Epigenomics and transcriptomics of systemic sclerosis CD4+ T cells reveal long-range dysregulation of key inflammatory pathways mediated by disease-associated susceptibility loci

Tianlu Li[1], Lourdes Ortiz-Fernández[2†], Eduardo Andrés-León[2†], Laura Ciudad[1], Biola M. Javierre[3], Elena López-Isac[2], Alfredo Guillén-Del-Castillo[4], Carmen Pilar Simeón-Aznar[4], Esteban Ballestar[1*†] and Javier Martin[2*†]

Abstract

Background: Systemic sclerosis (SSc) is a genetically complex autoimmune disease mediated by the interplay between genetic and epigenetic factors in a multitude of immune cells, with CD4+ T lymphocytes as one of the principle drivers of pathogenesis.

Methods: DNA samples exacted from CD4+ T cells of 48 SSc patients and 16 healthy controls were hybridized on MethylationEPIC BeadChip array. In parallel, gene expression was interrogated by hybridizing total RNA on Clariom™ S array. Downstream bioinformatics analyses were performed to identify correlating differentially methylated CpG positions (DMPs) and differentially expressed genes (DEGs), which were then confirmed utilizing previously published promoter capture Hi-C (PCHi-C) data.

Results: We identified 9112 and 3929 DMPs and DEGs, respectively. These DMPs and DEGs are enriched in functional categories related to inflammation and T cell biology. Furthermore, correlation analysis identified 17,500 possible DMP-DEG interaction pairs within a window of 5 Mb, and utilizing PCHi-C data, we observed that 212 CD4+ T cell-specific pairs of DMP-DEG also formed part of three-dimensional promoter-enhancer networks, potentially involving CTCF. Finally, combining PCHi-C data with SSc GWAS data, we identified four important SSc-associated susceptibility loci, *TNIP1* (rs3792783), *GSDMB* (rs9303277), *IL12RB1* (rs2305743), and *CSK* (rs1378942), that could potentially interact with DMP-DEG pairs cg17239269-*ANXA6*, cg19458020-*CCR7*, cg10808810-*JUND*, and cg11062629-*ULK3*, respectively.

(Continued on next page)

* Correspondence: eballestar@carrerasresearch.org; javiermartin@ipb.csic.es
†Lourdes Ortiz and Eduardo Andrés-León contributed equally to this work.
†Esteban Ballestar and Javier Martin share senior authorship.
[1]Epigenetics and Immune Disease Group, Josep Carreras Research Institute (IJC), 08916 Badalona, Barcelona, Spain
[2]Instituto de Parasitología y Biomedicina López-Neyra, Consejo Superior de Investigaciones Científicas (IPBLN-CSIC), Granada, Spain
Full list of author information is available at the end of the article

Li *et al. Genome Medicine*   (2020) 12:81

Page 2 of 17

(Continued from previous page)

Conclusion: Our study unveils a potential link between genetic, epigenetic, and transcriptional deregulation in CD4+ T cells of SSc patients, providing a novel integrated view of molecular components driving SSc pathogenesis.

Keywords: Systemic sclerosis, DNA methylation, Epigenetics, Long-distance regulation, Hi-C, Genetic susceptibility variants, CTCF

Background

Systemic sclerosis (SSc) is a chronic, progressive auto-immune disease of unknown etiology that is primarily characterized by extensive microvascular damage, de-regulation of immune cells, and systemic fibrosis of the skin and various organs. The estimated overall 10-year survival rate is 66%, which significantly decreases upon organ involvement [1]. SSc patients can be broadly categorized into three groups based on the extent of skin involvement: those with restricted involvement affecting the limbs distal to the elbows or knees with or without face and neck involvement are classified as limited cutaneous SSc (lcSSc) or those with proximal involvement affecting above the elbows and knees are classified as diffuse cutaneous SSc (dcSSc) and *sine* scleroderma with Raynaud's phenomenon, visceral involvement, and auto-antibodies but no skin involvement [2]. A proposed classification by the Royal Free Hospital in 1996, currently used by Spanish Scleroderma Registry (RESCLE), included early (eSSc) and very early scleroderma [3, 4]. The pathogenesis of SSc is not well-understood, in which disease onset and development appear to be multistep and multifactorial processes involving both genetic and environmental factors [5, 6].

Like most autoimmune diseases, genetics studies have revealed that SSc heritability is complex with the HLA loci playing an important contribution to disease risk [7]. Recent genome-wide association (GWAS) and immunochip array SNP studies revealed the presence of numerous non-HLA loci that associate with disease onset, including genes of type I interferon pathway, interleukin-12 pathway, and TNF pathway as well as B and T cell-specific genes [8–14]. Furthermore, recent chromatin interaction analyses have linked these genetic variants to potential target genes implicated in SSc disease progression [15]. However, genetic variants do not account for all of the genetic burden of SSc, where other factors, such as epigenetic dysregulation, play an indispensable role in disease pathogenesis [5, 16].

It is well-established that CD4+ T cells play a pivotal role in the pathogenesis of several autoimmune diseases. Abnormalities in the proportions of CD4+ T lymphocyte subpopulations were detected more than two decades ago in SSc patients [17, 18], and since then, several studies proposed mechanistic alterations in these lymphocyte populations that may contribute to disease manifestations.

Firstly, increased production of several CD4+ T cell-mediated cytokines, including IL-27, IL-6, TGF-β, and IL-17A, was detected in the serum of SSc patients, which may directly drive vascular dysregulation and fibrosis [19, 20]. Secondly, CD4+ T cells isolated from SSc patients were observed to be functionally impaired, as stimulation resulted in deregulated polarization towards Th17 expansion, as well as inherent diminished immune capacity of circulating Treg cells [21, 22]. Finally, aberrant interactions with other cell types, including mesenchymal stromal cells and fibroblasts, have been observed [23, 24]. The exact mechanism that drives CD4+ T cell deregulation in SSc is currently unknown; however, there is strong evidence that alterations in DNA methylation may be a primary culprit. DNA methylation plays a critical role in T cell polarization and activation, in which naïve T cells undergo reprograming of its methylome to increase accessibility of selective loci upon differentiation into different T helper lymphocyte subpopulations (reviewed in [25, 26]). Gene expression of DNA methylation-related proteins were observed to be deregulated in SSc [27]. Several studies have observed aberrant methylation of promoters of such genes as *CD40L* [28], *TNFSF7* [29], *FOXP3* [30], and IFN-associated genes [31], which may result in their aberrant expression.

In this study, we describe the relationship between DNA methylation and gene expression in SSc CD4+ T cells, and how aberrant DNA methylation potentially deregulates the expression of several important inflammatory genes through long-distance enhancer-promoter interactions. Furthermore, these alterations appear to correlate with the presence of genetic variants in nearby loci.

Methods

Patient cohort and CD4+ T cell isolation

This study included 48 SSc patients and 16 age- and sex-matched healthy donors (HD). Individuals included in this study gave both written and oral consent in regard to the possibility that donated blood would be used for research purposes. Samples were collected at Vall d'Hebron Hospital, Barcelona, in accordance with the ethical guidelines of the 1975 Declaration of Helsinki. The Committee for Human Subjects of the Vall d'Hebron Hospital and Bellvitge Hospital approved the study. Patients with SSc who fulfilled the 2013 American College of Rheumatology/ European League Against Rheumatism (ACR/EULAR)

criteria or the modified criteria proposed by LeRoy and Medsger in 1988 were included [32, 33]. Four clinical subsets were considered: diffuse SSc, limited SSc, sine SSc, and early SSc (patients who met the criteria for very early disease and also presented incipient visceral involvement or other manifestations (digital ulcers (DU) or pitting scars, telangiectasia, calcinosis, or arthritis)) [4]. Characteristics of patient cohort are summarized in Additional file 1: Table S1.

CD4+ T lymphocyte population was isolated from whole blood by fluorescence-activated cell sorting performed at the Unitat de Biologia (Campus de Bellvitge), Centres Científics i Tecnològics, Universitat de Barcelona (Spain). Briefly, peripheral blood mononuclear cells (PBMCs) were separated by laying on Lymphocytes Isolation Solution (Rafer, Zaragoza, Spain) and centrifuged without braking. PBMCs were stained with fluorochrom-conjugated antibody against CD4-APC (BD Pharmingen, New Jersey, USA) in staining buffer (PBS with 2 mM of EDTA and 4% FBS) for 20 min. Gating strategies were employed to eliminate doublets, cell debris, and DAPI+ cells. Lymphocytes were separated by forward and side scatter, in which CD4+ cells were separated by positive selection.

RNA/DNA isolation
RNA and DNA were isolated from the same cell pellet utilizing AllPrep DNA/RNA/miRNA Universal Kit (Qiagen, Hilden, Germany) according to manufacturer's instructions. RNA quality was assessed by the 2100 Bioanalyzer System (Agilent, CA, USA) carried out at the High Technology Unit (UAT), at Vall d'Hebron Research Institute (VHIR).

Illumina EPIC methylation assay and data processing
Bisulfite (BS) conversion was performed using EZ-96 DNA Methylation™ Kit (Zymo Research, CA, USA) according to manufacturer's instructions. Five hundred nanograms of BS-converted DNA was hybridized on Infinium MethylationEPIC Bead Chip array (Illumina, Inc., San Diego, CA, USA) following manufacturer's instructions to assess DNA methylation of 850,000 selected CpGs that cover 99% of annotated RefSeq genes. Fluorescence of probes was detected by BeadArray Reader (Illumina, Inc.), and image processing and data extraction were performed as previously described [34]. Downstream data processing and normalization were performed using the R statistical language. Probes were first filtered by detection p value ($p < 0.01$) and normalized by Illumina normalization provided by the *minfi* package. CpGs in single nucleotide polymorphism loci were eliminated. An additional filter of beta values (ratio of DNA methylation) was applied, in which top 5% of CpGs with the highest Δbeta between sample groups were retained for further analyses. ComBat adjustment, provided by the *sva* package, was performed to remove bias from batches. M values ($\log_2$-transformed beta values) were utilized to obtain p value and adjusted p value (Benjamini-Hochberg-calculated FDR) between sample groups by an eBayes-moderated paired t test using the *limma* package [35]. p value of < 0.01 and FDR of < 0.05 were considered statistically significant. Differentially methylated regions (DMRs) were identified using the bumphunter function from *minfi*, in which a p value of < 0.05 and a region containing 2 or more CpGs were considered DMRs. Raw DNA methylation dataset is available at GEO with accession number GSE146093 [36].

Clariom S gene expression array and data processing
One hundred nanograms of excellent quality RNA (RNA integrity number of > 9) was hybridized on Clariom™ S array, carried out at the High Technology Unit (UAT), at Vall d'Hebron Research Institute (VHIR), which interrogates the expression of $> 20,000$ transcripts. Data processing and normalization were carried out using the R statistical language. Background correction was performed using Robust Microarray Analysis (RMA) normalization provided by *oligo* package [37]. Annotation of probes was performed using *clariomshumantranscriptcluster.db* package [38], and the average expression level was calculated for probes mapped to the same gene. *ComBat* adjustment, provided by the *sva* package, was performed to remove bias from batches and other confounding variables. For comparisons between groups, the *limma* package was used to perform an eBayes-moderated paired t test provided in order to obtain $\log_2$ fold change ($\log_2$FC), p value, and adjusted p value (Benjamini-Hochberg-calculated FDR). Genes that displayed statistically significant tests (p value < 0.01 and FDR < 0.05) were considered differentially expressed. Deconvolution analysis was performed using the ABsolute Immune Signal (ABIS) deconvolution online tool [39]. Raw gene expression dataset is available at GEO with accession number GSE146093 [36].

