Published in final edited form as:

Expert Rev Clin Immunol. 2011 July ; 7(4): 463–473. doi:10.1586/eci.11.41.

Capturing the heterogeneity in systemic sclerosis with genome-wide expression profiling

Jennifer L Sargent¹ and Michael L Whitfield^{†,1}

¹Department of Genetics, Dartmouth Medical School, Hanover, NH 03755, USA

Abstract

Heterogeneity in the clinical presentation and basic science findings of systemic sclerosis (SSc) has hindered the understanding of pathogenesis and development of effective treatments. Genome-wide profiling of SSc has measured this heterogeneity. Gene expression studies of diffuse SSc skin have shown reproducible, disease-specific gene expression signatures when compared with healthy controls and, surprisingly, disease-specific gene expression was found in both lesional and non-lesional skin. SSc-specific gene expression in peripheral blood cells and the lungs has also been demonstrated. Hypothesis-driven approaches that assess the contribution of individual pathways provide insight into the etiology of gene expression subsets.

Keywords

animal models; gene expression profiling; intrinsic subsets; scleroderma; signaling; systemic sclerosis; TGF- β

Scleroderma (systemic sclerosis [SSc]) is a poorly understood systemic autoimmune disease of unknown etiology. A clear understanding of the fundamental biology underlying SSc pathogenesis has been impeded by the heterogeneous nature of the disease [1]. Molecular findings regarding the presence and identification of immune infiltrates, and variability of the fibrotic phenotype, have differed dramatically. Analysis of SSc with high-throughput, genome-wide gene expression methods has attempted to quantify this heterogeneity and identify the different pathways deregulated in the disease.

The most widely recognized classification of SSc divides patients into subgroups based largely on the degree of skin involvement. The two predominant forms of the disease are SSc with diffuse scleroderma (dSSc) and systemic sclerosis, with limited scleroderma (lSSc) [2]. Internal organ involvement across these two groups is variable with diffuse SSc patients being more likely to develop interstitial lung disease (ILD) [3] and limited patients being more likely to develop pulmonary arterial hypertension (PAH) [4].

The lack of robust, objective biomarkers has made it difficult to predict clinical end points or to assess response to therapy. The most common outcome measure in clinical trials is the modified Rodnan Skin Score (mRSS) [5–7]. The mRSS is a reliable measure of skin thickness and fibrosis [8,9], is considered correlative with disease severity [10–12], and has been used as a surrogate marker for renal involvement [13]. However, other studies have

shown that skin stiffness findings are not correlative with risk of developing severe internal organ involvement [14,15]. Since changes in skin are variable and skin score can ultimately improve as disease progresses, mRSS can be a misleading measure of disease activity [16,17].

Autoantibody profiles have been associated with disease subtypes and specific clinical outcomes in SSc [13]. Up to eight autoantibodies have been identified that are unique to scleroderma, and in many cases only one autoantibody is found in a given patient over the course of their disease [18,19]. In this respect, autoantibodies are stable and consistent within a patient over time. The predictive value of SSc autoantibodies has been reviewed extensively [13,20,21].

Additional efforts have been made to distinguish the heterogeneity among scleroderma patients using combinatorial and multivariate analysis of clinical parameters. [1,22–27]. Shand and colleagues used a latent trajectory model to classify patients into one of three subgroups based on skin score trajectories with different overall survival rates [28]. Another study analyzed patients with against topoisomerase I (anti-topo I) antibodies and divided them into five groups based on skin thickness progress rates [29]. However, to date, no disease classifications beyond those of diffuse and limited have been widely accepted.

Genome-wide technologies such as DNA microarrays and ultra-high throughput DNA sequencing allow us to broadly assess the gene expression profiles in tissues and cells affected or driving disease processes. Application of these technologies to SSc has the potential to rapidly advance our understanding of the molecular pathogenesis in end-target tissues and cells.

Gene expression profiling in cells & end-target tissues

Gene expression in end-target tissues has shown more robust and consistent differences than those found in cultured cells, a result that is true not only in scleroderma, but also in diseases such as cancer [30,31]. Capturing the gene expression established in affected tissues allows us to observe gene expression changes that occur in multiple cell types that may contribute to disease pathogenesis. A variety of relevant tissues have been analyzed including skin [32–34], peripheral blood cells (PBCs) [35–37], bronchoalveolar lavage fluid (BAL) [38,39] and lung [40].

Gene expression signatures in scleroderma skin

SSc skin shows robust and consistent changes in gene expression. Two studies of early dSSc skin focused on the gene expression differences between SSc patients and healthy controls [32,33]. In both studies, genes associated with increased deposition and synthesis of the extracellular matrix were found to be differentially expressed [32]. Whitfield *et al.* analyzed lesional forearm and non-lesional back biopsies from four patients with dSSc and healthy controls, as well as isolated healthy and SSc dermal fibroblast cultures. More than 2700 genes could be identified that were differentially expressed between normal and dSSc skin biopsies. A surprising result was that the lesional forearm samples and non-lesional back samples showed nearly identical, disease-specific patterns of gene expression indicating that the disease was truly systemic at the time the biopsies were taken [32]. A second important finding was that differences in gene expression could be mapped not only to fibroblasts but also to epithelial, endothelial, smooth muscle, T and B cells [32]. Gardner *et al.* analyzed lesional forearm skin biopsies from nine dSSc patients and nine healthy controls and identified 1800 differentially expressed probes with the SSc phenotype driving the groupings [33]. Genes representative of the TGF- β and Wnt pathways were identified among the differentially expressed genes.

Gene expression profiling of cultured scleroderma fibroblasts

Gene expression studies have also been performed on fibroblasts isolated from SSc skin and healthy controls [41–43]. Zhou *et al.* analyzed 11 dermal fibroblast cultures from SSc (eight dSSc and three lSSc) compared with seven healthy control fibroblast cultures using 4000 element DNA microarrays [41]. A total of 32 differentially expressed genes were identified, and seven encoded known autoantigens. When gene expression in muscle biopsies and in PBCs from three SSc patients was examined, the same increases in autoantigen gene expression were not observed, suggesting that expression of these autoantigen genes was fibroblast specific [41].

