A novel multi-network approach reveals tissue-specific cellular modulators of fibrosis in systemic sclerosis, pulmonary fibrosis and pulmonary arterial hypertension

Jaclyn N. Taroni[1], Casey S. Greene[2], Viktor Martyanov[1], Tammara A. Wood[1], Romy B. Christmann[3], Harrison W. Farber[4], Robert A. Lafyatis[3,5], Christopher P. Denton[6], Monique E. Hinchcliff[7], Patricia A. Pioli[8], J. Matthew Mahoney[9,§], Michael L. Whitfield[1, §]

Affiliations:

[1] Department of Molecular and Systems Biology, Geisel School of Medicine at Dartmouth, Hanover, NH 03755, USA
[2] Department of Systems Pharmacology & Translational Therapeutics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA
[3] Division of Rheumatology, Department of Medicine, Boston University School of Medicine, Boston, MA, USA
[4] Pulmonary Center, Department of Medicine, Boston University School of Medicine, Boston, MA, 02118, USA
[5] Division of Rheumatology and Clinical Immunology, Department of Medicine, University of Pittsburgh Medical Center, Pittsburgh, PA, 15261 USA
[6] Division of Medicine, University College London, London, UK
[7] Division of Rheumatology, Department of Medicine, Feinberg School of Medicine, Northwestern University, Chicago, IL, 60611, USA
[8] Department of Microbiology and Immunology, Geisel School of Medicine at Dartmouth, Lebanon, NH 03756, USA
[9] Department of Neurological Sciences, University of Vermont, Burlington, VT 05405, USA

Running Title: Multi-tissue functional genomic networks of fibrosis

[§]To whom correspondence should be addressed:

Michael L. Whitfield, Ph.D.
Department of Molecular and Systems Biology
Geisel School of Medicine at Dartmouth
7400 Remsen
Hanover, NH 03755
Email: michael.whitfield@dartmouth.edu

J. Matthew Mahoney, Ph.D.
Department of Neurological Sciences
University of Vermont College of Medicine
HSRF 426
149 Beaumont Avenue
Burlington, VT 05405
Email: John.M.Mahoney@uvm.edu

Abstract

We have used integrative genomics to determine if a common molecular mechanism underlies different clinical manifestations in systemic sclerosis (SSc), and the related conditions pulmonary fibrosis (PF) and pulmonary arterial hypertension (PAH). We identified a common pathogenic gene expression signature–an immune-fibrotic axis–indicative of pro-fibrotic macrophages (MØs) in multiple affected tissues (skin, lung, esophagus and PBMCs) of SSc, PF, and PAH. We used this disease-associated signature to query tissue-specific functional genomic networks. This allowed us to identify common and tissue-specific pathology of SSc and related conditions. We rigorously contrasted the lung- and skin-specific gene-gene interaction networks to identify a distinct lung resident MØ signature (LR-MØ) associated with lipid stimulation and alternative activation. In keeping with our network results, we find distinct MØ alternative activation transcriptional programs in SSc-PF lung and in the skin of patients with an 'inflammatory' SSc gene expression signature. Our results suggest that the innate immune system is central to SSc disease processes, but that subtle distinctions exist between tissues. Our approach provides a framework for examining molecular signatures of disease in fibrosis and autoimmune diseases and for leveraging publicly available data to understand common and tissue-specific disease processes in complex human diseases.

Author Summary

Human disease in part arises from aberrant interplay between tissues and from the interactions of gene products in tissue-specific microenvironments. Recent efforts have utilized 'big data' to build functional maps that model these interactions. We used these tools to study systemic sclerosis (SSc), a rare and clinically complex disease characterized by multi-organ involvement, high mortality, pulmonary fibrosis, and pulmonary arterial hypertension, and related fibrotic conditions. We developed a novel procedure to assess which processes are affected across multiple fibrotic organs and tissues. We found that patients with severe disease share molecular patterns that are indicative of dysregulated, immune and fibrotic processes. Placing these patterns into the context of functional maps allowed us to study severe disease manifestations that occur in a subset of patients. This study not only offers the potential to identify shared pathology in SSc and fibrosis, but a 'road map' for the use of tissue-specific networks to describe complex human diseases.

Introduction

Integrative genomics has yielded powerful tissue-specific functional networks that model the interaction of genes in these specialized 'microenvironments' (1). These tools hold promise for understanding how genes may contribute to human diseases (2) that arise, in part, out of an aberrant interplay of cell types and tissues. Network biology has played a crucial role in our understanding of complex human diseases such as cancer (3,4), and more recently, in disorders where the interactions among multiple tissues are dysregulated (5).

Analytical approaches that leverage biological 'big data' can be especially fruitful in rare and heterogeneous diseases (6), in which the risk of mortality is significant and no approved treatments exist. We performed an integrative, multi-tissue analysis for systemic sclerosis (SSc; scleroderma), a disease for which all of these tenets are true, and included samples from patients with pulmonary fibrosis (PF) and pulmonary arterial hypertension (PAH). SSc is characterized by abnormal vasculature, adaptive immune dysfunction (autoantibody production), and extracellular matrix (ECM) deposition in skin and internal organs. The etiology of SSc is unknown, but it has complex genetic risk (7) and postulated triggers include immune activation by cancer (8), infection (9), or dysbiosis (10). SSc is clinically heterogeneous with some patients experiencing rapidly progressive skin and internal organ disease, while others have stable disease that is largely limited to skin. Understanding the drivers of disease in multiple affected organ systems is critical to understand the pathogenesis of SSc and other complications, such as PF and PAH, that co-occur in these patients.

