Exploring the world of non-coding genes in stem cells and autoimmunity.

Messemaker, T.C.

Chapter 6

Antisense long non-coding RNAs are deregulated in skin tissue of patients with systemic sclerosis

J Invest Dermatol. 2017 Nov 24. pii: S0022-202X(17)33169-X. doi: 10.1016/j.jid.2017.09.053.

Tobias C. Messemaker, Loubna Chadli, Guoshuai Cai , Varshna S. Goelela, Maaike Boonstra, Annemarie L. Dorjée, Stefan N. Andersen, Harald M. M. Mikkers, Peter van ‘t Hof, Hailiang Mei, Oliver Distler, Harmen H. M. Draisma, Michael Johnson, Nicole Orzechowski, Robert W Simms, Rene E. M. Toes, Jamil Aarbiou, Tom W. Huizinga, Michael L. Whitfield, Jeroen DeGroot, Jeska de Vries-Bouwstra, Fina Kurreeman

ABSTRACT

Systemic sclerosis (SSc) is an autoimmune disease characterized by fibrosis of skin and multiple organs of which the pathogenesis is poorly understood. Here we studied differentially expressed coding and non-coding genes in relation to SSc pathogenesis with a specific focus on antisense non-coding RNAs. Skin biopsy-derived RNAs from fourteen early SSc patients and six healthy individuals were sequenced with ion-torrent and analysed using DEseq2.

Overall, 4901 genes with a fold change >1.5 and a false discovery rate < 5% were detected in patients versus controls. Upregulated genes clustered in immunological, cell adhesion and keratin-related processes. Interestingly, 676 deregulated non-coding genes were detected, 257 of which were classified as antisense genes. Sense genes expressed opposite of these antisense genes were also deregulated in 42% of the observed sense-antisense gene pairs. The majority of the antisense genes had a similar effect sizes in an independent North American dataset with three genes (CTBP1-AS2, OTUD6B-AS1 and AGAP2-AS1) exceeding the study-wide Bonferroni-corrected ρ-value (P_{Bonf}<0.0023, $P_{combined}$ = $1.1x10^{-9}$, $1.4x10^{-8}$, $1.7x10^{-6}$, respectively). In this study, we highlight that together with coding genes, (antisense) long non-coding RNAs are deregulated in skin tissue of SSc patients suggesting a novel class of genes involved in pathogenesis of SSc.

INTRODUCTION

Systemic sclerosis (SSc) is a heterogeneous complex autoimmune disease affecting connective tissues. Its pathogenesis remains elusive, but patients harbour vascular changes like Raynaud's phenomenon, autoimmunity with the presence of distinct autoantibodies, activation of both innate and adaptive immunity and active deposition of extracellular matrix leading to fibrosis. Progression of vascular and fibrotic organ damage accounts for a large proportion of the chronic morbidity and mortality up to 25% in the first five years after diagnosis in SSc (Rubio-Rivas *et al*. 2014).

In order to further understand the processes involved in SSc pathophysiology, several groups have performed gene expression studies in peripheral blood and skin of SSc patients (Gardner *et al*. 2006; Pendergrass *et al*. 2012; Milano *et al*. 2008; Whitfield *et al*. 2003). These studies have revealed that

gene expression profiles in skin from SSc patients not only differ from healthy skin but are associated with skin disease severity (Milano *et al*. 2008). Interestingly, several SSc-specific gene sets have been identified which include fibrosis related pathways involved in skin thickening (TGF-β related genes, collagen genes) as well as immunological and keratin-related pathways (interferon genes, activated macrophage genes, chemokine-related genes and keratin genes) (Mahoney *et al*. 2015; Assassi *et al*. 2015; Gardner *et al*. 2006; Mathes *et al*. 2014). These studies were all performed using microarrays, and focussed on the identification of protein coding genes and pathways that are differently regulated in SSc, and as a consequence missing an important component of non-coding genes involved in disease pathogenesis. With the use of next generation sequencing, transcriptomics studies can now shed light on the non-coding genome and the role of long non-coding RNAs (lncRNAs) in disease mechanisms.

lncRNAs represent an important layer of genome regulation and their role in the context of SSc is currently unknown. lncRNAs are transcripts over 200 nucleotides in length and come in diverse flavours including: antisense RNAs, long intergenic non-coding RNAs (lincRNAs) and pseudogenes (Derrien *et al*. 2012). Although the function of the majority of lncRNAs remains unknown, a role in regulating and shaping the genome has been proposed (Rinn JL 2013; Melé and Rinn 2016). Specifically, antisense RNAs can influence RNA levels of their sense counterpart (Faghihi and Wahlestedt 2010; Derrien *et al*. 2012; Chan *et al*. 2015; Peng *et al*. 2015; Kimura *et al*. 2013). In diseases like SSc, where deregulated gene expression signatures are present, identification of such regulatory genes may represent interesting candidates as biomarkers or unlock novel treatment avenues. In addition, compared to coding genes, lncRNAs display higher tissue specificity in their expression patterns (Derrien *et al*. 2012). Recently, deregulated lncRNA expression has been described in the skin of patients with psoriasis (Gupta *et al*. 2016) and in the regulation of TGF-β mediated processes (Richards *et al*. 2015) suggesting that lncRNAs may also be deregulated in skin of SSc patients.

In order to extend the current knowledge of the gene expression signature in SSc, we have performed RNA sequencing on skin biopsies of SSc patients and healthy controls and investigated deregulated expression of both coding and non-coding genes. Moreover, main findings on non-coding genes were replicated in an independent dataset.

RESULTS

DE genes in SSc patients are enriched in immunological, cell activation and keratinization pathways and overlap with previous studies.

In order to identify genes and pathways involved in SSc pathophysiology, we evaluated RNA expression levels in patients and controls. 4901 genes were DE with a minimum fold change of 1.5 and FDR *p*-value below 0.05 (Supplementary File 2). Hierarchical clustering on basis of these DE genes separates patients from healthy controls with the exception of 1 patient which displays a normal-like expression pattern (Supplementary Figure 1). Pathway analysis of overexpressed genes shows an enrichment in the immune response, cell activation and keratinization pathways (Supplementary File 3). Cross comparison with DE genes from a recent publication by Assassi *et al* indicates a small highly consistent (>96%) overlap with the most prominent common pathways belonging to the immunological and cell adhesion related processes (Figure 1a-c, Supplementary File 4).

In-depth analysis of specific SSc-related gene sets highlights additional candidate genes implicated in SSc and an inflammatory gene signature.

As an initial approach, we performed an in-depth analysis of several SSc gene sets which previously came forward from microarray studies including TGFβ signalling, collagen, keratin, interferon, alternative macrophage activation genes and chemokines (Figure 2, Supplementary File 1 and Supplementary File 5).

Similar to our GO-term enrichment analysis, a clear increased TGFβ expression profile that is involved in many fibrotic processes was not observed in our patient population as only 5 out of 86 TGFβ signalling genes were significantly increased (Figure 2a). On the other hand, TGFβ-gene COMP was found increased in patients as similar to previous reports (Farina *et al*. 2009; Assassi *et al*. 2015; Gardner *et al*. 2006). Moreover, many collagen and keratin associated genes are significantly increased in patients (Figure 2b and c). Also, 33 out of 97 genes from the interferon and macrophage gene sets were significantly increased in SSc patients (Figure 2d and e) indicating an increased inflammatory gene signature being present in early SSc patients (Assassi *et al*. 2015; Greenblatt *et al*. 2012; Mahoney *et al*. 2015). This observation is in line with previous studies showing that in early SSc (as is our population) the inflammatory signature is more prevalent (Assassi *et al*. 2015).

