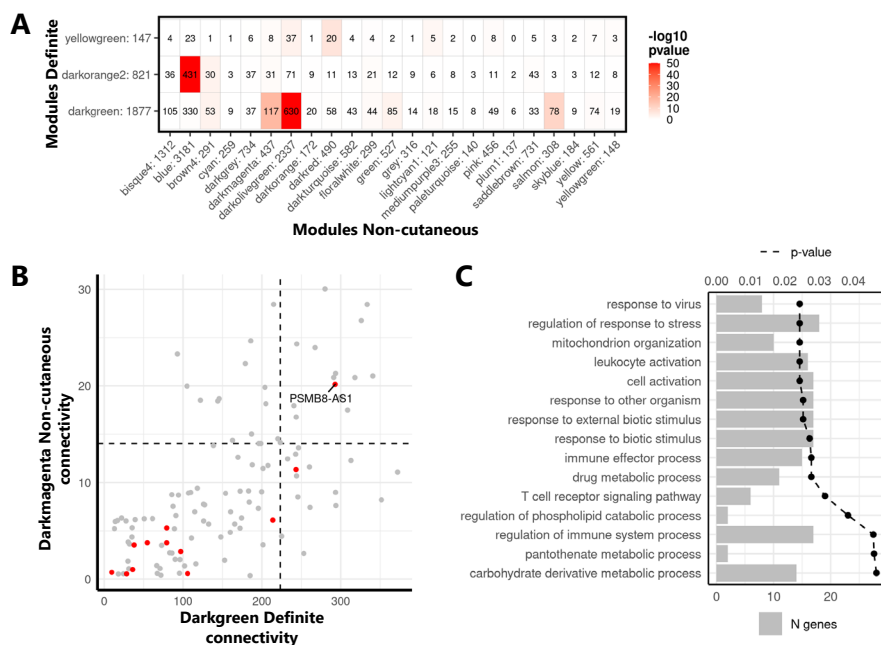
Figure 3. Identification of PSMB8-AS1 as a hub gene in replicated network modules. (A) Overlap of selected modules from the Definite (rows) and Non-cutaneous (columns) cohorts. Numbers behind module names indicate the total number of genes in the modules. Numbers in the table indicate the number of genes overlapping between two modules. Coloring indicates the significance of the overlap (Fisher's exact test, –10log (p-value)). (B) Intramodular connectivity of genes shared across the darkgreen module of the Definite cohort (x-axis) and the darkmagenta module of the Non-cutaneous cohort (y-axis). Each dot represents one gene, with lncRNAs highlighted in red. The top 25% most connected genes in both cohort modules (threshold indicated by black dashed lines) were considered as hub genes. (C) The GO-term enrichment of genes replicated across the darkgreen/darkmagenta modules. Bars depict the number of genes identified in the enrichment, the dotted line represents the B&H corrected p-value. Top 15 enriched terms are shown (B&H corrected p-value < 0.05).

Next, by comparing the intramodular connectivity of shared genes across the overlapping modules (Figure S3), we identified replicated hub genes with high connectivity within modules across both cohorts. While several lncRNAs were present in the replicated modules, PSMB8-AS1 was the only lncRNA identified among the top 25% most highly connected genes (Figure 3B). Specifically, PSMB8-AS1 was a replicated hub gene in the darkgreen and darkmagenta modules from the Definite and Non-cutaneous networks, respectively, of which overlapping genes are enriched in genes related to immune cell activation and to response to virus and external stimulus (Figure 3C). Overall, these results pointed to PSMB8-AS1 as a potential central player in the regulation of these molecular pathways that are also relevant processes for monocyte activation in SSc.

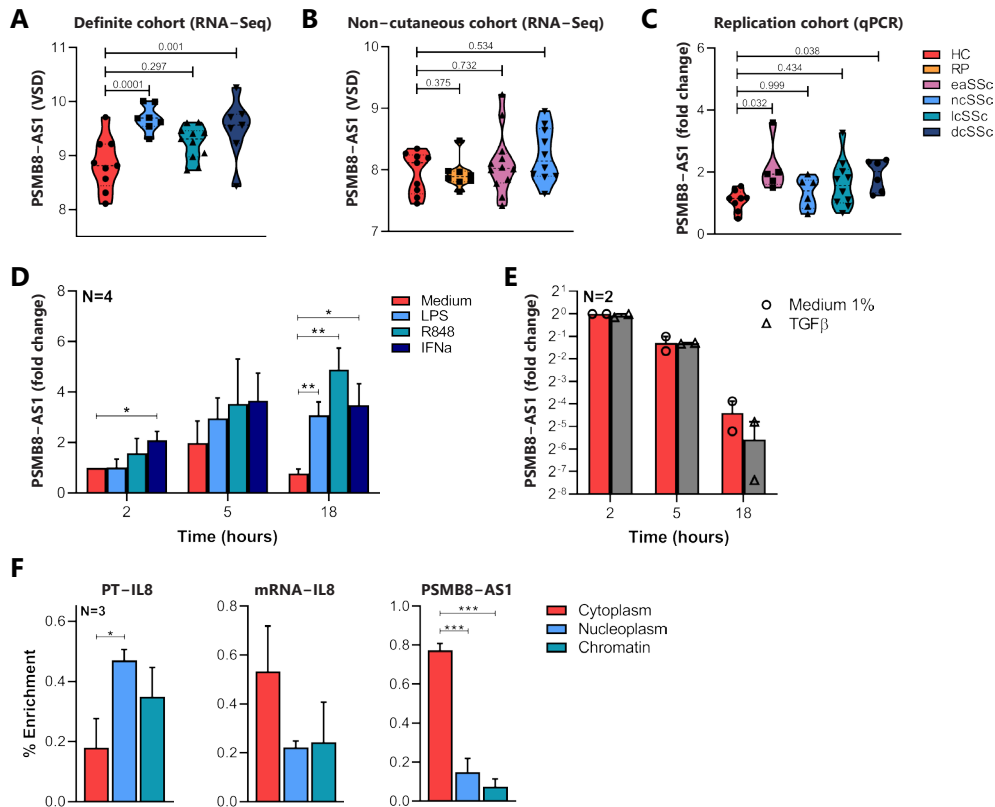## Characterization of PSMB8-AS1 Expression in SSc and Healthy Monocytes

Both the cis correlation and co-expression network analysis highlighted PSMB8-AS1 as a putative key regulator of biological processes relevant for SSc pathogenesis and monocyte activation. PSMB8-AS1 expression was significantly upregulated in ncSSc and dcSSc patients in the Definite cohort (Figure 4A). A trend for PSMB8-AS1 upregulation was also observed in the lcSSc patients in the Definite cohort and in eaSSc and ncSSc patients of the Non-cutaneous cohort (Figure 4A,B). To confirm the altered expression of PSMB8-AS1 across SSc patients, its expression was evaluated by a target specific RT-qPCR in an additional, independent cohort of SSc patients (Replication cohort). A statistically significant altered expression of PSMB8-AS1 was confirmed in SSc patients with the earliest symptoms (eaSSc) and the most severe phenotype (dcSSc patients) (p-value < 0.05, Figure 4C). In the other SSc groups, only some individuals displayed the upregulation of this lncRNA, but this, however, was not significantly changed when considering the whole group.

To identify the molecular pathways leading to PSMB8-AS1 induction in SSc, its expression was assessed in healthy human monocytes cultured for 2, 5, and 18 h in the presence of LPS (TLR4 ligand), R848 (TLR7/8 ligand), IFNα, and TGFβ, all stimuli linked to monocyte activation and fibrosis in SSc[32–35]. PSMB8-AS1 expression was strongly induced by LPS, R848, and IFNα, especially after 18 h of stimulation (Figure 4D, p-value < 0.05 compared to untreated cells). In contrast, PSMB8-AS1 expression was not altered by TGFβ treatment (Figure 4E; control for TGFβ stimulation is given in Figure S4). All together, these results demonstrated that PSMB8-AS1 expression can be induced by pro-inflammatory stimuli relevant for SSc and possibly indicate its implication in the regulation of the downstream pathways, as previously demonstrated for other lncRNAs induced by pro-inflammatory stimuli [20].

Next, as the intracellular localization of lncRNAs partially governs their function[36], the specific cellular location of PSMB8-AS1 was determined by subcellular fractionation of healthy monocytes (Figure 4F). The percentage of enrichment of IL-8 Primary Transcript and IL-8 mRNA were used as positive controls to verify the appropriate separation of different compartments, as they are expected to be found in the nucleus/chromatin or cytoplasm, respectively. In agreement with previous findings in epithelial, glioma, and pancreatic cells [37–39], PSMB8-AS1 was found to be enriched in monocyte cytoplasm (Figure 4F, third panel), suggesting that this lncRNA may regulate expression of its target(s) at the post-transcriptional level[29].

Characterization of PSMB8-AS1 Function in Monocytes

To further investigate the biological relevance of PSMB8-AS1, its expression was efficiently silenced in resting and R848 stimulated monocytes using a specific small interfering RNA (siRNA) (Figure 5A, p < 0.001). The impact of PSMB8-AS1 silencing on monocyte apoptosis was first investigated, as this lncRNA has previously been reported as a negative regulator of apoptotic signaling[37–39], and apoptotic signaling pathways were also identified by our GO-term enrichment analysis in the cis correlation analysis. However, FACS analysis demonstrated that PSMB8-AS1 silencing in monocytes did not affect apoptosis, neither in resting nor stimulated conditions (p-value > 0.05, Figure

Figure 4. Molecular characterization of PSMB8-AS1 in healthy and SSc monocytes. PSMB8-AS1 expression in the (A) Definite cohort and (B) Non-cutaneous cohort assessed by RNA-Seq. Data are expressed as variance stabilized data (VSD), mean +/– SEM is reported. For each comparison the p-value, calculated according to the Wald test, is shown. (C) PSMB8-AS1 expression was analyzed in the "Replication cohort" by RT-qPCR. Data are reported as the fold change of each donor versus one representative healthy control, and mean +/– SEM is reported. p-values, as determined by the Kruskall–Wallis test with post-hoc Dunn's test, are reported. (D) CD14+ monocytes were cultured for the indicated time points in the presence of LPS (100 ng/mL, light blue bars), R848 (5 μM, green bars), IFNα-a2 (1000 U/mL, dark blue bars), or (E) TGFβ (0.01 ng/μL, grey bars), or left untreated (medium control, red bars). PSMB8-AS1 expression was analyzed by RT-qPCR and expressed as a fold change over the medium control at 2 h. Data are shown as mean +/– SEM of (D) 4 and (E) 2 experiments. For (D), significance is indicated as * $p < 0.05$, ** $p < 0.01$ according to two-way ANOVA followed by Dunnett's post-hoc test. (F) RNA from cytoplasm (red), nucleoplasm (blue), and chromatin (green) of CD14+ monocytes were obtained as described in Materials & Methods. Expression of IL-8 primary transcript (PT-IL8), IL-8 mRNA, and PSMB8-AS1 in each fraction was analyzed by RT-qPCR. Data are reported as a percentage of transcript in each compartment compared to total cell lysates (2^–ΔCT*100, indicated by % Enrichment). Mean +/– SEM of 3 different experiments is shown. * $p < 0.05$, ***$p < 0.001$ according to one-way ANOVA followed by Tukey's post-hoc test.

S5A,B). Moreover, PSMB8-AS1 silencing also did not affect the expression of its cis correlated PCG PSMB8 (Figure S5C), showing that this lncRNA does not function via cis regulatory mechanisms, at least in the conditions tested here.
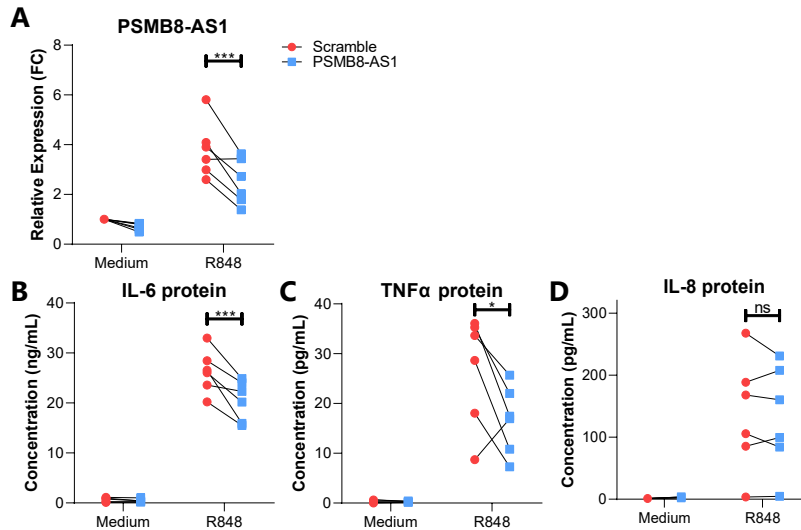
Figure 5. PSMB8-AS1 is involved in the regulation of cytokine levels in CD14+ monocytes. (A) CD14+ monocytes were transfected with si-PSMB8-AS1 (blue) or scramble siRNA (red), and stimulated with R848 for 18 h or left untreated (medium control). The expression of PSMB8-AS1 was analyzed by RT-qPCR and expressed as relative expression (FC = fold change) compared to medium scramble control. Protein levels of IL-6 (B), TNFα (C), and IL-8 (D) in the cell free supernatants of transfected monocytes were determined by ELISA. * = p-value < 0.05, *** = p-value < 0.001, ns = not significant, as determined by two-way ANOVA followed by Bonferroni's post-hoc test.

As the co-expression network analysis identified PSMB8-AS1 as a potential trans-acting hub lncRNA implicated in the regulation of immune cell activation and vesicle-related transport, its role in the secretion of pro-inflammatory cytokines was subsequently investigated. To this end, the impact of PSMB8-AS1 silencing was assessed on IL-6, IL-8, and TNFα secretion, three cytokines released by activated monocytes, and at higher levels by stimulated SSc monocytes[15, 16]. ELISA analysis showed a significant decrease of IL-6 and TNFα protein concentrations in the cell-free supernatant of si-PSMB8-AS1 monocytes stimulated with R848 for 18 h as compared to monocytes transfected with scramble siRNA (Figure 5B,C), whereas IL-8 protein levels were not affected (Figure 5D). Overall, these results demonstrated that PSMB8-AS1 regulates the secretion of specific cytokines in TLR-activated monocytes, and can thereby contribute to monocyte activation in SSc.

DISCUSSION

lncRNAs have been described as pivotal biological regulators of immune responses[19], and previous studies have suggested a role for lncRNAs in SSc[20–24]. However, a thorough characterization of the lncRNA profile in specific cell subsets obtained from SSc patients is still lacking. The aim of this study was to investigate the potential role of lncRNAs in the dysregulation of SSc monocytes, an important cell type implicated in SSc pathogenesis. Cis correlation and co-expression network analysis demonstrated that numerous lncRNAs are altered in SSc monocytes and are potentially implicated in the regulation of various biological processes relevant for SSc pathogenesis.

We have previously shown that the lncRNA NRIR plays an important role in IFN responses of SSc monocytes[20]. The results reported in the present study reinforce the concept that, besides NRIR, other lncRNAs might be strongly implicated in the regulation of IFN responses in these cells. Indeed, we identified a correlation between RP11-326I11.3, a lncRNA highly expressed in both the Definite and Non-cutaneous cohort, and IRF2, an important regulator of IFN responses[40]. Interestingly, a correlation between these two genes has previously been described in brain tissue, where they were linked to IFN signaling in central nervous system homeostasis[41]. Currently available clinical and molecular data suggest that type I IFN dysregulation is a major contributor to SSc pathogenesis[34], and the augmented expression of Type I IFN inducible genes in SSc monocytes has been linked to inflammation and the development of fibrosis[12]. Since our results showed that multiple lncRNAs are potentially involved in the amplified IFN signaling in SSc monocytes, further investigations should address the possible implication of these molecules in the disease pathogenesis.

Next to IFN signaling, cis correlated lncRNA-PCG pairs also linked this class of regulators to the negative regulation of apoptosis, a relevant process that, if inhibited, can potentially contribute to the increased numbers of circulating monocytes observed in SSc patients. RP11-356I2.4, for example, was strongly correlated with TNFAIP3, an apoptosis regulator that is induced through NF-κB signaling[42]. A correlation between RP11-356I2.4 and TNFAIP3 has also previously been observed in the inflammatory skin disorder chronic actinic dermatitis, where both genes were downregulated in comparison to healthy controls[43]. As actinic dermatitis, like SSc, is also characterized by inflamed and thickened skin, lncRNAs involved in monocyte apoptotic processes could potentially contribute to fibrotic processes in the skin.

Both the cis correlation analysis and co-expression network analysis pointed at the lncRNA PSMB8-AS1 as a key regulator of altered gene-expression in SSc monocytes. Although this lncRNA has previously been described in the context of influenza infection[44] and cancer[37–39], our results link PSMB8-AS1 dysregulation with autoimmunity for the first time. The cis correlation analysis predicted PSMB8, the protein coding gene located antisense to PSMB8-AS1, as a potential target of this lncRNA; however, the two genes were not present in the same co-expression modules, and PSMB8 mRNA levels were not altered upon PSMB8-AS1 silencing. These two genes are therefore not functionally related within SSc monocytes, while their expression correlation is probably the result of a shared transcriptional program given that their promoters share a binding site for the IFN inducible transcription factor IRF1[45]. Consistently, the protein PSMB8
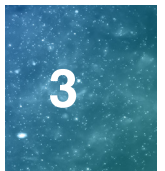
can be directly regulated by IRF1[46], and treatment of monocytes with IFNα and R848 (a TLR7/8 ligand activating IFN signaling) induces PSMB8-AS1 expression, an event likely mediated by IRF1. These findings demonstrated that IFN-mediated activation is, at least partially, responsible for the upregulation of PSMB8-AS1 in SSc monocytes, in agreement with previous studies showing that this lncRNA can be induced by influenza virus infections (triggering IFN signaling) in other human cells[44].

In line with previous observations of three other studies in pancreatic epithelial cell lines and glioma[37–39], our subcellular localization analysis demonstrated that PSMB8-AS1 is present in the cytoplasm of CD14+ monocytes. Given that cytoplasmic lncRNAs can regulate secretory and extracellular vesicles[47], and PSMB8-AS1 was present in co-expression modules annotated to vesicle related transport, it is possible that this lncRNA is involved in the regulation of cytoplasmic vesicles. A more detailed characterization by RNA-FISH of the precise subcellular compartments involved in intracellular vesicle transport (for example, the endoplasmic reticulum or Golgi)[36, 48] could provide better insights into the exact molecular processes related to PSMB8-AS1 activity. In addition, since different stimuli can modify the subcellular compartmentalization of lncRNAs[49], it would be interesting to verify whether factors relevant for SSc pathogenesis (e.g., TLR-agonists, IFNα, and CXCL4) can influence PSMB8-AS1 localization. The evidence that IL-6 and TNFα protein secretion was repressed by PSMB8-AS1 silencing in monocytes demonstrated that this lncRNA is involved in the positive regulation of these cytokines and possibly links its cytoplasmic localization to the control of cytokine secretion. In contrast, IL-8 release was unaffected, suggesting a possible cytokine-specific action of PSMB8-AS1. This is not surprising, given that specific cytokines can be secreted through distinct pathways in macrophages, adapted to suit specific stimulatory conditions[50]. Alternatively, the different impact of PSMB8-AS1 silencing on IL-6/TNFα and IL-8 secretion could be explained by distinct timeframes in the synthesis/expression of these factors or could be dependent on the specific stimuli used in our experiments. More detailed molecular investigations, exploiting a wide range of pro-inflammatory mediators, as well as stimulation times, are required to unravel the exact role of PSMB8-AS1 in cytokine release.

The positive regulation of IL-6 and TNFα secretion by PMSB8-AS1 directly link this lncRNA to SSc pathogenesis. Elevated levels of IL-6 and TNFα have indeed been reported in the serum of SSc patients in several studies[51, 52], and both cytokines are associated with fibrotic processes[53, 54], disease progression, and the occurrence of interstitial lung disease[55, 56], especially in dcSSc patients. Most importantly, different studies proposed monocytes as the potential source of these cytokines in SSc[15, 16, 57]. Interestingly, the highest upregulation of PSMB8-AS1 was observed in monocytes from dcSSc patients, where the increase of IL-6 and TNFα levels and the extent of skin fibrosis are the most severe. Even if the upregulation of PSMB8-AS1 was fluctuating in other disease subsets, possibly due to the extreme clinical heterogeneity of these patients, the upregulation of this lncRNA was also confirmed in early and ncSSc patients. Such evidence indicates that PSMB8-AS1 might be involved in the regulation of inflammatory processes from early stages of the disease onward, affecting the secretion of cytokines that eventually contribute to the development and perpetuation of fibrosis.

In conclusion, in-depth bioinformatics analysis unraveled numerous lncRNAs dysregulated in SSc monocytes and highlighted an important regulatory potential of these molecules both in immune activation and disease pathogenesis. Among these, we specifically discovered that the upregulation of PSMB8-AS1 can modulate the secretion of pro-inflammatory cytokines by monocytes, thereby potentially contributing to the increased activation of these cells in SSc. A more detailed understanding of lncRNAs and their contribution to disease pathogenesis could provide steppingstones for the identification of novel molecular targets for manipulating monocyte activity in SSc, in order to contrast disease onset and progression.

MATERIALS AND METHODS

Patient Demographics
Transcriptomic data from SSc patients and matched healthy controls from the Definite and Non-cutaneous cohorts were obtained from the gene expression omnibus (GSE124075). For the replication cohort, peripheral blood samples from SSc patients and age/sex matched healthy controls were obtained from the University Medical Center Utrecht. All participants enrolled in the study signed an informed consent form approved by the local institutional review boards prior to inclusion in this study (METC no. 12-466C, approved 2 October 2012), adherent of the Declaration of Helsinki Principles. Samples and clinical information were treated anonymously immediately after collection. SSc patients fulfilled the ACR/EULAR classification criteria[2] and were classified according to the extent of their skin fibrosis as lcSSc or dcSSc patients. Patients that fulfilled the classification criteria but did not present skin fibrosis are referred to as ncSSc patients throughout the manuscript. Finally, we also included eaSSc patients with Raynaud's Phenomenon and positivity for SSc-specific autoantibodies and/or typical nailfold capillaroscopy patterns, as defined by LeRoy *et al.*[3]. The demographics and clinical characteristics of the subjects enrolled in these cohorts are reported in Table 1. Ongoing treatment regimens are reported in Table S4.

Purification and Culture of CD14+ Monocytes from Healthy Control Blood and Buffy Coats
PBMCs were isolated from whole heparinized blood samples from SSc patients and healthy controls or from the buffy coats of healthy controls, by density gradient centrifugation using Ficoll-Paque Plus (GE Healthcare, Chicago, IL, USA). CD14+ monocytes were purified from PBMCs using the MACS Human Monocyte Isolation Kit II (Miltenyi Biotec, Bergisch Gladbach, Germany) on the autoMACs Pro Separator (Miltenyi Biotec) according to the manufacturer's instructions. For subsequent analysis, only cell preparations with more than 95% purity (measured by FACS analysis) for CD14+ cells were used.

For selected experiments, CD14+ monocytes purified from buffy coats were cultured in RPMI 1640 + 10% FCS (fetal calf serum, <0.5 EU/mL, Sigma-Aldrich, St. Louis, MO, USA) + 2 mM Glutamine at a concentration of 2 106 cells/mL. Cultured cells
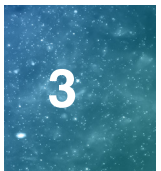
were left untreated (medium control) or treated with one of the following stimuli: 100 ng/mL ultra-pure E. coli lipopolysaccharide (LPS, strain O111:B4, Invivogen, San Diego, CA, USA), 5 μM R848 (Invivogen), 1000 U/mL IFNα-2a (Cell Sciences, Newburyport, MA, USA), and TGF-β2 (Bio-Techne, Minneapolis, MN, USA) according to the conditions and times indicated for each experiment in the results section.

RNA Purification

Total RNA was purified using the DNA/RNA/miRNA Universal kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. DNAse treatment (RNAse Free DNase I set, Qiagen) on column was performed. RNA was quantified with the Qubit® RNA Assay Kit (Life Technologies, Carlsbad, CA, USA) on the Qubit® Fluorometer (Invitrogen, Carlsbad, CA, USA) or on the Nanodrop 2000 spectrophotometer (Thermo Scientific, Waltham, MA, USA).

RNA-Sequencing Analysis

Raw sequencing data was obtained from GSE124075[17]. Sequencing reads were aligned to the GrCh38 reference human genome (Genome Reference consortium) and the H. sapiens transcriptome (Ensembl, version 77) using TopHat[58]. Summed exon read counts per gene were estimated using the HTSeq-count function provided in the HTSeq python package[59] (v. 0.6.1p1). Differential expression analysis was performed using the negative binomial distribution-based method implemented in DESeq2[60] (v. 1.6.3), and pair wise comparisons between SSc patients and HC groups were tested using the Wald test. Genes with a log2(FC) < -0.58 or > 0.58 and a p-value < 0.05 were considered significantly modulated. Normalized gene expression levels were expressed as variance stabilized data (VSD), calculated according to DESeq2 instructions. Gene types were annotated according to the Ensembl 77 database.

In Cis Correlation Analysis

Protein-coding genes (PCGs) that were localized within a region of 5 kb upstream or 5 kb downstream, regardless of the sense of transcription, for each differentially expressed lncRNA were recovered using the Biomart tool available on the Ensembl website. A Spearman rank-order correlation analysis of the expression of the lncRNAs and associated PCGs was then performed. Correlations with Spearman's Rho < -0.4 or > 0.4, and p-value < 0.05 were considered significant. Correlation analysis was performed using the *rcorr* function implemented in the Hmisc package in R[61].
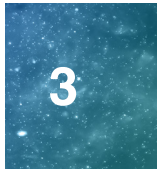
GO-Term Enrichment Analysis

Gene ontology (GO) enrichment analysis was performed using ToppFun[62]. Enrichment of biological process (BP) associated GO-terms was tested using the probability density function. p-value was adjusted according to Benjamini–Hochberg/FDR correction. BP terms significantly enriched (B&H corrected p-value ≤ 0.05) were considered.

Weighted Gene Co-Expression Network Analysis (WGCNA)

Weighted gene co-expression networks were constructed for the Definite and Non-cutaneous cohorts separately using the R package WGCNA[63]. The VSD data of all genes with at least 1 count in all samples were used as input. An unsigned network with a scale free topology was constructed, using a soft threshold power β = 13 for the Definite cohort and a soft threshold power β = 4 for the Non-cutaneous network. Modules were identified using the *cutreeDynamic* function with a minimum module size of 30. Closely related modules were merged using the *mergeCloseModules* function (cutHeight = 0.25). Gene expression profiles across the modules were summarized into module eigengene (ME) values based on the first principal component of each module. Fisher's exact test was used to calculate the extent of module overlap between the Definite and Non-cutaneous networks, as previously described[64]. Intramodular connectivity (i.e., the connectivity of a gene to genes nodes within the same module) was obtained using the *intramodularConnectivity* function from the WGCNA package.

Subcellular Fractionation

CD14+ monocytes were harvested and resuspended in cold RLN1 solution (50 mM Tris HCl pH 8.0; 140 mM NaCl; 1.5 mM MgCl2; 0.5% NP-40) supplemented with RNAse and protease inhibitors (1 U/μL RNAse Out, 5 μg/mL leupeptin, 5 μg/mL pepstatin, 20 μM PAO, 1 mM PMSF, 1 mM Na3VO4, 50 mM NaF, and 10 mM DTT) and incubated for 15 min on ice. After centrifugation at 4°C for 2 min at 200×g, the supernatant was saved as a "cytoplasmic fraction". The pellet was washed in cold RLN1 and resuspended in cold RLN2 solution (50 mM Tris HCl pH 8.0; 500 mM NaCl; 1.5 mM MgCl2; 0.5% NP-40) supplemented with RNAse and protease inhibitors (as described for RLN1) and incubated for 10 min on ice. After centrifugation at 4°C for 2 min at 500xg, the supernatant was saved as a "nuclear fraction". The remaining pellet was washed in RLN1 solution and saved as a "chromatin fraction". All fractions were resuspended in RLT+ (Qiagen) plus β-mercaptoethanol (Sigma-Aldrich, St. Louis, MO, USA) and processed for RNA extraction.

Transfection of CD14+ Monocytes Using siRNA

Purified CD14+ monocytes were transfected by means of electroporation, using the Amaxa™ Nucleofector™ II system (Lonza Basel, Switzerland) in combination with the Amaxa® Human Monocyte Nucleofector® Kit (Lonza). A minimum amount of 5x10^6, up to a maximum amount of 15x10^6 CD14+ monocytes were used for transfection, according to the manufacturer's protocol. Monocytes were transfected with 200 pmol of Silencer Select Pre-Designed siRNA (assay id n503525, PSMB8-AS1, Ambion, Austin, TX, USA) or 200 pmol of Silencer Negative control No.1 siRNA (Ambion). After transfection, the cells were plated in 50% RPMI 1640 + 10% FCS + 2 mM Glutamine, and 50% Iscove's Modified Dulbecco's Medium (IMDM, Lonza) + 10% FCS + 2 mM Glutamine, at a concentration of 3x10^6 cells/mL overnight. The next day, the medium was changed to RPMI 1640 + 10% FCS + 2 mM Glutamine, and the cells were stimulated for the times indicated for each separate experiment.

Reverse Transcription Quantitative Real-Time PCR (RT-qPCR)

Purified RNA (200–1000 ng) was reverse transcribed using the SuperScript® III Reverse Transcriptase kit (Invitrogen), according to the manufacturer's instructions. Gene expression was quantified, in duplicate, by RT-qPCR using 9 ng cDNA with SYBR Select Master Mix (Applied Biosystems, Foster City, CA, USA), in the presence of 400 nM gene- specific primers (Table 3), on the ViiA™ 7 Real-Time PCR System (Applied Biosystems). Relative expression of each gene was determined according to the comparative CT ($\Delta\Delta$CT) method using RPL32 as an endogenous control (where the ΔCT equals the CT of the mRNA of interest—the CT of RPL32)[65].

Gene	Forward Primer	Reverse Primer
PSMB8-AS1	CTTCTCTGCTCTCCCGTTATG	GTGTGTTACCTCCTTTCCAAG
RPL32	AGGGTTCGTAGAAGATTCAAGG	GGAAACATTGTGAGCGATCTC
IL-8	GCTCTGTGTGAAGGTGCAGT	CCAGACAGAGCTCTCTTCCA
PT-IL-8	ATTGAGAGTGGACCACACTG	ACTACTGTAATCCTAACACCTG
PSMB8	GAGGCGTTGTCAATATGTACC	CCTGGGGGAAATGCTTGTTC
MMP2	AGCGAGTGGATGCCGCCTTTAA	CATTCCAGGCATCTGCGATGAG

Table 3. Gene specific primer pairs used for RT-qPCR.

FACS Assessment of Monocyte Viability

Viability of CD14+ monocytes was studied by FACS analysis. At room temperature, 200,000 cells were stained for 15 min in an Annexin binding buffer (Invitrogen) using the following antibodies: CD14-PE (clone M5E2, cat. 561707, BD Biosciences, Franklin Lakes, NJ, USA), CD16-V500 (clone 3G8, cat. 561393, BD Biosciences), 7-AAD (cat. 559925, BD Biosciences), and Annexin V-APC (cat. 550474, BD Biosciences). Sample fluorescence was measured on the LSRFortessa (BD Biosciences). Data were analyzed using FlowJo (Version 10, Tree Star Inc., Ashland, OR, USA).

Assessment of Cytokine Levels Using ELISA

Concentrations of IL-6, TNFα, and IL-8 in cell-free supernatants from cultured CD14+ monocytes were measured by the sandwich enzyme linked immunosorbent assay (ELISA). IL-6 and IL-8 were quantified using the PeliKine compact human IL-6 and IL-8 ELISA kits (Sanquin Reagents, Amsterdam, The Netherlands), and TNFα was quantified using the Human TNF-α ELISA Set (Diaclone, Besançon, France), according to the manufacturer's instructions.

Statistical Analysis

Data are expressed as mean SEM unless otherwise indicated. Unless indicated otherwise, analysis of differences was performed using the Mann Whitney test. For multiple group comparisons, the one- or two-way analysis of variance (ANOVA) was used. p-values < 0.05 were considered statistically significant. Figures were produced using the R package ggplot2[66] or the GraphPad Prism software (v 8.3, www.graphpad.com, GraphPad Software, Inc., San Diego, CA, USA).

REFERENCES

1. Y. Allanore, R. Simms, O. Distler, *et al.* Systemic sclerosis. *Nat. Rev. Dis. Prim.* vol. 1, pp. 15002. Apr. 2015.

2. F. Van den Hoogen, D. Khanna, J. Fransen, *et al.* 2013 classification criteria for systemic sclerosis: An American college of rheumatology/European league against rheumatism collaborative initiative. *Ann. Rheum. Dis.* vol. 72, no. 11, pp. 1747–1755. Nov. 2013.

3. E.C. Leroy, T.A. Medsger. Criteria for the classification of early systemic sclerosis. *J. Rheumatol.* vol. 28, no. 7, pp. 1573–1576. Jul. 2001.

4. J.J. Belch. Raynaud's phenomenon: Its relevance to scleroderma. *Ann. Rheum. Dis.* vol. 50 (Suppl. 4), pp. 839–845. Nov. 1991.

5. M.B. Kahaleh. Vascular involvement in systemic sclerosis (SSc). *Clin. Exp. Rheumatol.* vol. 22, no. 3 (Suppl. 33), pp. 19–23. Jan-Feb. 2004.

6. D. Epattanaik, M. Brown, B.C. Postlethwaite, *et al.* Pathogenesis of Systemic Sclerosis. *Front. Immunol.* vol. 6, pp. 272. Jun. 2015.

7. O. Ishikawa, H. Ishikawa. Macrophage infiltration in the skin of patients with systemic sclerosis. *J. Rheumatol.* vol. 19, no. 8, pp. 1202–1206. Aug. 1992.

8. B.M. Kräling, G.G. Maul, S.A. Jimenez. Mononuclear cellular infiltrates in clinically involved skin from patients with systemic sclerosis of recent onset predominantly consist of monocytes/macrophages. *Pathobiology.* vol. 63, no. 1, pp. 48–56. 1995.

9. N. Higashi-Kuwata, M. Jinnin, T. Makino, *et al.* Characterization of monocyte/macrophage subsets in the skin and peripheral blood derived from patients with systemic sclerosis. *Arthritis Res. Ther.* vol. 12, no. 4, pp. R128. Jul. 2010.

10. M. van der Kroef, L.L. van den Hoogen, J.S. Mertens, *et al.* Cytometry by time of flight identifies distinct signatures in patients with systemic sclerosis, systemic lupus erythematosus and Sjögrens syndrome. *Eur. J. Immunol.* vol. 50, no. 1, pp. 119–129. Sep. 2019.

11. M.K.D. Scott, K. Quinn, Q. Li, *et al.* Increased monocyte count as a cellular biomarker for poor outcomes in fibrotic diseases: A retrospective, multicenter cohort study. *Lancet Respir Med.* vol. 7, no. 6, pp. 497–508. Jun. 2019.

12. Z. Brkic, L. van Bon, M. Cossu, *et al.* The interferon type I signature is present in systemic sclerosis before overt fibrosis and might contribute to its pathogenesis through high BAFF gene expression and high collagen synthesis. *Ann. Rheum. Dis.* vol. 75, no. 8, pp. 1567–1573. Sep. 2015.

13. M. Ciechomska, B. Wojtas, M. Swacha, *et al.* Global miRNA and mRNA expression profiles identify miRNA-26a-2-3p-dependent repression of IFN signature in systemic sclerosis human monocytes. *Eur. J. Immunol.* vol. 50, no. 7, pp. 1057–1066. Mar. 2020.

14. S.K. Mathai, M. Gulati, X. Peng, *et al.* Circulating monocytes from systemic sclerosis patients with interstitial lung disease show an enhanced profibrotic phenotype. *Lab. Investig.* vol. 90, no. 6, pp. 812–823. Jun. 2010.

15. T. Carvalheiro, S. Horta, J.A.G. van Roon, *et al.* Increased frequencies of circulating CXCL10-, CXCL8- and CCL4-producing monocytes and Siglec-3-expressing myeloid dendritic cells in systemic sclerosis patients. *Inflamm. Res.* vol. 67, no. 2, pp. 169–177. Feb. 2018.

16. T. Carvalheiro, A.P. Lopes, M. van der Kroef, *et al.* Angiopoietin-2 promotes inflammatory activation in monocytes of systemic sclerosis patients. *Int. J. Mol. Sci.* vol. 21, no. 24, pp. 9544. Dec. 2020.

17. M. van der Kroef, M. Castellucci, M. Mokry, *et al.* Histone modifications underlie monocyte dysregulation in patients with systemic sclerosis, underlining the treatment potential of epigenetic targeting. *Ann. Rheum. Dis.* vol. 78, no. 4, pp. 529–538. Feb. 2019.

18. Q. Liu, L.C. Zaba, A.T. Satpathy, *et al.* Chromatin accessibility landscapes of skin cells in systemic sclerosis nominate dendritic cells in disease pathogenesis. *Nat. Commun.* vol. 11, no. 1, pp. 5843. Nov. 2020.

19. M.R. Hadjicharalambous, M.A. Lindsay. Long Non-Coding RNAs and the Innate Immune Response. *Non Coding RNA.* vol 5, no.2, pp. 34. Apr. 2019.

20. B. Mariotti, N.H. Servaas, M. Rossato, *et al.* The long non-coding RNA NRIR drives IFN-response in monocytes: Implication for systemic sclerosis. *Front. Immunol.* vol. 10, pp. 100. Jan. 2019.

21. M. Dolcino, E. Tinazzi, A. Puccetti, *et al.* In systemic sclerosis, a unique long non coding RNA regulates genes and pathways involved in the three main features of the disease (vasculopathy, fibrosis and autoimmunity) and in carcinogenesis. *J. Clin. Med.* vol. 8, no. 3, pp. 320. Mar. 2019.

22. Z. Wang, M. Jinnin, K. Nakamura, *et al.* Long non-coding RNA TSIX is upregulated in scleroderma

3
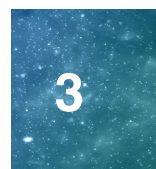
dermal fibroblasts and controls collagen mRNA stabilization. *Exp. Dermatol*. vol. 25, no. 2, pp. 131–136. Feb. 2016.

23. T.C. Messemaker, L. Chadli, G. Cai, *et al*. Antisense long non-coding RNAs are deregulated in skin tissue of patients with systemic sclerosis. *J. Investig. Dermatol*. vol. 138, no. 4, pp. 826–835. Apr. 2018.

24. M.A. Abd-Elmawla, M. Hassan, Y.A. Elsabagh, *et al*. Deregulation of long noncoding RNAs ANCR, TINCR, HOTTIP and SPRY4-IT1 in plasma of systemic sclerosis patients: SPRY4-IT1 as a novel biomarker of scleroderma and its subtypes. *Cytokine*. vol. 133, pp. 155124. May 2020.

25. I.M. Dykes, C. Emanueli. Transcriptional and post-transcriptional gene regulation by long non-coding RNA. *Genom. Proteom. Bioinform*. vol. 15, no. 3, pp. 177–186. Jun. 2017.

26. K.C. Wang, H.Y. Chang. Molecular mechanisms of long noncoding RNAs. *Mol. Cell*. vol. 43, no. 6, pp. 904–914. Sep. 2011.

27. L.L. Chen. Linking long noncoding RNA localization and function. *Trends Biochem. Sci*. vol. 41, no. 9, pp. 761–772. Sep. 2016.

28. C.-Y. Guh, Y.-H. Hsieh, H.-P. Chu. Functions and properties of nuclear lncRNAs-from systematically mapping the interactomes of lncRNAs. *J. Biomed. Sci*. vol. 27, no. 1, pp. 44. Mar. 2020.

29. J.H. Noh, K.M. Kim, W.G. McClusky, *et al*. Cytoplasmic functions of long noncoding RNAs. *Wiley Interdiscip. Rev. RNA*. vol. 9, no. 3, pp. e1471. May 2018.

30. B. Font-Cunill, L. Arnes, J. Ferrer, *et al*. Long non-coding RNAs as local regulators of pancreatic islet transcription factor genes. *Front. Genet*. vol. 9, pp. 524. Nov. 2018.

31. L. Statello, C.-J. Guo, L.-L. Chen, *et al*. Gene regulation by long non-coding RNAs and its biological functions. *Nat. Rev. Mol. Cell Biol*. vol. 22, no. 2, pp. 96–118. Feb. 2021.

32. L. Frasca, R. Lande. Toll-like receptors in mediating pathogenesis in systemic sclerosis. *Clin. Exp. Immunol*. vol. 201, no. 1, pp. 14–24. Feb. 2020.

33. M. Ciechomska, C.A. Huigens, T. Hügle, *et al*. Toll-like receptor-mediated, enhanced production of profibrotic TIMP-1 in monocytes from patients with systemic sclerosis: Role of serum factors. *Ann. Rheum. Dis*. vol. 72, no. 8, pp. 1382–1389. Dec. 2012.

34. B. Skaug, S. Assassi. Type I interferon dysregulation in Systemic Sclerosis. *Cytokine*. vol. 132, pp. 154635. Jan. 2019.

35. R. Lafyatis. Transforming growth factor β—At the centre of systemic sclerosis. *Nat. Rev. Rheumatol*. vol. 10, no.12, pp. 706–719. Dec. 2014.

36. M.N. Cabili, M.C. Dunagin, P.D. McClanahan, *et al*. Localization and abundance analysis of human lncRNAs at single-cell and single-molecule resolution. *Genome Biol*. vol.16, no. 1, pp. 20. Jan. 2015.

37. T. Hu, F. Wang, G. Han. LncRNA PSMB8-AS1 acts as ceRNA of miR-22-3p to regulate DDIT4 expression in glioblastoma. *Neurosci. Lett*. vol. 728, pp. 134896. Mar. 2020.

38. G. Shen, Y. Mao, Z. Su, *et al*. PSMB8-AS1 activated by ELK1 promotes cell proliferation in glioma via regulating miR-574-5p/RAB10. *Biomed. Pharmacother*. vol. 122, pp. 109658. Dec. 2019.

39. H. Zhang, C. Zhu, Z. He, *et al*. LncRNA PSMB8-AS1 contributes to pancreatic cancer progression via modulating miR-382-3p/STAT1/PD-L1 axis. *J. Exp. Clin. Cancer Res*. vol. 39, no. 1, pp. 179. Sep. 2020.

40. H. Harada, T. Taniguchi, N. Tanaka. The role of interferon regulatory factors in the interferon system and cell growth control. *Biochimie*. vol. 80, pp. 641–650. Aug-Sep. 1998.

41. Y. Zhou, P.-E. Lutz, Y.C. Wang, *et al*. Global long non-coding RNA expression in the rostral anterior cingulate cortex of depressed suicides. *Transl. Psychiatry*. vol. 8, no. 1, pp. 224. Oct. 2018.

42. K. Honma, S. Tsuzuki, M. Nakagawa, *et al*. TNFAIP3/A20 functions as a novel tumor suppressor gene in several subtypes of non-Hodgkin lymphomas. *Blood*. vol. 114, no. 12, pp. 2467–2475. Jul. 2009.

43. D. Lei, L. Lv, Li. Yang, *et al*. Genome-wide analysis of mRNA and long noncoding RNA profiles in chronic actinic dermatitis. *BioMed Res. Int*. vol. 2017, pp. 7479523. Nov. 2017.

44. S. More, Z. Zhu, K. Lin, *et al*. Long non-coding RNA PSMB8-AS1 regulates influenza virus replication. *RNA Biol*. vol. 16, no. 3, pp. 340–353. Jan. 2019.

45. L. Shi, J.C. Perin, J. Leipzig, *et al*. Genome-wide analysis of interferon regulatory factor I binding in primary human monocytes. *Gene*. vol. 487, no. 1, pp. 21–28. Jul. 2011.

46. M.A. El Hassan, K. Huang, M.B.K. Eswara, *et al*. Properties of STAT1 and IRF1 enhancers and the influence of SNPs. *BMC Mol. Biol*. vol. 18, no. 1, pp. 6. Mar. 2017.

47. M. Aillaud, L.N. Schulte. Emerging roles of long noncoding RNAs in the cytoplasmic milieu. *Non Coding RNA*. vol. 6, no. 4, pp. 44. Nov. 2020.

48. T. Santini, J. Martone, M. Ballarino. Visualization of nuclear and cytoplasmic long noncoding RNAs at single-cell level by RNA-FISH. *Methods Mol Biol*. vol. 2157, pp. 251–280. 2021.

49. P. Zhang, L. Cao, R. Zhou, *et al*. The lncRNA Neat1 promotes activation of inflammasomes in macrophages. *Nat. Commun*. vol. 10, no. 1, pp. 1495. Apr. 2019.

50. R. Z. Murray, J.L. Stow. Cytokine secretion in macrophages: SNAREs, rabs, and membrane trafficking. *Front. Immunol*. vol. 5, pp. 538. Oct. 2014.

51. E. Scala, S. Pallotta, A. Frezzolini, *et al*. Cytokine and chemokine levels in systemic sclerosis: Relationship with cutaneous and internal organ involvement. *Clin. Exp. Immunol*. vol. 138, no. 3, 540–546. Dec. 2004.

52. P. Gourh, F.C. Arnett, S. Assassi, *et al*. Plasma cytokine profiles in systemic sclerosis: Associations with autoantibody subsets and clinical manifestations. *Arthritis Res. Ther*. vol. 11, no. 5, pp. R147. Oct. 2009.

53. S. Sato, M. Hasegawa, K. Takehra, *et al*. Serum levels of interleukin-6 and interleukin-10 correlate with total skin thickness score in patients with systemic sclerosis. *J. Dermatol. Sci*. vol. 27, no. 2, pp. 140–146. Oct. 2001.

54. M. Hasegawa, M. Fujimoto, K. Kikuchi, *et al*. Elevated serum tumor necrosis factor-α levels in patients with systemic sclerosis: Association with pulmonary fibrosis. *J. Rheumatol*. vol. 24, no. 4, pp. 663–665. Apr. 1997.

55. A. De Lauretis, P. Sestini, P. Pantelidis, *et al*. Serum interleukin 6 is predictive of early functional decline and mortality in interstitial lung disease associated with systemic sclerosis. *J. Rheumatol*. vol. 40, no. 4, pp. 435–446. Apr. 2013.

56. K. Schmidt, L. Martinez-Gamboa, S. Meier, *et al*. Bronchoalveoloar lavage fluid cytokines and chemokines as markers and predictors for the outcome of interstitial lung disease in systemic sclerosis patients. *Arthritis Res. Ther*. vol. 11, no. 4, pp. R111. Jul. 2009.

57. B. Crestani, N. Seta, M. de Brandt, *et al*. Interleukin 6 secretion by monocytes and alveolar macrophages in systemic sclerosis with lung involvement. *Am. J. Respir. Crit. Care Med*. vol. 149, no. 5, pp. 1260–1265. May 1994.

58. C. Trapnell, L. Pachter, S.L. Salzberg. TopHat: Discovering splice junctions with RNA-Seq. *Bioinformatics*. vol. 25, no. 9, pp. 1105–1111. Mar. 2009.

59. S. Anders, P.T. Pyl, W. Huber. HTSeq—A Python framework to work with high-throughput sequencing data. *Bioinformatics*. vol. 31, no. 2, pp. 166–169. Sep. 2014.

60. M.I. Love, W. Huber, S. Anders. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. vol. 15, no. 12, pp. 550. 2014.

61. F.E. Harrell Jr., M.C. Dupont. The Hmisc Package; R Package Version 4.4.-0; *R Team: Vienna, Austria*, 2006.

62. J. Chen, E.E. Bardes, B.J. Aronow, *et al*. ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. *Nucleic Acids Res*. vol. 37, pp. W305–W311. May 2009.

63. P. Langfelder, S. Horvath. WGCNA: An R package for weighted correlation network analysis. *BMC Bioinform*. vol. 9, pp. 559. Dec. 2008.

64. P. Langfelder, R. Luo, M.C. Oldham, *et al*. Is my network module preserved and reproducible? *PLoS Comput. Biol*. vol. 7, no. 1, pp. e1001057. Jan. 2011.

65. T.D. Schmittgen, K.J. Livak. Analyzing real-time PCR data by the comparative C(T) method. *Nat. Protoc*. vol. 3, no. 6, pp. 1101–1108. 2008.

66. H. Wickham. ggplot2: Elegant Graphics for Data Analysis; *Springer: New York, NY, USA,* 2016.
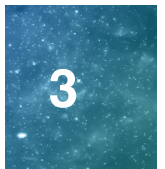
3

SUPPLEMENTARY INFORMATION

The Supplementary Tables for this article can be found online at:
https://www.mdpi.com/article/10.3390/ijms22094365/s1
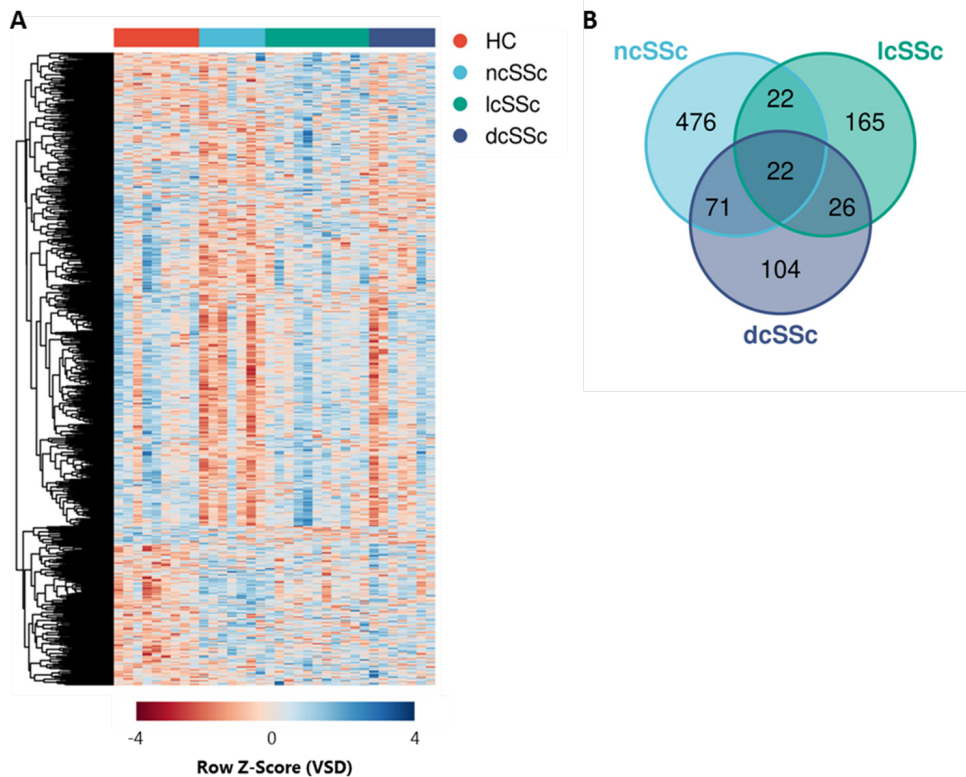
Figure S1. Expression of lncRNAs is altered in SSc monocytes. (A) Expression heatmap of lncRNAs significantly modulated in at least one comparison of SSc patients versus healthy controls (log2(FC) >0.58 or <-0.58, and p-value <0.05). lncRNAs expression is shown as row mean-centered Z-Score of the variance stabilized data (VSD) obtained from RNA-sequencing analysis. (B) Venn diagram showing the number of differentially expressed lncRNAs in each SSc patient subset (depicted by different colors) as well as the number of overlapping differential lncRNAs between patient subsets.
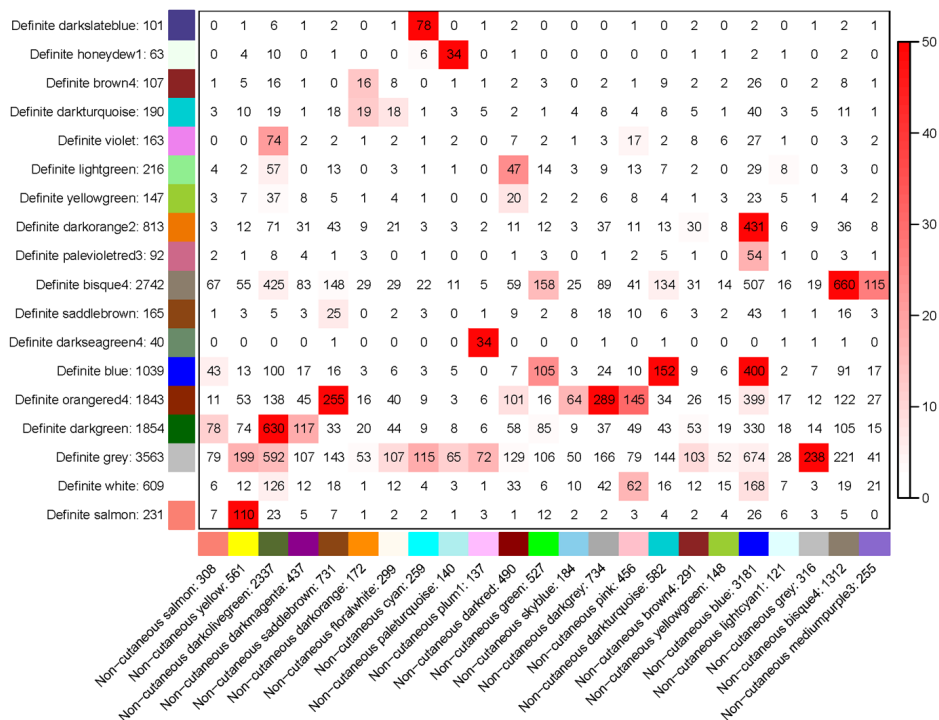
Figure S2. Module overlap between gene co-expression networks identified in the definite and non-cutaneous cohorts. Overlap of all modules from the definite (rows) and non-cutaneous cohorts (columns). Numbers behind modules names indicate the total number of genes in the modules. Numbers in the table indicate the number of genes overlapping between two modules. Coloring indicates the significance of the overlap (Fisher's exact test, – log10(p-value)).
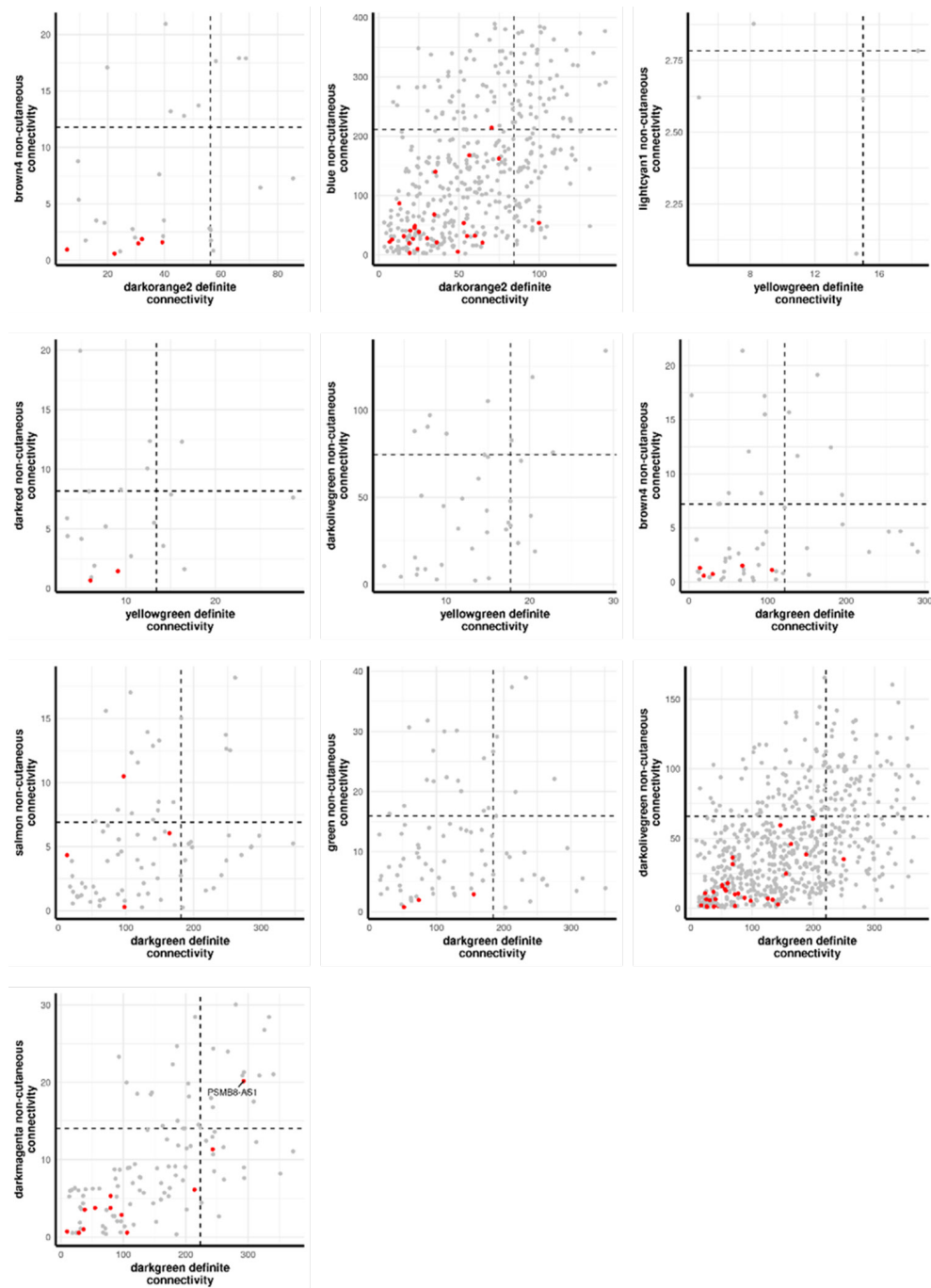
Figure S3. Connectivity overlap of selected reproduced co-expression modules. Intramodular connectivity of genes shared across the selected modules of the definite cohort (x-axis) and corresponding modules of the non-cutaneous cohort (y-axis). Each dot represents one gene, with lncRNAs highlighted in red. The top 25% most connected genes in both the definite and non-cutaneous modules (black dotted lines) were considered as hub-genes.
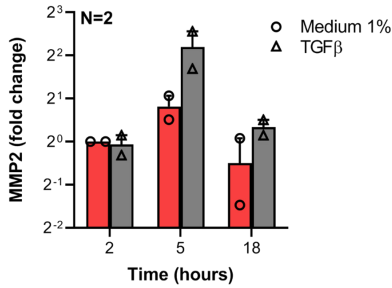
Figure S4. MMP2 expression is induced by TGFβ signaling in monocytes. CD14+ monocytes were cultured for the indicated time points in presence TGFβ (grey bars), or left untreated (medium control, red bars). PSMB8-AS1 expression was analyzed and expressed as fold change over the medium control at 2h. Data are shown as mean+/- SEM.
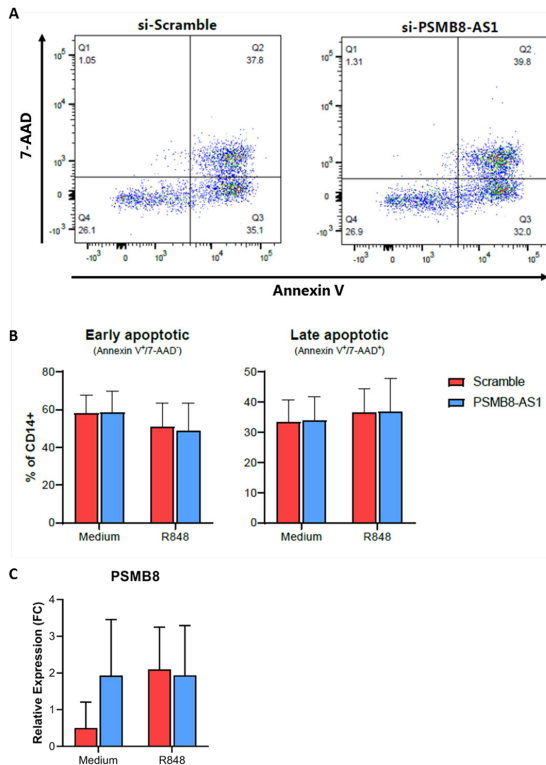
Figure S5. Silencing of PSMB8-AS1 does not affect apoptosis or PSMB8 expression in healthy monocytes. (A) Representative FACS plots of viability analysis of monocytes transfected with si-PSMB8-AS1 (right panel) or scramble siRNA (left panel). The percentage of viable (Q4), early apoptotic (Q3) and late apoptotic (Q2) cells, based on Annexin V (x-axis) and 7-AAD (y-axis) staining is indicated in the corresponding quadrants. (B) Bar graphs showing the percentage of early apoptotic (left panel) and late apoptotic (right panel) within the fraction of CD14+ monocytes. Cells were transfected with si-PSMB8-AS1 (blue) or scramble siRNA (red) , and either left unstimulated or treated with R848 (N=5). (C) Expression of PSMB8 was analyzed by RT-qPCR and expressed as relative expression (y-axis, FC=fold change) compared to medium scramble control (N=3).

Chapter 4

Epigenomic imprinting underlies disease specific dysregulation of monocytes in systemic sclerosis, systemic lupus erythematosus and rheumatoid arthritis
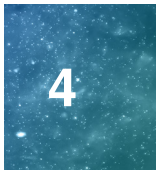
Authors:
N.H. Servaas[1]*, K. Woolcock[3]*, M. van der Kroef[1,2], J. Cole[3], L.L. van den Hoogen[1,2], T. Carvalheiro[1,2], E. Chouri[1,2], C.G.K. Wichers[1,2], C.P.J. Bekker[1,2], T.R.D.J. Radstake[1,2], C. Goodyear[3#], A. Pandit[1,2#]

*/# These authors contributed equally

Affiliations:
1. Laboratory of Translational Immunology, University Medical Center Utrecht, Utrecht, Netherlands.
2. Department of Rheumatology and Clinical Immunology, University Medical Center Utrecht, Utrecht, Netherlands.
3. Centre for Immunobiology, Institute of Infection, Immunity and Inflammation, 120 University Place, University of Glasgow, Glasgow, United Kingdom.

ABSTRACT

Background and objective Systemic sclerosis (SSc), Systemic Lupus Erythematosus (SLE) and Rheumatoid arthritis (RA) are rheumatic autoimmune diseases for which monocytes have been described to play a major role in disease pathogenesis. Given contribution of epigenomic alterations in the immunopathology of these diseases, studying the epigenomic landscape of SLE, RA and SSc monocytes should provide valuable insights into their roles in disease pathogenesis.

Methods Paired genome-wide transcriptomic profiling and chromatin immunoprecipitation followed by sequencing (ChIP-seq) for histone marks H3K4me3 and H3K27me3 was performed on monocytes of fifteen healthy controls and sixty patients with SSc, SLE or RA. Differential expression analysis and correlation of RNA expression with the levels of H3K4me3 and H3K27me3 located nearby their transcription start sites were performed to assess the impact of histone modifications on gene expression regulation.

Results We identified a total of 12,959 differentially expressed genes in SSc, SLE and RA patients, which were especially enriched for genes related to interferon (IFN), tumor necrosis factor (TNF), transforming growth factor beta (TGFβ) and collagen formation pathways. We identified an evident disease specific dysregulation of these signaling pathways, as SLE and SSc transcriptomes were marked by a strong dysregulation of IFN and TNFα signaling, while RA monocytes lacked the IFN signature and were highly enriched for TNF, TGFβ and collagen formation pathways. This was already imprinted at the histone level, showing that aberrances in histone marks selectively skew autoimmune monocytes towards distinct pro-inflammatory phenotypes. Moreover, we identified numerous poised promoters, characterized by the simultaneous presence of activating H3K4me3 and repressing H3K27me3, which were potentially involved in disease relevant signaling pathways and primed monocytes for enhanced activation in SSc and RA.