These 'big data' approaches integrate individual experiments measuring hundreds of disease states and biological perturbations. Integration of these data holds promise for understanding how genes contribute to organ specific manifestations of human diseases (2). We previously developed mutual information consensus clustering (MICC) to identify gene expression that is conserved across multiple, disparate datasets (11). Here we expanded MICC to perform an integrative, multi-tissue analysis of SSc and related fibrotic conditions. Following MICC, we used the Genome-scale Integrated Analysis of gene Networks in Tissues (GIANT) tissue-specific functional genomic networks (1) to identify gene-gene interactions among those expressed consistently across affected tissues. These

105 GIANT networks are a detailed, genome-scale representation of the functional interactions between
106 genes in different microenvironments. We included gene expression datasets from ten different cohorts
107 representing four different affected tissues from patients with SSc. We identified a pathogenic signature
108 – a common 'immune-fibrotic axis' – that is present in all tissues analyzed and is increased in the most
109 severe disease complications, including PF and PAH.

110 The immune-fibrotic axis implicates alternatively activated MØs as central drivers of fibrosis in all
111 solid organs studied. MØs are highly plastic cells implicated in a wide range of pathologic processes
112 (12-14). Using tissue-specific functional networks (1), we analyzed the nature of the immune-fibrotic
113 axis to understand the gene-gene interactions that underlie fibrosis across organ systems. Using
114 differential network analysis, we were able to identify skin- and lung-specific gene-gene interactions
115 relevant to MØ plasticity and SSc pathophysiology. We now propose a model that implicates
116 alternatively activated MØs as part of the immune-fibrotic axis that may drive fibrosis in multiple tissues.
117

118 **Results**

119 We performed an integrative analysis of ten independent gene expression datasets containing
120 samples from patients with systemic sclerosis and associated co-morbidities (Table 1). A total of 573
121 samples from 321 subjects recruited at seven independent centers were analyzed. These data
122 represent samples from four different affected tissues derived from seven different clinic centers in the
123 US and Europe. Data include SSc and control skin from a University of California, San Francisco cohort
124 (15), a Boston University cohort (16) and a Northwestern University cohort (17). Many patients in the
125 skin cohorts provided lesional (forearm) and non-lesional (back) skin biopsies; a subset of patients in
126 the Northwestern skin cohort provided biopsies longitudinally over time as part of a clinical trial for
127 mycophenolate mofetil (MMF). Peripheral blood mononuclear cell (PBMC) samples from patients with
128 and without SSc-PAH, patients with idiopathic PAH (IPAH) and healthy controls were included from a
129 Boston University cohort (18) and a University of Colorado PAH cohort (19). Lung data contained a
130 cohort of late or end-stage patients that underwent lung transplant at the University of Pittsburgh (20)
131 and a second cohort from open lung biopsies from early SSc-associated PF obtained in Brazil (21). The

132 lung biopsies included patients with SSc-associated PF, idiopathic PF (IPF), SSc-associated PAH, and
133 idiopathic PAH (IPAH). Data on previously unpublished samples were also included in these analyses.
134 These are two datasets of skin biopsies from patients with limited cutaneous SSc (LSSc) recruited from
135 University College London (UCL) / Royal Free Hospital and Boston University Medical Center. Only
136 data that were judged to be high quality were included in the analyses. To our knowledge, there was no
137 overlap between the patient cohorts beyond 5 patients recruited at Northwestern that provided both
138 skin and esophageal biopsies. We summarize all patient cohorts in S1 Table.

139 The primary goal of this study was to identify the fundamental processes that occur across end-
140 target and peripheral tissues of patients with SSc and related fibrotic conditions. Secondly, we aimed to
141 identify the presence or absence of common gene expression patterns that underlie the molecular
142 intrinsic subsets of SSc (15) in different organs. Analysis of multiple tissue biopsies from patients with
143 skin fibrosis, esophageal dysfunction, PF and PAH, allowed us to determine in an unbiased analysis
144 whether these tissues were perturbed in a similar manner on a genomic scale.

145 We applied MICC (11) to identify conserved, differentially co-expressed genes across all tissues
146 in our SSc compendium. MICC is a 'consensus clustering' procedure, meaning that it identifies the
147 *shared co-clustering of genes* present in multiple datasets. MICC identifies genes that are consistently
148 coexpressed in multiple tissues. Procedurally, MICC clusters gene expression data into coexpression
149 modules using weighted gene correlation network analysis (WGCNA) (Fig 1). Because this clustering is
150 purely data-driven, coexpression modules derived from different datasets necessarily differ from each
151 other. MICC integrates these coexpression modules across datasets by identifying significant overlaps
152 between modules from different datasets and forming a 'module overlap network'. MICC then parses
153 the module overlap network to find sets of modules (communities) that are strongly conserved across
154 many datasets (see Methods). These strongly overlapping modules correspond to molecular processes
155 that are conserved across multiple datasets.

