Antisense Long Non-Coding RNAs Are Deregulated in Skin Tissue of Patients with Systemic Sclerosis

Messemaker, T.C.; Chadli, L.; Cai, G.S.; Goelela, V.S.; Boonstra, M.; Dorjee, A.L.; ... ; Kurreeman, F.

Antisense Long Non-Coding RNAs Are Deregulated in Skin Tissue of Patients with Systemic Sclerosis

Tobias C. Messemaker[1,2], Loubna Chadli[3], Guoshuai Cai[4], Varshna S. Goelela[3], Maaike Boonstra[1], Annemarie L. Dorjée[1], Stefan N. Andersen[3], Harald M.M. Mikkers[2], Peter van 't Hof[5], Hailiang Mei[5], Oliver Distler[6], Harmen H.M. Draisma[7], Michael E. Johnson[4], Nicole M. Orzechowski[4], Robert W. Simms[8], Rene E.M. Toes[1], Jamil Aarbiou[3], Tom W. Huizinga[1], Michael L. Whitfield[4], Jeroen DeGroot[3], Jeska de Vries-Bouwstra[1] and Fina Kurreeman[1]

Systemic sclerosis is an autoimmune disease characterized by fibrosis of skin and multiple organs of which the pathogenesis is poorly understood. We studied differentially expressed coding and non-coding genes in relation to systemic sclerosis pathogenesis with a specific focus on antisense non-coding RNAs. Skin biopsy–derived RNAs from 14 early systemic sclerosis patients and six healthy individuals were sequenced with ion-torrent and analyzed using DEseq2. Overall, 4,901 genes with a fold change >1.5 and a false discovery rate <5% were detected in patients versus controls. Upregulated genes clustered in immunologic, 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 *P*-value ($P_{Bonf} < 0.0023$, $P_{combined} = 1.1 \times 10^{-9}$, 1.4×10^{-8}, 1.7×10^{-6}, respectively). In this study, we highlight that together with coding genes, (antisense) long non-coding RNAs are deregulated in skin tissue of systemic sclerosis patients suggesting a novel class of genes involved in pathogenesis of systemic sclerosis.

Journal of Investigative Dermatology (2018) **138,** 826–835; doi:10.1016/j.jid.2017.09.053

INTRODUCTION

Systemic sclerosis (SSc) is a heterogeneous complex autoimmune disease affecting connective tissues. Its pathogenesis remains elusive, but patients harbor 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 5 years after diagnosis in SSc (Rubio-Rivas et al., 2014).

In order to better understand the processes involved in SSc pathophysiology, several groups have performed gene expression studies in the peripheral blood and skin of SSc patients (Gardner et al., 2006; Milano et al., 2008; Pendergrass et al., 2012; 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 that include fibrosis-related pathways involved in skin thickening (transforming growth factor [TGF] β–related genes, collagen genes), as well as immunologic and keratin-related pathways (IFN genes, activated macrophage genes, chemokine-related genes, and keratin genes) (Assassi et al., 2015; Gardner et al., 2006; Mahoney et al., 2015; Mathes et al., 2014). These studies were all performed using microarrays, and focused on the identification of protein coding genes and pathways that are differently regulated in SSc and, as a consequence, are 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.

[1]*Department of Rheumatology, Leiden University Medical Center, Leiden, The Netherlands;* [2]*Department of Molecular Cell Biology, Leiden University Medical Center, Leiden, The Netherlands;* [3]*Charles River Nederland B.V., Leiden, The Netherlands;* [4]*Department of Genetics, Geisel School of Medicine at Dartmouth, Hanover, New Hampshire, USA;* [5]*Sequencing Analysis Support Core, Leiden University Medical Center, Leiden, The Netherlands;* [6]*Department of Rheumatology, University Hospital Zurich, Zurich, Switzerland;* [7]*Department of Human Genetics, Leiden University Medical Center, Leiden, The Netherlands; and* [8]*Boston University School of Medicine, Boston, Massachusetts, USA*

Correspondence: Tobias C. Messemaker, Departments of Rheumatology and Molecular Cell Biology, Leiden University Medical Center, 9600 2300 RC Leiden, The Netherlands. E-mail: TCMessemaker@lumc.nl

Abbreviations: lncRNA, long non-coding RNA; SSc, systemic sclerosis; TGF, transforming growth factor

Received 3 March 2017; revised 31 August 2017; accepted 23 September 2017; accepted manuscript published online 24 November 2017