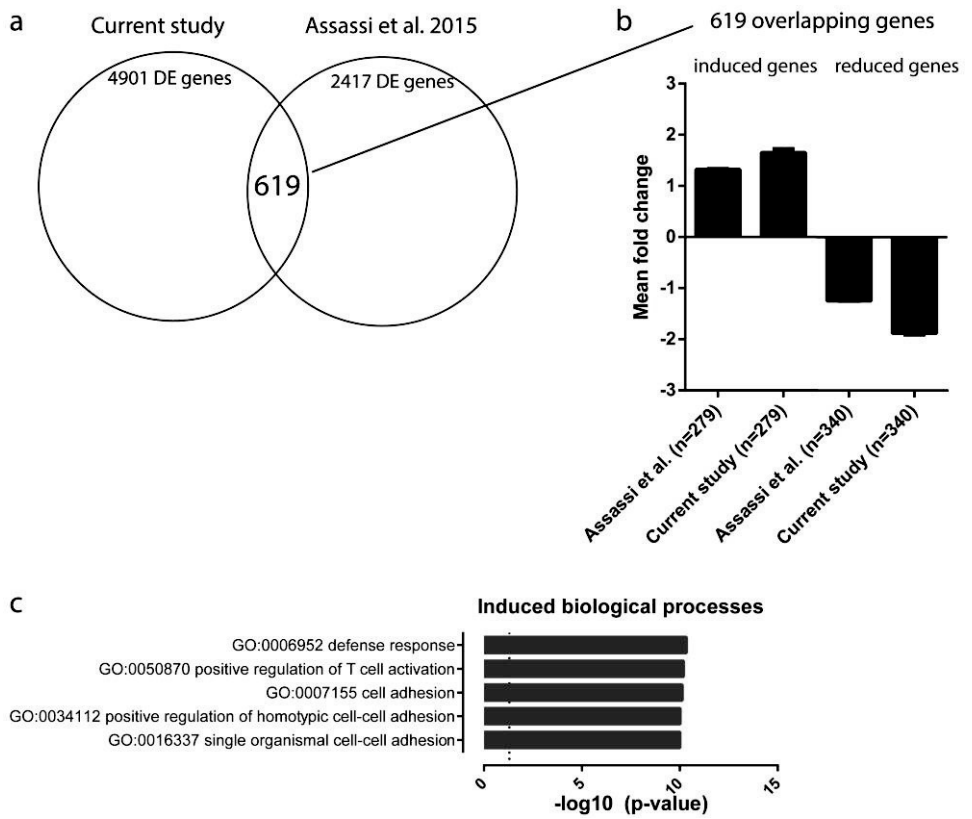

Figure 1. DE genes overlap with a previous microarray study and reveals consistent deregulated pathways. (a) Venn diagram comparing DE genes in SSc patients versus controls from the current study (n = 4901 DE genes) with a microarray study from Assassi *et al*. (n = 2417 DE genes). (b) Directionality of 619 consistently deregulated genes from the two studies displayed as mean fold change (mean ± SE). Genes up or down regulated from Assassi *et al* were selected and plotted. The concomitant fold changes of these genes from our study were also plotted indicating similar directionality in both studies. (c) Top 5 Biological processes GO-terms enriched using genes that are upregulated in SSc patients from the two studies.

Since skin paraffin sections were available for the patients under study, we stained skin sections for CD68, a marker for macrophages. In line with the observed inflammatory gene signature, clusters of macrophages were detected in the skin of SSc patients (Supplementary Figure 2). Besides these observations, several (to our knowledge previously unreported) genes including COL4A4, Keratin 4 and 9, TNFAIP3, CX3CR1, CXCL2 and PF4 were strongly deregulated in SSc patients (Figure 2b-k, Supplementary Table 1).

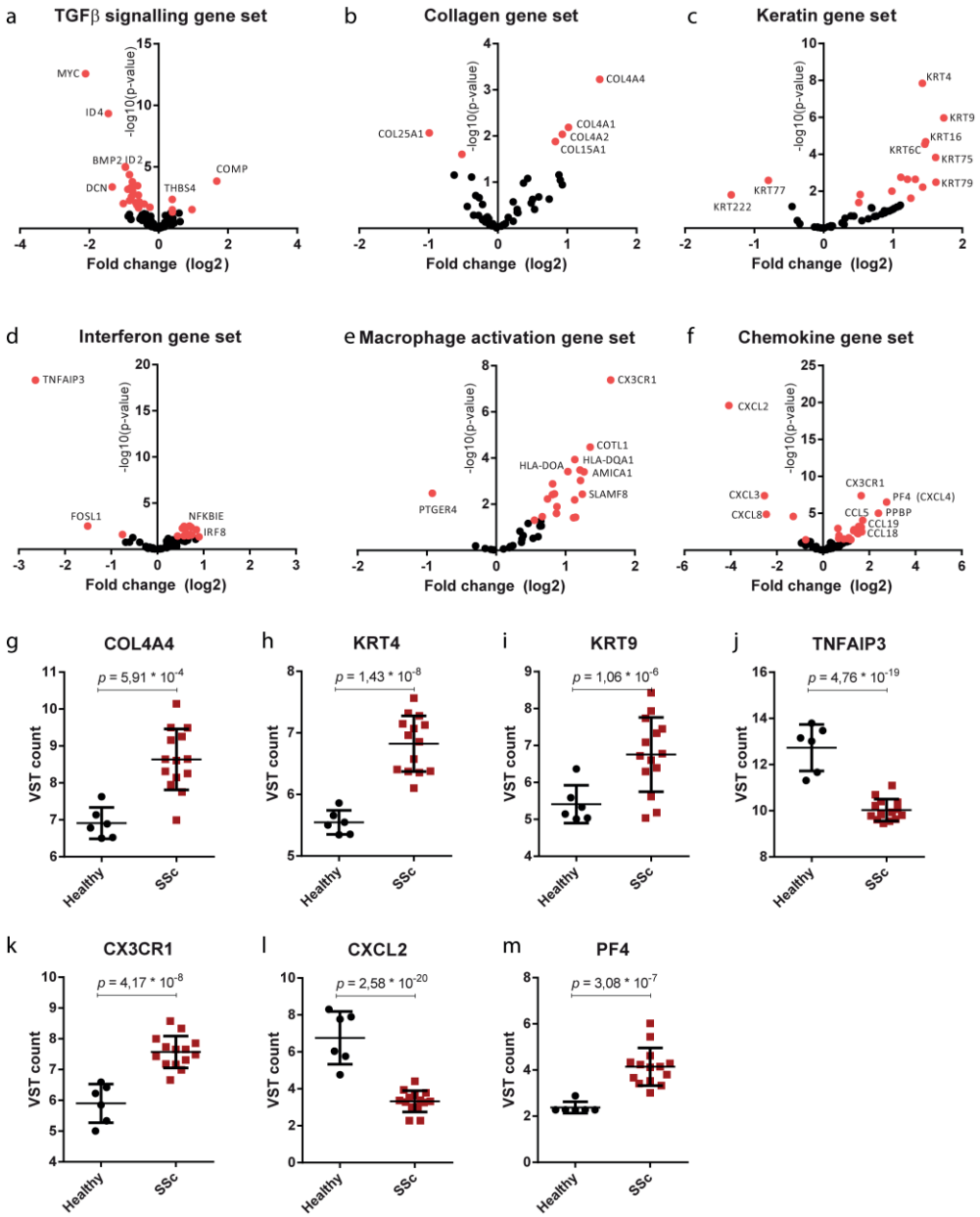

Figure 2. Analysis of DE genes of specific SSc gene sets. Volcano plots showing differential expression within the 6 genesets: TGFβ signalling (a, n=86 genes), collagen (b, n=46 genes), keratin (c, n=76 genes), interferon (d, n=50 genes), alternative macrophage activation (e, n=60 genes) and chemokine (f, n=84 genes). Genes depicted in red were significantly deregulated (Benjamini Hochberg-corrected *p*-value < 0.05). RNA levels (VST count) of individuals genes in healthy controls and SSc patients for COL4A4 (g), KRT4 (h), KRT9 (i), TNFAIP3 (j), CX3CR1 (k), CXCL2 (l), and PF4 (m). *p*-values represent Benjamini-Hochberg-corrected *p*-values. The mean ±SD of each group is depicted in the graphs.

