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

Tobias C. Messemaker1,2, Loubna Chadli3, Guoshuai Cai4, Varshna S. Goelela3, Maaike Boonstra1, Annemarie L. Dorjée1, Stefan N. Andersen3, Harald M.M. Mikkers2, Peter van ‘t Hof5, Hailiang Mei5, Oliver Distler6, Harmen H.M. Draisma7, Michael E. Johnson4, Nicole M. Orzechowski4, Robert W. Simms8, Rene E.M. Toes1, Jamil Aarbiou3, Tom W. Huizinga1, Michael L. Whitfield4, Jeroen DeGroot3, Jeska de Vries-Bouwstra1 and Fina Kurreeman1

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 ($P_{\text{Bonf}} < 0.0023$, $P_{\text{combined}} = 1.1 \times 10^{-9}, 1.4 \times 10^{-8}, 1.7 \times 10^{-6}$, 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.


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.

1Department of Rheumatology, Leiden University Medical Center, Leiden, The Netherlands; 2Department of Molecular Cell Biology, Leiden University Medical Center, Leiden, The Netherlands; 3Charles River Nederland B.V., Leiden, The Netherlands; 4Department of Genetics, Geisel School of Medicine at Dartmouth, Hanover, New Hampshire, USA; 5Sequencing Analysis Support Core, Leiden University Medical Center, Leiden, The Netherlands; 6Department of Rheumatology, University Hospital Zurich, Zurich, Switzerland; 7Department of Human Genetics, Leiden University Medical Center, Leiden, The Netherlands; and 8Boston 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
Intronic RNAs represent an important layer of genome regulation and their role in the context of SSc is currently unknown. IncRNAs 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 IncRNAs remains unknown, a role in regulating and shaping the genome has been proposed (Mele and Rinn, 2010; Kimura et al., 2013; Peng et al., 2015). 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, IncRNAs display higher tissue specificity in their expression patterns (Derrien et al., 2012). Specifically, deregulated IncRNA 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 IncRNAs 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...
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
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 IncRNAs 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 IncRNAs, 4,171 were expressed in our skin biopsies. Six hundred and seventy-six IncRNAs 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 IncRNAs are listed in Supplementary Material S6 (online). Out of 676 IncRNA genes, 122 were decreased, while the expression of 554 genes was increased in SSc patients compared to healthy controls. Interestingly, clustering analysis using different
In our dataset was therefore investigated. The relevance of the antisense genes expressed in the sense direction (Katayama et al., 2005; Villegas and Zaphiropoulos 2015) has recently been described to have important regulatory roles on their coding gene counterparts. Antisense lncRNAs have recently been described among the 676 deregulated lncRNAs, the largest proportion (38%) belongs to the antisense gene category (Figure 3c).

Differential expression and the top upregulated gene is A20 (Table 1). In this study, we selected 26 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).