156 All datasets were partitioned into coexpression modules using WGCNA, resulting in 549 modules
157 (Table 2). We constructed the 10-partite module overlap network (Fig 2) and identified eight
158 communities in the network using modularity-based community detection methods. Because the
159 community structure of the module overlap network was hierarchical, we used a hierarchical labeling

160 scheme, where numerals denote large communities and letters denote smaller sub-communities (Fig
161 2A). For each community, we used set theoretic formulae to derive a final gene set ('consensus genes')
162 associated with the modules in that community (see Methods and S2 Table; consensus gene sets
163 ranged from 64-9597 genes in size). The majority of the consensus gene sets pertain to biological
164 processes that are not disease-specific. These include processes such as telomere organization (1A)
165 and macromolecule localization (3A). *Disease-specific* consensus genes were identified by first
166 determining which communities contained modules associated with pathophenotypes under study and
167 then deriving consensus gene sets from those combined communities (see below).

168

169 **Severe pathophenotypes share a common immune-fibrotic axis**

170 The module overlap network is agnostic to the clinical phenotypes corresponding to each biopsy.
171 To associate communities in the module overlap network with SSc and fibrotic pathophenotypes, we
172 tested each of the 549 modules for differential expression in relevant pathophenotypes (see Methods).
173 For example, every lung module in the PAH cohorts was tested for differential expression in PAH.
174 Clusters 4A and 4B in the module overlap network contain modules with increased expression in all
175 pathophenotypes of interest: the inflammatory and proliferative subsets of SSc, PAH, and PF (Fig 2B).
176 Thus, the modules in these communities correspond to a common, broad disease signal that is present
177 in every pathophenotype under study. As with our prior studies, we did not find a strong association
178 with autoantibody subtype and the co-expression modules identified here.

179 Edges in the module overlap graph represent overlap between coexpression modules in different
180 datasets, so we identified the intersection of genes between adjacent modules. We then asked if these
181 'edge gene sets' were similar to known biological processes by computing the Jaccard similarity
182 between edges and canonical pathways from the Molecular Signatures Database (MSigDB; see
183 Methods) (22). Edges in 4A encode immune processes such as antigen processing and presentation
184 and cytotoxic T cell and helper T cell pathways (Table 3). This cluster also contains modules from all
185 tissues, including PBMCs (Fig 2B). Altered immunophenotypes have been reported in SSc-PAH and
186 SSc-PF (18-21). Here, we find that the immune processes with increased expression in these severe
187 pathophenotypes have substantial overlap with each other, as well as with the inflammatory subsets in

188 esophagus and skin (Fig 2B and S1). Notably, 4A is composed of modules with increased expression in
189 PAH in PBMCs and lung, and a module upregulated in end-stage PF (S1 Fig). This demonstrates a
190 commonality of molecular pathways between the inflammatory component of SSc and the most severe
191 end-organ complications at the expression level.

192 Edges in 4B encode pro-fibrotic processes including ECM receptor interaction, collagen
193 formation, and TGF-β signaling (Table 3). Cluster 4B consists of skin inflammatory and fibroproliferative
194 subset-associated modules as well as lung PAH-, late PF- and early PF-associated modules (Fig 2B
195 and S1). These results validate and expand what we have found in our prior meta-analysis of skin data
196 alone (11): the immune-fibrotic axis observed in the SSc intrinsic subsets are *connected* to and,
197 furthermore, *are found in all* other tissues and SSc-associated pathophenotypes.

198 To understand how the immune-fibrotic axis and these phenotypes are functionally related, we
199 identified the consensus genes in the combined 4A and 4B clusters (see Methods; 2079 unique genes;
200 S4 Table). Using a conservative measure, these consensus genes are enriched for genes with
201 increased expression in all disease manifestations (Significance Analysis of Microarrays or SAM (23),
202 FDR <5%) (PF in both lung datasets $p < 2.2e-16$; PAH lung, $p = 7.88x10^{-5}$; PAH in both PBMC
203 datasets, $p = 3.20x10^{-15}$, Fisher's exact test). This demonstrates that the tissue consensus genes are
204 highly relevant to all disease manifestations in this study. The tissue consensus gene sets allow us to
205 rigorously extrapolate from this conservative set a substantially broader, disease-associated signal.
206 This extrapolation is especially important for tissue studies that are underpowered to detect a large
207 number of significantly differentially expressed genes (see Discussion). We took the union of the tissue
208 consensus gene sets as a set of 'immune-fibrotic axis consensus genes' that are informative about
209 pathology in every tissue.

210

211 **The lung functional genomic network reveals a coupling of immune and fibrotic processes**

212 The GIANT functional networks infer functional relationships between genes by integrating
213 publicly available data including genome-wide human expression experiments, physical and genetic
214 interaction data, and phenotype and disease data (1). In these networks, genes are nodes and edges
215 are weighted by the estimated probability of a tissue-specific relationship between genes. GIANT