Identification of DE lncRNAs in SSc skin biopsies in comparison to healthy controls.
In addition to coding genes, RNA sequencing allows the query of non-coding genes. Among 15941 annotated lncRNAs, 4171 were expressed in our skin biopsies. 676 lncRNAs were DE (FDR < 0.05) between SSc patients and healthy controls and show a clear differential expression signature (Figure 3a).

All 676 DE lncRNAs are listed in Supplementary File 6. Out of 676 lncRNA genes, 122 genes were decreased, while the expression of 554 genes was increased in SSc patients as compared to healthy controls.Interestingly, clustering analysis using different selection criteria of lncRNAs all displayed a pattern where non-clinically active patients clustered within the patient population and separate from controls (Supplementary Figure 3). In total, 348 lncRNAs displayed over 2 fold differential expression and the top upregulated gene is CAPN10-AS1, an antisense lncRNA (Figure 3b). Interestingly, among the 676 deregulated lncRNAs, the largest proportion (38%) belongs to the antisense gene category (Figure 3c). nAntisense lncRNAs have recently been described to have important regulatory roles on their coding gene counterparts expressed in the sense direction (Pelechano and Steinmetz 2013; Werner 2013; Katayama *et al*. 2005; Villegas and Zaphiropoulos 2015). The relevance of the antisense genes in our data set was therefore investigated.

Identification of DE antisense genes in SSc patients and their link to sense coding genes.
In order to gain further insight into the possible role of antisense RNAs in SSc, we focused our analysis on antisense genes of which a sense gene was annotated (also known as sense-antisense (SAS) gene pairs). Close proximity of antisense genes with sense genes have been linked to co-expression and co-regulation within such a SAS gene pair (Villegas and Zaphiropoulos 2015; Katayama *et al*. 2005). Out of 257 DE antisense genes, 62 have an annotated sense gene. Interestingly, an important proportion (26 out of 62) of these SAS gene pairs includes both a significant DE antisense gene and a significant DE sense gene (FDR < 0.05) (Figure 3d). We further explored the relation between sense and antisense genes using correlation analysis by comparing the correlation of gene pairs where both genes are deregulated compared to gene pairs which were not deregulated in patients (consisting of gene pairs of which only one of the two genes was deregulated and of gene pairs of which neither the sense gene nor the

antisense gene was deregulated in patients). Here high correlations (median r > 0.7) were observed for gene pairs significantly deregulated in SSc (SSc gene pairs) and were significantly higher in comparison with non SSc-deregulated gene pairs ($P < 0.001$) (Figure 3e).

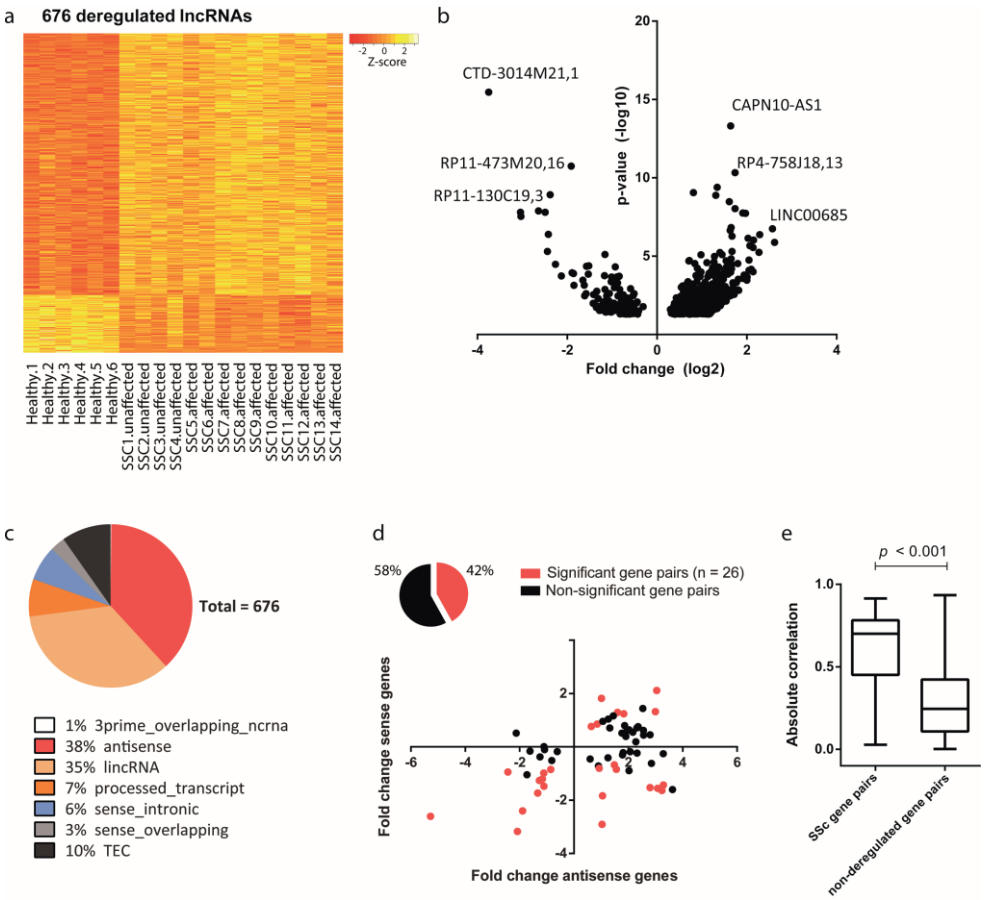

Figure 3. DE lncRNAs in SSc patients in comparison with healthy controls. (a) Heatmap depicting the Z-scores of 676 deregulated lncRNAs. Red colour indicates low expression and the yellow colour indicates high expression. (b) Volcano plot showing top deregulated lncRNAs by fold change (log2) on the x-axis and the *p*-value (-log10) on the y-axis. (c) Deregulated lncRNAs (n = 676) divided by subclasses. (d) Venn diagram and scatter plot showing the proportion of significant gene pairs (Benjamini Hochberg-corrected *p*-value < 0.05). Significant DE gene pairs are depicted in red and depicting the fold change (log2) of both the sense and antisense genes. e, Absolute spearman rank correlation between sense and antisense genes within SSc gene pairs and gene pairs not deregulated in SSc.

These data indicate that the identified antisense genes are either coexpressed with coding genes or involved in the regulation of their levels, illustrating a mechanism by which long non-coding (antisense) RNAs may play a role in SSc.

In order to obtain further evidence for the involvement in SSc of the selected 26 antisense genes, we acquired gene expression values from an independent dataset where RNA sequencing had been performed (14 SSc patients, 6 controls, Whitfield *et al*, unpublished data). 4 of the 26 genes were not present due to low expression in the independent dataset and were further excluded from the analysis. 12 out of 22 genes follow the same direction of association in both datasets (Table 1).

Table 1. Replication of 22 antisense genes in an independent RNA-seq dataset. The table includes, Fold changes (Log2FC) and *p*-values (P) from both studies and a combined *p*-value. Combined *p*-values were not calculated for the genes with opposite direction of association according to Rau *et al*. 2014 (Rau, Marot, and Jaffrézic 2014).