Conclusions Alterations in the epigenomic landscape of monocytes from SLE, SSc and RA patients correlate with the expression of distinct disease relevant signaling pathways. Our results offer new insights into epigenomic mechanisms underlying monocyte dysregulation in these autoimmune diseases, offering new avenues for future therapeutic targeting.

INTRODUCTION

Monocytes are circulating white blood cells that play critical roles in inflammation. They can initiate and contribute to local and systemic inflammation through the secretion of pro-inflammatory mediators[1], and are the precursors of myeloid lineage dendritic cells, macrophages, and osteoclasts[2, 3]. Monocytes are well known for their multipotency, and their capability to differentiate into distinct cell types is largely dependent on environmental cues and pro-inflammatory triggers[4]. These abilities are central to monocyte function and make them crucial cells in driving immune-related disorders.

Monocytes are described to be dysregulated in multiple autoimmune diseases including systemic sclerosis (SSc), systemic lupus erythematosus (SLE), and rheumatoid arthritis (RA), and are proposed to have an essential role in immunopathology. In SSc and RA, the number of circulating monocytes are higher compared to healthy individuals[5–7], and an increased infiltration of monocytes into affected tissues has been observed. In SSc skin lesions, monocytes are among the predominant infiltrating cells[5], while in RA large numbers of infiltrating monocytes and macrophages are observed in synovial tissues lining affected joints[8, 9]. Likewise, monocytes have been found to infiltrate kidneys of active lupus nephritis patients, where their frequencies are associated with disease activity[10]. After infiltration into affected tissues, monocytes can differentiate into pro-inflammatory macrophages[11], as well as specific disease relevant cell subsets, including bone-resorbing osteoclasts in RA[12], and extra-cellular matrix (ECM) producing myofibroblasts in SSc[13]. Moreover, the production of selective pro-fibrotic and pro-inflammatory cytokines by monocytes in the affected tissues contribute to the local and systemic inflammation observed in SLE, SSc and RA patients. Altogether, monocytes play a key role in the pathogenesis of the three autoimmune diseases. However, hitherto no study has been performed that compares the phenotypic and functional characteristics of monocytes in SLE, SSc and RA to study the commonalities and differences in mechanisms driving these diseases.

Environmental factors altering epigenetic factors such as histone modifications[14], non-coding RNA[15] and DNA methylation[16] play an important role in the development of SLE, SSc and RA. Thus, studying the epigenomic landscape of monocytes obtained from these diseases could provide valuable insights into their roles in disease pathogenesis. Previous research has already shown that post-transcriptional modifications of histones play a crucial role in monocyte differentiation into dendritic cells[17] and macrophages[18]. Moreover, changes in the levels of trimethylation of lysine 4 of histone 3 (H3K4me3) and acetylation on lysine 27 of histone 3 (H3K27ac) partially underlie the aberrant gene expression profiles observed in SSc monocytes[19]. Further exploration of the histone modification patterns affecting gene expression in SLE, SSc and RA monocytes can help to better characterize the mechanisms underlying monocyte dysregulation and identify potential novel therapeutic targets.

Here, we integrated paired transcriptomic and epigenomic data to evaluate how epigenetic imprinting affects the phenotype of circulating monocytes isolated from patients with SSc, SLE and RA. We used chromatin immunoprecipitation followed by high-throughput sequencing (ChIP-seq) for the activating histone mark H3K4me3 and

4

the repressing histone mark H3K27me3 in parallel with transcriptome analysis to show that distinct epigenetic imprinting results in the dysregulation of disease specific signaling pathways in SLE, SSc and RA monocytes.

MATERIALS AND METHODS

Study participants

Peripheral blood was collected from SSc, SLE and RA patients as well as age- and sex-matched healthy controls (HC). Informed consent was obtained from all patients and healthy donors enrolled in the study at the University Medical Center Utrecht, the Maasstad Medical Center Rotterdam and the Gartnavel Hospital Glasgow. All samples and clinical information were treated anonymously right after collection. All participants enrolled in the study signed a consent form approved by the local institutional review boards, in adherence to the Declaration of Helsinki Principles. All patients fulfilled their respective classification criteria (ACR/EULAR 2013 criteria for SSc[20], 1997 ACR criteria for SLE[21], and 2010 ACR criteria for RA[22]) (Table 1). SSc patients were subdivided into limited cutaneous SSc (lcSSc) and diffuse cutaneous (dcSSc) subsets based on the extent of skin fibrosis. Early SSc (eaSSc) patients presented with Raynaud's phenomenon (RP) in combination with either typical nailfold videocapillaroscopy (NVC) abnormalities or SSc-specific autoantibodies as defined by LeRoy *et al.*[23].

Monocyte isolation

Peripheral blood mononuclear cells (PBMCs) were isolated from whole blood samples collected in lithium heparin tubes using Ficoll Paque Plus (GE Healthcare). The CD14+ monocytes population was isolated from PBMCs using the MACS Human Monocyte Isolation Kit II (Miltenyi Biotech, Germany), according to the manufacturer's instructions, after which a purity of >95% (measured by FACS) was generally observed. Freshly isolated monocytes were immediately stored for RNA extraction and ChIP analysis performed later.

RNA extraction and RNA sequencing (RNA-seq)

Total RNA was isolated from monocytes using the AllPrep DNA/RNA/miRNA Universal Kit (Qiagen) according to the manufacturer's instructions. The quantity and quality of the obtained RNA was assessed using an Agilent 2100 bioanalyzer prior to sequencing. All samples used in the analysis had an RQN score of >8 with an average of 9.8. Library prep was performed by GenomeScan (Leiden, the Netherlands) using the NEBNext Ultra II Directional RNA Library Prep Kit (Illumina), according to the manufacturer's instructions. The quality and yield after sample preparation was measured with the Fragment Analyzer. Clustering and DNA sequencing using the NovaSeq6000 (Illumina) was performed according to manufacturer's protocols, generating 20-30 million 150bp paired ended reads for each sample.

RNAseq	HC	earlySSc	lcSSc	dcSSc	SLE	RA
ChIPseq H3K4me3	**HC**	**earlySSc**	**lcSSc**	**dcSSc**	**SLE**	**RA**
ChIPseq H3K27me3	**HC**	**earlySSc**	**lcSSc**	**dcSSc**	**SLE**	**RA**
N	15	10	10	10	10	9
	15	9	9	9	8	10
	14	9	9	9	7	10
Age (yr.)	54 (50-62)	45 (37-50)	57 (49-64)	53 (45-62)	45 (40-54)	68 (63-73
	54 (50-62)	45 (37-52)	57 (48-65)	51 (43-59)	44 (33-52)	66 (61-73)
	54 (50-60)	44 (37-52)	57 (48-65)	51 (43-59)	43 (30-49)	66 (61-73)
Female (n, %)	11 (73%)	8 (80%)	8 (80%)	6 (60%)	9 (90%)	7 (78%)
	11 (73%)	7 (78%)	7 (78%)	5 (56%)	8 (100%)	8 (80%)
	11 (78%)	7 (78%)	7 (78%)	5 (56%)	7 (100%)	8 (80%)
ANA (n pos, %)	-	8 (80%)	8 (80%)	9 (90%)	10 (100%)	-
	-	7 (78%)	7 (78%)	8 (89%)	8 (100%)	-
	-	7 (78%)	7 (78%)	8 (89%)	7 (100%)	-
ACA (n pos, %)	-	3 (30%)	4 (40%)	2 (20%)	-	-
	-	3 (33%)	4 (44%)	2 (22%)	-	-
	-	3 (33%)	4 (44%)	2 (22%)	-	-
Scl70 (n pos, %)	-	1 (10%)	0 (0%)	6 (60%)	-	-
	-	1 (11%)	0 (0%)	5 (56%)	-	-
	-	1 (11%)	0 (0%)	5 (56%)	-	-
mRSS	-	-	6.9 (2-12.5)	14.3 (10.5-21.5)	-	-
	-	-	7.5 (2-12.75)	14.7 (7-23)	-	-
	-	-	7.5 (2-12.75)	14.7 (7-23)	-	-
ILD n (%)	-	0 (0%)	1 (10%)	3 (30%)	-	-
	-	0 (0%)	1 (11%)	2 (22%)	-	-
	-	0 (0%)	1 (11%)	2 (22%)	-	-
SLEDAI	-	-	-	-	4 (0-7)	-
	-	-	-	-	4 (1-6)	-
	-	-	-	-	4 (0-6)	-
	-	-	-	-	-	4.2 (3.2-4.9)
DAS28	-	-	-	-	-	4.3 (3.5-5.2)
	-	-	-	-	-	4.3 (3.5-5.2)
PDN (n, %)	-	0 (0%)	1 (10%)	2 (20%)	3 (30%)	0 (0%)
	-	0 (0%)	0 (0%)	2 (22%)	2 (25%)	0 (0%)
	-	0 (0%)	0 (0%)	2 (22%)	2 (29%)	0 (0%)
MTX (n, %)	-	1 (10%)	1 (10%)	4 (40%)	0 (0%)	6 (67%)
	-	1 (11%)	1 (11%)	3 (33%)	0 (0%)	7 (70%)
	-	1 (11%)	1 (11%)	3 (33%)	0 (0%)	7 (70%)
CP (n, %)	-	0 (0%)	0 (0%)	1 (10%)	5 (50%)	0 (0%)
	-	0 (0%)	0 (0%)	1 (11%)	3 (38%)	0 (0%)
	-	0 (0%)	0 (0%)	1 (11%)	3 (43%)	0 (0%)
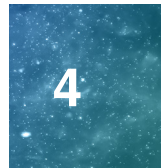
Table 1. Demographics and clinical characteristics of patients in the study. Values reported indicate the number (n) of patients and the mean for each parameter (Interquartile Range (IQR)), if not otherwise indicated. White and light and dark grey fields indicate features of patients of the same cohort *(continued)*

analysed either with RNAseq or ChIPseq for H3K4me3 and H3K27me3, respectively. Yr., years; ANA, antinuclear antibodies; ACA, anticentromere antibodies; Scl70, antitopoisomerase antibodies; mRSS, modified Rodnan Skin score; ILD, Interstitial Lung disease; HC, healthy controls; lcSSc, limited cutaneous SSc; dcSSc, diffuse cutaneous SSc; SLEDAI, Systemic Lupus Erythematosus Disease Activity Index; DAS28, Disease Activity Score-28; PDN, prednisone; MTX, methotrexate; CP, cyclophosphamide.

Chromatin immunoprecipitation and sequencing (ChIP-seq)

Freshly isolated monocytes for ChIP-seq were fixed using 1% formaldehyde (Sigma-Aldrich, Saint Louis, MO, USA) for 10 minutes at room temperature, followed by the addition of 0.125M Tris (pH 7.6) and subsequent incubation for 5 minutes at room temperature to stop the crosslinking reaction. After centrifugation at 500G the cell pellet was washed four times with cold PBS. Samples were stored at -80°C for use at a later time point. Cell pellets were lysed in dH2O containing 1M NaBut, 0.5M HEPES and 10% SDS supplemented with protease inhibitors (ThermoFisher Scientific cat#1862209). Sonication of the samples was performed using the Bioruptor Pico sonication device (Diagenode), in Bioruptor 0.65ml microtubes (Diagenode) to obtain chromatin fragments sized between 200-600bp. Samples were diluted in dH2O with 0.1% SDS, 1% Triton x-100 with a final concentration of 1.2mM EDTA, 16.7mM Tris pH 8 and 167 mM NaCl. The following antibodies were used in combination with Protein A dynabeads (Invitrogen) to perform the immunoprecipitation; H3K4me3 (Merck Millipore) and H3K27me3 (Merck Millipore). Nuclear extracts from 1 to 5 million monocytes were immunoprecipitated with 1.5 µg of antibody incubated with the protein A beads for 1 hour at room temperature. Antibody bound beads were added to the diluted chromatin fragments and incubated overnight at 4°C. De-crosslinking of the samples was performed by placing them into a thermomixer for 1h at 55°C followed by overnight at 65°C. DNA fragments were isolated using the MinElute PCR Purification Kit (Qiagen) following manufacturer's instructions. Single ended sequencing of the fragments was performed by Glasgow Polyomics (Glasgow UK) using an Illumina NextSeq500 sequencer according to the manufacturer's instructions generating ~30 million 75bp single-end reads for each sample.

Statistical analysis

RNA-sequencing analysis

Quality check of the raw sequences was performed using the FastQC tool. Sequencing reads were aligned using STAR aligner[24], using annotations from the GrCh38 (v79) built from the human genome (http://www.ensembl.org). On average 25 million uniquely mapped reads were obtained per sample. Summed exon read counts per gene were estimated using the HTSeq-count function provided in the HTSeq python package[25]. Between lane normalization using the upper quartile normalization was performed using the EDAseq Bioconductor/R package[26]. Differential expression analysis was performed using the negative binomial distribution-based method implemented in DESeq2[27], and pair wise comparisons between patients groups and healthy controls were tested using the Wald test. Genes with a nominal p-value < 0.05 as compared to healthy controls were considered differentially expressed. Variance

stabilizing transformation (VST) was calculated according to DESeq2 instructions to obtain the normalized read counts (NRC) for further analysis.

ChIP-sequencing analysis

Quality check of the raw sequences was performed using the FastQC tool. Sequencing reads were mapped against the reference genome GRCh38.p13 built from the human genome (NCBI) using bowtie2[28]. Peaks were called using MACS2[29], by comparing the IP samples to their matched input samples. Broad-peak mode was used to call peaks for both H3K4me3 and H3K27me3 (-g hs --broad --broad-cutoff 0.1). After calling, the obtained peaks were filtered based on q-value to only retain the high confidence peak regions. Additionally, peaks from ENCODE blacklist regions[30], and X- and Y-chromosome peaks were filtered out to reduce noise and exclude sex-specific peaks. Peaks were annotated to the nearest genes using the ChIPseeker Bioconductor/R package[31]. Peaks were associated with a gene when they were annotated within a 10kb range (up- or downstream) to the gene transcription start site (TSS).

For macrophage and osteoclast datasets, enriched regions peak regions (bed files) were retrieved from the Blueprint Consortium data access portal (http://dcc. blueprint-epigenome.eu/#/datasets)[32]. H3K4me3 and H3K27me3 peak sets were taken for two biological replicates from human macrophages (EGAD00001002504, donors S0022I and S00390), and osteoclasts (EGAD00001002391, donors BC2_0 and BC2_10). Overlapping peaks for biological replicates were identified using the *findOverlapsOfPeaks* function, implemented in the ChIPpeakAnno Bioconductor/R package[33]. Peaks overlapping between two biological replicates where considered for downstream analysis, and were further processed in the same way as the ChIP monocyte data.

Gene set enrichment and pathway analysis

Pre-ranked gene set enrichment analysis (GSEA, ranking based on log2 fold change) was performed using the fgsea Bioconductor/R package[34]. Gene sets were obtained from the molecular signatures database (MSigDB, https://www.gsea-msigdb. org/)[35]. The following gene sets were used: TNFα, "HALLMARK_TNFA_SIGNALING_ VIA_NFKB"; IFN, "REACTOME_INTERFERON_SIGNALING"; TGFβ, "HALLMARK_ TGF_BETA_SIGNALING"; collagen, "REACTOME_COLLAGEN_FORMATION". Enrichments with adjusted p-value <0.05 were considered significant. GO-term enrichment analysis was performed using *enrichGo* function from the clusterProfiler R package[36]. Biological processes with a B&H correct p-value <0.05 were considered significant.

RNA-ChIP correlation analysis

Spearman's rank correlation coefficient was calculated to assess correlations between the levels of H3K4me3 and H3K27me3 present within 10KB of a transcription start site and the expression level of the corresponding genes. Correlations were plotted using the R package ggplot2[37], and regression lines were added using the linear model implemented in the *geom_smooth* function from ggplot2.
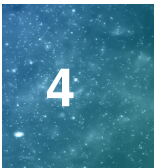
RESULTS

SLE, SSc and RA monocyte transcriptomes are characterized by disease specific patterns in pro-inflammatory and pro-fibrotic signaling pathways

To study the gene expression alterations in monocytes in the context of autoimmune disease, we performed bulk RNA-sequencing (RNA-seq) of CD14+ monocytes isolated from peripheral blood samples obtained from SLE (n=10), SSc (n=30) and RA (n=10) patients and matched healthy controls (n=15). We identified a total of 12,959 significantly differentially expressed genes (DEGs, nominal p-value <0.05) between all groups of autoimmune patients compared to healthy controls, demonstrating that the transcriptome of autoimmune monocytes is markedly different from healthy. This was confirmed by principal component analysis (PCA), which showed a clear separation between patients and HC (Figure 1A). In addition, we observed a strong separation between patients with SSc, SLE and especially RA patients. Overlap analysis showed that only a small fraction of the DEGs were shared between all disease groups (Figure 1B), of which the majority were related to interferon (IFN) signaling (Figure 1C). Next, we performed gene set enrichment analysis (GSEA) to identify potential signaling pathways that could segregate monocyte transcriptomes of SLE, SSc and RA patients (Figure 1D). While SLE monocytes displayed a strong enrichment of genes implicated in both IFN and tumor necrosis factor alpha (TNFα) signaling pathways, RA monocytes lacked an enrichment for IFN genes, while showing a strong enrichment for TNFα signaling (Figure 1D/E). dcSSc monocytes were enriched for both IFN and TNFα signaling genes, although to a lesser extent than SLE and RA monocytes, with the signature being weaker in lcSSc and eaSSc (in line with the decreased disease severity in these patient groups). Notably, only RA monocytes were significantly enriched in gene signatures related to TGFβ signaling and collagen formation, although this signature could also partially be observed in dcSSc monocytes, albeit not significant (Figure 1D/E). In line with the observation that SLE monocytes were strongly characterized by an increased expression of genes related to IFN signaling, SLE monocytes also displayed the strongest type I IFN score (Figure 1F). This also matches previous observations in literature, where SLE patients are marked by a much stronger type I IFN gene signature compared to SSc and RA[38]. Collectively, these results demonstrate that, based on their transcriptome, SLE, SSc and RA monocytes display disease specific phenotypes, marked by differences in pro-inflammatory (IFN/TNFα), and pro-fibrotic (TGFβ /collagen formation) signaling pathways.

Histone modification levels are correlated to gene expression levels in monocytes

To gain insights into the epigenomic histone landscape of patients with autoimmune diseases, and identify how such changes relate to gene expression, we performed genome-wide screening of the level of H3K4me3 and H3K27me3 histone modifications using ChIP-sequencing of monocytes from almost all individuals included in the transcriptomic cohort described above (Table 1). We identified a total of 24,187 peaks for H3K4me3 and 14,831 peaks for H3K27me3 marks (significantly enriched over negative control input samples). In line with previous observations[40], H3K4me3
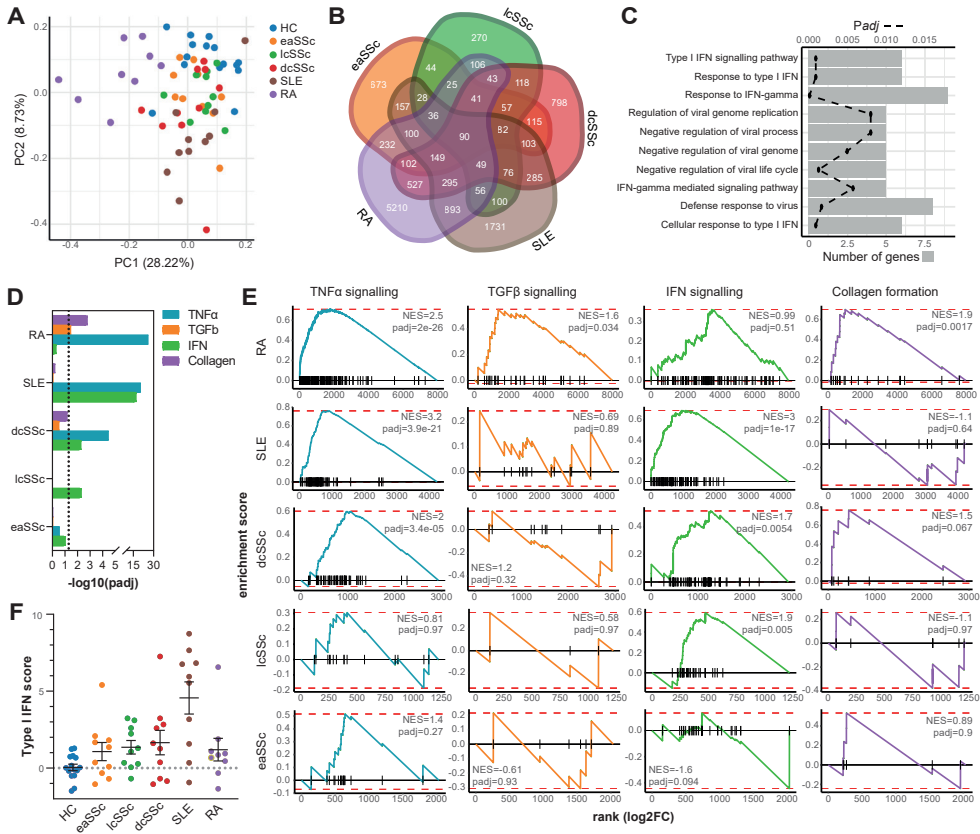
Figure 1. Disease specific patterns in pro-inflammatory and pro-fibrotic signaling pathways can be identified in SLE, SSc and RA monocytes. (A) Principal component analysis of the 12,959 differentially expressed genes identified in SLE, SSc and RA monocytes versus healthy controls (HC). (B) Venn diagram depicting the unique and overlapping differentially expressed genes identified in SLE, SSc and RA monocytes. (C) GO-term enrichment results of the 90 differentially expressed genes overlapping between all disease subsets. Bar graph represents the number of enriched genes, while dots/dashed line indicate the corresponding B&H corrected p-value. (D) Results from gene set enrichment analysis for the ranked differentially expressed genes from SLE, SSc and RA monocytes against hallmark genes signatures from TNFα, TGFβ, IFN and collagen formation pathways. Adjusted p-values for each geneset are indicated, with the dashed line representing p-value = 0.05. (E) Gene set enrichment plots of gene sets indicated in D. NES = normalized enrichment score. For every disease category, differentially expressed genes were ranked based on log2 fold change compared to healthy controls. (F) Type I interferon (IFN) gene signature scores, calculated as previously described[39]. Error bars indicate the mean and standard error of the mean for each distribution of IFN scores.

exhibited a typical profile with the majority of peaks being tightly localized around the transcription start site (TSS), while H3K27me3 profiles exhibited broader peaks around the TSS (Figure 2A). The majority of H3K4me3 as well as H3K27me3 peaks were detected within gene promoter regions (Figure 2A). In line with the previously described activating effects of H3K4me3 and repressing effects of H3K27me3 deposition[41], a positive correlation between gene expression levels and the presence of H3K4me3 near the TSS was commonly observed, while inverse correlations were observed for gene expression
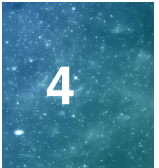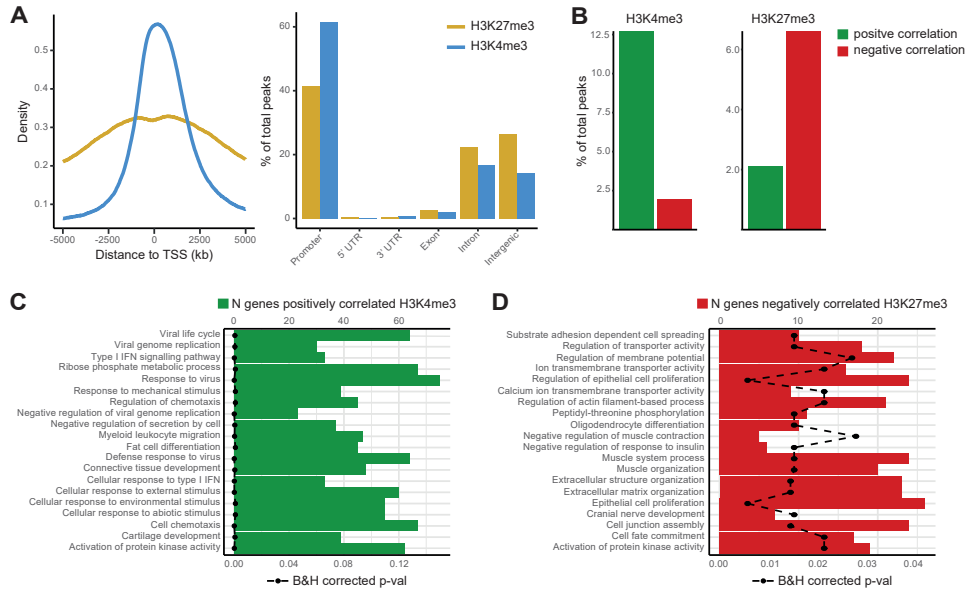
Figure 2. H3K4me3 and H3K27me3 histone modifications are involved in gene expression regulation of distinct functional pathways in monocytes. (A) Left: density profile of H3K4me3 and H3K27me3 peaks within a 5kb range around transcription start site (TSS). Right: percentage of peaks overlapping six categories of genomic locations, annotated using ChIPseeker. UTR = untranslated region. (B) Percentage of peaks that were significantly correlated (Spearman's rank correlation <0.05) in a positive or negative fashion to the level of H3K4me3 (left) or H3K27me3 (right) present near their TSS (<10kb). (C) GO-term enrichment results of the genes positively correlated to H3K4me3 or (D) negatively correlated to H3K27me3 near their TSS. Bar graph represents the number of enriched genes, while dots/dashed line indicate the corresponding B&H corrected p-value. Top 20 most significantly enriched terms are shown.

and H3K27me3 marks (Figure 2B). Pathway enrichment analysis showed that genes for which the expression was positively correlated to H3K4me3 were generally related to IFN signaling, response to stimuli, cell migration and connective tissue development (Figure 2C), while genes of which the expression was negatively correlated to H3K27me3 were generally related to cell adhesion, muscle system processes and extracellular matrix (ECM) organization (Figure 2D). These results demonstrate that H3K4me3 and H3K27me3 histone modifications in monocytes can be associated to changes in gene expression, and regulate distinct functional pathways in these cells.

Differentially expressed genes in SLE, SSc and RA monocytes are associated with promoter levels of H3K4me3 and H3K27me3

Combining the results of the genome-wide quantification of H3K4me3 and H3K27me3 levels with the transcriptome data allowed us to identify genes for which the dysregulated expression in SSc, SLE and RA monocytes is likely to be the result of altered histone modifications (Figure 3A). Given the respective activating and repressing functions of H3K4me3 and H3K27me3, we focused on genes correlated positively with H3K4me3 and negatively with H3K27me3. The expression levels of
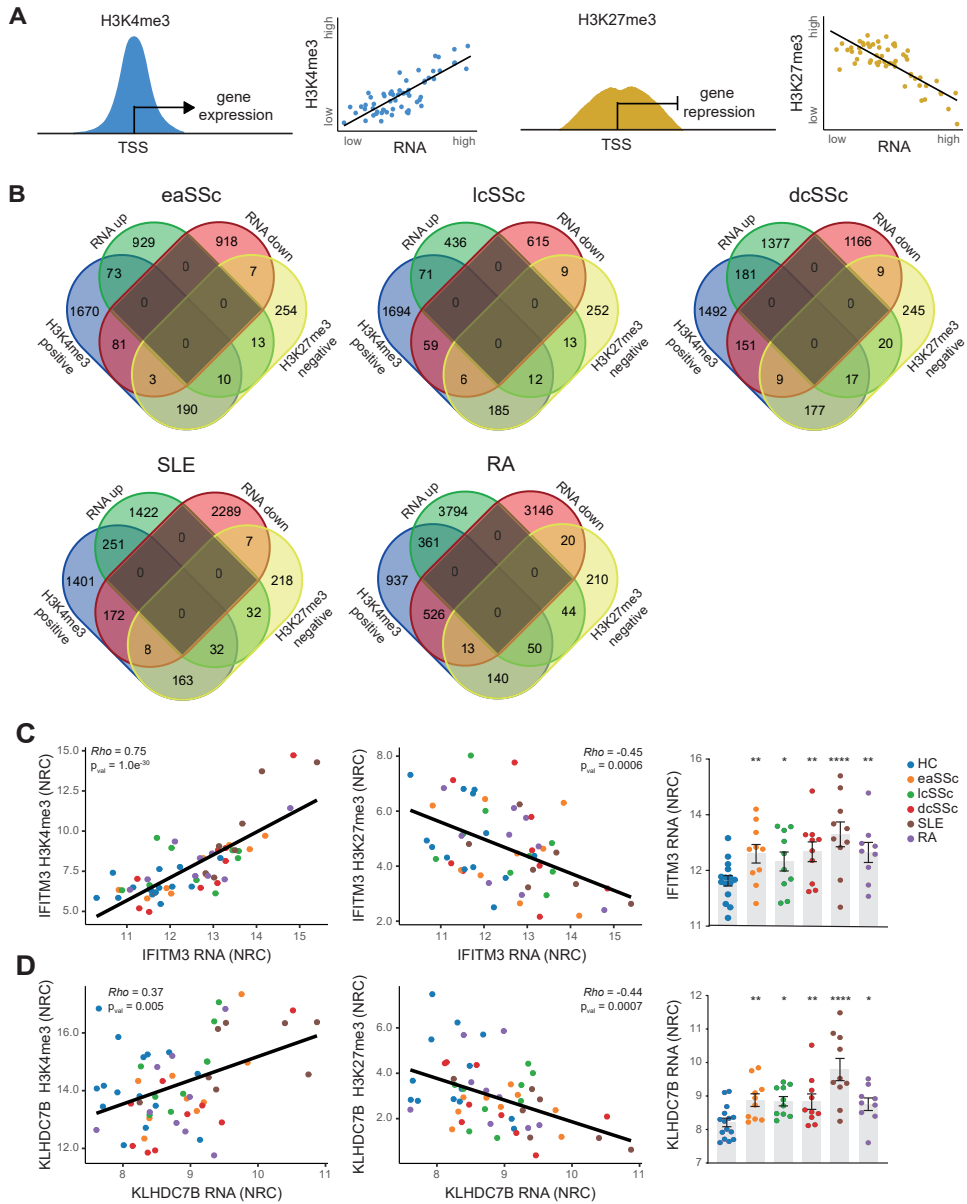
Figure 3. Levels of H3K4me3 and H3K27me3 partially underlie transcriptomic changes observed in monocytes SLE, SSc and RA monocytes. (A) Schematic overview of the effects of H3K4me3 and H3K27me3 at gene promoters on RNA expression. TSS = transcription start site. (B) Venn diagrams depicting overlap of DEGs with genes negatively correlated with H3K27me3 or positively with H3K4me3 levels at the gene promoter (+/- 10kb from the TSS). (C) Correlation of H3K4me3 (left panel) or H3K27me3 (middle panel) levels with RNA expression for IFITM3 and (D) KLHDC7B. Correlations were calculated using Spearman's rank correlation coefficient (Rho), Right panels show the expression levels of IFITM3 and KLHDC7B transcripts in SSc, SLE and RA in comparison to healthy controls (HC). Error bars indicate the mean and standard error of the mean. NRC = normalized read counts. P-value: * = <0.05, ** = <0.01, *** = <0.001, **** = <0.0001.
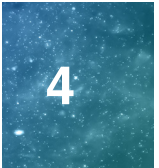
multiple DEGs in all patient groups were significantly correlated to changes in either H3K4me3, H3K27me3 or both marks (Figure 3B), indicating that changes in histone modifications at least partially underlie the transcriptomic changes observed in SLE, SSc and RA monocytes. GO-term enrichment analysis of upregulated DEGs with positively correlated H3K4me3 expression revealed that upregulated genes induced by increased H3K4me3 deposition were mostly involved in IFN and other pro-inflammatory signaling pathways in SSc and SLE patients (Supplementary Figure 1A-D), while again in RA a less strong enrichment for IFN was identified (Supplementary Figure 1E). Notably, RA was the only disease group where we also identified a significant enrichment for genes that were downregulated and positive correlated with H3K4me3, which were also related to IFN signaling as well as IL-1 production (Supplementary Figure 1F).This indicates that loss of H3K4me3 at gene promoters to IFN signaling in RA monocytes might also contribute to the lack of a clear IFN signature in these patients. For DEGs negatively correlated with H3K27me3 deposition, no significant enrichment was observed for any of the disease groups. Overlap analysis of the DEGs correlated positively to H3K4me3 and negatively to H3K27me3 (Supplementary Figure 2) revealed that there were only two genes upregulated in all disease subsets compared to healthy that correlated with both histone marks. These were interferon-induced transmembrane protein 3 (*IFITM3*, Figure 3C) and Kelch Domain Containing 7B (*KDC7B*, Figure 3D), which are both involved in IFN signaling[42, 43].

Disease specific gene expression patterns in pro-inflammatory signaling pathways are marked by altered H3K4me3 deposition at gene promoters

To see whether the altered expression of distinct signaling pathways in SLE, SSc and RA pathways are epigenetically imprinted, we repeated the gene set enrichment analysis on DEGs positively correlated with H3K4me3 or negatively correlated with H3K27me3. In line with the transcriptome analysis (Figure 1D/E), DEGs positively correlated with H3K4me3 from SLE monocytes displayed a strong enrichment for IFN as well as TNFα signaling, while DEGs from RA monocytes where exclusively enriched for genes involved TNFα signaling pathways (Figure 4A). Again SSc monocytes displayed an intermediate phenotype, with dcSSc monocytes being enriched in IFN and TNFα related genes (although less strong than SLE and RA), with weaker enrichment in lcSSc and eaSSc monocytes. For DEGs negatively correlated with H3K27me3 deposition, no enrichment was found. These results indicate that the upregulation of specific pro-inflammatory signaling pathways in SLE, SSc and RA monocytes is dependent on the altered deposition of the activating histone mark H3K4me3 near gene promoters.

Next, we studied the expression of IFN and TNFα related transcription factors present in the gene sets related TNFα and IFN signaling to further delineate the regulatory mechanisms underlying the differences in specific pro-inflammatory signaling pathways. After filtering for differential expression and positive correlation with H3K4me3, we identified a list of 22 human transcription factors (as defined in the human transcription factor database HumanTFDB[44]) involved in TNFα and/or IFN signaling pathways (Figure 4B). Again, transcription factors related to IFN signaling had a higher expression in SLE and SSc monocytes, while TNF related transcription factors were highly expressed
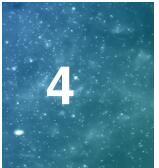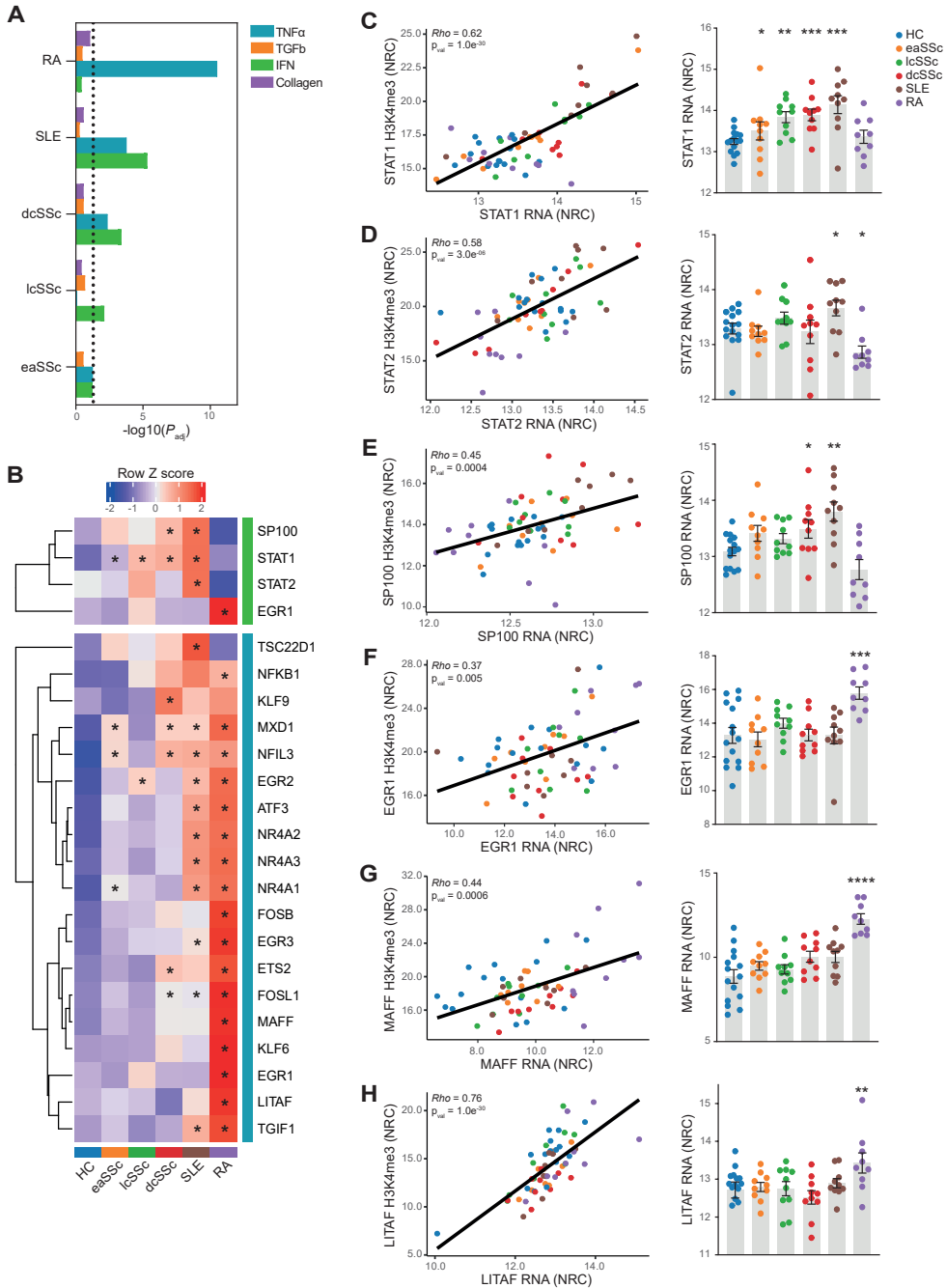
Figure 4. H3K4me3 deposition at promoters of TNF or IFN specific transcription factors drives the expression of distinct pro-inflammatory pathways in SLE, SSc and RA monocyte. (A) Gene set enrichment analysis for the ranked DEGs (positively correlated to H3K4me3) from SLE, SSc and RA monocytes against hallmark genes signatures from TNFα, TGFβ, IFN and collagen formation pathways. Adjusted p-values for each geneset are indicated, with the dashed line representing *(continued)*

p-value = 0.05. (B) Heatmap showing the expression of human transcription factors, implicated in IFN or TNFα signaling pathways. Differentially expressed (continued) transcription factors of which the expression positively correlates to the level of H3K4me3 at their gene promoters (+/- 10kb) are shown. * = significantly upregulated compared to healthy. (C-H) Left panels: correlation of H3K4me3 (y-axis) with RNA expression (x-axis). Right panels: expression levels of select transcription factors in the different disease subsets (SSc, SLE and RA) in comparison to healthy controls (HC). Error bars indicate the mean and standard error of the mean. NRC = normalized read counts. P-value: * = <0.05, ** = <0.01, *** = <0.001, **** = <0.0001.

in RA monocytes. In more detail, signal transducer and activator of transcription (STAT) family members *STAT1*, *STAT2* and SP100 Nuclear Antigen (*SP100*) were strongly upregulated in SLE monocytes and to some extent in SSc. monocytes (Figure 4C-E), while early growth response 1 (*EGR1*), V-Maf avian musculoaponeurotic fibrosarcoma oncogene homolog F (*MAFF*), and lipopolysaccharide induced TNF factor (*LITAF*), were exclusively upregulated in RA monocytes (Figure 4C-E). The expression of all these transcription factors was significantly correlated with the levels of H3K4me3 at their gene promoters. These results indicate that the increased deposition of the activating histone mark H3K4me3 leads to expression of pathway specific transcription factors, selectively skewing monocyte transcriptomes towards distinct pro-inflammatory phenotypes.

Increased deposition of H3K4me3 marks at poised promoters for extracellular matrix and fibrosis related genes in eaSSc, dcSSc and RA monocytes

Besides their respective roles in gene expression and repression, H3K4me3 and H3K27me3 can also occur simultaneously, marking bivalent gene domains. These bivalent domains have the ability to keep the chromatin in a poised state, where the chromatin is open due to H3K4me3 deposition, but gene transcription is repressed due to the concurrent presence of H3K27me3[45]. Following activation, gene transcription of poised promoters can increase rapidly with loss of H3K27me3, thus marking a fast switch between repression and expression[46].

To explore the role of bivalent promoters in monocytes, we compared the overall RNA expression profile of genes marked by a deposition of H3K4me3 only, H3K27me3 only or both (bivalent). In line with their respective activating and repressing functions, genes marked exclusively by H3K4me3 exhibited a high RNA expression, while genes marked exclusively by H3K27me3 showed a markedly lower expression (Figure 5A). Genes with bivalent promoters displayed an intermediate expression profile, with most genes being lowly expressed (Figure 5A). Next, we compared the relative levels of H3K4me3 and H3K27me3 at bivalent promoters and overlaid them with their corresponding RNA expression profiles (Figure 5B). Bivalent genes with high levels of H3K4me3, but low levels of H3K27me3 generally displayed a higher RNA expression (Figure 5B, genes closest to the x-axis), while, vice versa, genes with high H3K27me3 and low H3K4me3 displayed lower RNA expression levels (Figure 5B, genes closest to the y-axis). We also identified many genes with high levels of both H3K4me3 and H3K27me3, but low RNA expression levels, representing poised promoters. We then further characterized these poised promoters by taking genes marked by high levels of both H3K4me3 and H3K27me3 (Figure 5C, top 50%), which were not positively correlated with RNA, and displayed low RNA expression levels (Figure 5C, bottom 50%).
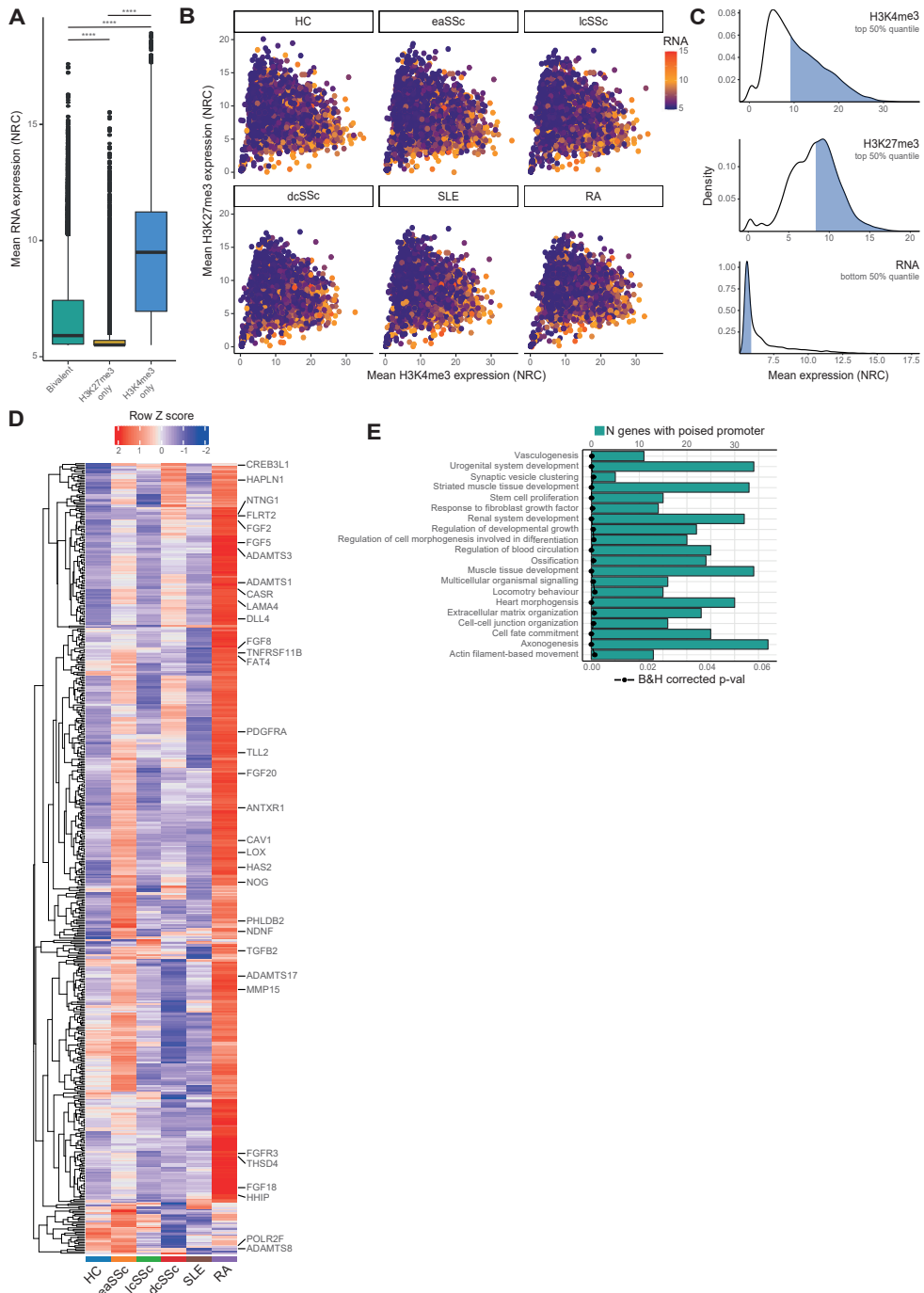
Figure 5. Identification of disease relevant poised promoters in monocytes from autoimmune patients. (A) RNA expression (y-axis, NRC = normalized read counts) for genes associated with single or bivalent histone modifications at their gene promoters (+/- 10kb). **** = p-value <0.0001 *(continued)*
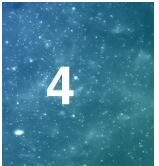
determined by Kruskal–Wallis test by ranks. (B) Scatter plots showing the expression levels of H3K4me3 (x-axis) and H3K27me3 at bivalent gene promoters. Every dot represents one gene. Colouring indicates the associated normalized RNA expression levels, ranging from low (purple) to high (red). (C) Density plots showing the distribution of H3K4me3 (top), H3K27me3 (middle) and RNA (bottom) expression of bivalent genes. Blue shading indicates the genes with the top 50% highest expression for H3K4me3 and H3K27me3, and the bottom 50% lowest expression for RNA. (D) Heatmap showing the levels of H3K4me3 expression at poised promoters acorss HC, eaSSc, lcSSc, dcSSc, SLE and RA patient monocytes. Genes involved in ECM organization and response to fibroblast growth factor have been highlighted. (E) GO-term enrichment results of genes with poised promoters. Bars represent the number of enriched genes, while dots/dashed line indicate the corresponding B&H corrected p-value. Top 20 most significantly enriched terms are shown. (F) Expression levels of H3K4me3 (left) and H3K27me3 (right) for selected genes with poised promoters. Error bars indicate the mean and standard error of the mean. NRC = normalized read counts.

This resulted in the selection of 532 genes with poised promoters in monocytes (Figure 5D). GO-term enrichment analysis revealed that these poised genes are majorly involved in regulation of developmental growth/response to growth factors, but also in ECM organization, vasculogenesis and ossification (bone formation) (Figure 5E), relevant in the context of SSc and RA pathogenesis.

Notably, the levels of H3K4me3 for the majority of poised promoters were higher in dcSSc, and especially in eaSSc and RA monocytes compared to healthy (Figure 5D), indicating that these genes will be induced with a particularly high expression following activation. Examples of such genes include the fibrosis related cytokines transforming growth factor beta 2 (*TGFB2*), the zinc-dependent metalloprotease tolloid-like protein 2 (*TTL2*), and matrix metalloproteinase 15 (*MMP15*) (Figure 5F). Thus, dcSSc and especially eaSSc and RA monocytes seem to be highly primed to respond strongly to activation, implicating genes in pathways relevant for pathogenesis of the respective diseases.

Identification of poised promoters active after monocyte differentiation to macrophages and osteoclasts

Since poised promoters can become activated in response to environmental stimuli or along cell type specific differentiation stages[46], we aimed to investigate whether poised promoters that we identified in circulating monocytes are active in differentiated macrophages or osteoclasts. To this end, we obtained H3K4me3 and H3K27me3 peak regions of macrophages and osteoclasts from the BLUEPRINT consortium, and overlapped these with poised promoters identified in our monocyte data. For the macrophage dataset, we identified 15 genes which had poised promoters in monocytes, but exclusively showed H3K4me3 expression in macrophages (Figure 6A), indicating that these genes can become activated in macrophages after loss of H3K27me3. Likewise, we identified 17 poised genes from monocytes that only showed H3K4me3 and lost H3K27me3 expression in osteoclasts (Figure 6B). Tripartite motif containing 47 (*TRIM47*) was active (as defined by the presence of H3K4me3 and absence of H3K27me3) in macrophages as well as osteoclasts. Notably, TRIM47 has been described to positively regulate pro-inflammatory signaling pathways as well as fibrosis[47, 48], two processes important in macrophage and osteoclast biology. Moreover, we identified Rhomboid 5 Homolog 1 (*RHBDF1*, also known as *iRhom1*), which regulates ossification through

the ADAM17/TGFα/EGFR signaling pathway[49], and the collagen processing enzyme Procollagen C-Endopeptidase Enhancer 2 (*PCOLCE2*)[50], two poised genes from monocytes that are marked by exclusive H3K4me3 deposition in osteoclasts. Altogether these results show that genes with poised promoters in monocytes can become activated in macrophages and osteoclasts, indicating that epigenetic imprinting, at least partially, underlies monocyte differentiation.
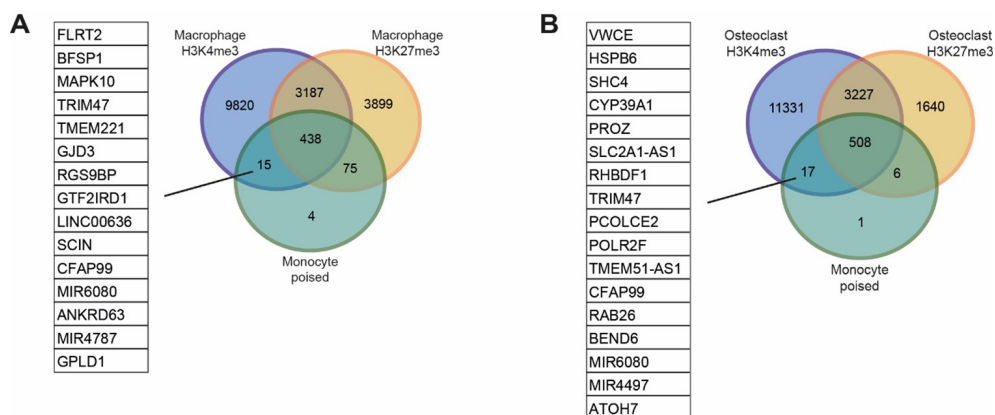
Figure 6. Identification of poised promoters active after monocyte differentiation to macrophages and osteoclasts. (A) Venn diagrams depicting overlap of with genes with poised promoters in monocytes versus genes with H3K4me3 or H3K27me3 at their promoters in macrophages and (B) osteoclasts.
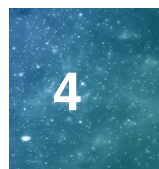
DISCUSSION

SSc, SLE and RA are rheumatic autoimmune diseases with pressing unmet clinical needs and a poorly understood etiology. The study of epigenomic alterations in immune cells of these patients may help to gain insights into disease specific immunopathology and identify potential novel therapeutic targets. Given the evidence of the role of monocytes in the pathogenesis of SSc, SLE and RA[5–10], in this study we aimed to identify the transcriptomic and epigenomic alterations characterizing these autoimmune monocytes. By performing paired genome-wide RNA- and ChIP-sequencing analysis of healthy, SSc, SLE and RA monocytes, we were able to characterize the transcriptomic landscape of monocytes in the context of autoimmunity, and directly correlate the expression of genes to the level of activating and repressing histone marks, H3K4me3 and H3K27me3, at the gene promoters. We identified numerous differentially expressed genes in SSc, SLE and RA monocytes for which the level of gene expression was highly proportional to the level of H3K4me3 and/or H3K27me3 at the gene promoter, confirming that epigenomic changes are directly associated with altered gene expression in these patients.

Transcriptomic analysis revealed that SSc, SLE and RA monocytes are characterized by differential the expression of genes involved in distinct pro-inflammatory and pro-fibrotic signaling pathways. Genes upregulated in SSc and particularly SLE monocytes were strongly enriched for pathways related to IFN signaling. Indeed, both SSc and SLE are considered type I interferon-mediated autoimmune diseases, sharing

the altered expression of IFN-related genes. Notably, the expression of IFN inducible factors have previously been be associated with disease severity in SSc patients[51, 52]. This is in line with our observations, as we observed an evidently weaker enrichment for IFN related pathways in eaSSc patients compared to lcSSc and dcSSc patients suffering from a more established form of the disease.

RA monocytes lacked an IFN signature but were highly enriched in genes related to the TNF signaling pathway, as well as pathways related to TGFβ and collagen formation. Previous research has also shown that the IFN inducible gene signature in RA is lower than that observed in SLE and SSc[38], and the peripheral blood IFN signature is not associated with clinical parameters in established RA[53]. We did, however, identify various genes related to IFN signaling that were upregulated in all disease groups, including RA. Interestingly, the expression of these IFN inducible genes, including *IFITM3*, has also been shown to be induced by TNFα though NF-κB signaling[54], suggesting that the partial IFN signature observed in RA patients is actually reflective of enhanced TNFα signaling. In fact, TNFα is widely recognized as a major cytokine governing autoimmune responses by activating monocytes to produce many pro-inflammatory mediators. This signaling through distinct pro-inflammatory pathways indicates a difference in immunopathology, implicating discrete immune pathway activation patterns in the dysregulation of monocytes in SSc, SLE and RA.

When repeating the gene set enrichment analysis only with those genes that showed a significant correlation to the increased amount of H3K4me3 at the promotor, we found the same pathways to be enriched, showing that the dysregulation of these pathways is already evident at the epigenomic level. Moreover, we identified numerous transcription factors implicated in TNF and IFN signaling pathways to be differentially expressed and correlated with activating H3K4me3. Transcription factors *STAT1*, *STAT2* and *SP100*, crucial for IFN signaling had a higher expression in SLE and SSc monocytes. Importantly, besides their well described role in the transcription of IFN dependent genes, members of the STAT family have previously been described to promote epigenetic changes leading to increased sensitivity to type I IFN stimulation[55, 56]. Interestingly, the upregulation of *STAT1* in SSc monocytes in our study was already observed in eaSSc patients, suggesting that STAT related signaling is involved in the potentiation of IFN signaling already from early disease stages onwards. For RA, on the other hand, we found very limited expression of transcription factors related to IFN signaling, while TNF related transcription factors, including *EGR1*, *MAFF*, *KLF6* and *LITAF*, were highly specifically upregulated and correlated to H3K4me3 profiles. Thus, whereas SSc and SLE monocyte inflammatory response are largely dependent on transcription factor pathways related to IFN, RA monocytes are clearly characterized by TNF driven signaling.

Parallel ChIP-seq analysis of the repressive histone mark H3K27m3 and the activating histone mark H3K27me3 from the same samples allowed us to identify poised promoters, characterized by the simultaneous presence of both marks. For poised promoters, H3K27me3 represses gene expression while the presence of H3K4me3 primes the promoter for rapid activation when needed[45, 46]. We identified numerous poised promoters in circulating monocytes, of which many were associated with genes related to response to fibroblast growth factors ECM organization, vasculogenesis and

ossification. These poised promoters showed increased levels of H3K4me3 in eaSSc, RA, and to a limited extend dcSSc monocytes, suggesting that these circulating monocytes display an enhanced poised status, enabling enhanced expression of disease relevant genes rapidly upon stimulation. Such stimuli may include signals from the inflammatory microenvironment (i.e. skin in SSc patients and synovium in RA patients), which can skew the selective differentiation of monocytes into disease relevant cell types including inflammatory macrophages, myofibroblasts and osteoclasts. Poised promoters are also highly interesting for eaSSc patients lacking the fibrotic features which typify the most severe disease stages of SSc. The presence of primed poised promoters in these early patients suggests that monocytes already display a skewed, "pre-fibrotic" phenotype before the onset of overt fibrosis, providing new avenues for disease classification and early intervention.

As monocytes are proposed to form a reservoir for disease relevant cell types in SSc[57, 58], and RA[59, 60], we also compared poised promoters from our ChIP-seq analysis to H3K4me3 and H3K27me3 data obtained from macrophages and osteoclasts. Here, we identified a number of poised genes from monocytes, which showed a "monovalent" H3K4me3 expression in macrophages or osteoclasts, reflecting an activating switch in gene expression. Poised promoters switching to monovalent H3K4me3 included some genes relevant for macrophage and osteoclast function including *TRIM47*, *RHBDF1* and *PCOLCE2*[47, 49, 50], implicating monocyte poised promoters in the differentiation to downstream cell types. To further delineate the role of poised promoters in monocyte differentiation, more in-depth analyses of the dynamics of the chromatin landscape during differentiation of monocytes to distinct cell types are needed. Additionally, the macrophage and osteoclast data that we used here were obtained from healthy individuals, which might be epigenetically distinct from the same cell types in autoimmune patients. Nevertheless, our data show that, based on the presence of poised promoters with enhanced H3K4me3 expression, RA and SSc monocytes have an increased potential to differentiate into disease relevant macrophages and osteoclasts.

In conclusion, we observed various differences in the chromatin landscapes of monocytes in autoimmune patients and found strong indications that these changes have implications for the transcription of genes relevant to disease pathogenesis. Given the reversible nature of histone modifications the targeting of histone modifying enzymes may proof beneficial in treating or preventing the symptoms of autoimmune disease.
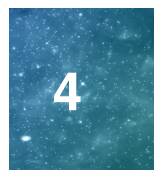
REFERENCES

1. C. S. Robbins, F. K. Swirski. The multiple roles of monocyte subsets in steady state and inflammation. *Cell. Mol. Life Sci.*, vol. 67, no. 16, pp. 2685–2693. Aug. 2010

2. L. B. Boyette, C. Macedo, K. Hadi, *et al.* Phenotype, function, and differentiation potential of human monocyte subsets. *PLoS One.* vol. 12, no. 4, p. e0176460. Apr. 2017

3. Y. Fujikawa, J. M. W. Quinn, A. Sabokbar, *et al.* The human osteoclast precursor circulates in the monocyte fraction. *Endocrinology.* vol. 137, no. 9, pp. 4058–4060. Sep. 1996.

4. R. Huber, D. Pietsch, J. Günther, *et al.* Regulation of monocyte differentiation by specific signaling modules and associated transcription factor networks. *Cell. Mol. Life Sci.* vol. 71, no. 1, pp. 63–92. Jan. 2014

5. N. Higashi-Kuwata, M. Jinnin, T. Makino, *et al.* Characterization of monocyte/macrophage subsets in the skin and peripheral blood derived from patients with systemic sclerosis. *Arthritis Res. Ther.* vol. 12, no. 4, p. R128. Jul. 2010.

6. M. van der Kroef, L.L. van den Hoogen, J.S. Mertens, *et al.* Cytometry by time of flight identifies distinct signatures in patients with systemic sclerosis, systemic lupus erythematosus and Sjögrens syndrome. *Eur. J. Immunol.* vol. 50, no. 1, pp. 119–129. Jan. 2020.

7. L. R. Coulthard, J. Geiler, R.J. Mathews, *et al.* Differential effects of infliximab on absolute circulating blood leucocyte counts of innate immune cells in early and late rheumatoid arthritis patients. *Clin. Exp. Immunol.* vol. 170, no. 1, pp. 36–46. Oct. 2012.

8. N. Kawanaka, M. Yamamura, T. Aita, *et al.* CD14+,CD16+ blood monocytes and joint inflammation in rheumatoid arthritis. *Arthritis Rheum.* vol. 46, no. 10, pp. 2578–2586. Oct. 2002.

9. R. Yano, M. Yamamura, K. Sunahori, *et al.* Recruitment of CD16+ monocytes into synovial tissues is mediated by fractalkine and CX3CR1 in rheumatoid arthritis patients. *Acta Med. Okayama.* vol. 61, no. 2, pp. 89–98. Apr. 2007.

10. S. Yoshimoto, K. Nakatani, M. Iwano, *et al.* Elevated Levels of Fractalkine Expression and Accumulation of CD16+ Monocytes in Glomeruli of Active Lupus Nephritis. *Am. J. Kidney Dis.* vol. 50, no. 1, pp. 47–58. Jul. 2007.

11. W.T. Ma, F. Gao, K. Gu, *et al.* The Role of Monocytes and Macrophages in Autoimmune Diseases: A Comprehensive Review. *Front. Immunol.* vol. 10. May 2019.

12. J. Xue, L. Xu, H. Zhu, *et al.* CD14+CD16– monocytes are the main precursors of osteoclasts in rheumatoid arthritis via expressing Tyro3TK. *Arthritis Res. Ther.* vol. 22, no. 1, p. 221. Dec. 2020.

13. N. Binai, S. O'Reilly, B. Griffiths, *et al.* Differentiation potential of CD14 + monocytes into myofibroblasts in patients with systemic sclerosis. *PLoS One.* vol. 7, no. 3, pp. 1–7. Mar. 2012.

14. T. T. Wada, Y. Araki, K. Sato, *et al.* Aberrant histone acetylation contributes to elevated interleukin-6 production in rheumatoid arthritis synovial fibroblasts. *Biochem. Biophys. Res. Commun.* vol. 444, no. 4, pp. 682–686. Feb. 2014.

15. K. Hur, S.-H. Kim, J.-M. Kim. Potential Implications of Long Noncoding RNAs in Autoimmune Diseases. *Immune Netw.*, vol. 19, no. 1. 2019.

16. S. Chen, W. Pu, S. Guo, *et al.* Genome-Wide DNA Methylation Profiles Reveal Common Epigenetic Patterns of Interferon-Related Genes in Multiple Autoimmune Diseases. *Front. Genet.* vol. 10, pp. 223. Apr. 2019.

17. R. J. Wierda, M. Goedhart, M.C.J.A. van Eggermond, *et al.* A role for KMT1c in monocyte to dendritic cell differentiation. *Hum. Immunol.* vol. 76, no. 6, pp. 431–437. Jun. 2015.

18. S. Saeed, J. Quintin, H.H.D. Kerstens, *et al.* Epigenetic programming of monocyte-to-macrophage differentiation and trained innate immunity. *Science.* vol. 345, no. 6204, pp. 1251086–1251086, Sep. 2014.

19. M. van der Kroef, M. Castellucci, M. Mokry, *et al.* Histone modifications underlie monocyte dysregulation in patients with systemic sclerosis, underlining the treatment potential of epigenetic targeting. *Ann. Rheum. Dis.* vol. 78, no. 4, pp. 529–538. Apr. 2019.

20. F. van den Hoogen, D. Khanna, J. Fransen, *et al.* 2013 classification criteria for systemic sclerosis: an American college of rheumatology/European league against rheumatism collaborative initiative. *Ann. Rheum. Dis.* vol. 72, no. 11, pp. 1747–1755. Nov. 2013.

21. M. C. Hochberg. Updating the American college of rheumatology revised criteria for the classification of systemic lupus erythematosus. *Arthritis Rheum.* vol. 40, no. 9, pp. 1725–1725. Sep. 1997.

22. D. Aletaha, T. Neogi, A.J. Silman, *et al.* 2010 Rheumatoid arthritis classification criteria: An American College of Rheumatology/European League Against Rheumatism collaborative initiative. *Arthritis*

and *Rheumatism*. vol. 62, no. 9, pp. 2569–2581. Sep. 2010.