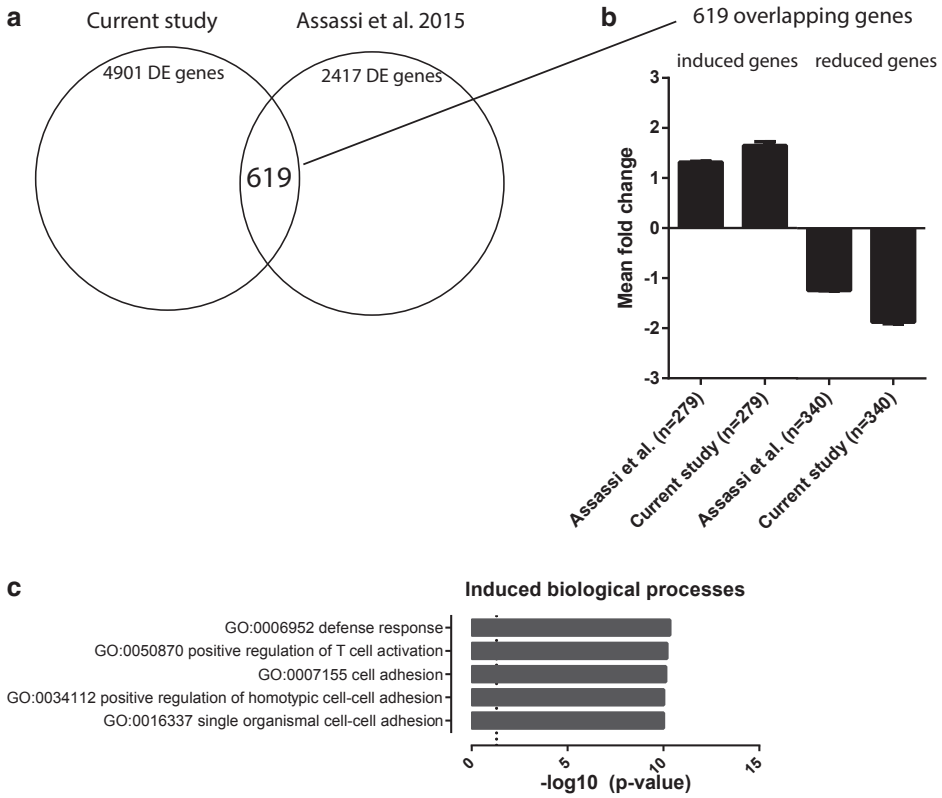

Figure 1. DE genes overlap with a previous microarray study and reveal consistent deregulated pathways. (**a**) Venn diagram comparing DE genes in SSc patients versus controls from the current study (n = 4,901 DE genes) with a microarray study from Assassi et al. (2015) (n = 2,417 DE genes). (**b**) Directionality of 619 consistently deregulated genes from the two studies displayed as mean fold change (mean ± standard error). Genes up- or downregulated from Assassi et al. (2015) 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 Gene Ontology (GO) terms enriched using genes that are upregulated in SSc patients from the two studies.

lncRNAs represent an important layer of genome regulation and their role in the context of SSc is currently unknown. lncRNAs are transcripts that are >200 nucleotides in length and come in diverse flavors, including antisense RNAs, long intergenic non-coding RNAs, 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 (Melé and Rinn 2016; Rinn and Chang, 2012). Specifically, antisense RNAs can influence RNA levels of their sense counterpart (Chan et al., 2015; Derrien et al., 2012; Faghihi and Wahlestedt 2010; Kimura et al., 2013; Peng et al., 2015). 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 TGFβ–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 immunologic, 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; 4,901 genes were DE with a minimum fold change of 1.5 and false discovery rate–adjusted *P*-value <0.05 (Supplementary Material S1 online). Hierarchical clustering on the basis of these DE genes separates patients from healthy controls with the exception of one patient that displayed a normal-like expression pattern (Supplementary Figure S1 online). Pathway analysis of overexpressed genes shows an enrichment in the immune response, cell activation, and keratinization pathways (Supplementary Material S3 online). Cross comparison with DE genes from a recent publication by Assassi et al. (2015) found a small, highly consistent (>96%) overlap with the most prominent common pathways belonging to the immunologic and cell adhesion–related processes (Figure 1a–1c, Supplementary Material S4 online).

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 signaling, collagen, keratin, IFN, alternative macrophage activation genes, and chemokines (Figure 2, Supplementary Material S2 online and Supplementary Material S5 online). Similar to our Gene