216 contains networks for multiple tissues, including skin and lung. To investigate the function of the
217 immune-fibrotic axis consensus genes in pulmonary manifestations of SSc, we extracted the
218 subnetwork of the GIANT whole genome lung network corresponding to the immune-fibrotic axis
219 consensus genes – the *lung network* (Fig 3 and S3). Similar to our previous analysis of SSc skin, we
220 find interconnected functional modules related to both immune (interferon (IFN)/antigen presentation
221 and innate immune/NF-κB/apoptotic processes) and fibrotic (response to TGF-β and ECM
222 disassembly/wound healing) processes (Fig 3A). This demonstrates that, like skin, there is functional
223 coupling between inflammatory and pro-fibrotic pathways in lung.
224
225 **The lung network distinguishes early and late events in SSc lung disease**
226 Our analysis includes two lung datasets derived from both early SSc-PF (open lung biopsies
227 obtained for diagnostic purposes (21)) and end-stage or late disease (SSc-PF patients that underwent
228 lung transplantation (20)). In addition to the differences in disease stage between these two datasets,
229 there is also some difference in the histological patterns of fibrosis in these cohorts. In the Bostwick
230 lung dataset (20), all patients with SSc-PF had usual interstitial pneumonia (UIP). This study used lung
231 tissues from patients who underwent lung transplantation (late disease). The Christmann lung dataset
232 (21) contains 5 patients with non-specific interstitial pneumonia (NSIP) and 2 patients with centrilobular
233 fibrosis (CLF). This study looked at early SSc-PF patients, used open lung biopsies, and specifically
234 avoided honeycombing areas.
235 Although NSIP and UIP have distinct clinical outcomes, they have been shown to be nearly
236 indistinguishable at the gene expression level (24). Furthermore, these datasets have overlapping
237 coexpression patterns as demonstrated by their shared community membership in the module overlap
238 network. Comparison of different datasets allows us to determine how genes with increased expression
239 at these different stages and histological subtypes of lung disease are distributed throughout the lung
240 network and to suggest an order of molecular events in SSc-PF progression. Genes overexpressed in
241 SSc-PF (SAM, PF vs. Normal comparison, FDR < 5%) are distributed throughout the lung network and
242 therefore are predicted to participate in all of the molecular processes identified in the network.
243 Quantification of the distribution of SSc-PF differentially expressed genes throughout the consensus

244 lung network (Fig 3B) demonstrates that molecular processes can be associated either with a disease
245 stage or transition between stages. The cell cycle module contains only early SSc-PF genes, the innate
246 immune response/NF-κB/apoptotic processes module contains more late SSc-PF genes, and the
247 response to TGF-β module contains genes from *both* disease stages (Fig 3A-B).
248
249 **Hub and bridge genes are highly relevant to the pathogenesis of pulmonary fibrosis**
250 Certain genes occupy privileged positions within molecular networks and these genes often have
251 critical biological function (25). *Module hub genes* are connected to a significant fraction of genes within
252 a functional module, whereas *bridge genes* are genes that connect to multiple functional modules and
253 thus 'bridge' them. We identified the hub and bridge genes within the lung network for their possible
254 roles in PF pathogenesis. We highlight the hubs and bridges of the lung network in Fig 3C-E and Fig
255 3F, respectively. The hubs of several of the functional modules in the consensus lung network show
256 increased expression at different disease stages (Fig 3C-E). For instance, *LAMC1* shows increased
257 expression in early SSc-PF and is highly connected within the response to TGF-β module (Fig 3C). The
258 gene Niemann-Pick disease, type C2 (*NPC2*) is upregulated in early disease and is connected to
259 cathepsins L and B (*CTSL, CTSB*) and *GLB1* in the lung network (Fig 3D). We tabulate information on
260 selected genes from the lung network in Table 4.
261 The innate immune response/NF-κB signaling/apoptotic process module contains genes that are
262 highly expressed in late SSc-PF, including the hub genes *CYR61* and *TM4SF1* (Fig 3A-B and S3). The
263 hub gene *TNFAIP3* (A20), which is increased in late SSc-PF (Fig 3E), is a negative regulator of NF-κB
264 signaling and inhibitor of TNF-mediated apoptosis. The innate immune response/NF-κB
265 signaling/apoptotic process and IFN/antigen presentation modules are bridged by *TNFSF10,* also
266 known as TRAIL (TNF-related apoptosis inducing ligand, Fig 3F). These results suggest that the
267 balance of apoptosis is altered in late SSc-PF. The upregulation of genes with anti-apoptotic function
268 was not reported in the original study (20), which demonstrates the strength of both the MICC method
269 and the study of functional interactions.
270 *CD44* and *PLAUR* (uPAR) bridge multiple functional modules in the lung network (Fig 3F) and
271 have been implicated in IPF (26,27). Because these genes link modules important in regulating disease

272 progression, therapeutic targeting of CD44 and uPAR may be an effective strategy in combatting SSc-
273 PF. Indeed, anti-CD44 treatment reduces fibroblast invasion and bleomycin-induced lung fibrosis (26),
274 and inhibition of uPAR ligation significantly reduces motility of pulmonary fibroblasts from patients with
275 idiopathic PF (28). These results are consistent with our identification of these genes as key genes in
276 the lung network.
277
278 **The lung microenvironment provides a distinct milieu for pro-fibrotic processes**
279 Pulmonary fibrosis is histologically distinct from skin fibrosis and occurs in a subset of patients
280 with SSc. We hypothesized that the lung microenvironment may have a distinct organization of
281 immune-fibrotic axis consensus genes when compared to skin. Indeed, for interactions (edge weight >
282 0.5) that are present in both the lung and skin networks, there are gene pairs that are much more likely
283 to interact in one tissue than the other (Fig 4A). In other words, the skin and lung networks are 'wired
284 differently'. To identify *highly lung-specific* and *highly skin-specific interactions*, we performed a
285 differential network analysis that identified gene pairs that are strongly predicted to interact in one
286 tissue but not the other (see Methods).
287 These highly specific interactions are displayed in Fig 4B, where a cell is red if it is lung-specific
288 or blue if it is skin-specific (cf. S4 Fig). The number of tissue-specific edges in each functional module is
289 quantified in Figs 4B and 4C, which illustrate that most functional modules in lung have fewer
290 interactions than in skin, with the exception of the cell cycle module. Of particular interest is the
291 relationship between the phagolysosome/ECM disassembly genes and response to TGF-β genes, as
292 strong differential connectivity can be observed in this module (Figs 4B and 4C). Thus, even though
293 ECM disassembly and TGF-β module genes are coordinately differentially expressed in both lung and
294 skin, they are differentially connected to each other suggesting that the microenvironment strongly
295 determines the functional consequences of upregulating these pro-fibrotic genes.
296 To summarize lung-specific biological processes in the immune-fibrotic axis, we clustered the
297 lung-specific interactions (differential lung network) to identify lung-specific pathways (S5 Fig). We
298 identified 23 clusters corresponding to biological processes such as type I IFN signaling (cluster 10),
299 antigen processing and presentation (cluster 4), REACTOME Cell surface interactions at the vascular