Gene ontology, motif, and regulon enrichment analyses
Gene ontology (GO) analysis of differentially methylated CpG positions (DMPs) and regions (DMRs) was performed using the GREAT online tool ((http://great.stanford.edu/public/html) [40], in which genomic regions were annotated by applying the basal plus extension settings. For DMPs, annotated CpGs in the EPIC array were used as background, and for DMRs, default background was used. GO terms with p value < 0.01 and fold change (FC) > 2 were considered significantly enriched. GO analysis of differentially expressed genes (DEGs) was carried out using the online tool DAVID (https://david.ncifcrf.gov) under Functional Annotation settings, in which annotated genes in the Clariom™ S array were

used as background. GO categories with p value of < 0.01 were considered significantly enriched.

Motif enrichment analysis of DMPs was performed using the findMotifsGenome.pl tool provided by the HOMER motif discovery software [41]. A window of ± 250 bp surrounding each DMP was applied, and CpGs annotated in the EPIC array were used as background. Transcription factor (TF) enrichment of DEGs was carried out using the DoRothEA (Discriminant Regulon Expression Analysis) v2 tool [42]. Regulons with confidence score of A–C were utilized for analysis, and a p value of < 0.05 and normalized enrichment score (NES) of ± 2 were considered significantly enriched.

Comparative analysis with public ChIP-seq datasets

DNase I hypersensitivity and ChIP-seq data of histone modifications H3K27ac, H3K4me1, H3K4me3, H3K27me3, H3K36me3, and H3K9me3 of total CD4+ T cells were downloaded from the BLUEPRINT portal (http://dcc.blue-print-epigenome.eu/) [43]. Five independent datasets were downloaded for each histone mark, and ChIP-seq peaks were consolidated using the MSPC program [44], in which peaks were first filtered by q value < 0.01 and FC ≥ 2 and using the parameters -w = 1E^{-4}, -s =1E^{-8}, and -c = 3.

ChIP-seq datasets of transcription factors were downloaded from ReMap database (http://pedagogix-tagc.univ-mrs.fr/remap) [45]. ChIP-seq data of STAT1, RUNX1, NFKB2, MYC, JUN, CTCFL, and CTCF were downloaded as merged peaks of all available data, whereas RELA data was generated in CD4+ T cells. Background of DMRs was generated from annotated CpGs from EPIC array by randomly permutating 1000 times, and an average p value of < 0.01 and average odds ratio > 1 were considered significantly enriched.

Methylation-expression quantitative trait loci and promoter capture Hi-C data analyses

Methylation-expression quantitative trait loci (meQTL) analysis to link varying gene expression to aberrant DNA methylation was carried out utilizing the *MatrixEQTL* package, including sex and age as covariates [46]. A window of 5 Mb was applied for *cis* interactions, and a Pearson correlation p value of < 0.01 was considered significant. CpG-gene interactions were then filtered by differential methylation and expression in SSc T cells compared to controls (FDR < 0.05 and p value < 0.01).

Promoter capture Hi-C (PCHi-C) datasets were generated previously in naïve, non-activated, and activated CD4+ T cells from healthy controls [47]. Briefly, cells are fixed by paraformaldehyde, lysed, and subjected to Hind III digestion. Restriction fragments (median size of 5 kb) were then biotinylated, and interacting fragments were ligased. "Captured fragments were then used to generate Hi-C libraries where fragments were mapped to

either a gene promoter or intergenic region [48]. Utilizing significant PCHi-C interactions, we first performed overlap between DEGs and mapped promoters from PCHi-C datasets. Second, we utilized pipelines provided by *GenomicRanges* package [49] to overlap the interacting PCHi-C fragments with DMP coordinates. Finally, we eliminated the DMPs found in gene promoters, in which 212 unique DMP-DEG interactions remained. Confirmed interactions were visualized using WashU Epigenome Browser (http://epigenomegateway.wustl.edu/legacy).

Association analysis and genotyping of risk variants

Genome-wide association study (GWAS) data was obtained from López-Isac et al. [15], in which SSc-associated susceptibility loci and their interacting genes were overlapped, by gene name, with DMP-DEG pairs identified from this study. Genome-wide genotyping was performed utilizing purified DNA obtained from peripheral blood mononuclear cells of our cohort of 48 SSc patients using standard methods and hybridized on Illumina GWAS platforms (HumanCytoSNP-12v2 and Illumina HumanCore) according to the manufacturer's instructions. Stringent quality controls were applied using PLINK [50]. Detailed information is described in López-Isac et al. [15].

Results

DNA methylation deregulation in SSc CD4 T cells

To gain insights into functional and molecular alterations of T lymphocytes in the context of SSc pathogenesis, we first isolated CD4+ T lymphocytes from the PBMCs of 48 SSc patients and 16 age- and sex-matched healthy donors (HD) by CD4+ positive cell sorting (Fig. 1a, b; Additional file 1: Table S1). To interrogate DNA methylation, we utilized bead arrays (see the "Methods" section). Subsequently, 9112 CpGs were found to be differentially methylated in SSc CD4+ T cells compared to HD (FDR < 0.05 and p value < 0.01), in which 7837 and 1275 CpGs were hyper- and hypo-methylated, respectively (Fig. 1c; Additional file 2: Table S2). Analysis of differentially methylated CpG positions (DMPs) was visualized by t-distributed stochastic neighbor embedding (t-SNE) analysis (Fig. 1d), and HD and SSc samples can be observed to separate along the t-SNE2 axis. DMPs were predominantly situated in open sea and intergenic regions (Additional file 3: Fig. S1A). Gene ontology analysis revealed relevant categories for both hyper- and hypomethylated DMPs (Fig. 1e). Hyper-methylated DMPs were enriched in inflammatory pathways, including IL-23 and IL-18 receptor activity and TLR3 signaling pathway, as well as pathways involved in T cell biology, such as memory and Th17 T cell differentiation (Fig. 1e, upper panel; Additional file 4: Table S3).
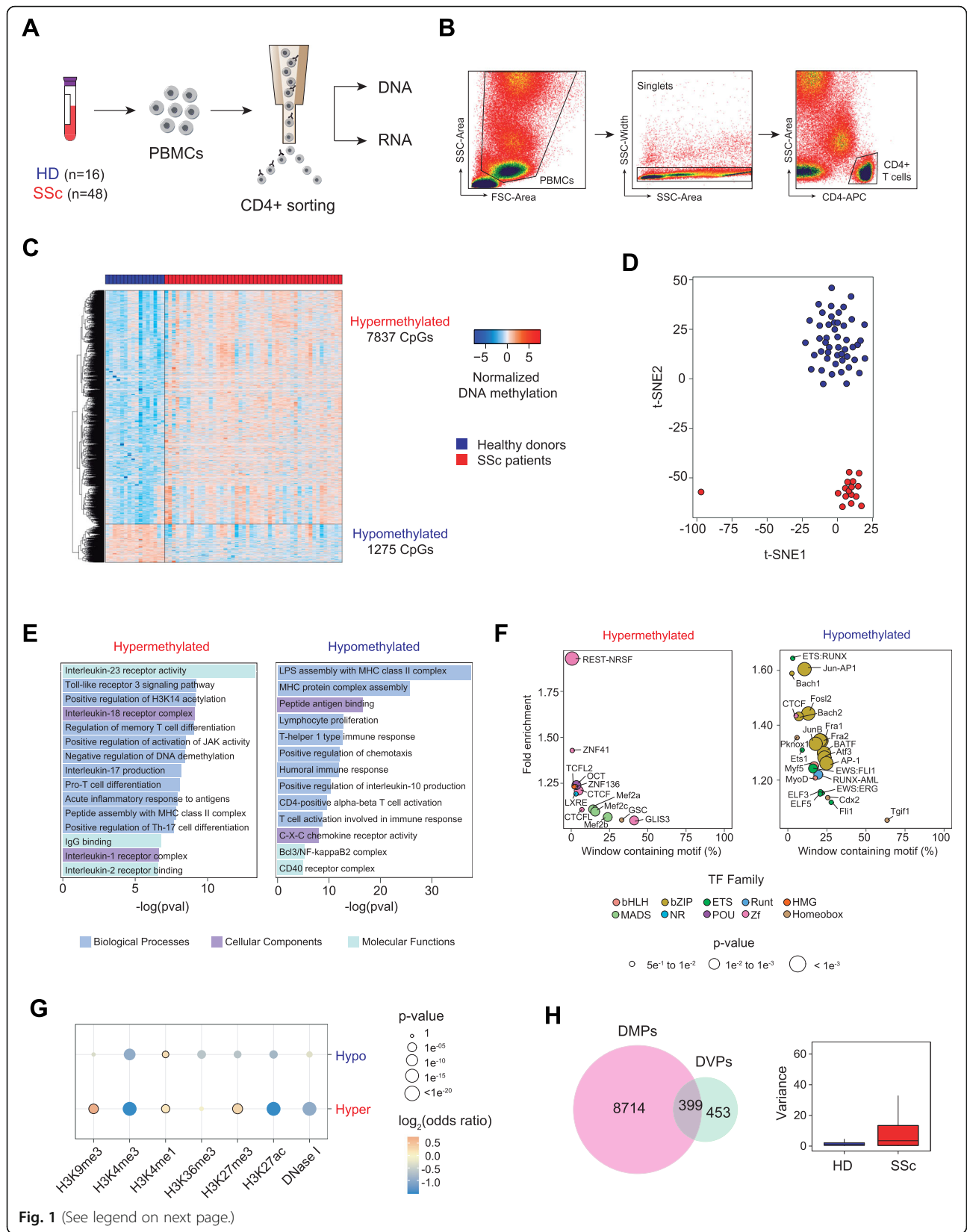
Fig. 1 (See legend on next page.)

Fig. 1 Aberrant DNA methylation in inflammatory loci of SSc CD4+ T cells. **a** Scheme depicting workflow including PBMC isolation, CD4+ T cell positive sorting, and DNA and RNA extraction. **b** Gating strategy to eliminate cell debris, doublets, and the isolation of CD4-APC+ T lymphocytes by side scatter. **c** Heatmap of differential methylated CpGs (DMPs), in which 7837 CpGs were hypermethylated (log2FC > 0, FDR < 0.05) and 1275 CpGs were hypomethylated (log2FC < 0, FDR < 0.05) in SSc patients ($n = 48$) compared to healthy control ($n = 16$). **d** t-SNE clustering of SSc and HD samples based on all identified DMPs. **e** Gene ontology of hyper- and hypomethylated DMPs, analyzed utilizing the GREAT online tool (http://great.stanford.edu/), in which CpGs annotated in the EPIC array were used as background. **f** HOMER motif enrichment of hyper- and hypomethylated CpGs, utilizing CpGs annotated in the EPIC array as background. A window of ± 250 bp centering around each DMP was applied. **g** Enrichment of DNase I hypersensitivity, H3K9me3, H3K4me3, H3K4me1, H3K36me3, H3K27me3, and H3K27ac ChIP-seq data, obtained from the BLUEPRINT portal, in hyper- and hypomethylated DMPs. Highlighted circles represent statistically significant comparisons (p value < 0.01 and odds ratio > 1) compared to background. **h** Overlap between DMPs and differentially variable CpGs (DVPs), identified utilizing the iEVORA algorithm (left), and graphical representation of variance of identified DVPs in HD and SSc