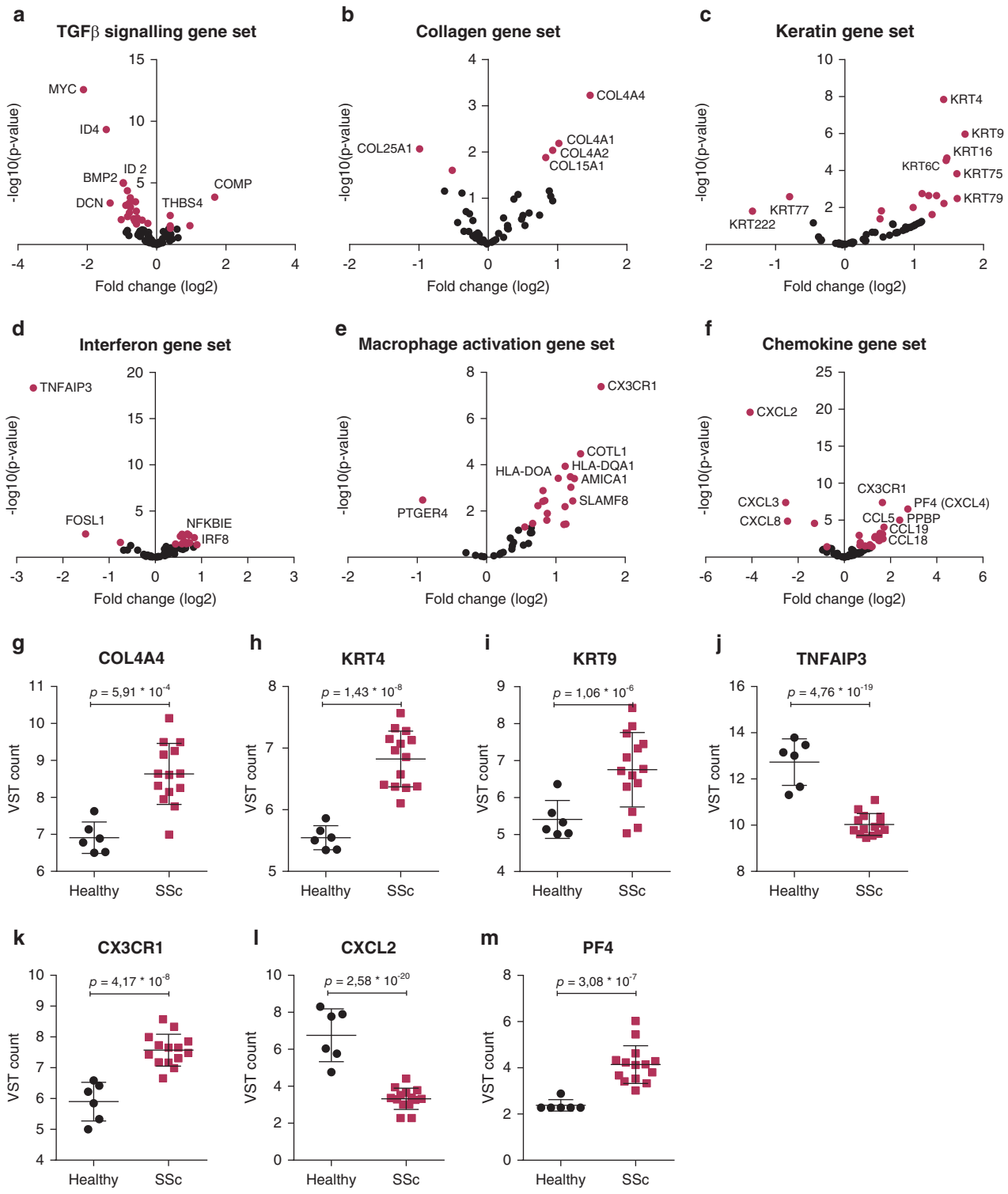

Figure 2. Analysis of DE genes of specific SSc gene sets. Volcano plots showing differential expression within the six gene sets: transforming growth factor−β signaling (**a**, n = 86 genes), collagen (**b**, n = 46 genes), keratin (**c**, n = 76 genes), IFN (**d**, n = 50 genes), alternative macrophage 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 ± standard deviation of each group is depicted in the graphs.

Ontology 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 of 86 TGFβ signaling genes were increased significantly (Figure 2a). On the other hand, TGFβ-gene COMP was found increased in patients as similar to previous reports (Assassi et al., 2015; Farina et al., 2009; Gardner et al., 2006). Moreover, many collagen- and keratin-associated genes are