	Dataset 1			Dataset 2		
Gene	Log2FC	P	FDR	log2FC	P	Combined
CTBP1-AS2	0,32	0,012	0,044	0,40	7,5E-07	1,1E-09
OTUD6B-AS1	-0,95	7,0E-05	0,001	-0,63	0,001	1,4E-08
AGAP2-AS1	0,50	0,006	0,027	0,34	0,002	1,7E-06
HAND2-AS1	-1,04	0,002	0,011	-0,63	0,007	2,1E-06
HMGN3-AS1	-0,64	0,009	0,034	-0,33	0,017	2,6E-05
ZBTB11-AS1	0,53	0,002	0,010	0,17	0,143	4,5E-05
NIFK-AS1	-0,56	0,006	0,027	-0,36	0,178	2,3E-04
WAC-AS1	-0,58	0,001	0,009	-0,17	0,217	5,9E-05
PIK3CD-AS2	1,50	5,1E-06	1,5E-04	0,18	0,407	3,1E-07
ARRDC1-AS1	0,43	0,012	0,045	0,13	0,411	0,001
ZNF252P-AS1	1,64	1,1E-04	0,001	0,19	0,422	7,8E-06
SBF2-AS1	-0,43	0,014	0,049	-0,06	0,715	0,002
UNC5B-AS1	1,52	5,6E-05	0,001	-0,76	0,007	NA
HOXA10-AS	-2,64	4,4E-11	1,3E-08	0,53	0,056	NA
SLC25A25-AS1	0,52	4,3E-04	0,004	-0,30	0,163	NA
RUNDC3A-AS1	0,92	0,001	0,005	-0,30	0,225	NA
ZBED5-AS1	0,45	0,012	0,044	-0,16	0,275	NA
LOXL1-AS1	0,80	5,6E-05	0,001	-0,22	0,408	NA
BRWD1-AS2	1,54	3,4E-07	2,0E-05	-0,14	0,514	NA
ZEB1-AS1	-0,67	0,003	0,015	0,06	0,738	NA
RGMB-AS1	0,78	0,005	0,023	-0,05	0,815	NA
TMPO-AS1	1,40	7,4E-08	5,8E-06	-0,02	0,923	NA

Three antisense genes CTBP1-AS2, OTUD6B-AS1 and AGAP2-AS1 reached beyond the study-wide replication *p*-value threshold (P < 0.0023) (Table 1 and Figure 4a-c). Verification using an second experimental approach confirmed that these three genes are significantly deregulated ((P < 0.01), Supplementary Figure 4).

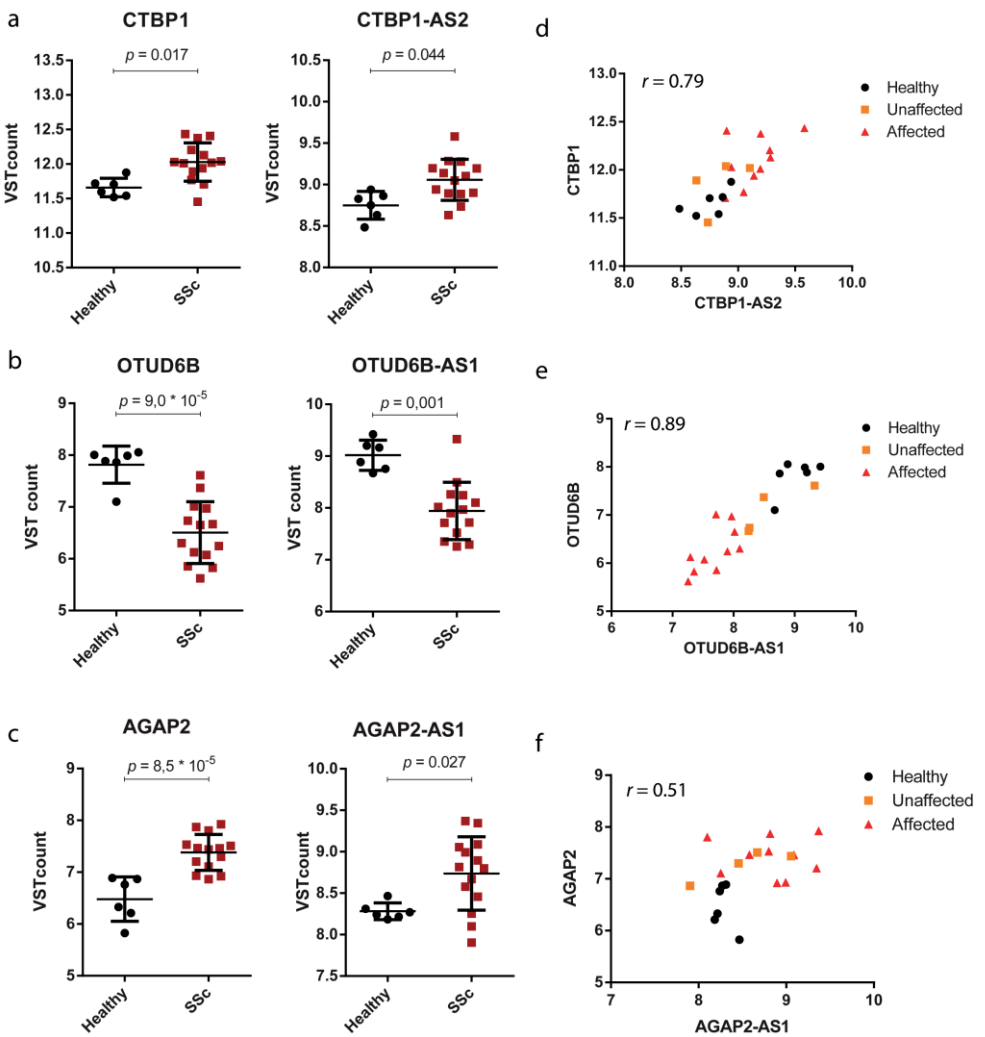

Figure 4. Top 3 replicated antisense genes show strong correlation with their sense coding gene. (a-c) VST count values of top 3 replicated SAS gene pairs: CTBP1 (a), OTUD6B (b) and AGAP2 (c) in SSc patients (n = 14) and controls (n = 6) *p*-values are Benjamini-Hochberg corrected and were generated via DEseq2. (d-f) Correlation between sense and antisense genes within a gene pair for CTBP1 (d), OTUD6B (e) and AGAP2 (f). Count values are divided into healthy, unaffected or affected skin tissue. Spearman rank test was used to calculate correlations between the sense and antisense gene.

We confirmed the non-coding nature of these antisense genes using a coding potential calculator which showed an overall low coding potential for CTBP1-AS2, OTUD6B-AS1 and AGAP2-AS1 (Supplementary Figure 5). We next evaluated the relationship of these non-coding antisense genes with their paired sense gene across our patients and controls. Interestingly, the identified antisense genes show a strong correlation with their paired sense gene across the 20 individuals, in particular for OTUD6B-AS1 and CTBP1-AS2 (r = 0.89, P < 0.001 and r = 0.79, P < 0.001, respectively, Figure 4D-F). As skin is composed of many cell types we took advantage of available cell type specific expression datasets to gain further insight into which cell types may be relevant for these candidates. CTBP1 and CTBP1-AS2 levels also positively correlate across specific cell types and this correlation is highest in immune cells (r = 0.7, P <0.001) (Figure 5a). The OTUD6B gene pair is expressed in dermal and immune cells, and shows a correlation that was similar as observed across patients (r = 0.6-0.8, P<0.01) (Figure 5b).

Interestingly, AGAP2 is only expressed in immune cells while AGAP2-AS1 is only expressed in dermal cell types (Figure 5c). Finally, we further investigated the correlation of these gene pairs in the replication dataset. These data show that the CTBP1 and OTUD6B gene pairs also display a significant correlation (r > 0.8, P <0.001 for both gene pairs) in the replication dataset (Supplementary Figure 6) while the correlation for AGAP2 is absent in the replication dataset (r = 0.21). These results seem to coincide with the tissue-specific expression data obtained from FANTOM5 were a positive correlation between AGAP2-AS1 and AGAP2 is also absent. Altogether, we identified non-coding genes that are expressed in cell-types relevant for SSc and of which the levels are altered in a disease specific manner in the skin of SSc patients.

DISCUSSION

Our results using next generation sequencing firstly confirmed previous studies using microarrays and confirmed an inflammatory signature in the skin of early SSC patients. In addition to the analyses on coding genes, we report an in-depth analysis of deregulated lncRNAs in skin tissue from SSc patients. The top-3 deregulated antisense genes included CTBP1-AS2, OTUD6B-AS1 and AGAP2-AS1, and these findings were replicated in an independent dataset and further validated by qPCR. The expression of these lncRNAs is clearly distinct in patients,

although the functional consequences of these deregulations are at this point difficult to infer given the limited information available on their potential functions. Future in-depth functional analyses are warranted on the functional roles of these genes to confirm their role in SSc pathogenesis. lncRNAs play an important role in development and disease (Batista and Chang 2013; Esteller 2011), but have not yet been described in relation to SSc.