On the other hand, hypomethylated DMPs were predominantly enriched in genes encoding the MHC protein complex (Fig. 1e, lower panel; Additional file 3: Fig. S1B; Additional file 4: Table S3). Other relevant categories, including proliferation, Th1-type immune response, and IL-10 production, were also enriched in hypomethylated DMPs. Motif enrichment analyses revealed zinc finger transcription factor CTCF to be common to both hyper- and hypomethylated DMPs (Fig. 1f). This observation is especially interesting given the importance of CTCF in mediating enhancer-gene interactions [51]; hence, aberrant DNA methylation of CTCF binding enhancer regions may affect the expression of interacting genes. Furthermore, other motifs of TF complexes important to T cell biology, such as the BAFT-JUN-AP1 complex, shown to play a role in CD4+ T cell differentiation [52], and TFs of the ETS family, whose deletion in CD4+ T cells result in autoimmunity in mice [53], were enriched in the hypomethylated cluster (Fig. 1f). Subsequent analyses revealed that hypermethylated DMPs showed a particular histone mark signature that was enriched in H3K9me3, H3K27me3, and H3K4me1 (Fig. 1g), which suggest the enrichment of poised enhancers, as was previously defined by Zentner et al. [54]. Conversely, hypomethylated DMPs were significantly enriched in H3K4me1 only, which is a hallmark of primed enhancers (Fig. 1g) [55]. Finally, among the DNA methylation alterations detected in SSc patients, we also observed 852 CpGs that changed their variance compared to healthy controls, and these are termed differentially variable positions (DVPs; Fig. 1h). Many of these CpGs were also differentially methylated; however, 453 CpGs were exclusively DVPs, and majority of them experienced an increase in variance in SSc patients (Fig. 1h). Increases in variance may be a consequence of differences in pathological evolution of the disease in the patient population. Overall, deregulation in DNA methylation of genes important to immune response and T cell biology was detected in peripheral blood-isolated CD4+ T cells. Given the possibility that pharmacological treatments may modulate DNA methylation in CD4+ T cells, we first stratified patients depending on whether

they were taking immunosuppressors or vasodilators at the time of sample collection. No significant DMPs (FDR > 0.05) were identified when comparing treated and untreated patients. Second, we evaluated whether treatment may be a confounding variable in identified SSc-associated DMPs, and observed no significant associations (Wilcoxon p value > 0.05) between treatment and DNA methylation (Additional file 3: Fig. S1C). Hence, DMPs identified were specifically associated with SSc disease. SSc patients were then stratified as diffuse (dcSSc), limited (lcSSc), or sine scleroderma (ssSSc), with the latter having no skin fibrosis. Early scleroderma (eSSc) was excluded from the analysis due to the very limited number of patients in this group. We analyzed aberrant DNA methylation of SSc subgroups and observed that many of the alterations (35%) were shared between at least two of the three subgroups (Additional file 3: Fig. S1C). Furthermore, subgroup-specific DMPs also displayed some degree of aberrancy in other SSc subgroups that did not reach statistical significance (Additional file 3: Fig. S1D), suggesting that the majority of alterations are shared among groups. Nevertheless, the lack of significant differences observed between SSc subtypes may be due to a lack of statistical power as a result of the small number of patient samples in each group. Altogether, despite the small cohort of patients, our results suggest that aberrant DNA methylation of CD4+ T cells may be a common hallmark of all subgroups of SSc.

Differentially methylated regions enriched in NF-κB signaling pathway

Although the identification of single DMPs can hint at possible alterations in the genomic landscape, differentially methylated regions (DMRs) may give functional relevance to DNA methylation, as DMRs have been described to correlate well with transcription, chromatin features, and phenotypic outcomes [56, 57]. Accordingly, we identified 1082 significant DMRs, in which 212 and 870 DMRs were hypo- and hypermethylated, respectively. More than 25% of DMRs were found to be situated in promoters of the nearest gene, and the same

proportion of DMRs were found more than 50 kb from TSS (Additional file 5: Fig. S2A). Many of the identified DMRs were mapped to such relevant genes as *COLEC11*, *GSTM1*, *HLA-C*, and *IL15RA* (Fig. 2a). Gene ontology analyses revealed that hypermethylated DMRs were significantly enriched in various categories relevant to immune functions, IL-10 secretion, Th2 cytokine production and differentiation, NLRP3 inflammasome

complex, and negative regulation of NF-κB signaling (Fig. 2b; Additional file 6. Table S4). On the other hand, hypomethylated DMRs were enriched in peptide antigen binding, MHC protein complex, IL-1 binding, and chemotaxis (Fig. 2b; Additional file 6: Table S4). Furthermore, both hyper- and hypomethylated DMRs overlapped significantly with H3K4me1 histone mark and DNase I hypersentivity regions, indicating the presence
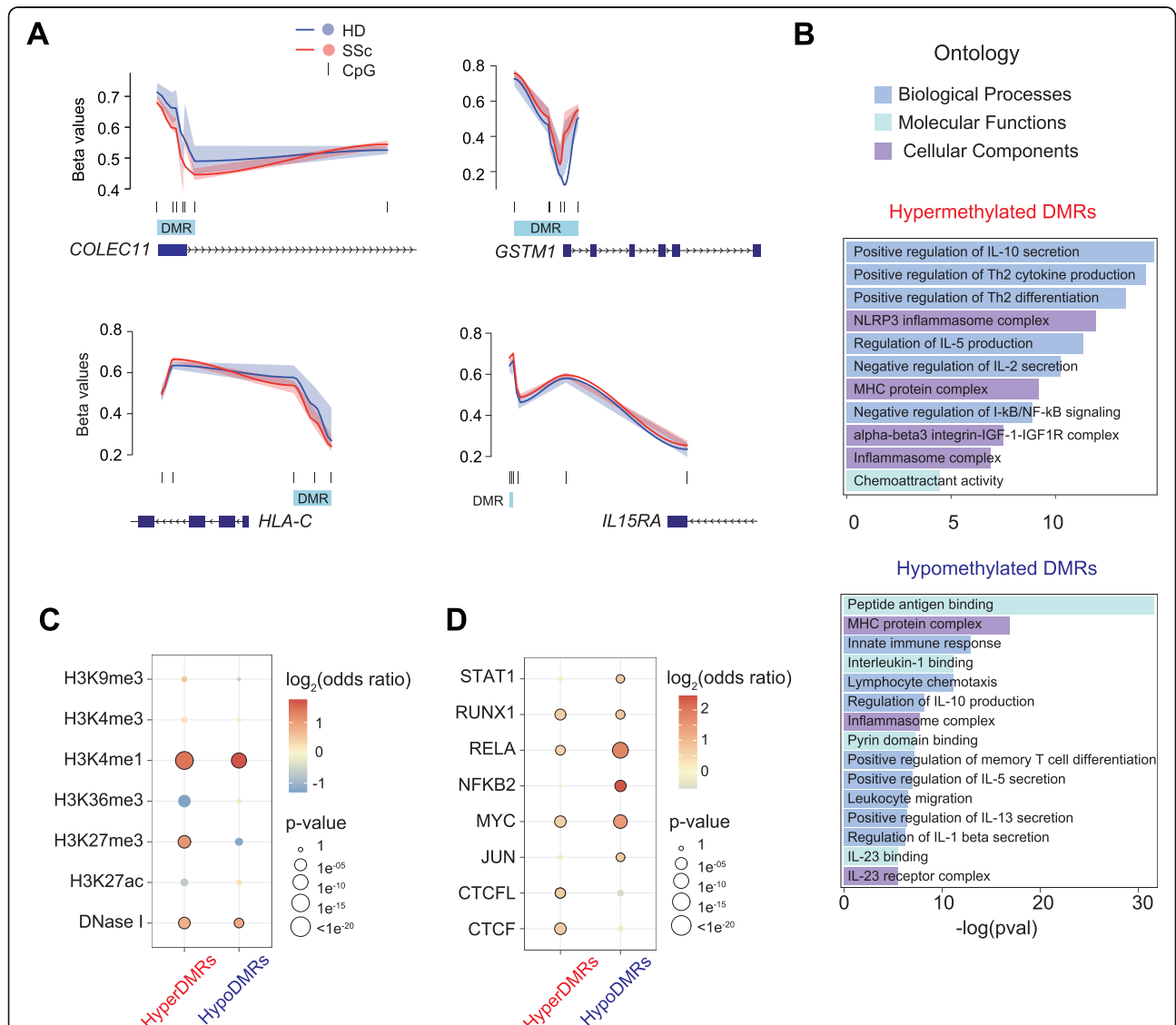

Fig. 2 Differentially methylated regions (DMRs) enrich in inflammatory loci and pro-inflammatory transcription factors. **a** Graphical representation of beta values of identified DMRs, utilizing *bumphunter* tool, mapped to relevant genes, including *COLEC11*, *GSTM1*, *HLA-C*, and *IL15RA*. Identified DMRs are highlighted in blue, and CpGs are depicted in black. Lines represent locally estimated scatterplot smoothing, and transparent areas represent confidence intervals. **b** GO analysis of hyper- and hypomethylated DMRs utilizing the GREAT online tool. DMRs were mapped to the nearest gene. **c** Enrichment of DNase I hypersensitivity, H3K9me3, H3K4me3, H3K4me1, H3K36me3, H3K27me3, and H3K27ac ChIP-seq data, obtained from the BLUEPRINT portal, in hyper- and hypomethylated DMRs. **d** Enrichment of TF ChIP-seq peaks in hyper- and hypomethylated DMRs. ChIP-seq data were downloaded from ReMap database, in which STAT1, RUNX1, NFKB2, MYC, JUN, CTCFL, and CTCF were downloaded as consolidated ChIP-seq peaks of all available datasets, whereas RELA ChIP-seq peaks were obtained from CD4+ T cells. **c**, **d** Background of the same length as DMRs was generated from EPIC array, in which 1000 resamplings were applied. *p* values and odds ratios are averages of 1000 permutations. Highlighted circles represent statistically significant comparisons (*p* value < 0.01 and odds ratio > 1) compared to background

Li *et al. Genome Medicine*   (2020) 12:81

Page 8 of 17

of enhancers (Fig. 2c). Additionally, hypermethylated DMRs were enriched in H3K27me3, which, together with H3K4me1, is a key mark for poised enhancers. Given the importance of CTCF in mediating long-range enhancer-DNA interactions and the enrichment of its motif in both hyper- and hypomethylated DMPs, we performed overlap between DMRs and CTCF and CTCF L ChIP-seq peaks obtained from ENCODE. We observed that hyperDMRs significantly overlapped with both CTCF and CTCFL binding regions (Fig. 2d). Moreover, significant overlap was also detected between both hyper- and hypoDMRs and binding sites of RUNX1, MYC, and RELA, all of which have important roles in T cell biology, suggesting that aberrant DNA methylation in SSc may alter their binding to these regions. Finally, hypoDMRs were selectively enriched in binding sites of NFKB2 and JUN (Fig. 2d).

Identification of aberrant gene expression in CD4 T cells of SSc patients

To characterize gene expression aberrancies in CD4+ T cells that drive disease pathogenesis of SSc, we performed genome-wide RNA expression analysis and found that a total 3929 genes displayed differential expression (differentially expressed genes (DEGs)) between HD and SSc, of which 1949 and 1980 were down- and upregulated, respectively (Fig. 3a, b; Additional file 7: Table S5). To identify the specific pathways that were altered, we performed DAVID gene ontology analysis and observed that downregulated genes were enriched in such signaling pathways as T cell receptor, IL-2 production, Fc-γ receptor, and interferon-gamma, as well as genes associated with systemic lupus erythematosus and viral carcinogenesis (Fig. 3c). Conversely, genes encoding proteins that propagate signaling pathways such as TNF, NF-κB, Fc-ε receptor, Wnt, and TLR-9 were upregulated (Fig. 3c). TF enrichment analysis revealed that upregulated DEGs were driven by such TFs as IRF2, NFKB1, FOXP1, and SOX10, and downregulated DEGs were regulated by TFs including EGR1, FOXA1, GATA2/3, and CTCF (Fig. 3d). Furthermore, the expression of several transcripts of the HLA cluster were also differentially expressed in SSc patients (Fig. 3e), which was in accordance with observed aberrant DNA methylation in these genomic regions. Interestingly, genes encoding several transcription factors whose motifs were enriched in aberrant DMPs were also found to be altered, and these genes include *JUN*, *FLI1*, *RUNX1*, and *CTCF* (Fig. 3e). The expression of other relevant TFs, such as *NFKB2*, *EGR1*, and *RELB*, were also dysregulated (Fig. 3e). Following deconvolution analyses of CD4 T cell populations, we observed that there were no significant differences in percentages of naïve and memory T cells between SSc and controls (Additional file 5: Fig. S2B).