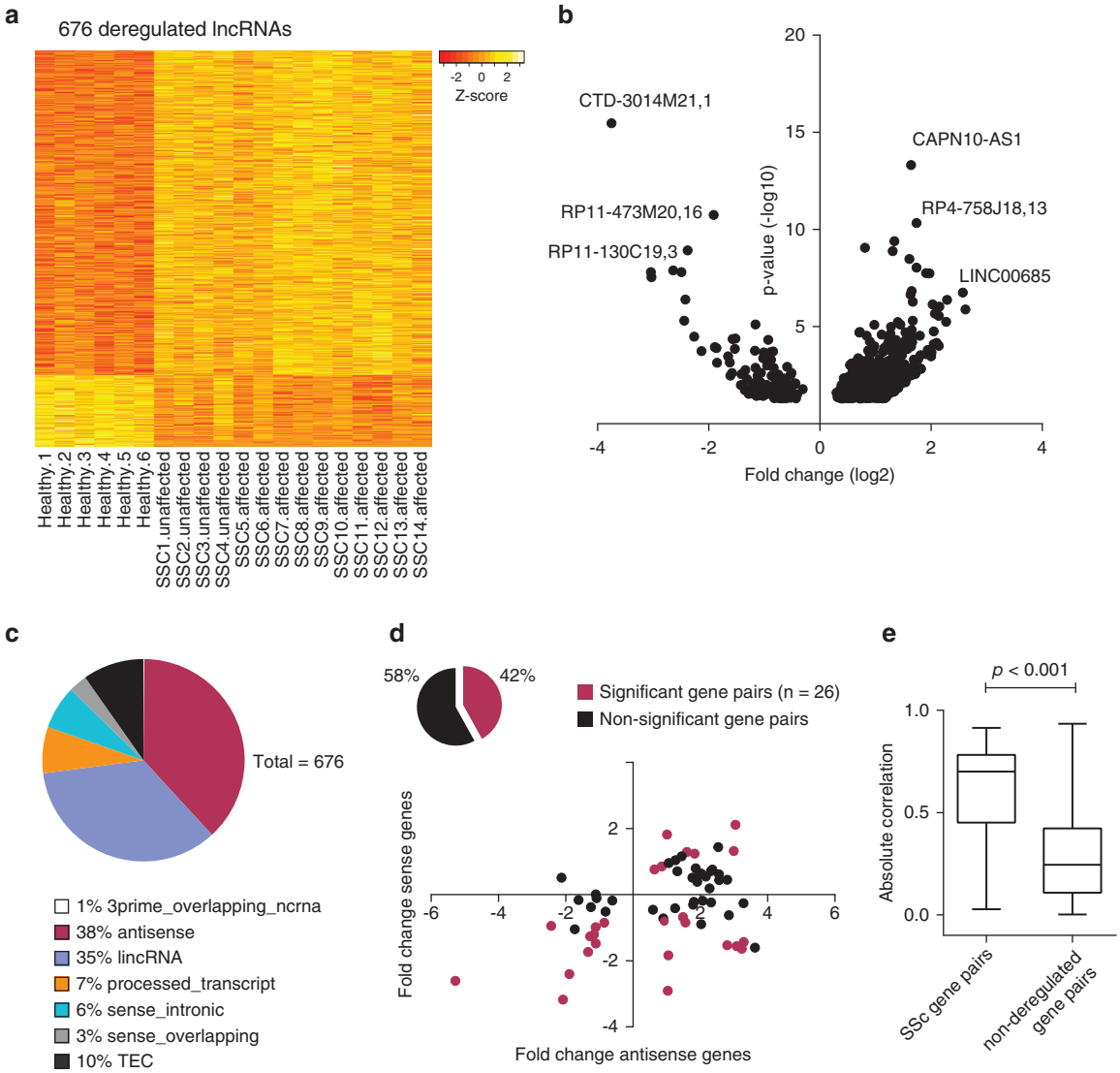

Figure 3. DE lncRNAs in SSc patients in comparison with healthy controls. (**a**) Heatmap depicting the *Z*-scores of 676 deregulated lncRNAs. Red color indicates low expression and the yellow color indicates high expression. (**b**) Volcano plot showing top deregulated lncRNAs by fold change (log2) 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.

increased significantly in patients (Figure 2b, 2c). Also, 33 of 97 genes from the IFN and macrophage gene sets were increased significantly in SSc patients (Figure 2d, 2e), indicating an increased inflammatory gene signature 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). Because 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 S2 online). Besides these observations, several (previously unreported, to our knowledge) genes including COL4A4, Keratin 4 and 9, TNFAIP3, CX3CR1, CXCL2, and PF4 were strongly deregulated in SSc patients (Figure 2b–2k, Supplementary Table S1 online).

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 15,941 annotated lncRNAs, 4,171 were expressed in our skin biopsies. Six hundred and seventy-six lncRNAs were DE (false discovery rate–adjusted *P*-value <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 Material S6 (online). Out of 676 lncRNA genes, 122 were decreased, while the expression of 554 genes was increased in SSc patients compared to healthy controls. Interestingly, clustering analysis using different

Table 1. Replication of 22 antisense genes in an independent RNA-seq dataset[1]

	Dataset 1			Dataset 2		
Gene	Log2FC	*P*-Value	FDR	Log2FC	*P*-Value	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

Abbreviations: FC, fold change; FDR, false discovery rate−adjusted *P*-value; NA, not available.
[1]The table includes fold changes (Log2FC) and *P*-values 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).

selection criteria of lncRNAs all displayed a pattern in which nonclinically active patients clustered within the patient population and separate from controls (Supplementary Figure S3 online). In total, 348 lncRNAs displayed >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). Antisense lncRNAs have recently been described to have important regulatory roles on their coding gene counterparts expressed in the sense direction (Katayama et al., 2005; Pelechano and Steinmetz 2013; Villegas and Zaphiropoulos 2015; Werner, 2013). The relevance of the antisense genes in our dataset 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 gene pairs). Close proximity of antisense genes with sense genes have been linked to co-expression and co-regulation within such a sense−antisense gene pair (Katayama et al., 2005; Villegas and Zaphiropoulos 2015). Out of 257 DE antisense genes, 62 have an annotated sense gene. Interestingly, an important proportion (26 out of 62) of these sense-antisense gene pairs includes both a significant DE antisense gene and a significant DE sense gene (false discovery rate <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 that 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). These data indicate that the identified antisense genes are either co-expressed 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, six controls, Whitfield et al, unpublished data). Four of the 26 genes were not present due to low expression in the independent dataset and were excluded from the analysis. Twelve of 22 genes follow the same direction of association in both datasets (Table 1). 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−4c). Verification using a second experimental approach confirmed that these three genes are significantly deregulated ($P < 0.01$) (Supplementary Figure S4 online).