300 wall (cluster 22), and mitotic cell cycle (cluster 16, shown in Fig S5B). Taken together, this suggests
301 that within the immune-fibrotic axis we find innate immune and cell proliferation processes that are
302 highly lung-specific. One of the largest of these clusters (cluster 13, Fig 4D and S5C) includes *NPC2*,
303 *S100A4*, and *CTSB*, which encode protein products that are highly expressed in normal lung-resident
304 MØs (LR-MØs) (29,30).

305 *NPC2*, is a hub of the ECM disassembly/wound healing module in the full lung network (Fig 3);
306 many of the genes in cluster 13 also belong to the ECM disassembly/wound healing module in the
307 whole network, including the cathepsins *CTSB* and *CTSL.* Alveolar MØs are the main source of
308 cathepsins in bleomycin-induced fibrotic lung tissue (31). Additional genes associated with
309 development and maintenance of alternative MØ activation include *TGFBI* (32), *NEU1* (33), *PRCP* (34),
310 and *DAB2* (35). Genes that are specifically associated with alternative activation of lung MØs include
311 *PLP2* (36) and *IFITM1* (37) (Fig 4D and S5C). Based on these genes and the complete lung network in
312 Figure 3, we identified an LR-MØ signature. These findings are consistent with previous reports of
313 alternative MØ activation in SSc (21,38).

314 To explore this signature further, we examined some genes from this cluster along with genes
315 identified in the Christmann, et al. study (21). Consistent with the primary publication (21), some
316 heterogeneity in SSc-PF gene expression is observed and is likely due to tissue sampling from various
317 lobes of the lung as well as the inclusion of patients with centrilobular fibrosis (Fig 5A, right dendrogram
318 branch). Nevertheless, the LR-MØ signature comprises genes that are highly correlated with canonical
319 markers of alternatively activated MØs that were validated by either PCR or immunohistochemistry in
320 the original study (e.g., *CD163* and CCL18) (21).

321 The LR-MØ cluster in the differential lung network also contains a number of genes implicated in
322 lipid storage disorders, including *HEXB*, *GLB1*, and *NPC2*. Several other LR-MØ cluster genes have
323 been shown to be important for regulating cholesterol trafficking genes in an animal model of obesity,
324 including *CTSB*, *CTSL*, and *NPC2* (39). It has been noted that lipid metabolism genes are upregulated
325 in lung MØs relative to other tissue-specific MØs (36). Furthermore, in the bleomycin injury mouse
326 model of pulmonary fibrosis, lipid-laden MØs have been observed to increase expression of markers
327 associated with alternative MØ activation and to secrete TGF-β (40).

328

329 **Distinct MØ gene expression programs are elevated in lung and skin**

330 We hypothesized that early SSc-PF lung samples may have evidence of both alternatively
331 activated and lipid-stimulated MØs and that this may differ from what is observed in skin. The presence
332 of alternatively activated MØs in the inflammatory subset of skin was inferred in our single tissue
333 analysis (11). To test this hypothesis, we used gene sets associated with classical activation of MØs,
334 alternative activation of MØs, or stimulation of MØs with a variety of activation stimuli, including free
335 fatty acids, taken from Xue, et al. (12). To summarize the expression of each MØ gene set (12) and
336 compare across tissues in these data, we computed the average expression of all genes in each gene
337 set (see Methods; see S5 Table for a mapping between Xue, et al. modules and our naming scheme).
338 Results are displayed for control and SSc-PF lung, as well as control and SSc-inflammatory skin (Fig
339 5B). As shown in Figure 5B, there is evidence of an increase in alternatively activated and free fatty
340 acid stimulated gene sets in SSc-PF and SSc-inflammatory skin. These data do not show statistically
341 significant differences in expression of gene sets associated with classical MØ activation between
342 controls and SSc-PF or SSc-inflammatory skin (see S6 Table for p-values of all modules tested).

343 The discovery of IFN (IFN)-related genes among the consensus genes indicates that these
344 pathways are increased in pathophenotypes of interest (e.g., SSc-PF and the skin inflammatory
345 subset). Christmann, et al. also noted a strong IFN-related gene signature in SSc-PF samples, although
346 the cellular compartment responsible for this signature was not described (21). Because stimulation
347 with IFN results in classical activation of MØs, we examined the expression of genes from CL 1, as it is
348 most strongly associated with IFN-γ treatment ("classical activation") in human MØs (12). However, CL
349 1 genes' expression is not different between disease and controls in either skin or lung (Wilcoxon p =
350 0.76 and 0.80, respectively; Fig 5B). This result is consistent with our inability to discern differences in
351 classical MØ activation markers between controls and SSc-PF and inflammatory skin and suggests that
352 classically activated MØs are not the source of the reported IFN signature.