Tan *et al.* analyzed non-lesional fibroblasts from dSSc patients compared with matched healthy controls on 16,659 element oligonucleotide arrays [42]. Approximately 5% of the 8324 genes that passed data pre-processing were found to be differentially expressed. Those genes with significant relative increases in expression included extracellular matrix components (*COL7A1*, *COL18A1* and *COMP*), lymphocyte homing antigen (CD44) and five metallothionein genes (*MTIX*, *MT1A*, *MT1B*, *MT2A* and *MT1F*). Subsequent testing of six multivariate classification models showed that 26 of the differentially expressed genes comprised the optimal set of genes for determining the class to which a given fibroblast sample belonged with a predictive accuracy of 99%.

Another study analyzed the gene expression in lesional and non-lesional dSSc fibroblasts compared with healthy controls [44]. Using a set of 4193 genes, they found that lesional and non-lesional fibroblasts from dSSc patients were more similar to each other than to healthy controls. This supports the findings of Whitfield *et al.* that showed gene expression in lesional and non-lesional SSc skin biopsies shared gene expression features distinct from healthy control biopsies [32].

Genetic versus non-genetic contributions to the SSc phenotype were examined by analyzing gene expression profiles of fibroblasts derived from monozygotic (MZ) and dizygotic twins discordant for SSc [43]. No significant differences were found when fibroblasts from lesional and non-lesional skin were compared, nor were significant differences observed between gene expression profiles of fibroblasts from affected and unaffected MZ twins. When gene expression profiles in dizygotic twin pairs were assessed, the differences were similar to those found when all SSc and all healthy individuals were compared (twins and controls). This study suggested that the unaffected MZ twins may have a genetic predisposition to SSc that is reflected in the gene expression profiles [43].

Scleroderma fibroblasts have been reported to lose their phenotype over time in culture, showing a marked reduction in collagen production in fibroblasts cultured for up to ten passages [45]. Gardner *et al.* directly addressed the extent to which cultured fibroblasts recapitulate the gene expression found in scleroderma skin biopsies [33]. Gene expression profiling was performed on skin biopsies (described earlier) and explanted fibroblasts from nine scleroderma patients and nine healthy control subjects. Although patient skin biopsies were clearly distinguishable from controls, the gene expression profiles from normal and SSc fibroblasts showed fewer differentially expressed genes and the changes were more variable [33]. This is consistent with the findings of Whitfield *et al.*, where large, consistent changes in gene expression were found in SSc skin relative to controls, but few significant differentially expressed genes were found when SSc and normal fibroblasts were analyzed by gene expression profiling [32].

Gene expression in lung

Microarray analysis of T cells isolated from BAL of SSc patients on cytokine microarrays demonstrated Th2 skewing of CD8⁺ T cells, consistent with the development of fibrotic

disease in these patients [39]. Gene expression analysis of cells in BAL fluid was able to identify patients at risk of developing ILD based on inflammatory markers [39]. While these findings are of importance, further analysis on comprehensive, whole genome DNA microarrays could yield additional information regarding mechanisms of pathogenesis.

Lung tissue from patients with SSc with pulmonary fibrosis or PAH were analyzed relative to normal donors, patients with idiopathic pulmonary fibrosis and patients with idiopathic PAH (IPAH) [40]. Lung tissue of SSc-pulmonary fibrosis and idiopathic pulmonary fibrosis show enrichment of collagen genes, insulin-like growth factor binding protein (*IGFBP*), matrix metalloproteinase 7 (*MMP7*) and caveolin-1 involved in fibrosis, caveolin-mediated endocytosis and insulin-like growth factor signaling. Lung tissues from individuals with SSc-PAH or IPAH showed enrichment of genes for interferon, *IL-4*, *IL-17*, *CCL-2* and *CXCL10*, associated with antigen presentation, chemokine activity and IL17 signaling.

Gene expression in peripheral blood cells

Gene expression in PBCs has shown upregulation of gene expression patterns associated with IFN- α and IFN- γ programs [35]. Duan and coworkers examined gene expression in purified monocytes and CD4⁺ T cells from dSSc, lSSc patients and healthy controls [36]. They identified 1800 genes differentially regulated in SSc monocytes compared with controls, and 863 genes differentially expressed in CD4⁺ T cells. Of these, 361 genes were differentially expressed in both cell types compared with controls, suggesting that the cells are responding similarly to the microenvironment of the SSc serum. Surprisingly, no significant differences were found when gene expression from monocytes and CD4⁺ T cells from dSSc patients was compared with those from lSSc patients. In the same study, the authors identified increased expression of an IFN type I-associated gene signature. Although no IFN- α was detectable in serum of these patients, increased levels of *IFN- α* mRNA were found in vascular and perivascular cells [36].

Patient subclassification using PBC gene expression has found a correlation between subtype and autoantibody profile [46–48]. Kim *et al.* demonstrated that the sera of SSc patients with autoantibodies anti-topo I induced IFN- α to a significantly greater extent than any other groups of patients [46]. Induction of IFN- α was from plasmacytoid dendritic cells and required uptake of immune complexes. Bos *et al.* showed that a subset of SSc patients with expression of a type I interferon signature in their PBCs were also negative for anti-centromere autoantibodies [47]. Assassi *et al.* examined the gene expression in PBCs from SSc and Lupus patients and also found an interferon signature that correlated with the presence of anti-topo I and anti-U1 RNP antibodies [48]. A higher interferon score was associated with a missense mutation in *IFNAR2*. Taken together, these studies show that autoantibody expression, or disease subtype as defined by autoantibody profiles, are strongly correlated with the presence of an IFN- α gene expression response signature.

A study examining gene expression in PBCs from patients with IPAH and from lSSc patients with PAH demonstrated a readily detectable PAH signature. Expression of the PAH-associated signature was enhanced in lSSc patient PBCs suggesting a causal relationship between SSc and PAH disease processes. Furthermore, subgroups of lSSc patients could be identified by gene expression correlated with PAH severity [37].

A second study examined patients with lSSc with and without PAH. Pendergrass *et al.* identified a gene expression signature that could clearly distinguish lSSc patients with PAH from those without [49]. These data showed a gradient of gene expression in the lSSc subset with the largest magnitude changes in those lSSc patients with PAH, intermediate in lSSc without PAH, and lowest in healthy controls. Increased circulating cytokine levels of inflammatory mediators such as TNF- α and IL1- β , and markers of vascular injury such as

VCAM-1 and VEGF, were also observed. Notably, similar inflammatory mediators were observed in lung tissue of SSc-PAH and IPAH [40]. Taken together, the gene expression in PBCs may provide additional insights into disease mechanisms for both dSSc and lSSc patients.