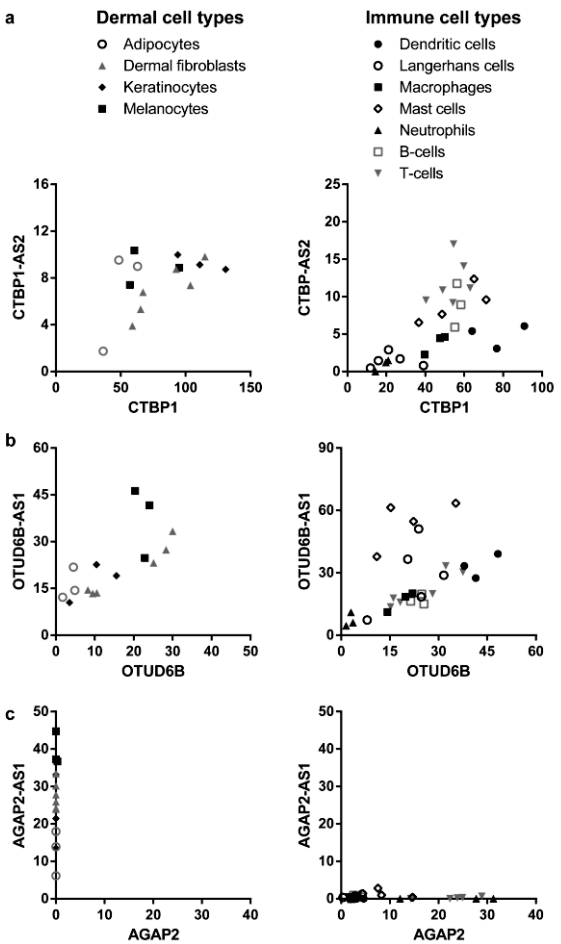

Figure 5. Cell type specific expression of SAS gene pairs in dermal and immune cells. Expression levels for CTBP1 and CTBP1-AS2 (a), OTUD6B and OTUD6B-AS1 (b) and AGAP2 and AGAP2-AS1 (c) in dermal and immune cell types. Expression values are shown as TPM for both the sense and antisense gene. Expression values of each cell type was measured in at least 3 donors. Correlation analysis was performed by spearman rank test.

Most lncRNAs are not yet available on microarrays and are therefore missed in the available data sets that were investigating SSc deregulated genes. More importantly, association of lncRNAs with inflammatory diseases like rheumatoid arthritis, diabetes and psoriasis are increasingly being reported, highlighting their potential role in disease mechanisms (Gupta *et al*. 2016; Messemaker, Huizinga, and Kurreeman 2015). Here, we identify 676 lncRNAs that are deregulated in skin from SSc patients as compared to healthy individuals. A large proportion of the deregulated lncRNAs belonged to the antisense RNA category. Antisense RNAs which reside in a locus with a sense gene (and often span part of this gene) and potentially function as co-regulators of the sense gene (Chan *et al*. 2015; Kimura *et al*. 2013; Peng *et al*. 2015). We identified 26 SAS gene pairs which displayed evidence of differential expression in SSc patients versus controls. From these gene pairs, 55% of the antisense genes showed similar direction of association in an independent data set. The top three deregulated antisense genes included CTBP1-AS2, OTUD6B-AS1 and AGAP2-AS1. OTUD6B is a deubiquitinating enzyme of which little is known. Its downregulation has been linked to cell proliferation in B cells following prolonged cytokine stimulation (Xu *et al*. 2011). CTBP1 is a C terminal binding protein which acts as a transcriptional corepressor and plays a role in epidermal development (Boxer *et al*. 2014). Increased CTBP1 levels were shown to disrupt skin homeostasis (Deng *et al*. 2014). AGAP2 was found upregulated in various cancers and is involved in focal adhesion and cell migration (Jia *et al*. 2016; Zhu *et al*. 2009). Interestingly, AGAP2-AS1 was also shown to be involved in cell migration and is able to repress transcription via interaction with EZH2 and LSD1 in cancer cells (Li *et al*. 2016).

Based on our data, we believe that future studies on functional roles of lncRNAs in SSc pathogenesis might focus on CTBP1-AS2, OTUD6B-AS1 and AGAP2-AS1 as these were significantly deregulated, the deregulation was also found in an independent dataset, and based on current knowledge a role in pathophysiology is plausible. Thereby, one should take into account that we have investigated deregulated polyA-positive lncRNAs, while also polyA-negative lncRNAs exists (Derrien *et al*. 2012). Although polyA-negative lncRNAs are less well-studied, we do hypothesize that also these lncRNAs might play important roles in SSc development and require further investigation (Yang *et al*. 2011). With respect to coding genes, we observe an inflammatory signature, in line with previous research that shows the presence of an interferon/inflammatory signature in early SSc patients (Johnson *et al*. 2015). In contrast to previous research, a clear

TGF-β signal did not come forward from our gene list, despite the increase of fibrosis related-genes as ACTA1 and COMP (Farina *et al*. 2009). When comparing genes from our study with a previous published dataset, a small proportion of genes (n = 619) overlaps suggesting that consistent deregulated genes exist despite SSc-well known disease heterogeneity, large differences in the mean age and disease duration of patients between both studies (Supplementary Table 2). Moreover, an additional comparison with 415 genes obtained from a meta-analysis performed by Lofgren *et al* show that 159 genes overlapped (38%)(Lofgren *et al*. 2016).

We investigated specific SSc-gene sets in more detail to identify genes deregulated in early SSc patients. Our study reports several coding genes which have not previously been highlighted in gene expression studies of SSc. COL4 (COL4A1, COL4A2 and COL4A4), is a gene in the collagen family and is a major component of the dermal-epidermal junction. Elevated levels of COL4 protein have been found in the serum of SSc patients (Gerstmeier H, Gabrielli A, Meurer M, Brocks D, Braun-Falco O 1988) and COL4 autoantibodies have been found in 31% of SSc patients highlighting that an increase of COL4 might play a role in SSc (Riente *et al*. 1995). KRT4 and KRT9, overexpressed genes from our study are normally not expressed in forearm skin. KRT4 is expressed in mucosal tissue and is increased upon inflammation (Bosch *et al*. 1989), while KRT9 is normally expressed in soles and hand palms (Rinn *et al*. 2008). KRT9 is required for structural integrity of the epidermis and KRT9 was found increased in psoriasis patients (Fu *et al*. 2014; Kim *et al*. 2016). The increased expression of these keratins in skin of early SSc patients highlights the possibility of aberrant activation of these genes early in disease.

Besides collagen and keratin genes, we also identified inflammatory genes. Some of these deregulated inflammatory genes are located in loci that are genetically associated to SSc including *HLA* and *TNFAIP3* (Dieudé *et al*. 2010). Interestingly, the expression of TNFAIP3 is strongly reduced in SSc skin tissue. Given the role of TNFAIP3 as a negative regulator of NF-κB signalling, its downregulation would be suggestive of an increased NF-κB activation, possibly further enhancing the increased pro-inflammatory environment. TNFAIP3 was also found deregulated in several other cell types and suggests that genes and pathways are deregulated across multiple tissues(Avouac *et al*. 2011). In line with this, we have also observed clusters of macrophages in our SSc skin biopsies. Also increased CX3CR1 expression came forward and likely contributes to skin

inflammation in SSc as CX3CR1 knockout experiments resulted in decreased skin inflammation (Morimura *et al*. 2016). Interestingly, the top deregulated chemokines were CXCL2 and PF4 (CXCL4). CXCL2, a neutrophil chemoattractant and pro-angiogenic factor (Raman, Sobolik-Delmaire, and Richmond 2011), was reduced and might influence vascular repair within skin of SSc patients (Hummers *et al*. 2009). PF4 (CXCL4) was increased at the RNA level and increased PF4 protein levels were found in SSc serum and skin (van Bon *et al*. 2014). Our study suggests that despite the short disease duration of the patients included in this study, distinct gene expression profiles already exist at an earlier stage in the disease process than investigated so far. Further studies in larger sample sets and long-term follow-up of patients should yield deeper insight into which relevant mechanisms are deregulated in what stage of the disease.