Evaluating whether vasodilator and immunosupressor treatments at the time of sample collection may affect gene expression, we performed limma analyses comparing treated and untreated patients. We only observed one gene, *RARA*, that was differentially expressed (FDR < 0.05) between patients treated with immunosuppressors and untreated patients, while no differences in gene expression were observed in patients on vasodilators. Furthermore, correlation analyses between treatments and first three principal components (PCs) of identified DEGs revealed that neither vasodilators nor immunosuppressors affected the first two PCs (Additional file 5: Fig. S2C). Vasodilators correlated significantly with gene expression of PC3; however, this only accounted for 0.3% of total variance. Hence, we discarded differences in treatments as confounding variables to SSc-associated DEGs.

DNA methylation changes establish short- and long-distance relationships with gene expression in SSc

We then investigated the relationship between DNA methylation and gene expression deregulation in SSc. TF binding can be both negatively or positively influenced by DNA methylation [58, 59]. DNA methylation of gene TSS associates negatively with gene expression [60], whereas methylation of CpGs located within gene body can also associate with active gene expression [61]. To investigate the potential relationship between DNA methylation alterations and gene expression in SSc, we searched for possible meQTLs (methylation-expression quantitative trait loci). Subsequently, we found 45 differentially methylated CpGs (DMPs) that interacted with differentially expressed genes (DEGs), in which the CpG was located near/in the promoter or TSS of the interacting gene (5 kb upstream and 1 kb downstream of TSS), of which 33 (73.3%) were negative interactions (Additional file 8: Table S6; Fig. 4a). GO analysis of interacting DEGs includes pathways involved in nucleotide-excision repair, nucleosome assembly, and positive regulation of interferon-beta production (Fig. 4b). Relevant genes include *ADAM20*, a member of the ADAM family of metalloproteases which are involved in T cell responses [62, 63]; *CD274*, which, upon binding to its receptor, mediates changes in T cell metabolism [64]; and *IRF1*, a key driver of Th1 cell differentiation [65] (Fig. 4c).

Given that SSc-associated differentially methylated CpGs and regions (DMPs and DMRs) were enriched in H3K4me1, a key histone mark characterizing enhancers, it is possible that aberrant expression of genes in SSc patient CD4+ T cells may influence or be influenced by DNA methylation in distal elements. To interrogate long-distance correlative interactions, an extended window of 5 Mb between CpG and gene was used to detect meQTLs followed by extensive multistep filtering
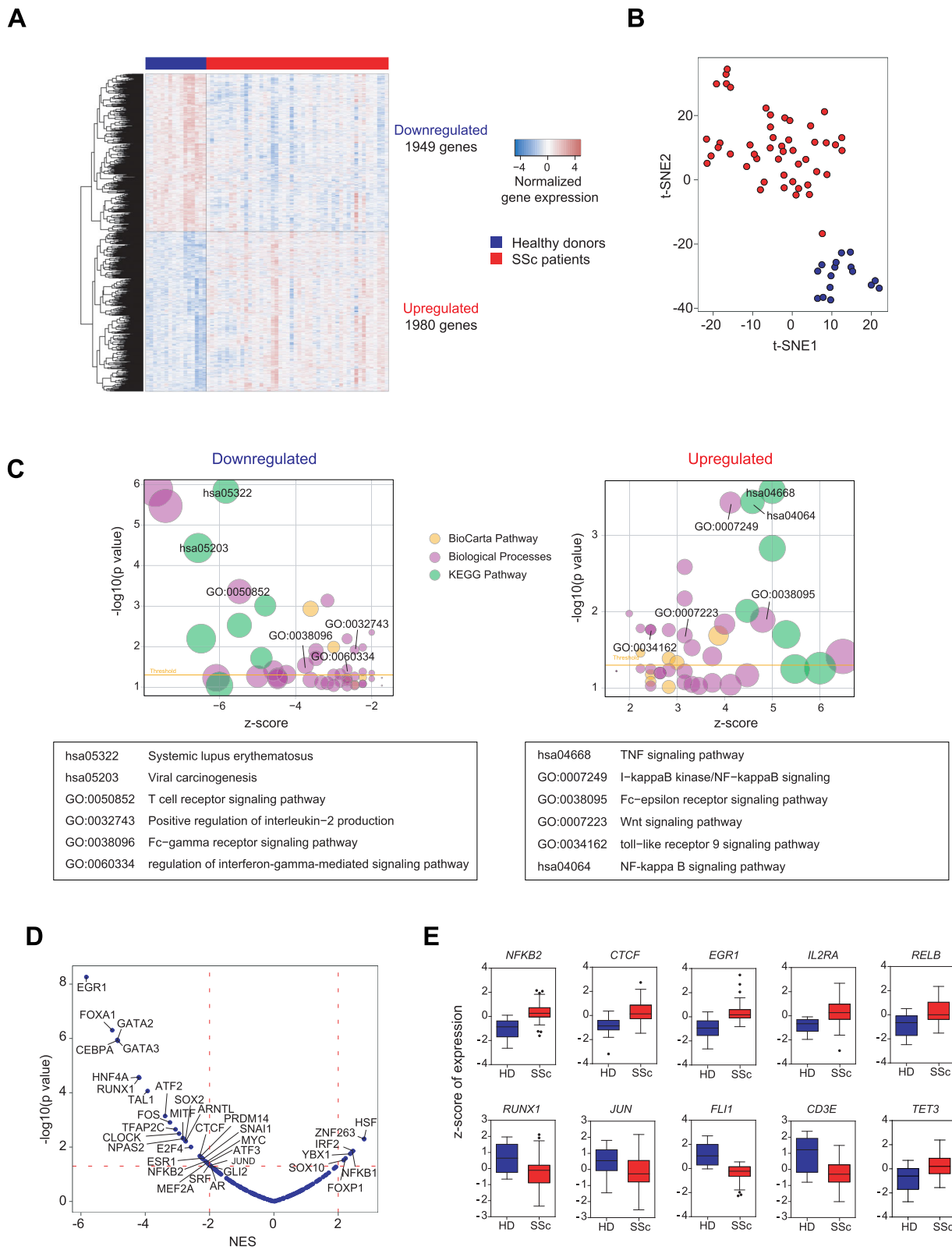
Fig. 3 (See legend on next page.)

(See figure on previous page.)

Fig. 3 Aberrant gene expression of SSc CD4+ T cells include inflammatory genes and relevant transcription factors. **a** Differentially expressed genes (DEGs) comparing RNA expression microarray data of CD4+ T cells isolated from SSc with HD, in which 1949 genes were downregulated ($\log_2$FC < 0, FDR < 0.05) and 1980 genes were upregulated ($\log_2$FC > 0, FDR < 0.05). **b** t-SNE clustering of identified DEGs in HD and SSc CD4 T cell samples. **c** GO analysis of down- and upregulated DEGs performed utilizing the DAVID online tool (https://david.ncifcrf.gov/). Bubble size refers to the size of GO term, and threshold represents the cutoff for statistical significance (*p* value < 0.05). Relevant GO terms are summarized in tables before the bubble plots. **d** TF enrichment analysis, utilizing the DoRothEA tool, of all interrogated genes ordered by normalized enrichment score (NES) of SSc compared to HD. Dotted red lines represent the cutoffs for statistical significance (NES < − 2 and *p* value < 0.01, NES > 2 and *p* value < 0.01). **e** Graphical representations of z-scores of genes relevant to T cell biology, including *CD3E* and *IL2RA*; transcription factors *NFKB2*, *CTCF*, *EGR1*, *RELB*, *RUNX1*, *JUN*, and *FLI1*; and DNA methylcytosine dioxygenase *TET3*

processes based on the causal inference test [66, 67] (Fig. 5a). Firstly, DNA methylation and gene expression correlations between all annotated CpG positions (M) and genes (E) were searched within a window of 5 Mb, in which a total of 99,525 significant (Pearson *p* value < 0.01) CpG-gene interactions were discovered. Secondly, CpG-gene pairs were then filtered by differential expression and methylation in SSc patients (Y) compared to healthy controls to yield 17,500 DMP-DEG pairs (5841 unique DMPs and 3237 unique genes), approximately half (9114 DMP-

DEG pairs) of which were negative correlations. Thirdly, utilizing previously generated promoter capture Hi-C (PCHi-C) data of healthy CD4+ T cells [47], we overlapped promoter-non-promoter interactomes with SSc-associated DMP-DEG pairs by first overlapping DEGs with annotated promoters in PCHi-C and then overlapping correlating DMPs with the interacting non-promoter PCHi-C fragments. Consequently, our analysis yielded 182, 170, and 162 DMP-DEG interactions utilizing datasets from naïve, non-activated, and activated CD4+ T cells, respectively.
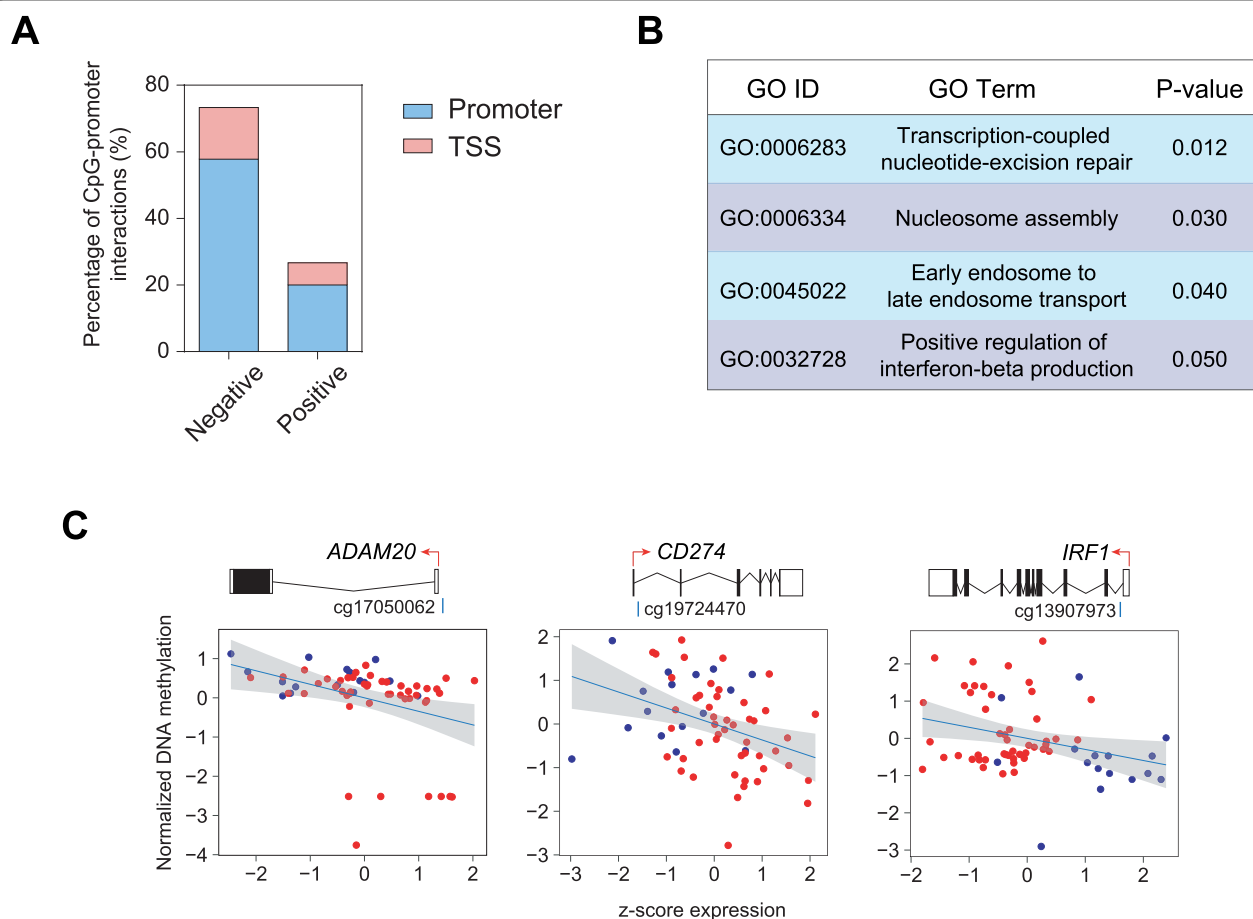