Quantifying heterogeneity in scleroderma

High-throughput gene expression data can be analyzed for data-driven groupings based solely on the expression profiles, or probed using pathway-specific signatures [50,51]. These analysis approaches have been described as ‘top-down’ or ‘bottom-up’, respectively (Figure 1) [50]. In the case of the top-down approach, biological information is inferred solely from gene expression in cells or in whole tissues. The data-driven gene expression groupings can be identified using different gene selection methods and clustering-based approaches to identify a gene expression-based model describing patient groups associated with specific clinical covariates (Figure 1). Ultimately, we investigate whether this information can aid in clinical decision making, such as targeting therapies to a specific group of patients.

Milano *et al.* analyzed the heterogeneity in gene expression from scleroderma skin biopsies [34]. Genome-wide expression profiling was performed on skin biopsies from dSSc, lSSc, localized (morphea) scleroderma and healthy controls. Both lesional forearm and non-lesional back skin biopsies were analyzed; 17 of 22 forearm back pairs again showed nearly identical, disease-specific gene expression. This systemic property of the biopsies was used to select 995 genes that showed the most consistent expression between forearm–back pairs for an individual, but the most variability across patients. This emphasized the consistent, patient-specific gene expression signatures rather than the differences between biopsies. Organizing the biopsies by these 995 genes alone identified three groups of patients with dSSc and two groups with lSSc, one of which was found in common. These groups have been termed the ‘intrinsic subsets’ (Figure 2). Each of the four major groups exhibited a unique gene expression profile and each could be associated with specific clinical variables such as ILD, gastrointestinal involvement and disease duration [34]. A group of dSSc patients showed a fibroproliferative signature, a second group of patients which included dSSc, lSSc and morphea, showed an inflammatory signature, a group of lSSc patients showed a heterogeneous gene signature and finally a number of lSSc and dSSc patients showed gene expression more similar to healthy controls and were termed ‘normal like’.

The finding that gene expression in SSc skin biopsies only loosely conforms to the major clinical subtypes of diffuse and limited is not surprising. Nevertheless, it demonstrates that heterogeneity in the disease can be captured objectively using gene expression profiling. More importantly, these findings show that scleroderma, like breast cancer [30,52,53], can be divided into multiple groups with different expression profiles that may be driven by different molecular mechanisms, thus requiring different therapies.

Deregulation of the TGF- β pathway in scleroderma

The hypothesis-driven, bottom-up approach examines gene expression in clinical samples using experimentally derived gene expression signatures identified when a specific signaling pathway is activated in cells in culture or in a target tissue [54]. To do this, one first defines a gene expression signature associated with the pathway or response of interest, and then uses that signature to provide an interpretive framework for microarray data from end-target tissues. Ultimately, the expectation and rationale for such a comparison is that disease in different subsets of patients is likely to be driven by different deregulated pathways, or respond differently to therapy. Identifying these pathways in the thousands of differentially expressed genes is non-trivial and using literature-derived gene signatures has proven to be error prone. Therefore, experimental signatures provide an unbiased set of genes that change

when the pathway is activated. Computational methods such as gene set enrichment analysis assess the enrichment of specific pathways and can be used to determine the statistical significance of experimentally derived gene expression signatures [55]. The specificity of the signature can be assessed using a randomly chosen list of the same size coupled with permutation of the clinical data labels to assess significance.

In order to fully understand the biology and molecular mechanisms underlying the intrinsic subsets of scleroderma, we have coupled the analysis of skin biopsies with a bottom-up strategy to specifically examine the role of cytokines and other soluble factors such as TGF- β that have been implicated in SSc pathogenesis [56]. Analysis of a TGF- β -responsive signature showed that a subset of dSSc skin biopsies that overlapped with the fibroproliferative intrinsic subset express this signature. Surprisingly, the TGF- β -responsive signature was not found in the inflammatory, limited or normal-like subsets. Thus a subset of dSSc skin biopsies, but not lSSc, morphea or healthy controls show expression of the TGF- β -responsive signature. The diffuse patients that showed high expression of the TGF- β -responsive signature were found more likely to have ILD and higher skin scores than patients in whom the signature was not expressed. Therefore, the fibroproliferative group may be driven, at least in part, by the TGF- β -pathway.

Surprisingly, no significant differences were found between the genes induced by TGF- β in SSc and normal fibroblasts [56]. This is consistent with a study of the genome-wide response to TGF- β in cultured lung fibroblasts, which did not find significant differences between the responses of SSc and normal fibroblasts [57]. The observed increases in collagen production, activation of TGF- β signaling and autoantigen synthesis may arise, in part, from fibroblasts responding to their microenvironment in resident tissues. Supporting this assertion, Zhou *et al.* found that application of serum from scleroderma patients to cultured healthy fibroblasts resulted in conversion of these cells into an activated scleroderma fibroblast phenotype resembling those isolated directly from scleroderma patients [43].

Assessment of response to therapy

A second way to utilize an expression signature approach is in assessing the response of a given patient to therapy. In this case, a specific response signature to a therapeutic agent can be measured and then analyzed in an SSc reference data set to determine which patients are most likely to respond. This was recently performed in a case report of imatinib mesylate (Gleevec™, Novartis) in SSc by Chung *et al.* [58]. Two patients were treated with imatinib mesylate and demonstrated reduction in fibrotic response measures including skin scores, digital ulcers and ILD. Gene expression analysis of lesional skin identified an imatinib mesylate responsive gene signature of >1000 genes [58]. Analysis of the gene expression dataset of Milano *et al.* showed that this expression signature is commonly deregulated in patients with dSSc [34].

This approach could be extended to the evaluation of other therapeutics in SSc and we propose the assessment of gene expression on a genome-wide scale be included as an outcome measure for all SSc clinical trials.

Next steps: quantitative assessment of murine models of scleroderma

The unknown etiology of SSc makes development of animal models a complex problem. A challenge for generating suitable models of SSc is that vascular abnormalities, immune dysfunction and fibrosis are difficult to recapitulate in a single model. Many mouse models have been proposed to resemble scleroderma on a phenotypic level (graft-versus-host disease, Tsk1, Tsk2, bleomycin-induced fibrosis and fibroblast-specific genetic disruptions

in TGF- β signaling) [59–69], however, none have been conclusively tested for concordance with the underlying molecular processes found in the human disease.