353 Modules ALT 1 and ALT 2 are both associated with IL-4 and IL-13 treatment, which are stimuli
354 associated with alternative activation of MØs (12). These two gene sets are non-overlapping
355 coexpression modules and therefore represent two "parts" of the alternatively activated MØ

356 transcriptional program. We performed functional enrichment analysis for ALT 1 and 2 to understand
357 which biological processes underlie these transcriptional signatures (see Methods). Module ALT 1 is
358 enriched for genes involved in oxidative phosphorylation (KEGG, $p < 0.0001$) and the citric acid cycle
359 (REACTOME, $p < 0.0001$) pathways. In lung, ALT 1 expression is higher in SSc-PF than in controls
360 (Wilcoxon $p = 0.0046$). There is no difference between healthy controls and the inflammatory subset in
361 skin (Wilcoxon $p = 0.41$). Module ALT 2 shows an opposite trend is enriched for genes implicated in the
362 positive regulation of response to wounding (GO BP, $p = 0.027$) and defense response (GO BP, $p =$
363 0.00035); this module includes alternatively activated MØ markers such as *CD14* and *CCL26* (41,42).
364 ALT 2 expression is increased in the inflammatory subset in skin (Wilcoxon $p = 0.041$) and trends
365 toward decreased expression in SSc-PF lung (Wilcoxon $p = 0.16$). Together, these pathways suggest a
366 metabolic "switch" associated with alternative activation in lung that is not found in skin (for review see
367 (43); Fig 5B).

368 We also analyzed modules associated with free fatty acids (FFA) stimulation, which are relevant
369 to the question of lipid signaling or exposure in SSc tissues (FFA 1, 2, and 3). We first performed
370 functional enrichment analysis for these modules to gain biological insight into these transcriptional
371 programs. FFA 1 is enriched for genes involved in the Unfolded Protein Response (REACTOME, $p =$
372 0.025). FFA 2 is enriched for Antigen processing-Cross presentation genes (REACTOME; $p =$
373 0.00101). FFA 3 is enriched for genes in the ER-Phagosome Pathway (REACTOME, $p = 0.0076$).
374 Expression of FFA 1 and 2 is significantly increased in lung (FFA 1: Wilcoxon $p = 0.046$; $p = 0.97$ in
375 skin; FFA 2: Wilcoxon $p = 0.0013$; $p = 0.63$ in skin), whereas FFA 3 is upregulated in SSc-PF lung
376 (Wilcoxon, $p = 0.0013$) and the SSc inflammatory subset in skin (Wilcoxon, $p = 0.00056$). These results
377 suggest that LR- MØs may have a distinct lipid exposure that strongly diverges from that in skin.

378 The differential network analysis (Fig. 4) allowed us to identify highly lung-specific interactions in
379 the immune-fibrotic axis that implicated lipid signaling as a distinct functional process in lung. The
380 higher expression of *multiple* free fatty acid-associated modules in lung suggests that the role of lipid
381 signaling in MØs may be more important in this tissue than in skin, consistent with what we would
382 predict based on *highly lung-specific* gene-gene interactions, and based on prior biomedical literature in

383 related conditions (36,40). Thus, a major difference between the lung and skin networks can be
384 attributed to the presence of a distinct MØ phenotype in lungs.
385

386 ## Discussion

387 SSc is a systemic disease that affects multiple internal organs but, to our knowledge, no one has
388 shown if there are distinct or common deregulated pathways between these organ systems, or their
389 relationship to other fibrotic conditions. In recent years, gene expression data have been collected for
390 multiple tissues. However, these data often have issues that are common to many diseases. First, SSc
391 is rare and patients with particular disease manifestations are still rarer, so there is a limit to the amount
392 of biopsy material available for study. Second, for practical and ethical reasons, internal organ biopsies
393 are seldom taken from healthy subjects making comparisons difficult. Thus, lung, esophagus, and other
394 affected internal organs are more difficult to study than blood and skin tissue. Therefore, there is a
395 critical need to leverage our biological prior knowledge with our understanding of well-studied tissues –
396 like blood and skin – to make plausible inferences about pathogenesis in tissues that are more difficult
397 to study.

398 The clinical heterogeneity of SSc, particularly the difficulty of predicting internal organ
399 involvement, raises an important question: are the fibrotic processes observed in multiple organs
400 derived from a common disease process, or is each organ manifestation effectively a distinct disease?
401 Our analyses demonstrate that there is a common gene expression signature underlying all severe
402 organ manifestations of SSc – the immune-fibrotic axis – in solid organs. The immune-fibrotic axis
403 underlies both SSc pulmonary manifestations of PF and PAH, and the intrinsic subsets of skin and
404 esophagus. Moreover, coexpression modules from peripheral blood, a mixture of innate and adaptive
405 immune cells, have significant overlap with modules associated with all pathophenotypes studied.
406 Thus, while fibrotic processes were largely associated with solid tissues, the inflammatory component
407 of the immune-fibrotic axis is only found in peripheral blood.