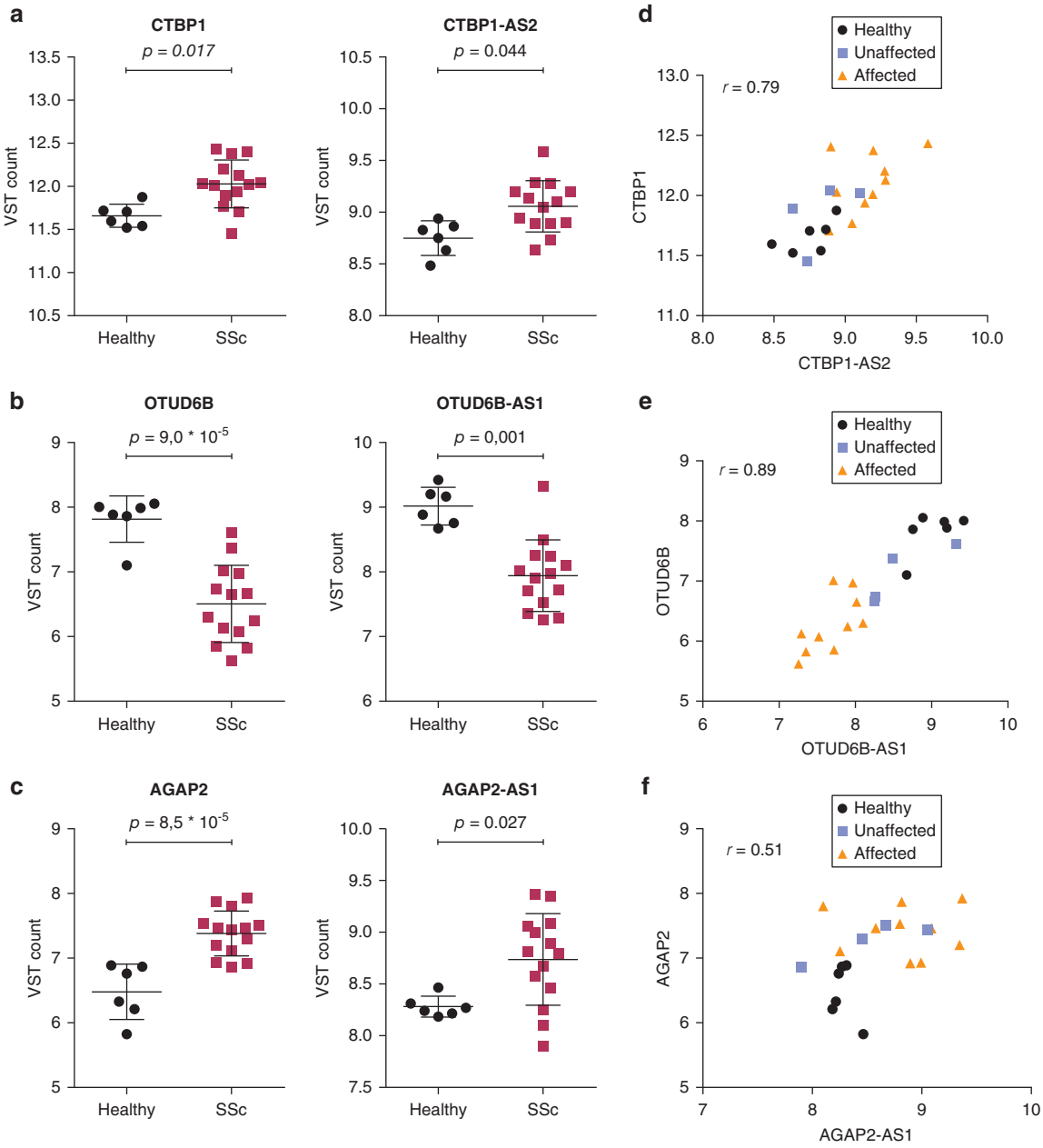

Figure 4. Top three replicated antisense genes show strong correlation with their sense coding gene. (**a**–**c**) VST count values of top three replicated sense–antisense gene pairs: CTBP1 (**a**), OTUD6B (**b**), and AGAP2 (**c**) in SSc patients (n = 14) and controls (n = 6) *P*-values are Benjamini-Hochberg–corrected 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 S5 online). 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 to that 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

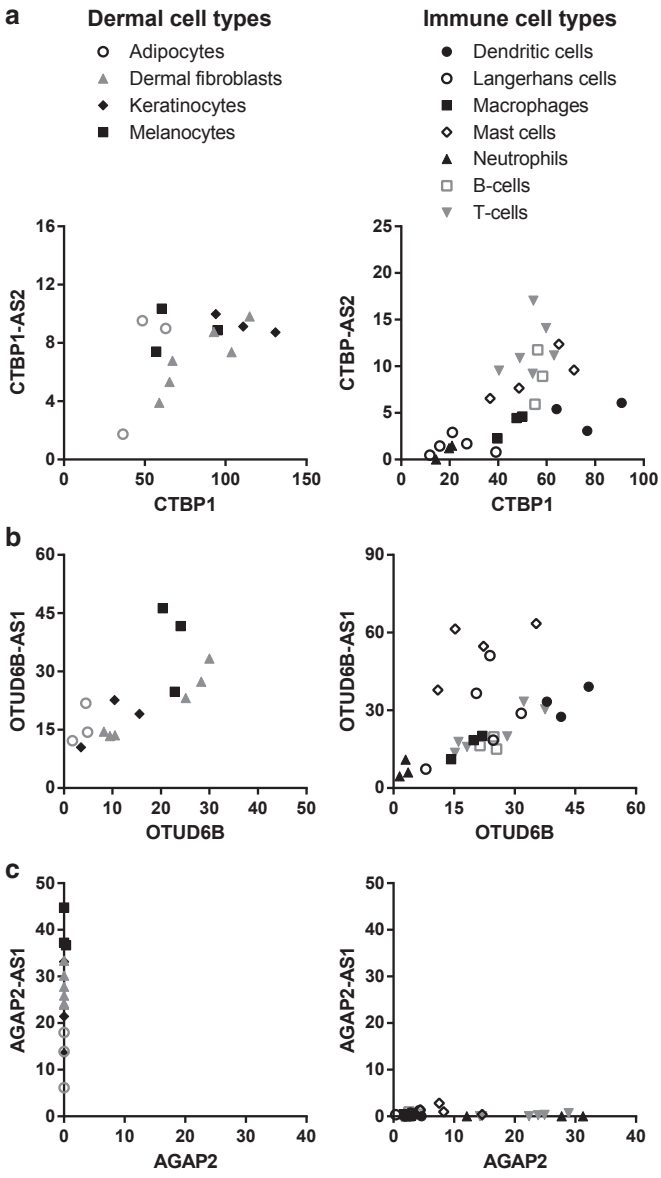

Figure 5. Cell-type−specific expression of sense−antisense gene pairs in dermal and immune cells. Expression levels for CTBP1 and CTBP1-AS2 (**a**), OTUD6B and OTUD6B-AS1 (**b**) and AGAP2 and AGAP2-AS1 (**c**) in dermal and immune cell types. Expression values are shown as transcripts per million for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