Genome-wide gene expression profiling allows for direct comparison of genes and pathways deregulated in SSc tissues with different mouse models of the disease (Figure 1). Interspecies comparisons using gene expression profiling have been used previously to define ‘stemness genes’ common to human and mouse cells [70], and to identify genes central to understanding mechanisms at play in prostate, breast and lung cancers and in hepatocellular carcinomas (HCCs) [71–74]. Herschkowitz and colleagues assessed gene expression in 13 different models of breast cancer and compared their results to gene expression in the molecular portraits of breast cancer [72,75]. Integrated analysis of gene expression data of orthologous genes from mouse tumors and from human breast cancer tumors showed conservation of features of the breast cancer molecular subsets in some models but not others [72].

In another study, gene expression was compared between characterized subtypes of human HCC tumors and seven HCC mouse models [74]. Again, cluster analysis showed that some models were better molecular representations of either subtype A, which demonstrate poor survival rates, or of subtype B, with better prognosis. Two models in this study were found to share features with only a small number of patient tumors and are probably not good models for studying HCC disease processes [74].

Given the diversity of mouse models currently used to study scleroderma phenotypes, we suggest a similar approach may quantify the degree to which each represents the human disease at a molecular level. Having identified the intrinsic subsets of scleroderma, analysis of mouse models suggest that different models approximate specific intrinsic subsets and share the same deregulated pathways [SARGENT JL, WHITFIELD ML, UNPUBLISHED DATA]. Characterization of the pathways contributing to pathogenesis in appropriate mouse models of scleroderma will further our understanding of the biology underlying the disease, provide the ability to more thoroughly test hypotheses regarding deregulated pathways and provide targets for the development of therapeutic agents.

Integration with genetic data

A recent genome-wide association study has identified several loci with significant associations with SSc [76]. This study confirmed genetic variation in the *MHC* region [77,78], *IRF5* [79] and *STAT4* [80,81] as SSc genetic risk factors, and further identified a new susceptibility locus associated with CD247 [76]. Associations were also identified in *CDH7* and *IRF4* but could not be confirmed in the replication cohort [76]. Another study identified a polymorphism in protein tyrosine phosphatase nonreceptor (*PTPN22*) involved in T-cell receptor signaling [82]. At this time it is very difficult to overlap the genetic studies with gene expression results since there are few cases where both genetic and gene expression data exists on the same patients. Assassi *et al.* examined a subset of single nucleotide polymorphisms and found that two different alleles at a single locus in the *IFNAR2* gene were strongly associated with a high type I interferon signature, whereas other SNPs in *IFNAR1*, *IFNAR2*, *IRF5*, *IRF7*, *STAT1* and *STAT4* did not correlate with interferon score [48]. Future studies should include not only gene expression measurements, but also the collection of DNA for the measurement of genetic variation that may influence genome-wide patterns of gene expression.

Conclusion

Gene expression profiling in SSc has shown robust changes in both skin, lung and PBCs. The gene expression changes in fibroblasts are more variable and imperfectly reflect what is

found *in vivo*. New advances show that gene expression profiling can capture the heterogeneity in SSc skin and demonstrate multiple gene expression groups. The TGF- β pathway is deregulated in the fibroproliferative group. We propose that quantitative expression profiling of mouse models and interspecies comparisons will likely identify appropriate mouse models in which to study each gene expression subset.

The nature of gene expression-based subsets in SSc is still being addressed. One hypothesis is that each group represents a discrete point in a progressive disease and that if we follow patients over time, they may change subsets. However, unpublished data suggest this is not the case, but rather that the different gene expression subsets are stable entities [PENDERGRASS SA, WHITFIELD ML. Stable gene expression of serial skin biopsies defines patient subsets in diffuse cutaneous systemic sclerosis. MANUSCRIPT SUBMITTED]. This would suggest that each gene expression group represents a fundamentally different disease, or that disease results from different etiologic mechanisms. Further analysis of carefully controlled patient cohorts is necessary to further understand the etiology of SSc gene expression subsets in different tissues, to rule out the treatment effects, and to determine how a patient's underlying genetics contribute to their gene expression phenotype.

The most important finding for clinical trials and treatment is that gene expression profiling can measure the response to therapy, as demonstrated in the case of imatinib mesylate, and therefore could be used as a clinical outcome measure. As the gene expression subsets are refined and responses to therapy are systematically characterized, a diagnostic test could be developed to target therapies to specific patients based on their gene expression profile.

Expert commentary

Systematic analysis of gene expression in skin biopsies, dermal fibroblasts, BAL fluids, lung and PBCs have provided insight into the systemic gene expression of SSc. Molecular profiling of SSc skin has shown that the heterogeneity of the disease can be broken down by gene expression and has been used to characterize the pathways that may underlie different subsets. Specific pathways implicated in disease pathogenesis, such as TGF- β , inflammatory responses, insulin-like growth factor (IGF) and interferon signaling, have been investigated across different tissues. In many cases, these pathways are heterogeneously expressed, but can be associated with specific clinical covariates, autoantibody profiles or genetic changes.

Identification of the 'intrinsic' molecular subsets of SSc based solely on gene expression in skin [34] represents a paradigm shift, and recent studies have identified variability in gene expression of SSc in PBCs [37,48,49] and the lungs [40]. These findings, coupled with pathway-specific interrogation of signaling pathways pave the way for the development of personalized treatment. Chung *et al.* demonstrate that gene expression profiling can identify a drug response in SSc and, thus, potentially those patients more likely to respond to specific treatment such as imatinib mesylate [58]. The subset and pathway-responsive gene expression signatures can be used to identify patient groups most likely to respond to a given therapeutic or as measureable outcome indicators in clinical trials.

Five-year view

A major challenge is understanding the mechanisms that drive disease in different subsets of SSc as well as the broad changes that may be central to disease pathogenesis. A key component of this will be identification of mouse models that represent different subsets, not only in skin but also in other tissues such as PBCs and lung. Many mouse models have been proposed to recapitulate aspects of SSc. Examining these at the gene expression level and how well they recapitulate the gene expression and deregulated pathways found in human

tissues allows us to quantify their relevance. Most importantly, these models can be used to test the hypotheses generated by analysis of the human disease.

Developing biomarkers that identify the gene expression subsets and can be used to stratify patients in the clinic is a major challenge. Hurdles that must be overcome include the need for validation data sets, smaller gene sets and cost-effective methods to measure them in a clinical setting, the need to gain a better understanding of the clinical associations with each subset, along with therapeutic response data.