408 The presence of a common gene expression signature across multiple tissues suggests a
409 common disease driver, but it does not resolve the possible tissue-specific processes that contribute to

410 disease in the internal organs. Indeed, there are many layers of biological regulation between gene
411 expression and whole tissue phenotypes. Resolving the relationship between molecular profiles and
412 phenotypes is a difficult biological problem underlying most biomedical inquiry. However, these
413 relationships have been approximated by integrating high-throughput genomic data into tissue-specific
414 functional networks using 'big data' machine learning strategies (1). We addressed tissue-specificity in
415 SSc pathology by interpreting the common expression signal – the immune-fibrotic axis – within these
416 tissue-specific functional networks. These networks allowed us to identify critical genes that occupy
417 important positions in molecular pathways in lung. It is clear from this work that the coupling of immune
418 and fibrotic processes is a hallmark of SSc that occurs in SSc-PF and SSc-PAH as well as skin.
419 However, we also find subtle, lung-specific functional differences that we attribute, in part, to the
420 plasticity of the myeloid cell lineage.
421
422 **The plasticity of the myeloid lineage may drive tissue-specific SSc disease processes**
423 By performing a combined analysis of SSc gene expression in multiple tissues, we are able to
424 observe and infer, in a genome-wide manner, commonalities in the complex mixture of cell types in a
425 tissue at the time of biopsy. Overwhelmingly, we detected a MØ signature associated with severe
426 disease. In the module overlap network, we find that PAH-associated modules from PBMCs (18,19)
427 have significant overlap with SSc inflammatory subset-associated modules from skin and esophagus
428 (Fig 2). Indeed, in Pendergrass et al. *(18)*, we observed that PBMCs from lcSSc patients have
429 significant enrichment in myeloid- and MØ-related gene sets as compared to healthy controls.
430 Christmann et al. (44) expanded on this, showing that highly expressed transcripts in lcSSc-PAH
431 $CD14^+$ monocytes were induced in IL-13-stimulated cells, i.e. that PAH monocytes are alternatively
432 activated. We assert that this MØ polarization is a significant part of the immune-fibrotic axis we find in
433 these data and, therefore, is likely *a common driver* of the complex pathophysiology of SSc. In support
434 of this, an independent study also identified MØs and dendritic cells (DCs) as possible sources of an
435 "inflammatory" signature in lesional SSc skin (45).
436 We found evidence for the contribution of LR-MØs to SSc-PF pathobiology, consistent with the
437 alternative activation of MØs and TGF-β production. In our prior analysis of skin, we inferred

438 alternatively activated MØs as modulators of the SSc inflammatory intrinsic subset in skin (11). Our
439 current study identifies a LR-MØ signature within the functional relationships of immune-fibrotic axis
440 consensus genes in lung (Fig 4D and 5A). We posit that the differences in fibrotic responses of skin
441 and lung tissue are due, in large part, to innate differences between tissue-resident MØs that have
442 been observed (46,47), as well as the interactions between infiltrating monocytes and tissue-resident
443 cell types (e.g., alveolar epithelial cells vs. keratinocytes). Because MØ phenotype and function are
444 plastic and readily modulated by the local tissue microenvironment, it is likely that differential activation
445 of MØs in these tissues is the result of exposure to distinct cytokine milieu. Indeed, we show that
446 distinct alternative activation gene expression programs have increased expression in SSc-PF lung and
447 inflammatory SSc skin (Fig. 5). In particular, there were multiple lipid-related signatures elevated in
448 SSc-PF lung alone.

449 We cannot rule out that the MØ changes we observe are a secondary response to the affected
450 organ pathology. Regardless, therapies that target MØ effectors such as IL6R have shown promise in
451 clinical trials (48) and MØ chemoattractants have been shown to be important in animal models of SSc
452 inflammatory disease (49), suggesting that MØs play a central role in SSc pathogenesis. We also
453 cannot rule out that DCs contribute to our results, as plasmacytoid DCs are observed to be important in
454 the Stiff Skin Syndrome mouse model (50). However, some skin-resident DCs have been shown to be
455 transcriptionally similar to peripheral blood monocytes in humans (51). We speculate that the circulation
456 of peripheral myeloid cells contributes to the multi-organ nature of SSc. Future studies may use *in silico*
457 and cell sorting techniques to deconvolve SSc expression data to identify changes in cell proportion
458 and transcriptome throughout disease course and to finely phenotype myeloid cells from SSc patient
459 tissue samples.

460

461 **An overview of SSc-PF disease processes**

462 The study of two different lung datasets that sampled early- and late-stage SSc-PF allows us to
463 describe differences between the disease processes found in these two datasets. The two datasets
464 each contained patients with different types of interstitial pneumonia (see Methods), which may limit
465 interpretation of these results. However, as stated in the results, we and others (24) find evidence of

466 highly similar gene expression patterns between UIP and NSIP. We do not have treatment information
467 for patients in these studies and acknowledge that late-stage patients are more likely to be treated with
468 immunosuppressive therapy. With these caveats in mind, we can nevertheless draw non-intuitive
469 conclusions through the combination of our data-driven approach and mechanistic insight from
470 disparate literature. We provide an overview of disease processes in Fig 6.