Figure 6 online), 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, where 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 first 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 three deregulated antisense genes included CTBP1-AS2, OTUD6B-AS1, and AGAP2-AS1, and these findings were replicated in an independent dataset and further validated by quantitative PCR. 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. Most lncRNAs are not yet available on microarrays and are therefore missed in the available datasets that were investigating SSc deregulated genes. More importantly, association of lncRNAs with inflammatory diseases like rheumatoid arthritis, diabetes, and psoriasis are being reported increasingly, highlighting their potential role in disease mechanisms (Gupta et al., 2016; Messemaker et al., 2015). Here, we identified 676 lncRNAs that are deregulated in skin from SSc patients compared to healthy individuals. A large proportion of the deregulated lncRNAs belonged to the antisense RNA category. Antisense RNAs that 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 sense−antisense gene pairs that 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 dataset. The top three deregulated antisense genes included CTBP1-AS2, OTUD6B-AS1, and AGAP2-AS1. OTUD6B is a deubiquitinating enzyme about 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 that 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 these lncRNAs might also play important roles in SSc development and they 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 IFN/inflammatory signature in early SSc patients (Johnson et al., 2015). In contrast to previous research, a clear TGFβ signal did not come forward from our gene list, despite the increase of fibrosis-related genes like 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) overlap, suggesting that consistent deregulated genes exist despite SSc well-known disease heterogeneity and large differences in mean age and disease duration of patients between both studies (Supplementary Table S2 online). Moreover, an additional comparison with 415 genes obtained from a meta-analysis performed by Lofgren et al. (2016) show that 159 genes overlapped (38%).

We investigated specific SSc gene sets in more detail to identify genes deregulated in early SSc patients. Our study reports several coding genes that have not been highlighted previously 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 et al., 1988) and COL4 autoantibodies have been found in 31% of SSc patients, indicating 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 signaling, its downregulation would be suggestive of 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 et al., 2011), was reduced and might influence vascular repair within the 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 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 not, to our knowledge, been described in the 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 Supplementary Material (online).

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 of the American Rheumatism Association Diagnostic and Therapeutic Criteria Committee, 1980). Patient characteristics can be found in Supplementary Table S3 (online). Institutional Review Board approval and written informed consent were obtained before patients entered this study. Two 4-mm skin biopsies were taken; in 10 patients the skin biopsy came from a clinically affected area and in 4 patients the skin was unaffected locally. Skin biopsies from healthy individuals were commercially sourced (Tissue Solutions, Glasgow, UK), came from surgeries of arm and leg and were age and sex-matched.

Transcriptome characterization 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, Leiden, 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 European Genome-phenome Archive database under nr: EGAO00000000316 (https://www.ebi.ac.uk/ega/organisations/EGAO00000000316).

Publicly available gene expression datasets and gene sets analysis

DE genes (false discovery rate < 0.05, fold change > 1.5) were compared with a publicly available dataset obtained from Assassi et al. (2015). DE genes were investigated via Gene Ontology term analysis using Toppgene, version 23, and in specific gene sets. Gene sets were obtained from Hugo Gene Nomenclature Committee or by

additionally compiled SSc gene sets from alternative sources. Genes in the IFN and alternative macrophage activation signature were obtained from Mahoney et al. (2015). Genes involved in TGFβ signaling were obtained from the Broad Institute (Cambridge, MA). All genes included in these gene sets are outlined in Supplementary Material S1.

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 six healthy individuals. An overall combined *P*-value was calculated using Fisher's exact test. The top three sense and antisense genes were visualized in integrative genome viewer to ensure strand specificity and non-overlapping reads (Supplementary Figure S7 online). 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 stabilized transformed counts by Spearman rank test.

CONFLICT OF INTEREST

Oliver Distler had consultancy relationship and/or has received research funding from Actelion, Bayer, Boehringer Ingelheim, ChemomAb, espeRare foundation, Genentech/Roche, GSK, Inventiva, Italfarmaco, Lilly, medac, MedImmune, Mitsubishi Tanabe Pharma, Novartis, Pfizer, Sanofi, Sinoxa, and UCB in the area of potential treatments for scleroderma and its complications. In addition, Oliver Distler has a patent mir-29 for the treatment of systemic sclerosis licensed.

ACKNOWLEDGMENTS

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 United Nations Educational, Scientific and Cultural Organization–L'Oreal for Women in Science Fellowship, Marie Curie FP7 Outgoing Fellowship and a Leiden University Medical Center Research Fellowship. Oliver Distler has support by the Swiss National foundation (project number 310030_166259) for studying the role of lncRNAs in SSc.

AUTHOR CONTRIBUTIONS

All authors were involved in drafting the article or revising it critically for important intellectual content, and all authors approved the final version to be published. TCM had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. TCM, LC, MB, OD, TWH, MLW, REMT, JA, JD, JVB, and FK designed the study; LC, MB, MLW, JD, and JVB collected patient and control material and clinical data; TCM, LC, GC, VSG, MB, AD, MEJ, NMO, RWS, MLW, JVB, and FK performed experiments and acquired the data; TCM, LC, GC, VSG, MB, SNA, HMM, PH, HM, OD, HHMD, TWH, MLW, REMT, JA, JD, JVBm and FK were involved in analysis and interpretation of the data.

SUPPLEMENTARY MATERIAL

Supplementary material is linked to the online version of the paper at www.jidonline.org, and at https://doi.org/10.1016/j.jid.2017.09.053.

REFERENCES

Assassi S, Swindell WR, Wu M, Tan FD, Khanna D, Furst DE, Tashkin DP, et al. Dissecting the heterogeneity of skin gene expression patterns in systemic sclerosis. Arthritis Rheum 2015;67:3016–26.

Avouac J, Cagnard N, Distler JH, Schoindre Y, Ruiz B, Couraud PO, et al. Insights into the pathogenesis of systemic sclerosis based on the gene expression profile of progenitor-derived endothelial cells. Arthritis Rheum 2011;63:3552–62.