Finally, the amount of data available probing SSc and related diseases in the public domain is continuing to increase at a rapid rate. Gene expression data sets, genome-wide association studies and proteomic data sets can now be mined by continued metaanalysis of publicly available data for new hypotheses regarding deregulated pathways.

Acknowledgments

Michael L Whitfield and Jennifer L Sargent were supported by a grant from the Scleroderma Research Foundation. Michael L Whitfield received additional support from an Arthritis Investigator Award from the Arthritis Foundation, and from NIH grant U01AR055063. Michael L Whitfield and Jennifer L Sargent have filed patent applications for gene expression biomarkers in scleroderma. Michael L Whitfield is a scientific founder and holds an interest in Celdara Medical, LLC.

No writing assistance was utilized in the production of this manuscript.

References

Papers of special note have been highlighted as:

• of interest

•• of considerable interest

1. Barnett AJ, Miller MH, Littlejohn GO. A survival study of patients with scleroderma diagnosed over 30 years (1953–1983): the value of a simple cutaneous classification in the early stages of the disease. *J. Rheumatol.* 1988; 15(2):276–283. [PubMed: 3361537]
2. LeRoy EC, Black C, Fleischmajer R, et al. Scleroderma (systemic sclerosis): classification, subsets and pathogenesis. *J. Rheumatol.* 1988; 15(2):202–205. [PubMed: 3361530]
3. Bolster MB, Silver RM. Lung disease in systemic sclerosis (scleroderma). *Baillieres Clin. Rheumatol.* 1993; 7(1):79–97. [PubMed: 8519079]
4. Black CM, Welsh KI, Maddison PJ, Jayson MI, Bernstein RM. HLA antigens, autoantibodies and clinical subsets in scleroderma. *Br. J. Rheumatol.* 1984; 23(4):267–271. [PubMed: 6593111]
5. Akesson A, Wollheim FA. Organ manifestations in 100 patients with progressive systemic sclerosis: a comparison between the CREST syndrome and diffuse scleroderma. *Br. J. Rheumatol.* 1989; 28(4):281–286. [PubMed: 2787182]
6. Merkel PA, Silliman NP, Denton CP, et al. Validity, reliability, and feasibility of durometer measurements of scleroderma skin disease in a multicenter treatment trial. *Arthritis Rheum.* 2008; 59(5):699–705. [PubMed: 18438905]
7. Rodnan GP, Lipinski E, Luksick J. Skin thickness and collagen content in progressive systemic sclerosis and localized scleroderma. *Arthritis Rheum.* 1979; 22(2):130–140. [PubMed: 420706]
8. Clements P, Lachenbruch P, Siebold J, et al. Inter and intraobserver variability of total skin thickness score (modified Rodnan TSS) in systemic sclerosis. *J. Rheumatol.* 1995; 22(7):1281–1285. [PubMed: 7562759]
9. Furst DE, Clements PJ, Steen VD, et al. The modified Rodnan skin score is an accurate reflection of skin biopsy thickness in systemic sclerosis. *J. Rheumatol.* 1998; 25(1):84–88. [PubMed: 9458208]

10. Clements PJ, Hurwitz EL, Wong WK, et al. Skin thickness score as a predictor and correlate of outcome in systemic sclerosis: high-dose versus low-dose penicillamine trial. *Arthritis Rheum.* 2000; 43(11):2445–2454. [PubMed: 11083267]
11. Steen VD, Medsger TA Jr. Improvement in skin thickening in systemic sclerosis associated with improved survival. *Arthritis Rheum.* 2001; 44(12):2828–2835. [PubMed: 11762943]
12. Verrecchia F, Laboureaux J, Verola O, et al. Skin involvement in scleroderma—where histological and clinical scores meet. *Rheumatology (Oxford).* 2007; 46(5):833–841. [PubMed: 17255134]
13. Steen VD. Autoantibodies in systemic sclerosis. *Semin. Arthritis Rheum.* 2005; 35(1):35–42. [PubMed: 16084222]
14. Herrick A, Rooney B, Finn J, Silman A. Lack of relationship between functional ability and skin score in patients with systemic sclerosis. *J. Rheumatol.* 2001; 28(2):292–295. [PubMed: 11246664]
15. Hanitsch LG, Burmester GR, Witt C, et al. Skin sclerosis is only of limited value to identify SSc patients with severe manifestations—an analysis of a distinct patient subgroup of the German Systemic Sclerosis Network (DNSS) Register. *Rheumatology (Oxford).* 2009; 48(1):70–73. [PubMed: 19056798]
16. Medsger, TA.; Steen, VD. Classification, prognosis. In: Clements, PJ.; Furst, DE., editors. *Systemic Sclerosis*. Williams & Wilkins; Baltimore, MD, USA: 1996. p. 51–64.
17. Siebold, JR. Chapter 79: Scleroderma. In: Harris, ED., Jr; Budd, RC.; Firestein, GS.; Genovese, MC.; Sargent, JS.; Ruddy, S.; Sledge, CB., editors. *Kelley's Textbook of Rheumatology*. Elsevier & Saunders; Amsterdam, The Netherlands: 2005. p. 1279–1308.
18. Hildebrandt S, Jackh G, Weber S, Peter HH. A long-term longitudinal isotypic study of anti-topoisomerase I autoantibodies. *Rheumatol. Int.* 1993; 12(6):231–234. [PubMed: 8387225]
19. Dick T, Mierau R, Bartz-Bazzanella P, et al. Coexistence of antitopoisomerase I and anticentromere antibodies in patients with systemic sclerosis. *Ann. Rheum. Dis.* 2002; 61(2):121–127. [PubMed: 11796397]
20. Grassegger A, Pohla-Gubo G, Frauscher M, Hintner H. Autoantibodies in systemic sclerosis (scleroderma): clues for clinical evaluation, prognosis and pathogenesis. *Wien Med. Wochenschr.* 2008; 158(1–2):19–28. [PubMed: 18286246]
21. Koenig M, Dieude M, Senecal JL. Predictive value of antinuclear autoantibodies: the lessons of the systemic sclerosis autoantibodies. *Autoimmun Rev.* 2008; 7(8):588–593. [PubMed: 18617021]
22. Scussel-Lonzetti L, Joyal F, Raynauld JP, et al. Predicting mortality in systemic sclerosis: analysis of a cohort of 309 French Canadian patients with emphasis on features at diagnosis as predictive factors for survival. *Medicine (Baltimore).* 2002; 81(2):154–167. [PubMed: 11889414]
23. Maricq HR, Valter I. A working classification of scleroderma spectrum disorders: a proposal and the results of testing on a sample of patients. *Clin. Exp. Rheumatol.* 2004; 22(3 Suppl. 33):S5–S13. [PubMed: 15344590]
24. Ferri C, Valentini G, Cozzi F, et al. Systemic sclerosis: demographic, clinical, and serologic features and survival in 1,012 Italian patients. *Medicine (Baltimore).* 2002; 81(2):139–153. [PubMed: 11889413]
25. Barnett AJ, Miller M, Littlejohn GO. The diagnosis and classification of scleroderma (systemic sclerosis). *Postgrad. Med. J.* 1988; 64(748):121–125. [PubMed: 3050937]
26. Medsger TA Jr, Silman AJ, Steen VD, et al. A disease severity scale for systemic sclerosis: development and testing. *J. Rheumatol.* 1999; 26(10):2159–2167. [PubMed: 10529133]
27. Steen VD, Medsger TA Jr. Severe organ involvement in systemic sclerosis with diffuse scleroderma. *Arthritis Rheum.* 2000; 43(11):2437–2444. [PubMed: 11083266]
28. Shand L, Lunt M, Nihtyanova S, et al. Relationship between change in skin score and disease outcome in diffuse cutaneous systemic sclerosis: application of a latent linear trajectory model. *Arthritis Rheum.* 2007; 56(7):2422–2431. [PubMed: 17599771]
29. Perera A, Fertig N, Lucas M, et al. Clinical subsets, skin thickness progression rate, and serum antibody levels in systemic sclerosis patients with anti-topoisomerase I antibody. *Arthritis Rheum.* 2007; 56(8):2740–2746. [PubMed: 17665460]
30. Perou CM, Sorlie T, Eisen MB, et al. Molecular portraits of human breast tumours. *Nature.* 2000; 406(6797):747–752. [PubMed: 10963602]