Fig. 4 Correlation of DNA methylation and gene expression of DMPs situated in gene promoters/TSS. **a** Percentage of detected short-range interactions that present either negative or position correlation between DNA methylation and gene expression, in which the DMP is situated in promoter or TSS of interacting DEG. **b** GO analysis utilizing the DAVID online tool of the DEGs with short-range interacting DMPs. **c** Graphical representation of correlation of DNA methylation and gene expression of DMP-DEG pairs. Red and blue dots represent SSc and HD individuals, respectively
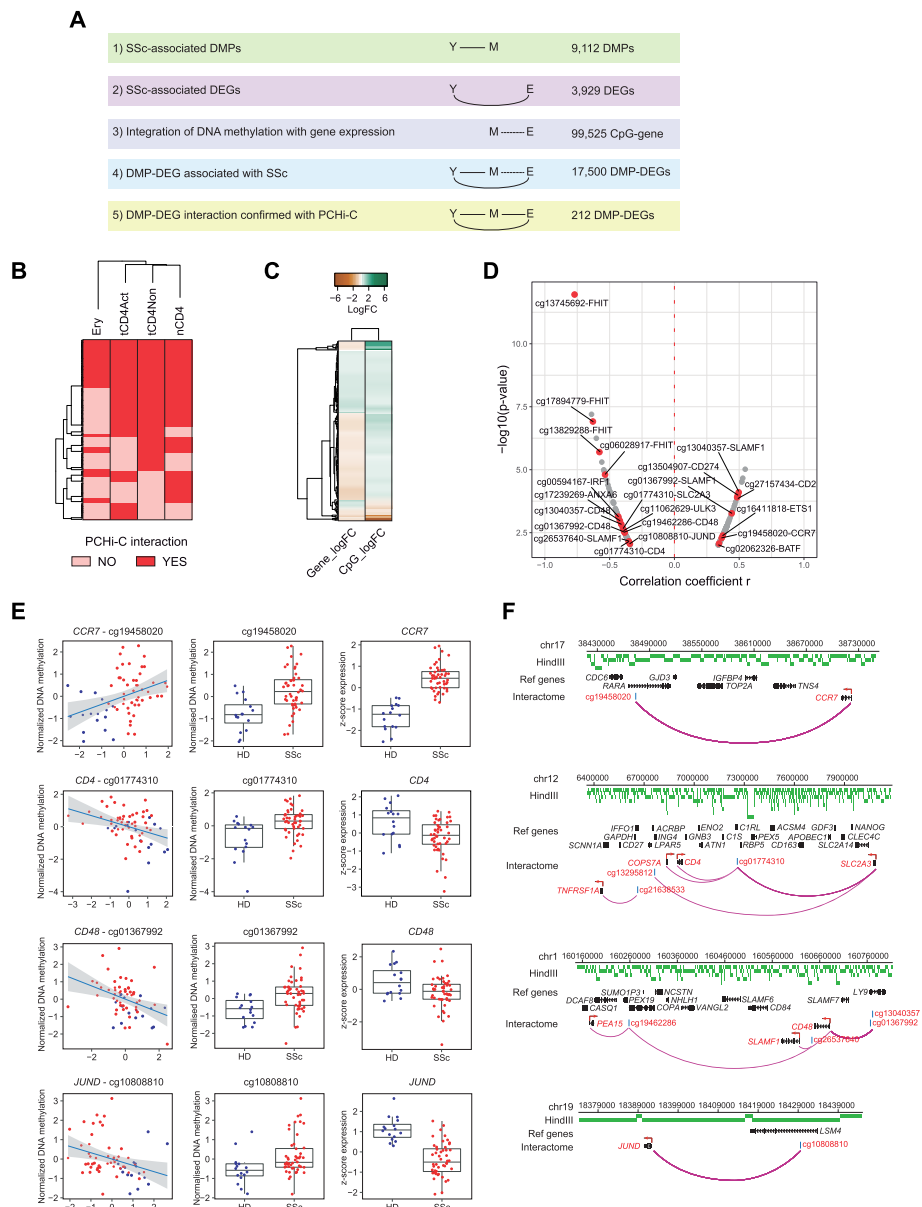
Fig. 5 Long-range interactions identified by promoter capture Hi-C (PCHi-C). **a** Correlations of possible long-range interactions between DMPs and DEGs were detected using the *MatrixEQTL* package by applying a maximum interaction distance of 5 Mb with a Pearson *p* value cutoff of 0.01. Overview summarizing multistep filtering process based on the causal inference test. Y, SSc phenotype; M, identified DMPs; E, identified DEGs; S, SSc-associated susceptibility loci. **b** Heatmap showing the presence (red) or absence (pink) of promoter capture Hi-C (PCHi-C) interactions of 212 identified DMP-DEGs identified utilizing naïve (nCD4), non-activated (tCD4Non), and activated (tCD4Act) PCHi-C datasets. The presence or absence of the same interactions was interrogated in erythroblasts (Ery). **c** Heatmap of log$_2$FC of gene expression and DNA methylation of the 212 confirmed interacting DMP-DEGs. **d** Volcano plot of the 212 confirmed DMP-DEG interactions. Correlation coefficient *r* of DNA methylation and gene expression values is plotted on the *x*-axis against −log$_{10}$(*p* value) on the *y*-axis. Genes relevant to T cell biology are highlighted in red. **e** Graphical representations of DMP-DEG pairs with significant correlation between DNA methylation and gene expression (Pearson *p* value < 0.01). **f** Overlap between DMP-DEG pairs and PCHi-C data was performed to confirm long-range interactions. Arc plot of long-distance interactions between DMPs and DEGs as confirmed by PCHi-C, in which confirmed interactions are depicted as purple arcs. HindIII fragments are depicted in green, and interacting genes and CpGs are highlighted in red

Following consolidation of the three datasets, we detected a total of 212 unique DMP-DEG interactions (Additional file 9: Table S7), in which 128 interactions were shared between all three datasets (Fig. 5b). Furthermore, these interactions appeared to be specific to CD4+ T cells, as comparison with PCHi-C data obtained from erythroblasts displayed little overlap (Fig. 5a). Additionally, of the 212 DMP-DEG interactions, more than half displayed a

negative correlation between gene expression and DNA methylation (Fig. 5c, d). Among these DMP-DEG interactions, aberrantly expressed genes relevant to T cell biology, such as *ANXA6*, *CCR7*, *CD274*, *CD4*, *CD48*, *IRAK2*, *JUND*, and *NFKB2*, were observed to be part of the same PCHi-C interactome with CpGs that displayed differential DNA methylation in SSc patients (Fig. 5d–f).

DMPs and SSc-associated genetic susceptibility loci form part of the same interactome with DEGs

To fully unravel the relevance of DNA methylation in SSc pathogenesis, we can hypothesize that SSc-associated genetic variants may at least partially contribute to aberrant DEG-interacting CpGs. Hence, we utilized a large GWAS dataset generated by López-Isac et al. [15], which included 26,679 individuals that identified 27 SSc-associated risk loci (S) physically interacting with 43 target genes, as validated by HiChIP data in CD4+ T cells. We therefore interrogated the presence of interacting SNPs in association with identified DMPs and DEGs. We first overlapped the SNP-interacting genomic regions with identified DEGs and observed that 36 SNP-interacting genes identified by López-Isac et al. were aberrantly expressed in our cohort of SSc (Fig. 6a). Second, of these 36 DEGs, 5 of them correlated with a SSc-associated DMP that formed part of the same promoter-non-promoter interactome, as observed by PCHi-C. Finally, we observed that 4 of the identified candidate SNP-DEGs further interacted with the gene in which the associated DMP was located according to HiChIP data by López-Isac et al. [15] (Fig. 6b; Additional file 10: Fig. S3A). Specifically, the SSc-associated susceptibility loci *TNIP1* (rs3792783), *GSDMB* (rs9303277), *IL12RB1* (rs2305743), and *CSK* (rs1378942) were possible candidates that interacted with both DMPs, situated in *TNIP1*, *RARA*, *LSM4*, and *MPI*, respectively, and their associated DEGs, namely *ANXA6*, *CCR7*, *JUND*, and *ULK3*, respectively (Fig. 6b). Furthermore, we observed that the vicinity of the identified SNPs and DMPs was particularly enriched in H3K4me1, a classical mark for enhancers (Additional file 10: Fig. S3B).

Further genotyping of the SSc cohort was performed, and gene expression and DNA methylation of DEGs and DMPs were evaluated in patients harboring risk variants compared to patients without risk variants. We observed a clear correlation between gene expression and DNA methylation with the presence of risk alleles for *ANXA6* and cg17239269 associated with SNP rs3792783 (*TNIP1*). For rs9303277 (*GSDMB*), only DNA methylation correlated with risk allele presence, whereas DNA methylation and gene expression associated with SNPs rs2305743 (*IL12RB1*) and rs1378942 (*CSK*) were not found to correlate well with the presence of risk variants (Additional file 10: Fig. S3C).

Furthermore, visualizing ChIP-seq data of CTCF and p65, obtained from the ReMap database, we observed that the SNP-DMP-DEG interaction regions were extensively covered by CTCF binding sites. Interestingly, we observed that p65 binding is predominantly restricted to the promoters/TSS of DEGs *ANXA6*, *CCR7*, and *JUND*. Furthermore, some ChIP-seq peaks were detected near the interacting CpGs and SNPs, compatible with a potential role for p65 in the transcriptional regulation of these genes following correct chromatin looping between genomic loci (Fig. 6b).

Discussion

In this study, the integrated analysis of DNA methylation, gene expression, promoter capture Hi-C, and genetic data potentially unveils novel functional relationships in CD4+ T cells of patients with systemic sclerosis. First, we report DNA methylation and gene expression alterations in SSc CD4+ T lymphocytes associated with essential pathways implicated in T cell differentiation and function. Aberrant DNA methylation in distinct loci could be directly influencing aberrant gene expression through long-distance interactions that involve CTCF. Finally, we identified four important SSc-associated susceptibility loci, *TNIP1* (rs3792783), *GSDMB* (rs9303277), *IL12RB1* (rs2305743), and *CSK* (rs1378942), that form part of the same interactomes with cg17239269-*ANXA6*, cg19458020-*CCR7*, cg10808810-*JUND*, and cg11062629-*ULK3*, respectively.