Batista PJ, Chang HY. Long noncoding RNAs: cellular address codes in development and disease. Cell 2013;152:1298–307.

Bosch FX, Ouhayoun JP, Bader BL, Collin C, Grund C, Lee I, et al. Extensive changes in cytokeratin expression patterns in pathologically affected human gingiva. Virchows Arch B Cell Pathol Incl Mol Pathol 1989;58:59–77.

Boxer LD, Barajas B, Tao S, Zhang J, Khavari PA. ZNF750 interacts with KLF4 and RCOR1, regulators to repress epidermal progenitor genes and induce differentiation genes. Genes Dev 2014;28:2013–26.

Chan J, Atianand M, Jiang Z, Carpenter S, Aiello D, Elling R, et al. Cutting edge: a natural antisense transcript, AS-IL1α, controls inducible transcription of the proinflammatory cytokine IL-1α. J Immunol 2015;195:1359–63.

Deng H, Li F, Li H, Deng Y, Liu J, Wang D, et al. CtBP1 overexpression in keratinocytes perturbs skin homeostasis. J Invest Dermatol 2014;134:1323–31.

Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, Tilgner H, et al. The GENCODE v7 catalog of human long noncoding RNAs?: Analysis of their gene structure, evolution, and expression. Genome Res 2012;22:1775–89.

Dieudé P, Guedj M, Wipff J, Ruiz B, Riemekasten G, Matucci-Cerinic M. Association of the TNFAIP3 rs5029939 variant with systemic sclerosis in the European Caucasian population. Ann Rheum Dis 2010;69:1958–64.

Esteller M. Non-coding RNAs in human disease. Nat Rev Genet 2011;12:861–74.

Faghihi MA, Wahlestedt C. Regulatory roles of natural antisense transcripts. Nat Rev Mol Cell Biol 2010;10:637–43.

Farina G, Lemaire R, Pancari P, Bayle J, Widom RL, Lafyatis R. Cartilage oligomeric matrix protein expression in systemic sclerosis reveals heterogeneity of dermal fibroblast responses to transforming growth factor beta. Ann Rheum Dis 2009;68:435–41.

Fu DJ, Thomson C, Lunny DP, Dopping-Hepenstal PJ, McGrath JA, Smith FJD, et al. Keratin 9 is required for the structural integrity and terminal differentiation of the palmoplantar epidermis. J Invest Dermatol 2014;134:754–63.

Gardner H, Shearstone JR, Bandaru R, Crowell T, Lynes M, Trojanowska M, et al. Gene profiling of scleroderma skin reveals robust signatures of disease that are imperfectly reflected in the transcript profiles of explanted fibroblasts. Arthritis Rheum 2006;54:1961–73.

Gerstmeier H, Gabrielli A, Meurer M, Brocks D, Braun-Falco O, Krieg T. Levels of type IV collagen and laminin fragments in serum from patients with progressive systemic sclerosis. J Rheumatol 1988;15:969–72.

Greenblatt MB, Sargent JL, Farina G, Tsang K, Lafyatis R, Glimcher LH, et al. Interspecies comparison of human and murine scleroderma reveals IL-13 and CCL2 as disease subset-specific targets. J Pathol 2012;180:1080–94.

Gupta R, Ahn R, Lai K, Mullins E, Debbaneh M, Dimon M, et al. Landscape of long noncoding RNAs in psoriatic and healthy skin. J Invest Dermatol 2016;136:603–9.

Harrow J, Frankish A, Gonzalez JM, Tapanari E, Diekhans M, Kokocinski F, et al. GENCODE: the reference human genome annotation for the ENCODE Project. Genome Res 2012;22:1760–74.

Hummers LK, Hall A, Wigley FM, Simons M. Abnormalities in the regulators of angiogenesis in patients with scleroderma. J Rheumatol 2009;36:576–82.

Jia W, Feng YI, Sanders AJ, Davies EL, Jiang WG. Phosphoinositide-3-kinase enhancers, PIKEs?: their biological functions and roles in cancer. Anticancer Res 2016;36:1103–9.

Johnson ME, Mahoney JM, Taroni J, Sargent JL, Marmarelis E, Wu MR, et al. Experimentally-derived fibroblast gene signatures identify molecular pathways associated with distinct subsets of systemic sclerosis patients in three independent cohorts. PLoS One 2015;10:e0114017.

Katayama S, Tomaru Y, Kasukawa T, Waki K, Nakanishi M, Nakamura M, et al. Antisense transcription in the mammalian transcriptome. Science 2005;309(5740):1564–6.

Kim D, Hossain MZ, Nieves A, Gu L, Ratliff TS, Mi Oh S, et al. To control site-specific skin gene expression, autocrine mimics paracrine canonical Wnt signaling and is activated ectopically in skin disease. Am J Pathol 2016;186:1140–50.

Kimura T, Jiang S, Nishizawa M, Yoshigai E, Hashimoto I, Nishikawa M, et al. Stabilization of human interferon-α 1 mRNA by its antisense RNA. Cell Mol Life Sci 2013;70:1451–67.

Kong L, Zhang Y, Ye ZQ, Liu XQ, Zhao SQ, Wei L, et al. CPC: Assess the protein-coding potential of transcripts using sequence features and support vector machine. Nucleic Acids Res 2007;35(Web Server issue): W345–9.

Li W, Sun M, Zang C, Ma P, He J, Zhang M, et al. 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 Dis 2016;7:e2225.

Lizio M, Harshbarger J, Shimoji H, Severin J, Kasukawa T, Sahin S, et al. Gateways to the FANTOM5 promoter level mammalian expression atlas. Genome Biol 2015;16:22.