31. Ross DT, Scherf U, Eisen MB, et al. Systematic variation in gene expression patterns in human cancer cell lines. *Nat. Genet.* 2000; 24(3):227–235. [PubMed: 10700174]
32. Whitfield ML, Finlay DR, Murray JJ, et al. Systemic and cell type-specific gene expression patterns in scleroderma skin. *Proc. Natl Acad. Sci. USA.* 2003; 100(21):12319–12324. [PubMed: 14530402] • The first genome-wide analysis of gene expression in systemic sclerosis (SSc) skin.
33. Gardner H, Shearstone JR, Bandaru R, 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(6):1961–1973. [PubMed: 16736506] • Demonstrated that gene expression in SSc skin was not accurately reflected in paired fibroblasts grown in cell culture.
34. Milano A, Pendergrass SA, Sargent JL, et al. Molecular subsets in the gene expression signatures of scleroderma skin. *PLoS ONE.* 2008; 3(7):e2696. [PubMed: 18648520] • First demonstration of ‘intrinsic’ gene expression-based subsets in SSc skin.
35. Tan FK, Zhou X, Mayes MD, et al. Signatures of differentially regulated interferon gene expression and vasculotrophism in the peripheral blood cells of systemic sclerosis patients. *Rheumatology (Oxford).* 2006; 45(6):694–702. [PubMed: 16418202] • Demonstrated an interferon gene expression signature in peripheral blood cells.
36. Duan H, Fleming J, Pritchard DK, et al. Combined analysis of monocyte and lymphocyte messenger RNA expression with serum protein profiles in patients with scleroderma. *Arthritis Rheum.* 2008; 58(5):1465–1474. [PubMed: 18438864]
37. Grigoryev DN, Mathai SC, Fisher MR, et al. Identification of candidate genes in scleroderma-related pulmonary arterial hypertension. *Transl. Res.* 2008; 151(4):197–207. [PubMed: 18355767]
38. Luzina IG, Atamas SP, Wise R, Wigley FM, Xiao HQ, White B. Gene expression in bronchoalveolar lavage cells from scleroderma patients. *Am. J. Respir. Cell Mol. Biol.* 2002; 26(5):549–557. [PubMed: 11970906]
39. Luzina IG, Atamas SP, Wise R, et al. Occurrence of an activated, profibrotic pattern of gene expression in lung CD8⁺ T cells from scleroderma patients. *Arthritis Rheum.* 2003; 48(8):2262–2274. [PubMed: 12905481]
40. Hsu E, Shi H, Jordan RM, Lyons-Weiler J, Pilewski JM, Feghali-Bostwick CA. Lung tissues in patients with systemic sclerosis have gene expression patterns unique to pulmonary fibrosis and pulmonary hypertension. *Arthritis Rheum.* 2011; 63(3):783–794. [PubMed: 21360508] • First genome-wide analysis of gene expression in SSc lung.
41. Zhou X, Tan FK, Xiong M, et al. Systemic sclerosis (scleroderma): specific autoantigen genes are selectively overexpressed in scleroderma fibroblasts. *J. Immunol.* 2001; 167(12):7126–7133. [PubMed: 11739535]
42. Tan FK, Hildebrand BA, Lester MS, et al. Classification analysis of the transcriptome of nonlesional cultured dermal fibroblasts from systemic sclerosis patients with early disease. *Arthritis Rheum.* 2005; 52(3):865–876. [PubMed: 15751056]
43. Zhou X, Tan FK, Xiong M, Arnett FC, Feghali-Bostwick CA. Monozygotic twins clinically discordant for scleroderma show concordance for fibroblast gene expression profiles. *Arthritis Rheum.* 2005; 52(10):3305–3314. [PubMed: 16200604]
44. Fuzii HT, Yoshikawa GT, Junta CM, et al. Affected and non-affected skin fibroblasts from systemic sclerosis patients share a gene expression profile deviated from the one observed in healthy individuals. *Clin. Exp. Rheumatol.* 2008; 26(5):866–874. [PubMed: 19032821]
45. Vuorio T, Makela JK, Vuorio E. Activation of type I collagen genes in cultured scleroderma fibroblasts. *J. Cell Biochem.* 1985; 28(2):105–113. [PubMed: 4077926]
46. Kim D, Peck A, Santer D, et al. Induction of interferon- α by scleroderma sera containing autoantibodies to topoisomerase I: association of higher interferon- α activity with lung fibrosis. *Arthritis Rheum.* 2008; 58(7):2163–2173. [PubMed: 18576347]
47. Bos CL, van Baarsen LG, Timmer TC, et al. Molecular subtypes of systemic sclerosis in association with anti-centromere antibodies and digital ulcers. *Genes Immun.* 2009; 10(3):210–218. [PubMed: 19129850] • Demonstrated that an interferon signature showed decreased expression in patients that were positive for anti-centromere autoantibodies.
48. Assassi S, Mayes MD, Arnett FC, et al. Systemic sclerosis and lupus: points in an interferon-mediated continuum. *Arthritis Rheum.* 2010; 62(2):589–598. [PubMed: 20112391] • Identified