23. E. C. LeRoy, T. A. Medsger. Criteria for the classification of early systemic sclerosis. *J. Rheumatol*. vol. 28, no. 7, pp. 1573–1576. Jul. 200.

24. A. Dobin, C.A. Davis, F. Schlesinger, *et al*. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. vol. 29, no. 1, pp. 15–21. Jan. 2013.

25. S. Anders, P. T. Pyl, W. Huber. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics*. vol. 31, no. 2, pp. 166–169. Jan. 2015.

26. D. Risso, K. Schwartz, G. Sherlock, *et al*. GC-Content Normalization for RNA-Seq Data. *BMC Bioinformatics*. vol. 12, no. 1, pp. 480, 2011. Dec. 2011.

27. M. I. Love, W. Huber, S. Anders. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. vol. 15, no. 12, pp. 550. Dec. 2014.

28. B. Langmead, S. L. Salzberg. Fast gapped-read alignment with Bowtie 2. *Nat. Methods*. vol. 9, no. 4, pp. 357–359. Apr. 2012.

29. J. M. Gaspar. Improved peak-calling with MACS2. *bioRxiv*. 2018

30. H. M. Amemiya, A. Kundaje, A. P. Boyle. The ENCODE Blacklist: Identification of Problematic Regions of the Genome. *Sci. Rep.* vol. 9, no. 1, pp. 9354. Dec. 2019.

31. G. Yu, L.-G. Wang, Q.-Y. He. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics*. vol. 31, no. 14, pp. 2382–2383, Jul. 2015.

32. D. Adams, L. Altucci, S.E. Antonarakis, *et al*. BLUEPRINT to decode the epigenetic signature written in blood. *Nat. Biotechnol*. vol. 30, no. 3, pp. 224–226. Mar. 2012.

33. L. J. Zhu, C. Gazin, N.D. Lawson, *et al*. ChIPpeakAnno: a Bioconductor package to annotate ChIP-seq and ChIP-chip data. *BMC Bioinformatics*. vol. 11, no. 1, p. 237. Dec. 2010.

34. G. Korotkevich, V. Sukhov, N. Budin, *et al*. Fast gene set enrichment analysis.*bioRxiv*, 2021.

35. A. Liberzon, A. Subramanian, R. Pinchback, *et al*. Molecular signatures database (MSigDB) 3.0. *Bioinformatics*. vol. 27, no. 12, pp. 1739–1740. Jun. 2011.

36. G. Yu, L.-G. Wang, Y. Han, *et al*. clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. *Omi. A J. Integr. Biol*. vol. 16, no. 5, pp. 284–287. May 2012.

37. H. Wickham. ggplot2: Elegant Graphics for Data Analysis. *Springer-Verlag New York*, 2016.

38. B. W. Higgs, Z. Liu, B. White, *et al*. Patients with systemic lupus erythematosus, myositis, rheumatoid arthritis and scleroderma share activation of a common type I interferon pathway. *Ann. Rheum. Dis.* vol. 70, no. 11, pp. 2029–2036. Nov. 2011.

39. K. A. Kirou, C. Lee, S. George, *et al*. Coordinate overexpression of interferon-alpha-induced genes in systemic lupus erythematosus. *Arthritis Rheum*. vol. 50, no. 12, pp. 3958–3967. Dec. 2004.

40. T.Y. Roh, S. Cuddapah, K. Cui, *et al*. The genomic landscape of histone modifications in human T cells. *Proc. Natl. Acad. Sci*. vol. 103, no. 43, pp. 15782–15787. Oct. 2006.

41. M. Lawrence, S. Daujat, R. Schneider. Lateral Thinking: How Histone Modifications Regulate Gene Expression. *Trends Genet*. vol. 32, no. 1, pp. 42–56. Jan. 2016.

42. G. Jeong, H. Bae, D. Jeong, *et al*. A Kelch domain-containing KLHDC7B and a long non-coding RNA ST8SIA6-AS1 act oppositely on breast cancer cell proliferation via the interferon signaling pathway. *Sci. Rep*. vol. 8, no. 1, pp. 12922. Dec. 2018.

43. D. Jiang, J.M. Weidner, M. Qing, *et al*. Identification of Five Interferon-Induced Cellular Proteins That Inhibit West Nile Virus and Dengue Virus Infections. *J. Virol*. vol. 84, no. 16, pp. 8332–8341. Aug. 2010.

44. H. Hu, Y.-R. Miao, L.-H. Jia, *et al*. AnimalTFDB 3.0: a comprehensive resource for annotation and prediction of animal transcription factors. *Nucleic Acids Res*. vol. 47, no. D1, pp. D33–D38. Jan. 2019.

45. P. Voigt, W.-W. Tee, D. Reinberg. A double take on bivalent promoters. *Genes Dev*.vol. 27, no. 12, pp. 1318–1338. Jun. 2013.

46. A. M. Pietersen, M. van Lohuizen. Stem cell regulation by polycomb repressors: postponing commitment. *Curr. Opin. Cell Biol*. vol. 20, no. 2, pp. 201–207. Apr. 2008.

47. M.-Q. Hao, L.-J. Xie, W. Leng, *et al*. Trim47 is a critical regulator of cerebral ischemia-reperfusion injury through regulating apoptosis and inflammation. *Biochem. Biophys. Res*. Commun. vol. 515, no. 4, pp. 651–657. Aug. 2019.

48. L. Li, S. Zhang, L. Wei, *et al*. Anti-fibrotic effect of melittin on TRIM47 expression in human embryonic lung fibroblast through regulating TRIM47 pathway. *Life Sci*. vol. 256, pp. 117893. Sep. 2020.

49. R. Fang, C. Haxaire, M. Otero, *et al*. Role of iRhoms 1 and 2 in Endochondral Ossification. *Int. J. Mol. Sci*. vol. 21, no. 22, pp. 8732. Nov. 2020.

4
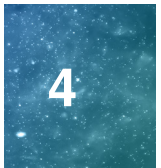
50. B. M. Steiglitz, D. R. Keene, D. S. Greenspan. PCOLCE2 Encodes a Functional Procollagen C-Proteinase Enhancer (PCPE2) That Is a Collagen-binding Protein Differing in Distribution of Expression and Post-translational Modification from the Previously Described PCPE1. *J. Biol. Chem.* vol. 277, no. 51, pp. 49820–49830. Dec. 2002.
51. X. Liu, M.D. Mayes, F.K. Tan, *et al.* Correlation of interferon-inducible chemokine plasma levels with disease severity in systemic sclerosis. *Arthritis Rheum.* vol. 65, no. 1, pp. 226–235. Jan. 2013.
52. G. Farina, D. Lafyatis, R. Lemaire, *et al.* A four-gene biomarker predicts skin disease in patients with diffuse cutaneous systemic sclerosis. *Arthritis Rheum.* vol. 62, no. 2, pp. 580–588. Feb. 2010.
53. T. D. de Jong, M. Blits, S. de Ridder, *et al.* Type I interferon response gene expression in established rheumatoid arthritis is not associated with clinical parameters. *Arthritis Res. Ther.* vol. 18, no. 1, p. 290. Dec. 2016.
54. A. Nakajima, D. Ibi, T. Nagai, *et al.* Induction of interferon-induced transmembrane protein 3 gene expression by lipopolysaccharide in astrocytes. *Eur. J. Pharmacol.* vol. 745, pp. 166–175. Dec. 2014.
55. E. Kaleviste, M. Saare, T.R. Leahy, *et al.* Interferon signature in patients with STAT1 gain-of-function mutation is epigenetically determined. *Eur. J. Immunol.* vol. 49, no. 5, pp. 790–800. May 2019.
56. G. Vahedi, H. Takahashi, S. Nakayamada, *et al.* STATs Shape the Active Enhancer Landscape of T Cell Populations. *Cell.* vol. 151, no. 5, pp. 981–993. Nov. 2012.
57. S. Soldano, A.C. Trombetta, P. Contini, *et al.* Increase in circulating cells coexpressing M1 and M2 macrophage surface markers in patients with systemic sclerosis. *Ann. Rheum. Dis.* vol. 77, no. 12, pp. 1842–1845. Dec. 2018
58. M. R. York, T. Nagai, A. J. Mangini, *et al.* A macrophage marker, siglec-1, is increased on circulating monocytes in patients with systemic sclerosis and induced by type i interferons and toll-like receptor agonists. *Arthritis Rheum.* vol. 56, no. 3, pp. 1010–1020. Mar. 2007.
59. S. Fukui, N. Iwamoto, A. Takatani, *et al.* M1 and M2 Monocytes in Rheumatoid Arthritis: A Contribution of Imbalance of M1/M2 Monocytes to Osteoclastogenesis. *Front. Immunol.* vol. 8. pp. 1958. Jan. 2018.
60. F. Liote, B. Boval-Boizard, D. Weill, *et al.* Blood monocyte activation in rheumatoid arthritis: increased monocyte adhesiveness, integrin expression, and cytokine release. *Clin. Exp. Immunol.* vol. 106, no. 1, pp. 13–19. Oct. 1996.

SUPPLEMENTARY INFORMATION
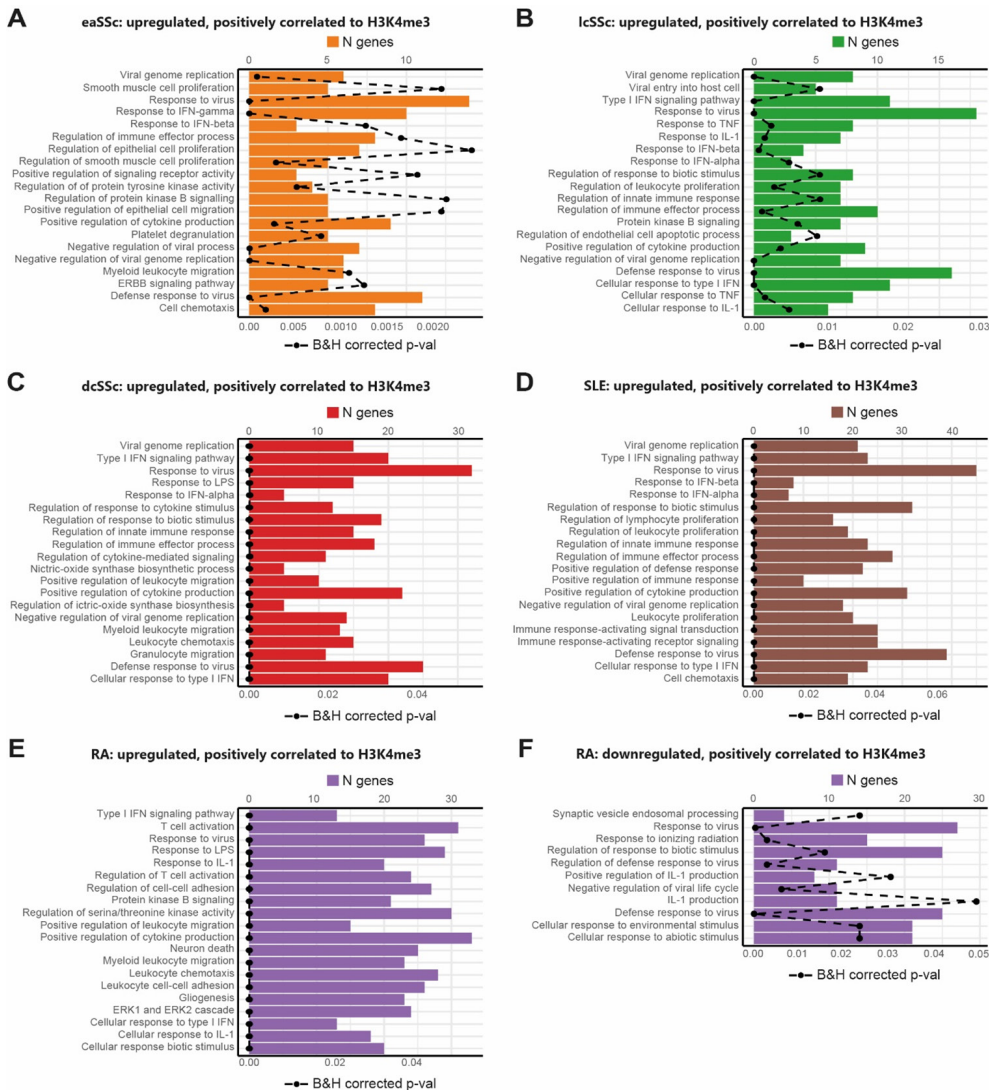
Supplementary Figure 1. Pathway enrichment analysis of differentially expressed genes correlated with H3K4me3 at their transcription start site. Bar graphs represents the number of enriched DEGs, while dots/dashed line indicate the corresponding B&H corrected p-value. For A-E, top 20 most significantly enriched terms are shown. For DEGs correlated to H3K27me3 expression, no significant enrichments were found.
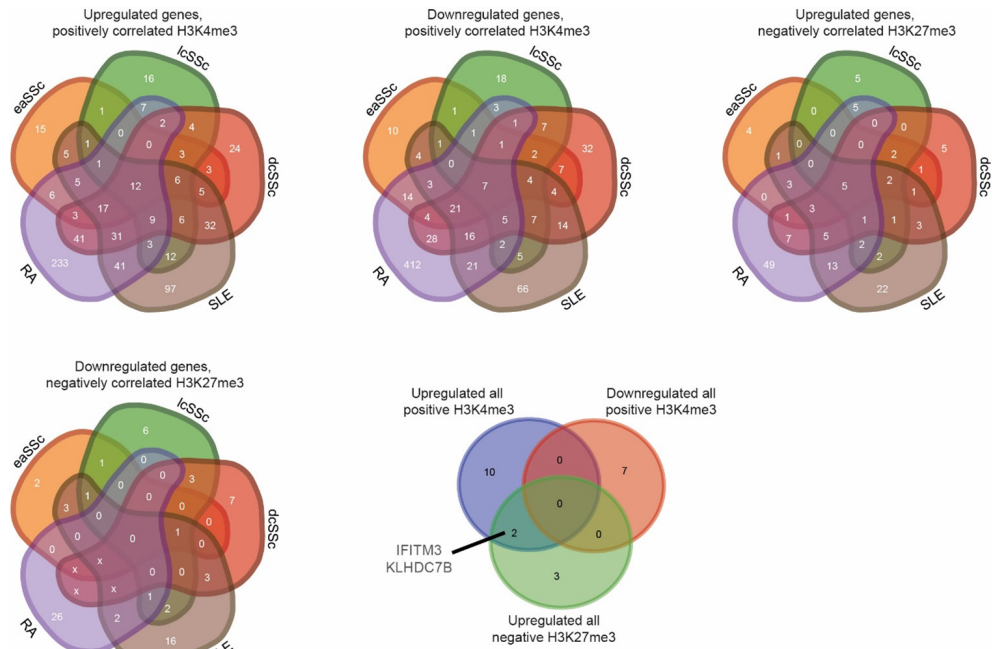
Supplementary Figure 2. Identification of DEG genes overlapping between all disease groups, correlated to H3K4me3 or H3K27me3 levels at their promoter (+/- 10kb). Venn diagrams depicting the unique and overlapping differentially expressed genes identified in SLE, SSc and RA monocytes, either positively correlated to the acting histone mark H3K4me3 or negatively correlated to the repressing histone mark H3K27me3.

4
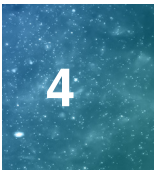
Chapter 5

Network-based multi-omics of CD1c+ dendritic cells implicates nuclear receptor subfamily 4A signaling as a key disease pathway in systemic sclerosis

5
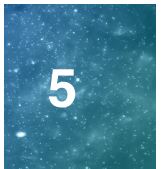
Authors:
N.H. Servaas[1], S. Hiddingh[1,2], E. Chouri[1,3], C.G.K. Wichers[1,3], A.J. Affandi[1,3], A. Ottria[1,3], C.P.J. Bekker[1,3], M. Cossu[1,3], S. Silva-Cardoso[1,3], M. van der Kroef[1,3], A.C. Hinrichs[1,3], T. Carvalheiro[1,3], N. Vazirpanah[1,3], L. Beretta[4], M. Rossato[1,3,5], F. Bonte-Mineur[6], T.R.D.J. Radstake[1,3], J.J.W. Kuiper[1,2]*, M. Boes[1,7]*, A. Pandit[1,3]*

*These authors contributed equally

Affiliations:
1. Center for Translational Immunology, University Medical Center Utrecht, Utrecht University, Utrecht, The Netherlands
2. Ophthalmo-Immunology Unit, University Medical Center Utrecht, Utrecht, The Netherlands
3. Department of Rheumatology & Clinical Immunology, University Medical Center Utrecht, Utrecht University, Utrecht, The Netherlands
4. Scleroderma Unit, Referral Center for Systemic Autoimmune Diseases, Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico di Milano, Milan, Italy
5. Department of Biotechnology, University of Verona, Verona, Italy
6. Department of Rheumatology and Clinical Immunology, Maasstad Hospital, Rotterdam, the Netherlands
7. Department of Pediatrics, University Medical Center Utrecht, Utrecht University, The Netherlands.

Manuscript in preparation

ABSTRACT

Objectives: To identify key pathways, transcriptional regulators and epigenetic mechanisms underlying conventional dendritic cell (cDC) alteration in Systemic Sclerosis (SSc).

Methods: Transcriptomic profiling was performed on CD1c+ cDCs isolated from peripheral blood samples obtained from 12 healthy donors and 48 SSc patients with all major disease subtypes. Differential expression analysis comparing the different SSc subtypes and healthy donors was performed to uncover genes dysregulated in SSc. To identify biologically relevant pathways, a gene co-expression network was built using the Weighted Gene Correlation Network Analysis. ChIP-sequencing and in vitro functional assays were performed to independently validate the role of key transcriptional regulators identified by network analysis.

Results: We identified 17 modules of co-expressed genes in cDCs that correlated with SSc subtypes and key clinical traits. A module of immune regulatory genes was markedly down regulated in patients suffering from a highly progressive subtype with severe fibrosis. This immune regulatory module was predicted to be controlled by genes belonging to the NR4A (nuclear receptor 4A) subfamily (*NR4A1*, *NR4A2*, *NR4A3*), which are master transcriptional mediators of inflammation. Indeed, ChIP-sequencing analysis of cDCs supported that these NR4A members target numerous differentially expressed genes in SSc cDCs. Functional experiments using agonists showed that dysregulation of NR4As affects cytokine production by cDCs and modulates T-cells activation.

Conclusions: We identified *NR4A1*, *NR4A2* and *NR4A3* as important regulators of immunosuppressive and fibrosis-associated pathways in SSc cDCs. Thus, the NR4A family represent novel potential targets to restore cDC homeostasis in SSc.

INTRODUCTION

Systemic Sclerosis (SSc) is a complex, chronic autoimmune disease mainly characterized by vascular abnormalities, immunological abnormalities, and fibrosis of the skin and internal organs[1]. According to the American College of Rheumatology (ACR) criteria, and their extent of skin fibrosis, patients can be classified into four subsets: early (eaSSc), non-cutanous (ncSSc), limited (lcSSc) and diffuse (dcSSc)[2, 3]. Vascular injury appears to be the earliest events in the pathogenesis of SSc[4]. Such vascular damage leads to the recruitment and activation of various immune cells, which secrete pro-inflammatory cytokines and growth factors[5]. The resulting mix of inflammatory molecules induces the differentiation of resident epithelium, endothelium, monocytes and fibroblasts into myofibroblast that deposit excessive amounts of extracellular matrix, eventually leading to permanent tissue scarring[6, 7].

Conventional dendritic cells (cDCs) are a heterogeneous population of antigen presenting cells that regulate adaptive immune cell responses[8], but also vascular tissue, and fibroblasts[9, 10]. Consequently, cDCs have been suggested to play an important role in the pathogenesis of SSc. Indeed, in the early phases of the disease cDCs migrate to the skin[11, 12] and cDCs from patients display an enhanced pro-inflammatory cytokine production upon Toll like receptor (TLR) stimulation[13]. As such, in SSc pathogensis, cDCc have been hypothesized to dictate T-cell responses and activate pathways that promote fibrosis[14].

Although data from previous studies supports an important role for cDCs in the pathogenesis of SSc, it remains incompletely understood what molecular processes drive their dysregulation in the disease. To better understand the role of cDCs in SSc, we performed transcriptional profiling of circulating CD1c+ cDCs obtained from peripheral blood of SSc patients, and compared these profiles to those from cDCs obtained from healthy donors. Using weighted gene correlation network analysis (WGCNA), we show that a cluster of immune regulatory genes is down-regulated in cDCs obtained from SSc patients. Moreover, using cell culture-based experimentation, we demonstrate that manipulation of NR4As, major regulators of the gene cluster, attenuates pro-inflammatory responses in cDCs.

MATERIALS AND METHODS

Patient demographics

Peripheral blood samples were collected from patients with SSc and age/sex matched healthy volunteers. Informed consent was obtained from all patients and healthy volunteers included in this study at the University Medical Center Utrecht (The Netherlands), the Maasstad Medical Center Rotterdam (The Netherlands), and the IRCCS Policlinico of Milan (Italy). Samples and clinical information were immediately anonymized upon collection. All participants enrolled in the study signed an informed consent form approved by the local institutional review boards prior to inclusion in this study (METC no. 12-466C, approved 2 October 2012), adherent of the Declaration of Helsinki Principles. All SSc patients fulfilled the American College of Rheumatology/

5
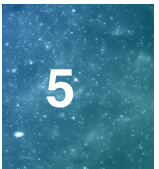
European League Against Rheumatism (ACR/EULAR) 2013 classification criteria[15]. We also included in our studies non-cutaneous (ncSSc) patients who fulfilled the classification criteria but did not present skin fibrosis, and early (eaSSc) patients with Raynaud's Phenomenon and positivity for SSc-specific autoantibodies and/or typical nailfold capillaroscopy patterns, as defined by LeRoy *et al.*[3]. The demographics of the SSc patients and healthy volunteers included in this study, for RNA-sequencing as well as RT-qPCR validation and FACS analysis are provided in Table 1.

	RNA-sequencing cohort					RT-qPCR validation		FACS cohort	
	HC	eaSSc	ncSSc	lcSSc	dcSSc	HC	dcSSc	HC	dcSSc
n	12	12	12	12	12	7	6	11	12
Age (years)	53 (49-66)	60 (42-77)	53 (49-60)	61 (52-65)	53 (48-58)	46 (38-51)	45 (37-55)	44 (35-54)	56 (50-64)
Female, n (%)	10 (83)	10 (83)	12 (100)	12 (100)	7 (58)	7 (100)	6 (100)	3 (25)	3 (23)
ANA, n pos (%)	-	10 (83)	10 (83)	11 (92)	11 (92)	-	5 (83)	-	12 (100)
ACA, n pos (%)	-	7 (58)	7 (58)	10 (83)	2 (17)	-	0 (0)	-	1 (8)
Scl70, n pos (%)	-	1 (8)	1 (8)	3 (25)	3 (25)	-	4 (67)	-	5 (42)
mRSS	-	-	-	4 (3-6)	17 (10-22)	-	10 (7-11)	-	12 (10-13)
Disease duration	-	-	5 (2-7)	20 (11-30)	7 (2-12)	-	10 (2-15)	-	5 (2-6)

Table 1. Clinical characteristics of subjects enrolled in the study. Reported values indicate the number (n) of individuals, with the (%) for each parameter. For age, mRSS and disease duration, the median with (IQCR) is given. ANA, anti-nuclear antibodies; ACA, anti-centromere antibodies; Scl70, anti-topoisomerase antibodies; mRSS, modified Rodnan skin score; HC, healthy controls; eaSSc, early SSc; ncSSc, non-cutaneous SSc; lcSSc, limited cutaneous SSc; dcSSc, diffuse cutaneous SSc; pos, positivity.

CD1c+ cDC purification

Peripheral blood mononuclear cells (PBMCs) were isolated from whole heparinized-blood samples obtained from SSc patients and healthy volunteers, by density gradient centrifugation on Ficoll-PlaqueTM Plus (GE Healthcare Life Sciences). The CD1c+ cDC population was purified from PBMCs using the MACS human CD1c (CD1c+) dendritic cell isolation kit (Miltenyi Biotec) on the autoMACs Pro Separator (Miltenyi Biotec) according to the manufacturer's protocol. For the RNA-sequencing cohort as well as the RT-qPCR validation cohort, freshly isolated cDCs were immediately lysed in RLTplus buffer (Qiagen) supplemented with β-mercaptoethanol, and stored at -20°C until further processing.

RNA-sequencing and analysis

Total RNA was purified from RLTplus lysates using the DNA/RNA/miRNA Universal Kit (Qiagen), according to manufacturer's instructions. Purified RNA was quantified with the Qubit® RNA Assay Kit (Life Technologies) on the Qubit® Fluorometer (Invitrogen). RNA-sequencing was performed at the Beijing Genomics Institute (BGI). cDNA libraries were generated from total RNA using the TruSeq RNA sample preparation kit (Illumina), specifically selecting for polyadenylated transcripts. The libraries were sequenced on the HiSeq 2000 system (Illumina), using 100bp paired-end reads.

We obtained at least 20 million raw reads for each sample. After quality filtering

according to the BGI pipeline, reads were aligned to the GrCh38 reference human genome and the homo sapiens transcriptome (Ensembl, version 79), using the STAR aligner[16]. Summed exon read counts per gene were calculated using the Python package HTSeq[17], using annotations from the GrCh38 built from the human genome (http://www.ensembl.org, version 79). Between lane normalization using the upper quartile normalization method was performed using the Bioconductor/R package EDAseq[18]. To account for batch effects arising from the inclusion of samples from different geographic locations (The Netherlands and Italy), we applied the generalized linear model from the Bioconductor/R package RUVseq[19] using the *RUVr* function for k = 2 factors of unwanted variation. Subsequently, differential expression analysis was performed using the negative binomial distribution-based method implemented in DESeq2[20] on the normalized summed exon read counts per gene. Pair wise comparisons between patients and HC groups were tested using the Wald test, and genes with a nominal p-value <0.05 were considered to be significantly differential. Gene expression levels are given as variance stabilised data (VSD), calculated according to DESeq2 instructions.

Weighted gene co-expression network analysis

Weighted gene co-expression networks were constructed using the R package WGCNA[21], using the VSD data of all genes with an normalized read count expression higher than 5.53 (translates to at least 1 raw count) in all samples as input. We used a soft threshold power of 5 to construct an unsigned network with scale free topology. Modules were identified using the *cutreeDynamic*() function with a minimum module size of 50 genes. Next, closely related modules were merged using the *mergeCloseModules* function, with a cutHeight of 0.25, according to the WGCNA manual.

cDC cultures

For functional experiments, CD1c+ cDCs were purified from healthy donor buffy coat blood samples (Sanquin, Amsterdam, The Netherlands), according to the cDC purification method described in section 2.2. Freshly isolated cDCs were cultured in RPMI 1640 medium with GlutaMAX™ (Life Technologies), supplemented with 10% heat inactivated fetal bovine serum (Biowest) and 1% penicillin streptomycin (Life Technologies). For stimulation experiments, cDCs were cultured at a concentration of 0.5 × 10^5 cells/mL in 100μl in a 96-well round-bottom plate. Cells were either left untreated, or treated for 18 hours with one of the following stimuli: TLR7/8 ligand Resiquimod (R848, 100ng/mL, Invivogen), GM-CSF (800U/mL, R&D), CXCL4 (10ug/mL, Peprotech), TNFα (100ng/mL, Tebu-bio), IFNα-2a (1000U/mL, Cell Sciences), LPS EB Ultrapure (100ng/mL, Invivogen), IL-6, (50ng/mL, ImmunoTools), IL-15 (200 ng/mL, ImmunoTools) or TGFβ-b2 (100ng/ml, R&D). For hypoxia experiments, cells were cultured in atmospheric or hypoxic conditions (Rasquinn invivO2 1000 hypoxic chamber, set at 1% O2 and 5% CO2) for 24 hours. For experiments using NR4A agonists, cells were pre-treated for one hour with either dimethyl sulfoxide (DMSO, Sigma) or NR4A agonists C-DIM5 (Tocris Bioscience) or C-DIM12 (Tebu-Bio) prior to stimulation. cDC cultures were incubated at 37°C in the presence of 5% CO2, for the time points indicated in each single experiment in the results section. Supernatants were stored at -80°C until further use, while cDCs

5
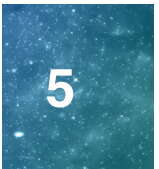
were lysed in RLTplus buffer and stored at -20°C until further use.

PBMC cultures

For functional experiments, three batches (individual days) of randomly selected dcSSc and healthy control samples of liquid nitrogen stored PBMCs were thawed in RPMI 1640 (Thermo Fisher Scientific) supplemented with 20% FCS (Sigma), and washed with PBS. Cells were resuspended in RPMI 1640 medium with GlutaMAX™ (Life Technologies), supplemented with 10% heat inactived fetal bovine serum (Biowest) and 1% penicillin streptomycin (Life Technologies), and plated at a concentration of $0,75 \times 10^6$ cells in $200 \mu l$ in round-bottom 96-wells plates. PBMCs were pre-treated for one hour with 10uM dimethyl sulfoxide (DMSO, Sigma), C-DIM5 (Tocris Bioscience) or C-DIM12 (Tebu-Bio) prior to stimulation with R848 (100ng/mL, Invivogen) and GolgiStop (1500x, BD biosciences), and incubated at 37°C in the presence of 5% CO_2, for 4 hours. IL-6 production by CD1c+ cDCs was measured using intracellular cytokine staining by FACS.

Reverse Transcription Quantitative Real-Time PCR (RT-qPCR)

Purified RNA was reverse transcribed using the SuperScript® IV Reverse Transcriptase kit (Invitrogen), according to the manufacturer's instructions. Gene expression was quantified, in duplicate, by RT-qPCR using the SYBR Select Master Mix (Applied Biosystems), using gene-specific primers (Supplementary Table 1) on the QuantStudio 12k flex System (Applied Biosystems). Relative gene expression was determined according to the comparative CT (ΔΔCT) method using GUSB as an endogenous control (where the ΔCT equals the CT of the mRNA of interest—the CT of GUSB). The fold change (FC) of each sample was calculated in relation to the ΔCt of the medium control according to the formula $FC = 2^{-\Delta\Delta Ct}$, where ΔΔCt = ΔCt sample—ΔCt reference.

Assessment of IL-6 production using ELISA

Concentrations of IL-6 in cell-free supernatants from cultured CD1c+ cDCs were measured by sandwich enzyme linked immunosorbent assay (ELISA). IL-6 was quantified using the PeliKine compact human IL-6 (Sanquin Reagents, Amsterdam, The Netherlands), according to the manufacturer's instructions.

cDC / CD4+ T-cell co-cultures

CD1c+ cDCs and CD4+ T-cells were isolated in parallel from PBMCs obtained from healthy donor buffy coats using the MACS human CD1c dendritic cell isolation kit and CD4+ T Cell Isolation Kit (Miltenyi Biotec) on the autoMACs Pro Separator (Miltenyi Biotec) according to the manufacturer's instructions. cDCs and T-cells were cultured, in parallel, in culture medium (RPMI 1640 medium with GlutaMAX™ (Life Technologies), supplemented with 10% heat inactived fetal bovine serum (Biowest) and 1% penicillin streptomycin (Life Technologies)). cDCs were pre-treated for one hour with either dimethyl sulfoxide (DMSO, Sigma) or specific NR4A agonists C-DIM5 (Tocris Bioscience) or C-DIM12 (Tebu-Bio), and were either left untreated or treated with R848 (100ng/mL, Invivogen). After overnight incubation, cDCs were washed twice with sterile PBS,

resuspended in culture medium and added to the CD4+ T-cells in a 1:5 ratio (10.000 cDCs : 50.000 CD4+ T-cells, in 100μl in a round bottom 96-wells plate). Cells were cultured for 3 days at 37°C in the presence of 5% CO2. After 3 days, co-cultures were re-stimulated with phorbol myristate acetate (PMA, 50ng/ml, Sigma-Aldrich) and ionomycin (500ng/ml, Sigma-Aldrich) for 6 hours. During the last 3 hours of incubation, GolgiStop (1500x, BD biosciences) was added to block cytokine release. IFNy production by CD4+ T-cells was measured using intracellular cytokine staining by FACS.

Flow cytometry

For cDC/T-cell co-cultures and PBMC cultures, cells were washed with cold PBS and incubated with Fixable Viability Dye eF780 (eBioscience) at room temperature for 10 minutes. Cells were then transferred to V-bottomed plates (Greiner Bio-one), washed with PBS and incubated for 30 minutes at 4°C in the dark with the surface staining antibodies provided in Supplementary Table 2. Next, the cells were washed in FACS buffer (1% bovine serum albumin and 0.1% sodium azide in phosphate buffered saline), and fixed/permeabilized for 30 minutes at 4°C in the dark with 100 μl Fixation/ Permeabilization Concentrate and Diluent (00-5123-43, 00-5223-56, eBioscience), followed by intracellular staining using the antibodies provided in Supplementary Table 2. After 60 minutes of staining at 4°C in the dark, the cells were washed and taken up in FACS buffer and flow cytometric analyses were performed on the BD LSRFortessa with four lasers (405, 488, 561, and 635 nm) using DIVA software version 8.0.1. Analysis of FCS files were performed using FlowJo software (TreeStar inc.).

Chromatin immunoprecipitation sequencing (ChIP-seq)

CD1c+ cDCs were isolated from PBMCs obtained from healthy donor buffy coats using the MACS human CD1c dendritic cell isolation kit. 1.5×10^6 freshly isolated cDCs were cultured overnight in RPMI 1640 medium with GlutaMAX™ (Life Technologies), supplemented with 10% heat inactived fetal bovine serum (Biowest) and 1% penicillin streptomycin (Life Technologies), and were either left untreated or treated with TLR7/8 agonist R848 (Resiquimod, 100ng/mL, Invivogen). After culture, cDCs were washed with PBS and crosslinked using the truChIP® Ultra-Low Chromatin Shearing Kit (Covaris), according to the manufacturer's instructions. Chromatin shearing was performed using AFA Fiber Pre-Slit Snap-Cap microTUBEs (Covaris) by sonication (peak incident power: 105, duty factor: 2%, cycles per burst: 200, treatment time: 12 minutes) on the Covaris S220 focused ultrasonicator (Woburn, MA, USA). After shearing, the chromatin was transferred into pre-chilled microcentrifuge tubes and centrifuged at 10.000 x g, 4°C for 5 minutes to pellet insoluble material, and the chromatin was stored at -20°C for downstream processing. For every condition we obtained three biological replicates, for which material from three to four donors was pooled to obtain enough chromatin for chromatin immunoprecipitation. Chromatin immunoprecipitation was performed with 3μg anti-NR4A1 (NB100-56745, Novusbio), anti-NR4A2 (NB110-40415, Novusbio) or anti-NR4A3 (NLS2341, Novusbio) antibodies, using the low cell ChIP-seq kit (Active Motif), according to the manufacturer's instructions. For all conditions, 10% of the input chromatin was removed prior to addition of the antibodies and used to normalize the amount of

5

immunoprecipitated DNA (input control). Following immunoprecipitation, ChIP DNA was de-crosslinked in the presence of NaCl and Proteinase K (Active Motif) in a thermocycler at 65°C overnight, and DNA was extracted by penol:chloroform:isoamylalcohol precipitation and dissolved in Low-EDTA TE buffer (Active Motif). ChIP-seq libraries were generated by GenomeScan (Leiden, the Netherlands) with the NEBNext® Ultra II DNA Library Prep kit (Illumina), and were sequenced using Illumina NovaSeq6000 generating ~20 million 150bp paired ended reads for each sample.

ChIP-seq analysis

Quality check of the raw sequencing reads was performed using the FastQC tool. Sequencing reads were mapped against the reference genome GRCh38.p13 built from the human genome (NCBI) using bowtie2[22]. Peaks were called using MACS2[23], by comparing the IP samples to their matched input samples in paired end mode (BAMPE). After calling, peaks from ENCODE blacklist regions[24], and X- and Y-chromosome peaks were filtered out to reduce noise and exclude sex-specific peaks. Peaks were annotated to the nearest genes using the ChIPseeker Bioconductor/R package[25]. Peaks were considered to be associated to a gene when they were annotated within a 10kb range (up- or downstream) to the gene transcription start site (TSS).

Statistical analysis

The Mann Whitney test was used to compare any combination of two groups. For multiple group comparisons, the one- or two-way analysis of variance (ANOVA) was used. A p-value < 0.05 was considered statistically significant. Figures were produced using the R package ggplot2[26]. GO-term and KEGG enrichment analyses were performed using enrichGo and enrichKEGG functions from the clusterProfiler R package[27]. Terms with a B&H correct p-value <0.05 were considered significant. Spearman's rank correlation coefficient was calculated to assess correlations.

RESULTS

SSc cDCs are transcriptionally distinct from healthy

RNA sequencing of purified blood cDCs from 48 SSc patients and 12 age/sex matched healthy individuals was performed to assess differences in their transcriptomic profiles. In total, 6,568 genes were found to be differentially expressed in at least one SSc subset versus healthy controls (P<0.05), of which 2,931 were down-regulated in SSc and 3,663 genes were upregulated (Figure 1A/B). Principal component analysis (PCA) revealed that the profile of differentially expressed genes clearly separated SSc patients from healthy donors, with eaSSc and dcSSc patients displaying the most distinct clustering patterns (Figure 1C). GO-term enrichment analysis showed that differentially expressed genes were predominantly enriched in pathways related to immune cell activation, interferon (IFN) signaling and translation (Supplementary Figure 1, Supplementary Table 3). These results demonstrate that SSc patient cDCs have a distinct transcriptional profile from healthy cDCs. Accordingly, the identified differentially
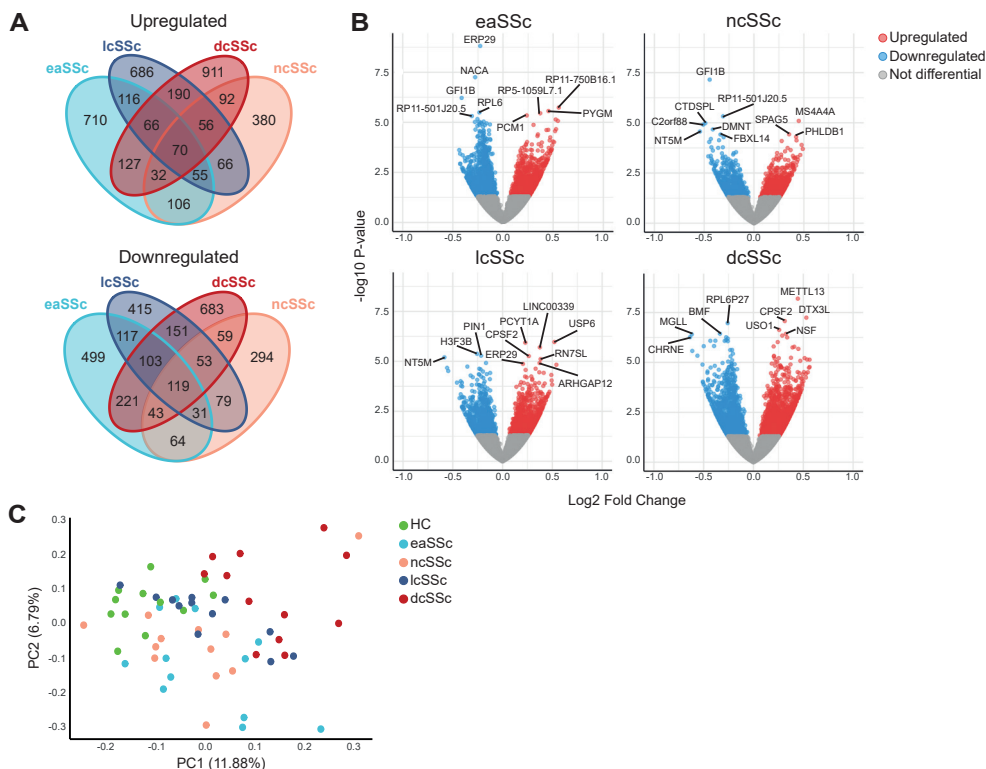
Figure 1. SSc cDCs are transcriptomically distinct from healthy cDCs. a) Number of differentially expressed genes (P < 0.05) identified in cDCs from different SSc subsets versus healthy donors. b) Volcano plots highlighting transcriptional differences between different SSc subsets and healthy donors. Red dots represent significantly down-regulated genes (P < 0.05, log2 fold change < 0), while blue dots represent significantly up-regulated genes (P < 0.05, log2 fold change > 0). Top 10 differentially expressed genes based on p-value are highlighted. c) Principal component analysis (PCA) of the differentially expressed genes from all comparisons of SSc patients to healthy donors.

expressed genes may elude to pathways relevant for SSc pathogenesis.

Co-expression network analysis identifies NR4As as key regulators of functionally relevant pathways altered in SSc cDCs

In order to clarify the molecular pathways dysregulated in SSc cDCs, we conducted genome-wide gene co-expression analysis using Weighted Gene Co-expression Network Analysis (WGCNA)[21]. Using this approach, we identified 42 modules of tightly co-expressed genes in cDCs. Comparison of module gene composition with differentially expressed genes, and correlation between module eigengenes and clinical traits for SSc allowed us to identify 17 clinically relevant modules significantly enriched in differentially expressed genes and correlated with clinical traits (Figure 2A). Functional annotation of these modules using pathway enrichment analysis showed that multiple modules were associated with molecular pathways relevant in the context of cDC biology and SSc pathogenesis. These included the blue module (associated with viral pathways
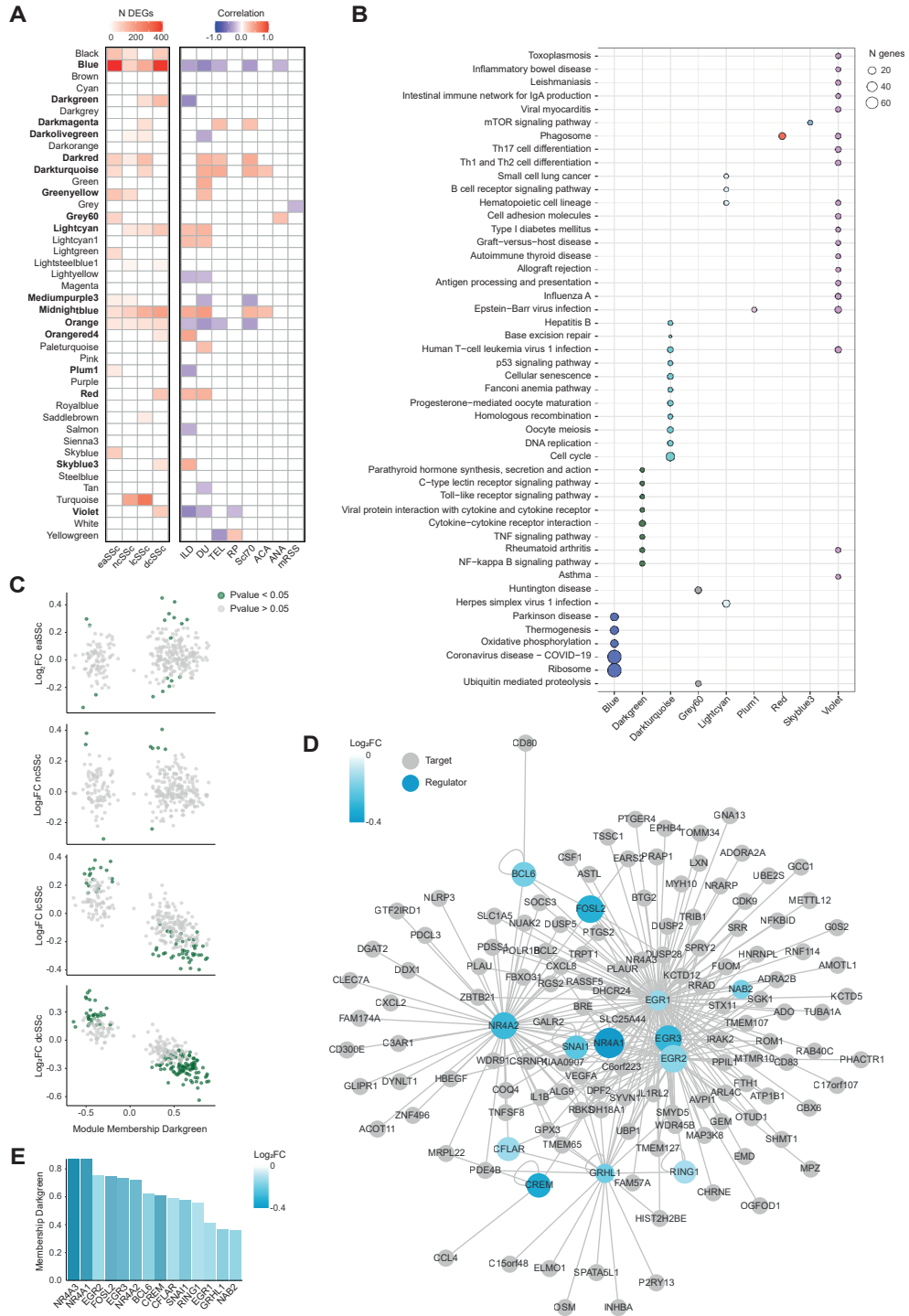
Figure 2 *(Left)*. Co-expression network analysis identifies NR4A transcription factors of regulatory pathways decreased in SSc cDCs. (A) Heatmaps depicting overlap of differentially expressed genes with co-expression modules (right) and correlation of module eigengenes (MEs) to SSc clinical traits (left). Cells of significant enrichments for differentially expressed genes (Fisher's exact test p-value <0.05) or significant correlations with clinical traits (Pearson, p-value <0.05) are highlighted. N DEGs = number of differentially expressed genes. Modules significantly enriched in DEGs and correlated with clinical traits are highlighted in bold. (B) KEGG enrichment of relevant modules. Circle size denotes the number of module genes associated to enriched pathways. Top 10 pathways are shown (B&H corrected p-value < 0.05). (C) Module membership (x-axis) and the Log2FC (fold change) in gene expression compared to healthy controls (y-axis) for genes in the darkgreen module in patients with eaSSc, ncSSc, lcSSc and dcSSc. Significantly differentially expressed genes are highlighted in green (p-value <0.05). (D) Transcription factor network obtained for the darkgreen module. Transcriptional regulators (blue) are connected to their targets (grey) based on known interactions from REGNET and TTRUST databases. Blue shading of indicate fold change between dcSSc and healthy cDCs for transcriptional regulators. Node size denotes module membership in the darkgreen module. (E) Module membership (y-axis) and Log2(fold change) (healthy versus dcSSc, colour scale) of transcriptional regulators (x-axis) identified in the darkgreen module.

and ribosomes), the darkgreen module (associated with immune cell activation) and the violet module (associated with antigen presentation and inflammation) (Figure 2B).

Closer inspection of the darkgreen module revealed that genes with the highest module membership (MM, i.e. best representative of the overall module gene expression pattern), were also the most strongly downregulated in dcSSc patients (Figure 2C). This observation supports that genes downregulated in dcSSc cDCs are likely driving this module. Because transcription factors are key regulators of gene expression, we next performed a transcription factor network analysis to see whether transcription factors and their targets were enriched within the darkgreen module. Following this approach, we identified 13 transcription factors that had target genes present in the darkgreen module (Figure 2D). Most notable were members of the NR4A family of nuclear receptors, *NR4A1* and *NR4A2*, which have previously been described as important regulators of inflammation and fibrosis[30–34]. Out of the 13 transcription factors identified, *NR4A1* displayed the highest module membership (MM = 0.87) and was strongly downregulated in dcSSc patients (log2FC = -0.38, P = 0.004, central node in Figure 2D). In addition, *NR4A3*, which was not identified in the transcription factor network analysis but is also known to regulate inflammatory pathways[35], also displayed high module membership in the darkgreen module (MM = 0.87) and was strongly downregulated in dcSSc patients (log2FC = -0.44, P = 0.0006). These, results support that NR4As are important regulatory factors in cDCs, and their expression is down-regulated in dcSSc patients.

NR4As are downregulated in cDCs obtained from dcSSc patients, but their expression is induced by pro-inflammatory stimuli.

To validate the downregulation of the NR4A family members observed in dcSSc patients in our RNA-sequencing cohort (Figure 3A), we evaluated their expression by RT-qPCR in another, independent cohort of 6 dcSSc patients and 7 healthy controls (Table 1). We found *NR4A2* and *NR4A3* to be significantly downregulated in dcSSc patients in this cohort (P = 0.002 and 0.035 respectively, Figure 3B), while for *NR4A1* we observed a trend for downregulation, albeit not significant (P = 0.18, Figure 3B). Nonetheless, the
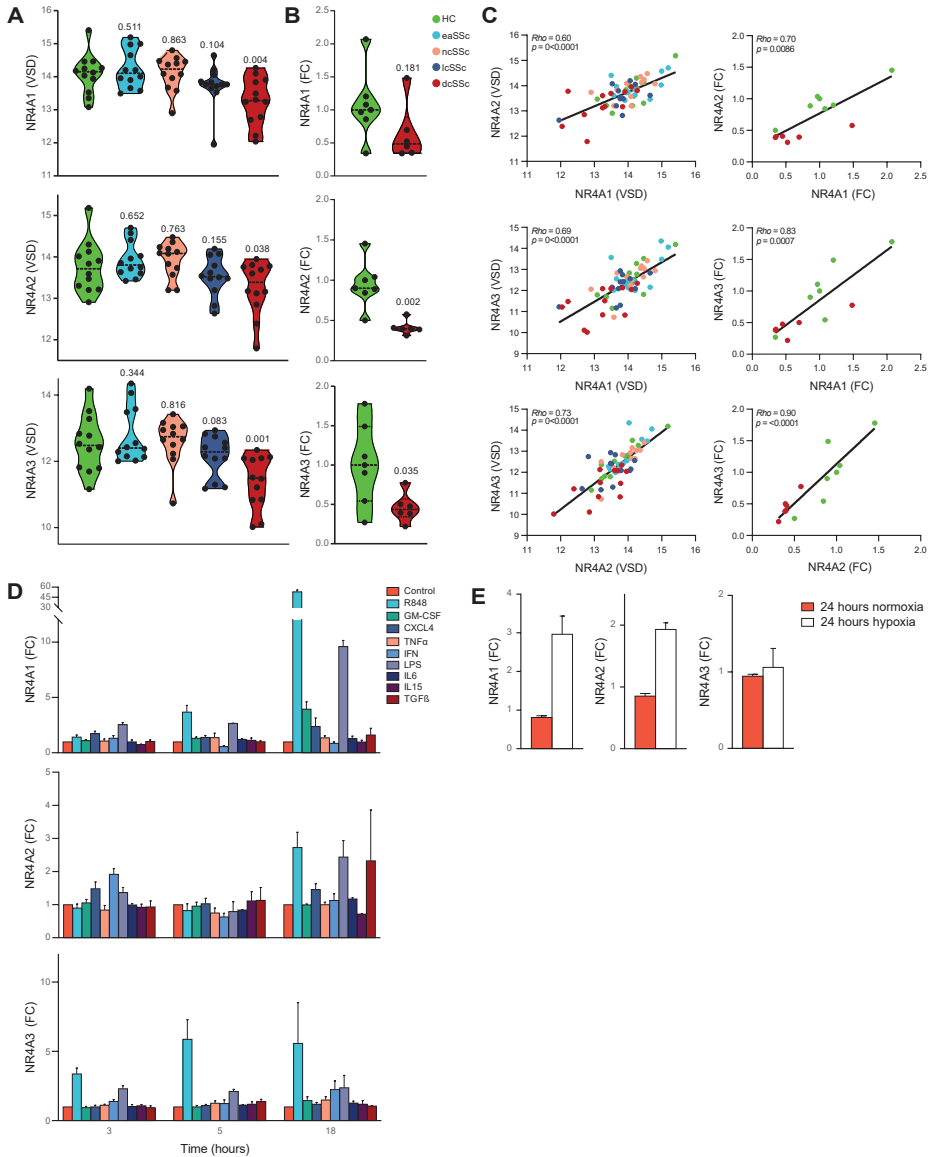
Figure 3. Characterization of NR4A expression in SSc and healthy cDCs. (A) Violin plots depicting NR4A expression in SSc patients and healthy donors from the RNA-sequencing cohort (dashed lines indicates mean). For each comparison of SSc patients versus healthy controls, the p-value, calculated according to the Wald test implemented in DESeq2, is shown. (B) Violin plots showing NR4A expression in the validation cohort measured by target specific RT-qPCR. Data are reported as the fold change of each donor versus one representative healthy control. P-values (Kruskall–Wallis with post-hoc Dunn's test), are reported. (C) Scatterplots showing the correlation (regression line) of NR4A expression in the RNA-sequencing cohort (left) and the validation cohort (right). Correlations were calculated using Spearman's rank correlation coefficient (Rho). (D) NR4A expression in CD1c+ cDCs from healthy donors stimulated for 3, 5 or 18 hours with Toll-like receptor ligands and cytokines implicated in dendritic cell biology and systemic sclerosis pathogenesis. Data are shown as mean with SEM of 3-4 (continued) experiments. (E) NR4A expression in CD1c+ cDCs from healthy donors cultured in normoxic or hypoxic conditions. Data are shown as mean with SEM of 3 experiments.

expression of levels of all NR4A genes in the RNA-sequencing cohort and the RT-qPCR validation cohort were strongly correlated to each other (Figure 3C), showing that all dcSSc donors display consistent downregulation of all NR4As.

Next, we sought to identify factors underlying the downregulation of NR4As observed in dcSSc patients. Because dcSSc patients display enhanced levels of pro-inflammatory cytokines and TLR stimulation[13], we postulated that the downregulation of NR4As might be caused by cDC activation through these pro-inflammatory factors. To explore this hypothesis, we cultured freshly isolated CD1c+ cDCs obtained from peripheral blood of healthy volunteers and stimulated them with various pro-inflammatory factors for 3, 6, or 18 hours. NR4A expression was either induced (i.e. LPS and R848), or unaffected in all culture conditions studied (Figure 3D). For all three NR4As, the strongest induction of expression was observed upon stimulation with R848 (TLR7/8 agonist) for 18 hours of culture (Figure 3D). Additionally, since hypoxia has been described as an important factor underlying immune cell dysregulation in SSc[36, 37], we also investigated the effect of hypoxia on NR4A expression in cDCs. Similar to what was observed in the pro-inflammatory conditions, the expression of *NR4A1* and *NR4A2* was induced in hypoxia, while *NR4A3* expression was unaffected (Figure 3E). To explore whether induction of NR4A expression by pro-inflammatory stimuli was different between healthy donors and SSc patients, we repeated this analysis on 5 dcSSc patients and 5 sex/aged matched healthy donors. No significant difference in NR4A expression was observed between R848 stimulated cDCs from healthy and SSc donors (Supplementary Figure 2). These results support that the activation of cDCs through pro-inflammatory cytokines/TLR signaling or hypoxia does not underlie the decreased expression of NR4As observed in dcSSc patients.

Genome-wide ChIP-sequencing analysis further implicates NR4As as crucial regulators of dendritic cell dysregulation in SSc

To further investigate the regulatory potential of NR4As in cDCs, we set out to characterize the genome-wide binding sites of NR4A1, NR4A2 and NR4A3 by chromatin immunoprecipitation followed by parallel DNA sequencing (ChIP-seq). We cultured freshly isolated CD1c+ cDCs obtained from healthy donors in the overnight presence or absence of R848, and performed ChIP-seq analysis to identify NR4A specific binding sites under resting and stimulated conditions. We identified numerous NR4A binding sites in the promoter regions of genes in cultured cDCs, of which some were specific to resting (medium-treated control) conditions, while others were specific to stimulated (R848) conditions (Supplementary Figure 3). In line with their known roles in neuronal development[42] and cardiac tissue development[43], pathway analysis of NR4A binding sites identified enrichment for numerous genes related to these processes (Figure 4A). Moreover, we identified NR4A binding in the promoters of genes related to cell morphogenesis, and, in R848 stimulated cDCs, we identified an enrichment of NR4A2 and NR4A3 binding sites for genes related to extracellular matrix organization (Figure 4A). Overall, these results demonstrate that, besides their role in pro-inflammatory processes and T-cell activation by cDCs, NR4As are also involved in the regulation of various other processes in cDCs, including morphogenesis and fibrosis, which are highly
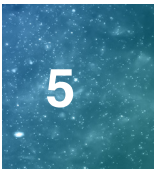
5
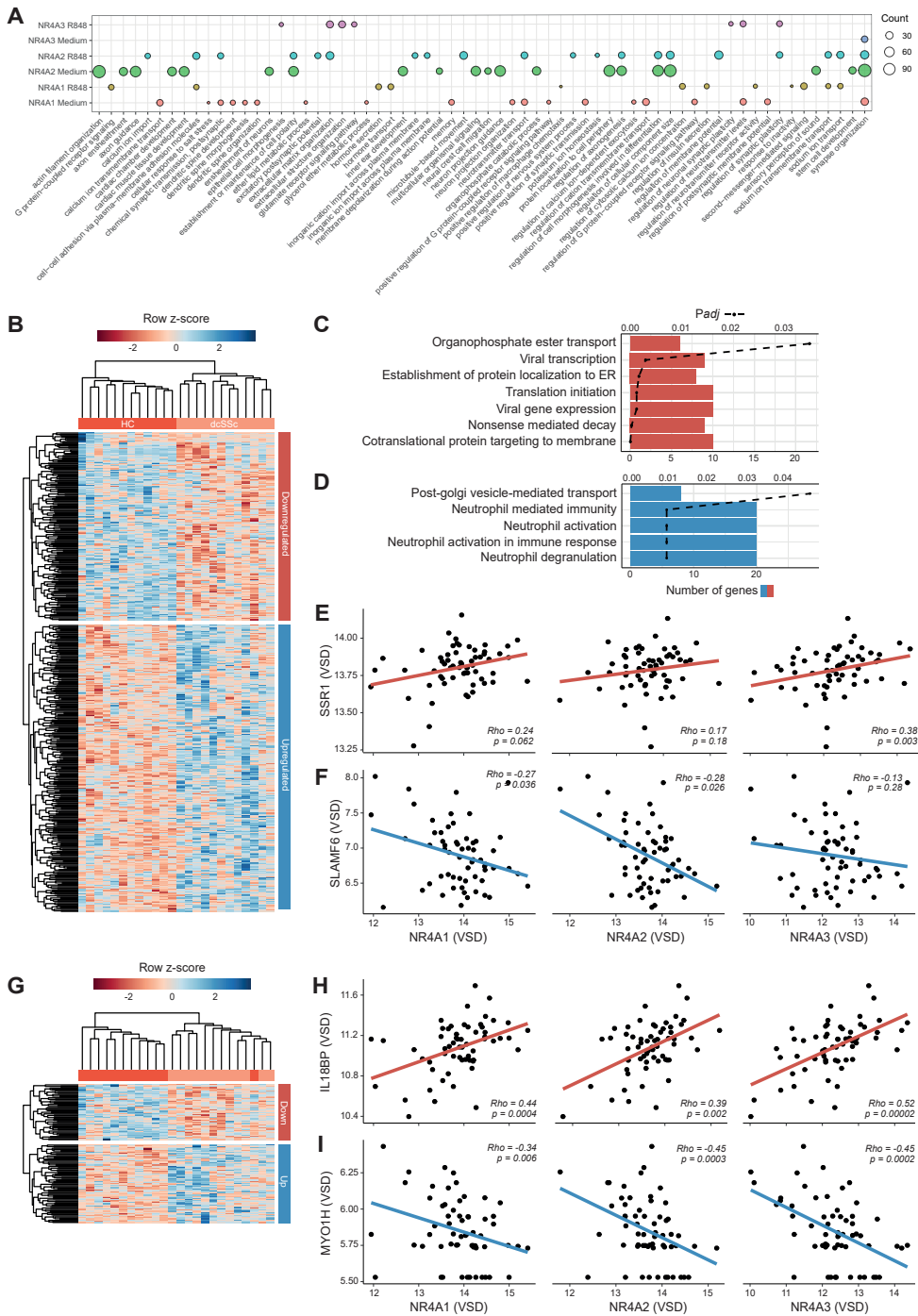
Figure 4. ChIP-sequencing of transcriptional regulation of cDCs by NR4As. (A) GO-term enrichment of genes that show NR4A binding within their promoter region. Circle size denotes the number of genes associated to enriched biological processes. Top 20 are shown (B&H corrected P < 0.05). (B) Heatmap of differentially expressed genes in dcSSc that also display binding of NR4As at their promoters *(continued)*

in resting cultured cDCs. (C) Pathway enrichment of genes downregulated and (D) upregulated in dcSSc with NR4A binding at their gene promoters in resting cultured cDCs. Terms for significantly enriched biological processes are given on the y-axis. Bars depict the number of genes identified within the enriched pathway (N genes, bottom x-axis), dashed line indicates B&H corrected P-value of the enrichment (p-value, top x-axis). (E) Scatterplots showing the correlation of *NR4A1-3* expression (x-axis) with *SSR1*, or (F) *SLAMF6* expression (y-axis). Correlations were calculated using Spearman's rank correlation coefficient (Rho). VSD = variance stabilized data. (G) Heatmap of differentially expressed genes in dcSSc that also display binding of NR4As at their promoters in R848 stimulated cDCs. (H) Scatterplots showing the correlation of *NR4A1-3* expression (x-axis) with *IL18BP*, or (I) *MYOH1* expression (y-axis). The red line on the scatter plot represents the regression line or "line of best fit" for positive correlations, and blue for negative correlations.

relevant in the context of SSc pathology.

We further aimed to identify which transcriptomic changes observed in dcSSc cDCs could be directly attributed to transcriptional regulation by NR4As. Overlap analysis of NR4A binding sites with differentially expressed genes in dcSSc patients revealed that 267 genes upregulated and 175 genes downregulated in dcSSc were directly bound by NR4As in resting cDCs (Figure 4B). Pathway enrichment analysis demonstrated that NR4A bound genes downregulated in dcSSc cDCs are implicated in the regulation of biological processes broadly related to transcription, translation and endoplasmic reticulum (ER) processing (Figure 4C), while NR4A bound genes upregulated in dcSSc cDCs are implicated in immune activation pathways (Figure 4D). Given the significantly lower expression of NR4As in dcSSc cDCs compared to healthy cDCs (Figure 3A/B), these results suggest that the induction of immune pathways in dcSSc cDCs is a direct effect of loss of transcriptional inhibition by NR4As, while the repression of pathways related to transcription/translation is a direct effect of loss of transcriptional activation. These results were substantiated by the fact that, in the RNA-sequencing cohort, NR4A expression was directly correlated with genes associated with ER processing, including *SSR1* - which encodes a receptor associated with protein translocation across the ER membrane[44] (Figure 4E), while NR4A expression was inversely correlated with genes related to immune activation, including *SLAMF6*[45] (Figure 4F). Moreover, we observed a strikingly lower overlap of differentially expressed genes with NR4A binding sites obtained from R848 stimulated cDCs as compared to resting cDCs (Figure 4G), reflecting loss of NR4A binding during cDC activation. Among the genes directly bound by NR4As in R848 stimulated cDCs, we identified the anti-inflammatory IL18 binding protein (*IL18BP*)[46], which was downregulated in dcSSc cDCs and displayed a direct correlation with NR4As in the RNA-sequencing cohort (Figure 4H). On the other hand, Myosin IH (*MYO1H*), a molecule involved cytokinesis, maintenance of cell shape, and cell motility[47], was upregulated in dcSSc cDCs and negatively correlated to NR4A expression (Figure 4I). Taken together, these results demonstrate that NR4As are directly involved in transcriptional programs limiting cDC activation in vitro, at least in the culture-based experimental conditions used here. Accordingly, the loss of NR4A expression in SSc cDCs leads to an increased activation of these cells.