Despite the increasing number of genetic variants associated with SSc, their functional relevance is still a challenge. In addition, alongside genetic predisposition, environmental factors and epigenetic deregulation also contribute to the pathogenesis of this disease (reviewed in [68]). Our comprehensive approach has unveiled a high number of DNA methylation and expression changes in SSc CD4+ T cells compared to control (9112 DMPs, 1082 DMRs, and 3929 DEGs). The vast majority of DNA methylation changes were SSc-associated hypermethylation. Conversely, only around 10% of changes corresponded to aberrant hypomethylation, which is perhaps associated with the increased gene expression of *TET3*, as detected by our RNA expression analysis. TET3 has been previously linked to T cell differentiation, and its deletion resulted in decreased proportions of progenitor and naïve CD4+ T cells [69]. Therefore, we cannot discard the possibility that changes in DNA methylation may be a result of alterations in CD4+ T cell populations. However, deconvolution analysis showed no significant differences in the proportions of memory and naïve T cell populations between SSc and controls.

Analysis of functional categories of the identified DMPs, DMRs, and DEGs revealed the enrichment of numerous relevant pathways, in which many were in accordance to previous studies, including circadian signaling, Rho protein signaling, thyroid hormone secretion, T cell activation, and cytokine-mediated responses [70, 71]; however,
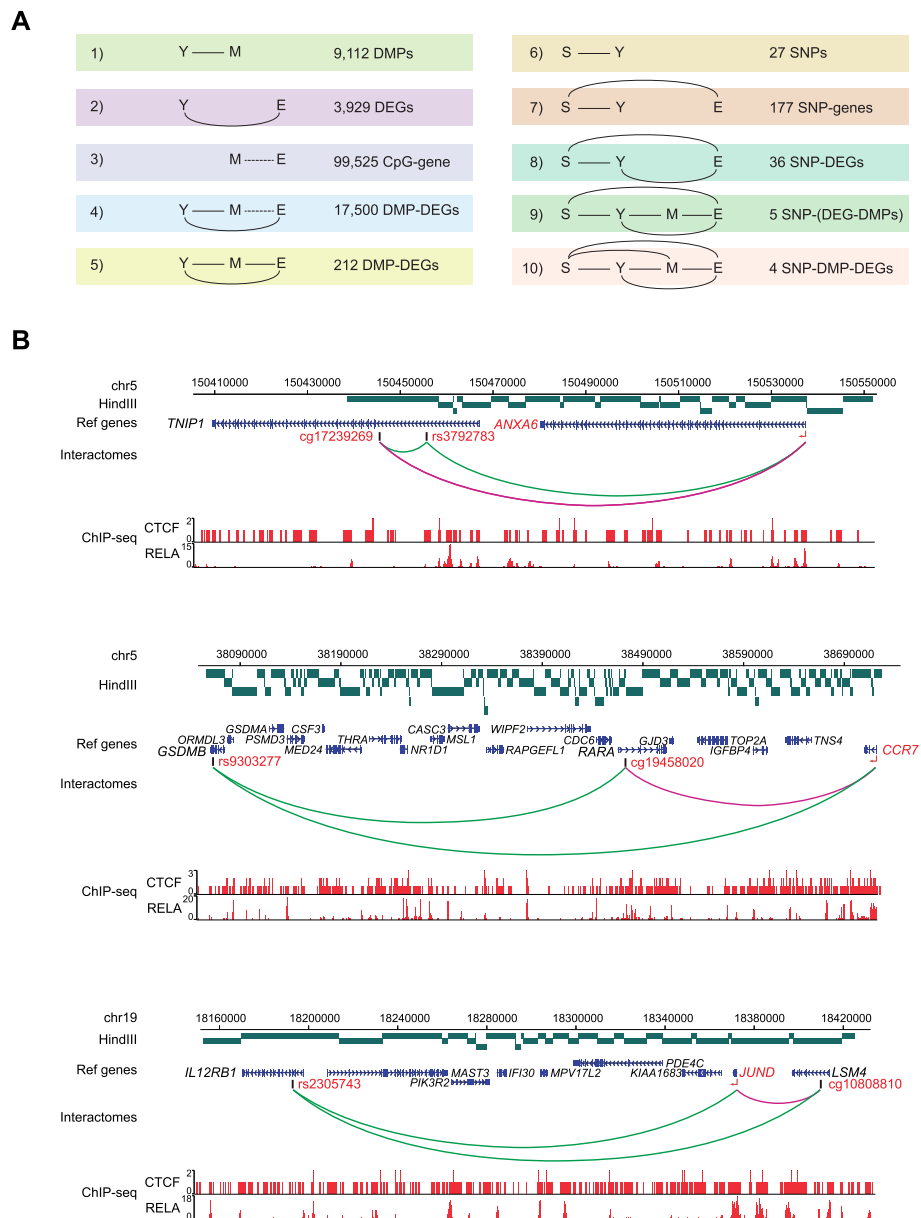
Fig. 6 Presence of SSc-associated genetic variants associated with PCHi-C-identified DMP-DEG interactomes. **a** Multistep filtering process based on the causal inference test extended from Fig. 5, summarizing the filtering steps adopted to identified four interactomes involving SSc-associated SNPs, DMPs, and DEGs. Y, SSc phenotype; M, identified DMPs; E, identified DEGs; S, SSc-associated susceptibility loci. **b** Arcs in green represent interactions of SSc-associated SNPs with distant genes, as confirmed by López-Isac et al. [15], and arcs in purple represent DMP-DEG interactions identified and confirmed in this study. HindIII fragments are depicted in green, and interacting genes and CpGs are highlighted in red. Density plots represent ChIP-seq data of RELA and CTCF. ChIP-seq data were downloaded from ReMap database, in which CTCF were downloaded as consolidated ChIP-seq peaks of all available datasets, whereas RELA ChIP-seq peaks were obtained from CD4+ T cells

we were able to identify several novel pathways, including several cytokine receptor-propagated pathways, MHC complex assembly, T cell polarization, and NF-κB signaling pathway, among others. Interestingly, several of these deregulated pathways were enriched in both hyper- and hypomethylated DMP/DMR datasets. Although we do not know how aberrant DNA methylation affects the activation of these pathways, nonetheless, alterations in these

loci may have implications their correct functions. One such example is the HLA loci, which have been previously described to associate with SSc disease by genetic studies [7, 72].

The relationship between DNA methylation and gene expression is complex, and there are studies describing DNA methylation as both a cause and a consequence of gene expression. Furthermore, DNA methylation

patterns are highly tissue-specific and are established during dynamic differentiation events by site-specific re-modeling at regulatory regions [73]. Nevertheless, methylation of CpGs located in gene promoter, first exon, and intron robustly correlates to gene expression in an inverse manner [74–76]. First, we identified 45 DMPs located within or near the promoter or TSS of DEGs with statistically significant correlations, with majority displaying negative correlations. Hence, aberrant DNA methylation observed in SSc patients may directly cause aberrant gene expression in CD4+ T cells when located in/near promoter or TSS regions. Second, gene expression deregulation of several TFs was observed in SSc CD4+ T cells, including *JUN*, *CTCF*, *FLI1*, and *RUNX1*, whose motifs were enriched in identified DMPs. Therefore, it is also plausible that aberrant gene expression of TFs may mediate altered recruitment to their binding sites, which may consequently shape DNA methylation patterns at these sites. Several TFs have been shown to bind unmethylated regions to block de novo methylation, and one such factor is CTCF, in which it acts as a boundary element to directly impede methylation of regulatory regions [77, 78]. Conversely, other TFs were observed to actively recruit DNA (de) methylation enzymes. One such example is PU.1, which has been shown to recruit both TET2 and DNMT3A to promote DNA demethylation and methylation, respectively [79].

Within the last decade, it has become increasingly clear the existence of long-range looping interactions between regulatory elements and promoters, in which only ~ 7% of all looping interactions are with the nearest gene [80]. These interactions do not only physically exist, but play essential molecular roles in regulating distant gene expression in both biological and pathological settings [47, 81]. Long-range integration of DNA methylation and gene expression has already been explored in other disease contexts, in which one study involving a large cohort of colon cancer patients showed that methylation of distal CpGs controlled genes at a distance of > 1 Mb [82]. Furthermore, several studies have identified essential long-range associations between susceptibility loci with gene expression (GWAS) or with DNA methylation (EWAS) mediating several autoimmune diseases including multiple sclerosis [83], rheumatoid arthritis [84], and SSc [15]. Our study represents the first to integrate genetic risk with epigenome and transcriptome deregulations within the same interactomes in the context of autoimmune disease. First, we identified four SSc-DMPs, cg17239269 (*TNIP1*), cg19458020 (*RARA*), cg10808810 (*LSM4*), and cg11062629 (*MPI*), whose DNA methylation correlated with the expression of four distant SSc-DEGs, *ANXA6*, *CCR7*, *JUND*, and *ULK3*. Second, these interactions were identified to exist within the same

interactome in healthy individuals by previously published promoter capture Hi-C data from CD4+ T cells [47]. Third, from the previous study by López-Isac and colleagues [15], four SSc risk variants, *TNIP1* (rs3792783), *GSDMB* (rs9303277), *IL12RB1* (rs2305743), and *CSK* (rs1378942), were identified to physically interact with *ANXA6*, *CCR7*, *JUND*, and *ULK3*, respectively, as well as with the genes in which the interacting SSc-DMPs, cg17239269 (*TNIP1*), cg19458020 (*RARA*), cg10808810 (*LSM4*), and cg11062629 (*MPI*), respectively, were located. Further genotyping of our SSc cohort showed that the presence of risk alleles in SNP rs3792783 (*TNIP1*) correlated well with differential DNA methylation and gene expression of associated DMPs and DEGs, namely *ANXA6*, which has been described to be essential to CD4+ T cell proliferation via interleukin-2 signaling [85]. Other evaluated risk variants were not observed to correlate with both associated DNA methylation and gene expression. We cannot disregard that this may be due to the small cohort of patient samples, coupled with the disparity in numbers of patients versus healthy controls. Therefore, it is plausible that the number of patients harboring risk variants was too small to yield conclusive statistical significance, in which a larger cohort is required to validate the correlation between risk variants, DNA methylation, and gene expression. Nevertheless, although we observed significant changes in DNA methylation and gene expression, in SSc CD4+ T cells, we cannot conclusively speculate on the effects these alterations have in regard to chromatin accessibility and structure without further validation in patients. However, previous studies do show a direct correlation between DNA methylation and chromatin states [86–88]; therefore, it is possible that chromatin structure is also altered in SSc CD4+ T cells.

One limitation of this study is we cannot accurately predict whether aberrant DNA methylation is a cause or consequence of changes in gene expression. However, given that DMPs were found to enrich in CTCF binding motifs, which may be a consequence of aberrant upregulation of the *CTCF* gene in SSc CD4+ T cells, and the importance of CTCF in enabling chromatin loop formation [89], it is therefore possible that aberrant DNA methylation deregulates CTCF recruitment, which has been previously described to be DNA methylation-dependent [90], in turn affecting long-range interactions with distant genes to alter their expression. However, further validation in SSc CD4+ T cells would be required to fully determine the role of the formation of these interactomes in mediating SSc disease risk.

Collectively, our results showed the occurrence of widespread DNA methylation and expression alterations in SSc CD4+ T cells, which at least are in part determined through long-distance interactions. Additionally,

Li *et al. Genome Medicine*   (2020) 12:81

Page 15 of 17

we have detected the presence of four causal risk variants which may be associated with aberrant DNA methylation and gene expression. Although we do not directly validate these interactions in our SSc cohort, integrative analyses of GWAS and PCHi-C suggest that these SNPs, DEGs, and DMPs form part of the same interactomes.

Conclusions

In the present study, we have shown that CD4+ T cells from SSc patients display widespread changes in their methylomes and transcriptomes in relation to healthy controls. Many of the changes enrich in functional categories associated with T cell biology and inflammation. We observed significant correlations between DNA methylation and gene expression alterations. In fact, using promoter capture Hi-C data, we observed that numerous DMP-DEG pairs form interactomes in healthy CD4+ T cells. Finally, we show that these alterations appear to be associated with the presence of SSc-susceptibility loci that form part of the same interactomes with DMP-DEG pairs. In summary, our study established a novel link between genetic susceptibility in SSc and DNA methylation-associated transcriptional changes in CD4+ T cells, providing a perspective on the relationship between genetic and epigenetic factors contributing to the aberrant behavior of CD4+ T cells in SSc.