In conclusion, we here report a gene list of 619 genes consistently deregulated over two studies accounting for direction of association and providing a basis of consistent gene expression changes. We show that the expression of keratin genes is increased and that patients display enhanced levels of genes originating from inflammatory gene signatures. In addition, we here provide a blueprint of DE lncRNAs which may play a role as underlying regulators disturbing processes contributing to SSc. Interestingly, even though many of these DE lncRNAs have to our knowledge not yet been described in context of SSc, we show strong correlations with coding genes for several antisense genes. Given the replication in an independent cohort, future studies on the functional role of these specific lncRNAs in SSc pathogenesis are warranted.

MATERIALS AND METHODS

For full details of methods see online supplementary material.

Patient information

Early SSc patients (with a disease duration < 2 years) were recruited at the Department of Rheumatology of the Leiden University Medical Center (Leiden, The Netherlands) and all patients met the American Rheumatism Association classification criteria for SSc (Subcommittee for scleroderma criteria 1980). Patient characteristics can be found in Supplementary Table 3. Institutional review board approval and written informed consent was obtained before patients entered this study. Two 4 mm skin biopsies were taken and from 10

patients the skin biopsy came from a clinically affected area and in 4 patients the skin was locally unaffected. Skin biopsies from healthy individuals were commercially sourced (Tissue Solutions, UK), came from surgeries of arm and leg and were age and sex-matched.

Transcriptome characterisation and analysis

RNA was isolated from skin biopsies and sequenced using polyA selection and a stranded protocol using Ion Torrent next generation sequencing technology (Service XS, The Netherlands). Reads were aligned to the human genome (*Homo sapiens GRh38.78)* using Bowtie2 and STAR and differential expression analysis was carried out using HTseq and DEseq2. All genes with a minimum base mean expression value of 2.3 were included in the differential expression analysis. RNA sequencing files are deposited at the EGA-database under nr: EGAO00000000316 (https://www.ebi.ac.uk/ega/organisations/EGAO00000000316).

Publicly available gene expression datasets and gene sets analysis

DE genes [FDR < 0.05, FC > 1.5] were compared with a publicly available dataset obtained from Assassi *et al*. (Assassi *et al*. 2015). DE genes were investigated via Gene Ontology (GO)-term analysis using Toppgene [version 23 may 2016] and in specific gene sets. Gene sets were obtained from Hugo Gene Nomenclature Committee (HGNC) or by additionally compiled SSc gene sets from alternative sources. Genes in the interferon and alternative macrophage activation signature were obtained from Mahoney *et al*. (Mahoney *et al*. 2015). Genes involved in TGFβ signalling were obtained from the Broad Institute. All genes included in these gene sets are outlined in Supplementary File 1.

Long non-coding RNAs

Genes annotated as lncRNAs (and sub classifications) were obtained from GENCODE (Ensemble version 82) (Harrow *et al*. 2012). Antisense genes were linked to sense genes on the basis of annotations from GENCODE (Harrow *et al*. 2012). Antisense genes with a concomitant DE sense gene were investigated in an as yet unpublished RNA sequencing dataset of skin biopsies of 14 SSc patients and 6 healthy individuals. An overall combined *p*-value was calculated using Fisher's exact test. The top three sense and antisense genes were visualised in IGV to ensure strand specificity and non-overlapping reads (Supplementary Figure 7). The coding potential of antisense genes was determined using an in-

silico coding potential calculator (Kong *et al*. 2007) and analysis of cell specific expression was performed using publicly available FANTOM5 datasets (http://fantom.gsc.riken.jp/5/) (Lizio *et al*. 2015; Severin *et al*. 2014). Correlations between antisense and sense genes were calculated using variance stabilised transformed (VST) counts by spearman rank test.

SUPPLEMENTARY INFORMATION

Supplementary information is available online on the website of the journal of investigative dermatology: Supplementary Figure 1-7 and Supplementary Table 1-4.

ACKNOWLEDGEMENTS

We are grateful for SSc patients for participating in this study. We are also thankful for Koen Quint who aided in the analysis of skin sections. This work was supported by the Dutch Arthritis Foundation, The Netherlands. F. Kurreeman has support from UNESCO-L'Oreal for Women in Science Fellowship, Marie Curie FP7 Outgoing Fellowship and an LUMC Research Fellowship. O. Distler has support by the Swiss National foundation (project number 310030_166259) for studying the role of lncRNAs in SSc.

REFERENCES

Assassi, Shervin, William R Swindell, Minghua Wu, Filemon D Tan, Dinesh Khanna, Daniel E Furst, Donald P Tashkin, *et al*. 2015. "Dissecting the Heterogeneity of Skin Gene Expression Patterns in Systemic Sclerosis." *Arthritis & Rheumatology (Hoboken, N.J.)* 67 (11): 3016–2
6. doi:10.1002/art.39289.

Avouac, Jérôme, Nicolas Cagnard, Jörg H Distler, Yoland Schoindre, Barbara Ruiz, Pierre Olivier Couraud, Georges Uzan, Catherine Boileau, Gilles Chiocchia, and Yannick Allanore. 2011. "Insights into the Pathogenesis of Systemic Sclerosis Based on the Gene Expression Profile of Progenitor-Derived Endothelial Cells." *Arthritis and Rheumatism* 63 (11): 3552–62. doi:10.1002/art.30536.

Batista, Pedro J, and Howard Y Chang. 2013. "Long Noncoding RNAs: Cellular Address Codes in Development and Disease." *Cell* 152 (6). Elsevier Inc.: 1298–1307. doi:10.1016/j.cell.2013.02.012.

Bon, Lenny van, Alsya J Affandi, Jasper Broen, Romy B Christmann, Renoud J Marijnissen, Lukasz Stawski, Giuseppina a Farina, *et al*. 2014. "Proteome-Wide Analysis and CXCL4 as a Biomarker in Systemic Sclerosis." *The New England Journal of Medicine* 370 (5): 433–43. doi:10.1056/NEJMoa1114576.

Bosch, Franz X, Jean-pierre Ouhayoun, Bernhard L Bader, Christine Collin, Christine Grund, Inchul Lee, and Werner W Franke. 1989. "Extensive Changes in Cytokeratin Expression Patterns in Pathologically Affected Human Gingiva." *Virchows Archiv B Cell Pathology*, 59–77.

Boxer, Lisa D, Brook Barajas, Shiying Tao, Jiajing Zhang, and Paul A Khavari. 2014. "ZNF750 Interacts with KLF4 and RCOR1 , Regulators to Repress Epidermal Progenitor Genes and Induce Differentiation Genes," 2013–26. doi:10.1101/gad.246579.114.Bao.

Chan, Jennie, Maninjay Atianand, Zhaozhao Jiang, Susan Carpenter, Daniel Aiello, Roland Elling, Katherine a Fitzgerald, and Daniel R Caffrey. 2015. “Cutting Edge: A Natural Antisense Transcript, AS-IL1α, Controls Inducible Transcription of the Proinflammatory Cytokine IL-1α.” *Journal of Immunology (Baltimore, Md. : 1950)* 195 (4): 1359–63. doi:10.4049/jimmunol.1500264.

Deng, Hui, Fulun Li, Hong Li, Yu Deng, Jing Liu, Donna Wang, Gangwen Han, Xiao-Jing Wang, and Qinghong Zhang. 2014. “CtBP1 Overexpression in Keratinocytes Perturbs Skin Homeostasis.” *The Journal of Investigative Dermatology* 134 (5). Elsevier Masson SAS: 1323–31. doi:10.1038/jid.2013.504.