5

Activation of NR4As limits pro-inflammatory cytokine production in healthy and dcSSc cDCs

Since NR4As have been described as critical regulators of pro-inflammatory responses[30–35], which was also confirmed in our ChIP-seq analysis, we sought to identify whether activation of NR4A signaling could attenuate pro-inflammatory cytokine production by healthy and dcSSc cDCs. We first treated freshly isolated CD1c+ cDCs obtained from peripheral blood of healthy donors with increasing concentrations of DMSO (negative control), C-DIM5 (NR4A1 agonist), or C-DIM12 (NR4A2 agonist) and measured the gene expression of the housekeeping gene *GUSB* to evaluate the effect of NR4A activation on cDC viability. For C-DIM5 and C-DIM12, concentrations up to 10uM and 25uM were well tolerated (Figure 5A), thus for subsequent experiments, treatment concentrations of 10uM were used. Pre-incubation of cDCs with 10uM C-DIM5 and C-DIM12 before stimulation with R848, led to a significant decrease in IL-6 production, both on the mRNA and protein level (Figure 5B), confirming the suppressive role of NR4As in pro-inflammatory cytokine production. To substantiate these results, and to evaluate the effect of NR4A agonists on dcSSc cDCs, we repeated this experiment using PBMC samples obtained from 12 dcSSc patients and 11 matched healthy controls. Intracellular FACS staining and gating on CD1c+ cDCs (according to the gating strategy provided in Supplementary Figure 4) again showed that NR4A activation led to a significant decrease of IL-6 production by cDCs (Figure 5C). Notably, NR4A activation also led to a significant decrease of IL-6 production in dcSSc cDCs, demonstrating that NR4A activation can effectively attenuate pro-inflammatory cytokine production in these patients.

Activation of NR4As decreases the CD4+ T-cell stimulatory capacity of cDCs

Since cDCs are indispensable for CD4+ T-cell activation, and multiple lines of evidence imply an important role for CD4+ T-cells in SSc pathogenesis[38–41], we next investigated the role of NR4As in cDC priming for CD4+ T-cell activation. We performed co-cultures with healthy, autologous CD4+ T-cells and CD1c+ cDCs pretreated with C-DIM5 and C-DIM12 and evaluated the production of IFNγ by activated CD4+ T-cells by FACS (according to the gating strategy provided in Supplementary figure 5) after 3 days of co-culture followed by 6 hours PMA/Ionomycin stimulation (Figure 6A). cDCs pretreated with C-DIM5 and C-DIM12 were less capable of inducing IFNγ production by CD4+ T-cells compared to control cDCs pretreated with DMSO (Figure 6B). These results held true for cDCs stimulated with R848, showing that NR4A activation also attenuates CD4+ T-cell induction by cDCs under pro-inflammatory conditions (Figure 6C). These data show that, beside controlling the expression of pro-inflammatory cytokines, NR4As also have the capacity to control T-cell activation by CD1c+ cDCs.

DISCUSSION

In this study, we provide detailed transcriptomic profiling of CD1c+ cDCs from a cohort of 48 SSc patients and 12 healthy controls, which allowed us to characterize the transcriptomic landscape of cDCs in SSc. Using network analysis, we identified various clinically relevant modules of tightly co-expressed genes enriched in genes from
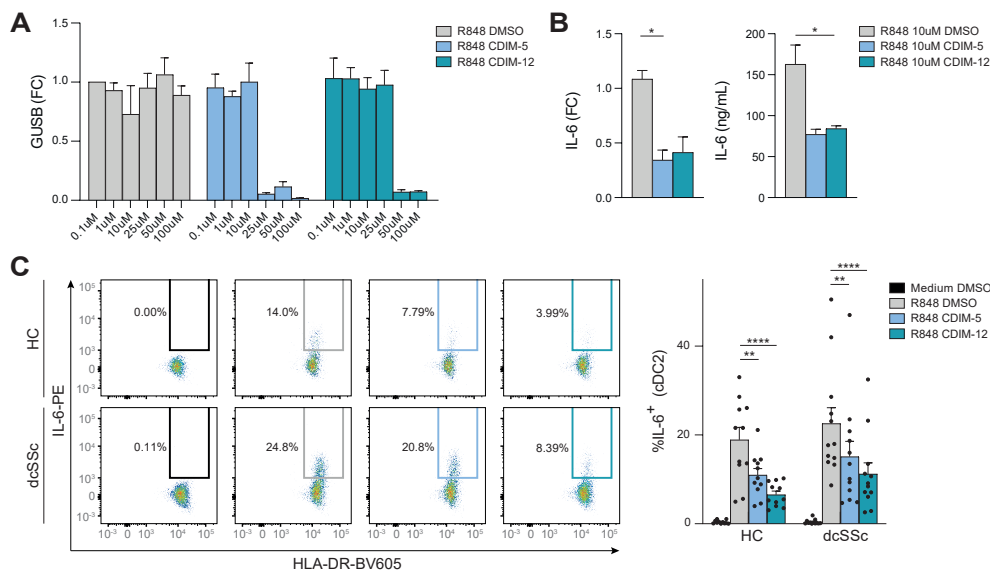
Figure 5. NR4A activation leads to a decrease in the production of IL-6 in healthy and dcSSc cDCs. (A) RT-qPCR of *GUSB* mRNA expression by freshly isolated cDCs after pre-incubation of increasing concentrations of DMSO (negative control), or NR4A agonists C-DIM5 and C-DIM12, followed by overnight stimulation with R848. Fold change (FC) in comparison to $1\mu M$ DMSO is shown. Data are shown as mean with SEM of 3 experiments. (B) *IL-6* mRNA (left) and protein expression (right) of cDCs pretreated with 10uM DMSO, C-DIM5 or C-DIM12, followed by overnight stimulation with R848. Relative mRNA expression levels (FC) shown are normalized to *GUSB* housekeeping levels. Data are shown as mean with SEM of 3 experiments. * = P <0.05, calculated by one-way Anova followed by Friedman test for multiple comparisons. (C) Representative FACS plots of the percentage of IL-6 positive cDC within the cDC (cDC2) fraction in PBMC cultures pretreated with 10uM DMSO, C-DIM5 or C-DIM12, followed by overnight stimulation with R848. Barplots (mean + SEM) depict the quantification of FACS data obtained from 12 healthy controls (HC) and 13 dcSSc patients. ** = P <0.01, **** = P <0.001, calculated by two-way Anova followed by Dunnett's test for multiple comparisons.

pathways highly relevant for SSc and cDC biology, including immune cell activation, antigen presentation, anti-viral mechanisms, and ribosomes. Pathways related to viral processes mainly included IFN inducible genes, which have previously been described to be highly expressed in SSc, referred to as the type I IFN signature[48]. The ribosome associated module included many ribosomal genes involved in translation, which were strongly downregulated in all subsets of SSc patients. Interestingly, downregulation of ribosomal protein mRNAs has previously been described in plasmacytoid DCs (pDCs) of primary Sjögren's syndrome patients[49], as well as in healthy pDCs and monocyte derived DCs (moDCs) activated by TLR ligands[49–51]. These data suggest that the downregulation of ribosomal protein mRNAs reflects an activated transcriptional phenotype of cDCs, highlighting the fact that circulating CD1c+ cDCs from SSc patients are transcriptionally primed towards an activated state, already from early stages of the disease.

We identified a module of immune regulatory genes, that is particularly downregulated in cDCs of dcSSc patients. Transcription factor network analysis pointed
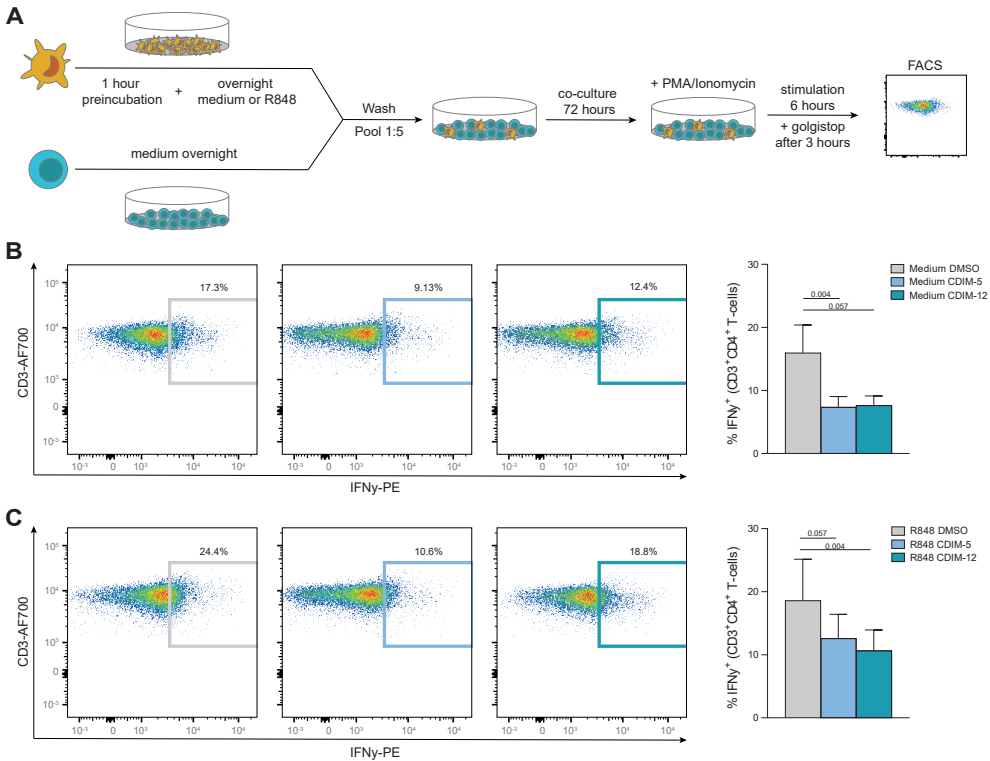
Figure 6. NR4A activation leads to a decrease in CD4+ T-cell activation by cDCs. (A) Schematic overview of the co-culture setup of freshly isolated autologous CD4+ T-cells and cDCs. (B) Representative FACS plots of the percentage of IFNγ positive T-cells within the CD4+ T-cell fraction after 3 days of co-culture with cDCs pretreated with 10uM DMSO, C-DIM5 or C-DIM12, followed by overnight culture in medium or (C) overnight stimulation with R848. Barplots (mean + SEM) depict the quantification of FACS data obtained from 5 donors. P-value in comparison to DMSO treatment is shown, calculated by one-way ANOVA followed by Friedman test for multiple comparisons.

towards the NR4A family of orphan nuclear receptors (*NR4A1*, *NR4A2* and *NR4A3*) as potentially important regulators of this module. NR4As are transcription factors that regulate gene expression in a ligand-independent manner, meaning that their activity is largely dependent on their expression levels and posttranslational modifications[52, 53]. These nuclear receptors have previously been shown to play critical roles in the regulation of immune cell activation[54–56], and NR4A1 in particular, has been established as regulator of pro-inflammatory responses in dendritic cells[33]. In line with these observations, we show that activation of NR4As in CD1c+ cDCs by selective agonists attenuates release of the pro-inflammatory cytokine IL-6 and downstream activation of CD4+ T-cells. Importantly, we show that, although the expression of *NR4A1-3* is downregulated in cDCs obtained from dcSSc patients, activation of NR4A signaling by NR4A agonists C-DIM5 and C-DIM12 inhibits IL-6 production, showing that modulation of NR4A activity can attenuate the pro-inflammatory phenotype displayed by dcSSc cDCs. Thus, small molecule agonists can overcome the reduced expression of

NR4A transcription factors in dcSSc cDCs and may represent very interesting targets for immunotherapy in SSc. Given the fact that SSc patients with early diffuse phenotypes have also previously been shown to display sign of enhanced dendritic cell activation[13], targeting NR4As early in SSc pathogenesis might prevent DC activation at early stages and limit disease progression.

The factors that underlie the downregulated expression of NR4As in dcSSc cDCs remain to be resolved. However, our experiments do provide new insights. In line with the roles of NR4As as immediate early response genes[52], stimulation of freshly isolated BCA1+ cDCs from healthy donors with TLR ligands R848 and LPS, as well as hypoxia largely induced the expression of *NR4A1-3*. Also, cytokines known to be increased in peripheral blood of SSc patients or related to SSc pathogenesis, including CXCL4, IFNα, TGFβ, GM-CSF, IL-6 and IL-15 did not reduce NR4A expression, at least for the concentrations and time points included here. Moreover, stimulation of freshly isolated CD1c+ cDCs from dcSSc patients with TLR7/8 ligand R848 also led to an induction of NR4A expression, comparable to the levels in healthy donors, suggesting that the upstream transcriptional regulation of NR4As is not defective in dcSSc cDCs. Given the well described heterogeneity of the CD1c+ cDC subset[57, 58], one might propose that the downregulation of NR4A expression that we observed in bulk CD1c+ cDCs from dcSSc patients is actually reflective of a disbalance of distinct cDC populations within the CD1c+ subset as compared to healthy controls. Indeed the expression of *NR4A2* and *NR4A3* have previously been shown to be low in CD1c+Tbet- cDC2Bs, an inflammatory DC population within the CD1c+ cDC compartment[59]. However, recent analysis of the composition of CD1c+ cDC subsets in the blood of SSc patients by Dutertre *et al.*, showed that the proportion of distinct CD1c+ cDC subpopulations in SSc was not significantly different from healthy[58]. Thus, the downregulation of NR4As that we observed in dcSSc CD1c+ cDCs is not likely to be attributed to heterogeneity within the DC compartment. Thus, it remains to be investigated what exact molecular mechanisms cause NR4A downregulation in dcSSc cDCs. These might include alterations in the chromatin landscape or regulation at the post-transcriptional level.

To unravel exactly how NR4As regulate the transcriptional programs leading to cDC activation and dysregulation in SSc, we performed genome-wide ChIP-sequencing analysis of *NR4A1-3* binding in resting and activated CD1c+ cDCs. We identified numerous genes that were characterized by binding of NR4A at their promoters. Importantly, a large part of the differentially expressed genes that we found in dcSSc in our RNA-sequencing analysis also displayed binding of NR4A at their promoters in our ChIP-sequencing analysis, suggesting that NR4As are strongly involved in transcriptional programs underlying dendritic cell dysregulation in SSc. Additionally, we identified genes involved in dendritic cell morphology and, under stimulated conditions, ECM production that were directly bound by NR4As. Given the reduced expression of NR4As in circulating CD1c+ cDCs from dcSSc patients, these results suggest that dcSSc cDCs might show an enhanced expression of ECM related genes once they get stimulated, for example upon migration to the skin. Interestingly, inflammatory DCs have been proposed to contribute to fibrosis in SSc through the increased secretion of ECM molecules and promotion of myofibroblast differentiation[60], suggesting that besides driving a pro-

5

inflammatory phenotype, downregulation of NR4As might also drive dcSSc cDCs to a pro-fibrotic phenotype. Although more detailed ChIP-sequencing analyses to quantify NR4A binding in dcSSc in comparison to healthy donor cDCs are needed to validate these results, our analysis points towards NR4As as major transcriptional regulators of pathways implicated in cDC dyregulation in SSc.

In conclusion, we show that the NR4A transcription factor family members NR4A1, NR4A2 and NR4A3 are important regulators underlying DC dysregulation in SSc. We propose that pharmacological activation of NR4As is an attractive therapeutic option to attenuate pro-inflammatory and pro-fibrotic responses in SSc patients, as such or if necessary using DC-targeting based approaches.

REFERENCES

1. Gabrielli, E. V. Avvedimento, T. Krieg. Scleroderma. *N. Engl. J. Med.* vol. 360, no. 19, pp. 1989–2003, May 2009.
2. F. van den Hoogen, D. Khanna, J. Fransen, *et al*. 2013 classification criteria for systemic sclerosis: an American college of rheumatology/European league against rheumatism collaborative initiative. *Ann. Rheum. Dis*. vol. 72, no. 11, pp. 1747–1755. Nov. 2013.
3. E. C. LeRoy, T. A. Medsger. Criteria for the classification of early systemic sclerosis. *J. Rheumatol*. vol. 28, no. 7, pp. 1573–1576. Jul. 2001.
4. S. Guiducci, R. Giacomelli, M. M. Cerinic. Vascular complications of scleroderma. *Autoimmun. Rev*. vol. 6, no. 8, pp. 520–523, Sep. 2007.
5. M. Cutolo, S. Soldano, V. Smith. Pathophysiology of systemic sclerosis: current understanding and new insights. *Expert Rev. Clin. Immunol*. vol. 15, no. 7, pp. 753–764. Jul. 2019.
6. V. S. Rajkumar, K. Howell, K. Csiszar, *et al*. Shared expression of phenotypic markers in systemic sclerosis indicates a convergence of pericytes and fibroblasts to a myofibroblast lineage in fibrosis. *Arthritis Res. Ther.* vol. 7, no. 5, pp. R1113–R1123. Jul. 2005
7. N. Binai, S. O'Reilly, B. Griffiths, *et al*. Differentiation potential of CD14 + monocytes into myofibroblasts in patients with systemic sclerosis. *PLoS One*. vol. 7, no. 3, pp. 1–7. Mar. 2012.
8. P. Guermonprez, J. Valladeau, L. Zitvogel, *et al*. Antigen presentation and T cell stimulation by dendritic cells. *Annu. Rev. Immunol*. vol. 20, no. 1, pp. 621–667. Apr. 2002.
9. T.C. Tzeng, S. Chyou, S. Tian, *et al.*, CD11c hi Dendritic Cells Regulate the Re-establishment of Vascular Quiescence and Stabilization after Immune Stimulation of Lymph Nodes. *J. Immunol*. vol. 184, no. 8, pp. 4247–4257. Apr. 2010.
10. L. Ding, T. Liu, Z. Wu, *et al*. Bone Marrow CD11c + Cell–Derived Amphiregulin Promotes Pulmonary Fibrosis. *J. Immunol*. vol. 197, no. 1, pp. 303–312. Jul. 2016.
11. A. L. Mathes, R.B. Christmann, G. Stifano, *et al*. Global chemokine expression in systemic sclerosis (SSc): CCL19 expression correlates with vascular inflammation in SSc skin. *Ann. Rheum. Dis.*, vol. 73, no. 10, pp. 1864–1872. Oct. 2014.
12. S. Mokuda et al. CD1a+ survivin+ dendritic cell infiltration in dermal lesions of systemic sclerosis. Arthritis Res. Ther. vol. 17, no. 1, pp. 275. Dec. 2015.
13. L. van Bon, C. Popa, R. Huijbens, et al. Distinct evolution of TLR-mediated dendritic cell cytokine secretion in patients with limited and diffuse cutaneous systemic sclerosis. *Ann. Rheum. Dis.*, vol. 69, no. 8, pp. 1539–1547, Aug. 2010.
14. T. Carvalheiro, M. Zimmermann, T. R. D. J. Radstake, *et al*. Novel insights into dendritic cells in the pathogenesis of systemic sclerosis. *Clin. Exp. Immunol*. vol. 201, no. 1, pp. 25–33. Jul. 2020.
15. F. van den Hoogen, D. Khanna, J. Fransen, *et al*. 2013 classification criteria for systemic sclerosis: an American college of rheumatology/European league against rheumatism collaborative initiative. *Ann. Rheum. Dis*. vol. 72, no. 11, pp. 1747–1755. Nov. 2013.
16. A. Dobin, C.A. Davis, F. Schlesinger, *et al*. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. vol. 29, no. 1, pp. 15–21. Jan. 2013.
17. S. Anders, P. T. Pyl, W. Huber. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics*. vol. 31, no. 2, pp. 166–169. Jan. 2015.
18. D. Risso, K. Schwartz, G. Sherlock, *et al*. GC-Content Normalization for RNA-Seq Data. *BMC Bioinformatics*. vol. 12, no. 1, pp. 480. 2011.
19. D. Risso, J. Ngai, T. P. Speed, *et al*. Normalization of RNA-seq data using factor analysis of control genes or samples. *Nat. Biotechnol*. vol. 32, no. 9, pp. 896–902. Sep. 2014.
20. M. I. Love, W. Huber, S. Anders. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* vol. 15, no. 12, pp. 550. Dec. 2014.
21. P. Langfelder, S. Horvath. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*. vol. 9, no. 1, pp. 559. Dec. 2008.
22. B. Langmead, S. L. Salzberg. Fast gapped-read alignment with Bowtie 2. *Nat. Methods*. vol. 9, no. 4, pp. 357–359. Apr. 2012.
23. J. M. Gaspar. Improved peak-calling with MACS2. *bioRxiv*. 2018
24. H. M. Amemiya, A. Kundaje, A. P. Boyle. The ENCODE Blacklist: Identification of Problematic Regions of the Genome. *Sci. Rep.* vol. 9, no. 1, pp. 9354. Dec. 2019.
25. G. Yu, L.-G. Wang, Q.-Y. He. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics*. vol. 31, no. 14, pp. 2382–2383. Jul. 2015.

5

26. H. Wickham. ggplot2: Elegant Graphics for Data Analysis. *Springer-Verlag New York*, 2016.

27. G. Yu, L.-G. Wang, Y. Han, *et al.* clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. *Omi. A J. Integr. Biol.* vol. 16, no. 5, pp. 284–287. May 2012 .

28. J. Haub, N. Roehrig, P. Uhrin, *et al.* Intervention of Inflammatory Monocyte Activity Limits Dermal Fibrosis. *J. Invest. Dermatol.* vol. 139, no. 10, pp. 2144–2153. Apr. 2019.

29. X. Zeng, Z. Yue, Y. Gao, *et al.* NR4A1 is Involved in Fibrogenesis in Ovarian Endometriosis. *Cell. Physiol. Biochem.* vol. 46, no. 3, pp. 1078–1090. Apr. 2018.

30. S. Mahajan, A. Saini, V. Chandra, *et al.* Nuclear Receptor Nr4a2 Promotes Alternative Polarization of Macrophages and Confers Protection in Sepsis. *J. Biol. Chem.* vol. 290, no. 30, pp. 18304–18314. Jul. 2015.

31. N. Tel-Karthaus, E. D. Kers-Rebel, M. W. Looman, *et al.* Nuclear Receptor Nur77 Deficiency Alters Dendritic Cell Function. *Front. Immunol.* vol. 9, pp. 3389. Aug. 2018.

32. K. Palumbo-Zerr, P. Zerr, A. Distler, *et al.* Orphan nuclear receptor NR4A1 regulates transforming growth factor-β signaling and fibrosis. *Nat. Med.* vol. 21, no. 2, pp. 150–158. Feb. 2015.

33. C. Ma, L. Wu, L. Song, *et al.* The pro-inflammatory effect of NR4A3 in osteoarthritis. *J. Cell. Mol. Med.* vol. 24, no. 1, pp. 930–940. Jan. 2020.

34. J. L. Silverstein. Cutaneous Hypoxia in Patients With Systemic Sclerosis (Scleroderma). *Arch. Dermatol.* vol. 124, no. 9, pp. 1379. Sep. 1988.

35. T. W. van Hal, L. van Bon, T. R. D. J. Radstake. A System Out of Breath: How Hypoxia Possibly Contributes to the Pathogenesis of Systemic Sclerosis. *Int. J. Rheumatol.* vol. 2011, pp. 1–7. Nov. 2011.

36. J. B. Eells, J. Wilcots, S. Sisk, *et al.* NR4A Gene Expression Is Dynamically Regulated in the Ventral Tegmental Area Dopamine Neurons and Is Related to Expression of Dopamine Neurotransmission Genes. *J. Mol. Neurosci.* vol. 46, no. 3, pp. 545–553. Mar. 2012.

37. L. Medzikovic, C. J. M. de Vries, V. de Waard. NR4A nuclear receptors in cardiac remodeling and neurohormonal regulation. *Trends Cardiovasc. Med.* vol. 29, no. 8, pp. 429–437. Nov. 2019.

38. F. Vogel, E. Hartmann, D. Gorlich, and T. A. Rapoport, "Segregation of the signal sequence receptor protein in the rough endoplasmic reticulum membrane," Eur. J. Cell Biol., vol. 53, no. 2, pp. 197–202, 1990.

39. M. Chatterjee, T. Rauen, K. Kis-Toth, *et al.* Increased Expression of SLAM Receptors SLAMF3 and SLAMF6 in Systemic Lupus Erythematosus T Lymphocytes Promotes Th17 Differentiation. *J. Immunol.* vol. 188, no. 3, pp. 1206–1212. Feb. 2012.

40. C. A. Dinarello, D. Novick, S. Kim, et al. Interleukin-18 and IL-18 Binding Protein. *Front. Immunol.*, vol. 4, pp. 289. Oct. 2013.

41. J. L. Maravillas-Montero, L. Santos-Argumedo. The myosin family: unconventional roles of actin-dependent molecular motors in immune cells. *J. Leukoc. Biol.* vol. 91, no. 1, pp. 35–46. Jan. 2012.

42. A. Kalogerou, E. Gelou, S. Mountantonakis, *et al.* Early T cell activation in the skin from patients with systemic sclerosis. *Ann. Rheum. Dis.,* vol. 64, no. 8, pp. 1233–1235. Aug. 2005.

43. G. Papp, I.F. Horvath, S. Barath, *et al.* Altered T-cell and regulatory cell repertoire in patients with diffuse cutaneous systemic sclerosis. *Scand. J. Rheumatol.* vol. 40, no. 3, pp. 205–210. May 2011.

44. S. O'Reilly, T. Hugle, J. M. van Laar. T cells in systemic sclerosis: a reappraisal. *Rheumatology*. vol. 51, no. 9, pp. 1540–1549. Sep. 2012.

45. N. H. Servaas, F. Zaaraoui-Boutahar, C.G.K. Wichers, *et al.* Longitudinal analysis of T-cell receptor repertoires reveals persistence of antigen-driven CD4+ and CD8+ T-cell clusters in systemic sclerosis. *J. Autoimmun.* vol. 117, pp. 102574. Feb. 2021.

46. Z. Brkic, L. van Bon, M. Cossu, *et al.* The interferon type I signature is present in systemic sclerosis before overt fibrosis and might contribute to its pathogenesis through high BAFF gene expression and high collagen synthesis. *Ann. Rheum. Dis.* vol. 75, no. 8, pp. 1567–1573. Aug. 2016.

47. M. R. Hillen, A. Pandit, S.L.M. Blokland, *et al.* Plasmacytoid DCs From Patients With Sjögren's Syndrome Are Transcriptionally Primed for Enhanced Pro-inflammatory Cytokine Production. *Front. Immunol.* vol. 10, pp. 2096, Sep. 2019.

48. T. S. M. Mathan, J. Textor, A.E. Sköld, *et al.* Harnessing RNA sequencing for global, unbiased evaluation of two new adjuvants for dendritic-cell immunotherapy. *Oncotarget.* vol. 8, no. 12, pp. 19879–19893. Mar. 2017.

49. M. Ceppi, G. Clavarino, E. Gatti, *et al.* Ribosomal protein mRNAs are translationally-regulated during human dendritic cells activation by LPS. *Immunome Res.* vol. 5, no. 1, pp. 5. Nov. 2009.

50. M. A. Maxwell, G.E.O. Muscat. The NR4A Subgroup: Immediate Early Response Genes with

Pleiotropic Physiological Roles. *Nucl. Recept. Signal*. vol. 4, no. 1, pp. nrs.04002. Jan. 2006.

51. T.J. Fahrner, S.L. Carroll, J. Milbrandt. The NGFI-B protein, an inducible member of the thyroid/steroid receptor family, is rapidly modified posttranslationally. *Mol. Cell. Biol.* vol. 10, no. 12, pp. 6454–6459. Dec. 1990.

52. D.S. Koenis, L. Medzikovic, P.B. van Loenen, *et al*. Nuclear Receptor Nur77 Limits the Macrophage Inflammatory Response through Transcriptional Reprogramming of Mitochondrial Metabolism. *Cell Rep*. vol. 24, no. 8, pp. 2127-2140.e7. Aug. 2018.

53. J. Chen, I.F. López-Moyado, *et al*. NR4A transcription factors limit CAR T cell function in solid tumours.*Nature*. vol. 567, no. 7749, pp. 530–534. Mar. 2019.

54. C. McEvoy, M. de Gaetano, H.E. Giffney, *et al.* NR4A Receptors Differentially Regulate NF-κB Signaling in Myeloid Cells. *Front. Immunol*. vol. 8, pp. 7. Jan. 2017.

55. A.-C. Villani, R. Satija, G. Reyolds, *et al*. Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. *Science*. vol. 356, no. 6335, pp. eaah4573. Apr. 2017.

56. C.-A. Dutertre, E. Becht, S.E. Irac, *et al*. Single-Cell Analysis of Human Mononuclear Phagocytes Reveals Subset-Defining Markers and Identifies Circulating Inflammatory Dendritic Cells. *Immunity*. vol. 51, no. 3, pp. 573-589.e8. Sep. 2019.

57. C.C. Brown, H. Gudjonson, Y. Pritykin, *et al*. Transcriptional Basis of Mouse and Human Dendritic Cell Heterogeneity. *Cell*. vol. 179, no. 4, pp. 846-863.e24. Oct. 2019.

58. S.C. Silva-Cardoso, W. Tao, C. Angiolilli, *et al*. CXCL4 Links Inflammation and Fibrosis by Reprogramming Monocyte-Derived Dendritic Cells in vitro. *Front. Immunol*. vol. 11, pp. 2149. Sep. 2020.
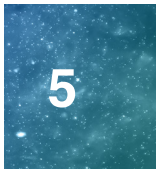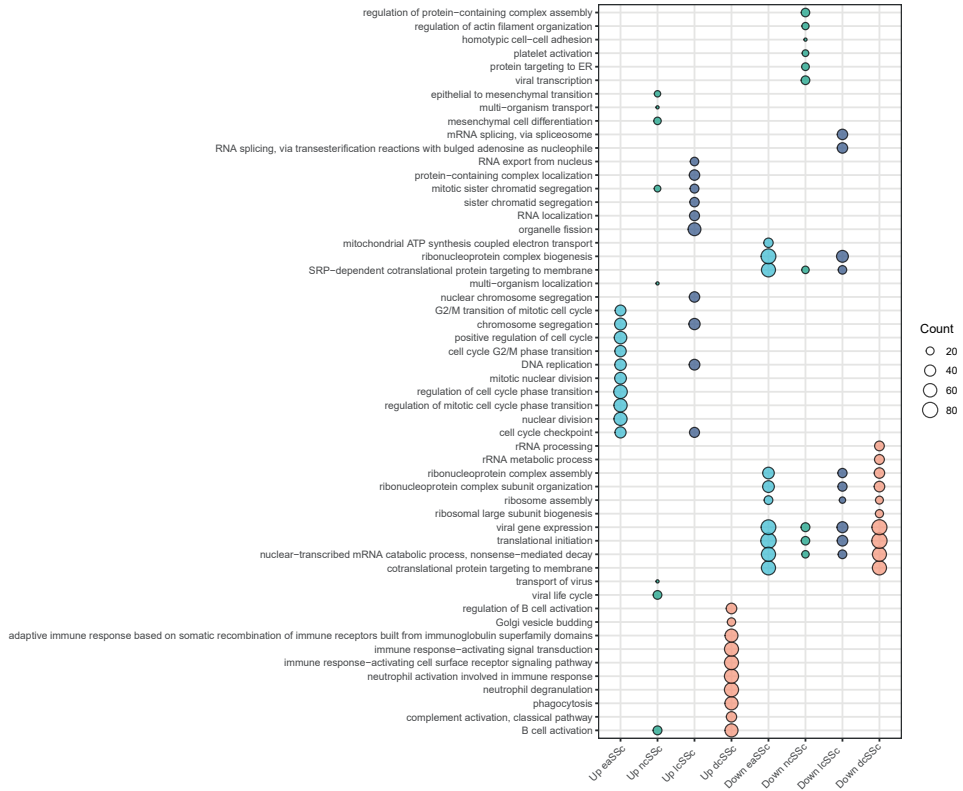
5

SUPPLEMENTARY INFORMATION

Supplementary Figure 1. Pathway enrichment analysis of differentially expressed genes identified in SSc patients versus healthy controls. Circles size denotes the number of differentially expressed genes genes associated to enriched pathways. Top 10 pathways are shown (B&H corrected p-value < 0.05).
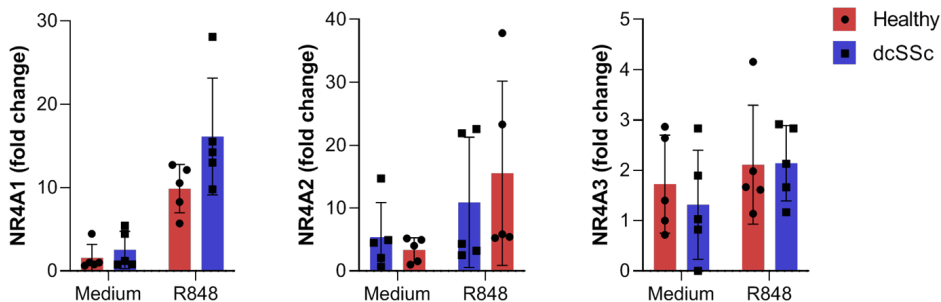
Supplementary Figure 2. RT-qPCR of NR4A expression in stimulated CD1c+ cDCs from healthy controls and dcSSc patients. *NR4A1*, *NR4A2* and *NR4A3* for untreated (medium) or treated (R848) cDCs following 18 hours of culture. Relative mRNA expression levels (FC) shown are normalized to *GUSB* housekeeping levels. Bars depict mean, error bars depict SEM.
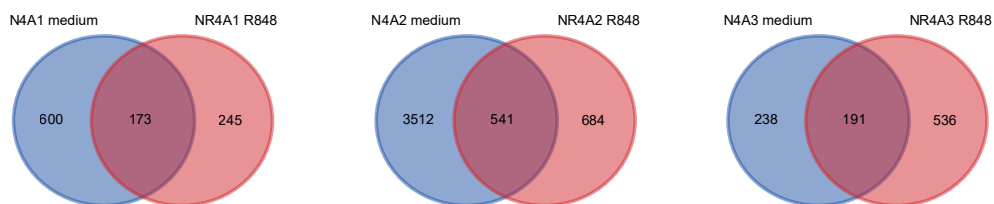
Supplementary Figure 3. Genome-wide identification of NR4A binding sites in resting and activated CD1c + cDCs. Venn diagrams depicting the overlap of NR4A1, NR4A2 and NR4A3 transcription factor binding sites (within 10kb of the nearest gene promoter region) in resting (blue, medium) and activated (red, R848) cDCs.
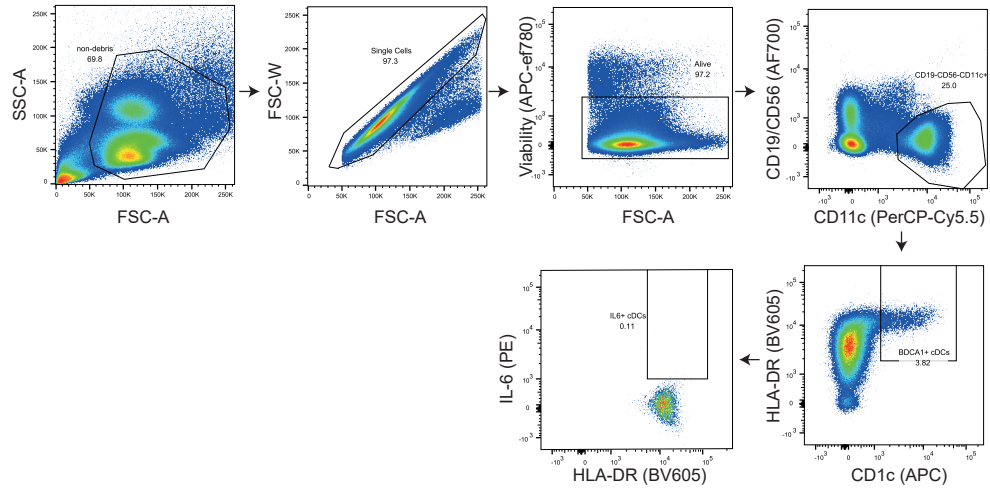
Supplementary Figure 4. Gating strategy for measuring IL-6 production in the CD1c+ cDC fraction in PBMC cultures.
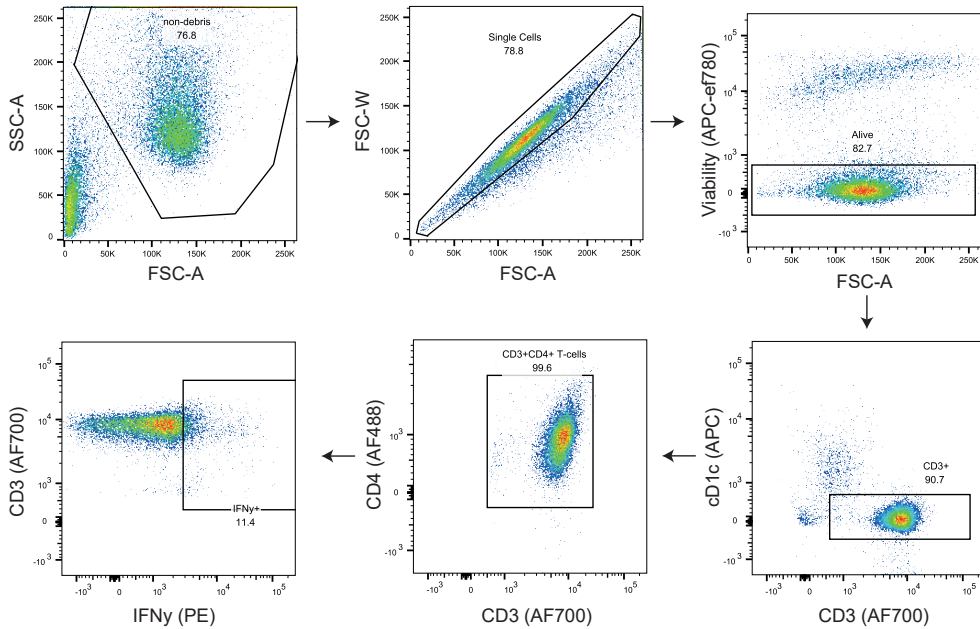
Supplementary Figure 5. Gating strategy for measuring IFNy production by CD4+ T-cells in cDC/T-cell co-cultures.

Gene	Forward	Reverse
GUSB	CACCAGGGACCATCCAATACC	GCAGTCCAGCGTAGTTGAAAAA
NR4A1	ACTGCCCTGTGGACAAGAG	CTGTTCGGACAACTTCCTTC
NR4A2	GGACTCCCCATTGCTTTTC	AGGCGAGGACCCATACTG
NR4A3	CTGCCCAGTAGACAAGAGAC	CTCCTCCCTTTCAGACTATC
IL-6	GGCACTGGCAGAAAACAACC	GCAAGTCTCCTCATTGAATCC

Supplementary Table 1. Gene specific primers used for RT-qPCR

	Antibody	Label	Clone	Company
PBMC panel	Fixable viability	APC-ef780	N.A.	eBioscience
	Surface CD11c	PerCP-Cy5.5	3.9	Thermo Scientific
	Surface CD1c	APC	AD5-8E7	Miltenyi
	Surface CD19	AF700	HIB19	eBioscience
	Surface CD56	AF700	B159	BD
	Surface CD3	PE/Cy7	UCHT1	Biolegend
	Surface CD4	BV510	RPA-T4	Biolegend
	Surface HLA-DR	BV605	G46-6	BD
	Surface CD14	BV785	M5E2	Biolegend
	Intracellular IL-6	PE	MQ2-6A3	BD
Co-culture panel	Fixable viability	APC-ef780	N.A.	eBioscience
	Surface CD3	AF700	UCHT1	Biolegend
	Surface CD4	AF488	RPA-T4	BD
	Surface CD1c	APC	AD5-8E7	Miltenyi
	Intracellular IFNy	PE	4S.B3	BD

Supplementary Table 2. Antibody panels for used for flow cytometry
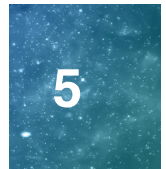
5

Subset	Description	GeneRatio	p.adjust
Up dcSSc	complement activation, classical pathway	34/1134	5.13E-09
Up dcSSc	B cell activation	54/1134	5.13E-09
Up dcSSc	humoral immune response mediated by circulating immunoglobulin	35/1134	6.14E-09
Up dcSSc	phagocytosis	57/1134	6.46E-08
Up dcSSc	neutrophil degranulation	68/1134	6.46E-08
Up dcSSc	neutrophil activation involved in immune response	68/1134	6.55E-08
Up dcSSc	neutrophil mediated immunity	69/1134	6.55E-08
Up dcSSc	immune response-activating cell surface receptor signaling pathway	66/1134	9.65E-08
Up dcSSc	neutrophil activation	68/1134	1.21E-07
Up dcSSc	Golgi vesicle budding	22/1134	4.63E-07
Up dcSSc	regulation of B cell activation	35/1134	4.64E-07
Up dcSSc	lymphocyte mediated immunity	52/1134	7.43E-07
Up dcSSc	vesicle budding from membrane	24/1134	2.13E-06
Up dcSSc	vesicle organization	48/1134	2.75E-06
Up dcSSc	protein activation cascade	35/1134	2.75E-06
Up dcSSc	leukocyte migration	64/1134	2.88E-06
Up dcSSc	regulation of complement activation	25/1134	4.73E-06
Up dcSSc	regulation of protein activation cascade	25/1134	5.44E-06
Up dcSSc	Golgi vesicle transport	50/1134	1.69E-05
Up dcSSc	regulation of humoral immune response	26/1134	2.34E-05
Down dcSSc	cotranslational protein targeting to membrane	68/859	1.26E-59
Down dcSSc	establishment of protein localization to endoplasmic reticulum	68/859	1.79E-55
Down dcSSc	nuclear-transcribed mRNA catabolic process, nonsense-mediated decay	67/859	8.72E-55
Down dcSSc	translational initiation	81/859	4.95E-54
Down dcSSc	viral gene expression	78/859	6.62E-51
Down dcSSc	viral transcription	74/859	4.04E-49
Down dcSSc	ribosome assembly	20/859	8.46E-10
Down dcSSc	ribosomal large subunit biogenesis	20/859	9.42E-09
Down dcSSc	ribonucleoprotein complex subunit organization	39/859	4.56E-07
Down dcSSc	rRNA processing	30/859	1.10E-05
Down dcSSc	positive regulation of release of cytochrome c from mitochondria	9/859	0.000938357
Down dcSSc	ciliary basal body-plasma membrane docking	16/859	0.00109231
Down dcSSc	regulation of ubiquitin protein ligase activity	7/859	0.006110243
Down dcSSc	positive regulation of hemopoiesis	21/859	0.015943419
Down dcSSc	myeloid cell homeostasis	18/859	0.017807193
Down dcSSc	positive regulation of leukocyte cell-cell adhesion	23/859	0.022355059
Down dcSSc	T cell activation	39/859	0.025806037
Down dcSSc	regulation of lymphocyte differentiation	19/859	0.033649155
Down dcSSc	protein insertion into mitochondrial membrane involved in apoptotic signaling pathway	7/859	0.036411064
Down dcSSc	cellular response to osmotic stress	8/859	0.041013014
Up eaSSc	nuclear division	58/861	4.01E-11
Up eaSSc	DNA replication	46/861	4.01E-11
Up eaSSc	mitotic nuclear division	45/861	4.01E-11

Up eaSSc	regulation of cell cycle phase transition	62/861	1.82E-10
Up eaSSc	cell cycle checkpoint	39/861	1.82E-10
Up eaSSc	regulation of mitotic cell cycle phase transition	59/861	1.82E-10
Up eaSSc	cell cycle G2/M phase transition	42/861	1.56E-09
Up eaSSc	sister chromatid segregation	34/861	3.90E-09
Up eaSSc	regulation of chromosome organization	46/861	2.28E-08
Up eaSSc	DNA damage checkpoint	27/861	1.39E-07
Up eaSSc	covalent chromatin modification	51/861	3.44E-06
Up eaSSc	histone modification	49/861	5.33E-06
Up eaSSc	negative regulation of mitotic cell cycle	39/861	2.17E-05
Up eaSSc	regulation of mitotic nuclear division	25/861	2.17E-05
Up eaSSc	mRNA export from nucleus	20/861	2.17E-05
Up eaSSc	mRNA-containing ribonucleoprotein complex export from nucleus	20/861	2.17E-05
Up eaSSc	regulation of nuclear division	27/861	2.17E-05
Up eaSSc	mitotic cell cycle checkpoint	25/861	2.17E-05
Up eaSSc	multi-organism transport	16/861	2.17E-05
Up eaSSc	multi-organism localization	16/861	2.17E-05
Down eaSSc	cotranslational protein targeting to membrane	68/723	3.74E-65
Down eaSSc	SRP-dependent cotranslational protein targeting to membrane	67/723	3.74E-65
Down eaSSc	establishment of protein localization to endoplasmic reticulum	69/723	3.26E-62
Down eaSSc	nuclear-transcribed mRNA catabolic process, nonsense-mediated decay	68/723	1.90E-61
Down eaSSc	translational initiation	82/723	2.02E-61
Down eaSSc	viral gene expression	74/723	9.60E-52
Down eaSSc	viral transcription	70/723	1.20E-49
Down eaSSc	ribosome biogenesis	52/723	9.00E-18
Down eaSSc	ribonucleoprotein complex assembly	49/723	6.82E-17
Down eaSSc	ribonucleoprotein complex subunit organization	50/723	1.00E-16
Down eaSSc	mitochondrial ATP synthesis coupled electron transport	28/723	4.96E-15
Down eaSSc	respiratory electron transport chain	29/723	1.11E-13
Down eaSSc	rRNA processing	33/723	1.53E-09
Down eaSSc	purine nucleoside triphosphate metabolic process	38/723	6.21E-07
Down eaSSc	mitochondrial gene expression	23/723	6.45E-06
Down eaSSc	mitochondrial respiratory chain complex assembly	17/723	1.22E-05
Down eaSSc	neutrophil activation	43/723	6.25E-05
Down eaSSc	cellular protein complex disassembly	25/723	8.43E-05
Down eaSSc	neutrophil activation involved in immune response	41/723	0.000206438
Down eaSSc	neutrophil mediated immunity	40/723	0.000807639
Up lcSSc	organelle fission	54/824	1.02E-07
Up lcSSc	chromosome segregation	41/824	1.53E-06
Up lcSSc	mitotic sister chromatid segregation	25/824	1.03E-05
Up lcSSc	DNA replication	35/824	1.23E-05
Up lcSSc	mitotic nuclear division	34/824	1.26E-05
Up lcSSc	RNA localization	29/824	0.00015651
Up lcSSc	regulation of chromosome segregation	18/824	0.000211293
Up lcSSc	protein-containing complex localization	32/824	0.000338403
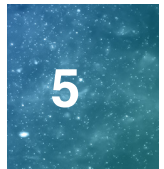
5

Up lcSSc	RNA export from nucleus	20/824	0.000657521
Up lcSSc	regulation of mitotic cell cycle phase transition	42/824	0.000931009
Up lcSSc	spindle organization	22/824	0.001514295
Up lcSSc	mRNA-containing ribonucleoprotein complex export from nucleus	17/824	0.001612556
Up lcSSc	ribonucleoprotein complex localization	18/824	0.002488175
Up lcSSc	meiotic cell cycle	27/824	0.002702056
Up lcSSc	regulation of chromosome organization	33/824	0.00366235
Up lcSSc	mitotic cell cycle checkpoint	20/824	0.005885022
Up lcSSc	DNA damage checkpoint	18/824	0.009383219
Up lcSSc	meiotic nuclear division	20/824	0.009383219
Up lcSSc	multi-organism transport	12/824	0.009383219
Up lcSSc	multi-organism localization	12/824	0.009383219
Down lcSSc	viral gene expression	39/655	1.19E-15
Down lcSSc	translational initiation	36/655	3.09E-13
Down lcSSc	viral transcription	34/655	5.75E-13
Down lcSSc	SRP-dependent cotranslational protein targeting to membrane	23/655	1.18E-09
Down lcSSc	nuclear-transcribed mRNA catabolic process, nonsense-mediated decay	24/655	1.94E-09
Down lcSSc	establishment of protein localization to endoplasmic reticulum	24/655	2.51E-09
Down lcSSc	ribosomal large subunit biogenesis	16/655	7.08E-07
Down lcSSc	ribosome assembly	14/655	6.35E-06
Down lcSSc	ribonucleoprotein complex subunit organization	29/655	9.36E-05
Down lcSSc	RNA splicing, via transesterification reactions with bulged adenosine as nucleophile	34/655	0.000107062
Down lcSSc	mRNA splicing, via spliceosome	34/655	0.000107062
Down lcSSc	oxidative phosphorylation	16/655	0.007197812
Down lcSSc	pentose-phosphate shunt	5/655	0.015245683
Down lcSSc	regulation of cellular amine metabolic process	11/655	0.015245683
Down lcSSc	neutrophil activation	34/655	0.019838415
Down lcSSc	glucose 6-phosphate metabolic process	6/655	0.019838415
Down lcSSc	release of cytochrome c from mitochondria	9/655	0.021155956
Down lcSSc	respiratory electron transport chain	13/655	0.024382183
Down lcSSc	neutrophil activation involved in immune response	33/655	0.026577717
Down lcSSc	mitochondrial translational termination	11/655	0.027366409
Up ncSSc	B cell activation	24/517	0.028037426
Up ncSSc	multi-organism transport	10/517	0.041913325
Up ncSSc	multi-organism localization	10/517	0.041913325
Up ncSSc	transport of virus	9/517	0.041913325
Up ncSSc	viral life cycle	23/517	0.041913325
Up ncSSc	ferric iron transport	7/517	0.044980462
Up ncSSc	trivalent inorganic cation transport	7/517	0.044980462
Up ncSSc	mitotic sister chromatid segregation	14/517	0.044980462
Up ncSSc	regulation of lymphocyte activation	29/517	0.045390335
Up ncSSc	NLS-bearing protein import into nucleus	5/517	0.045390335
Down ncSSc	viral gene expression	25/421	3.50E-09
Down ncSSc	viral transcription	24/421	3.50E-09

Down ncSSc	protein targeting to ER	19/421	1.99E-08
Down ncSSc	SRP-dependent cotranslational protein targeting to membrane	18/421	1.99E-08
Down ncSSc	establishment of protein localization to endoplasmic reticulum	19/421	2.10E-08
Down ncSSc	translational initiation	23/421	3.73E-08
Down ncSSc	nuclear-transcribed mRNA catabolic process, nonsense-mediated decay	18/421	8.78E-08
Down ncSSc	platelet activation	15/421	0.000543591
Down ncSSc	homotypic cell-cell adhesion	9/421	0.017107301
Down ncSSc	negative regulation of viral transcription	5/421	0.03209776

Supplementary Table 3. GO-term enrichment analysis of genes dysregulated in SSc cDCs.

5
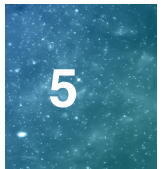
Part 2

**Adaptive immunity in
Systemic Sclerosis**

Chapter 6

Longitudinal analysis of T-cell receptor repertoires reveals persistence of antigen-driven CD4+ and CD8+ T-cell clusters in systemic sclerosis

Authors:

N.H. Servaas[1,2], F. Zaaraoui-Boutahar[3], C.G.K. Wichers[1,2], A. Ottria[1,2], E. Chouri[1,2], A.J. Affandi[1,2], S. Silva-Cardoso[1,2], M. van der Kroef[1,2], T. Carvalheiro[1,2], F. van Wijk[2], T.R.D.J. Radstake[1,2], A.C. Andeweg[3], A. Pandit[1,2]

Affiliations:

1. Department of Rheumatology and Clinical Immunology, University Medical Center Utrecht, Utrecht, Netherlands.
2. Center for Translational Immunology, University Medical Center Utrecht, Utrecht University, Utrecht, the Netherlands.
3. Department of Vioscience, Erasmus Medical Center Rotterdam, Rotterdam, the Netherlands.

ABSTRACT

The T-cell receptor (TCR) is a highly polymorphic surface receptor that allows T-cells to recognize antigenic peptides presented on the major histocompatibility complex (MHC). Changes in the TCR repertoire have been observed in several autoimmune conditions, and these changes are suggested to predispose autoimmunity. Multiple lines of evidence have implied an important role for T-cells in the pathogenesis of Systemic Sclerosis (SSc), a complex autoimmune disease. One of the major questions regarding the roles of T-cells is whether expansion and activation of T-cells observed in the diseases pathogenesis is antigen driven.

To investigate the temporal TCR repertoire dynamics in SSc, we performed high-throughput sequencing of CD4+ and CD8+ TCRβ chains on longitudinal samples obtained from four SSc patients collected over a minimum of two years. Repertoire overlap analysis revealed that samples taken from the same individual over time shared a high number of TCRβ sequences, indicating a clear temporal persistence of the TCRβ repertoire in CD4+ as well as CD8+ T-cells. Moreover, the TCRβs that were found with a high frequency at one time point were also found with a high frequency at the other time points (even after almost four years), showing that frequencies of dominant TCRβs are largely consistent over time. We also show that TCRβ generation probability and observed TCR frequency are not related in SSc samples, showing that clonal expansion and persistence of TCRβs is caused by antigenic selection rather than convergent recombination. Moreover, we demonstrate that TCRβ diversity is lower in CD4+ and CD8+ T-cells from SSc patients compared with memory T-cells from healthy individuals, as SSc TCRβ repertoires are largely dominated by clonally expanded persistent TCRβ sequences. Lastly, using "Grouping of Lymphocyte Interactions by Paratope Hotspots" (GLIPH2), we identify clusters of TCRβ sequences with homologous sequences that potentially recognize the same antigens and contain TCRβs that are persist in SSc patients.

In conclusion, our results show that CD4+ and CD8+ T-cells are highly persistent in SSc patients over time, and this persistence is likely a result from antigenic selection. Moreover, persistent TCRs form high similarity clusters with other (non-)persistent sequences that potentially recognize the same epitopes. These data provide evidence for an antigen driven expansion of CD4+/CD8+ T-cells in SSc.

INTRODUCTION

Systemic Sclerosis (SSc) is a complex chronic autoimmune disease, characterized by vascular abnormalities and widespread fibrosis affecting the skin and internal organs[1]. Although the pathogenic mechanisms underlying SSc remain largely unknown, multiple lines of evidence imply an important role for CD4+ and CD8+ T-cells in the progression of the disease. Activated T-cells infiltrate the skin of SSc patients already in the early phase of the disease[2, 3]. These infiltrating T-cells can cross-talk with fibroblasts, inducing fibroblast activation and apoptosis through secretion of pro-inflammatory cytokines and fas/fas ligand engagement[4, 5]. Moreover, T-cells isolated from SSc patients undergo clonal expansion when cultured together with autologous fibroblasts, suggesting that auto-antigens presented by SSc fibroblasts can induce auto-reactive T-cell responses[6]. Apart from skin, peripheral blood T-cells from SSc patients also exhibit signs of activation and express activation markers, including IL-2R, HLA-DR, and CD29 7–9], and secrete pro-inflammatory and pro-fibrotic factors[10–12].

The T-cell receptor (TCR) is a highly polymorphic surface receptor that allows T-cells to recognize antigenic peptides presented on the major histocompatibility complex (MHC)[13]. CD4+ T-cells recognize peptides presented on the MHC class II complex, while CD8+ T-cells recognize peptides presented on the MHC class I complex. Classical TCRs are heterodimers consisting of a paired α- and β-chain. These chains making up the TCR are generated through somatic recombination of V (variable), D (diversity) and J (joining) gene segments accompanied by pseudorandom insertions and deletions of nucleic acids at their joining regions[14], thereby giving rise to an enormously diverse TCR repertoire in every individual. By this process of VDJ recombination, the small set of genes that encode the TCR can be used to create over 10^{15} potential TCR clonotypes[13, 15]. Previous estimates of number of unique T-cells in a human range from 10^6 to 10^{11} [16–18], meaning that every individual only carries a small fraction of the potential repertoire.

6

High throughput sequencing of TCR repertoires is emerging as a valuable tool to unravel the exact role of T-cells in autoimmune diseases. The TCR repertoire has been proposed to serve as diagnostic biomarker for various autoimmune diseases, and recent studies have identified disease-associated TCR sequences in autoimmune diseases including autoimmune encephalomyelitis (AE), systemic lupus erythematosus (SLE), and rheumatoid arthritis (RA)[19–21]. Moreover, changes in T-cell repertoire diversity have been suggested to predispose the pathological manifestations in RA patients[22]. Prior studies examining the TCR repertoire in SSc have shown that there is an oligoclonal expansion of T-cells in the skin, lungs and blood of SSc patients[23–25], suggesting that expanded T-cells are involved in the disease pathogenesis. However, there are limitations to the results of previous studies that have examined the TCR repertoire in SSc patients. These include: a) lack of the use of high-throughput techniques; b) consideration of either CD4+ or CD8+ or unsorted T-cell populations; and c) the study of T-cells obtained only from a single time point thereby providing a static snapshot of the TCR repertoire in SSc.

Two major hypotheses have been postulated to explain mechanism of the expansion of T-cells in the context of autoimmunity[26, 27]. The first hypothesis states that T-cells might expand non-specifically or by chance (bystander activation) due to

chronic inflammation observed in autoimmune patients[27–29]. In this case, proliferation of T-cells is induced through non-specific activation in the presence of TLR ligands and cytokines during an immune response. Due to inherent biases in the V(D)J recombination process, some TCRβ sequences are more prevalent as they have a high generation probability[30]. As a result of this bias, during bystander activation, naïve T-cell clones with TCRβs that have high generation probabilities have a larger chance of being at the site of action due to their increased prevalence, and therefore have a higher chance to expand. In this case, expansion is a result of chance. The second hypothesis states that clonal expansion in autoimmunity is driven by chronic, and specific responses to antigens that selectively skew the TCR repertoire[26]. Here expansion is driven by antigen specific selection rather than chance. It remains to be unraveled which of these two mechanisms contributes to the activation and expansion of autoreactive T-cells in SSc.

To better understand T-cell responses in SSc pathogenesis, here we investigate the temporal TCR repertoire dynamics in SSc patients. We performed high-throughput sequencing of TCRβ chains of sorted CD4+ and CD8+ non-naive T-cells isolated from longitudinal samples from four SSc patients collected over a minimum of two years.

MATERIALS AND METHODS

Sample collection

Whole heparinized blood samples from SSc patients were obtained from the University Medical Center Utrecht. SSc patients were classified according to the ACR/EULAR criteria[31]. This study was conducted in accordance with the Declaration of Helsinki and was performed with approval of the Institutional Review Board of the University Medical Centre Utrecht, The Netherlands. The medical ethics committee of the UMC Utrecht approved the study (METC nr. 13–697). All participants enrolled in the study signed informed consent, and patient samples were anonymized upon collection. Peripheral blood mononuclear cells (PBMCs) were isolated by density gradient centrifugation on Ficoll-PlaqueTM Plus (GE Healthcare, Uppsala, Sweden). pDCs, mDCs, B-cells and monocytes were first depleted by magnetic bead sorting using the autoMACs Pro Separator (Miltenyi Biotec, Bergisch Gladbach, Germany) according to the manufacturer's protocol. The remaining peripheral blood lymphocytes (PBLs) were resuspended in freezing medium (80% FCS, (Sigma-Aldrich, Saint Louis, Missouri, USA)/20% DMSO, (Sigma-Aldrich)) and stored in sterile cryovials in liquid nitrogen (-196°C) until further use. From all patients PBLs were collected at baseline (T0), at least one year after inclusion (T1, ranging from 12 to 19 months) and at least two years after inclusion (T2, ranging from 24 to 46 months) (Fig. 1a, Table 1). High-resolution HLA typing was performed on DNA obtained from PBMCs from one time point for each patient by next generation sequencing.
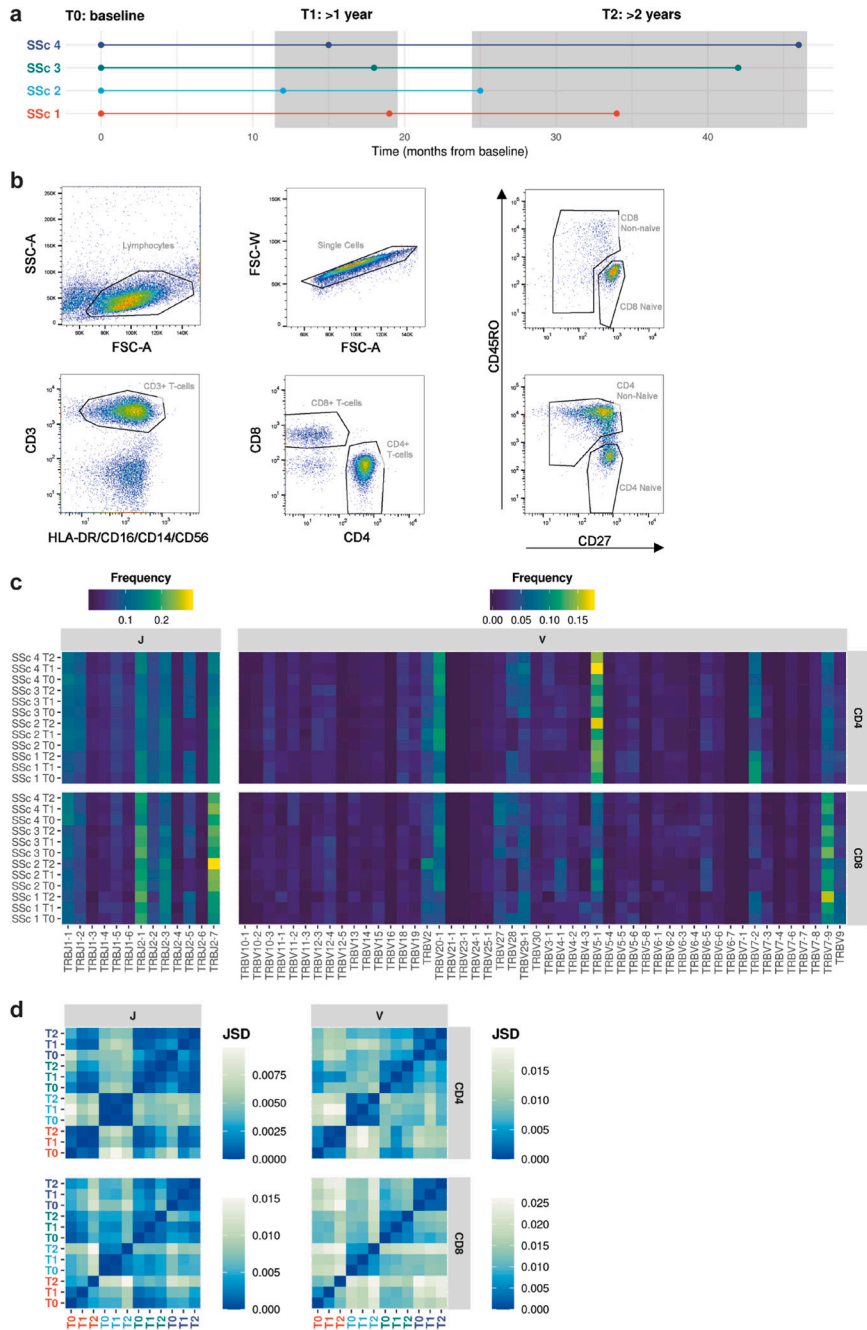
Figure 1. High-throughput TCRβ sequencing of SSc patients over time shows temporal persistence of Vβ and Jβ gene segment usage. a) Experimental design of longitudinal sampling. b) Gating strategy for FACS sorting of CD4+ non-naïve and CD8+ non-naïve T-cells for PBL samples. c) Heatmap showing the frequency of Vβ and Jβ gene usage across different samples. d) Heatmap of Jensen-Shannon divergence of Vβ and Jβ gene usage between different samples. Lower divergence indicates similar Vβ/Jβ gene usage.