Supplementary information

Supplementary information accompanies this paper at https://doi.org/10.1186/s13073-020-00779-6.

Additional file 1: Table S1. Clinical characteristics of cohort.

Additional file 2: Table S2. Differentially methylated positions (DMPs) between SSc and healthy control CD4+ T cells.

Additional file 3: Figure S1. Additional analyses of DNA methylation datasets.

Additional file 4: Table S3. GREAT GO analyses of hypermethylated and hypomethylated DMPs.

Additional file 5: Figure S2. Additional analyses of DMRs and gene expression datasets.

Additional file 6: Table S4. GREAT GO analyses of hypermethylated and hypomethylated DMRs.

Additional file 7: Table S5. Differential expression (Differentially Expressed Genes: DEGs) between HD and SSc.

Additional file 8: Table S6. Differentially expressed genes (DEGs) that interact with DMPs situated near/in its promoter.

Additional file 9: Table S7. Unique DMP-DEG interactions, as esteimated from promoter capture Hi-C (PCHi-C) data of CD4+ T cells.

Additional file 10: Figure S3. Association of DEG-DMP interactomes with SSc-associated susceptibility alleles.

Availability of data and materials
DNA methylation and expression data for this publication have been deposited in the NCBI Gene Expression Omnibus and are accessible through GEO Series accession number GSE146093 [36].

Ethics approval and consent to participate
The Committee for Human Subjects of the Vall d'Hebron Hospital and Bellvitge Hospital approved the study (PR275/17), which was conducted in accordance with the ethical guidelines of the 1975 Declaration of Helsinki. All samples were in compliance with the written informed consent to participate in the study.

Consent for publication
Written informed consent for publication was provided by the participants.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]Epigenetics and Immune Disease Group, Josep Carreras Research Institute (IJC), 08916 Badalona, Barcelona, Spain. [2]Instituto de Parasitología y Biomedicina López-Neyra, Consejo Superior de Investigaciones Científicas (IPBLN-CSIC), Granada, Spain. [3]3D Chromatin Organization, Josep Carreras Research Institute (IJC), 08916 Badalona, Barcelona, Spain. [4]Unit of Systemic Autoimmunity Diseases, Department of Internal Medicine, Vall d'Hebron Hospital, Barcelona, Spain.

References
1. Denton CP, Khanna D. Systemic sclerosis. Lancet. 2017;390(10103):85–1699.
2. Allanore Y, Simms R, Distler O, Trojanowska M, Pope J, Denton CP, et al. Systemic sclerosis. Nat Rev Dis Prim; 2015;1.
3. Denton CP, Black CM, Korn JH, De Crombrugghe B. Systemic sclerosis: current pathogenetic concepts and future prospects for targeted therapy. Lancet. 1996;347:1453–8.
4. Trapiella-Martínez L, Díaz-López JB, Caminal-Montero L, Tolosa-Vilella C, Guillén-Del Castillo A, Colunga-Argüelles D, et al. Very early and early systemic sclerosis in the Spanish scleroderma Registry (RESCLE) cohort. Autoimmun Rev. 2017;16(8):796–802.
5. Angiolilli C, Marut W, van der Kroef M, Chouri E, Reedquist KA, Radstake TRDJ. New insights into the genetics and epigenetics of systemic sclerosis. Nat Rev Rheumatol. 2018;14(11):657–73.
6. Acosta-Herrera M, López-Isac E, Martín J. Towards a better classification and novel therapies based on the genetics of systemic sclerosis. Curr Rheumatol Rep. 2019;21(9).

7.   Bossini-Castillo L, López-Isac E, Martín J. Immunogenetics of systemic sclerosis: defining heritability, functional variants and shared-autoimmunity pathways. J Autoimmun. 2015;64:53–65.

8.   Radstake TRDJ, Gorlova O, Rueda B, Martin JE, Alizadeh BZ, Palomino-Morales R, et al. Genome-wide association study of systemic sclerosis identifies CD247 as a new susceptibility locus. Nat Genet. 2010;42:426–9.

9.   Allanore Y, Saad M, Dieudé P, Avouac J, Distler JHW, Amouyel P, et al. Genome-wide scan identifies TNIP1, PSORS1C1, and RHOB as novel risk loci for systemic sclerosis. PLoS Genet. 2011;7(7).

10. Gorlova O, Martin JE, Rueda B, Koeleman BPC, Ying J, Teruel M, et al. Identification of novel genetic markers associated with clinical phenotypes of systemic sclerosis through a genome-wide association strategy. PLoS Genet. 2011;7(7).

11. Mayes MD, Bossini-Castillo L, Gorlova O, Martin JE, Zhou X, Chen WV, et al. Immunochip analysis identifies multiple susceptibility loci for systemic sclerosis. Am J Hum Genet. 2014;94:47–61.

12. Márquez A, Kerick M, Zhernakova A, Gutierrez-Achury J, Chen WM, Onengut-Gumuscu S, et al. Meta-analysis of immunochip data of four autoimmune diseases reveals novel single-disease and cross-phenotype associations. Genome Med. 2018;10(1).

13. Acosta-Herrera M, Kerick M, González-Serna D, Wijmenga C, Franke A, Gregersen PK, et al. Genome-wide meta-analysis reveals shared new loci in systemic seropositive rheumatic diseases. Ann Rheum Dis. 2018;78(3).

14. Martin JE, Broen JC, David Carmona F, Teruel M, Simeon CP, Vonk MC, et al. Identification of CSK as a systemic sclerosis genetic risk factor through genome wide association study follow-up. Hum Mol Genet. 2012;21:2825–35.

15. López-Isac E, Acosta-Herrera M, Kerick M, Assassi S, Satpathy AT, Granja J, et al. GWAS for systemic sclerosis identifies multiple risk loci and highlights fibrotic and vasculopathy pathways. Nat Commun. 2019;10:4955.

16. Tsou PS, Sawalha AH. Unfolding the pathogenesis of scleroderma through genomics and epigenomics. J Autoimmun. 2017;83:73–94.

17. Frieri M, Angadi C, Paolano A, Oster N, Blau SP, Yang S, et al. Altered T cell subpopulations and lymphocytes expressing natural killer cell phenotypes in patients with progressive systemic sclerosis. J Allergy Clin Immunol. 1991; 87:773–9.

18. Fiocco U, Rosada M, Cozzi L, Ortolani C, De Silvestro G, Ruffatti A, et al. Early phenotypic activation of circulating helper memory T cells in scleroderma: correlation with disease activity. Ann Rheum Dis. 1993;52:272–7.

19. Yoshizaki A, Yanaba K, Iwata Y, Komura K, Ogawa A, Muroi E, et al. Elevated serum interleukin-27 levels in patients with systemic sclerosis: association with T cell, B cell and fibroblast activation. Ann Rheum Dis. 2011;70:194–200.

20. Krasimirova E, Velikova T, Ivanova-Todorova E, Tumangelova-Yuzeir K, Kalinova D, Boyadzhieva V, et al. Treg/Th17 cell balance and phytohaemagglutinin activation of T lymphocytes in peripheral blood of systemic sclerosis patients. World J Exp Med. 2017;7:84–96.

21. Fenoglio D, Battaglia F, Parodi A, Stringara S, Negrini S, Panico N, et al. Alteration of Th17 and Treg cell subpopulations co-exist in patients affected with systemic sclerosis. Clin Immunol. 2011;139:249–57.

22. Liu X, Gao N, Li M, Xu D, Hou Y, Wang Q, et al. Elevated levels of CD4(+)CD25(+)FoxP3(+) T cells in systemic sclerosis patients contribute to the secretion of IL-17 and immunosuppression dysfunction. PLoS One. 2013; 8:e64531.

23. Cipriani P, Di Benedetto P, Liakouli V, Del Papa B, Di Padova M, Di Ianni M, et al. Mesenchymal stem cells (MSCs) from scleroderma patients (SSc) preserve their immunomodulatory properties although senescent and normally induce T regulatory cells (Tregs) with a functional phenotype: implications for cellular-based therapy. Clin Exp Immunol. 2013;173:195–206.

24. Saigusa R, Asano Y, Nakamura K, Hirabayashi M, Miura S, Yamashita T, et al. Systemic sclerosis dermal fibroblasts suppress Th1 cytokine production via galectin-9 overproduction due to Fli1 deficiency. J Invest Dermatol. 2017; 137:1850–9.

25. Sawalha AH. Epigenetics and T-cell immunity. Autoimmunity. 2008;41(4): 245–52.

26. Phan AT, Goldrath AW, Glass CK. Metabolic and epigenetic coordination of T cell and macrophage immunity. Immunity. Cell Press. 2017;46(5):714–29.

27. Lei W, Luo Y, Yan K, Zhao S, Li Y, Qiu X, et al. Abnormal DNA methylation in CD4+ T cells from patients with systemic lupus erythematosus, systemic sclerosis, and dermatomyositis. Scand J Rheumatol. 2009;38:369–74.

28. Lian X, Xiao R, Hu X, Kanekura T, Jiang H, Li Y, et al. DNA demethylation of CD40L in CD4+ T cells from women with systemic sclerosis: a possible explanation for female susceptibility. Arthritis Rheum. 2012;64:2338–45.

29. Jiang H, Xiao R, Lian X, Kanekura T, Luo Y, Yin Y, et al. Demethylation of TNFSF7 contributes to CD70 overexpression in CD4+ T cells from patients with systemic sclerosis. Clin Immunol. 2012;143:39–44.

30. Wang YY, Wang Q, Sun XH, Liu RZ, Shu Y, Kanekura T, et al. DNA hypermethylation of the forkhead box protein 3 (FOXP3) promoter in CD4+ T cells of patients with systemic sclerosis. Br J Dermatol. 2014;171:39–47.

31. Ding W, Pu W, Wang L, Jiang S, Zhou X, Tu W, et al. Genome-wide DNA methylation analysis in systemic sclerosis reveals hypomethylation of IFN-associated genes in CD4+ and CD8+ T cells. J Invest Dermatol. 2018; 138:1069–77.

32. Van Den Hoogen F, Khanna D, Fransen J, Johnson SR, Baron M, Tyndall A, et al. 2013 classification criteria for systemic sclerosis: an American College of Rheumatology/European League Against Rheumatism collaborative initiative. Ann Rheum Dis. 2013;72:1747–55.

33. Carwile LeRoy E, Black C, Fleischmajer R, Jablonska S, Krieg T, Medsger TA, et al. Scleroderma (systemic sclerosis): classification, subsets and pathogenesis. J Rheumatol. 1988;15:202–5.

34. Garcia-Gomez A, Li T, Kerick M, Català-Moll F, Comet NR, Rodríguez-Ubreva J, et al. TET2- and TDG-mediated changes are required for the acquisition of distinct histone modifications in divergent terminal differentiation of myeloid cells. Nucleic Acids Res. 2017;45:10002–17.

35. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. Limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43:e47.

36. Li T, Ortiz L, Andrés-León E, Ciudad L, Javierre BM., López-Isac E, et al. Epigenomic and transcriptomic analysis of systemic sclerosis CD4+ T cells. Datasets. Gene Expr. Omnibus Ser. GSE146093. 2020 [cited 2020 Sep 8]. Available from: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc= GSE146093.

37. Carvalho BS, Irizarry RA. A framework for oligonucleotide microarray preprocessing. Bioinformatics. 2010;26:2363–7.

38. MacDonald JW. clariomshumantranscriptcluster.db: Affymetrix clariomshuman annotation data (chip clariomshumantranscriptcluster). R package version 8.7.0; 2017.