Derrien, Thomas, Rory Johnson, Giovanni Bussotti, Andrea Tanzer, Sarah Djebali, Hagen Tilgner, Gregory Guernec, *et al*. 2012. “The GENCODE v7 Catalog of Human Long Noncoding RNAs : Analysis of Their Gene Structure , Evolution , and Expression.” *Genome Research*, 1775–89. doi:10.1101/gr.132159.111.

Dieudé, P, M Guedj, J Wipff, B Ruiz, G Riemekasten, M Matucci-Cerinic, I Melchers, *et al*. 2010. “Association of the TNFAIP3 rs5029939 Variant with Systemic Sclerosis in the European Caucasian Population.” *Annals of the Rheumatic Diseases* 69 (11): 1958–64. doi:10.1136/ard.2009.127928.

Esteller, Manel. 2011. “Non-Coding RNAs in Human Disease.” *Nature Reviews. Genetics* 12 (12). Nature Publishing Group: 861–74. doi:10.1038/nrg3074.

Faghihi, Mohammad Ali, and Claes Wahlestedt. 2010. “Regulatory Roles of Natural Antisense Transcripts” 10 (9): 637–43. doi:10.1038/nrm2738.Regulatory.

Farina, G, R Lemaire, P Pancari, J Bayle, R L Widom, and R Lafyatis. 2009. “Cartilage Oligomeric Matrix Protein Expression in Systemic Sclerosis Reveals Heterogeneity of Dermal Fibroblast Responses to Transforming Growth Factor Beta.” *Annals of the Rheumatic Diseases* 68 (3): 435–41. doi:10.1136/ard.2007.086850.

Fu, Dun Jack, Calum Thomson, Declan P Lunny, Patricia J Dopping-Hepenstal, John a McGrath, Frances J D Smith, W H Irwin McLean, and Deena M Leslie Pedrioli. 2014. “Keratin 9 Is Required for the Structural Integrity and Terminal Differentiation of the Palmoplantar Epidermis.” *The Journal of Investigative Dermatology* 134 (3). Elsevier Masson SAS: 754–63. doi:10.1038/jid.2013.356.

Gardner, Humphrey, Jeffrey R Shearstone, Raj Bandaru, Tom Crowell, Matthew Lynes, Maria Trojanowska, Jaspreet Pannu, *et al*. 2006. “Gene Profiling of Scleroderma Skin Reveals Robust Signatures of Disease That Are Imperfectly Reflected in the Transcript Profiles of Explanted Fibroblasts.” *Arthritis and Rheumatism* 54 (6): 1961–73. doi:10.1002/art.21894.

Gerstmeier H, Gabrielli A, Meurer M, Brocks D, Braun-Falco O, Krieg T. 1988. “Levels of Type IV Collagen and Laminin Fragments in Serum from Patients with Progressive Systemic Sclerosis.” *J Rheumatol.* Jun;15(6):

Greenblatt, Matthew B, Jennifer L Sargent, Giuseppina Farina, Kelly Tsang, Robert Lafyatis, Laurie H Glimcher, Michael L Whitfield, and Antonios O Aliprantis. 2012. “Interspecies Comparison of Human and Murine Scleroderma Reveals IL-13 and CCL2 as Disease Subset-Specific Targets.” *The American Journal of Pathology* 180 (3). Elsevier Inc.: 1080–94. doi:10.1016/j.ajpath.2011.11.024.

Gupta, Rashmi, Richard Ahn, Kevin Lai, Elizabeth Mullins, Maya Debbaneh, Michelle Dimon, Sarah Arron, and Wilson Liao. 2016. “Landscape of Long Noncoding RNAs in Psoriatic and Healthy Skin.” *The Journal of Investigative Dermatology* 136 (3). The Authors: 603–9. doi:10.1016/j.jid.2015.12.009.

Harrow, Jennifer, Adam Frankish, Jose M Gonzalez, Electra Tapanari, Mark Diekhans, Felix Kokocinski, Bronwen L Aken, *et al*. 2012. “GENCODE: The Reference Human Genome Annotation for the ENCODE Project.” *Genome Research* 22 (9): 1760–74. doi:10.1101/gr.135350.111.

Hummers, Laura K, Amy Hall, Fredrick M Wigley, and Michael Simons. 2009. “Abnormalities in the Regulators of Angiogenesis in Patients with Scleroderma.” *The Journal of Rheumatology* 36 (3): 576–82. doi:10.3899/jrheum.080516.

Jia, Weijuan, Y I Feng, Andrew J Sanders, Eleri L Davies, and W E N G Jiang. 2016. “Phosphoinositide-3-Kinase Enhancers , PIKEs : Their Biological Functions and Roles in Cancer” 1110: 1103–9.

Johnson, Michael E., J. Matthew Mahoney, Jaclyn Taroni, Jennifer L. Sargent, Eleni Marmarelis, Ming-Ru Wu, John Varga, Monique E. Hinchcliff, and Michael L. Whitfield. 2015. “Experimentally-Derived Fibroblast Gene Signatures Identify Molecular Pathways Associated with Distinct Subsets of Systemic Sclerosis Patients in Three

Independent Cohorts." Edited by Stamatis-Nick Liossis. *Plos One* 10 (1): e0114017. doi:10.1371/journal.pone.0114017.

Katayama, S, Y Tomaru, T Kasukawa, K Waki, M Nakanishi, M Nakamura, H Nishida, *et al*. 2005. "Antisense Transcription in the Mammalian Transcriptome." *Science (New York, NY)* 309 (5740): 1564–66. doi:10.1126/science.1112009.

Kim, Dongwon, M Zulfiquer Hossain, Ashley Nieves, Lihong Gu, Tabetha S Ratliff, Seung Mi Oh, Angela Park, *et al*. 2016. "To Control Site-Specific Skin Gene Expression, Autocrine Mimics Paracrine Canonical Wnt Signaling and Is Activated Ectopically in Skin Disease." *The American Journal of Pathology* 186 (5): 1140–50. doi:10.1016/j.ajpath.2015.12.030.

Kimura, Tominori, Shiwen Jiang, Mikio Nishizawa, and Emi Yoshigai. 2013. "Stabilization of Human Interferon- a 1 mRNA by Its Antisense RNA," 1451–67. doi:10.1007/s00018-012-1216-x.

Kong, Lei, Yong Zhang, Zhi-Qiang Ye, Xiao-Qiao Liu, Shu-Qi Zhao, Liping Wei, and Ge Gao. 2007. "CPC: Assess the Protein-Coding Potential of Transcripts Using Sequence Features and Support Vector Machine." *Nucleic Acids Research* 35 (Web Server issue): W345-9. doi:10.1093/nar/gkm391.

Li, W, M Sun, C Zang, P Ma, J He, M Zhang, Z Huang, Y Ding, and Y Shu. 2016. "Upregulated Long Non-Coding RNA AGAP2-AS1 Represses LATS2 and KLF2 Expression through Interacting with EZH2 and LSD1 in Non-Small-Cell Lung Cancer Cells." *Cell Death & Disease* 7 (300): e2225. doi:10.1038/cddis.2016.126.

Lizio, Marina, Jayson Harshbarger, Hisashi Shimoji, Jessica Severin, Takeya Kasukawa, Serkan Sahin, Imad Abugessaisa, *et al*. 2015. "Gateways to the FANTOM5 Promoter Level Mammalian Expression Atlas." *Genome Biology* 16 (January): 22. doi:10.1186/s13059-014-0560-6.

Lofgren, Shane, Monique Hinchcliff, Mary Carns, Tammara Wood, Kathleen Aren, Esperanza Arroyo, Peggie Cheung, *et al*. 2016. "Integrated, Multicohort Analysis of Systemic Sclerosis Identifies Robust Transcriptional Signature of Disease Severity." *JCI Insight* 1 (21): e89073. doi:10.1172/jci.insight.89073.