Patient	Time	Months	Subset	Age	Sex	ILD	ANA	mRSS	HLA-A	HLA-B	HLA-C	HLA-DRB1	HLA-DQB1	HLA-DPB1
SSc 1	T0	0	dcSSc	65	F	1	1	11	*11:01 *31:01	*55:01 *56:01	*01:02 *03:03	*08:03 *11:01	*03:01 *03:01	*02:01 *04:01
	T1	19	dcSSc			1	1	11						
	T2	34	dcSSc			1	1	11						
SSc 2	T0	0	dcSSc	31	M	0	1	16	*23:01 *33:03	*08:01 *15:10	*03:04 *03:04	*08:06 *13:04	*03:01 *03:19	*01:01 *131:01
	T1	12	dcSSc			0	1	13						
	T2	25	dcSSc			0	1	13						
SSc 3	T0	0	lcSSc	46	F	1	1	4	*01:01 *02:01	*07:02 *37:01	*06:02 *07:02	*15:01 unk	*06:02 *06:unk	*03:01 *05:01
	T1	18	lcSSc			1	1	4						
	T2	42	lcSSc			1	1	8						
SSc 4	T0	0	lcSSc	49	F	0	1	2	*01:01 *24:02	*08:01 *14:02	*02:02 *07:01	*01:01 *03:01	*02:01 *05:01	*04:01 *04:02
	T1	15	lcSSc			0	1	2						
	T2	46	lcSSc			0	1	2						

Table 1. Clinical characteristics of patients included. Time from T0 is indicated in months after the first sample was taken. Abbreviations: dcSSc = diffuse SSc, lcSSc = limited SSc, M = male, F = female, ILD = interstitial lung disease (1 = yes, 0 = no), ANA = anti-nuclear antibodies (1 = yes, 0 = no), mRSS = modified rodman skin score, unk = unkown.

T-cell sorting

PBLs were thawed in RPMI 1640 (Gibco, Thermo Fisher Scientific, Waltham, Massachusetts, USA) supplemented with 20% FCS (Sigma- Aldrich), and washed with PBS. Subsequently, the cells were resuspended in FACS buffer (PBS supplemented with 1% BSA (Sigma-Aldrich) and 0.1% sodium azide (Sigma-Aldrich)) and stained using the following antibodies: CD3-AF700 (clone UCHT1, Biolegend, San Diego, California, USA), CD4-BV711 (clone OKT4, Sony Biotechnology, San Jose, California, USA), CD8-V500 (clone RPA-T8, BD Bioscience, San Jose, California, USA), CD56-PE-CF594 (clone B159, BD Bioscience), CD16- BV785 (clone 3G8, Sony Biotechnology), CD14-PerCP-Cy5.5 (clone HCD14, Biolegend), HLADR-BV421 (clone L243, Biolegend), CD45RO-PE-Cy7 (clone L243, BD Bioscience), CD27-APC-eFluor780 (clone O323, eBioscience, San Diego, California, USA). Multiparametric flow cytometry sorting of non-naive CD4+ T-cells (CD3+CD4+CD45RO+/-CD27+/-) and non-naive CD8+ T-cells (CD3+CD8+CD45RO+/-CD27+/-) was performed on the BD FACSAria II (BD Bioscience), according to the gating strategy described in Fig. 1b. After sorting, cells were washed in PBS, lysed in TRIzol™ Reagent (Invitrogen, Carlsbad, California, USA), and stored at -20°C until RNA isolation.

RNA isolation and TCR sequencing

RNA was isolated using the RNeasy Mini Kit (Qiagen, Venlo, Netherlands) by adding ethanol to the upper aqueous phase of processed TRIzol samples and transferring directly to the RNeasy spin columns. TCR amplification was performed according to the protocol published by Mamedov *et al.*[32]. Primer and barcode sequences are provided in Supplementary Table 1. Briefly, cDNA was generated by RACE using a primer directed to the TCRβ constant region. Thirteen nucleotide long unique molecular identifiers (UMIs) were incorporated during cDNA synthesis. Subsequently, two-stage semi-nested PCR

amplification was performed including a size selection/agarose gel purification step after the first PCR. To minimize cross-sample contamination, 5-nucleotide sample specific barcodes were introduced at two steps during the library preparation process[32]. Resulting TCR amplicons were subjected to high-throughput sequencing using the Ovation Low Complexity Sequencing System kit (NuGEN, San Carlos, California, USA) according to the manufacturer's instructions, and the Illumina MiSeq system (Illumina, San Diego, California, USA), using indexed paired-end 300 cycle runs.

TCR repertoire analysis

Raw paired-end reads were assembled using Paired-End reAd mergeR (PEAR)[33]. Sample specific barcode correction was performed using the '*umi_group_ec*' command from the Recover T Cell Receptor (RTCR) pipeline[34], allowing zero mismatches in the barcode seed sequence for UMI detection (sample specific barcodes are provided in Supplementary Table 1). This strict barcode selection resulted in about 50% loss of reads, but ensured that there was minimal cross-sample contamination. Subsequently, barcode sequence reads having the same UMI were collapsed into consensus sequences using the RTCR pipeline to accurately recover TCRβ sequences. Downstream data analysis of TCRβ repertoires was performed using the tcR R package[35].

Healthy longitudinal sequencing data[36] was obtained from the immuneACCESS portal of Adaptive Biotechnologies repository at: *https://clients.adaptivebiotech.com/pub/healthy-adult-time-course-TCRB*. Healthy data for validation of the diversity analysis was obtained from the immuneACCESS repository for the datasets from Emerson *et al.*[37], Rowe *et al.*[38], Tourino *et al.*[39], Lindau *et al.*[40], Soto *et al.*[41], Savola *et al.*[42], and De Neuter *et al.*[43]. The processed healthy data from Wang *et al.*[44] was obtained from the supplementary data provided in their publication.

Statistical analyses

Statistical analyses were performed using R version 3.4.1[45], and figures were produced using the R package ggplot2[46]. Generation probabilities of TCRβ amino acid sequences were computed using the generative model of V(D)J recombination implemented by OLGA (Optimized Likelihood estimate of immunoGlobulin Amino-acid sequences)[47], using the default parameters. Diversity estimates were calculated by sample-size-based rarefaction and extrapolation using the R package iNEXT (iNterpolation/EXTrapolation)[48]. Clustering analysis was performed using the GLIPH2 [49] webtool (*http://50.255.35.37:8080/*). Significant clusters were considered based on the following parameters: number of samples 3, number of CDR3 3, vb_score<0.05, length_score<0.05. After filtering for significance, clusters were ordered based on final_score obtained from GLIPH2. Network graphs of clusters were produced using the R package igraph[50]. Unless indicated otherwise, analysis of differences was performed using Student t-test. For multiple group comparisons, one-way anova was used. P-values <0.05 were considered statistically significant.

6

Availability of data

The TCRβ sequencing data presented in this study have been deposited in NCBI's Gene Expression Omnibus (GEO) database under GEO: GSE156980. Both raw data and processed data are available.

RESULTS

High-throughput TCR sequencing of SSc patients

To investigate the TCR repertoire dynamics in SSc, we performed high-throughput sequencing (HTS) of TCRβ chains on longitudinal samples obtained from four SSc patients collected over a minimum of two years. The clinical characteristics and HLA haplotypes of the SSc patients are included in Table 1. Among the SSc patients included, two were limited SSc (lcSSc), and two were diffuse SSc patients (dcSSc). From all patients PBLs were collected at baseline (T0), at least one year after inclusion (T1, ranging from 12 to 19 months) and at least two years after inclusion (T2, ranging from 24 to 46 months) (Fig. 1a). Frozen PBL samples were subjected to FACS sorting and sorted non-naive CD4+ T-cells (CD3+CD4+CD45RO+/-CD27+/-) and non-naive CD8+ T-cells (CD3+CD8+CD45RO+/-CD27+/-) were used for TCRβ sequencing (Fig. 1b). Sample specific barcodes and UMIs to barcode individual mRNA molecules were used to accurately recover TCRβ sequences using the RTCR pipeline[34]. We produced a total of 906 448 and 125 962 TCRβ UMI corrected amino acid (AA) sequence reads for CD4+ and CD8+ T-cells respectively. The number of UMI corrected AA sequence reads per sample was on average 75 000 for CD4+ T-cells and on average 10 500 for CD8+ T-cells (details see Supplementary Table 2).

Frequency of Vβ and Jβ gene segment usage indicates temporal persistence of TCRβ sequences in SSc patients

We first assessed the frequency of Vβ and Jβ gene segment usage in SSc patients over time (Fig. 1c). The most frequently used Vβ segments across all samples were V20–1, V5-1 and V7-9, for both CD4+ and CD8+ T-cells. When looking at Jβ segment usage, J2-7, J2-1 and J2-3 were most frequently observed. In previous studies, these Vβ (V20–1, V5-1 and V7-9) and Jβ genes (J2-7, J2-1 and J2-3) were also identified as the most frequently used genes in both healthy and diseased individuals[51, 52], reflecting known intrinsic biases in the V-D-J rearrangement process[53]. Additionally, Vβ2, which we identified with a relatively high frequency in CD4+ and CD8+ T-cells, was previously found to be one of the most frequent Vβ chains in peripheral blood T-cells in SSc patients in another study[54], showing that disease specific patterns are also present. Furthermore, we also observed individual specific patterns of Vβ and Jβ segments. As an example, SSc patient 2 displayed a lower frequency of V-28, V7-2 and J2-5 in both CD4+ and CD8+ T-cells across all time points compared to the other patients (Fig. 1c).

In order to quantify the relative similarity of Vβ and Jβ gene segment usage across all samples, we calculated the Jensen-Shannon divergence (JSD) between them (Fig. 1d). The JSD ranges from 0 to 1. A JSD of 0 indicates identical Vβ and Jβ

segment usage, while a JSD of 1 indicates that the Vβ and Jβ segment usage is distinct between two samples. When comparing the JSD between samples taken from the same patient, we observed that the use of Vβ and Jβ segments for both CD4+ and CD8+ T-cells was rather consistent over time, while samples taken from different patients displayed a higher divergence amongst each other (Fig. 1d). Moreover, when comparing the differences in JSD between Vβ and Jβ segment usage, we observed that Vβ usage (Fig. 1d, right panels) was more variable then Jβ segment usage (Fig. 1d, left panels) across different individuals. This difference in variability between Vβ and Jβ usage is to be expected since the TCRβ locus has more Vβ than Jβ gene segments (according to the ImMunoGeneTics (IMGT) database[55]), resulting in a greater potential variability for Vβ segment usage. Overall this analysis shows that, Vβ and Jβ gene segment usage is largely persistent within SSc patients over time.

SSc TCRβ repertoires are highly stable over time

Apart from examining persistence in the use of Vβ and Jβ segments, we wanted to further examine the persistence of full CDR3 amino acid sequences within SSc patients over time. In order to quantify the overlap in TCRβ repertoire between different samples, Morisita's overlap index was calculated to intersect amino acid CDR3 sequences. This index ranges from 0 (no overlapping sequences) to 1 (identical repertoires). Overlap analysis revealed that samples taken from the same patient shared a high number of sequences, indicating a clear temporal persistence of the repertoire within patients, while overlap was extremely limited between samples taken from different patients. This pattern was consistent over all time points, for both CD4+ and CD8+ T-cells (Fig. 2a and Fig. 2b respectively). These results demonstrate that the TCRβ repertoire in SSc patients is highly unique and stable over long periods of time, also at the level of exact CDR3 amino acid sequences.

Next we examined whether the frequencies of TCRβ sequences were also consistent over time. We observed that TCRβ sequences that were found with a high frequency at one time point were also found with a high frequency at the other time points collected from the same patient, for both CD4+ and CD8+ T-cells (Fig. 2c and d respectively). This shows that the frequencies of highly abundant TCRβs within SSc patients are largely consistent over long periods of time, even after almost four years of follow-up for SSc patient 4. Persistence of dominant TCRβ sequences was observed for all the SSc patients included in our study, for both CD4+ and CD8+ T-cells (Supplementary Figure S1). The exact number of TCRβ amino acid sequences overlapping between samples taken from the same patient are shown in Fig. 2e. Although the absolute number of sequences that are overlapping within patients over time (i.e. persistent sequences) are low, they make up a substantial part of the samples in terms of abundance, based on UMI corrected reads, as shown in Fig. 2f. These results clearly demonstrate that SSc repertoires are largely dominated by persistent sequences.

6
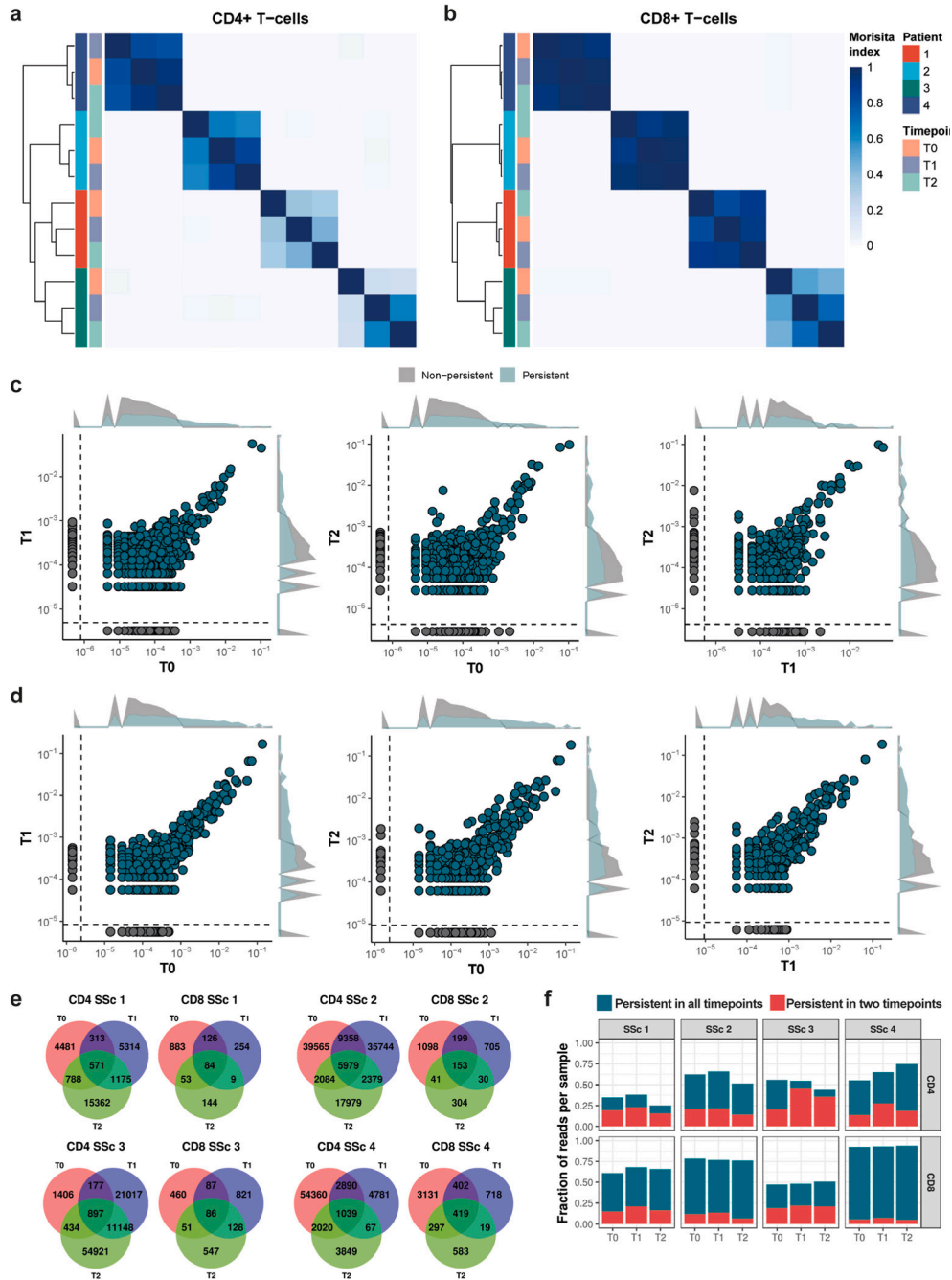
Figure 2. SSc TCRβ repertoires are highly stable over time and are dominated by persistent TCRβ sequences. a) Heatmap showing high overlap (dark blue, morisita index) of TCR sequences within individuals and limited overlap (light blue) of TCR clones between individuals for CD4+ T-cells and b) CD8+ T-cells. c) Overlap between T-cell sequences within SSc patient 4 over time for CD4+ T-cells and d) CD8+ T-cells. Each dot on the scatter plots indicates a single TCR sequence. Axes denote frequency of sequences. Persistent sequences are shown in blue and non-persistent sequences in grey. *(continued)*

Frequency was calculated based on the total reads for each sample. Density curves indicate the distribution of sequences across the samples. e) Venn diagram showing the exact number of persistent and non-persistent TCR sequences within each patient for CD4 and CD8 T-cells. f) Boxplots (continued) showing the fraction of total reads per sample (y-axis) across the different samples. Majority of reads is coming from sequences persistent across all time points (blue) or occurring in at least two time points (red).

        In order to investigate whether persistent TCRβ sequences have any known antigen specificity, we queried the sequences that were persistently present in all three time points for every patient in VDJdb (a curated database of TCR sequences with known antigen specificities)[56]. The results of this analysis are shown in Table 2. In this table we show the hits for peptides presented on MHC II for CD4+ T-cells and peptides presented on MHC I for CD8+ T-cells, matching with the HLA haplotype of the patient from which the sequences were obtained.

        For persistent TCRβs from CD8+ T-cells, we mainly found associations with peptides related to viral antigens including EBV, CMV, influenza and HIV (Table 2). For SSc patient 2, we identified one TCRβ sequence (CASSRLAGGTDTQYF) associated with both CMV and HIV-1. This TCRβ sequence most likely represents a hit against CMV in patient 2, as all the patients included in our study are known to be HIV-1 negative. For persistent TCRβs from CD4+ T-cells, we identified two persistent sequences (CASSLEETQYF and CASSLGGEETQYF) associated with CMV. No hits against TCRβ sequences associated with human autoantigens where identified in VDJdb. However, the vast majority of the records present in VDJdb are based on studies of viral and cancer epitopes, while TCRβ sequences associated with autoantigens remain understudied.

        Overall, these results reveal a clear temporal persistence of clonally expanded CD4+ and CD8+ T-cells in SSc patients, and we show that the TCRβ repertoire in SSc patients is highly stable over time.

6

Persistent clones display common features across SSc patients

        Previously, TCRs in the context of autoimmune disease have been associated with certain characteristics such as shorter length and a bias in Vβ/Jβ-gene segment usage[21, 39, 57]. To further investigate the potential involvement of persistent TCRβs identified in SSc patients in autoimmune responses, we computed the distribution of lengths of all the TCRβ amino acid sequences and compared the lengths of persistent and non-persistent TCRβs. The lengths were calculated based on both incidence (without weighing sequences by their abundance) and abundance (also taking into account the frequency of the sequences). A Gaussian distribution of lengths was observed for both persistent and non-persistent sequences, in CD4+ as well as CD8+ T-cells, when looking at incidence (Fig. 3a), and distribution of lengths of sequences was similar between persistent and non-persistent sequences in all samples (two sample Kolmogorov-Smirnov tests >0.05 for all comparisons). When comparing the distribution of CDR3 lengths of persistent and non-persistent sequences based on abundance, we again observed no significant differences in the distributions neither for CD4+ nor CD8+ T-cells (two sample Kolmogorov-Smirnov tests >0.05 for all comparisons, Fig. 3b). Although the CDR3 length distributions where not significantly different between persistent and

CD4 - MHC II					
	CDR3	**MHC A**	**MHC B**	**Epitope**	**Species**
SSc 3	CASSLEETQYF	HLA-DRA*01:01	HLA-DRB1*15	pp65	CMV
SSc 3	CASSLGGEETQYF	HLA-DRA*01:01	HLA-DRB1*15	pp65	CMV
CD8 - MHC I					
	CDR3	**MHC A**	**MHC B**	**Epitope**	**Species**
SSc 1	CASSWGQGSNYGYTF	HLA-A*11:01	B2M	EBNA3B	EBV
SSc 2	CASSLGQAYEQYF	HLA-B*08:01	B2M	EBNA3A	EBV
SSc 2	CASSPGQGEGYEQYF	HLA-B*08:01	B2M	BZLF1	EBV
SSc 2	CASSPGTGEGYEQYF	HLA-B*08:01	B2M	BZLF1	EBV
SSc 2	CASSRLAGGTDTQYF*	HLA-B*08:01	B2M	Gag	HIV-1
SSc 2	CASSRLAGGTDTQYF*	HLA-B*08:01	B2M	IE1	CMV
SSc 3	CSARDRTGNGYTF*	HLA-A*02:01	B2M	BMLF1	EBV
SSc 3	CSARDRTGNGYTF*	HLA-A*02:01	B2M	M	Influenza
SSc 3	CSARDRTGNGYTF*	HLA-A*02:01	B2M	BMLF1	EBV
SSc 3	CASSPTDTQYF	HLA-A*02	B2M	pp65	CMV
SSc 4	CASSVGQAYEQYF	HLA-B*08:01	B2M	EBNA3A	EBV

Table 2. Epitope specificity of persistent TCRβs (occurring in all three time points within a patient), according to VDJdb. For CD4+ TCRβ sequences, specificities for peptides presented on MHC II are given and for CD8+ TCRβ sequences, specificities for peptides presented on MHC I are given. Only results for HLA molecules matching the HLA haplotype of the specific patients are given. Sequences indicated with an asterisk (*) are identical.

non-persistent sequences, in SSc patient 1 and 2, more shorter length sequences were observed in the persistent TCRβs, while in SSc patient 4, longer sequences were found (Fig. 3b). However, this skewness in lengths is mainly caused by the expansion of few dominant TCRβ sequences.

Next to comparing lengths, we also compared the frequencies of Vβ and Jβ segment usage between persistent and non-persistent TCR se- quences to see if there was any preferential usage (Fig. 3c). Although most differences observed were small, we identified various Vβ and Jβ gene segments that had either higher or lower frequencies in persistent sequences across all SSc patients. As an example, TRBJ1-2 had a significantly lower frequency in persistent sequences in CD4+ as well as CD8+ T-cells across all SSc patients, while the frequency of TRBV7-2 and TRBV7-3 was higher in persistent sequences in CD4+ and CD8+ T-cells respectively (Fig. 3c). This analysis demonstrates that, although the number of exact sequences that are shared between SSc patients is low, TCRβ sequences that are persistently present in SSc patients over time show similarities in terms of Vβ and Jβ usage. Moreover, these are significantly different from non-persistent sequences, showing that persistent sequences display preferential usage of Vβ and Jβ segments across SSc patients. Given that similar TCR sequences are thought to be involved in T-cell responses to similar antigens[58–61], the preferential segment Vβ and Jβ segment usage across SSc patients over time might reflect chronic immune responses against antigens that are commonly present across patients.
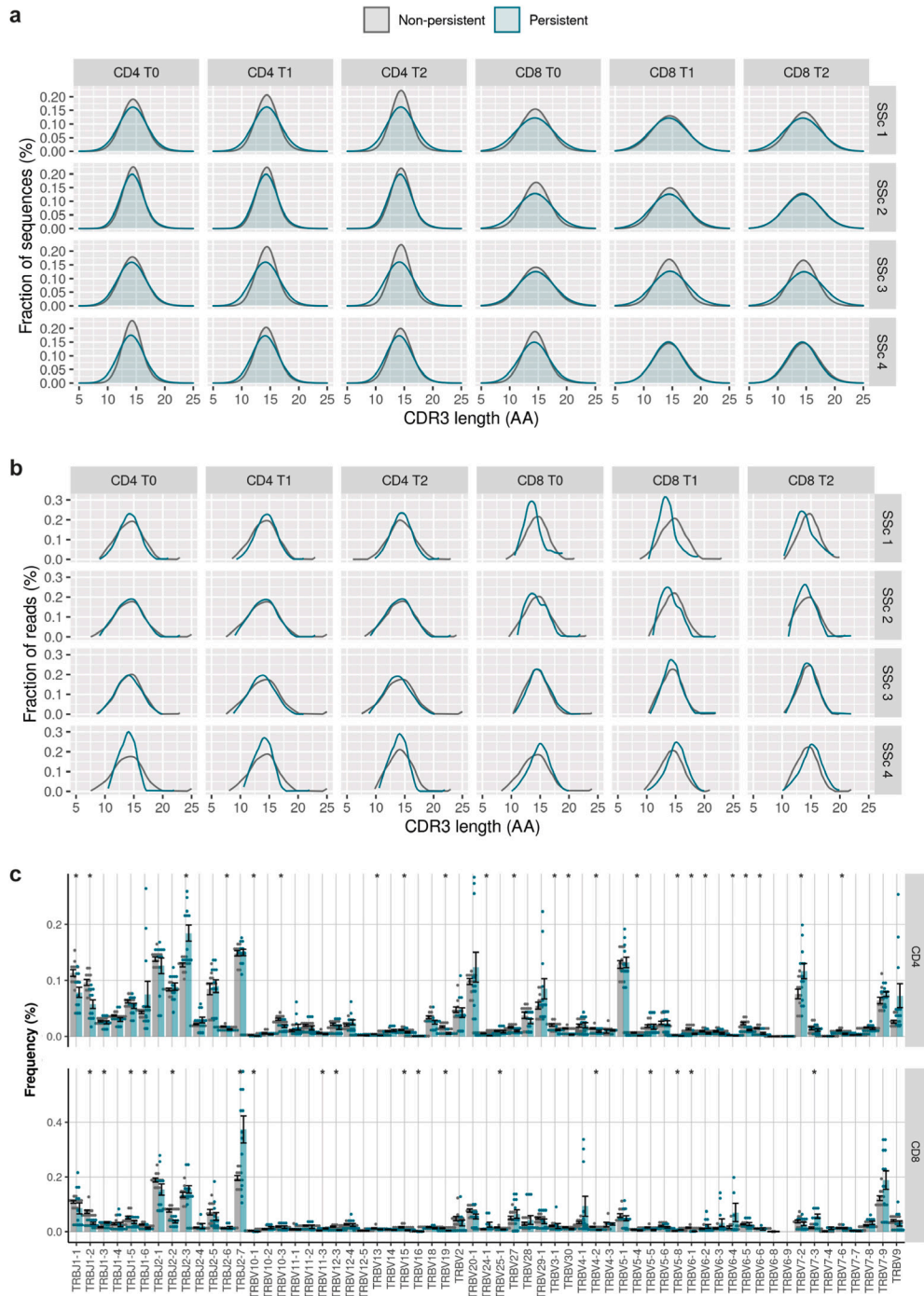
Figure 3. Comparison of TCRβ characteristics between persistent and non-persistent sequences.
a) Density plots showing the distribution of lengths of TCR amino acid sequences (x-axis), based on
incidence. Distribution of lengths of persistent (blue) and non-persistent (grey) TCR sequences *(continued)*

are shown. b) Line plots showing the distribution of lengths of TCR amino acid sequences (x-axis), based on abundance. c) Bar plot showing the frequency (y-axis) of TCR Vβ and Jβ gene segments (x-axis) of persistent sequences (blue) compared to non-persistent (grey) sequences. Every dot represents one sample. T-tests were performed to test if the gene usage was significantly different between persistent and non-persistent sequences (* = $p < 0.05$).

SSc TCRβ repertoires are less diverse than healthy memory repertoires

The persistence of highly abundant TCRs is not necessarily unique to autoimmune repertoires and has in fact previously been observed in healthy individuals[36]. Therefore, we also compared our SSc data to a public dataset of longitudinal TCRβ sequences from healthy donors[36]. To investigate whether TCRβs in SSc patient repertoires are aberrantly expanded compared to healthy repertoires, we compared the Shannon diversities of healthy memory repertoires to those of SSc. As the samples differed in their sizes, to compare the diversity between samples we performed rarefaction analysis. The healthy control dataset from Chu *et al*. included here did not contain singletons (sequences represented by one read). We therefore performed rarefaction analysis and estimated Shannon diversity in our SSc patient data by including and excluding singletons. In Fig. 4a, we show the estimated Shannon diversity based on the rarefied and extrapolated data excluding singletons. The estimated species richness and Shannon diversity was significantly lower in SSc CD4+ and CD8+ T-cell repertoires as compared to the healthy memory repertoire (p-value <0.05, Fig. 4b). To estimate the effect of excluding singletons on the diversity, the same analysis was also performed including the SSc singletons. In this analysis, the estimated Shannon diversity of TCRβ sequences obtained from SSc CD8+ T-cells were still significantly lower than the diversity of healthy memory cells, while the significant difference in diversity with the CD4+ T-cells was no longer observed (Supplementary Figure S2a-b).
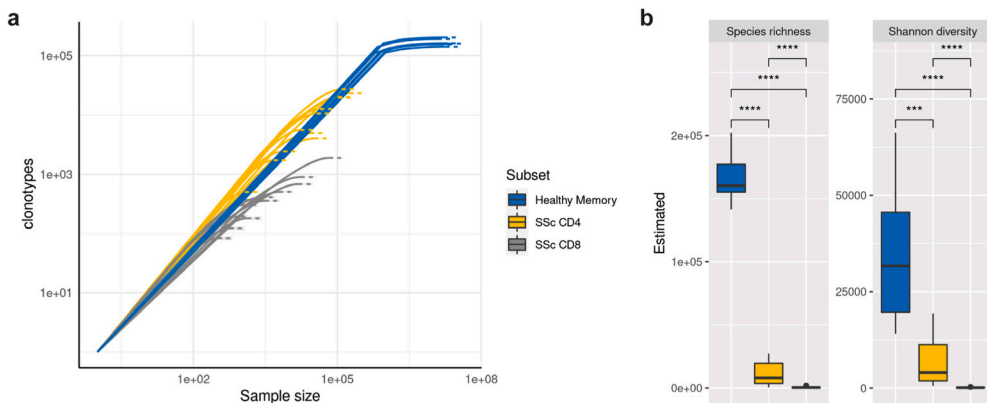
Figure 4. SSc TCRβ repertoires have a decreased diversity compared to healthy memory repertoires. a) Sample based rarefaction and extrapolation curves. Solid lines depict observed data, dashed line depict extrapolated data. Calculated for all healthy memory samples (blue), SSc CD4 (yellow) and SSc CD8 samples (grey). Every line represents one sample. b) Boxplots show median of asymptotic diversity estimates calculated from the rarefied and extrapolated data shown in G. ***p < 0.001, ****p < 0.0001 (one way anova).

Additionally, next to the decreased diversity of SSc repertoires versus healthy repertoires, we also observed a decreased diversity of the SSc CD8+ repertoire versus the SSc CD4+ repertoire. Given the fact that there was a substantial difference in sequencing depth between CD4+ and CD8+ T-cells in our analysis (Supplementary Table 2), we wanted to confirm the results from the rarefaction analysis in a separate subsampling analysis. We repeated the diversity analysis for 100 random subsamples obtained from SSc CD4+ T-cells and compared these results to the results obtained from SSc CD8+ T-cells. In all subsamples analyzed, CD8+ T-cells had a significantly lower diversity than the CD4+ T-cells (Supplementary Figure S2c), corroborating with the results from the rarefaction analysis from the original/complete data.

As the healthy control dataset from Chu *et al.* contains some technical differences from our SSc dataset (e.g. DNA versus RNA, no distinction between CD4+/CD8+ T-cells, exclusion of singletons), we compared the diversity observed in SSc repertoires versus healthy donors in eight additional datasets. The eight additional datasets included TCRβ repertoires from sorted CD4+/CD8+ T-cells isolated from healthy individuals (see Supplementary Table 3). We compared Shannon diversity, estimated by rarefaction analysis, for all eight datasets with our SSc dataset (Supplementary Figure 3). For datasets excluding singletons, we also excluded singletons from the SSc data. We were able to validate the decreased diversity observed in SSc patients for both CD4+ and CD8+ T-cells for the vast majority of the datasets analyzed. In more detail, in six out of seven datasets containing CD4+ T-cells, Shannon diversity was significantly lower in SSc as compared to CD4+ T-cells from healthy individuals. For all five datasets containing CD8+ T-cells, Shannon diversity was significantly lower in SSc as compared to CD8+ T-cells from healthy individuals.

Overall, these results demonstrate that the TCRβ repertoire diversity is lower in SSc patients compared to healthy individuals. This provides evidence for a skewed, clonally expanded repertoire in SSc, potentially due to chronic antigen driven T-cell responses.

6

Persistent frequency of dominant TCR sequences is driven by antigenic selection rather than bystander activation

The hypothesis of 1) bystander activation due to chronic inflammation and 2) clonal expansion driven by the chronic presence of antigens skewing the TCR repertoire have been proposed. To test the hypothesis of bystander activation, we investigated whether easy to generate TCRβ sequences (having high generation probabilities) are present at high frequencies in SSc patients. To this extend, we calculated the generation probabilities (p_{gens}) of the TCRβs identified in SSc patients using OLGA[47]. We then used linear regression to model the relationship between TCRβ frequency and p_{gen}. Using this analysis, we show that in SSc CD4+ and CD8+ T-cells, TCRβ frequency and p_{gen} are not related (p-value >0.05, Fig. 5a and Fig. 5b for CD4+ and CD8+ T-cells respectively). This indicates that T-cells persistent in SSc patients have not expanded because of random bystander effects.

For naive T-cells isolated from healthy individuals, we found a significant positive correlation between the p_{gen} and abundance of TCRβs (p-value <0.05, Fig.
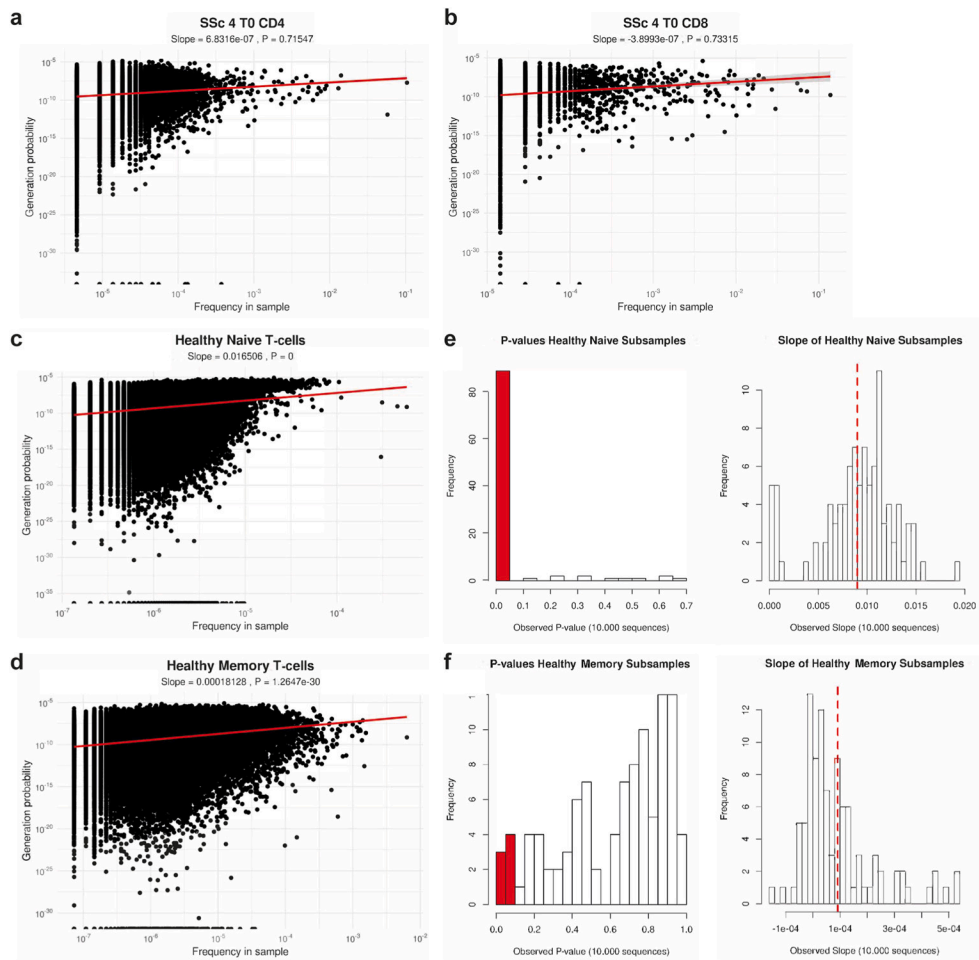
Figure 5. Frequencies of dominant TCRβ sequences are not driven by generation probabilities in SSc patients. a) Linear regression plot between frequency (x-axis) and generation probability (y-axis) of one representative plot of taken from SSc CD4+ T-cells (patient 4, time point 0). Red line indicates the linear regression model fit. b) Linear regression plot between frequency and generation probability of one representative plot of taken from SSc CD8+ T-cells (patient 4, time point 0). c) Linear regression plot between frequency and generation probability of one representative plot of taken from healthy naïve T-cells (donor 1, time point 4). d) Linear regression plot between frequency and generation probability of one representative plot of taken from healthy memory T-cells (donor 1, time point 4). e) Histograms showing the distribution of p-values (left) and slopes (right) for linear regression between frequency and generation probability (computed using OLGA) of subsamples of healthy donor naïve T-cell samples (donor 1, time point 4). Distribution over 100 different subsamples is shown. Red bars in the p-value histogram indicates p-value ≤ 0.05. Red dashed line in the slope histogram indicates the mean slope for the 100 subsamples. f) Distribution of p-values (left) and slopes (right) for linear regression between frequency and generation probability of subsamples of healthy donor memory T-cell samples (donor 1, time point 4).

5c). For memory T-cells isolated from the same healthy individuals, we also observed a significant positive correlation between p_{gen} and abundance of TCRβs (p-value <0.05, Fig. 5d). Notably, the slope for memory T-cells was lower than that observed for the

naive T-cells (0.00018 versus 0.0165, respectively). From this analysis we show that for naïve and memory TCRβs obtained from healthy individuals, p_{gen} and abundance are positively correlated, whereas in non-naive SSc T-cells no relationship between p_{gen} and abundance is observed. However, as the samples obtained from healthy individuals were sequenced more deeply than our SSc samples, this observed difference in correlation might be confounded by sequencing depth. Therefore, we repeated the linear regression analysis for 100 random subsamples obtained from the healthy dataset and compared these results to the results obtained from SSc samples. Upon subsampling of the naïve healthy T-cells, 89% of the slopes observed the in linear regression were significantly different from zero (linear regression p-value <0.05 and slope >0, as indicated by the red bars in Fig. 5e), confirming that indeed there is a positive relationship between p_{gen} and abundance for TCRβs obtained from healthy naïve T-cells. However, when looking at healthy memory T-cells, we did not confirm the correlation between frequency and p_{gen} that was observed in the full sample (p-value <0.05 in 7% of subsamples, Fig. 5f), showing that in healthy memory T-cells there is no clear correlation between TCRβ pgen and abundance. Thus, subsampling analysis shows that in both healthy memory T-cells and SSc non-naïve CD4+ and CD8+ T-cells, TCRβ p_{gen} and abundance are not correlated. These results suggest that the more abundant TCRβs in SSc repertoires are likely there because of selection, similar to what is observed in healthy memory repertoires, rather than bystander activation of naive T-cells that are not antigen specific.

Clusters of similar TCRβ sequences can be identified in SSc patients over time

In order to identify TCRβs in SSc patients that are potentially involved in chronic autoimmune responses, we used "Grouping of Lymphocyte Interactions by Paratope Hotspots" (GLIPH2), that employs sequence similarity and motif analysis to group TCR sequences that potentially recognize the same epitope[49]. To screen for antigen specific TCRβ clusters, we used all sequences obtained from every time point for each individual SSc patient as an input for GLIPH2. In order to exclude false positives, for each patient we considered the clusters that contained sequences from all time points, had at least three unique CDR3s, had similar CDR3 lengths (length score <0.05), and shared similar Vβ-gene frequency distributions (Vβ score <0.05). The number of clusters that were obtained for every patient for CD4+ and CD8+ T-cells are given in Table 3.

Significant clusters were identified in all patients, either based on global similarity (CDR3 sequence differing by maximum one amino acid) or local similarity (shared motif within CDR3 amino acid region). All clusters identified by GLIPH2 are given in Supplementary Table 4. In Fig. 6a, an example network of TCRβ sequences based on clustering analysis by GLIPH2 in CD4+ T-cells is given. Red and purple nodes within this network represent TCRβs that are persistently present in three or two time points respectively, while blue shaded nodes represent sequences are found only in one time point. Nodes are connected when they are part of the same cluster as identified by GLIPH2. Global and local motifs are indicated in green and orange respectively. A "%" sign within the motif indicates a variable amino acid across the sequences in which that particular motif is present.

6

Cell type	Patient	Total clusters	Significant clusters
CD4	SSc 1	7800	204
	SSc 2	50365	3735
	SSc 3	34882	554
	SSc 4	22674	550
CD8	SSc 1	335	2
	SSc 2	534	23
	SSc 3	453	22
	SSc 4	1500	106

Table 3. Number of clusters identified by GLIPH2 within SSc patients. All sequences from all time points from each patient (for the two cell types) were used as input for GLIPH2. The total clusters column represents the total number of clusters identified by GLIPH2. The significant clusters column represents the number of significant clusters (consisting of sequences from all three time points, at least three unique CDR3s, length score <0.05, and Vβ score <0.05).

Within the network, persistent TCRβs share motifs with other persistent and non-persistent TCRβs sequences. This shows that persistent TCRβ sequences cluster together with other, similar TCRβ sequences, potentially representing groups of T-cells responding to the same antigen. The top 10 clusters, based on final cluster score outputted by GLIPH2, and their corresponding motifs for this network are given in Fig. 6b. Some clusters also shared TCRβ sequences with other clusters within a patient, showing that clusters also display convergence between each other (Fig. 6c, top 3 clusters are shown). Similar results were obtained for CD8+ T-cells, where we also identified many clusters of persistent and non-persistent TCRβs sharing motifs within the CDR3 region (Fig. 6d–f). These results demonstrate that, apart from the presence of individual persistent clonally expanded T-cells, clusters of T-cells with potentially similar specificities are present within SSc patients over time.

Lastly, we performed an overlap analysis of all motifs from significant clusters identified by GLIPH2 between the different SSc patients (Fig. 6g and h). We did not observe any motifs for CD4+ T-cells or for CD8+ T-cells that overlap between all four patients. For the CD4+ T-cells, there were 11 motifs that were identified in clusters from three out of four SSc patients (Fig. 6g). These represent groups of T-cells that are likely to recognize the same or highly similar antigens across SSc patients, which could potentially be involved in SSc pathogenesis. We also performed the GLIPH2 analysis on the data from memory T-cells for healthy donors obtained from Chu *et al.*[36]. For this analysis, the same parameters as for SSc were used, however, as for healthy donor 1 and healthy donor 2, memory T-cells were obtained from only two time points, for these donors we also considered clusters that had at least two unique CDR3s to be significant. The number of significant clusters that were obtained for the memory T-cells obtained from healthy donors is given in Supplementary Figure 4a. In order to determine whether the 11 motifs that we found overlapping in CD4+ TCRβ clusters between SSc patients were specific for SSc, we overlapped them with the motifs obtained from the healthy donor memory T-cells. Of the 11 motifs, 2 (S%TTDT and S%DRAYE) were found to be uniquely present in SSc patients (Supplementary Figure 4b). In order to investigate
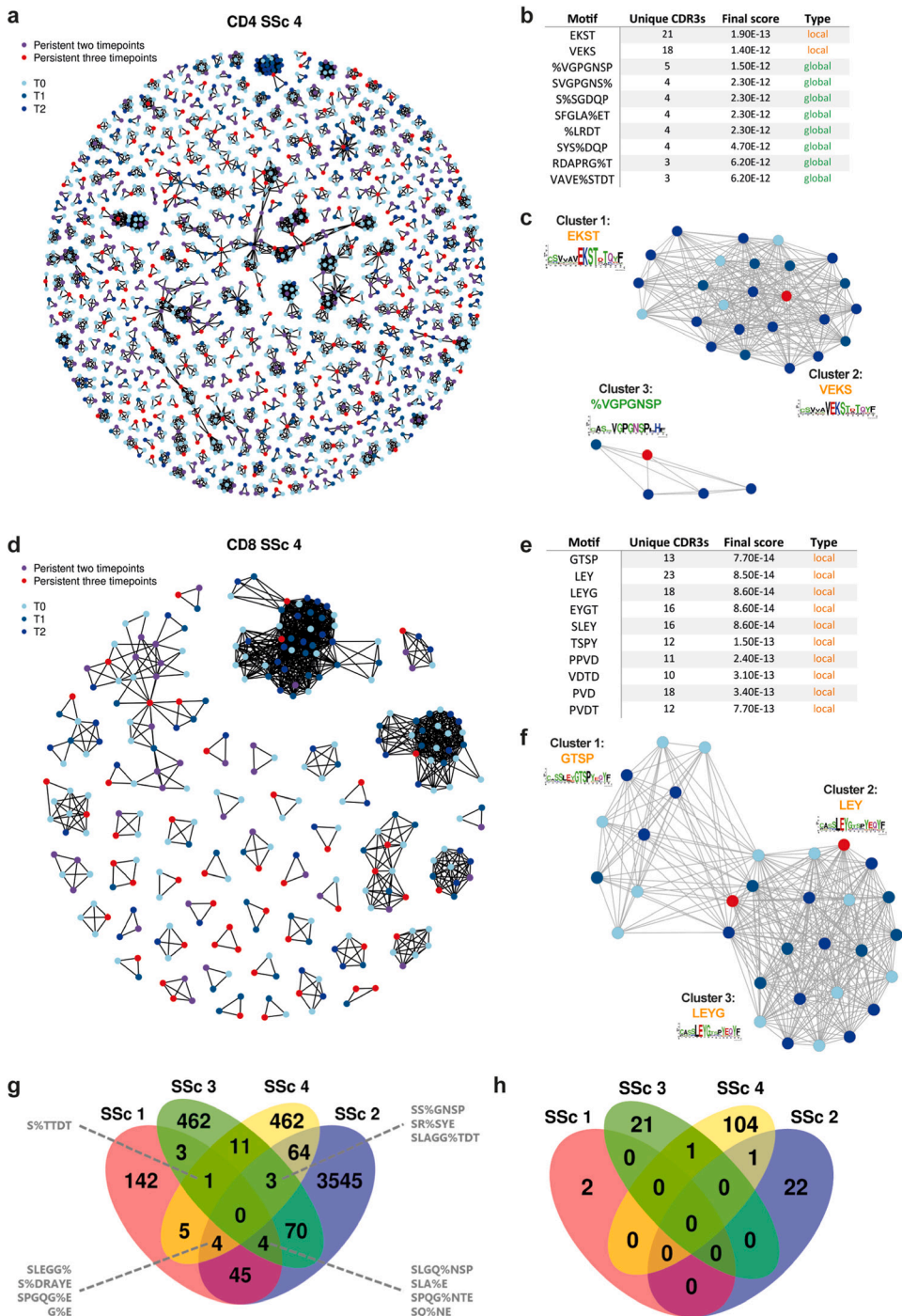
Figure 6. Clustering analysis of similar TCRβ sequences in SSc patients over time. a) Network of TCRβ clusters identified by GLIPH2 in CD4+ T-cells from SSc patient 4. Every node represents one TCRβ sequence. Red nodes represent TCRβs persistent across three time points within a patient, purple nodes represent TCRβs persistent across two time points within a patient, and blue shaded nodes *(continued)*

represent TCRβs present at a single time point. Nodes are connected if they share a motif or a have similar CDR3 region sequence. b) Top ten clusters identified by GLIPH2 in CD4+ T-cells from SSc patient 4. For each cluster, the motif identified by GLIPH2 is given. (continued) Motifs with global similarity (CDR3 sequence differing by maximum one amino acid) are indicated in green, while motifs with local similarity (shared motif within CDR3 amino acid region) are indicated in orange. c) Top three clusters in CD4+ T-cells of SSc 4. For each cluster a sequence logo is given based on the TCRβ sequences present in the cluster. d) Network of TCRβ clusters identified by GLIPH2 in CD8+ T-cells from SSc patient 4. e) Top ten clusters identified by GLIPH2 in CD8+ T-cells from SSc patient 4. f) Top three clusters in CD8+ T-cells of SSc 4.

whether these motifs have been associated to known antigens, we queried them against the VDJdb motif database, which contains several curated CDR3 motifs specific to different epitopes[56]. No hits were identified for these motifs.

DISCUSSION

Identification of TCR sequences that are associated with the chronic autoimmune response in SSc will help to get more insights into the autoimmune pathogenesis of the disease, and will help to identify the antigenic triggers that underlie these responses. Our analysis reveals that the peripheral blood TCRβ repertoire of SSc patients is highly stable over time. Moreover, the TCRβ sequences that were found within a patient with a high frequency at one time point were also found with a high frequency at the other time points (even after four years), showing that frequencies of dominant TCRβs are largely consistent over time. These persistent, clonally expanded CD4+ and CD8+ T-cells are potentially involved in the autoimmune responses underlying SSc pathogenesis. Furthermore, we have shown that the persistent expansion of these T-cells is likely a result of antigenic selection rather than recombination bias, as TCRβ frequencies were not related to their respective generation probabilities.

When we queried the persistent TCRβs found in our SSc patients in VDJdb, we obtained several hits for TCRβ sequences that are known to be associated with viral antigens from influenza, CMV and EBV, especially in the CD8+ T-cell compartment. SSc patient 4 has a CMV and EBV positive status, and TCRβ sequences associated with CMV and EBV were identified in this patient. For the other SSc patients included in this study, the CMV and EBV status are unknown. Interstingly, EBV and CMV infections have been shown to be environmental risk factors for SSc[62–65], and molecular mimicry between chronic viral antigens and human autoantigens has been proposed as a potential driver for autoimmune disease[66].

Chu *et al.* have previously shown that subsets of persistent TCRβs are also present within healthy individuals[36]. Thus, persistence of TCR sequences in itself is likely not just a characteristic of autoimmune related repertoires. However, the temporal dynamics of the TCR repertoire in healthy individuals in the aforementioned study have only been investigated over a period of one year, so it remains to be seen whether this stability is also maintained in healthy individuals over longer periods of time, as is observed in SSc in our study. Moreover, whereas the previous study looked into the total pool of memory T-cells, we show that persistent TCRβs can be identified in both the CD4+ and CD8+ memory T-cell compartments separately.

When further comparing the TCR repertoires of CD4+ and CD8+ T-cells from SSc patients to repertoires obtained from healthy memory T-cells, we found that SSc repertoires have lower diversity. Indeed, decreased diversities of TCRβ repertoires as compared to healthy have been observed in other autoimmune diseases[20, 52, 67], and have been proposed as a characteristic of autoimmune repertoires. Interestingly, in SSc, differences in the diversity of the TCR repertoire have also been observed between responders and non-responders after autologous hematopoietic stem-cell transplantation (AHSCT, the only therapy with long-term clinical benefit in SSc), with non-responders having a less diverse repertoire[68]. This provides further evidence that decreased TCR repertoire diversity contributes to the autoimmune pathogenesis of SSc.

Determinant spreading has been proposed as a pathogenic event in various autoimmune diseases. During determinant spreading, an antigenic epitope can induce an immune response against other, distinct epitopes on the same protein or other proteins in the same tissue. Those epitopes then become additional targets for the immune response. In systemic sclerosis, vascular damage or skin injury could induce an autoimmune response against epitopes that are normally sequestered[69]. It has been proposed that during determinant spreading, diversification of the TCR repertoire occurs as various TCRs would be expected to respond against the novel epitopes. However, this response is thought to be dynamic and although diversification of the repertoire may occur in early stages of the disease, at later time points the response might become more restricted[70]. Following this line of thought, as the SSc patients in our study have more established forms of the disease (lcSSc and dcSSc), we expect to see a more skewed TCR repertoire. Indeed when comparing the diversity of the TCR repertoire of our SSc patients to healthy donors, we observe a decrease in the repertoire diversity in SSc. In the context of determinant spreading, it would be interesting to repeat this analysis in patients at earlier stages of the disease (for example eaSSc patients) and see if the repertoire is more diverse, which could imply that determinant spreading is ongoing and contributing to disease pathogenesis.

In our generation probability analysis, we show that in both healthy memory T-cells and SSc non-naïve CD4+ and CD8+ T-cells, TCRβ p_{gen} and abundance are not correlated. Therefore it is likely that in SSc patients, similar as to healthy donors, T-cell expansion is caused by antigen specific selection rather than bystander effects. Bystander activation of T-cells has been demonstrated to be driven by excessive production of cytokines including type I IFN, IL-15 and IL-18 for CD8+ T-cells[71], and IL-2 for CD4+ T-cells[72] during immunopathology. Interestingly, type I IFNs are implicated in SSc pathogenesis[73], and IL-15 and IL-18 have been found to be increased in circulation of SSc patients as compared to healthy individuals[74, 75]. This shows that SSc patients display a skewed pro-inflammatory milieu that favors bystander activation of T-cells. Therefore, one might expect an increased number of bystander expanded T-cells within these patients as compared to healthy individuals. However, since we observe no difference between SSc and healthy memory T-cells within our generation probability analysis, we show that bystander activation does not significantly contribute to the skewed clonal expansion of T-cells that we observed in SSc patients.

Predicting T-cell reactivity towards antigens is one of the major areas currently

6

investigated in the field of TCR research. Since prediction of TCR binding to a specific antigen is extremely challenging, current efforts are more focused on identifying groups of TCRs that contain certain motifs within their CDR3 region[58–61]. These groups of TCRs comprise clones that potentially respond to the same antigen. Apart from exact sharing of TCRβ sequences between samples, we also identified clusters of TCRβs that share sequence motifs and were persistent within patients. This indicates that antigen selection reshapes the TCRβ repertoire in SSc. Potential antigens could include self-antigens, or chronic infections with pathogens (e.g., CMV or EBV, for which we identified associated persistent TCRβs). Interestingly, we also found clusters of TCRβs from CD4+ T-cells within patients that shared motifs with other TCRβ clusters between patients, even though these patients did not all share common HLA alleles. Notably, two of these motifs could not be detected in TCRβ clusters identified from memory T-cells of healthy donors. These could represent clusters of similar TCRβs that might contribute to more public autoimmune responses underlying SSc pathogenesis. The antigens that these clusters of TCRβs are potentially responding to remain to be identified. Although some epitope associated TCR CDR3 motifs have recently been identified in VDJdb[56], current information on antigen specific motifs is extremely sparse, especially for MHC II associated epitopes. Sequencing of larger (longitudinal) cohorts and MHC Class II tetramer studies including autoantigens are needed to further identify which TCR motifs can be associated with autoimmune disease, and link them to their potential pathogenic targets. For CD8+ T-cells, we did not find any clusters overlapping between more than two patients. However, in general we obtained less clusters in CD8+ T-cells than we did in CD4+ T-cells. This could be due to the fact that we sequenced less CD8+ than CD4+ T-cells, and thus this difference might be explained by a difference in sequencing depth between the two cell types.

To validate our findings and further study the potential pathogenic role of antigen specific TCRβ clusters in SSc, larger patient cohorts should be studied. In this cohort we included a limited number of patients with similar clinical characteristics which makes it difficult to account for factors such as age, sex and ethnicity influencing the immune system. Thus, studying larger longitudinal cohorts are needed to further define disease specific clusters of autoimmune associated TCRβs. Lastly, it would also be interesting to perform immune sequencing of SSc skin to see if these TCRβ clusters/motifs can be traced back in the skin (the major organ affected by the disease) of SSc patients.

In conclusion, our data provide evidence for an antigen driven expansion of CD4+/CD8+ T-cells in SSc. We have identified clusters of T-cell clones that are highly persistent over time, and we have shown that this persistence likely is a result of antigenic selection.

REFERENCES

1. Gabrielli, E. V. Avvedimento, T. Krieg. Scleroderma. *N. Engl. J. Med.* vol. 360, no. 19, pp. 1989–2003, May 2009.
2. A.D. Roumm, T.L. Whiteside, T.A. Medsger, *et al*. Lymphocytes in the skin of patients with progressive systemic sclerosis. *Arthritis Rheum*. vol. 27, no. 6, pp. 645–653. Jun. 1984.
3. A. Kalogerou, E. Gelou, S. Mountantonakis, *et al*. Early T cell activation in the skin from patients with systemic sclerosis. *Ann. Rheum. Dis.,* vol. 64, no. 8, pp. 1233–1235. Aug. 2005.
4. R. de Palma, E. D'aiuto, S. Vettori, *et al*. Peripheral T cells from patients with early systemic sclerosis kill autologous fibroblasts in co-culture: is T-cell response aimed to play a protective role? *Rheumatology*. vol. 49, no. 7, pp. 1257–1266. Jul. 2010.
5. T. Hügle, S. O'Reilly, R. Simpson, *et al*. Tumor necrosis factor-costimulated T lymphocytes from patients with systemic sclerosis trigger collagen production in fibroblasts. *Arthritis Rheum*. vol. 65, no. 2, pp. 481–491. Feb. 2013.
6. R. De Palma, F. Del Galdo, S. Lupoli, *et al*. Peripheral T lymphocytes from patients with early systemic sclerosis co-cultured with autologous fibroblasts undergo an oligoclonal expansion similar to that occurring in the skin. *Clin. Exp. Immunol*. vol. 144, no. 1, pp. 169–176. Apr. 2006.
7. A. Kahan, A. Kahan, F. Picard, *et al*. Abnormalities of T lymphocyte subsets in systemic sclerosis demonstrated with anti-CD45RA and anti- CD29 monoclonal antibodies. *Ann. Rheum. Dis.* vol. 50, no. 6, pp. 354–358. Jun. 1991.
8. L.P. Ercole, M. Malvezzi, A.C. Boaretti, *et al*. Analysis of lymphocyte subpopulations in systemic sclerosis. *J Investig. Allergol. Clin. Immunol*. vol. 13, no. 2, pp. 87–93. 2003.
9. G. Papp, I.F. Horvath, S. Barath, *et al*. Altered T-cell and regulatory cell repertoire in patients with diffuse cutaneous systemic sclerosis. *Scand. J. Rheumatol*. vol. 40, no. 3, pp. 205–210. May 2011.
10. P. Fuschiotti, A.T. Larregina, J. Ho, *et al.* Interleukin-13-producing CD8 T cells mediate dermal fibrosis in patients with systemic sclerosis. *Arthritis Rheum*. vol. 65, no. 1, pp. 236–246. Jan. 2013.
11. M.E. Truchetet, N.C. Brembilla, E. Montanari, *et al*. Increased frequency of circulating Th22 in addition to Th17 and Th2 lymphocytes in systemic sclerosis: association with interstitial lung disease. *Arthritis Res. Ther*. vol. 13, no. 5, pp. R166. Oct. 2011.
12. S. O'Reilly, T. Hügle, J.M. Van Laar. T Cells in Systemic Sclerosis: A Reappraisal. *Rheumatology (Oxford)*. vol. 51, no. 9, pp. 1540–1549. Sep. 2012.
13. M.M. Davis, P.J. Bjorkman. T-cell antigen receptor genes and T-cell recognition, *Nature*. vol. 334, no. 6181, pp. 395–402. Aug. 1988.
14. C.H. Bassing, W. Swat, F.W. Alt. The mechanism and regulation of chromosomal V(D)J recombination. *Cell*. vol. 109, pp. S45–S55. 2002.
15. H.S. Robins, S.K. Srivastava, P.V. Campregher, *et al*. Overlap and effective size of the human CD8+ T cell receptor repertoire. *Sci. Transl. Med*. vol. 2, no. 47, pp. 47ra64. Sep. 2010.
16. T.P. Arstila, A. Casrouge, V. Baron, *et al*. A direct estimate of the human αβ T cell receptor diversity, *Science*. vol. 286, no. 5441, pp 958–961. Oct. 1999.
17. Q. Qi, Y. Liu, Y. Cheng, *et al*. Diversity and clonal selection in the human T-cell repertoire. *Proc. Natl. Acad. Sci*. vol. 111, no. 36, pp. 13139–13144. Sep. 2014.
18. T. Mora, A.M. Walczak. How many different clonotypes do immune repertoires contain? *Curr. Opin. Struct. Biol*. vol. 18, pp. 104–110. Dec. 2019.
19. Y. Zhao, P. Nguyen, J. Ma, *et al*. Preferential use of public TCR during autoimmune encephalomyelitis. *J. Immunol*. vol. 196, no. 12, pp. 4905–4914. Jun. 2016.
20. C.M. Chang, Y.W. Hsu, H.S.C. Wong, *et al*. Characterization of T-cell receptor repertoire in patients with rheumatoid arthritis receiving biologic therapies. *Dis. Markers*. vol. 2019, pp. 2364943. Jul. 2019.
21. X. Liu, W. Zhang, M. Zhao, *et al*. T cell receptor β repertoires as novel diagnostic markers for systemic lupus erythematosus and rheumatoid arthritis. *Ann. Rheum. Dis*. vol. 78, no. 8, pp. 1070–1078. May 2019.
22. J.J. Goronzy, C.M. Weyand. T-cell senescence and contraction of T-cell repertoire diversity - catalysts of autoimmunity and chronic inflammation. *Arthritis Res. Ther*. vol. 5, no. 5, pp. 225–234. Aug. 2003.
23. L.I. Sakkas, B. Xu, C.M. Artlett, S. Lu, S.A. Jimenez, C.D. Platsoucas, Oligoclonal T cell expansion in the skin of patients with systemic sclerosis, J. Immunol. 168 (2002) 3649–3659.
24. V.V. Yurovsky, F.M. Wigley, R.A. Wise, *et al*. Skewing of the CD8 T-cell repertoire in the lungs of patients with systemic sclerosis. *Hum. Immunol*. vol. 48, no. 1–2, pp. 84–97. Jun–Jul. 1996.

6

25. B. White, V. V Yurovsky. Oligoclonal expansion of V delta 1+ gamma/delta T-cells in systemic sclerosis patients. *Ann. N. Y. Acad. Sci*. vol. 756, pp. 382–391. Jul. 1995.

26. M. Attaf, E. Huseby, A.K. Sewell. αβ T cell receptors as predictors of health and disease. *Cell. Mol. Immunol.* vol. 12, no. 4, pp. 391–399. Jan. 2015.

27. Y. Pacheco, Y. Acosta-Ampudia, D.M. Monsalve, *et al.* Bystander Activation and Autoimmunity. *J Autoimmun*. vol. 103, pp. 102301. Sep. 2019.

28. H.G. Lee, J.U. Lee, D.H. Kim, *et al.* Pathogenic function of bystander-activated memory-like CD4+ T cells in autoimmune encephalomyelitis, *Nat. Commun*. vol. 10, no. 1, pp. 709. Feb. 2019.

29. N.J. Burroughs, M. Ferreira, B.M.P.M. Oliveira, *et al.* Autoimmunity arising from bystander proliferation of T cells in an immune response model. *Math. Comput. Model*. vol. 53, no. 7–8, pp. 1389–1393. Apr. 2011.

30. V. Venturi, K. Kedzierska, D.A. Price, *et al.* Sharing of T cell receptors in antigen-specific responses is driven by convergent recombination. *Proc. Natl. Acad. Sci. U. S. A.* vol. 103, no. 49, pp. 18691–18696. Dec. 2006.

31. F. Van Den Hoogen, D. Khanna, J. Fransen, *et al.* Classification criteria fo rsystemic sclerosis: an American college of rheumatology/European league against rheumatism collaborative initiative. *Ann. Rheum. Dis*. vol. 72, no. 11, pp. 1747–1755. Nov. 2013.

32. I.Z. Mamedov, O.V. Britanova, I.V. Zvyagin, *et al.* Preparing unbiased T-cell receptor and antibody cDNA libraries for the deep next generation sequencing profiling. *Front. Immunol*. vol. 4, pp. 456. Dec. 2013.

33. J. Zhang, K. Kobert, T. Flouri, *et al.* PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics*. vol. 30, no. 5, pp. 614–620. Oct. 2013.

34. B. Gerritsen, A. Pandit, A.C. Andeweg, *et al.* RTCR: a pipeline for complete and accurate recovery of T cell repertoires from high throughput sequencing data. *Bioinformatics*. vol. 32, no. 20, pp. 3098–3106. Oct. 2016.

35. V.I. Nazarov, M.V. Pogorelyy, E.A. Komech, *et al.* tcR: an R package for T cell receptor repertoire advanced data analysis. *BMC Bioinf.* vol. 16, no. 1, pp. 175. May 2015.