- commonalities between the SSc and lupus interferon signatures. A significant association with a missense mutation in *IFNAR2* and high expression of the interferon signature was also shown.
49. Pendergrass SA, Hayes E, Farina G, et al. Limited systemic sclerosis patients with pulmonary arterial hypertension show biomarkers of inflammation and vascular injury. *PLoS ONE*. 2010; 5(8):e12106. [PubMed: 20808962] •• Demonstrated that gene expression subsets could be identified in limited SSc that identified individuals with pulmonary arterial hypertension. Similar changes could also be found among circulating factors in the blood.
 50. Liu ET. Mechanism-derived gene expression signatures and predictive biomarkers in clinical oncology. *Proc. Natl Acad. Sci. USA*. 2005; 102(10):3531–3532. [PubMed: 15738396]
 51. Wong DJ, Chang HY. Learning more from microarrays: insights from modules and networks. *J. Invest. Dermatol.* 2005; 125(2):175–182. [PubMed: 16098025]
 52. Sorlie T, Perou CM, Tibshirani R, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc. Natl Acad. Sci. USA*. 2001; 98(19):10869–10874. [PubMed: 11553815]
 53. Julka PK, Chacko RT, Nag S, et al. A Phase II study of sequential neoadjuvant gemcitabine plus doxorubicin followed by gemcitabine plus cisplatin in patients with operable breast cancer: prediction of response using molecular profiling. *Br. J. Cancer*. 2008; 98(8):1327–1335. [PubMed: 18382427]
 54. Sargent JL, Milano A, Connolly MK, Whitfield ML. Scleroderma gene expression and pathway signatures. *Curr. Rheumatol. Rep.* 2008; 10(3):205–211. [PubMed: 18638428]
 55. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA*. 2005; 102(43):15545–15550. [PubMed: 16199517]
 56. Sargent JL, Milano A, Bhattacharyya S, et al. A TGF β -responsive gene signature is associated with a subset of diffuse scleroderma with increased disease severity. *J. Invest. Dermatol.* 2010; 130(3): 694–705. [PubMed: 19812599] •• Demonstrated that one of the ‘intrinsic’ subsets of SSc was associated with a **TGF- β -responsive gene signature**.
 57. Chambers RC, Leoni P, Kaminski N, Laurent GJ, Heller RA. Global expression profiling of fibroblast responses to transforming growth factor- β 1 reveals the induction of inhibitor of differentiation-1 and provides evidence of smooth muscle cell phenotypic switching. *Am. J. Pathol.* 2003; 162(2):533–546. [PubMed: 12547711]
 58. Chung L, Fiorentino DF, Benbarak MJ, et al. Molecular framework for response to imatinib mesylate in systemic sclerosis. *Arthritis Rheum.* 2009; 60(2):584–591. [PubMed: 19180499] •• Demonstrated an imatinib mesylate response signature in patients that responded to therapy, and suggested a subset of diffuse SSc patients may benefit from treatment.
 59. Zhang Y, McCormick LL, Desai SR, Wu C, Gilliam AC. Murine sclerodermatous graft-versus-host disease, a model for human scleroderma: cutaneous cytokines, chemokines, and immune cell activation. *J. Immunol.* 2002; 168(6):3088–3098. [PubMed: 11884483]
 60. Ruzek MC, Jha S, Ledbetter S, Richards SM, Garman RD. A modified model of graft-versus-host-induced systemic sclerosis (scleroderma) exhibits all major aspects of the human disease. *Arthritis Rheum.* 2004; 50(4):1319–1331. [PubMed: 15077316]
 61. Green MC, Sweet HO, Bunker LE. Tight-skin, a new mutation of the mouse causing excessive growth of connective tissue and skeleton. *Am. J. Pathol.* 1976; 82(3):493–512. [PubMed: 176891]
 62. Asano N, Fujimoto M, Yazawa N, et al. B Lymphocyte signaling established by the CD19/CD22 loop regulates autoimmunity in the tight-skin mouse. *Am. J. Pathol.* 2004; 165(2):641–650. [PubMed: 15277237]
 63. Hasegawa M, Hamaguchi Y, Yanaba K, et al. B-lymphocyte depletion reduces skin fibrosis and autoimmunity in the tight-skin mouse model for systemic sclerosis. *Am. J. Pathol.* 2006; 169(3): 954–966. [PubMed: 16936269]
 64. Christner PJ, Peters J, Hawkins D, Siracusa LD, Jimenez SA. The tight skin 2 mouse. An animal model of scleroderma displaying cutaneous fibrosis and mononuclear cell infiltration. *Arthritis Rheum.* 1995; 38(12):1791–1798. [PubMed: 8849351]