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

| Gene              | Dataset 1 Log2FC | P-Value   | FDR | Dataset 2 Log2FC | P-Value | Combined
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>CTBP1-AS2</td>
<td>0.32</td>
<td>0.012</td>
<td>0.044</td>
<td>0.40</td>
<td>7.5E-07</td>
<td>1.1E-09</td>
</tr>
<tr>
<td>OTUD6B-AS1</td>
<td>-0.95</td>
<td>7.0E-05</td>
<td>0.001</td>
<td>-0.63</td>
<td>0.001</td>
<td>1.4E-08</td>
</tr>
<tr>
<td>AGAP2-AS1</td>
<td>0.50</td>
<td>0.006</td>
<td>0.027</td>
<td>0.34</td>
<td>0.002</td>
<td>1.7E-06</td>
</tr>
<tr>
<td>HAND2-AS1</td>
<td>-1.04</td>
<td>0.002</td>
<td>0.011</td>
<td>-0.63</td>
<td>0.007</td>
<td>2.1E-06</td>
</tr>
<tr>
<td>HMGON3-AS1</td>
<td>-0.64</td>
<td>0.009</td>
<td>0.034</td>
<td>-0.33</td>
<td>0.017</td>
<td>2.6E-05</td>
</tr>
<tr>
<td>ZBTB11-AS1</td>
<td>0.53</td>
<td>0.002</td>
<td>0.010</td>
<td>0.17</td>
<td>0.143</td>
<td>4.5E-05</td>
</tr>
<tr>
<td>NIFK-AS1</td>
<td>-0.56</td>
<td>0.006</td>
<td>0.027</td>
<td>-0.36</td>
<td>0.178</td>
<td>2.3E-04</td>
</tr>
<tr>
<td>WAC-AS1</td>
<td>-0.58</td>
<td>0.001</td>
<td>0.009</td>
<td>-0.17</td>
<td>0.217</td>
<td>5.9E-05</td>
</tr>
<tr>
<td>PIK3CD-AS2</td>
<td>1.50</td>
<td>5.1E-06</td>
<td>1.5E-04</td>
<td>0.18</td>
<td>0.407</td>
<td>3.1E-07</td>
</tr>
<tr>
<td>ARRDC1-AS1</td>
<td>0.43</td>
<td>0.012</td>
<td>0.045</td>
<td>0.13</td>
<td>0.411</td>
<td>0.001</td>
</tr>
<tr>
<td>ZNF252P-AS1</td>
<td>1.64</td>
<td>1.1E-04</td>
<td>0.001</td>
<td>0.19</td>
<td>0.422</td>
<td>7.8E-06</td>
</tr>
<tr>
<td>SF2B-AS1</td>
<td>-0.43</td>
<td>0.014</td>
<td>0.049</td>
<td>-0.06</td>
<td>0.715</td>
<td>0.002</td>
</tr>
<tr>
<td>UNC5B-AS1</td>
<td>1.52</td>
<td>5.6E-05</td>
<td>0.001</td>
<td>-0.76</td>
<td>0.007</td>
<td>NA</td>
</tr>
<tr>
<td>HOXA10-AS</td>
<td>-2.64</td>
<td>4.4E-11</td>
<td>1.3E-08</td>
<td>0.53</td>
<td>0.056</td>
<td>NA</td>
</tr>
<tr>
<td>SLC25A25-AS1</td>
<td>0.52</td>
<td>4.3E-04</td>
<td>0.004</td>
<td>-0.30</td>
<td>0.163</td>
<td>NA</td>
</tr>
<tr>
<td>RUNDC3A-AS1</td>
<td>0.92</td>
<td>0.001</td>
<td>0.005</td>
<td>-0.30</td>
<td>0.225</td>
<td>NA</td>
</tr>
<tr>
<td>ZBED5-AS1</td>
<td>0.45</td>
<td>0.012</td>
<td>0.044</td>
<td>-0.16</td>
<td>0.275</td>
<td>NA</td>
</tr>
<tr>
<td>LOXL1-AS1</td>
<td>0.80</td>
<td>5.6E-05</td>
<td>0.001</td>
<td>-0.22</td>
<td>0.408</td>
<td>NA</td>
</tr>
<tr>
<td>BRWD1-AS2</td>
<td>1.54</td>
<td>3.4E-07</td>
<td>2.0E-05</td>
<td>-0.14</td>
<td>0.514</td>
<td>NA</td>
</tr>
<tr>
<td>ZEB1-AS1</td>
<td>-0.67</td>
<td>0.003</td>
<td>0.015</td>
<td>0.06</td>
<td>0.738</td>
<td>NA</td>
</tr>
<tr>
<td>RGM8-AS1</td>
<td>0.78</td>
<td>0.005</td>
<td>0.023</td>
<td>-0.05</td>
<td>0.815</td>
<td>NA</td>
</tr>
<tr>
<td>TMPO-AS1</td>
<td>1.40</td>
<td>7.4E-08</td>
<td>5.8E-06</td>
<td>-0.02</td>
<td>0.923</td>
<td>NA</td>
</tr>
</tbody>
</table>

Abbreviations: FC, fold change; FDR, false discovery rate–adjusted P-value; NA, not available.

1The 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).
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 S5 online).
The replication dataset (Figure 5. Cell-type Analysis of deregulated lncRNAs in SSc TC Messemaker et al.). Our results using next-generation sequencing first confirmed disease-specific manner in the skin of SSc patients. In 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-negative 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/immunflatory 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 TNAIP3 (Diederé et al., 2010). Interestingly, the expression of TNAIP3 is strongly reduced in SSc skin tissue. Given the role of TNAIP3 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. TNAIP3 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: EGA00000000316 (https://www.ebi.ac.uk/ega/organisations/EGA00000000316).

**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 IncRNAs (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, Sinora, 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 IncRNAs 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, ME, 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**