Lofgren S, Hinchcliff M, Carns M, Wood T, Aren K, Arroyo E, et al. Integrated, multicohort analysis of systemic sclerosis identifies robust transcriptional signature of disease severity. JCI Insight 2016;1: e89073.

Mahoney JM, Taroni J, Martyanov V, Wood TA, Greene CS, Pioli PA, et al. Systems level analysis of systemic sclerosis shows a network of immune and profibrotic pathways connected with genetic polymorphisms. PLoS Comput Biol 2015;11:e1004005.

Mathes AL, Christmann RB, Stifano G, Affandi AJ, Radstake TR, Farina GA, et al. Global chemokine expression in systemic sclerosis (SSc): CCL19 expression correlates with vascular inflammation in SSc skin. Ann Rheum Dis 2014;73:1864–72.

Melé M, Rinn JL. 'Cat's cradling' the 3D genome by the act of LncRNA transcription. Mol Cell 2016;62:657–64.

Messemaker TC, Huizinga TW, Kurreeman F. Immunogenetics of rheumatoid arthritis: understanding functional implications. J Autoimmun 2015;64: 74–81.

Milano A, Pendergrass SA, Sargent JL, George LK, McCalmont TH, Connolly MK, et al. Molecular subsets in the gene expression signatures of scleroderma skin. PLoS One 2008;3:e2696.

Morimura S, Oka T, Sugaya M, Sato S. CX3CR1 deficiency attenuates imiquimod-induced psoriasis-like skin inflammation with decreased M1 macrophages. PLoS One 2016;82:175–88.

Pelechano V, Steinmetz LM. Gene regulation by antisense transcription. Nat Rev Genet 2013;14:880–93.

Pendergrass SA, Lemaire R, Francis IP, Mahoney JM, Lafyatis R, Whitfield ML. Intrinsic gene expression subsets of diffuse cutaneous systemic sclerosis are stable in serial skin biopsies. J Invest Dermatol 2012;132:1363–73.

Peng H, Liu Y, Tian J, Ma J, Tang X, Rui K, et al. The long noncoding RNA IFNG-AS1 promotes T helper type 1 cells response in patients with Hashimoto's thyroiditis. Sci Rep 2015. 4;5:17702.

Raman D, Sobolik-Delmaire T, Richmond A. Chemokines in health and disease. Exp Cell Res 2011;317:575–89.

Rau A, Marot G, Jaffrézic F. Differential meta-analysis of RNA-Seq data from multiple studies. BMC Bioinform 2014;15:91.

Richards EJ, Zhang G, Li ZP, Permuth-Wey J, Challa S, Li Y. Long non-coding RNAs (LncRNA) regulated by transforming growth factor (TGF) β: LncRNA-hit-mediated TGFβ-induced epithelial to mesenchymal transition in mammary epithelia. J Biol Chem 2015;290: 6857–67.

Riente L, Marchini B, Dolcher MP, Puccetti A, Bombardieri S, Migliorini P. Anti-collagen antibodies in systemic sclerosis and in primary Raynaud's phenomenon. Clin Exp Immunol 1995;102:354–9.

Rinn JL, Wang JK, Liu H, Montgomery K, van de Rijn M, Chang HY. A systems biology approach to anatomic diversity of skin. J Invest Dermatol 2008;128:776–82.

Rinn JL, Chang HY. Genome regulation by long noncoding RNAs. Ann Rev Biochem 2012;81:145–66.

Rubio-Rivas M, Royo C, Simeón CP, Corbella X, Fonollosa V. Mortality and survival in systemic sclerosis: systematic review and meta-analysis. Semin Arthritis Rheum 2014;44:208–19.

Severin J, Lizio M, Harshbarger J, Kawaji H, Daub CO, Hayashizaki Y, et al. Interactive visualization and analysis of large-scale sequencing datasets using ZENBU. Nat Biotechnol 2014;32:217–9.

Subcommittee for Scleroderma Criteria of the American Rheumatism Association Diagnostic and Therapeutic Criteria Committee. Preliminary criteria for the classification of systemic sclerosis (scleroderma). Arthritis Rheum 1980;23:581–909.

van Bon L, Affandi AJ, Broen J, Christmann RB, Marijnissen RJ, Stawski L, et al. Proteome-wide analysis and CXCL4 as a biomarker in systemic sclerosis. N Engl J Med 2014;370:433–43.

Villegas VE, Zaphiropoulos PG. Neighboring gene regulation by antisense long non-coding RNAs. Int J Mol Sci 2015;16:3251–66.

Werner A. Biological functions of natural antisense transcripts. BMC Biol 2013;11:31.

Whitfield ML, Finlay DR, Murray JI, Troyanskaya OG, Chi JT, Pergamenschikov A, et al. Systemic and cell type-specific gene expression patterns in scleroderma skin. Proc Natl Acad Sci USA 2003;100: 12319–24.

Xu Z, Zheng Y, Zhu Y, Kong X, Hu L. Evidence for OTUD-6B participation in B lymphocytes cell cycle after cytokine stimulation. PLoS One 2011;6: e14514.

Yang L, Duff MO, Graveley BR, Carmichael GG, Chen LL. Genomewide characterization of non-polyadenylated RNAs. Genome Biol 2011;12: R16.

Zhu Y, Wu Y, Kim JI, Wang Z, Daaka Y, Nie Z. Arf GTPase-activating protein AGAP2 regulates focal adhesion kinase activity and focal adhesion remodeling. J Biol Chem 2009;284:13489–96.