39. Monaco G, Lee B, Xu W, Zippelius A, Ao J, De Magalh~ P, et al. RNA-seq signatures normalized by mRNA abundance allow absolute deconvolution of human immune cell types. CellReports. 2019;26:1627–1640.e7.

40. McLean CY, Bristor D, Hiller M, Clarke SL, Schaar BT, Lowe CB, et al. GREAT improves functional interpretation of cis-regulatory regions. Nat Biotechnol. 2010;28:495–501.

41. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell. 2010;38:576–89.

42. Garcia-Alonso L, Holland CH, Ibrahim MM, Turei D, Saez-Rodriguez J. Benchmark and integration of resources for the estimation of human transcription factor activities. Genome Res. 2019;29:1363–75.

43. Fernā Ndez JM, De La Torre V, Richardson D, Torrents D, Carrillo E, Pau S, et al. The BLUEPRINT data analysis portal. 2016;.

44. Jalili V, Matteucci M, Masseroli M, Morelli MJ. Using combined evidence from replicates to evaluate ChIP-seq peaks. Bioinformatics. 2015;31:2761–9.

45. Chèneby J, Gheorghe M, Artufel M, Mathelier A, Ballester B. ReMap 2018: an updated atlas of regulatory regions from an integrative analysis of DNA-binding ChIP-seq experiments. Nucleic Acids Res. 2018;46:D267–75.

46. Shabalin AA. Matrix eQTL: ultra fast eQTL analysis via large matrix operations. Bioinformatics. 2012;28:1353–8.

47. Javierre BM, Burren OS, Wilder SP, Kreuzhuber R, Hill SM, Sewitz S, et al. Lineage-specific genome architecture links enhancers and non-coding disease variants to target gene promoters. Cell. 2016;167:1369–84 e19.

48. Schoenfelder S, Javierre BM, Furlan-Magaril M, Wingett SW, Fraser P. Promoter capture Hi-C: High-resolution, genome-wide profiling of promoter interactions. J Vis Exp; 2018;2018(136).

49. Lawrence M, Huber W, Pagès H, Aboyoun P, Carlson M, Gentleman R, et al. Software for computing and annotating genomic ranges. Prlic A, editor. PLoS Comput Biol. 2013;9:e1003118.

50. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet Cell Press. 2007;81:559–75.

51. Kim S, Yu NK, Kaang BK. CTCF as a multifunctional protein in genome regulation and gene expression. Exp Mol Med. 2015;47:e166.

52. Li P, Spolski R, Liao W, Wang L, Murphy TL, Murphy KM, et al. BATF-JUN is critical for IRF4-mediated transcription in T cells. Nature. 2012;490:543–6.

Li *et al. Genome Medicine*   (2020) 12:81

Page 17 of 17

53. Kim CJ, Lee C-G, Jung J-Y, Ghosh A, Hasan SN, Hwang S-M, et al. The transcription factor Ets1 suppresses T follicular helper type 2 cell differentiation to halt the onset of systemic lupus erythematosus. Immunity. 2018;49:1034–1048.e8.

54. Zentner GE, Tesar PJ, Scacheri PC. Epigenetic signatures distinguish multiple classes of enhancers with distinct cellular functions. 21(8).

55. Sharifi-Zarchi A, Gerovska D, Adachi K, Totonchi M, Pezeshk H, Taft RJ, et al. DNA methylation regulates discrimination of enhancers from promoters through a H3K4me1-H3K4me3 seesaw mechanism. https://doi.org/10.1186/s12864-017-4353-7.

56. Yagi S, Hirabayashi K, Sato S, Li W, Takahashi Y, Hirakawa T, et al. DNA methylation profile of tissue-dependent and differentially methylated regions (T-DMRs) in mouse promoter regions demonstrating tissue-specific gene expression. Genome Res. 2008;18:1969–78.

57. Rakyan VK, Down TA, Thorne NP, Flicek P, Kulesha E, Gräf S, et al. An integrated resource for genome-wide identification and analysis of human tissue-specific differentially methylated regions (tDMRs). Genome Res. 2008;18:1518–29.

58. Lea AJ, Vockley CM, Johnston RA, Del Carpio CA, Barreiro LB, Reddy TE, et al. Genome-wide quantification of the effects of DNA methylation on human gene regulation. Elife. 2018;7.

59. Stephens DC, Poon GMK. Differential sensitivity to methylated DNA by ETS-family transcription factors is intrinsically encoded in their DNA-binding domains. Nucleic Acids Res. 2016;44:8671–81.

60. Ando M, Saito Y, Xu G, Bui NQ, Medetgul-Ernar K, Pu M, et al. Chromatin dysregulation and DNA methylation at transcription start sites associated with transcriptional repression in cancers. Nat Commun. 2019;10.

61. Arechederra M, Daian F, Yim A, Bazai SK, Richelme S, Dono R, et al. Hypermethylation of gene body CpG islands predicts high dosage of functional oncogenes in liver cancer. Nat Commun. 2018;9.

62. Link MA, Lücke K, Schmid J, Schumacher V, Eden T, Rose-John S, et al. The role of ADAM17 in the T-cell response against bacterial pathogens. PLoS One. 2017;12.

63. Gilardin L, Delignat S, Peyron I, Ing M, Lone YC, Gangadharan B, et al. The ADAMTS13 1239-1253 peptide is a dominant HLA-DR1-restricted CD4 + T-cell epitope. Haematologica Ferrata Storti Foundation. 2017;102:1833–41.

64. Patsoukis N, Bardhan K, Chatterjee P, Sari D, Liu B, Bell LN, et al. PD-1 alters T-cell metabolic reprogramming by inhibiting glycolysis and promoting lipolysis and fatty acid oxidation. Nat Commun. 2015;6.

65. Giang S, La Cava A. IRF1 and BATF: key drivers of type 1 regulatory T-cell differentiation. Cell. Mol Immunol. 2017;14(8):652–4.

66. Millstein J, Zhang B, Zhu J, Schadt EE. Disentangling molecular relationships with a causal inference test. BMC Genet. 2009;10:23.

67. Liu Y, Aryee MJ, Padyukov L, Fallin MD, Hesselberg E, Runarsson A, et al. Epigenome-wide association data implicate DNA methylation as an intermediary of genetic risk in rheumatoid arthritis. Nat Biotechnol. 2013;31:142–7.

68. Mazzone R, Zwergel C, Artico M, Taurone S, Ralli M, Greco A, et al. The emerging role of epigenetics in human autoimmune disorders. Clin Epigenetics. 2019;11(1).

69. Issuree PD, Day K, Au C, Raviram R, Zappile P, Skok JA, et al. Stage-specific epigenetic regulation of CD4 expression by coordinated enhancer elements during T cell development. Nat Commun. 2018;9.

70. Lu T, Klein KO, Colmegna I, Lora M, Greenwood CMT, Hudson M. Whole-genome bisulfite sequencing in systemic sclerosis provides novel targets to understand disease pathogenesis. BMC Med Genomics; 2019;12.

71. Chen S, Pu W, Guo S, Jin L, He D, Wang J. Genome-Wide DNA methylation profiles reveal common epigenetic patterns of interferon-related genes in multiple autoimmune diseases. Front Genet. Frontiers; 2019;10:223.

72. Gutierrez-Arcelus M, Baglaenko Y, Arora J, Hannes S, Luo Y, Amariuta T, et al. Allele-specific expression changes dynamically during T cell activation in HLA and other autoimmune loci. Nat Genet. 2020;52(3).

73. Blake LE, Roux J, Hernando-Herraez I, Banovich N, Garcia Perez R, Hsiao CJ, et al. A comparison of gene expression and DNA methylation patterns across tissues and species. Genome Res. 2020;gr.254904.119.

74. Li L, Gao Y, Wu Q, Cheng ASL, Yip KY. New guidelines for DNA methylome studies regarding 5-hydroxymethylcytosine for understanding transcriptional regulation. Genome Res. 2019;29:543–53.

75. Anastasiadi D, Esteve-Codina A, Piferrer F. Consistent inverse correlation between DNA methylation of the first intron and gene expression across tissues and species. Epigenetics Chromatin. 11(1).

76. Korthauer K, Irizarry RA. Genome-wide repressive capacity of promoter DNA methylation is revealed through epigenomic manipulation. bioRxiv; 2018;381145.

77. Wang H, Maurano MT, Qu H, Varley KE, Gertz J, Pauli F, et al. Widespread plasticity in CTCF occupancy linked to DNA methylation. Genome Res. 2012;22:1680–8.

78. Lee DH, Singh P, Tsai SY, Oates N, Spalla A, Spalla C, et al. CTCF-dependent chromatin bias constitutes transient epigenetic memory of the mother at the H19-igf2 imprinting control region in prospermatogonia. PLoS Genet. 2010;6.

79. de la Rica L, Rodríguez-Ubreva J, García M, Islam ABMMK, Urquiza JM, Hernando H, et al. PU.1 target genes undergo Tet2-coupled demethylation and DNMT3b-mediated methylation in monocyte-to-osteoclast differentiation. Genome Biol 2013;14.

80. Sanyal A, Lajoie BR, Jain G, Dekker J. The long-range interaction landscape of gene promoters. Nature. 2012;489:109–13.

81. Orlando G, Law PJ, Cornish AJ, Dobbins SE, Chubb D, Broderick P, et al. Promoter capture Hi-C-based identification of recurrent noncoding mutations in colorectal cancer. Nat Genet; 2018. p. 1375–1380.

82. Lemire M, Zaidi SHE, Ban M, Ge B, Aïssi D, Germain M, et al. Long-range epigenetic regulation is conferred by genetic variation located at thousands of independent loci. Nat Commun. 2015;6.

83. Shin J, Bourdon C, Bernard M, Wilson MD, Reischl E, Waldenberger M, et al. Layered genetic control of DNA methylation and gene expression: a locus of multiple sclerosis in healthy individuals. Hum Mol Genet. 2015;24:5733–45.

84. Martin P, McGovern A, Orozco G, Duffus K, Yarwood A, Schoenfelder S, et al. Capture Hi-C reveals novel candidate genes and complex long-range interactions with related autoimmune risk loci. Nat Commun. 2015;6.

85. Cornely R, Pollock AH, Rentero C, Norris SE, Alvarez-Guaita A, Grewal T, et al. Annexin A6 regulates interleukin-2-mediated T-cell proliferation. Immunol Cell Biol. 2016;94:543–53.

86. Li T, Garcia-Gomez A, Morante-Palacios O, Ciudad L, Özkaramehmet S, Van Dijck E, et al. SIRT1/2 orchestrate acquisition of DNA methylation and loss of histone H3 activating marks to prevent premature activation of inflammatory genes in macrophages. Nucleic Acids Res. 2020;48:665–81.

87. Charlton J, Jung EJ, Mattei AL, Bailly N, Liao J, Martin EJ, et al. TETs compete with DNMT3 activity in pluripotent cells at thousands of methylated somatic enhancers. Nat Genet; 2020;52(8).

88. Du J, Johnson LM, Jacobsen SE, Patel DJ. DNA methylation pathways and their crosstalk with histone methylation. Nat Rev Mol Cell Biol. 2015;16:519–32.

89. Li Y, Haarhuis JHI, Cacciatore ÁS, Oldenkamp R, van Ruiten MS, Willems L, et al. The structural basis for cohesin-CTCF-anchored loops. Nature. 2020;578(7795).

90. Wiehle L, Thorn GJ, Raddatz G, Clarkson CT, Rippe K, Lyko F, et al. DNA (de) methylation in embryonic stem cells controls CTCF-dependent chromatin boundaries. Genome Res. 2019;29:750–61.

Publisher's Note