65. Gentiletti J, McCloskey LJ, Artlett CM, Peters J, Jimenez SA, Christner PJ. Demonstration of autoimmunity in the tight skin-2 mouse: a model for scleroderma. *J. Immunol.* 2005; 175(4):2418–2426. [PubMed: 16081813]
66. Barisic-Dujmovic T, Boban I, Clark SH. Regulation of collagen gene expression in the Tsk2 mouse. *J. Cell Physiol.* 2008; 215(2):464–471. [PubMed: 17960558]
67. Yamamoto T, Takagawa S, Katayama I, et al. Animal model of sclerotic skin. I: Local injections of bleomycin induce sclerotic skin mimicking scleroderma. *J. Invest. Dermatol.* 1999; 112(4):456–462. [PubMed: 10201529]
68. Sonnylal S, Denton CP, Zheng B, et al. Postnatal induction of transforming growth factor β signaling in fibroblasts of mice recapitulates clinical, histologic, and biochemical features of scleroderma. *Arthritis Rheum.* 2007; 56(1):334–344. [PubMed: 17195237]
69. Denton CP, Zheng B, Evans LA, et al. Fibroblast-specific expression of a kinase-deficient type II transforming growth factor β (TGF β) receptor leads to paradoxical activation of TGF β signaling pathways with fibrosis in transgenic mice. *J. Biol. Chem.* 2003; 278(27):25109–25119. [PubMed: 12707256]
70. Sato N, Sanjuan IM, Heke M, Uchida M, Naef F, Brivanlou AH. Molecular signature of human embryonic stem cells and its comparison with the mouse. *Dev. Biol.* 2003; 260(2):404–413. [PubMed: 12921741]
71. Ellwood-Yen K, Graeber TG, Wongvipat J, et al. Myc-driven murine prostate cancer shares molecular features with human prostate tumors. *Cancer Cell.* 2003; 4(3):223–238. [PubMed: 14522256]
72. Herschkowitz JI, Simin K, Weigman VJ, et al. Identification of conserved gene expression features between murine mammary carcinoma models and human breast tumors. *Genome Biol.* 2007; 8(5):R76. [PubMed: 17493263]
73. Sweet-Cordero A, Mukherjee S, Subramanian A, et al. An oncogenic *KRAS2* expression signature identified by cross-species gene-expression analysis. *Nat. Genet.* 2005; 37(1):48–55. [PubMed: 15608639]
74. Lee JS, Chu IS, Mikaelyan A, et al. Application of comparative functional genomics to identify best-fit mouse models to study human cancer. *Nat. Genet.* 2004; 36(12):1306–1311. [PubMed: 15565109]
75. Bennett CN, Green JE. Unlocking the power of cross-species genomic analyses: identification of evolutionarily conserved breast cancer networks and validation of preclinical models. *Breast Cancer Res.* 2008; 10(5):213. [PubMed: 18828875]
76. Radstake TR, Gorlova O, Rueda B, et al. Genome-wide association study of systemic sclerosis identifies CD247 as a new susceptibility locus. *Nat. Genet.* 2010; 42(5):426–429. [PubMed: 20383147]
77. Zhou X, Lee JE, Arnett FC, et al. HLA-*DPB1* and *DPB2* are genetic loci for systemic sclerosis: a genome-wide association study in Koreans with replication in North Americans. *Arthritis Rheum.* 2009; 60(12):3807–3814. [PubMed: 19950302]
78. Gilchrist FC, Bunn C, Foley PJ, et al. Class II HLA associations with autoantibodies in scleroderma: a highly significant role for HLA-DP. *Genes Immun.* 2001; 2(2):76–81. [PubMed: 11393660]
79. Ito I, Kawaguchi Y, Kawasaki A, et al. Association of a functional polymorphism in the *IRF5* region with systemic sclerosis in a Japanese population. *Arthritis Rheum.* 2009; 60(6):1845–1850. [PubMed: 19479858]
80. Rueda B, Broen J, Simeon C, et al. The *STAT4* gene influences the genetic predisposition to systemic sclerosis phenotype. *Hum. Mol. Genet.* 2009; 18(11):2071–2077. [PubMed: 19286670]
81. Tsuchiya N, Kawasaki A, Hasegawa M, et al. Association of *STAT4* polymorphism with systemic sclerosis in a Japanese population. *Ann. Rheum. Dis.* 2009; 68(8):1375–1376. [PubMed: 19605749]
82. Diaz-Gallo LM, Gourh P, Broen J, et al. Analysis of the influence of *PTPN22* gene polymorphisms in systemic sclerosis. *Ann. Rheum. Dis.* 2010; 70(3):454–462. [PubMed: 21131644]

Key issues

- Heterogeneity in scleroderma has complicated the understanding of disease pathogenesis and is likely a confounding factor in clinical trials. Gene expression in skin, lung and peripheral blood cells (PBCs) provides a way to measure and quantify this heterogeneity, and use the information to subset patients.
- Gene expression analysis of systemic sclerosis (SSc) PBCs has shown a consistent interferon signature, which has been associated with autoantibody subsets and a missense mutation in *IFNAR2*. The interferon signature in SSc shows similarities to that observed in other SSc tissues and lupus, suggesting common mechanisms.
- Analysis of gene expression in lesional and non-lesional skin biopsies of SSc has shown consistent disease-specific patterns of gene expression that are distinct from healthy controls. Gene expression subsets can be identified by analysis of SSc skin, and each group can be associated with different pathways and clinical covariates.
- A subset of diffuse SSc patients show activation of TGF- β -responsive genes concomitant with higher modified Rodnan Skin Score and an increased occurrence of interstitial lung disease relative to those diffuse SSc patients without the TGF- β -responsive gene signature.
- The discovery of molecular subsets in SSc provides the opportunity for development of therapeutics targeting each subset along with a companion diagnostic. Clinical trials should benefit from subtyping patients into pathway-centric groups.

Figure 1. Gene expression profiling approaches to understanding systemic sclerosis

The top-down approach utilizes gene expression profiling from SSc tissues to derive information about genes and pathways deregulated in the disease. Stratification of patients into molecular subsets, each associated with different clinical covariates, may aid in treatment decisions. The bottom-up approach provides a framework to analyze gene expression in SSc target tissues using gene expression signatures derived by stimulating specific pathways such as TGF- β . The contribution of a specific cytokine or signaling pathway can then be assessed using the gene expression in the end-target tissues. Analysis of gene expression in end-target tissues of animal models allows for comparison of gene expression and pathways in human SSc tissues.

SSc: Systemic sclerosis.

Figure 2. Molecular subsets in the gene expression of systemic sclerosis skin

Molecular, or ‘intrinsic’, subsets identified by Milano *et al.* [34] are shown. The clustering dendrogram is colored to reflect the major gene expression groups of fibroproliferation (blue and red), inflammatory (purple), limited (gold) and normal like (green). The diagnosis of each patient is indicated by the color of the hash mark below the dendrogram tree. A subset of the 995 intrinsic genes is shown and delineates the five gene groups based on similarity of the expression profiles. These are the limited, fibroproliferative, immunoglobulins (IgG) and inflammatory and normal-like expression groups. A subset of genes in each cluster is labeled on the right of the figure. For a full description of the sample groupings and gene expression groups, refer to [34].

Modified with permission from [34]. © 2008 Milano *et al.*