Mahoney, J Matthew, Jaclyn Taroni, Viktor Martyanov, Tammara a Wood, Casey S Greene, Patricia a Pioli, Monique E Hinchcliff, and Michael L Whitfield. 2015. "Systems Level Analysis of Systemic Sclerosis Shows a Network of Immune and Profibrotic Pathways Connected with Genetic Polymorphisms." *PLoS Computational Biology* 11 (1): e1004005. doi:10.1371/journal.pcbi.1004005.

Mathes, Allison L, Romy B Christmann, Giuseppina Stifano, Alsya J Affandi, Timothy R D J Radstake, G Alessandra Farina, Cristina Padilla, Sarah McLaughlin, and Robert Lafyatis. 2014. "Global Chemokine Expression in Systemic Sclerosis (SSc): CCL19 Expression Correlates with Vascular Inflammation in SSc Skin." *Annals of the Rheumatic Diseases* 73 (10): 1864–72. doi:10.1136/annrheumdis-2012-202814.

Melé, Marta, and John L Rinn. 2016. "'Cat's Cradling' the 3D Genome by the Act of LncRNA Transcription." *Molecular Cell* 62 (5): 657–64. doi:10.1016/j.molcel.2016.05.011.

Messemaker, Tobias C, Tom W Huizinga, and Fina Kurreeman. 2015. "Immunogenetics of Rheumatoid Arthritis: Understanding Functional Implications." *Journal of Autoimmunity* 64 (July). Elsevier Ltd: 7–14. doi:10.1016/j.jaut.2015.07.007.

Milano, Ausra, Sarah a Pendergrass, Jennifer L Sargent, Lacy K George, Timothy H McCalmont, M Kari Connolly, and Michael L Whitfield. 2008. "Molecular Subsets in the Gene Expression Signatures of Scleroderma Skin." *PloS One* 3 (7): e2696. doi:10.1371/journal.pone.0002696.

Morimura, Sohshi, Tomonori Oka, Makoto Sugaya, and Shinichi Sato. 2016. "CX3CR1 Deficiency Attenuates Imiquimod-Induced Psoriasis-like Skin Inflammation with Decreased M1 Macrophages." *Journal of Dermatological Science* 82 (3). Japanese Society for Investigative Dermatology: 175–88. doi:10.1016/j.jdermsci.2016.03.004.

Pelechano, Vicent, and Lars M Steinmetz. 2013. "Gene Regulation by Antisense Transcription." *Nature Reviews. Genetics* 14 (12). Nature Publishing Group: 880–93. doi:10.1038/nrg3594.

Pendergrass, Sarah a, Raphael Lemaire, Ian P Francis, J Matthew Mahoney, Robert Lafyatis, and Michael L Whitfield. 2012. "Intrinsic Gene Expression Subsets of Diffuse Cutaneous Systemic Sclerosis Are Stable in Serial Skin Biopsies." *The Journal of Investigative Dermatology* 132 (5). Elsevier Masson SAS: 1363–73. doi:10.1038/jid.2011.472.

Peng, Huiyong, Yingzhao Liu, Jie Tian, Jie Ma, Xinyi Tang, Ke Rui, and Xinyu Tian. 2015. “The Long Noncoding RNA IFNG-AS1 Promotes T Helper Type 1 Cells Response in Patients with Hashimoto ’ S Thyroiditis.” *Nature Publishing Group*, no. December. Nature Publishing Group: 1–9. doi:10.1038/srep17702.

Raman, Dayanidhi, Tammy Sobolik-Delmaire, and Ann Richmond. 2011. “Chemokines in Health and Disease.” *Experimental Cell Research* 317 (5). Elsevier B.V.: 575–89. doi:10.1016/j.yexcr.2011.01.005.

Rau, Andrea, Guillemette Marot, and Florence Jaffrézic. 2014. “Differential Meta-Analysis of RNA-Seq Data from Multiple Studies,” 1–10.

Richards, Edward J, Gu Zhang, Zhu-peng Li, Jennifer Permuth-wey, Sridevi Challa, Yajuan Li, William Kong, *et al*. 2015. “Long Non-Coding RNAs (LncRNA) Regulated by Transforming Growth Factor (TGF) ” 290 (11): 6857–67. doi:10.1074/jbc.M114.610915.

Riente, L, B Marchini, M P Dolcher, A Puccetti, S Bombardieri, and P Migliorini. 1995. “Anti-Collagen Antibodies in Systemic Sclerosis and in Primary Raynaud â€™ S Phenomenon.” *Clin. Exp. Immunol.*, 354–59.

Rinn, John L, Jordon K Wang, Helen Liu, Kelli Montgomery, Matt van de Rijn, and Howard Y Chang. 2008. “A Systems Biology Approach to Anatomic Diversity of Skin.” *The Journal of Investigative Dermatology* 128 (4). Elsevier Masson SAS: 776–82. doi:10.1038/sj.jid.5700986.

Rinn JL, Chang HY. 2013. “Genome Regulation by Long Noncoding RNAs.” *Annu. Rev. Biochem.*, 8–12. doi:10.1146/annurev-biochem-051410-092902.Genome.

Rubio-Rivas, Manuel, Cristina Royo, Carmen Pilar Simeón, Xavier Corbella, and Vicent Fonollosa. 2014. “Mortality and Survival in Systemic Sclerosis: Systematic Review and Meta-Analysis.” *Seminars in Arthritis and Rheumatism* 44 (2). Elsevier: 208–19. doi:10.1016/j.semarthrit.2014.05.010.

Severin, Jessica, Marina Lizio, Jayson Harshbarger, Hideya Kawaji, Carsten O Daub, Yoshihide Hayashizaki, Nicolas Bertin, and Alistair R R Forrest. 2014. “Interactive Visualization and Analysis of Large-Scale Sequencing Datasets Using ZENBU.” *Nature Biotechnology* 32 (3): 217–19. doi:10.1038/nbt.2840.

Subcommittee for scleroderma criteria, 1980. 1980. “Preliminary Criteria for the Classification of Systemic Sclerosis (Scleroderma).” *Arthritis and Rheumatism* 23 (5): 581–90. doi:10.1002/art.1780230510.

Villegas, Victoria E, and Peter G Zaphiropoulos. 2015. “Neighboring Gene Regulation by Antisense Long Non-Coding RNAs.” *International Journal of Molecular Sciences* 16 (2): 3251–66. doi:10.3390/ijms16023251.

Werner, Andreas. 2013. “Biological Functions of Natural Antisense Transcripts.” *BMC Biology*, 2–4.

Whitfield, Michael L, Deborah R Finlay, John Isaac Murray, Olga G Troyanskaya, Jen-tsan Chi, Alexander Pergamenschikov, Timothy H McCalmont, Patrick O Brown, David Botstein, and M Kari Connolly. 2003. “Systemic and Cell Type-Specific Gene Expression Patterns in Scleroderma Skin.” *Proceedings of the National Academy of Sciences of the United States of America* 100 (21): 12319–24. doi:10.1073/pnas.1635114100.

Xu, Zhongping, Yufang Zheng, Yufei Zhu, Xiangyin Kong, and Landian Hu. 2011. “Evidence for OTUD-6B Participation in B Lymphocytes Cell Cycle after Cytokine Stimulation.” *PloS One* 6 (1): e14514. doi:10.1371/journal.pone.0014514.

Yang, Li, Michael O Duff, Brenton R Graveley, Gordon G Carmichael, and Ling-ling Chen. 2011. “Genomewide Characterization of Non-Polyadenylated RNAs,” 1–14.

Zhu, Yunjuan, Yuanjun Wu, Jae I Kim, Zhimin Wang, Yehia Daaka, and Zhongzhen Nie. 2009. “Arf GTPase-Activating Protein AGAP2 Regulates Focal Adhesion Kinase Activity and Focal Adhesion Remodeling.” *The Journal of Biological Chemistry* 284 (20): 13489–96. doi:10.1074/jbc.M900469200.