36. N.D. Chu, H.S. Bi, R.O. Emerson, *et al.* Longitudinal immunosequencing in healthy people reveals persistent T cell receptors rich in highly public receptors. *BMC Immunol.* vol. 20, no. 1, pp. 19. Jun. 2019.

37. R. Emerson, A. Sherwood, C. Desmarais, et al. Estimating the ratio of CD4+ to CD8+ T cells using high-throughput sequence data. J. Immunol. Methods. vol. 391, no. 1–2, pp. 14–21. Feb. 2013.

38. J.H. Rowe, O.M. Delmonte, S. Keles, *et al.* Patients with CD3G mutations reveal a role for human CD3g in Treg diversity and suppressive function. *Blood*. vol. 131, no. 21, pp. 2335–2344. May 2018.

39. I. Gomez-Tourino, Y. Kamra, R. Baptista, *et al.* T cell receptor β-chains display abnormal shortening and repertoire sharing in type 1 diabetes. *Nat. Commun*. vol. 8, pp. 1792. Nov. 2017.

40. P. Lindau, R. Mukherjee, M.V. Gutschow, et al. Cytomegalovirus exposure in the elderly does not reduce CD8 T cell repertoire diversity. J. Immunol. vol. 202, no. 2, pp. 476–483. Jan. 2019.

41. C. Soto, R.G. Bombardi, A. Branchizio, *et al.* High frequency of shared clonotypes in human B cell receptor repertoires. *Nature*. vol. 566, no. 7744, pp. 398–402. Feb. 2019.

42. P. Savola, T. Kelkka, H.L. Rajala, *et al.* Somatic mutations in clonally expanded cytotoxic T lymphocytes in patients with newly diagnosed rheumatoid arthritis. *Nat. Commun*. vol. 8, pp. 15869. Jun. 2017.

43. N. De Neuter, E. Bartholomeus, G. Elias, *et al.* Memory CD4+ T cell receptor repertoire data mining as a tool for identifying cytomegalovirus serostatus. *Gene Immun*. vol. 20, no. 3, pp. 255–260. Mar. 2019.

44. L. Wang, P. Zhang, J. Li, H. *et al.* High-throughput sequencing of CD4+ T cell repertoire reveals disease-specific signatures in IgG4-related disease. *Arthritis Res. Ther*. vol. 21, no. 1, pp. 295. Dec. 2019.

45. R Development Core Team, A Language and Environment for Statistical computing., Vienna, Austria, 2017, ISBN 3-900051-07-0. R Foundation for Statistical Computing, Vienna, Austria, http://www.R-project.org.

46. H. Wickham. ggplot2: Elegant Graphics for Data Analysis, S*pringer-Verlag, New York*, 2016.

47. Z. Sethna, Y. Elhanati, C.G. Callan, *et al.* OLGA: fast computation of generation probabilities of B-and T-cell receptor amino acid sequences and motifs. *Bioinformatics*. vol. 35, no. 17, pp. 2974–2981. Sep. 2019.

48. T.C. Hsieh, K.H. Ma, A. Chao. iNEXT: an R package for rarefaction and extrapolation of species

diversity (Hill numbers). Methods Ecol. Evol. vol. 7, no. 12, pp. 1451–1456. Jun. 2016.

49. H. Huang, C. Wang, F. Rubelt, *et al.* Analyzing the Mycobacterium tuberculosis immune response by T-cell receptor clustering with GLIPH2 and genome-wide antigen screening. *Nat. Biotechnol.* vol. 38, no. 10, pp. 1–9. Oct. 2020.

50. G. Csardi, T. Nepusz. The Igraph Software Package for Complex Network Research. *InterJournal Complex Syst.* Nov. 2006.

51. L. Ma, L. Yang, B. Shi, *et al.* Analyzing the CDR3 repertoire with respect to TCR - beta chain V-D-J and V-J rearrangements in peripheral T cells using HTS. *Sci. Rep.* vol. 6, pp. 29544. Jul. 2016.

52. J.H. Cui, Y.B. Jin, K.R. Lin, *et al.* Characterization of peripheral blood TCR repertoire in patients with ankylosing spondylitis by high-throughput sequencing. *Hum. Immunol.* vol. 79, no. 6, pp. 485–490. Jun. 2018.

53. J.E. Park, R.A. Botting, C.D. Conde, *et al.* A cell atlas of human thymic development defines T cell repertoire formation. *Science.* vol. 367, no. 6480, pp. eaay3224. Feb. 2020.

54. D. Farge, L.C.M. Arruda, F. Brigant, *et al.* Long-term immune reconstitution and T cell repertoire analysis after autologous hematopoietic stem cell transplantation in systemic sclerosis patients. *J. Hematol. Oncol.* vol.10, no. 1, pp. 21. Jan. 2017.

55. M.P. Lefranc. IMGT, the international ImMunoGeneTics database®. Nucleic Acids Res. vol. 29, no .1, pp. 207–209. Jan. 2001.

56. D.V. Bagaev, R.M.A. Vroomans, J. Samir, *et al.* VDJdb in 2019: database extension, new analysis infrastructure and a T-cell receptor motif compendium. *Nucleic Acids Res.* vol. 48, pp. D1057–D1062. Jan. 2020.

57. X. Hou, P. Zeng, X. Zhang, *et al.* Shorter TCR β-chains are highly enriched during thymic selection and antigen-driven selection. *Front. Immunol.* vol. 10, pp. 299. Feb. 2019.

58. M.V. Pogorelyy, A.A. Minervina, M. Shugay, et al. Detecting T cell receptors involved in immune responses from single repertoire snapshots. PLoS Biol. vol. 17, no. 6. pp. e3000314. Jun. 2019.

59. J. Glanville, H. Huang, A. Nau, *et al.* Identifying specificity groups in the T cell receptor repertoire. *Nature.* vol. 547, no. 7661, pp. 94–98. Jul. 2017.

60. P. Dash, A.J. Fiore-Gartland, T. Hertz, *et al.* Quantifiable predictive features define epitope-specific T cell receptor repertoires. *Nature.* vol. 547, no. 7661, pp. 89–93. Jun. 2017.

61. A. Madi, A. Poran, E. Shifrut, *et al.* T Cell Receptor Repertoires of Mice and Humans Are Clustered in Similarity Networks Around Conserved Public CDR3 Sequences. *Elife.* vol. 6, pp. e22057. Jul. 2017.

62. C. Lunardi, C. Bason, R. Navone, *et al.* Systemic sclerosis immunoglobulin G autoantibodies bind the human cytomegalovirus late protein UL94 and induce apoptosis in human endothelial cells. *Nat. Med.* vol. 6, no. 10, pp. 1183–1186. Oct. 2000.

63. Y. Arnson, H. Amital, S. Guiducci, *et al.* The role of infections in the immunopathogensis of systemic sclerosis-evidence from serological studies. *Ann. N. Y. Acad. Sci.* vol. 1173 , pp. 627–632. Sep. 2009.

64. L. Sternbæk, A.H. Draborg, M.T. Østerlund, *et al.* Increased antibody levels to stage-specific Epstein–Barr virus antigens in systemic autoimmune diseases reveal a common pathology. *Scand. J. Clin. Lab. Invest.* vol. 79, no. 1–2, pp. 7–16. Feb–Apr. 2019.

65. L.C.M. Arruda, E. Clave, C. Douay, *et al.* CMV-specific clones may lead to reduced TCR diversity and relapse in systemic sclerosis patients treated with AHSCT. *Rheumatology.* vol. 59, no. 9, pp. e38–e40. Sep. 2020.

66. M.F. Cusick, J.E. Libbey, R.S. Fujinami. Molecular mimicry as a mechanism of autoimmune disease. *Clin. Rev. Allergy Immunol.* vol. 42, no. 1, pp. 102–111. Feb. 2012.

67. D.R. Thapa, R. Tonikian, C. Sun, *et al.* Longitudinal analysis of peripheral blood T cell receptor diversity in patients with systemic lupus erythematosus by next-generation sequencing. *Arthritis Res. Ther.* vol. 17, no. 1, pp. 132. May 2015.

68. L.C.M. Arruda, K.C.R. Malmegrim, J.R. Lima-Júnior, *et al.* Immune rebound associates with a favorable clinical response to autologous HSCT in systemic sclerosis patients. *Blood Adv.* vol. 2, no. 2, pp. 126–141. Jan. 2018.

69. L.S. Chan, C.J. Vanderlugt, T. Hashimoto, *et al.* Epitope spreading: lessons from autoimmune skin diseases. *J. Invest. Dermatol.* vol. 110, no. 2, pp. 103–109. Feb. 1998.

70. P.V. Lehmann, E.E. Sercarz, T. Forsthuber, *et al.* Determinant spreading and the dynamics of the autoimmune T-cell repertoire. *Immunol. Today.* vol. 14, no. 5, pp. 203–208. May 1993.

71. T.S. Kim, E.C. Shin. The activation of bystander CD8+ T cells and their roles in viral infection. *Exp. Mol. Med.* vol. 51, no. 12, pp. 1–9. Dec. 2019.

6

72. O. Boyman. Bystander activation of CD4 T cells. *Eur. J. Immunol.* vol. 40, no. 4, pp. 936–939. Apr. 2010.

73. B. Skaug, S. Assassi. Type I Interferon Dysregulation in Systemic Sclerosis. *Cytokine*, vol. 132, pp. 154635. Aug. 2020.

74. D.M. Wuttge, M. Wildt, P. Geborek, *et al.* Serum IL-15 in patients with early systemic sclerosis: a potential novel marker of lung disease. *Arthritis Res. Ther.* vol. 9, no. 5, pp. R85. 2007.

75. E. Lin, F.B. Vincent, J. Sahhar, *et al.* Analysis of serum interleukin(IL)-1α, IL-1β and IL-18 in patients with systemic sclerosis. *Clin. Transl. Immunol.* vol. 8, no. 4, pp. e1045. Apr. 2019.

SUPPLEMENTARY INFORMATION
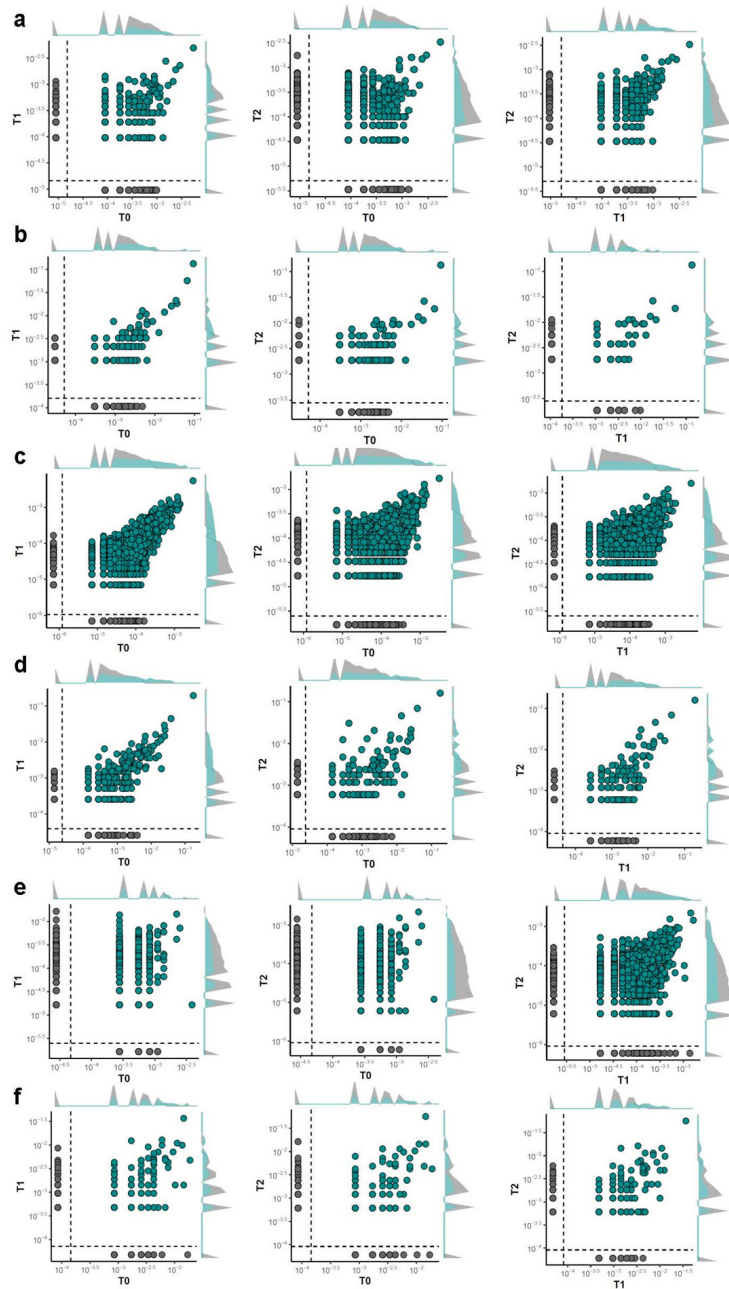
Figure S1 Overlap between frequencies of TCR sequences for a) CD4+ T-cells from SSc patient 1, b) CD8+ T-cells from SSc patient 1, c) CD4+ T-cells from SSc patient 2, d) CD8+ T-cells from SSc patient 2, e) CD4+ T-cells from SSc patient 3, f) CD8+ T-cells from SSc patient 3. Axes denote frequency of sequences. Persistent sequences are shown in blue and non-persistent sequences in grey. Frequency was calculated based on the total UMI corrected reads for each sample. Density curves indicate the distribution of sequences across the samples.
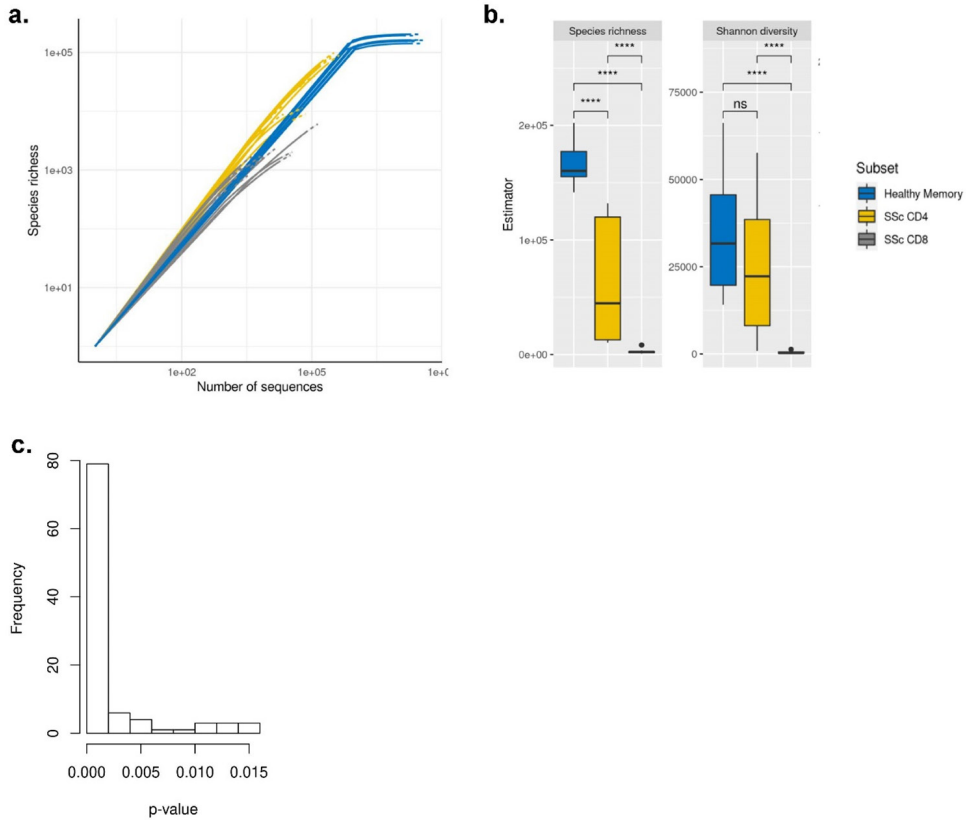
Figure S2 a) Sample based rarefaction and extrapolation curves. SSc samples include singletons, healthy samples exclude singletons. Solid lines depict observed data, dashed line depict extrapolated data. Calculated for all healthy memory samples (blue), SSc CD4 (yellow) and SSc CD8 samples (grey). Every line represents one sample. b) Boxplots show median of asymptotic diversity estimates calculated from the rarefied and extrapolated data shown in G. ns: non-significant, **** p < 0.0001 (one way anova). c) Histogram showing the distribution of p-values for difference in diversity of subsamples of CD4+ T-cells versus CD8+ T-cells Distribution over 100 different subsamples is shown.
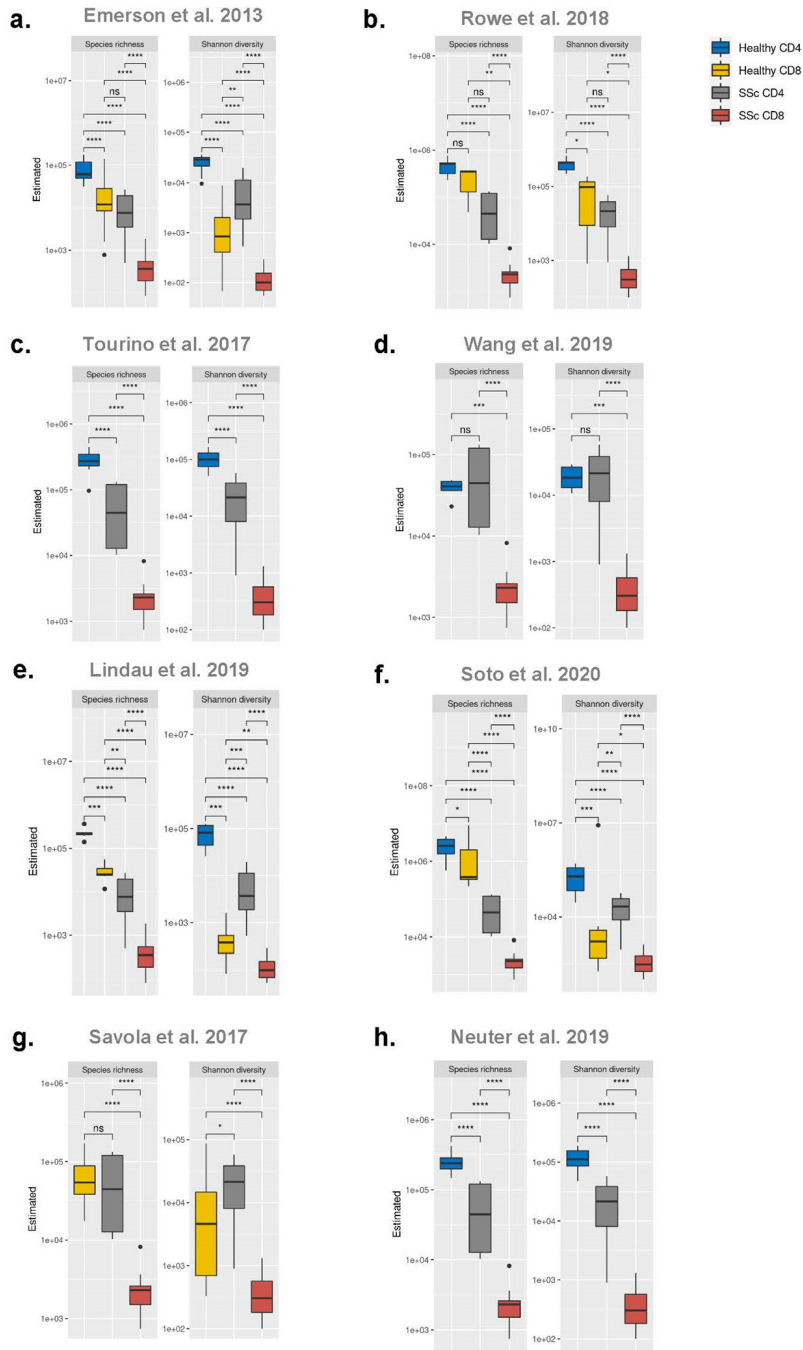
Figure S3 Validation of the decrease in diversity in SSc repertoires versus healthy donors in eight additional datasets. a-h) Boxplots show median of asymptotic diversity estimates calculated using rarefaction analysis. Colours depict different samples obtained from healthy CD4+ T-cells (blue), healthy CD8+ T-cells (yellow), SSc CD4+ T-cells (grey), and SSc CD8+ T-cells (red). Asterisks indicate p-values, where NS=not significant, * p<0.05, ** p<0.01, *** p<0.001, **** p < 0.0001 (one way anova).

a.

Cell type	Donor	Total clusters	Significant clusters
Healthy Memory	HC 1	152663	50483
	HC 2	112848	41288
	HC 3	153063	27183

b.
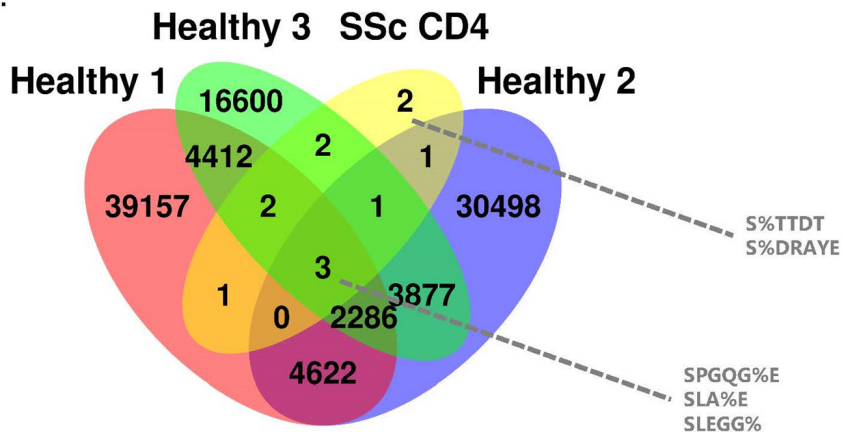
Figure S4 a) Number of clusters identified by GLIPH2 within memory T-cells obtained from healthy donors. All sequences from all time points from each donor were used as input for GLIPH2. The total clusters column represents the total number of clusters identified by GLIPH2. The significant clusters column represents the number of significant clusters. b) venn diagram showing the overlap in significant motifs obtained by GLIPH2 in healthy donors, as well as motifs identified to be overlapping in 3 out of 4 SSc patients (indicated by SSc CD4). Motifs overlapping between all comparisons, and motifs that were uniquely found in SSc CD4 have been highlighted.

6

Chapter 7

Compartmentalization and persistence of dominant (regulatory) T cell clones indicates antigen skewing in juvenile idiopathic arthritis

Authors:

G. Mijnheer[1]*, N.H. Servaas[1]*, J.Y. Leong[2], A. Boltjes[1], E. Spierings[1], P. Chen[2], L. Lai[2], A. Petrelli[1], S. Vastert[1,3], R.J. de Boer[4], S. Albani[2], A. Pandit[1#], F. van Wijk[1#]

*/# These authors contributed equally

Affiliations:
1. Center for Translational Immunology, University Medical Center Utrecht, Utrecht University, Utrecht, the Netherlands
2. Translational Immunology Institute, Singhealth/Duke-NUS Academic Medical Centre, the Academia, 20 college road, Discovery tower level 8, Singapore
3. Pediatric Immunology & Rheumatology, Wilhelmina Children's Hospital , University Medical Center Utrecht, Utrecht University, Utrecht, the Netherlands
4. Theoretical Biology, Utrecht University, Utrecht, the Netherlands

7

ABSTRACT

Objective Autoimmune inflammation is characterized by tissue infiltration and expansion of antigen-specific T cells. This inflammation is often limited to specific target tissues, but within tissues, multiple sites can be affected. However, it remains yet to be explored whether distinct affected sites are infiltrated with the same T cell clones and whether these clones persist over time.

Methods Here we performed CyToF analysis and T cell receptor (TCR) sequencing to study immune cell composition and (hyper-)expansion of circulating and joint-derived Tregs and non-Tregs in Juvenile Idiopathic Arthritis (JIA, N=9). We studied different joints affected at the same time, as well as over the course of relapsing remitting disease.

Results CyToF analysis revealed that the composition and functional characteristics of the immune infiltrates are strikingly similar between joints within one patient. Furthermore, we observed a strong overlap between dominant T cell clones, especially Treg, in inflamed joints affected at the same time, of which some could also be detected in circulation. Finally, these dominant T cell clones were found to persist over the course of relapsing remitting disease. Moreover, T cell clones found across two distinct joints were characterized by a high degree of sequence similarity, indicating the presence of TCR clusters responding to the same antigens.

Conclusions Together, these data suggest that in localized autoimmune disease there is auto-antigen driven expansion of both Teffector and Treg clones, that are highly persistent and are (re)circulating. These dominant clones might represent interesting therapeutic targets.

INTRODUCTION

Inflammation, often localized to specific target tissues, is a hallmark of autoimmune diseases. Within specific inflamed tissues, multiple sites can be affected, however. Examples of this phenomenon include the inflammation of several sites within the intestine in Inflammatory Bowel Diseases (IBD), or the inflammation of multiple joints in Rheumatoid Arthritis (RA) and Juvenile Idiopathic Arthritis (JIA). Investigation of the pathophysiological mechanisms underlying these autoimmune diseases have implicated T cells as key players of inflammation in specific target tissues. Firstly, many autoimmune diseases are associated with the expression of specific MHC (HLA) class II alleles, which is hypothesized to lead to altered antigen presentation and enhanced CD4+ T cell activation[1]. Moreover, highly activated CD4+ T cells often accumulate in affected tissue sites[2]. Next to the presence of these highly activated CD4+ T cells, increased numbers of CD4+CD25+CD127^{low}FOXP3+ (forkhead box P3) regulatory T cells (Tregs), capable of suppressing immune responses and fundamental to immune homeostasis, are also present[3, 4].

It is increasingly appreciated that tissue resident T cells display an array of distinct trafficking and functional markers compared to circulating T cells[5]. Previous flow cytometric studies have already identified a heterogeneous pool of CD4+ T cells and Tregs at inflamed sites in human autoimmune diseases, which are phenotypically distinct from their circulating counterparts[6–10]. Novel technologies such as mass cytometry allow for more detailed, high resolution analysis of the cellular heterogeneity with inflamed tissues. This information can be used to reveal potential pathogenic T cell populations characterized by unique phenotypes within these tissues. Next to the identification of activated T cell subsets using flow cytometry, studies assessing the T cell receptor (TCR) repertoire have generated evidence for the presence of clonally expanded T cells in specific tissues in a wide spectrum of autoimmune diseases[11–15]. These findings suggest that the T cell response at the site of autoimmune inflammation is mounted by the presentation of specific local antigens that selectively induces activation, expansion and/or migration of antigen-specific T cell clones.

Similar to conventional T cells, Tregs that leave the thymus typically express a unique TCR. While Tregs only represent a small fraction of the total CD4+ T cell pool, the TCR repertoire of peripheral Tregs, consisting of both thymus derived and peripheral induced Tregs, is as diverse as that of conventional CD4+ T cells[16–18]. Through interaction of their TCR with a cognate peptide-MHC complex, Tregs can recognize and respond to specific (auto-)antigens, which is central to their thymic development as well as their suppressive function once they leave the thymus[19, 20]. Several studies, including investigations in transgenic mice, previously showed that a restricted TCR repertoire of the Treg compartment can lead to the development of autoimmune disease [21–23], for example through a loss of tolerance towards commensal bacteria[24]. However, Tregs with a single TCR specificity can also inhibit autoimmune responses, thereby also providing some degree of protection against autoimmunity[25]. In humans, hyper-expanded Treg TCRβ clones can be found at the site of inflammation in JIA[26–28], and in refractory JIA patients hyper-expanded Tregs can even be found in circulation[29]. This expansion is likely caused by a dominance of specific (auto)antigens present at

7

target tissues. However, the exact antigen specificity and temporal and spatial dynamics of hyper-expanded effector T cells and Tregs in chronic inflammation and their relation to disease relapses remains to be established. Defining the specific CD4+ T cell subsets that are expanding in JIA patients is critical to decipher disease pathogenesis, and hyper-expanded T cells may represent novel therapeutic targets. Moreover, insight into the antigen specificity of local T cells may aid the discovery of disease-associated autoantigens.

Here, we had the unique opportunity to study autoimmune inflammation 1) within different affected sites at one single time point (spatial dynamics), and 2) over time (temporal dynamics), to get a detailed understanding of T cell dynamics during human autoimmune inflammation. To this end, we profiled the T cell composition of inflammatory exudate as well as peripheral blood obtained from JIA patients using mass cytometry. In addition, we performed TCRβ repertoire sequencing of Tregs as well as conventional CD4+ T cells (non-Tregs) derived from inflamed sites of JIA patients over time and space. Hyper-expanded TCRβ clones identified at the site of inflammation were highly overlapping between two affected joints that were sampled from the same individual at the same time. Moreover, the cellular distribution of two inflamed joints within a patient was almost identical, whereas distinct patterns were observed between patients. Furthermore, when comparing affected joints during multiple disease relapses, a large degree of overlap was observed in the dominant TCRβ clones, in the Treg as well as non-Treg compartments. Together, these data indicate a presence of re-circulating T(reg)-cells that very likely react to dominant auto-antigens continuously present in affected joints in JIA patients.

METHODS

Collection of SF and PB Samples

Patients with Juvenile Idiopathic Arthritis (JIA) were enrolled at the University Medical Center of Utrecht (The Netherlands). A total number of 9 JIA patients were included in this study. Of the JIA patients included, n = 2 were diagnosed with extended oligo JIA, n = 2 with rheumatoid factor negative poly-articular JIA, and n = 5 with oligo JIA, according to the revised criteria for JIA[30]. The average age at the time of (first) sample inclusion was 13,1 years (range 3,2 – 18,1 years) with a disease duration of 7,3 years (range 0.4 – 14.2 years).

Peripheral blood (PB) of JIA patients was obtained via veni-puncture or intravenous drip, while synovial fluid (SF) was obtained by therapeutic joint aspiration of affected joints. Informed consent was obtained from all patients either directly or from parents/guardians when the patients were younger than 12 years of age. The study was conducted in accordance with the Institutional Review Board of the University Medical Center Utrecht (approval no. 11-499/C). PB from n = 3 healthy children (average age 15,1 years with range 14,7 - 15,4 years) was obtained from a cohort of control subjects for a case-control clinical study. Samples were collected in compliance with the Declaration of Helsinki.

For cell isolation, SF was incubated with hyaluronidase (Sigma-Aldrich) for 30 min at 37°C to break down hyaluronic acid. Synovial fluid mononuclear cells (SFMCs) and peripheral blood mononuclear cells (PBMCs) were isolated using Ficoll Isopaque density gradient centrifugation (GE Healthcare Bio-Sciences, AB), and were used after freezing in Fetal Calf Serum (FCS) (Invitrogen) containing 10% DMSO (Sigma-Aldrich).

Flow cytometry and cell sorting

For TCR sequencing purposes, $CD3^+CD4^+CD25^{high}CD127^{low}$ Tregs and $CD3^+CD4^+CD25^{low/int}CD127^{int/high}$ non-Tregs were isolated from frozen PBMC and SFMC, using the FACS Aria III (BD). Antibodies used for sorting were: anti human CD3-BV510 (Biolegend), CD4-FITC (eBioscience), CD25-PE/Cy7 (BD), CD127-AF647 (Biolegend). To check for FOXP3 expression of the sorted populations anti human FOXP3-eF450 (eBioscience) was used.

CyToF

Frozen PBMCs and SFMCs were thawed and stained with a T cell focused panel of 37 heavy metal-conjugated antibodies (Supplemental Table 1), as previously described[31], and analyzed by CyToF-Helios. A detailed description of the methods for CyToF and analysis are provided in the supplementary material.

TCR sequencing

Tregs and non-Tregs were lysed in RLT buffer and frozen at -80°C. Between $0.15x10^6$ and $1x10^6$ Tregs, and between $0.46x10^6$ and $1x10^6$ non-Tregs were obtained for TCR sequencing. A detailed description of the methods for TCR sequencing and analysis are provided in the supplementary material

Statistical analyses

Nonparametric Mann Whitney (two-tailed) statistical test was performed in the manual gating of cellular subsets in FlowJo; p-values <0.05 were considered statistically significant. The correlation matrix for the node frequency was calculated using Spearman's rank-order correlation. Generation probabilities (P_{gens}) of TCRβ amino acid sequences were computed using OLGA[32]. Figures were produced using the R package ggplot2[33]. Venn diagrams were made on: *http://bioinformatics.psb.ugent.be/webtools/Venn/*.

7

RESULTS

Immune architecture of cellular infiltrates is similar between anatomically distinct inflamed sites

To study the peripheral and tissue specific immune cell composition in autoimmune disease, we profiled PBMCs and SFMCs from JIA patients with both knees affected at the time of sampling using CyToF. We set up a T cell centric panel (Supplemental Table 1) comprising 37 phenotypical and functional markers, as well as CD45 barcoding markers

to be able to study matched samples simultaneously. T-distributed stochastic neighbor embedding (t-SNE) and k-means clustering identified 22 immune cell populations in the SF/PB compartments (Figure 1A, p < 1e-21, and Supplemental Figure 1A and B). These populations could be broadly segregated into Treg subsets (CD25+/FoxP3+), naïve (CD45RA+), effector/memory T cells (CD45RA−), and non T cell populations (CD3−/CD4−/CD8−). Preliminary clustering of the median marker expression on T cells revealed a clear demarcation of SFMCs and PBMCs (Figure 1B), and a strong association of immune phenotypes between intra-individual paired knee SFMCs. Furthermore, density maps of immune cell populations within the t-SNE indicate strong dichotomy in the locations of SFMC and PBMC subsets (Figure 1C), with SFMCs occupying the bottom right and the PBMCs occupying the top left regions. Comparison of the node fingerprints between SFMC and PBMC samples (Figure 1D) revealed that SMFCs were enriched in CD4+CD25+FoxP3+ Tregs (node 2), and CD4+CD45RA− memory T cells (nodes 5, 9, 10), while PBMCs were enriched in CD45RA+ naïve T cell subsets (nodes 3, 6, 7, 13, 15). Next to this, a strikingly similar density cellular distribution profile was observed in the left and right knee joints of each JIA individual (Figure 1C), characterized by nearly identical node fingerprints (Figure 1D). The correlation matrix of the entire spectrum of node frequencies demonstrated a strong positive correlation between the SFMCs and their left and right joints, and an inverse negative correlation as compared with the PBMC populations (Supplemental Figure 1C). These results demonstrate that, while distinct differences in T cell signatures can be identified between PB and SF compartments, the phenotypic T cell architecture of distinct inflamed sites (left and right knees) are remarkably similar, indicating commonality in underlying disease etiology.

Effector T cells and Tregs are phenotypically similar across distinct inflamed sites

Next, we further functionally characterized the specific T cell subsets present in the inflamed joints. SF CD4+ and CD8+ T cell subsets displayed an increased expression of pro-inflammatory cytokines (TNFα, IFNγ and IL-6), indications of chronic TCR activation (PD1 and LAG3)[34] and a memory phenotype (CD45RA−), compared to their PBMC counterparts (Supplemental Figure 2A and 2B, p <0.05). Remarkably, the cytokine diversity of CD4+ memory T cells revealed nearly identical profiles for the left and right knee joints for each individual (Figure 2A), with minor inter-individual differences. This trend in cytokine profile was also reflected in the CD8+CD45RA- compartment (data not shown). The Treg (CD25+FOXP3+) population was significantly enriched in the synovial environment (Figure 2B, p<0.05 and Supplemental Figure 2C-D) with enhanced expression of T cell memory (CD45RA−) and Treg activation markers (HLA-DR and ICOS) compared with PBMC counterparts. Additionally, SF memory Tregs displayed a significantly higher proliferation (Ki67 staining) as compared to synovial effector memory T cells (Figure 2B, p<0.05), which was further confirmed by flow cytometry analysis (Supplemental Figure 2E). This indicates that Tregs belong to the most proliferative T cell subset in the inflamed environment. Moreover, memory Tregs showed very similar CTLA4/HLA-DR/ICOS/PD1 expression profiles in the left and right knee joints for each individual with some inter-individual differences (Figure 2C). Altogether, these data demonstrate that within JIA patients, there is an identical T cell phenotypic and functional
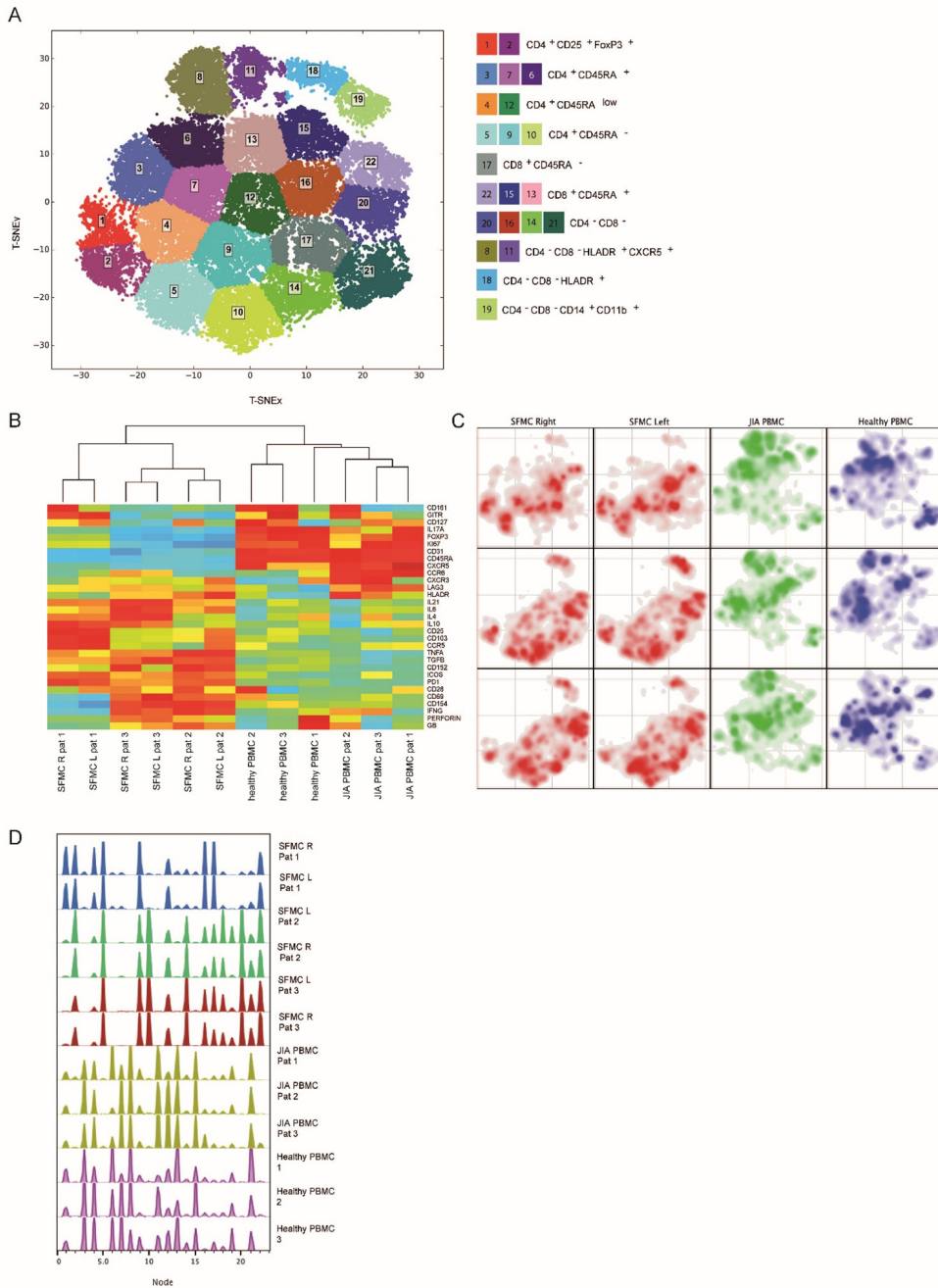
Figure 1. Overall immune architecture in left and right aff ected joint is very similar but distinct from peripheral blood. A Density maps based on T-SNE dimensional reduction and k-means clustering analysis on SFMC and PBMC samples, resulting in 22 cellular nodes. B Preliminary hierarchal clustering on the median expression of all markers, excluding lineage markers. C Density maps of immune cellular populations within the T-SNE maps. D Node frequency fingerprints showing the distribution across the nodes of SFMCs an PBMCs.
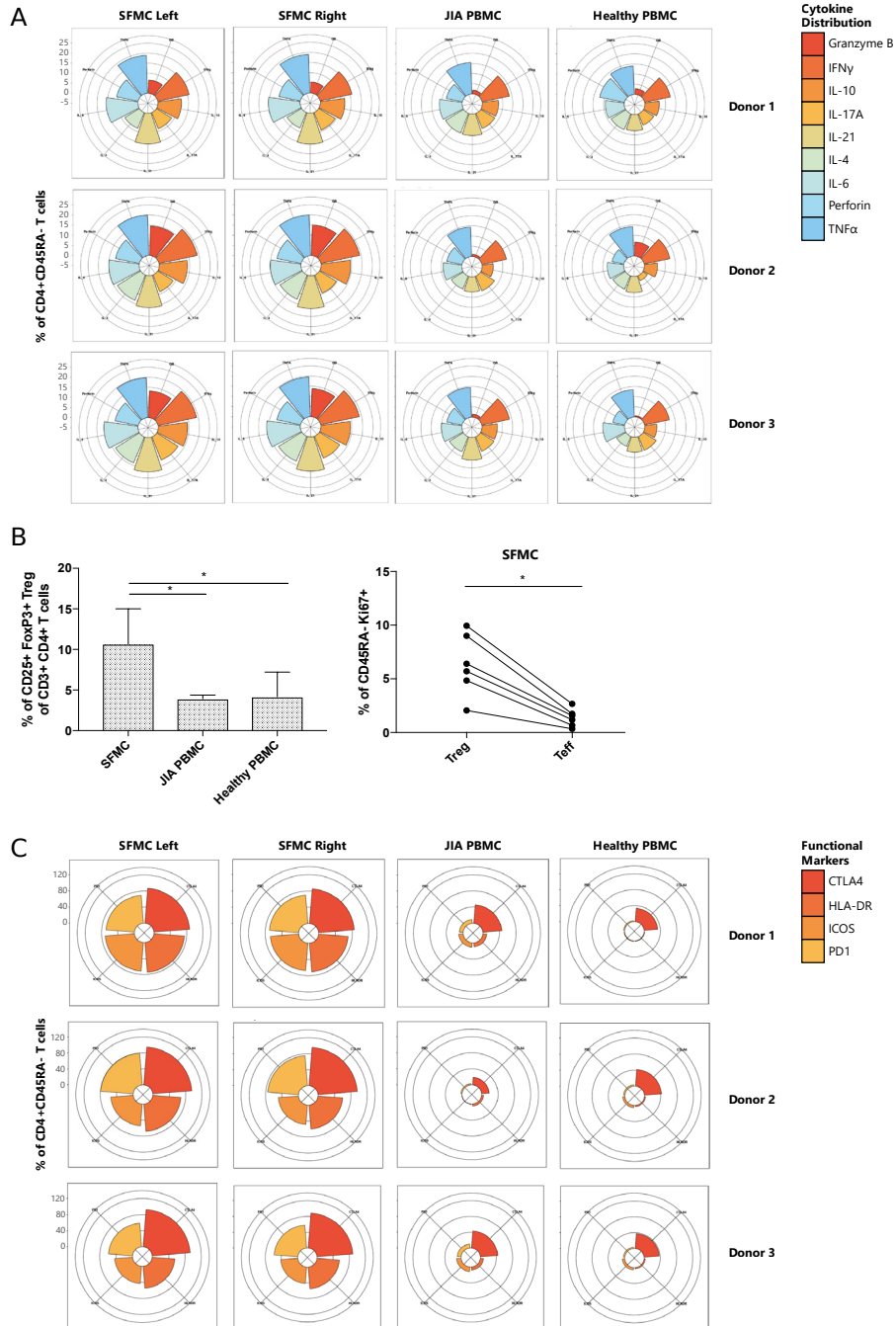
Figure 2. T cells display similar phenotypical and functional profiles at distinct inflamed locations. A Cytokine production of CD4+ CD45RA- memory cells depicted in radarplots. B Percentage CD25+FOXP3+ Treg of CD3+CD4+ cells in SFMC and PBMC of JIA patients and healthy children, and percentage of Ki67+ cells within CD45RA- cells in Treg and non-Treg in SFMC (non-parametric Mann-Whitney, * = p <0.05). C Expression of functional markers by CD25+ FOXP3+ CD45RA- cells.

profile present at separated inflamed locations, with increased amounts of activated and proliferating Treg populations.

Hyper-expanded T cell clones are shared between left and right joints

We hypothesized that the similarities in T cell phenotype and functional characteristics between two knees can be explained by infiltration and expansion of identical T cell clones present in both affected joints. To study whether the same cells infiltrate multiple joints, we sorted similar numbers of $CD3^+CD4^+CD25^+CD127^{low}$ Tregs and $CD3^+CD4^+CD25^-CD127^+$ non-Tregs from affected joints of JIA patients. The samples were derived from the same donors and time points as the ones used for CyToF analysis, regarding the first two patients. We examined the TCRβ repertoires of the isolated cells subsets by next generation sequencing. As expected, within the inflamed joints, clonally expanded cells were detected, which was more pronounced for Tregs compared to non-Tregs (Figure 3A). In line with the CyToF analysis, the distribution of the T cell clones was highly similar between left and right joint, both for Tregs and non-Tregs. Hyper-expanded T cells were further studied by sequential intersection of the most abundant TCRβ clonotypes across samples. Notably, we found a high degree of sharing between the two affected joints, while a small fraction of clones was shared between the SF compartment and PB (Figure 3B). Moreover, in addition to the increased clonality within the Treg compartment, sharing of clones between two locations was also more evident for Tregs compared to non-Tregs (Figure 3B).

More detailed analysis of dominant TCRs revealed that frequencies of hyper-expanded T cells were highly conserved between distinct anatomical sites, with the most dominant clones also detectable in PB (Figure 3C). To assess whether dominant clones were shared as a result of high generation probability (P_{gen}, convergent recombination[32]), or in response to antigen (convergent selection), we calculated the P_{gens} of shared and non-shared sequences and correlated these with their respective frequencies. This analysis revealed that frequencies of shared sequences were not positively correlated with P_{gen} (Figure 3D), while frequencies of non-shared sequences (i.e. those only appearing in the left or right knee joint) showed a significant positive correlation with P_{gen} (Figure 3E). Notably, this correlation was more pronounced for non-Tregs compared to Tregs (Figure 3E), indicating either bystander activation or non-antigen specific circulation of the non-shared TCR sequences in the non-Treg compartment. In summary, both non-Treg and Treg hyper-expanded T cell clones are shared between inflamed joints. This overlap is most pronounced for Treg, with the highly dominant Treg clones in SF also being detectable in circulation, likely driven by responses to shared antigens.

Dominant clones persist over time during relapsing remitting disease

Next, we questioned whether dominant clones are persistently present in inflamed joints over the relapsing-remitting course of JIA. To this end, we profiled the Treg and non-Treg TCRβ repertoire of SF as well as PB samples obtained from five JIA patients over time (Supplemental Figure 3). Repertoire overlap analysis (Jaccard index) showed that TCRβs of Treg samples obtained from SF were highly shared within patients over time (Figure 4A). Remarkably, this degree of sharing was also conserved across

7
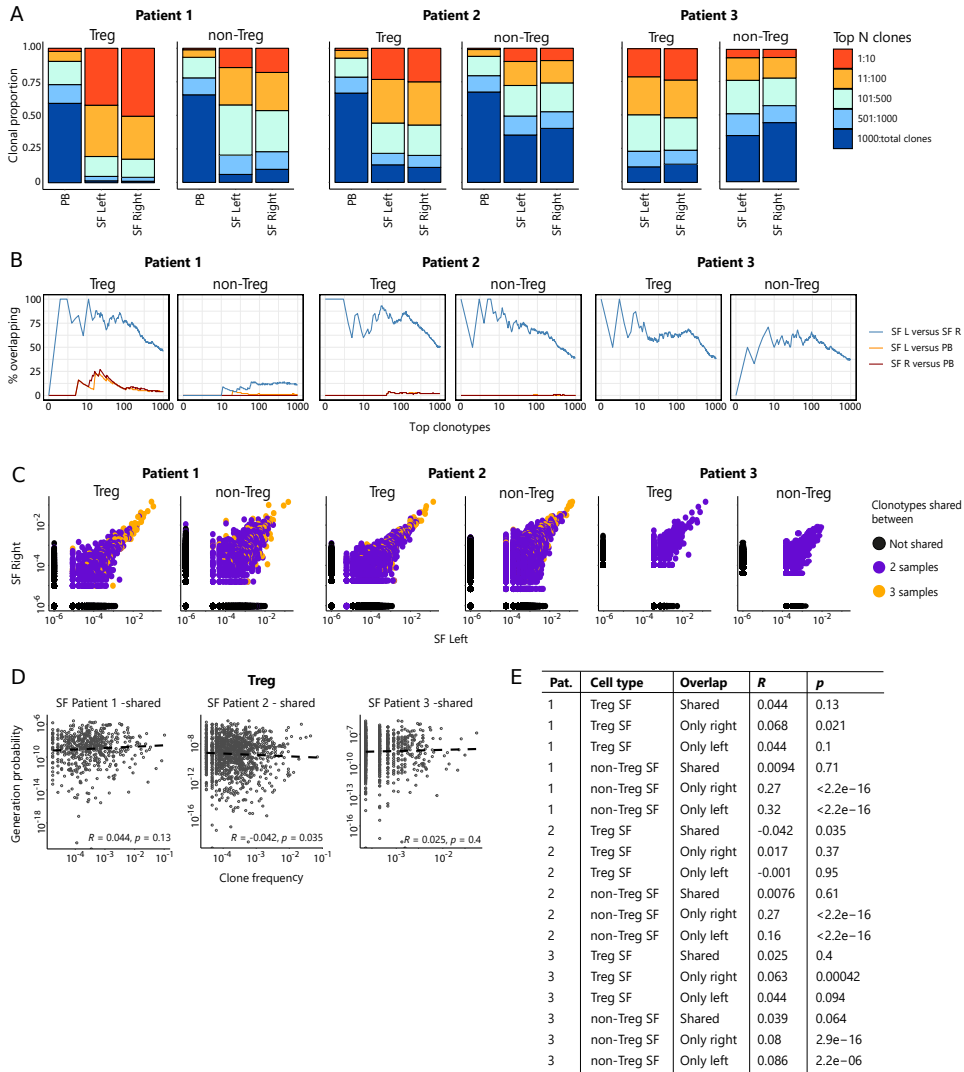
Figure 3. Highly dominant T cell clones are shared in SF from left and right joint and peripheral blood. A Clonal proportions of the TCRβ clones as detected in Treg and non-Treg sorted from PBMC, SFMC left joint, SFMC right joint of two different JIA patients. B Sequential intersection of abundant TCRβ clonotypes (based on amino acid sequence) across samples. Top clonotypes (ranging from 1-1000) are given on the x-axis, with the percentage of sequences overlapping between two given samples on the y-axis. C Frequency plots showing the overlapping Treg and non-Treg clones between left joint derived SF (x-axis) and right joint derived SF (y-axis), with color coding highlighting the clones that are shared with none of the other samples (black circle), shared in two samples (purple) and all three samples (PB, SF left, SF right; yellow). D Correlation (linear regression, dashed line) between frequency (x-axis) and generation probability (y-axis) of TCR clones shared across SF two samples. E Results of correlation between frequency and generation probability across all samples. Pat. = patient, R = Spearman's Rho, p = p-value.

different joints over time (Figure 4A/B, Supplemental Figure 4A). In contrast, TCRβs from PB did not cluster together over time, and showed much less overlap with their synovial counterparts (Figure 4A). More detailed analysis showed that the frequencies of shared TCRβs were also consistent over time, with the most dominant T cell clones having the highest degree of sharing (Figure 4C). Again, this phenomenon was more pronounced in Tregs obtained from SF compared to PB (Figure 4C), although the most dominant clones from synovial fluid were also detectable in PB (Supplemental Figure 5). Moreover, correlation of frequencies and P_{gens} of persistent sequences showed that persistent TCRβs with high abundance were not driven by recombination bias (Figure 4D), similar to what was observed for T cell clones shared between two knees sampled at the same time point (Figure 3D).

Next, we repeated our analysis on TCRβ sequences of non-Tregs from the same samples. Repertoire overlap analysis showed that, although non-Tregs also display sharing of TCRβ sequences over time (Figure 5A/B, Supplemental Figure 4B), the degree of sharing was less pronounced compared to Tregs (Figure 4A). Frequencies of highly shared TCRβs in non-Tregs were also consistent over time (Figure 5C), and frequencies and P_{gens} of persistent non-Treg sequences were not correlated (Figure 5D). Collectively, these data show that during relapsing-remitting disease, persistent dominant T cell clones are taking part in the local immune response in the inflamed joints of JIA patients, and this phenomenon is more pronounced for Tregs than non-Tregs as Tregs display a larger degree of repertoire persistence over time.

Patterns in similar TCR sequences are shared between JIA patient knees

Recent studies have demonstrated that immune responses against a particular antigen, involve T cell clones with similar TCR sequences[35–37]. This has led to the hypothesis that the degree of sequence similarity within the TCR repertoire is related to the antigen recognition coverage, with a high degree of sequence similarity pointing towards repertoire skewing by dominant antigens. Since we observed that the TCR repertoire of Tregs from JIA SF is persistent, both spatially (across two knees) and temporally (over time), we hypothesized that persistent T cell clones cluster together with other, similar T cell clones that are involved in responses against the same antigens. To investigate this, we performed TCR similarity analysis, focusing on SF samples obtained from two affected knees. Using k-mer overlap analysis, we constructed TCR similarity networks for JIA patients and compared these to networks generated from random repertoires with the same number of TCRβ sequences (Figure 6A). In line with our expectations, TCR networks from JIA patients were highly connected (much more than expected by chance), showing that patient repertoires exhibit a high degree of sequence similarity (Figure 6B). Moreover, clusters within JIA networks were composed of TCR sequences coming from distinct knees, as well as overlapping sequences, indicating that non-persistent sequences are homologous to persistent sequences and therefore also likely involved in the same antigen specific responses. Notably, in the random repertoires, clusters were less mixed (as indicated by a high cluster purity) compared to JIA networks (Figure 6C), again highlighting that TCRs from JIA samples display higher sequence similarity than expected by chance. Overall, these results show that the SF Treg repertoire is highly

7
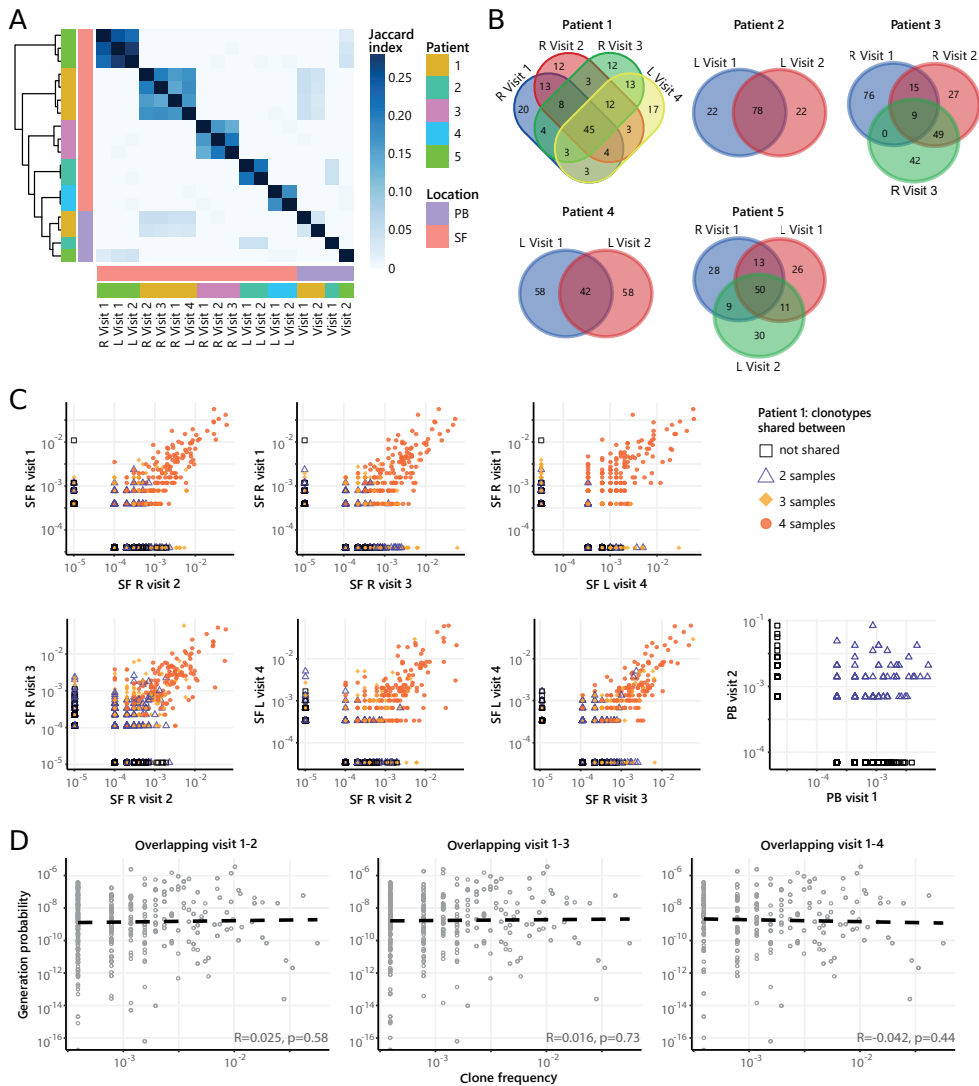
Figure 4. Persistence of Treg clones over the course of relapse remitting disease. A Heatmap showing overlap (Jaccard index, light blue = limited overlap, darkblue = high overlap) of Treg derived TCR sequences obtained from SF or PB from JIA patients over time. L = left knee, R = right knee. B Venn diagrams displaying the 100 most abundant unique TCRβ clones, defined by amino acid sequence, for longitudinal SF samples from all patients. C Frequency plots showing the overlapping Treg clones between visits for SF and PB, with color coding and shapes highlighting the number of samples in which unique clones are found. R = right, L = left. D Correlation (linear regression, dashed line) between frequency (x-axis) and generation probability (y-axis) of TCR clones shared across two visits for SF samples.
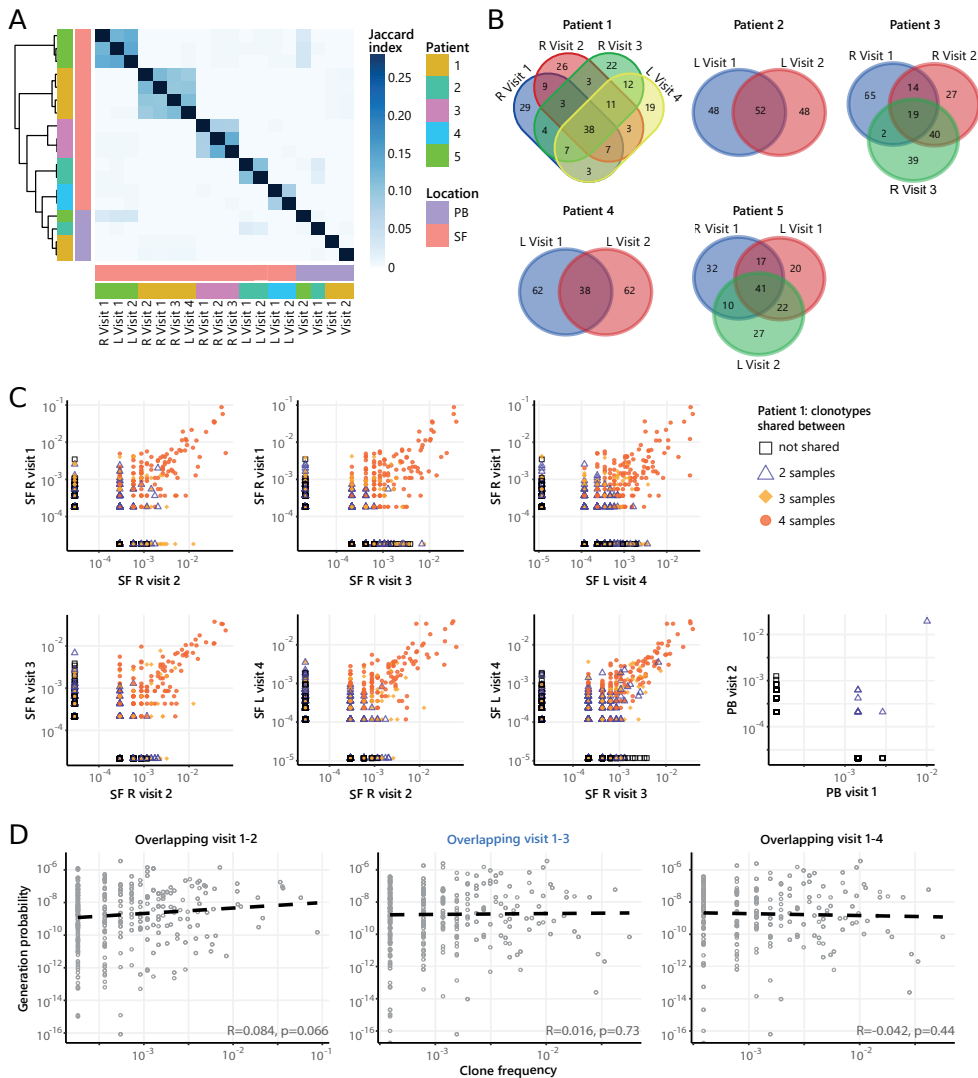
Figure 5. Persistence of non-Treg clones over the course of relapse remitting disease. A Heatmap showing overlap (Jaccard index, light blue = limited overlap, darkblue = high overlap) of non-Treg derived TCR sequences obtained from SF or PB from JIA patients over time. L = left knee, R = right knee. B Venn diagrams displaying the 100 most abundant unique TCRβ clones, defined by amino acid sequence, for longitudinal SF samples from all patients. C Frequency plots showing the overlapping non-Treg clones between visits for SF and PB, with color coding and shapes highlighting the number of samples in which unique clones are found. R = right, L = left. D Correlation (linear regression, dashed line) between frequency (x-axis) and generation probability (y-axis) of TCR clones shared across two visits for SF samples.
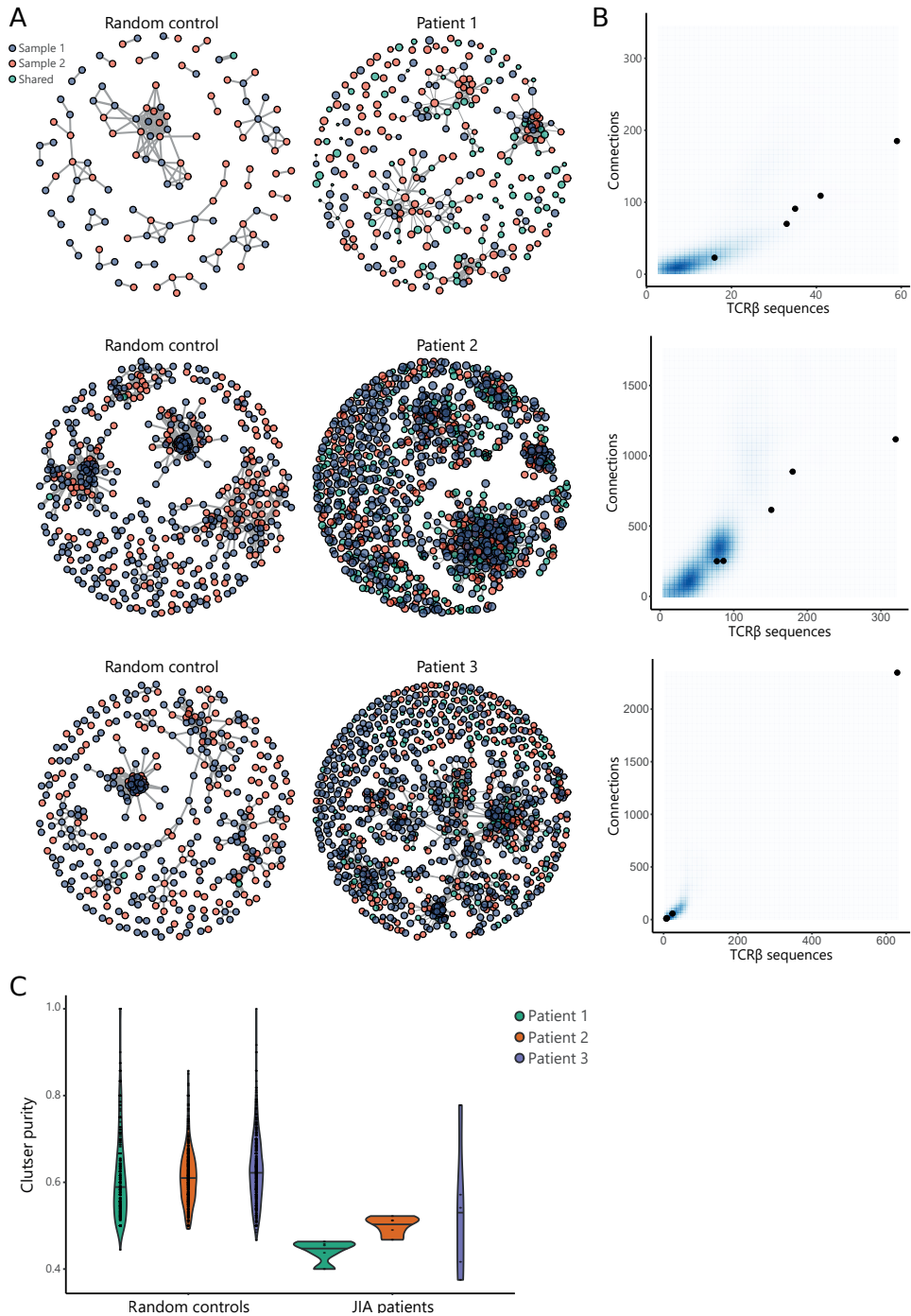
Figure 6. TCR similarly analysis of sequences found across distinct JIA patient knees. A TCR similarity networks based on amino acid k-mer sharing (k = 3) between TCR sequences. Every node represents one TCRβ sequence, with sequences present in one sample (SF from left or right knees) highlighted in blue and orange, and sequences shared across two samples highlighted in *(continued)*

green. Nodes are connected if TCRs share at least 8 k-mers. Networks from JIA patient repertoires (right) are compared to random repertoires (left), with the same repertoire size. B Number of TCR sequences (x-axis) and their connections (y-axis) to other TCR sequences of the top five similarity clusters identified in A. Blue density maps depict clusters identified in random repertoires (N=100), while black circles depict clusters identified in JIA patients. C Cluster purity (y-axis, %) for the top five clusters identified in random repertoires (RC), and JIA patient TCR similarity networks. Numbers indicate p-value of difference between RC and JIA (Mann-Whitney).

skewed by antigenic selection.

DISCUSSION

In this study, we provide the first CyToF and TCRβ sequencing analysis of purified Tregs and non-Tregs, uncovering their spatial and temporal behavior in a human autoimmune disease setting. We show that the architecture of T cell responses from distinct inflamed joints is remarkably similar, even though inflammation often starts at different times in the two joints. This response is characterized by a pro-inflammatory cytokine production by effector cells, as well as expanding Tregs with an activated profile. Moreover, the local immune response is dominated by a limited number of TCRβ clones, of which especially Tregs are locally expanding. During relapsing-remitting disease course, hyper-expanded T cell clones are persistently present, a process likely driven by chronic antigen exposure. Altogether, these data indicate that there is a strong driving force locally in the joint that heavily skews the TCRβ repertoire, which can be explained by the presence of dominant (auto-)antigens.

Although the antigen(s) driving T cell activation and expansion in JIA remain elusive, our data provide strong support for the presence of ubiquitously expressed auto-antigens given the observed overlap in dominant clones over time and in space. Given the tissue restrictive character of the JIA, it is tempting to speculate that the potential antigen would be joint-specific, although it has been shown that ubiquitously expressed auto-antigens can also induce joint-specific autoimmune disease[38, 39]. We show that SF Tregs have high expression of Ki67 (marking proliferation and thus recent antigen encounter), suggesting that these cells actively respond to synovial antigens. Moreover, we show that the expansion of dominant TCR clones is not driven by convergent recombination (i.e. dependent on generation probabilities), further highlighting that antigen are driving T cell activation. Further support for the hypothesis that persistent, hyper-expanded Tregs found in JIA SF are auto-reactive is provided by a recent study performed in mice with type 1 diabetes, where Tregs with a high degree of self-reactivity were found to be expanding locally in affected pancreatic islets and displayed a specific profile with elevated levels of GITR, CTLA-4, ICOS and Ki67, very similar to our observations[40].

Our data demonstrated that dominant T cell clones in SF can be traced back in circulation. Together with observations that similar T cell clones are detected in multiple affected joints and the obvious overlap in immune cell composition, this strongly suggests that T cells migrate from the joint to peripheral blood and vice versa. This could mean that Tregs are either recirculating, or actively being replenished from circulating (precursor)

7

T cells. These observations are in line with other recent studies in arthritis showing that synovial CD4+ T cells and Treg clones can also be detected in peripheral blood[27, 41], where their presence correlates with disease activity and response to therapy[27, 42]. Next to that, in refractory JIA and juvenile dermatomyositis (JDM) patients that underwent an autologous hematopoietic stem cell transplantation (aHSCT) as a last resort therapy, the peripheral T cell compartment was shown to be heavily skewed towards dominant T cells clones prior to transplantation, especially in the Treg compartment[29]. After transplantation, the TCR repertoire diversified in responders whereas in the only non-responder a clonal repertoire remained. A renewal of the TCR repertoire of CD4+ T cells and an increased diversity of the Treg in peripheral blood has also been observed in other autoimmune patients undergoing aHSCT. The latter appears to be important for inducing successful remission post-transplantation[42]. This knowledge, combined with our findings that the same T cell clones dominate the immune response at different sites of inflammation and the persistence of the same clones in the relapsing-remitting course of disease, strengthen the possibility to use circulating disease-associated T cell clones for disease monitoring or prognostic purposes. However, to accurately monitor and predict which T cell clones from peripheral blood are also implicated in active immune processes in joints, more detailed phenotyping is needed to fully characterize the functional profile and origins of dominant clones. Multi-omic single-cell profiling to link TCR antigen specificity with gene expression will give novel and required information to bring this closer to the clinic.

The existence of a temporal and spatially persistent clonal Treg TCR repertoire, and its correlation with disease severity as outlined above, raises questions about the role of Tregs in (localized) autoimmune responses. An important question that needs to be addressed here is to what degree clonally expanded Tregs can modulate inflammation over the course of an autoimmune response. Various studies have shown that Tregs in JIA maintain their suppressive capacity, but local effector T cells are resistant to this suppression[9, 43]. Thus, the clonotypic expansion in SF Treg cells might reflect an insufficient attempt to control expanding effector T cells. The importance of a diverse Treg repertoire is shown in several mouse models[21–24]. Föhse *et al.* showed that Tregs with a higher diversity are able to expand more efficiently compared to Treg with a lower diversity in mice with TCR restricted conventional T cells[22]. It has been suggested that this is due to the TCR diverse Tregs having access to more ligands and as a result being able to out-compete the TCR-restricted Treg cells[16]. However, this applies for circulating Treg, and whether this would also be important for Treg in tissues is not known. The finding that tissue Treg residing in healthy tissues also show a considerable oligoclonality regarding their TCR repertoire may indicate that this is a normal feature[44, 45]. Additionally, it was recently shown that a diverse Treg repertoire in mice is especially needed to control Th1 responses, whereas Th2 and Th17 responses were still suppressed by single Treg clones[25]. This could be an explanation why the Th1 rich SF environment is poorly controlled by the large amount of clonally expanded Tregs. Thus, hyper-expanded Tregs alone might not be sufficient to prevent or inhibit autoimmune responses, and future Treg centric therapies should take this into account.

It should be noted that in this study we sequenced the β-chain of the TCR and

not the α-chain. The identified dominant TCRβ clones can pair with several α-chains, possibly leading to less overlapping TCR repertoire and a different Ag specificity. Future sequencing of both TCR chains will provide insight into the total TCR repertoire. Next to that, we are aware of a possible amplification bias because of a difference in efficiency of PCR primers. However, in our analysis approach we attempted to control as much as possible for such biases. An interesting next step would be to combine single cell RNA-sequencing with identification of the TCR to directly link the expression profile of a given cell to its TCR clonotype and facilitate the identification of the antigenic target and its HLA class II restriction.

In conclusion, we show that in SF the immune cell architecture is marked by inflammatory responses of activated effector T cells as well as activated and highly expanding Tregs. The remarkable overlap in immune cell composition as well as the dominant clones over time and in space provide indications for a powerful driving force that shapes the local T cell response during joint inflammation. The presence of these inflammation-associated clones in the circulation provide promising perspectives for use in disease monitoring. Moreover, the high degree of sequence similarity observed between Treg clones obtained from distinct inflamed joints indicates that antigen selection significantly reshapes the local Treg repertoire. Further research is needed to pinpoint these driving antigens and to create opportunities to target disease-specific T cells.

7

REFERENCES

1. T. David, S.F. Ling, A. Barton. Genetics of immune-mediated inflammatory diseases. *Clin. Exp. Immunol*. vol. 193, no. 1, pp. 3–12. Jul. 2018.

2. A.P.B. Black, H. Bhayani, C.A.J. Ryder, *et al*. T-cell activation without proliferation in juvenile idiopathic arthritis. *Arthritis Res. Ther*. vol. 4, no. 3, pp. 177–183. 2002.

3. E.J. Wehrens, B.J. Prakken, F. van Wijk. T cells out of control—impaired immune regulation in the inflamed joint. *Nat. Rev. Rheumatol*. vol. 9, pp. 34–42. 2013.

4. S.A. Long, J.H. Buckner. CD4 + FOXP3 + T Regulatory Cells in Human Autoimmunity: More Than a Numbers Game. *J. Immunol*. vol. 187, no. 5, pp. 2061–2066. Sep. 2011.

5. B.V. Kumar, W. Ma, M. Miron, *et al*. Human Tissue-Resident Memory T Cells Are Defined by Core Transcriptional and Functional Signatures in Lymphoid and Mucosal Sites. *Cell Rep*. vol. 20, no. 12, pp. 2921–2934. Sep. 2017.

6. K. Nistala, S. Adams, H. Cambrook, *et al*. Th17 plasticity in human autoimmune arthritis is driven by the inflammatory environment. *Proc. Natl. Acad. Sci*. vol. 107, no. 33, pp. 14751–14756. Aug. 2010.

7. L. Cosmi, R. Cimaz, L. Maggi, et al. Evidence of the transient nature of the Th17 phenotype of CD4+CD161+ T cells in the synovial fluid of patients with juvenile idiopathic arthritis. *Arthritis Rheum*. vol. 63, no. 8, pp. 2504–2515. Aug. 2011.

8. K. Ohl, H. Nickel, H. Moncrieffe, *et al*. The transcription factor CREM drives an inflammatory phenotype of T cells in oligoarticular juvenile idiopathic arthritis. *Pediatr. Rheumatol*. vol. 16, no.1, pp. 39. Jun. 2018.

9. E.J. Wehrens, G. Mijnheer, C.L. Duurland, *et al*. Functional human regulatory T cells fail to control autoimmune inflammation due to PKB/c-akt hyperactivation in effector cells. *Blood*. vol 118, no. 13, pp. 3538–3548. Sep. 2011.

10. C.L. Duurland, C.C. Brown, R.F.L. O'Shaughnessy, *et al*. CD161$^+$ Tconv and CD161$^+$ Treg Share a Transcriptional and Functional Phenotype despite Limited Overlap in TCRβ Repertoire. *Front. Immunol*. vol. 8, pp. 103. Mar. 2017.

11. P.A. Muraro, R. Cassiani-Ingoni, K. Chung, *et al*. Clonotypic analysis of cerebrospinal fluid T cells during disease exacerbation and remission in a patient with multiple sclerosis. *J. Neuroimmunol*. vol. 171, no. 1–2, pp. 177–183. Feb. 2006.

12. C.G. Chapman, R. Yamaguchi, K. Tamura, *et al*. Characterization of T-cell Receptor Repertoire in Inflamed Tissues of Patients with Crohn's Disease Through Deep Sequencing. *Inflamm. Bowel Dis*. vol. 22, no. 6, pp. 1275–1285. Jun. 2016.

13. M.E. Doorenspleet, L. Westera, C.P. Peters, *et al*. Profoundly Expanded T-cell Clones in the Inflamed and Uninflamed Intestine of Patients With Crohn's Disease. *J. Crohn's Colitis*. vol. 11, no. 7, pp. 831–839.

14. S. Günaltay, D. Repsilber, G. Helenius, *et al*. Oligoclonal T-cell Receptor Repertoire in Colonic Biopsies of Patients with Microscopic Colitis and Ulcerative Colitis. *Inflamm. Bowel Dis*. vol. 23, no. pp. 932–945. Jun. 2017.

15. A. Musters, P.L. Klarenbeek, M.E. Doorenspleet, *et al*. In Rheumatoid Arthritis, Synovitis at Different Inflammatory Sites Is Dominated by Shared but Patient-Specific T Cell Clones. *J. Immunol*. vol. 201, no. 2, pp. 417–422. Jul. 2018.

16. J.B. Wing, S. Sakaguchi. TCR diversity and Treg cells, sometimes more is more. *Eur. J. Immunol*. vol. 41, no. 11, pp. 3097–3100. Oct. 2011.

17. M.W.L. Leung, S. Shen, J.J. Lafaille. TCR-dependent differentiation of thymic Foxp3+ cells is limited to small clonal sizes. *J. Exp. Med*. vol. 206, no. 10, pp. 2121–2130. Sep. 2009.

18. C. Wang, C.M. Sanders, Q. Yang, *et al*. High throughput sequencing reveals a complex pattern of dynamic interrelationships among human T cell subsets. *Proc. Natl. Acad. Sci*. vol. 107, no. 4, pp. 1518–1523. Jan. 2010.

19. A.G. Levine, A. Arvey, W. Jin, *et al*. Continuous requirement for the TCR in regulatory T cell function. *Nat. Immunol*. vol. 15, no. 11, pp. 1070–1078. Nov. 2014.

20. M.O. Li, A.Y. Rudensky. T cell receptor signalling in the control of regulatory T cell differentiation and function. *Nat. Rev. Immunol*. vol. 16, no.4, pp. 220–233. Apr. 2016.

21. D. Adeegbe, T. Matsutani, J. Yang, *et al*. CD4 + CD25 + Foxp3 + T Regulatory Cells with Limited TCR Diversity in Control of Autoimmunity. *J. Immunol*. vol. 184, no. 1, pp. 56–66. Jan. 2010.

22. L. Föhse, J. Suffner, K. Suhre, *et al*. High TCR diversity ensures optimal function andhomeostasis of Foxp3 + regulatory Tcells. *Eur. J. Immunol*. vol. 41, no. 11, pp. 3101–3113. Nov. 2011.

23. A. Yu, M.J. Dee, D. Adeegbe, et al. The Lower Limit of Regulatory CD4 + Foxp3 + TCRβ Repertoire Diversity Required To Control Autoimmunity. *J. Immunol.* vol. 198, no. 8. pp. 3127–3135. Apr. 2017.
24. J. Nishio, M. Baba, K. Atarashi, *et al.* Requirement of full TCR repertoire for regulatory T cells to maintain intestinal homeostasis. *Proc. Natl. Acad. Sci.* vol. 112, no. 41, pp. 12770–12775. Oct. 2015.
25. A.G. Levine, S. Hemmers, A.P. Baptista, *et al.* Suppression of lethal autoimmunity by regulatory T cells with a single TCR specificity. *J. Exp. Med.* vol. 214, no. 3, pp. 609–622. Mar. 2017.
26. D. Bending, E. Giannakopoulou, H. Lom, *et al.* Synovial Regulatory T Cells Occupy a Discrete TCR Niche in Human Arthritis and Require Local Signals To Stabilize FOXP3 Protein Expression. *J. Immunol.* vol. 195, no. 12, pp. 5616–5624. Dec. 2015.
27. M. Rossetti, R. Spreafico, A. Consolaro, *et al.* TCR repertoire sequencing identifies synovial Treg cell clonotypes in the bloodstream during active inflammation in human arthritis. *Ann. Rheum. Dis.* vol. 76, no. 2, pp. 435–441. Feb. 2017.
28. L.A. Henderson, S. Volpi, F. Frugoni, *et al.* Next-Generation Sequencing Reveals Restriction and Clonotypic Expansion of Treg Cells in Juvenile Idiopathic Arthritis. *Arthritis Rheumatol.* vol. 68, no. 7, pp. 1758–1768. Jul. 2016.
29. E.M. Delemarre, T. van den Broek, G. Mijnheer, *et al.* Autologous stem cell transplantation aids autoimmune patients by functional renewal and TCR diversification of regulatory T cells. *Blood.* vol. 127, no. 1, pp. 91–101. Jan. 2016.
30. R.E. Petty, T.R. Southwood, J. Baum, *et al.* Revision of the proposed classification criteria for juvenile idiopathic arthritis: Durban, 1997. *Journal of Rheumatology.* vol. 25, no. 10, pp. 1991–1994. Oct. 1998.
31. V. Chew, Y.H. Lee, L. Pan, *et al.* Immune activation underlies a sustained clinical response to Yttrium-90 radioembolisation in hepatocellular carcinoma. *Gut.* vol. 68, no. 2, pp. 335–346. Feb. 2019.
32. Z. Sethna, Y. Elhanati, C.G. Callan, *et al.* OLGA: fast computation of generation probabilities of B- and T-cell receptor amino acid sequences and motifs. *Bioinformatics.* vol. 35, no. 17, pp. 2974–2981. Sep. 2019.
33. H. Wickham. ggplot2: Elegant Graphics for Data Analysis. *Springer-Verlag New York.* 2016.
34. A. Petrelli, G. Mijnheer, D.P. Hoytema van Konijnenburg, *et al.* PD-1+CD8+ T cells are clonally expanding effectors in human chronic inflammation. *J. Clin. Invest.* vol. 128, no. 10, pp. 4669–4681. Oct. 2018.
35. P. Dash, A.J. Fiore-Gartland, T. Hertz, *et al.* Quantifiable predictive features define epitope-specific T cell receptor repertoires. *Nature.* vol. 547, no. 7661, pp. 89–93. Jul. 2017.
36. J. Glanville, H. Huang, A. Nau, *et al.* Identifying specificity groups in the T cell receptor repertoire. *Nature.* vol. 547, no. 7661, pp. 94–98. Jul. 2017.
37. M.V. Pogorelyy, A.A. Minervina, M. Shugay, *et al.* Detecting T cell receptors involved in immune responses from single repertoire snapshots. P*LOS Biol.* vol. 17, no. 6, pp. e3000314. Jun. 2019.
38. Y. Ito, M. Hashimoto, K. Hirota, *et al.* Detection of T cell responses to a ubiquitous cellular protein in autoimmune disease. *Science.* vol. 346, no. 6207, pp. 363–368. Oct. 2014.
39. L. Mandik-Nayak, B.T. Wipke, F.F. Shih, *et al.* Despite ubiquitous autoantigen expression, arthritogenic autoantibody response initiates in the local lymph node. *Proc. Natl. Acad. Sci.* vol. 99, no. 22, pp. 14368–14373. Oct. 2002.
40. M.L. Sprouse, M.A. Scavuzzo, S. Blum, *et al.* High self-reactivity drives T-bet and potentiates Treg function in tissue-specific autoimmunity. *JCI Insight.* vol. 3, no. 2, pp. e97322. Jan. 2018.
41. R. Spreafico, M. Rossetti, J. van Loosdregt, *et al.* A circulating reservoir of pathogenic-like CD4 + T cells shares a genetic and phenotypic signature with the inflamed synovial micro-environment. *Ann. Rheum. Dis.* vol. 75, no. 2, pp. 459–465. Feb. 2016.
42. L. Lutter, J. Spierings, F.C.C. van Rhijn-Brouwer, *et al.* Resetting the T Cell Compartment in Autoimmune Diseases With Autologous Hematopoietic Stem Cell Transplantation: An Update. *Front. Immunol.* vol. 9, pp. 767. Apr. 2018.
43. S. Haufe, M. Haug, C. Schepp, *et al.* Impaired suppression of synovial fluid CD4+CD25– T cells from patients with juvenile idiopathic arthritis by CD4+CD25+ Treg cells. *Arthritis Rheum.* vol. 63, no. 10, pp. 3153–3162. Oct. 2011.
44. D. Burzyn, W. Kuswanto, D. Kolodin, *et al.* A Special Population of Regulatory T Cells Potentiates Muscle Repair. *Cell.* vol. 155, no. 6, pp. 1282–1295. Dec. 2013.
45. R. Sanchez Rodriguez, M.L. Pauli, I.M. Neuhaus, *et al.* Memory regulatory T cells reside in human skin. *J. Clin. Invest.* vol. 124, no. 3, pp. 1027–1036. Mar. 2014.

7

SUPPLEMENTARY INFORMATION

Supplementary Methods

CyToF and CyToF data analysis

Briefly, PBMCs were stimulated with or without phorbol 12-myristate 13-acetate (150 ng/ml, Sigma-Aldrich) and ionomycin (750 ng/ml, Sigma-Aldrich) for 4 hours, and blocked with secretory inhibitors, brefeldin A (1:1000, eBioscience) and monensin (1:1000, Biolegend) for the last 2 hours. The cells were then washed and stained with cell viability dye cisplatin (200 µM, Sigma-Aldrich). Each individual sample was barcoded with a unique combination of anti-CD45 conjugated with either heavy metal 89, 115, 141 or 167, as previously described(Lai et al., 2015). Barcoded cells were washed and stained with the surface antibody cocktail for 30 min on ice, and subsequently washed and re-suspended in fixation/permeabilization buffer (permeabilization buffer, eBioscience) for 45 min on ice. Permeabilized cells were subsequently stained with an intra-cellular antibody cocktail for 45 min on ice, followed by staining with a DNA intercalator Ir-191/193 (1:2000 in 1.6% w/v paraformaldehyde, Fluidigm) overnight at 4°C or for 20 min on ice. Finally, the cells were washed and re-suspended with EQ™ Four Element Calibration beads (1:10, Fluidigm) at a concentration of 1x106 cells/ml. The cell mixture was then loaded and acquired on a Helios mass cytometer (Fluidigm) calibrated with CyToF Tunning solution (Fluidigm). The output FCS files were randomized and normalized with the EQ™ Four Element Calibration beads against the entire run, according to the manufacturer's recommendations.

Normalized CyToF output FCS files were de-barcoded manually into individual samples in FlowJo (v.10.2), and down-sampled to equal cell events (5000 cells) for each sample. Batch run effects were assessed using an internal biological control (PBMC aliquots from the same healthy donor for every run). Normalized cells were then clustered with MarVis (Kaever et al., 2009), using Barnes Hut Stochastic Neighbor Embedding (SNE) nonlinear dimensionality reduction algorithm and k-means clustering algorithm, as previously described (Chew et al., 2019). The default clustering parameters were set at perplexity of 30, and p<1e-21. The cells were then mapped on a 2-dimensional t-distributed SNE scale based on the similarity score of their respective combination of markers, and categorized into nodes (k-means). To ensure that the significant nodes obtained from clustering were relevant, we performed back-gating of the clustered CSV files and supervised gating of the original FCS files with FlowJo as validation. Visualizations (density maps, node frequency fingerprint, node phenotype, radar plots) were performed through R scripts and/or Flow Jo (v.10.2). Correlation matrix and node heatmaps were generated using MarVis (Kaever et al., 2009) and PRISM (v 7.0).

TCR sequencing and analysis

Total RNA was isolated using the RNeasy Mini Kit (Qiagen) for cell fractions ≥0.2x106 cells and the RNeasy Micro Kit (Qiagen) for fractions ≤0.2x10^6 cells, following the manufacturer's instructions. cDNA was synthesized using the SMARTer RACE cDNA Amplification kit (Clontech). Amplification of the TCRβ VDJ region was performed using

previously described primers and amplification protocols (Zhou et al., 2006). PCR product fragment size was analyzed using the QIAxcel Advanced System (Qiagen). End repair and barcode adapter ligation were performed with the NGSgo®-LibrX and NGSgo®-IndX (GenDx) according to the manufacturer's instructions. Cleanup of the samples was performed after each step using HighPrep PCR beads and following the manufacturer's instructions (GC Biotech). Paired-end next-generation sequencing was performed on the Illumina MiSeq system 500 (2 x250 bp) (Illumina). TCR sequencing analysis was performed using RTCR as previously described (Gerritsen et al., 2016).

TCR network analysis

For sequence similarity analysis, we counted the presence of overlapping 3-mer amino acid segments (defined as k-mers) in the TCRβ (CDR3) sequences. TCR sequences were considered similar when they shared at least 8 k-mers, independent of the total sequence length. Random repertoires were generated using the generative model of V(D)J recombination implemented in OLGA (Sethna et al., 2019). For equal comparison to biological samples, random repertoires were down sampled to equal the number of TCR sequences. Cluster purity was calculated as the ratio of number of TCR sequences from the most abundant sequence within the cluster and the total number of TCR sequences in the cluster.

References

V. Chew, Y.H. Lee, L. Pan, *et al.* Immune activation underlies a sustained clinical response to Yttrium-90 radioembolisation in hepatocellular carcinoma. *Gut*. vol. 68, no. 2, pp. 335–346. Feb. 2019.

B. Gerritsen, A. Pandit, A.C. Andeweg, *et al.* RTCR: A pipeline for complete and accurate recovery of T cell repertoires from high throughput sequencing data. *Bioinformatics*. vol. 32, no. 20, pp. 3098–3106. Oct. 2016.

A. Kaever, T. Lingner, K. Feussner, *et al.* MarVis: a tool for clustering and visualization of metabolic biomarkers. *BMC Bioinformatics*. vol. 10, no. 1, pp. 92. Mar. 2009.

L. Lai, R. Ong, J. Li, *et al.* A CD45-based barcoding approach to multiplex mass-cytometry (CyTOF). *Cytometry Part A*. vol. 87, no. 4, pp. 369–374. Apr. 2015.

Z. Sethna, Y. Elhanati, C.G. Callan, *et al.* OLGA: fast computation of generation probabilities of B- and T-cell receptor amino acid sequences and motifs. *Bioinformatics*. vol. 35, no. 17, pp. 2974–2981. Sep. 2019.

D. Zhou, R. Srivastava, V. Grummel, *et al.* High throughput analysis of TCR-β rearrangement and gene expression in single T cells. *Laboratory Investigation*. vol. 86, no. 3, pp. 314–321. Mar. 2006.

7
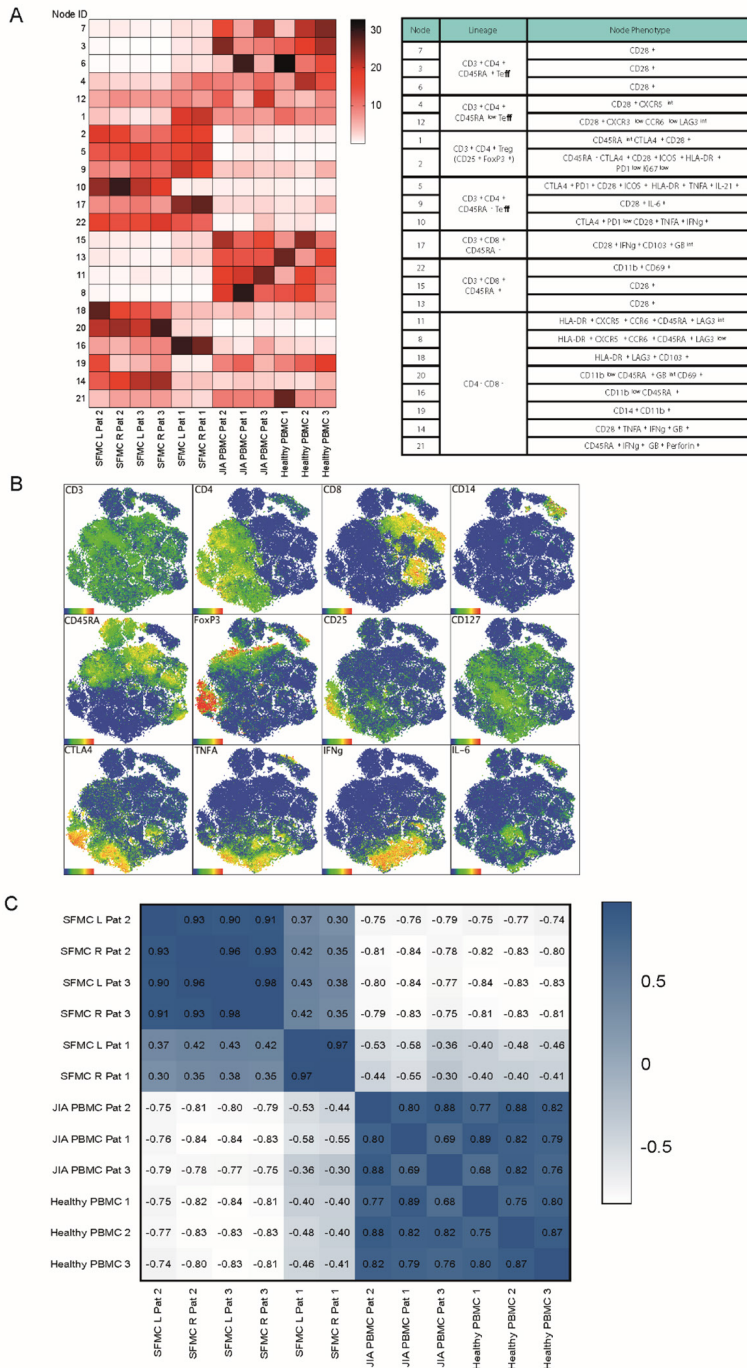
Supplemental Figure 1. Preliminary analysis reveals correlation between SFMC from distinct joints. A Node frequency showing the distribution of T cell markers across the nodes of SFMCs an PBMCs in the CyToF analysis. B Marker expression of t-SNE dimensional reduction and k-means clustering analysis on SFMC and PBMC samples C Correlation matrix using spearman correlation of the entire spectrum of node frequency given in A.
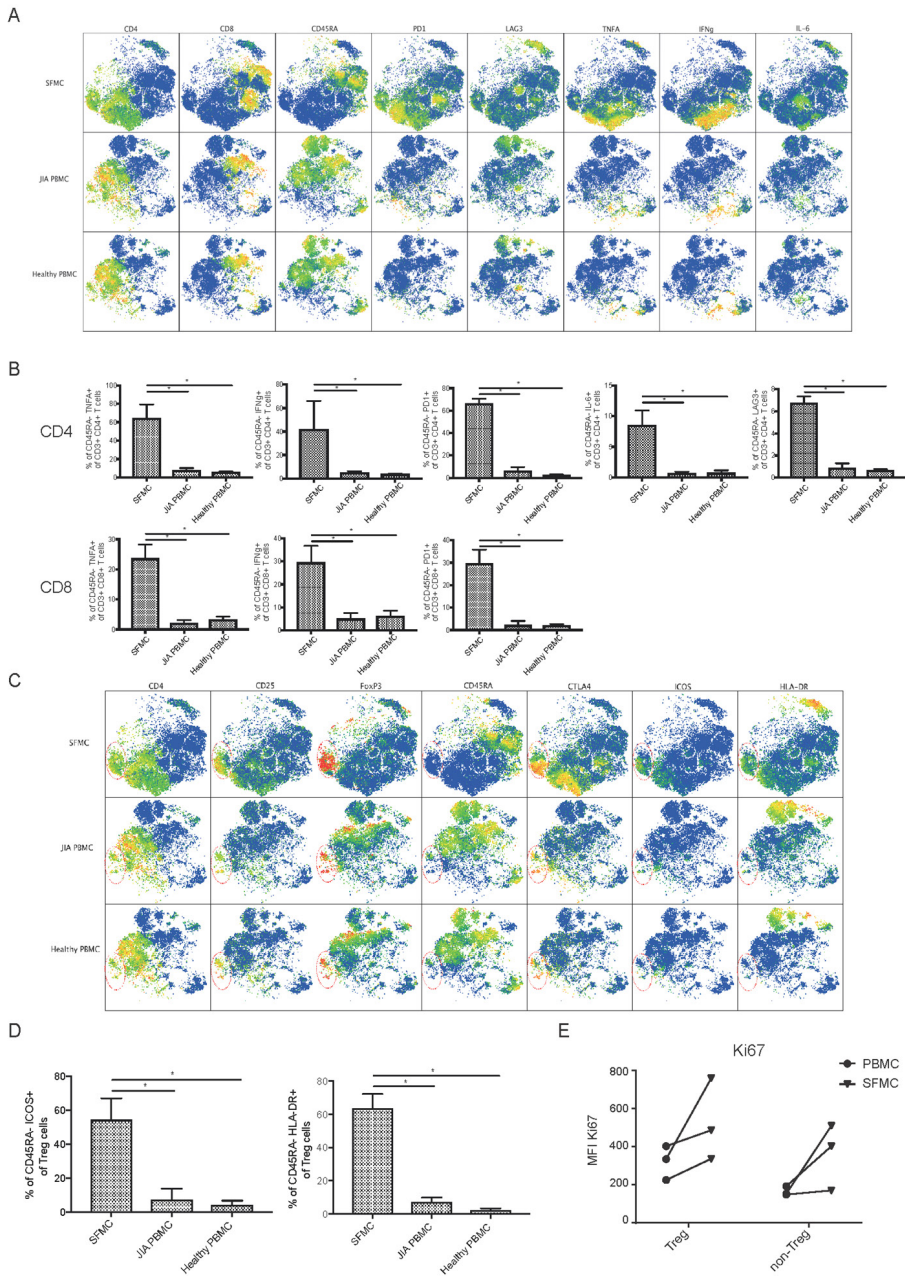
Supplemental Figure 2. SFMC display an activated expression profile. A T-SNE plots showing the expression profile of phenotypical and functional markers in SFMC, PBMC from JIA patients and PBMC from healthy children. B Bar charts showing the percentage of specific cell populations within CD4+CD45RA- and CD8+CD45RA- cells (non-parametric Mann-Whitney, * = p <0.05). C T-SNE plots showing the expression profile of phenotypical and functional Treg markers in SFMC, PBMC from JIA patients and PBMC from healthy children. D Quantification of CD45RA-ICOS+ and CD45RA-HLA-DR+ expression on CD25+ FOXP3+ Treg (non-parametric Mann-Whitney, * = p <0.05). E MFI of Ki67 protein expression in Treg and non-Treg as determined by flow cytometry.
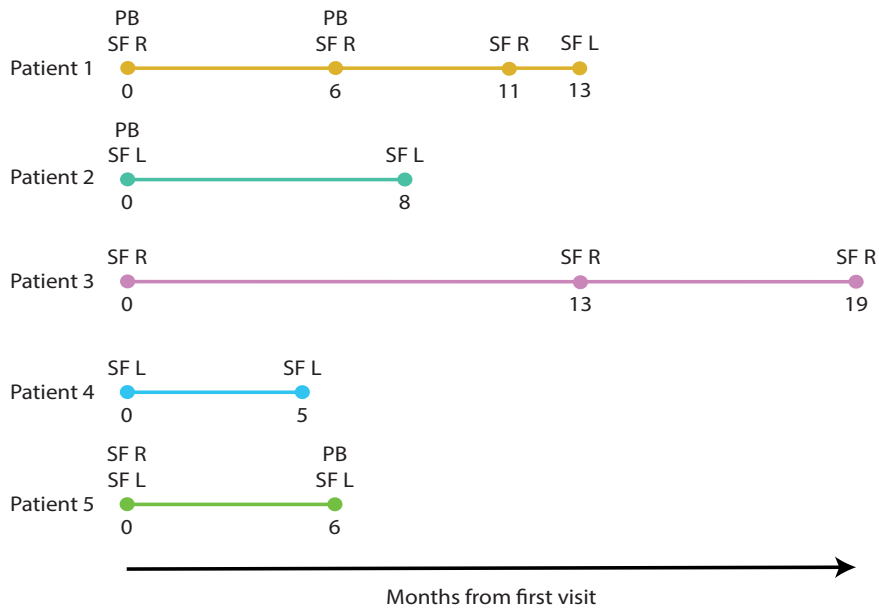
Supplemental Figure 3. Longitudinal sampling timelines of JIA patients. PB = peripheral blood, SF = synovial fluid, L = left, R = right.
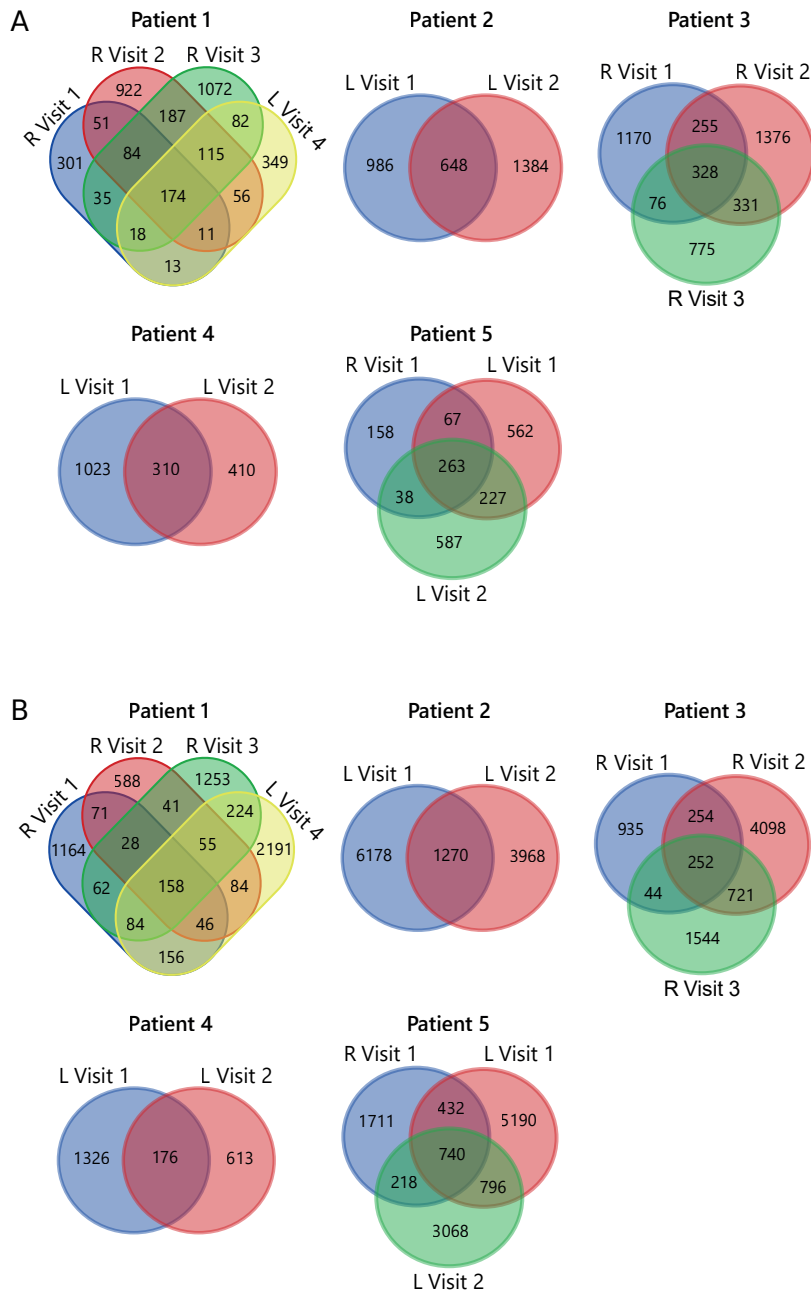
Supplemental Figure 4. TCR overlap analysis. A Venn diagrams displaying the overlap of all unique TCRβ clones, defined by amino acid sequence, for longitudinal SF samples from all patients for Tregs and B non-Tregs.
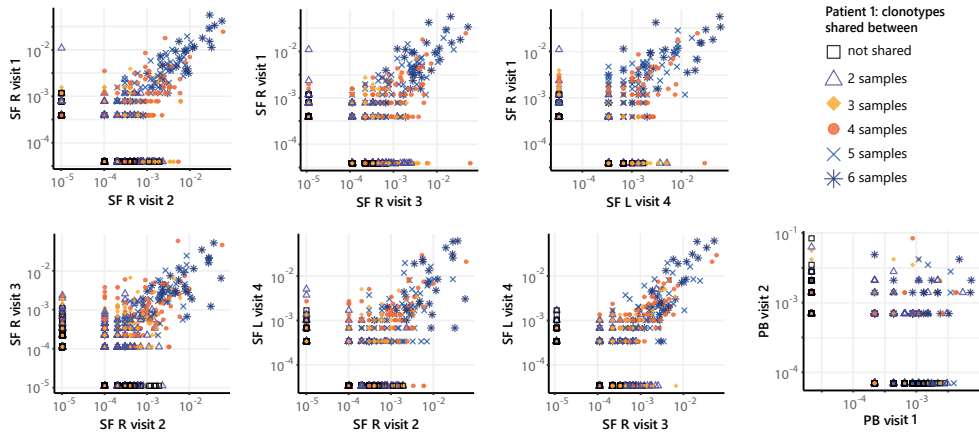
Supplemental Figure 5. Frequencies of TCRs from persistent Tregs shared across SF and PB samples. Frequency plots showing the overlapping Treg clones between visits for SF and PB, with color coding and shapes highlighting the number of samples in which unique clones are found.

Targets	Metal Channel	Clone	Antibody Vendor/Catalogue number
Lineage markers			
CD3	139	UCHT1	Biolegend (300402)
CD4	148	SK3	Biolegend (344625)
CD8	144	SK1	Biolegend (344727)
CD11b	161	ICRF44	Biolegend (301302)
CD16	209	3G8	Fluidigm (3209002B)
CD14	112/114	M5E2	Biolegend (301843)
T helper subsets			
IL-4	156	8D4-8	Biolegend (500707)
IFN-g	168	B27	Biolegend (506513)
IL-17A	169	BL168	Biolegend (512302)
IL-21	151	3A4-N2	Biolegend (513009)
CD161	157	HP-3G10	Biolegend (339902)
T cell functional markers			
CD45RA	171	HI100	Biolegend (304102)
CD69	176	FN50	Biolegend (310902)
CD28	146	CD28.2	Biolegend (302923)
CD152 (CTLA4)	155	BNI3	Biolegend (555851)
CD154 (CD40L)	149	24-31	Biolegend (310835)
HLA-DR	143	L243	Biolegend (307612)
LAG3	159	17B4	Abcam (ab40466)
PD1	147	EH12.2H7	Biolegend (329941)
Ki67	166	20Raj1	Thermofisher/ebioscience (14-5699-82)
ICOS	154	C398.4A	Biolegend (313512)
CD31	172	WM59	Biolegend (303102)
CD103	142	B-Ly7	Thermofisher/ebioscience (14-1038-82)
Chemokine receptors			
CXCR3	163	G025H7	Biolegend (353718)
CXCR5	160	RF8B2	BD biosciences (552032)
CCR5	145	NP-6G4	Abcam (ab115738)
CCR6	170	G034E3	Biolegend (353402)
Treg markers			
CD25	150	M-A251	BD biosciences (555429)
CD127	153	A019D5	Biolegend (351302)
FoxP3	165	PCH10L	Thermofisher/ebioscience (14-4776-82)
GITR	164	621	Biolegend (311602)
TGF-B (LAP)	175	TW4-2F8	Biolegend (349602)
IL-10	158	JES3-9D7	Biolegend (501402)
Cytokines/Enzymes			
TNF-alpha	152	Mab11	Biolegend (502902)
IL-6	162	MQ2-13A5	Thermofisher/ebioscience (16-7069-85)
Granzyme B	173	CLB-GB11	Abcam (ab103159)
Perforin	174	B-D48	Abcam (ab47225)
Barcodes			
CD45-A	89	HI30	Fluidigm (3089003B)
CD45-B,C or D	115, 141, 167	HI30	Biolegend (304002)
Live/Dead /Singlets			
DNA (Singlets)	191/193	Nil	Fluidigm Cell-ID Intercalator-Ir (201192
Cisplatin (Live/Dead)	195	Nil	Sigma-aldrich (479306-1G)

Supplemental Table 1. Overview of the T cell panel with 37 markers

7

Chapter 8

General Discussion

8

Summary of major findings

One of the current important outstanding questions in immunology is how the immune system can strike the right balance between activation and tolerance, providing protection against pathogens while also maintaining body homeostasis. How is this balance kept and what molecular mechanisms can cause the immune system to go awry, leading to autoimmune disease? To address these questions, high-throughput studies of different molecular layers from immune cells in health and disease are needed. Therefore, the work presented in this thesis aimed to explore the molecular mechanisms contributing to immune cell dysregulation in the autoimmune disease Systemic Sclerosis (SSc) using a multi-omics approach. The thesis describes the application of various omics techniques to study the epigenomic, transcriptomic and proteomic landscape of monocytes, DCs and T cells of SSc patients as well as individuals suffering from other rheumatic autoimmune diseases including systemic lupus erythematosus (SLE), rheumatoid arthritis (RA), and juvenile idiopathic arthritis (JIA). This chapter summarizes the major findings of the research described in this thesis, and discusses the clinical implications for SSc, as well as considerations for future studies.

Molecular alterations priming innate immune cells towards hyper-activation in SSc

Cells of the innate immune system, including monocytes and DCs, are crucial in sensing danger signals and mounting an effective immune response, but also play indispensable roles in the dampening and resolution of inflammation. Moreover, the formation of adaptive immune responses is largely instructed by co-stimulatory signals provided by these innate immune cells, making them a crucial component that partially determines the susceptibility to develop autoimmune disease. Given the evidence for the involvement of innate immune cells (including increased frequencies, activated profiles and infiltration into skin) in the earliest stages of SSc, even before the onset of fibrosis, it has been proposed that these cells are driving factors in disease pathogenesis[1–8]. However, this causal relationship remains to be established, as alterations in these cells might also be driven by chronic activation after autoimmunity has already been established. Nonetheless, it is likely that the dysregulation of innate immune cells in SSc is likely a result of altered signaling pathways downstream of a triggered danger-sensing receptor. Many regulators are potentially involved in keeping these signaling pathways from derailing. These include the recently described class of long non-coding RNAs (lncRNAs), as well as histone modifications and other complex networks of immune activating and immune regulatory transcription factors, which were further investigated in this thesis.

lncRNAs as novel regulators of immune tolerance in SSc

Research from our group and others previously showed that monocytes obtained from SSc patients secrete aberrantly high levels of pro-inflammatory and pro-fibrotic cytokines upon TLR stimulation as compared to the same cells obtained from healthy individuals[9–11]. The data presented in **chapter 2** and **chapter 3** of this thesis show that lncRNAs are important molecules involved in the regulation of these TLR mediated cytokine signaling pathways. We identified the lncRNA NRIR (Negative

Regulator of the IFN Response) as regulator of IFN responses downstream of TLR4, and the lncRNA PSMB8-AS1 as a regulator of cytokine secretion downstream of TLR7/8 in monocytes. Both NRIR and PSMB8-AS1 were reproducibly upregulated in monocytes of SSc patients compared to healthy monocytes, highlighting lncRNAs as novel molecular factors contributing to monocyte dysregulation in SSc. These results show that lncRNAs are actively involved in the establishment of peripheral tolerance and have a critical role in priming innate immune cells towards a hyper-activated state in SSc. Indeed, recent studies have already revealed that lncRNAs act as critical regulators of immune cell function and are involved in the maintenance of immune tolerance. As an example, the lncRNAs MALAT1 and NEAT1 are described to be involved in the induction of tolerogenic DCs[12, 13], as well as the skewing of monocytes towards distinct pro-inflammatory states[14, 15]. These studies further underline the potential for lncRNAs in the priming and activation of immune cells, and highlight their roles as novel tolerance regulators.

Histone modifications: rewiring the immune system to a pro-inflammatory state

For a long time, the distinct transcriptional programs involved in cellular activation and tolerance were thought to be hardwired during differentiation. However, it is becoming more and more clear that differentiated immune cells, including monocytes, can be rewired through the loss or acquirement of specific histone modifications following exposure to pathogen-associated molecular patterns (PAMPs) and damage-associated molecular patterns (DAMPs)[16]. What makes these epigenomic changes especially interesting is that, as opposed to other cell signaling transducers such as phosphorylation, they can persist long after the initial stimulus is eliminated, and can even be stably transmitted throughout cell divisions[17]. This phenomenon is often referred to as 'trained immunity', and normally confers resistance to secondary infections. In **chapter 4** of this thesis we observed an increased deposition of activating histone marks at the promoters of genes relevant for disease specific pro-inflammatory pathways in monocytes of SSc, SLE and RA patients, demonstrating that epigenomic modifications contribute to the rewiring of monocytes in these patients. This rewiring, or epigenomically mediated trained immunity, might be induced by DAMPs as a result of tissue damage early in disease pathogenesis. Indeed, various DAMPs have already been associated with the induction of trained immunity in immune related disorders, including oxidized low-density lipoprotein (oxLDL) in atherosclerosis[18] and uric acid in gout[19]. In SSc, it would be interesting to study to what extent DAMPs released upon vascular damage (one of the earliest events in SSc pathogenesis[20]) are capable of inducing epigenomically mediated trained immunity. More knowledge on the exact (early) signals inducing epigenomic rewiring in immune cells could help to better understand and potentially reverse this process.

Immune dysregulation in SSc is not merely driven by enhanced activation, but also a loss of negative feedback

To limit inflammation and maintain immune tolerance, endogenous negative feedback mechanisms are in place. Characterization of these feedback mechanisms can help to identify new targets to prevent or reverse excessive inflammation. Therefore, in **chapter 5**, using transcriptomic data of cDCs from healthy donors and SSc, we applied

8

a co-expression network approach as well as transcription factor ChIP-sequencing to identify biologically relevant transcriptional regulators contributing to cDC dysregulation in SSc. Following this approach, we identified the NR4A (nuclear receptor 4A) subfamily (NR4A1, NR4A2, NR4A3), to be strongly involved in transcriptional programs underlying cDC dysregulation in SSc. These NR4A receptors are induced by pro-inflammatory stimuli and are involved in negative feedback mechanisms to dampen immune responses via transcriptional regulation of various genes. Indeed, our functional experiments using agonists targeting NR4As showed that they are involved in cytokine production by and modulation of T cells activation by cDCs. These data implicate NR4As as important negative regulators of immune pathways in cDCs, and NR4A downregulation potentially contributes to the dysregulation of cDCs in SSc patients. This demonstrates that immune cell dysregulation in SSc is not merely a consequence of enhanced activation but can also be attributed to a loss of negative feedback.

A loss of innate immune tolerance may lead to T cell receptor repertoire skewing in SSc

In the first part of this thesis, we show that a variety of molecular mechanisms may induce a loss of tolerance in the innate immune system in SSc and rewire immune cells to a hyper-activated state. The resulting inflammatory milieu may provide a perfect niche for the aberrant activation expansion of antigen specific T cell clones that perpetuate tissue damage and inflammation, further contributing to the pathogenesis of SSc. To further investigate this, in the second part of this thesis, the dynamics of the T cell repertoire in SSc were studied to better understand the role of antigen specific T cell responses in the disease pathogenesis. To this end, in **chapter 6**, we performed high-throughput sequencing of T cell receptors (TCRs) of circulating CD4+ and CD8+ T cells from longitudinal samples obtained from SSc patients. Here we show that the TCR repertoire in SSc is highly stable over time, and this persistence is likely a result of antigenic selection rather than bystander activation. These observations indeed suggest that aberrances in the innate immune compartment facilitate the generation of a highly oligo clonal T cell repertoire in SSc. In line with this, we show that SSc TCR repertoires are less diverse than T cell memory repertoires from healthy individuals, demonstrating that T cells are highly clonally expanded in SSc, potentially due to chronic antigen activation. To determine to what extent the longitudinal persistence of circulating CD4+ and CD8+ T cells are characteristic of SSc, in **chapter 7** we studied the immune cell architecture and TCR repertoire dynamics of peripheral blood and affected joints of JIA patients. Because unlike SSc patients, which are characterized by systemic inflammation, the JIA patients included in this study suffered from localized inflammation in the knee joints, we were provided with the unique opportunity to compare the T cell landscape of systemic versus localized autoimmune disease. Whereas in SSc patients, circulating CD4+ and CD8+ T cells were highly clonally expanded (as shown in **chapter 6**), only T cells obtained from affected joints from JIA patients exhibited an expanded profile while circulating T cells did not. These results indicate that tissue specific dominant (auto-)antigens in JIA patients heavily skew the TCR repertoire, while in SSc, the potential antigens might be more ubiquitously expressed. However, it should be noted that in JIA patients, the T cell

clones that were found with the highest frequency in affected joints could also be traced back in circulation, and the same T cell clones were present in distinctly affected joints within one patient, strongly suggesting that reactive clones are recirculating. Additionally, JIA patients were characterized by a strong expansion and persistence of regulatory T cells (Tregs) rather than effector T cells, whereas in SSc, effector CD4+ and CD8+ T cells were highly expanded over time. Thus, T cell activation and expansion in JIA might be a result of a failure to suppress autoreactive T cells by Tregs, while in SSc an aberrant activation of effector T cells by innate immune cells is more likely to drive T cell hyper expansion. However, more detailed investigations of the functional profiles, origins of expanded T cell clones, for example through lineage tracking and single cell sequencing, are needed to substantiate this hypothesis.

Novel approaches to target immune cell activation in SSc

Considering the importance of innate immune cell activation in SSc pathogenesis[21, 22], modulation or inhibition of these cells represents a potential therapeutic option to restore immune homeostasis in SSc. However, broad suppression, for example through targeting TLRs, may lead to increased infectious disease. Thus, the major challenge here is to dampen innate immune signaling pathways to ameliorate the immune response just enough, without completely shutting it down. Therefore, the modulation of downstream targets regulators that tweak signaling pathways instead of abolishing them, represent promising novel therapeutic targets. The studies represented in this thesis highlight an important role for novel regulators driving immune cell hyper-activation in SSc. These include lncRNAs (**chapter 2 and 3**), histone modifications (**chapter 4**) and immune regulatory transcription factors (**chapter 5**). Importantly, rather than representing binary "on-off" switches, these regulators have subtle immune modulatory effects. This makes them potentially relevant clinical targets to help reset innate immune cells in SSc towards an immunotolerant state.

lncRNAs: a novel therapeutic option for SSc?

Given the immune-regulatory potential of lncRNAs, lncRNA-orientated next-generation drugs might represent a novel therapeutic avenue for SSc. Indeed, their ability to fine-tune the expression of immune-related genes, including cytokines and inflammation-related transcription factors as shown in **chapter 2 and 3**, has potential therapeutic implications. The use of antisense oligonucleotide drugs that target ncRNAs in other diseases have already been approved by the FDA, and there are various ongoing studies for the development of ncRNA-based therapeutics, including testing in animal models and clinical trials[23]. An advantage of using such synthetic oligonucleotides targeting lncRNAs is that they can very specifically modulate single lncRNA activity as a result of their precise sequence complementarity.

However, lncRNAs are still new kids on the block when it comes to the regulation of immune responses, and a lot is still to be discovered about this intricate class of molecules. Moreover, lncRNA are known to have cell type specific functions, so overall targeting might not be desirable. As an example, in **chapter 2**, we identified the lncRNA NRIR as a positive regulator of IFN responses in monocytes, while NRIR has also been

8

described as a negative regulator of IFN responses in hepatocytes[24]. On the other hand, this cell-type specificity could also be exploited to modulate lncRNA activity in a very targeted way to avoid unwanted side effects. More detailed studies of lncRNAs and their effects in different cell types/phenotypes are required to better understand their functions. Single-cell sequencing studies might help to gain more insights into the specific expression of lncRNAs in various cell types and tissues. These could also be combined with knock-out studies to monitor the effect of lncRNA modulation of specific cell types within one system. Next to experimental approaches, future studies can be performed using computational biology based approaches including lncRNA function prediction based on primary sequence, secondary structure conservation or prediction of binding partners[25]. Lastly, RNA hybridization-based approaches can be performed to precisely dissect the biological targets and molecular mechanisms through which lncRNAs exert their functions[26]. Such studies would provide critical insights into the (cell type) specific functions of lncRNAs, and further aid their translation into a clinical setting.

Targeting histone modifying enzymes to rewrite the epigenomic code

Histone modifying enzymes are critical regulatory proteins that can bind specific sites marked by histone acetylation or methylation. They act as writers or erasers to increase or decrease the deposition of histone modifications, thereby affecting downstream gene expression[27]. Targeting these histone modifying enzymes can help to rewrite the epigenomic code and potentially reverse epigenomically mediated trained immunity characterizing innate immune cells in SSc. An example of a histone modifying enzyme that could hold therapeutic value in SSc is the histone demethylase Jumonji-C domain 3 (JMJD3), that specifically removes the inhibiting histone mark H3K27me3[28], [29]. Interestingly, a role for JMJD3 in fibroblast activation via the removal of H3K27me3 at fibrosis related genes has already been described, and pharmacological inhibition of this demethylase ameliorates fibrosis in mouse models of SSc[29]. Moreover, JMJD3 has an important role in monocyte to macrophage differentiation, where it regulates the demethylation of H3K27me3 at the promoters of genes important for M2 polarization, including IRF4[30]. Next to JMJD3, the methyl transferase enhancer of zeste homologue 2 (EZH2), which trimethylates H3K27, resulting in transcriptional repression, has been shown to be involved in fibroblast activation[31] and metalloprotease activity in monocytes from SSc patients[32]. Given their roles in fibrosis, monocytes, and macrophage polarization, it would be highly interesting to further delineate the effects of JMJD3 or EZH2 modulation in circulating monocytes, especially in the control of bivalent genes. Since in **chapter 4** we also identified many bivalent promoters primed for high activation in monocytes from early SSc patients, it would be highly interesting to study the effect of JMJD3 or EZH2 modulation in this patient group to determine whether epigenomic targeting can inhibit or delay the onset of fibrosis.

Interestingly, histone modifying enzymes have been shown to actively remodel transcription factor networks, and are known to regulate NR4A expression[33, 34], which we found to be downregulated in cDCs from SSc patients in **chapter 5**. This suggests that the downregulation of NR4As cDCs in SSc might be a consequence of an altered epigenomic landscape of these cells, again highlighting a role for epigenomic remodeling

and trained immunity in the dysregulation of immune cells in SSc. Thus rewiring the epigenomic landscape in SSc by targeting histone modifying enzymes may also help to modulate transcriptional networks implicated in SSc pathogenesis. To further investigate this, detailed characterizations of the epigenomic landscape of cDCs from SSc patients are needed. However, it should be noted that cDCs are a very rare cell population, and genome-wide ChIP-sequencing of these cells might prove difficult due to limited amounts of material, especially when studying patient samples. Moreover, it should be noted that the epigenetic basis of inflammatory responses is extremely complex, involving the interplay of multiple histone modifications, DNA methylation, regulation by long non-coding RNAs, miRNAs and many more. We are just beginning to scratch the surface of this intriguingly complex regulatory network and much still remains to be elucidated. The application and integration of multiple omics technologies to further uncover the dynamics of inflammatory responses is crucial in this. Since epigenetic modifications are reversible, a better understanding of the epigenomic dynamics of inflammation in health and autoimmune disease should aid in the discovery of new therapeutic targets with the ultimate aim to restore immune tolerance.

Targeting adaptive immunity through generation of tolerogenic DCs

As highlighted various times in this thesis, innate control of adaptive immunity is an important paradigm, and blocking interactions between the innate and adaptive immune system might ameliorate downstream adaptive immune responses in SSc patients. Indeed, in **chapter 5** of this thesis we show that the activation of anti-inflammatory transcription factors of the NR4A family in cDCs leads to a decrease in the downstream activation of T cells. Thus, the generation of cDCs with a tolerogenic phenotype might hold therapeutic value to halt aberrant T cell activation in SSc. Various molecules such as anti-inflammatory cytokines (e.g. IL-10), vitamin D3, rapamycin, glucocorticoids and many more possess tolerogenic properties that induce the generation of tolerogenic DCs. These tolerogenic DCs may either be generated *in vitro*[35] followed by infusion, or *in vivo*, through the administration of tolerogenic immunotherapy. The *in vitro* generation and infusion of tolerogenic DCs has already been explored in various autoimmune diseases, such as RA[36], multiple sclerosis (MS)[37], and type I diabetes (T1D)[38]. In these patients, infusion of autologous tolerogenic dendritic cells was shown to be safe and well tolerated, highlighting the clinical potential for these approaches. Currently, two clinical trials with tolerogenic DC-based vaccines are ongoing for MS (NCT02903537 and NCT02618902), and one for T1D (NCT04590872). Results from these studies should provide further insights into the potential for harnessing DCs in the treatment of autoimmune disease.

Besides *ex vivo* generation and infusion of tolerogenic DCs, *in vivo* induction of tolerogenic DCs also forms an attractive therapeutic option. This can be achieved through the administration of biologicals or hormones with anti-inflammatory characteristics. One such molecule which has recently gained a lot of attention for its ability to induce a tolerogenic phenotype in DCs is vitamin D3[39]. Interestingly, vitamin D3 levels have been shown to be lower in SSc patients as compared to healthy individuals in various studies[40–42], leading to the hypothesis that reduced vitamin D3 is associated with a

8

loss of immune tolerance in SSc patients. However, reports on the association of vitamin D3 levels and clinical severity of SSc are conflicting, and the exact relation between low vitamin D3 and SSc remains unclear. Thus, there is a need for the exploration of other therapeutic avenues to induce tolerogenic DCs in SSc. In this context, it would be interesting to explore NR4As as novel targets to induce tolerance in DCs from SSc patients. In **chapter 5**, we show that NR4A activity can be modulated using agonists, providing initial proof for the potential of NR4A targeting to induce tolerance. Other pharmaceutical compounds activating NR4As, including synthetic bisindole-derived compounds (C-DIMs), cytosporone B (Csn-B) and mercaptopurine (6-MP), could also hold therapeutic potential in SSc. Interestingly, Csn-B has already been shown to ameliorate collagen deposition and myofibroblast differentiation in mouse models of fibrosis[43], highlighting the therapeutic potential of NR4A targeting and providing a new mechanism to exploit the induction of tolerogenic DCs in SSc.

Stem cell transplantation in SSc: is the innate immune system crucial for a reset to tolerance?

Autologous hematopoietic stem-cell transplantation (AHSCT) currently is the only therapy with long-term clinical benefit in rapidly progressive SSc. The main rationale for applying AHSCT to SSc is to restore immune homeostasis. This is achieved by first applying an intensive immunoablative conditioning regimen, which eliminates the pathogenic self-reactive immune cells, followed by the reconstitution of a new immune system from reinfused hematopoietic precursors. The success of AHSCT is proposed to be largely dependent on the eradication of clonally expanded auto-reactive T cells. Indeed, after AHSCT, TCR diversities increase significantly, reflecting the reconstitution of a new, more tolerant immune system[44–46]. Notably, differences in the clonality of the TCR repertoire have also been observed between responders and non-responders to AHSCT, with non-responders having a less diverse repertoire[44, 45]. These data further highlight that decreased TCR repertoire diversity contributes to the autoimmune pathogenesis of SSc, as also shown in **chapter 6** of this thesis.

Both adaptive and innate immune responses are modulated after AHSCT in SSc and contribute to the generation of a tolerant immune system. Although the adaptive immune system has been studied in more detail, the innate immune system also has a potential immunosuppressive/modulating role in regulating adaptive responses after AHSCT. In support of this theory, CD14+ monocytes have already been shown to be capable of regulating T cell responses following AHSCT[47]. One important question that remains unanswered is whether the innate immune system, which reconstitutes much faster than the adaptive immune system[48], creates a permissive environment for the regeneration of a tolerant adaptive immune system after AHSCT. In other words, is the reconstitution of innate immune cells such as monocytes, DCs after AHSCT a prerequisite for the formation of a more diverse TCR repertoire? In order to answer this fundamental question, innate immune cells should be studied in more detail, especially at early time-points after AHSCT. Furthermore, it would be interesting to investigate if innate immune reconstitution and priming of adaptive responses underlies the efficacy of AHSCT and whether this is different in responders versus non-responders. To further

investigate to what extent the innate immune system facilitates the expansion of antigen specific T cells, it would be interesting to study the TCR repertoire over the course of immune system reconstitution after AHSCT. Moreover, it would be interesting to perform TCR profiling of T cells co-cultured with monocytes or DCs from SSc patients, before and after AHSCT, to see whether these cells indeed display a high potency to induce oligo clonal T cell expansion and how this is affected by AHSCT.

Thus, further investigation into the exact role of innate immune cells and their importance in regulating immune responses after AHSCT is needed. Better insights into these responses may help to further improve patient care and predict transplant outcomes more accurately, as well as helping to gain a better understanding of SSc pathogenesis.

Directly targeting the adaptive immune system through T cell therapy

Given the evidence for T cells in the pathogenesis of SSc, and the fact that the AHSCT is proposed to be largely dependent on the eradication of auto-reactive T cells, directly targeting T cells might also hold therapeutic value for SSc. Depletion of T cells using alemtuzumab (an antibody targeting CD52, a molecule is highly expressed on the surface of B and T cells) has previously been explored in a case report for a single dcSSc patient, where treatment with alemtuzumab led to a rapid and sustained improvement of skin score[49]. However, although treatment with this antibody is approved for patients with relapsing forms of MS, side effects related to the development of secondary autoimmunity have been observed[50], putting alemtuzumab in a not very desirable position. Moreover, T cell depleting therapies like alemtuzumab are in the gray zone between immunosuppression and immunoablation, with a high potential to drive the immune system in an undesirable lymphopenic state. Thus, targeting specific autoreactive T cell clones might hold better therapeutic potential. Since in **chapter 6** of this thesis we show that specific T cell clones are highly expanded in SSc patients over time, specific targeting of these T cells seems plausible. As an example, adoptive immunotherapy to transfer genetically engineered-Tregs expressing TCRs against the same antigens as autoreactive CD4+ and/or CD8+ T cells could potentially suppress T cell mediated autoimmune responses[51]. However, since the (auto)antigens targeted by T cells in SSc remain yet to be elucidated, engineering antigen specific Tregs is not possible yet. On the other hand, clonally expanded T cells from SSc patients could be converted into Tregs through upregulation of the Foxp3 gene, thereby potentially retaining their antigen specificity while gaining a suppressive phenotype[52]. Another approach of targeting autoreactive T cells is through T cell vaccination (TCV). The concept of TCV is based on the finding that inactivated autoreactive T cells can induce inhibition of T cell dependent autoimmune responses[53]. Here, the target antigen against which immunity is induced is (parts of) the TCR of autoreactive T cells, thereby eliminating them. However, the challenge here is to elucidate exactly which TCRs are involved in the response to disease related (auto)antigens.

Our analyses in **chapter 6** as well as **chapter 7** show that expanded T cell clones in SSc as well as JIA are characterized by a high sequence similarity, indicating that pools of highly expanded T cell clones respond to the same (auto)antigens. This

8

knowledge can be used to obtain motifs in antigen specific TCRs that are shared across patients and design broad pools of TCVs that can potentially ameliorate autoreactive T cell responses. In order to achieve this, studies of large patients cohorts with high-throughput deep sequencing of a large number of T cells are necessary. The design of novel machine learning methods to unravel motifs shared across autoreactive T cell clones should also aid in the discovery and classification of autoimmune related TCRs. Such efforts are currently underway, by our group as well as others[54, 55], and promise to bring exciting new developments to the field of TCR research.

Challenges and considerations for future studies

The studies presented in this thesis provide new insights into the molecular mechanisms underlying aberrant immune cell activation in SSc. From these results, it is clear that immune cell dysregulation in SSc pathogenesis is mediated through the derailment of various factors at different levels of gene expression regulation, including previously unannotated lncRNAs, histone modifications and transcription factor regulatory networks. However, exactly how these separate factors interplay with each other and to which extent their regulatory modes of action are cell type specific remains to be uncovered. Here, potential future studies to answer these questions are proposed.

Multi-omics integration to generate extensive immune regulatory networks

To obtain a more complete picture of the full spectrum of factors contributing to immune cell activation, and loss of immune tolerance, future studies will require the integration of individual omics datasets to get a more complete understanding of the immunopathology of SSc. In other words: a shift from single-omics to integrated multi-omics is required. Such multi-omics studies should include the information on DNA sequence variation, open chromatin regions, histone modifications and DNA methylation, transcription factor binding and activity, RNA expression variation, and post-transcriptional and translational regulation. Information from these different omics layers can be used to build extensive gene regulatory networks (GRNs)[56]. GRNs provide a map for the molecular processes and interactions at different levels of cellular regulation, and have the potential to predict the outcomes of immune cell perturbation (for example as a result of genomic alterations or stimulation). Therefore, future studies applying multi-omics approaches paired with computational modelling and GRN construction should provide better insights into the factors that skew the immune system in SSc towards an hyper-activated status. Such studies should especially focus on the epigenomic landscape, novel regulators such as lncRNAs, and transcription factor feedback loops since these were demonstrated to be important factors contributing to loss of immune tolerance in SSc in this thesis.

One way to perform high throughput analyses in a multi-omics approach is to simultaneously apply multiple omics techniques to measure a single biological sample. In this way, an unbiased view of the complex relationships between genomics, different epigenomic markers and their effect on gene-expression measured by transcriptomics can be obtained. The usefulness of this approaches has recently been highlighted by Ai *et al.*, who combined ATAC-seq, ChIP-seq of six different histone marks, RNA-seq, and WGBS

for DNA methylation, to comprehensively study the epigenomic landscape of rheumatoid arthritis fibroblast-like synoviocytes (FLS)[57]. Because the authors used FLS from RA patients and healthy donors that were grown out for multiple passages, enough material could be obtained to perform these multi-omics analyses from material obtained from a single sample. In the case of more rare immune cell populations, the limited amounts of primary biological material that can be obtained will not suffice. However, multiple analyses combining two or three omics approaches for primary immune cells have been performed already[58–60] (also in **chapter 4** and **chapter 5** of this thesis) and these provide valuable insights into how multiple omics layers interplay and affect each other. Further such studies should provide more data for the construction of immune regulatory GRNs and help to get better insights into signaling pathways leading to immune cell dysregulation in SSc.

Single cells approaches to delineate cell type specific dysregulation in SSc

Using traditional, bulk omics approaches, the contribution of heterogeneous, rare cell populations are often not identified because their signal gets drowned out by the presence of larger cell pools. To address this, new single-cell technologies have recently been developed that can assess the genome, epigenome, transcriptome, proteome and metabolome at a single-cell resolution[61]. These single-cell omics can help to pinpoint specific immune cell subsets that contribute to disease pathogenesis in more detail. Moreover, in addition to obtaining a better understanding of the contribution of various cell populations to SSc pathogenesis, single-cell profiling can also help to identify specific drug targets uniquely expressed in specific immune cell populations. Moreover, single cell approaches can also help to overcome the limited amount of material that is available for multi-omics studies, as single cell multi-omics approaches only require a small amount of cells as compared to bulk approaches. A few single cell studies have recently been performed using material from SSc patients, which have helped to uncover pathogenesis associated signatures in distinct cell populations. These include the activation of macrophages and pDCs in skin[62], as well as an enrichment of SSc-associated single-nucleotide polymorphisms and open chromatin regions in DCs in skin lesions[63]. However, thus far, these single cell studies in SSc mainly focus on the analysis of single layers of omics data, thereby not capturing the full spectrum of data needed to construct cell type specific GRNs. Thus, future single-cell studies should also focus on the integration of multiple omics techniques.

Sample size and reproducibility

As SSc is a relatively rare disease, current studies into SSc pathogenesis (including the ones presented in this thesis) are often limited by small sample sizes. Given the heterogeneity characterizing SSc patients, small sample sizes make it challenging to find coherent signatures within or between different cohorts of SSc patients. Ideally, for future studies, larger sample sizes are needed to find distinct differences in immune cell signatures between SSc patients and healthy donors, as well as between different subtypes of SSc. As this might not always be feasible due to limited sample availability, the development of collective standardized protocols and computational tools to allow

8

for the integration and comparison of multiple diverse datasets from different research groups could offer a potential solution. To this end, thorough benchmarking needs to be performed to evaluate the performance of diverse experimental and computational tools. In the field of cancer research, gold standard multi-omics datasets such as The Cancer Genome Atlas (TCGA)[64] already exist. Comparable initiatives in the field of SSc and other (rheumatic) autoimmune diseases could help to standardize the way of performing omics studies within this field, and provide a solid base for omics datasets to build upon and complement each other.

Conclusions

Using different omics techniques, the work presented in this thesis shows that immune dysregulation in SSc can be attributed to aberrances at various levels of molecular organization. These include the regulation of TLR signaling by lncRNAs, epigenomic imprinting of histone modifications and downregulation of immune regulatory transcription factors in monocytes and DCs. Enhanced activation of these innate immune cells has the potential to cue the adaptive immune system and orchestrate the generation of highly clonal autoreactive T cell repertoire. Future studies using standardized multi-omics and single cell approaches should help to further unravel the regulatory pathways contributing to immune cell dysregulation in SSc.

REFERENCES

1. N. Higashi-Kuwata, M. Jinnin, T. Makino, *et al*. Characterization of monocyte/macrophage subsets in the skin and peripheral blood derived from patients with systemic sclerosis. *Arthritis Res. Ther.* vol. 12, no. 4, pp. R128. Jul. 2010.
2. N. Higashi-Kuwata, T. Makino, Y. Inoue, *et al.* Alternatively activated macrophages (M2 macrophages) in the skin of patient with localized scleroderma. *Exp. Dermatol*. vol. 18, no. 8, pp. 727–729. 2009.
3. M. Kroef, L.L. van den Hoogen, J.S. Mertens, *et al.* Cytometry by time of flight identifies distinct signatures in patients with systemic sclerosis, systemic lupus erythematosus and Sjögrens syndrome. *Eur. J. Immunol.* vol. 50, no. 1, pp. 119–129. Jan. 2020.
4. O. Ishikawa, H. Ishikawa. Macrophage infiltration in the skin of patients with systemic sclerosis. *J. Rheumatol.* vol. 19, no. 8, pp. 1202–6. Aug. 1992.
5. B.M. Kräling, G.G. Maul, S. A. Jimenez. Mononuclear cellular infiltrates in clinically involved skin from patients with systemic sclerosis of recent onset predominantly consist of monocytes/macrophages. *Pathobiology*. vol. 63, no. 1, pp. 48–56. 1995.
6. S. Mokuda, T. Miyazaki, Y. Ubara, *et al*. CD1a+ survivin+ dendritic cell infiltration in dermal lesions of systemic sclerosis. *Arthritis Res. Ther.* vol. 17, no. 1, pp. 275. Dec. 2015.
7. A.L. Mathes, R.B. Christmann, G. Stifano *et al*. Global chemokine expression in systemic sclerosis (SSc): CCL19 expression correlates with vascular inflammation in SSc skin. *Ann. Rheum. Dis*. vol. 73, no. 10, pp. 1864–1872. Oct. 2014.
8. L. van Bon, C. Popa, R. Huijbens, *et al*. Distinct evolution of TLR-mediated dendritic cell cytokine secretion in patients with limited and diffuse cutaneous systemic sclerosis. *Ann. Rheum. Dis*. vol. 69, no. 8, pp. 1539–1547. Aug. 2010.
9. S.K. Mathai, M. Gulati, X. Peng, *et al.* Circulating monocytes from systemic sclerosis patients with interstitial lung disease show an enhanced profibrotic phenotype. *Lab. Investig*. vol. 90, no. 6, pp. 812–823. Jun. 2010.
10. T. Carvalheiro, S. Horta, J.A.G. van Roon, *et al*. Increased frequencies of circulating CXCL10-, CXCL8- and CCL4-producing monocytes and Siglec-3-expressing myeloid dendritic cells in systemic sclerosis patients. *Inflamm. Res*. vol. 67, no. 2, pp. 169–177. Feb. 2018.
11. T. Carvalheiro, A.P. Lopes, M. van der Kroef, *et al*. Angiopoietin-2 Promotes Inflammatory Activation in Monocytes of Systemic Sclerosis Patients. *Int. J. Mol. Sci*. vol. 21, no. 24, p. 9544. Dec. 2020.
12. J. Wu, H. Zhang, Y. Zheng, *et al*. The long noncoding RNA MALAT1 induces tolerogenic dendritic cells and regulatory T cells via miR155/dendritic cell-specific intercellular adhesion molecule-3 grabbing nonintegrin/IL10 axis. *Front. Immunol*. vol. 9, pp. 1847, Aug. 2018.
13. M. Zhang, Y. Zheng, Y. Sun, *et al*. Knockdown of NEAT1 induces tolerogenic phenotype in dendritic cells by inhibiting activation of NLRP3 inflammasome. *Theranostics*. 2019. vol. 9, no. 12, pp. 3425-3442. May 2019.
14. F. Zhang, L. Wu, J. Qian, *et al*. Identification of the long noncoding RNA NEAT1 as a novel inflammatory regulator acting through MAPK pathway in human lupus. *J. Autoimmun*. vol. 75, pp. 96–104. Dec. 2016.
15. H. Cui, S. Banerjee, S. Guo, *et al*. Long noncoding RNA Malat1 regulates differential activation of macrophages and response to lung injury. *JCI insight*. vol. 4, no. 4, pp. e124522, Feb. 2019.
16. S. Mehta, K.L. Jeffrey. Beyond receptors and signaling: Epigenetic factors in the regulation of innate immunity. *Immunol Cell Biol*. vol. 93, no. 3, pp. 233–44. Mar. 2015.
17. P. Byvoet, G.R. Shepherd, J.M. Hardin, *et al*. The distribution and turnover of labeled methyl groups in histone fractions of cultured mammalian cells. *Arch. Biochem. Biophys.* vol. 148, no. 2, pp. 558–67. Feb. 1972.
18. S. Bekkering, J. Quintin, L. A. B. Joosten, *et al*. Oxidized low-density lipoprotein induces long-term proinflammatory cytokine production and foam cell formation via epigenetic reprogramming of monocytes. *Arterioscler. Thromb. Vasc. Biol*. vol. 34, no. 8, pp. 1731–1738. Jun. 2014.
19. T.O. Crişan, M.C.P. Cleophas, B. Novakovic, *et al*. Uric acid priming in human monocytes is driven by the AKT-PRAS40 autophagy pathway. *Proc. Natl. Acad. Sci. U. S. A.* vol. 114, no. 21, pp. 5485–5490. May 2017.
20. P.M. Campbell, E. C. LeRoy. Pathogenesis of systemic sclerosis: A vascular hypothesis. *Semin. Arthritis Rheum*. vol. 4, no. 4, pp. 351–368. May 1975.
21. L. Frasca, R. Lande. Toll-like receptors in mediating pathogenesis in systemic sclerosis. *Clin. Exp. Immunol*. vol. 201, no. 1, pp. 14–24. Jul. 2020.

8

22. P. Laurent, V. Sisirak, E. Lazaro, *et al.* Innate Immunity in Systemic Sclerosis Fibrosis: Recent Advances. *Front. Immunol.* vol. 9, pp. 1702. Jul. 2018.

23. C.K. Huang, S. Kafert-Kasting, T. Thum. Preclinical and Clinical Development of Noncoding RNA Therapeutics for Cardiovascular Disease. *Circulation Research.* vol. 126, no. 5, pp. 663-678. Feb. 2020.

24. H. Kambara, F. Niazi, L. Kostadinova, *et al.* Negative regulation of the interferon response by an interferon-induced long non-coding RNA. *Nucleic Acids Res.* vol. 42, no. 16, pp. 10668–10680. Aug. 2014.

25. J. Li, X. Zhang, C. Liu. The computational approaches of lncRNA identification based on coding potential: Status quo and challenges. *Comput. Struct. Biotechnol. J.* vol. 18, pp. 3666–3677. Nov. 2020.

26. M. Cao, J. Zhao, G. Hu. Genome-wide methods for investigating long noncoding RNAs. *Biomed. Pharmacother.* vol. 111, pp. 395–401. Mar. 2019.

27. M.A.J. Morgan, A. Shilatifard. Reevaluating the roles of histone-modifying enzymes and their associated chromatin modifications in transcriptional regulation. *Nat. Genet.* vol. 52, no. 12, pp. 1271–1281. Dec. 2020.

28. F.M. Davis, A. denDekker, A.D. Joshi, *et al.* Palmitate-TLR4 signaling regulates the histone demethylase, JMJD3, in macrophages and impairs diabetic wound healing. *Eur. J. Immunol.* vol. 50, no. 12, pp. 1929–1940. Dec. 2020.

29. C. Bergmann, A. Brandt, B. Merlevede, *et al.* The histone demethylase Jumonji domain-containing protein 3 (JMJD3) regulates fibroblast activation in systemic sclerosis. *Ann. Rheum. Dis.* vol. 77, no. 1, pp. 150–158. Jan. 2018.

30. T. Satoh, O. Takeuchi, A. Vandenbon, *et al.* The Jmjd3-Irf4 axis regulates M2 macrophage polarization and host responses against helminth infection. *Nat. Immunol.* vol. 11, no. 10, pp. 936–944. Oct. 2010.

31. M. Krämer, C. Dees, J. Huang, *et al.* Inhibition of H3K27 histone trimethylation activates fibroblasts and induces fibrosis. *Ann. Rheum. Dis.* vol. 72, no. 4, pp. 614–620. Apr. 2013.

32. M. Ciechomska, S. O'Reilly, S. Przyborski. Histone Demethylation and Toll-like Receptor 8-Dependent Cross-Talk in Monocytes Promotes Transdifferentiation of Fibroblasts in Systemic Sclerosis Via Fra-2. *Arthritis Rheumatol.* vol. 68, no. 6, pp. 1493–1504. Jun. 2016.

33. J.L. Kwapis, Y. Alaghband, A.J. López, *et al.* HDAC3-mediated repression of the Nr4a family contributes to age-related impairments in long-term memory. *J. Neurosci.* vol. 39, no. 25, pp. 4999–5009. Jun. 2019.

34. Y. Zhao, T. Nomiyama, H.M. Findeisen, *et al.* Epigenetic regulation of the NR4A orphan nuclear receptor NOR1 by histone acetylation. *FEBS Lett.* vol. 588, no. 24, pp. 4825–4830. Dec. 2014.

35. S. Yoo, S.-J. Ha. Generation of Tolerogenic Dendritic Cells and Their Therapeutic Applications. *Immune Netw.* vol. 16, no. 1, pp. 52–60. Feb. 2016.

36. G.M. Bell, A.E. Anderson, J. Diboll, *et al.* Autologous tolerogenic dendritic cells for rheumatoid and inflammatory arthritis. *Ann. Rheum. Dis.* vol. 76, no. 1, pp. 227–234. Jan. 2017.

37. I. Zubizarreta, G. Flórez-Grau, G. Vila, *et al.* Immune tolerance in multiple sclerosis and neuromyelitis optica with peptide-loaded tolerogenic dendritic cells in a phase 1b trial. *Proc. Natl. Acad. Sci. U. S. A.* vol. 116, no. 7, pp. 8463–8470. Apr. 2019.

38. T. Nikolic, J.J. Zwaginga, B.S. Uitbeijerse, *et al.* Safety and feasibility of intradermal injection with tolerogenic dendritic cells pulsed with proinsulin peptide—for type 1 diabetes. *Lancet Diabetes Endocrinol.* vol. 8, no. 6, pp. 470–472. Jun. 2020.

39. G.B. Ferreira, A.-S. Vanherwegen, G. Eelen, *et al.* Vitamin D3 induces tolerance in human dendritic cells by activation of intracellular metabolic pathways. *Cell Rep.* vol. 10, no. 5, pp. 711–725. Feb. 2015.

40. L. Groseanu, V. Bojinca, T. Gudu, *et al.* Low vitamin D status in systemic sclerosis and the impact on disease phenotype. *Eur. J. Rheumatol.* vol. 3, no. 2, pp. 50–55. Jun. 2016.

41. L. An, M.H. Sun, F. Chen, et al. Vitamin D levels in systemic sclerosis patients: A meta-analysis. Drug Des Devel Ther. vol. 11, pp. 3119–3125. Oct. 2017.

42. D. Giuggioli, M. Colaci, G. Cassone, *et al.* Serum 25-OH vitamin D levels in systemic sclerosis: analysis of 140 patients and review of the literature. *Clin. Rheumatol.* vol. 36, no. 3, pp. 83–590. Mar. 2017.

43. K. Palumbo-Zerr, P. Zerr, A. Distler, *et al.* Orphan nuclear receptor NR4A1 regulates transforming growth factor-β signaling and fibrosis. *Nat. Med.* vol. 21, no. 2, pp. 150–158. Feb. 2015.

44. D. Farge, C. Henegar, M. Carmagnat, *et al*. Analysis of immune reconstitution after autologous bone marrow transplantation in systemic sclerosis. *Arthritis Rheum*. vol. 52, no. 5, pp. 1555–1563. May 2005.
45. L.C.M. Arruda, K.C.R. Malmegrim, J.R. Lima-Júnior, *et al*. Immune rebound associates with a favorable clinical response to autologous HSCT in systemic sclerosis patients. *Blood Adv*. vol. 2, no. 2, pp. 126–141. Jan. 2018.
46. D. Farge, L.C.M. Arruda, F. Brigant, *et al*. Long-term immune reconstitution and T cell repertoire analysis after autologous hematopoietic stem cell transplantation in systemic sclerosis patients. *J. Hematol. Oncol*. vol. 10, no. 1, pp. 21. Jan. 2017.
47. U. Hainz. Monocyte-mediated T-cell suppression and augmented monocyte tryptophan catabolism after human hematopoietic stem-cell transplantation. *Blood*. vol. 105, no. 10, pp. 4127–4134, May 2005.
48. L. Stern, H. McGuire, S. Avdic, *et al*. Mass Cytometry for the Assessment of Immune Reconstitution After Hematopoietic Stem Cell Transplantation. *Front. Immunol*. vol. 9, p. 1672. Jul. 2018.
49. J. D. Isaacs, B. L. Hazleman, K. Chakravarty, et al. Monoclonal antibody therapy of diffuse cutaneous scleroderma with CAMPATH-1H. *J. Rheumatol*. vol. 23, no. 6, pp. 1103–1106. Jun. 1996.
50. L. Costelloe, J. Jones, A. Coles. Secondary autoimmune diseases following alemtuzumab therapy for multiple sclerosis. *Expert Rev Neurother*. vol. 12, no. 3, pp. 335-341. Mar. 2012.
51. M. Tenspolde, K. Zimmermann, L.C. Weber, *et al*. Regulatory T cells engineered with a novel insulin-specific chimeric antigen receptor as a candidate immunotherapy for type 1 diabetes. *J. Autoimmun*. vol. 103, pp. 102289. Sep. 2019.
52. M. Akamatsu, N. Mikami, N. Ohkura, *et al*. Conversion of antigen-specific effector/memory T cells into Foxp3-expressing Tregcells by inhibition of CDK8/19. *Sci. Immunol*. vol. 4, no. 40, pp. eaaw2707. Oct. 2019.
53. A. Ben-nun, H. Wekerle, I.R. Cohen. Vaccination against autoimmune encephalomyelitis with T-lymphocite line cells reactive against myelin basic protein. *Nature*. vol. 292, no. 5818, pp. 60–61. Jul. 1981.
54. J.-W. Sidhom, H.B. Larman, D.M. Pardoll, et al. DeepTCR is a deep learning framework for revealing sequence concepts within T-cell repertoires. *Nat. Commun*. vol. 12, no. 1, pp. 1605. Dec. 2021.
55. Z. Zhang, D. Xiong, X. Wang, *et al*. Mapping the functional landscape of T cell receptor repertoires by single-T cell transcriptomics. *Nat. Methods*. vol. 18, no. 1, pp. 92–99. Jan. 2021.
56. K. Yugi, H. Kubota, A. Hatano, *et al*. Trans-Omics: How To Reconstruct Biochemical Networks Across Multiple 'Omic' Layers. *Trends Biotechnol*. vol. 34, no. 4, pp. 276–290. Apr. 2016.
57. R. Ai, T. Laragione, D. Hammaker, *et al*. Comprehensive epigenetic landscape of rheumatoid arthritis fibroblast-like synoviocytes. *Nat. Commun*. vol. 9, no. 1, pp. 1921. Dec. 2018.
58. F. Zhang, K. Wei, K. Slowikowski, *et al*. Defining inflammatory cell states in rheumatoid arthritis joint synovial tissues by integrating single-cell transcriptomics and mass cytometry. *Nat. Immunol*. vol. 20, no. 7, pp. 928–942. Jul. 2019.
59. M. van der Kroef, M. Castellucci, M. Mokry, *et al*. Histone modifications underlie monocyte dysregulation in patients with systemic sclerosis, underlining the treatment potential of epigenetic targeting. *Ann. Rheum. Dis*. vol. 78, no. 4, pp. 529–538. Apr. 2019.
60. C.M. Lanata, I. Paranjpe, J. Nititham, *et al*. A phenotypic and genomics approach in a multi-ethnic cohort to subtype systemic lupus erythematosus. *Nat. Commun*. vol. 10, no. 1, pp. 3902. Dec. 2019.
61. D. Wang, S. Bodovitz. Single cell analysis: the new frontier in 'omics'. *Trends Biotechnol*. vol. 28, no. 6, pp. 281–290. Jun. 2010.
62. D. Xue, T. Tabib, C. Morse, *et al*. Expansion of FCGR3A + macrophages, FCN1 + mo-DC, and plasmacytoid dendritic cells associated with severe skin disease in systemic sclerosis. *Arthritis Rheumatol*. no. 2, p. art.41813 *(online ahead of print)*. May 2021.
63. Q. Liu, L.C. Zaba, A.T. Satpathy, *et al*. Chromatin accessibility landscapes of skin cells in systemic sclerosis nominate dendritic cells in disease pathogenesis. *Nat. Commun*. vol. 11, no. 1, pp. 5843. Dec. 2020.
64. J.N. Weinstein, E.A. Collisson, G.B. Mills, *et al*. The Cancer Genome Atlas Pan-Cancer analysis project. *Nat. Genet*. vol. 45, no. 10, pp. 1113–1120. Oct. 2013.

8

Appendix

English Summary
Nederlandse Samenvatting
Acknowledgments
Curriculum Vitae
List of Publications

A

ENGLISH SUMMARY

Background and aims of this thesis

Systemic sclerosis (SSc) is a complex, heterogeneous autoimmune disease characterized by vascular abnormalities, immune involvement and extensive fibrosis of the skin and internal organs. Immune system dysregulation is recognized as one of the main culprits of SSc pathogenesis, however, it remains unclear what molecular mechanisms underlie this. By applying various high-throughput omics approaches, the studies presented in this thesis aim to uncover the molecular mechanisms contributing to immune cell dysregulation in SSc.

Factors driving innate immune cell dysregulation in SSc

In **chapter 2**, we performed transcriptomic profiling of monocytes from healthy individuals and SSc patients to identify lncRNAs involved in the regulation of toll-like receptor (TLR) induced pro-inflammatory responses in these cells. We found the lncRNA NRIR (Negative Regulator of the IFN Response) to be upregulated in SSc as compared to healthy monocytes. Characterization of NRIR function by siRNA mediated knockdown showed that this lncRNA is involved in the positive regulation of interferon (IFN) inducible genes. Notably, we also found NRIR to be upregulated in a publicly available RNA-sequencing dataset of monocytes from systemic lupus erythematosus (SLE) patients, which, like SSc, are also characterized by the presence of an IFN signature. These results indicate a role for NRIR in controlling IFN responses in monocytes from SSc patients, and potentially other autoimmune diseases.

In **chapter 3**, the lncRNA profile of SSc monocytes was investigated in more detail. Here we show that lncRNAs are also implicated in regulation of other pathways important for monocyte biology, including monocyte apoptosis and cytokine secretion. We identified the lncRNA PSMB8-AS1 as a potential regulator of immune related pathways in SSc monocytes. siRNA mediated knockdown of PSMB8-AS1 reduced the secretion of the cytokines IL-6, TNFα and IL-8 by stimulated monocytes, further highlighting lncRNAs as novel molecular factors contributing to monocyte dysregulation in SSc.

In **chapter 4**, we investigated the epigenomic landscape of monocytes from fifteen healthy controls and sixty patients with SSc, SLE or rheumatoid arthritis (RA). Whereas SSc and SLE monocytes were marked by a strong dysregulation of IFN and TNFα signaling pathways, RA monocytes lacked the IFN signature and were highly enriched for TNFα, TGFβ, and collagen formation pathways. The dysregulation of these disease specific pathways was already imprinted at the histone level, showing that aberrances in histone marks selectively skew SSc, SLE and RA monocytes towards distinct pro-inflammatory phenotypes. We also identified numerous bivalent promoters, of which many displayed increased levels of H3K4me3 in RA and eaSSc, and to a limited extent dcSSc monocytes. Interestingly, a large number of bivalent domains were identified in promoters of genes related to response to fibroblast growth factors, ECM organization, and vasculogenesis.

In **chapter 5**, we identified the nuclear receptor 4A subfamily (NR4A1, NR4A2, NR4A3), as important transcriptional repressors of inflammation in conventional dendritic

cells (cDCs) from SSc patients. Characterization of the genome-wide binding sites of NR4As in resting and stimulated cDCs by ChIP-sequencing showed that NR4As are strongly involved in transcriptional programs underlying cDC dysregulation in SSc. Indeed, functional experiments using agonists targeting NR4As showed that these receptors are involved in cytokine production and modulation of T cell activation by cDCs. Thus, NR4As are important negative regulators of immune pathways in cDCs, and NR4A downregulation potentially contributes to the dysregulation of these cells in SSc patients.

Skewing of the adaptive immune system in SSc

In the second part of this thesis, the dynamics of the T cell repertoire in SSc were studied to better understand the role of antigen specific T cell responses in SSc pathogenesis. In **chapter 6**, we show that the TCR repertoire in SSc patients is highly persistent over time, which is likely driven by antigenic selection. Moreover, we identified clusters of TCRs with similar specificities in SSc patients over time, representing groups of T cells that are likely to recognize the same or highly similar antigens.

To determine to what extent the longitudinal persistence of circulating T cells are characteristic of SSc, in **chapter 7** we studied the immune cell architecture and TCR repertoire dynamics of peripheral blood and affected joints of juvenile idiopathic arthritis (JIA) patients. Whereas in SSc patients T cells from peripheral blood were highly clonally expanded, only T cells obtained from affected joints from JIA patients exhibited an expanded profile while circulating T cells did not. These results indicate that tissue specific dominant (auto-)antigens in JIA patients heavily skew the TCR repertoire, while in SSc, the potential antigens might be more ubiquitously expressed. Additionally, JIA patients were characterized by a strong expansion and persistence of regulatory T cells (Tregs) rather than effector T cells, whereas in SSc, effector CD4+ and CD8+ T cells were highly expanded over time.

Concluding remarks

Altogether, the work presented in this thesis shows that immune dysregulation in SSc can be attributed to aberrances at various levels of molecular organization. These include the regulation of TLR signaling by lncRNAs, epigenomic imprinting of histone modifications and downregulation of immune regulatory transcription factors in monocytes and DCs. Enhanced activation of these innate immune cells has the potential to cue the adaptive immune system and orchestrate the generation of highly clonal autoreactive T cell repertoire. These insights offer new avenues for the development of novel therapeutics for SSc as well as other autoimmune diseases.

A

NEDERLANDSE SAMENVATTING

Achtergrond en doelstelling van dit proefschrift

Systemische sclerose (SSc) is een complexe, heterogene auto-immuunziekte die wordt gekenmerkt door vasculaire afwijkingen, activatie van het immuunsysteem en uitgebreide fibrose van de huid en inwendige organen. Ontregeling van het immuunsysteem wordt gezien als een van de belangrijkste oorzaken van SSc-pathogenese, maar het blijft onduidelijk welke moleculaire mechanismen hieraan ten grondslag liggen. De studies in dit proefschrift, waar verschillende omics-technieken worden toegepast, hebben als doel de moleculaire mechanismen bloot te leggen die bijdragen aan de ontregeling van het immuunsysteem in SSc.

Ontregeling van het aangeboren immuunsysteem in SSc

In **hoofdstuk 2** hebben we het transcriptoom van monocyten van gezonde controles en SSc patiënten onderzocht om lange niet-coderende RNAs (lncRNAs) te identificeren die betrokken zijn bij de regulatie van door toll-like receptor (TLR) geïnduceerde pro-inflammatoire processen. De lncRNA NRIR had een hogere expressie in monocyten van SSc patiënten dan gezonde monocyten. Met behulp van knockdown experimenten hebben wij aangetoond dat NRIR betrokken is bij de positieve regulatie van interferon (IFN) induceerbare genen die belangrijk zijn in de pathogenese van SSc. Daarnaast was de expressie van NRIR ook verhoogd in monocyten van patiënten met systemische lupus erythematodes (SLE), die, net als SSc patiënten, worden gekenmerkt door een verhoogde expressie van IFN induceerbare genen. Deze resultaten wijzen op een rol voor NRIR bij het controleren van IFN-responsen in monocyten van SSc patiënten en mogelijk ook bij andere auto-immuunziekten.

In **hoofdstuk 3** tonen we aan dat lncRNAs ook betrokken zijn bij de regulatie van andere processen die belangrijk zijn voor de biologie van monocyten, waaronder apoptose en cytokine secretie. Daarnaast identificeerden we PSMB8-AS1 als een potentiële regulator van immuun gerelateerde processen in monocyten van SSc patiënten. Verder laten we zien dat knockdown van PSMB8-AS1 in monocyten leidt tot een verminderde secretie van de pro-inflammatoire cytokines IL-6, TNFα en IL-8. Deze resultaten tonen aan dat lncRNAs nieuwe moleculaire factoren zijn die bijdragen aan ontregeling van monocyten in SSc.

In **hoofdstuk 4** hebben we gekeken naar de correlatie tussen het transcriptoom en het epigenoom in monocyten van vijftien gezonde controles en zestig patiënten met SSc, SLE of reumatoïde artritis (RA). Terwijl monocyten van SSc en SLE patiënten werden gekenmerkt door een sterke ontregeling van IFN en TNFα signalering, misten monocyten van RA patiënten de IFN-signatuur en waren sterk verrijkt voor genen gerelateerd aan TNFα- en TGFβ signalering en collageenvorming. Deze ontregeling was al op het niveau van histonen ingeprent, wat aantoont dat selectieve afwijkingen in het epigenoom van SSc-, SLE- en RA-monocyten verschillende pro-inflammatoire fenotypes onder ligt. We vonden ook talrijke genen met bivalente promotors, waarvan vele verhoogde niveaus van H3K4me3 vertoonden in monocyten van RA en specifieke subtypes van SSc patiënten (eaSSc en dcSSc). Een groot aantal van deze bivalente

domeinen werden geïdentificeerd in promotors van genen geassocieerd met de respons op fibroblast groeifactoren, ECM-organisatie en vasculogenese.

In **hoofdstuk 5** laten we zien dat de transcriptiefactoren NR4A1, NR4A2, en NR4A3 belangrijke remmers zijn van pro-inflammatoire responsen in conventionele dendritische cellen (cDCs). Door middel van ChIP-seq tonen we aan dat een groot deel van de genen die differentieel tot expressie komen in cDCs van dcSSc patiënten direct gereguleerd worden door NR4As. Dit suggereert dat NR4As sterk betrokken zijn bij transcriptionele programma's die ten grondslag liggen aan de ontregeling van cDCs in SSc patiënten. Functionele experimenten met agonisten voor NR4As toonden aan dat deze transcriptie factoren betrokken zijn bij cytokineproductie en T-cel activatie door cDCs. NR4As zijn dus belangrijke negatieve regulatoren van de immuunrespons in cDCs en hun verlaagde expressie draagt mogelijk bij aan de ontregeling van deze cellen in SSc patiënten.

Ontregeling van het adaptieve immuunsysteem in SSc

In het tweede deel van dit proefschrift hebben we de dynamiek van het T-cel repertoire in SSc bestudeerd om de rol van antigeen specifieke T-cel reacties in SSc-pathogenese beter te begrijpen. Hiervoor onderzochten we in **hoofdstuk 6** het T-celreceptor (TCR) repertoire van circulerende T-cellen uit longitudinale monsters verkregen van vier SSc-patiënten. Het TCR repertoire in SSc-patiënten is zeer stabiel over tijd. Deze stabiliteit wordt waarschijnlijk gedreven door antigene selectie. Verder vonden we ook clusters van T-cellen in SSc patiënten die gekenmerkt werden door een hoge overeenkomst in hun TCR sequenties. Deze T-cellen herkennen hierdoor waarschijnlijk dezelfde of sterk vergelijkbare antigenen.

Om te bepalen in hoeverre de stabiliteit van circulerende TCR repertoire kenmerkend is voor SSc, hebben we in **hoofdstuk 7** de immuun cel compositie en TCR repertoire dynamiek van perifeer bloed en aangetaste gewrichten van patiënten met Juveniele Idiopathische Artritis (JIA) bestudeerd. Terwijl bij SSc-patiënten T-cellen uit perifeer bloed sterk klonaal geëxpandeerd waren, vertoonden bij JIA-patiënten alleen T-cellen verkregen uit aangetaste gewrichten een geëxpandeerd profiel. Deze resultaten geven aan dat weefselspecifieke dominante (auto-)antigenen bij JIA-patiënten T-cellen activeren, terwijl in SSc de potentiële antigenen waarschijnlijk door het hele lichaam tot expressie worden gebracht. Daarnaast werden JIA-patiënten gekenmerkt door een sterke expansie regulatoire T-cellen (Tregs), terwijl in SSc de effector T-cellen sterk waren geëxpandeerd. T-cel activering en expansie in JIA kan dus het gevolg zijn van het niet onderdrukken van autoreactieve T-cellen door Tregs, terwijl de activatie van effector-T-cellen in SSc waarschijnlijk wordt gedreven door een disregulatie van aangeboren immuun cellen.

Conclusie

Al met al laat het werk gepresenteerd in dit proefschrift zien dat de disregulatie van het immuunsysteem in SSc kan worden toegeschreven aan afwijkingen op verschillende niveaus van moleculaire organisatie in immuuncellen. Deze omvatten de regulatie van TLR-signalering door lncRNAs, epigenomische imprinting van histon-

A

modificaties en verlaagde expressie van immuun regulerende transcriptiefactoren in monocyten en DC's. De ontregeling van het aangeboren immuunsysteem leidt tot overactiviteit van het adaptieve immuunsysteem en het ontstaan van een zeer klonaal autoreactief T-celrepertoire. Deze inzichten in vormen belovende aangrijpingspunten voor het ontwikkelen van nieuwe therapieën voor SSc en andere auto-immuunziekten.

A

ACKNOWLEDGEMENTS

As I'm writing these acknowledgments, my PhD journey is finally coming to an end. The past four and a bit years have been filled with exciting moments as well as some struggles. Luckily I never had to go through these moments alone, as I had the opportunity to share some of my tears, but mostly smiles, with some amazingly talented scientists and friends.

Supervisors

I want to start by thanking my promotor, prof. dr. **Timothy Radstake**. Dear **Tim**, starting from my time as a master student in your group, you have provided me with numerous opportunities to nurture my scientific interests. From the first internship at the UMCU, to my time in Italy, to the switch from wet lab to computational biology during my PhD, you have always given me the room to discover different sides of research and find my place in the world of science. For this I am truly grateful. Thank you for sharing your expertise, but most of all your unlimited enthusiasm and passion for clinical research and, of course, systems medicine.

Then my first co-promotor, dr. **Aridaman Pandit**. Dear **Aridaman**, thank you for adopting me as your PhD student and guiding me as I took my first steps in the world of computational immunology. I remember us laughing together because I was writing down my first R scripts in a Word document, and you would stand behind me and tell me to do things by proper coding in the terminal instead of using the "clicky" interface. Thank you for your time and patience in teaching me all about bioinformatics and systems immunology. Another thing that I should thank you for is infecting me with your enthusiasm for T-cells, which I think spread even faster than COVID-19. In the beginning of my PhD I would complain to you about how we were always discussing T-cell related papers in our journal clubs, and I always said that T-cells were "just boring". Eventually, your passion for T-cell biology and TCR sequencing won me over, and I'm now even doing a postdoc on gene regulation in… T-cells! I also really enjoyed our extensive brainstorming sessions, and you were never afraid to try out new hypotheses and experiments. Although the experimental side could sometimes be extremely challenging, you always encouraged me to keep going and helped to think of new ways to address our biological questions. Aridaman, thank you for taking me in and being my mentor for these four years!

Next, I want to thank my second co-promotor, dr. **Marianne Boes**. Dear **Marianne**, I remember our first talk in your office, when we were discussing the possibility of me becoming part of your research group. I was looking for someone to help me with the aspects of the wet lab part of my PhD, and I found a great supervisor in you. You would always found different ways of looking at my projects, and helped to address the computational questions from a molecular immunology point of view. Besides your knowledge on immunology, one thing that you said that will always stick with me, is that you consider your week successful if you helped at least three people. I think this might

be an even more important lesson than all the immunological knowledge that you passed on to me. Thank you for all your expertise and guidance, it was a pleasure to be part of the Boes group!

The Winchester

Some might know The Winchester as a tavern where the main characters in the movie Shaun of the Dead take refuge during a zombie apocalypse. Although I never saw the movie, I was lucky to have a similar place of refuge (*De Winnie*) during the sometimes apocalyptic feeling PhD. **Rianne**, **Anneline**, **Wouter** and **Safae**, thank you for all the hours we spend in our own little Winchester tavern! It was so nice to have our safe space where we could literally share everything.

Rianne (*Rianus*), it was a pleasure sharing an office with you, and I already miss your crazy random raptor sounds and fietsopa quotes. I will never forget your face during the handstand push up CrossFit lesson from Willy, the millions of sambal jars you brought, krokettendag, your super matching scrunchies, and our hunt for the best Zwift machines (and all other random shit we doubted about buying online). You are an amazingly sweet person, sometimes even too sweet for the rest of the world, and I'm sure that there are great things out there for you in the future. **Anneline** (*Annelineke*, I will spare you from your other nickname), thank you for always being there and always being honest. No matter the problem, you always offered a listening ear and I felt safe to discuss really anything with you. I'm super proud that you have become a real lab*****, with the cherry on top that you taught me how to use the plate reader! You are going to be an awesome rheumatologist, and remember, we need more female rheumatology profs so I'm counting on you ;). **Wouter** (my beer buddy), I have never met anyone that is so passionate about arts and crafts… with CRISPR! It was really fun to hear about all your experiments and creative ways to do gene-editing. But many times we had even more important stuff to discuss: beer! Thank you for organizing the awesome trips to De Molen, getting all the delicious IPAs (on discount) from your favourite shop in Houten, and of course the nice conversations that we had during beer nights. I'm really looking forward to our future Dutch-German beer exchanges (no… more…. Reinheitsgebot…). **Safae** (*Safoe*) although you joined our little tavern a bit later, it feels like you have been there for the whole of my PhD. I don't think I have ever met anyone who owns more makeup palettes than you, and I will miss your beautiful sparkly Pat McGrath eyeshadow as much as your sparkly personality. Apart from the makeup outside, I think you are a beautiful person inside as you always take the time to listen and speak up for what is right. Thank you for all the amazing Hector stories, arranging my first Anastasia Beverly Hills palette, and of course for introducing us to Kunefe.

Computational biology and Systems Immunology (CoSI) group

Dear CoSIs (or was it SICos?), you were the first computational biologists that I worked with, and I have learned so much from you guys. Although we all have totally different backgrounds and worked on many different projects, we always had very nice scientific discussions. It was great to be part of such a diverse group of people, not only

from the scientific side, but also from the cultural side (Chinese, Indian, Dutch). Thanks for all the good times!

Weiyang, you truly are our R wizard and you would always take the time to help out with any coding problem. Besides that you are also quite the culinary wizard, with your most amazing food invention ever being the "KroCro" (kroket on a croissant). I still don't know if you are a genius or mad for this. Thank you for all your lessons about machine learning and I hope you, Yicheng and Bohui are doing well in China! **Abhinandan** (*Abhi*), I think I bothered you the most at the beginning of my PhD, when I was still learning how to code and would tap on your shoulder to ask for help all the time. It was great to share an office with you, and I will never forget your attempt at doing the cinnamon challenge. Thanks for all the laughs and help with computational problems. **Jingwen**, you are an amazingly kind and hardworking person, and I'm sad that we did not get to spend more hours at the WKZ together due to COVID. It was great meeting you and your son Tim (the other Tim), and enjoying the food at your house. I still have to steal your recipe for the sweet and sour fish someday! My computational master students, **Joeri** and **Nandhini**, thank you for the work during your internships. I hope you both have a bright future ahead of you. All other (ex-)members of the CoSI group, **Wannisa**, **Terry**, **Zora**, **Laura**, **Hidde**, **Bram**, **Charles** and **Carlotta**, thanks for sharing your expertise and stories, and for sitting through all those dendritic cell journal clubs!

Radstake Group

I have had the pleasure of being a part of the Radstake group not only during my PhD, but also during my first master internship. During this time, many people came and went, but everybody was always willing to help and in for some good times.

Cowboy Maarten, it was really awesome to work with you, and you are a great scientist, colleague, gardener, guild leader (McMarsh), neighbour and friend. Despite the fact that you almost ruined my PhD by getting me addicted to Legendary Games of Heroes, I am really glad that I got to meet you and spend time with you in Verona, at the UMCU and of course at De Kanaalstraat. Thank you for trusting me with the analysis of your final PhD project. It was an honour to work with you and I will always hold our friendship dearly. **Sanne** (*Sannekoek*), my fellow buffycoat slayer. Although 99% of our experiments failed (cDCs fault, not us), I always enjoyed spending time in the lab with you. I will really miss our cDC brainstorming sessions, lunch at the camels and gossiping at the cell culture hood. Thank you for listening to all my problems and bullshit stories, and of course, thank you for all the hours of cDC isolation and FACS sorting. I'm already sad not to be doing experiments with you anymore! **Rina** (*Rinus*), when I started my PhD we were in a small subgroup together working on the cDC differentiation model. Although the subgroup didn't last so long, I was really happy to spend time with you during the rest of my PhD. No matter the problem or the timing, you were always willing to help and I learned a lot from you in the lab. Thanks for all your help and of course the kudo's on Strava. **Sarita** (*Sartita*), thank you for organizing everything in the lab and finances, and making sure that there were always enough MACS beads for the many, many cDC

experiments. You were a really fun person to work with, and I will miss your stupid jokes, Pokémon battles, and great outfits including all your awesome Adidas and Nike sneakers. **Elena**, my supervisor! Thank you for having me as your master student, and also for being a great colleague after that. You were the first person to guide me around the lab and teach me how to do qPCR, with the most important thing being that everything should be super, super, super clean. **Ana** (*Ana banana*), thank you for sharing all your knowledge on cDCs and helping me whenever I had questions about any protocols. It was really nice to discuss cDCs, lncRNA experiments and other (not so scientific stuff) with you. **Chiara**, thanks for all your help around the lab, great scientific discussions, and always keeping me motivated to keep learning Italian. **Nanette**, kipnugget warrior. It was so much fun to work with you, with your love for stupid WhatsApp stickers, pizza met vlees en zilverfolie, and your amazing tolerance for tequila. Thanks for all the laughs, and for introducing me to CrossFit (with Willy)! I hope you, Jakob and Veerle are doing well. **Maarten Hillen**, you always organized so many fun activities around the lab, with the icing on the cake being our awesome lab retreat, okTIMberfest. Besides your scientific knowledge, you also have an incredible amount of Pokémon knowledge, and you would always make sure that Team Blauw + spion had the best teams to win the raids at the WKZ gym. **Juliette**, you are one of the most hardworking persons I have ever met. You are always in to help anybody, even when you don't really have the time. Besides that, you are always in for fun stuff and helped to organize so many activities around the lab, including the early mentioned legendary okTIMberfest. Thanks for all the support and fun times! Monday morning meeting group, **Wiola**, **Maili**, **Samu**, **Tiago**, **Andrea** and **Ralph**, thank you for all the scientific discussions and sharing your knowledge on SSc, but above that, thanks for all the crazy cat stories. I hope you are all doing well, and **Wiola**, I hope you did not get any more cats ;). **Jonas**, I have never met anyone who has some much knowledge about basically anything. I always really enjoyed brainstorming with you about cDCs, and you always came up with great ideas for experiments and storylines for papers. Thanks for all your support over these four years. **Joel**, thank you for all the scientific discussions and always asking many interesting questions. You made me look at my projects from totally different perspectives and also helped to make sure that I could communicate my science to different types of audiences, from molecular biologists to clinicians. **Michel**, **Cornelis** and **Bea**, thank you for all your help and technical support around the lab. **Diana**, thanks for all your help with the bureaucracy and making sure everything went smoothly at the end of my PhD. All other past and present members of the Radstake group, **Roos**, **Kamil**, **Fleurieke**, **Emmerik**, **Nienke**, **Lude**, **Els**, **Sandra**, **Alsya**, **Luuk**, **Sofie**, **Nadia**, **Tessa**, **Anne-Mieke**, **Kris**, **Jorre**, **Pawel**, **Marta**, **Marzia**, thank you for all the amazing memories.

Boes Group

Dear Boes group members, it was so much fun to be part of such a diverse research group. I learned a lot from you guys, and it was always great to listen to you and learn from your research projects.

Mischa, thank you for showing us why IgA is more awesome than the other

Igs, and thank you for all the jokes in the protein lab when you were doing your million ELISAs. **Patrick**, you always made me laugh when you popped up on screen during the digital Boes meetings, looking a bit bewildered because your neighbours had been partying all night. I hope they are a bit more quiet by now. **Angela**, my trainee, it was really fun to be doing FACS experiments with you in the short time that we had together at the UMC. It was an honour to be your trainer. **Quirine** and **Niels**, thank you for the hard work during your internships. You are both very bright and fun to work with, and I'm sure you are going to be great scientists. All other past and present Boes lab members, **Juliette**, **Joel**, **Sarah**, **kleine Sarah**, **Bas**, **Sandra**, **Fransesca**, **Lotte**, **Xiaogang**, thanks for all the scientific discussions and great times at the Boes BBQs.

Center for Translational Immunology (CTI)

I want to thank everybody in the CTI for the great times in the lab, at meetings and of course at the borrels. Prof. **Linda Meyaard**, head of the CTI and inspirational leader. **Mareille**, **Saskia**, **Yvonne**, **Margje**, **Sigrid** and **Gerrit**, thank you for all the administrative help, IT help and guidance in and around the labs. BorrelCie, **Rianne**, **Marlot**, **Lucas**, **Marwan**, thanks for all the borrels, parties, eggcellent easter egg hunts, and all the other fun times. **Lucas**, **Marwan**, **Tiago**, **Rianne**, **Abhi**, **Na**, **Erinke**, thanks for being awesome roomies, and thanks for putting up with me and Rianne talking all the time. It was great to share the office with you. **Lucas**, thanks for all the great nerdy conversations about epigenetics and RNA, das toch goud! **Marwan**, I really enjoyed our scientific discussions, even though they could sometimes get a little heated ;). Too bad we couldn't fill the movie board with all of our tickets because of COVID. **Femke**, and **Gerdien**, thank you for involving me in your work trusting me with the final analyses for the JIA T-cell paper. It was great to work together with you. **Kristin**, your course during the bachelor BMW got me really enthusiastic about immunology. Thank you for the nice talks in the coffee room, and all your efforts in the educational programs. **Ellen Kaan**, **Ellen van Vroonhoven**, **Laura**, thanks for being awesome teammates during the I&I PhD retreat. **Akashdip**, **Ines**, **Suus**, **Guilio**, **Selma**, **Saskia**, **Julia**, **Vanya**, thanks for the nice talks in the hallways and at the coffee machines.

IJCLUB members

I want to thank all (ex) members of the immuno journal club (IJCLUB) for all the great scientific discussions, explanations about mathematical modelling, and help with preparing manuscripts and presentations. **Rob** and **Can**, thank you for your help in the work regarding the TCR sequencing, your scientific feedback and great critical questions. **Juliane**, **Peter**, **Arpit**, and **Erdem**, thank you for all the great discussions during IJCLUBs and ImmuniTea. It was really nice to see what you guys were working on, especially because it was so out of my comfort zone. Thank you for taking your time to explain everything and introducing me to the world of computational modelling. **Juliane** and **Erdem**, thank you for your efforts in organizing the IJCLUB every week. **Peter**, thanks for the awesome peanut butter brownies and help with my TCR paper. **Arpit**, thanks for all the discussions and great laughs.

Verona

Care **Marzia**, **Flavia**, **Monica** e **Barbara**, grazie per tutto quello che mi avete insegnato, e per i belli momenti che ho passato con voi a Verona. Dear **Marzia**, you were my first internship supervisor in Utrecht, and thanks to you I got the opportunity to do a second internship in Verona. I am very grateful for all the time that you spend teaching me about miRNAs, lncRNAs, and SSc and all the great discussions that we had (often till very late in the night). I have two chapters in my thesis that are based on the work in Verona, and I am very thankful for all your help in this. It was a pleasure to work with you, and I hope maybe one day we can meet again for a wine at Hostaria Verona. Dear **Babi**, I had such a great time with you in Verona, with all the laughs that we shared. You were the first person that taught me about bioinformatics and showed me how to do my first R analysis. I will always remember our great times together in the lab, till very late in the night and early in the morning. Thank you for everything you have done, and I hope one day I can come to visit you in Trento. Dear **Flavia**, thank you for offering me a place in your group for my internship, and for all your help with the manuscripts that we published during my PhD. I wish you all the BEST ;)! Grazie per tutto!

Friends and Family

Lieve **Morgan** (*Momo*) en **Jeanneke** (*JJ*), Ik heb even geteld, en in de tijd dat ik mijn PhD begon en nu zijn er minstens 16 seizoenen van RPDR uitgekomen… GASP! Bedankt voor alle leuke watch nights en bier avondjes, en bedankt dat jullie altijd wilden luisteren naar mijn PhD life problemen. Zonder jullie en RuPaul had ik het echt nooit overleefd. **Anouk** (was het nou *snork* of toch *broodje ORI*?), **Eline** (*elinalien*) en **Eliza** (*enieza*), het lijkt zo kort en tegelijkertijd zo lang geleden dat we elkaar ontmoetten als kersverse eerstejaars BMW studenten. Ik ben heel blij dat we al die jaren vriendinnen zijn gebleven, en dat we altijd onze verhalen met elkaar kunnen delen ondanks dat we elkaar maar een paar keer per jaar zien. Bedankt voor het aanhoren van alle ups en downs van mijn PhD tijd! **Fleur** en **Farah**, bedankt dat ik jullie bodypump clubje mocht joinen. Het was altijd super fijn om na een lange dag op het lab lekker tegen jullie te kunnen klagen en met (net iets te zware) barbells te rammen.

Lieve **Marion**, bedankt voor alle steun en interesse die je altijd toonde in mijn werk. **Harry**, omdat je zo van dankwoorden houdt, hierbij: bedankt. Lieve **Heidi**, bedankt voor alle gekkigheid en leuke gesprekjes die we hebben gehad over cellen en het immuunsysteem. Lieve **Pap** en **Mam**, bedankt voor alle steun tijdens mijn studie en PhD tijd. Jullie hebben er voor gezorgd dat Lotte, Dirk en ik altijd konden doen wat we wilden en staan altijd voor ons klaar. Zonder jullie steun had ik dit zeker nooit bereikt. Lieve **Lotte** en **Dirk** (*prak*), addergebroed, bedankt voor alle afleiding, kattenfilmpjes en lekkere biertjes, en onthoudt: vo! voor de academie!

Lieve **Rianne**, bedankt dat je de afgelopen jaren altijd voor me klaar hebt gestaan. Jij wist mij altijd op te vrolijken als ik weer eens chagrijnig van het lab thuis kwam omdat een experiment was mislukt of een meeting niet helemaal vlekkeloos was verlopen. Ik vond het super fijn om samen met jou te brainstormen over mijn projecten,

A

natuurlijk geheel in consultancy stijl door samen flowcharts te maken of dingen uit te tekenen op post-its. Als ik het er met jou samen over had leken de bergen werk die ik nog moest verzetten altijd iets minder hoog, en als ik het even niet meer zag zitten gaf jij mij altijd het gevoel dat ik de wereld wel aan kon. Bedankt voor alle lieve woorden, prutjes pasta, samen avondjes bankhangen en samen avondjes werken, en je onvoorwaardelijke steun. Ik heb ontzettend veel zin in het nieuwe avontuur in Heidelberg samen met jou!

SSc patients, clinicians and research nurses

Lastly, I would like to thank all the **SSc patients** for donating their precious time and samples, as well as all the **clinicians** and **research nurses** involved in sample collection. The data presented in this thesis are the results of years of hard work and sample collection, and I hope our research can contribute to a better understanding of this rare autoimmune disease.

A

CURRICULUM VITAE

Nila Hendrika Servaas was born on the 30th of June 1993 in Gouda, The Netherlands. From 2005 till 2011 she attended high school at the Cals College in Nieuwegein. During this time, she developed a great interest in biology and uncovering the mechanisms behind disease. In 2011, Nila started the bachelor Biomedical Sciences at Utrecht University, after which she further explored her passion for research during her master Drug Innovation at Utrecht University. In her Bachelor's and Master's studies, she became fascinated by the immune system and its role in human health and disease. This led her to do an internship in the group of Prof. Timothy Radstake at the Center for Translational Immunology of the University Medical Center Utrecht in 2015. Here she focussed on circulating miRNAs as prognostic and diagnostic biomarkers for Systemic Sclerosis. After this, Nila got the opportunity to continue her research in the group of Prof. Flavia Bazzoni at the University of Verona, Italy, where she worked on the characterization of long-noncoding RNAs dysregulated in monocytes of Systemic Sclerosis patients. After graduating with honors from her Master's program, in 2017 she got a position as a PhD candidate in the research group of prof. Timothy Radstake, under the joint supervision of dr. Aridaman Pandit and dr. Marianne Boes. During her PhD, Nila focussed on integrating computational and experimental approaches to study immune cell dysregulation in Systemic Sclerosis. Alongside her research, she was the president of the Regional Student Group (RSG) Netherlands, part of the International Society of Computational Biology (ISCB), with the aim to promote interaction among young researchers in the field of bioinformatics and computational biology. In June 2021, Nila took the next step in her scientific career and started as a postdoctoral research fellow in the group of dr. Judith Zaugg at the European Molecular Biology Laboratory (EMBL) in Heidelberg, Germany. Here she is studying gene regulatory networks in human T cells.

LIST OF PUBLICATIONS

This thesis

- **N.H. Servaas***, B. Mariotti*, M. van der Kroef, C.G.K. Wichers, A. Pandit, F. Bazzoni, T.R.D.J. Radstake, M. Rossato. "Characterization of Long Non-Coding RNAs in Systemic Sclerosis Monocytes: A Potential Role for PSMB8-AS1 in Altered Cytokine Secretion", Int. J. Mol. Sci. 2021. *Authors contributed equally.

- **N.H. Servaas**, F. Zaaraoui-Boutahar, C.G.K. Wichers, A. Ottria, E. Chouri, A. Affandi, S. Silva-Cardoso, M. van der Kroef, T. Carvalheiro, F. van Wijk, T.R.D.J. Radstake, A.C. Andeweg, A. Pandit. "Longitudinal analysis of T-cell receptor repertoires reveals persistence of antigen-driven CD4+ and CD8+ T-cell clusters in Systemic Sclerosis", J. Autoimmunity, 2021.

- **N.H. Servaas**, J. Spierings, A. Pandit, J.M. van Laar. "The role of innate immune cells in systemic sclerosis in the context of autologous hematopoietic stem cell transplantation", Clin Exp Immunol., 2020.

- **N.H. Servaas***, K. Woolcock*, M. van der Kroef, J. Cole, L.L. van den Hoogen, T. Carvalheiro, E. Chouri, C.G.K. Wichers, C.P.J. Bekker, T.R.D.J. Radstake, C. Goodyear#, A. Pandit#. "Epigenomic imprinting underlies disease specific dysregulation of monocytes in systemic sclerosis, systemic lupus erythematosus and rheumatoid arthritis". In preparation. */#Authors contributed equally.

- **N.H. Servaas**, S. Hiddingh, E. Chouri, C.G.K. Wichers, A.J. Affandi, A. Ottria, C.P.J. Bekker, M. Cossu, S. Silva-Cardoso, M. van der Kroef, A.C. Hinrichs, T. Carvalheiro, N. Vazirpanah, L. Beretta, M. Rossato, F. Bonte-Mineur, T.R.D.J. Radstake, J.J.W. Kuiper#, M. Boes#, A. Pandit#. "Network-based multi-omics of CD1c+ dendritic cells implicates nuclear receptor subfamily 4A signaling as a key disease pathway in systemic sclerosis". In preparation. #Authors contributed equally.

- G. Mijnheer*, **N.H. Servaas***, J.Y. Leong, A. Boltjes, E. Spierings, P. Chen, L. Lai, A. Petrelli, S. Vastert, R.J. de Boer, S. Albani, A. Pandit#, F. van Wijk#. "Compartmentalization and persistence of dominant (regulatory) T cell clones indicates antigen skewing in juvenile idiopathic arthritis". Submitted for pulication. */#Authors contributed equally.

- B. Mariotti, **N.H. Servaas**, M. Rossato, N. Tamassia, M.A. Cassatella, M. Cossu, L. Beretta, M. van der Kroef, T.R.D.J. Radstake, F. Bazzoni. "The Long Non-coding RNA NRIR Drives IFN-Response in Monocytes: Implication for Systemic Sclerosis", Front. Immunol., 2019.

A

237

Other Publications

- E. Chouri, **N.H. Servaas**, C.P.J. Bekker, A.J. Affandi, M. Cossu, M.R. Hillen, C. Angiolilli, J.S. Mertens, L.L. van den Hoogen, S. Silva-Cardoso, M. van der Kroef, N. Vazirpanah, C.G.K. Wichers, T. Carvalheiro, S.L.M. Blokland, B. Giovannone, L. Porretti, W. Marut, B. Vigone, J.A.G. van Roon, L. Beretta, M. Rossato, T.R.D.J. Radstake, "Serum microRNA screening and functional studies reveal miR-483-5p as a potential driver of fibrosis in systemic sclerosis", J. Autoimmunity, 2018.

- T. Carvalheiro*, A.P. Lopes*, M. van der Kroef, B. Malvar-Fernandez, C. Rafael-Vidal, A.C. Hinrichs, **N.H. Servaas**, F. Bont-Minuer, M.R. Kok, L. Beretta, M. Zimmerman, W. Marut, J.M. Pego-Reigosa, T.R.D.J. Radstake, S. Garcia Perez, "Angiopoietin-2 Promotes Inflammatory Activation in Monocytes of Systemic Sclerosis Patients", Int. J. Mol. Sci. 2020. *Authors contributed equally.

- M.R. Hillen*, A. Pandit*, S.L.M. Blokland, S.A.Y. Hartgring, C.P.J. Bekker, E.H.M. van der Heijden, **N.H. Servaas**, M. Rossato, A.A. Kruize, J.A.G. van Roon, T.R.D.J. Radstake, "Plasmacytoid DCs From Patients With Sjögren's Syndrome Are Transcriptionally Primed for Enhanced Pro-inflammatory Cytokine Production", Front. Immunol., 2019. *Authors contributed equally.