471 Christmann and coworkers identified an increase in IFN- and TGF-β-regulated genes in biopsies
472 from early SSc-PF (21). It was also noted that there was more CCL18 at the protein-level and a higher
473 level of *CD163* transcript in SSc-ILD lungs, suggestive of the presence of alternatively activated MØs
474 (21). However, it was unclear which cell types were responsible for the IFN signature or if there was
475 evidence of distinct subpopulations of MØs. We found that gene signatures that are upregulated in
476 alternatively-activated human MØs and MØs treated with free fatty acids are enriched in early SSc-PF
477 patients and that there is no evidence for enrichment of a pro-inflammatory, IFN-stimulated MØ
478 signature (Fig 5) (12).

479 The LR-MØ signature identified in our differential network analysis consisted of genes with
480 increased expression in early SSc-PF that participate in lipid and cholesterol trafficking (Figs 4D, S6,
481 differential lung network). The expression of these genes is correlated with "canonical" MØ genes
482 identified in the primary publication (21) (Fig 5). We find elevated gene expression programs associated
483 with MØ alternative activation (specifically metabolic "reprogramming") and lipid exposure in this
484 dataset (Fig 5). In the bleomycin injury mouse model of pulmonary fibrosis, lipid-laden MØs, or foam
485 cells, have been observed to upregulate markers associated with alternative MØ activation and to
486 secrete TGF-β (59). Oxidized phospholipid treatment also causes alternative activation and TGF-β
487 secretion in human MØs (40). Consistent with this report, recent work demonstrates that foam cell
488 formation *in vivo* favors the development of a pro-fibrotic MØ activation profile (52,53). These studies,
489 along with our results, suggest that lipid exposure or uptake in MØs may be important.

490 TGF-β signaling is a hallmark of fibrotic disease, and was noted in the initial analysis of both lung
491 datasets (20,21). Similarly, we find genes from both datasets in the response to TGF-β module of the
492 lung network. However, we also find evidence that the type I IFN signature is present in the Bostwick
493 dataset(Fig 3). The functional module most strongly associated with late stage disease/UIP is the

494 innate immune, NF-κB, and apoptotic processes module. This module is connected to the TGF-β
495 module through components of the fibrinolysis pathway such as PAI-1 (*SERPINE1*) (Fig 3). PAI-1 is
496 upregulated in late stage SSc-PF and is known to be important in pulmonary fibrosis (54-56). One
497 mechanism by which fibrinolysis may contribute to the resolution of fibrosis is through the induction of
498 fibroblast apoptosis (57). Both TGF-β1 and PAI-1 have been shown to inhibit lung fibroblast apoptosis
499 (57).

500 We found evidence for a shift in the balance of apoptosis in the Bostwick dataset, perhaps in
501 myofibroblasts (58), in our network analyses (Fig 6). Long-lived myofibroblasts are thought to
502 continually deposit collagen and contribute to persistent fibrosis (59). This apoptotic-resistance
503 phenotype is related to the stiffness of the matrix (60), suggesting that a shift in apoptotic processes
504 may occur once the deposition of excess collagen begins. Moreover, impaired phagocytosis of
505 apoptotic cells, or efferocytosis, has been observed in the alveolar MØs of IPF patients (61). We find
506 genes involved in efferocytosis, specifically in receptors (*CD44*) and endocytic machinery associated
507 with this process, in the lung network (Figs 3, 6) (62). If the shift in apoptosis and efferocytosis occurs,
508 we speculate that the fibrotic and inflammatory processes in our network will also be altered.
509 Efferocytosis by alveolar MØs plays a key role in the resolution of inflammation in the lung through the
510 subsequent release of TGF-β (63). We hypothesize that following initial injury, TGF-β signaling,
511 antifibrinolytic factors, and the disruption of apoptosis and efferocytosis may contribute to progressive
512 fibrosis in SSc-PF (Fig 6).

513

514 **Conclusions**

515 In this study, we have utilized data from multiple tissues to examine the systemic nature of SSc.
516 Our integrative analysis allowed us to leverage well-studied tissues to inform us about SSc
517 manifestations that are under-studied molecularly. This study rigorously tests the notion that patients
518 with severe disease have shared immunological and fibrotic alterations. The common immune-fibrotic
519 axis shows evidence for alternatively activated MØs in multiple SSc tissues. However, there are subtle
520 differences in the MØ gene expression programs detected in skin and lung. Different
521 microenvironments likely provide distinct stimuli to infiltrating MØs that determine the pro-fibrotic

522 character of these cells. The plasticity of this lineage is likely central to the divergence of fibrotic
523 processes in multiple SSc-affected tissues and is a central component of an immune-fibrotic axis
524 driving disease.
525

526 # Methods

527 **Patients and datasets**

528 Eight out of 10 datasets included in this study were previously published (see Table 1) and
529 descriptions of the patient populations and criteria for inclusion can be found in those publications. We
530 used the patient disease label (e.g., PAH) as published in the original work for all of these sets. In Table
531 S1, we summarize the patient information to which we had access on a per array basis as that is what
532 is required for comparison to the expression data. Below, we note some important characteristics (for
533 the purposes of this work) of the included patient populations. As noted in the Results section, the two
534 lung datasets contained patients with different histological patterns of lung disease. Some patients
535 included in the PBMC dataset, including those with PAH, also had interstitial lung disease, though
536 exclusion of these patients does not significantly change the interpretation as put forth in (18). As
537 illustrated in S1 Table, two datasets (ESO, LSSc) did not contain healthy control samples and three
538 datasets (UCL, LSSc, and PBMC) were comprised entirely of lcSSc patients.

539

540 **Ethics statement on previously unpublished datasets**

541 The LSSc and UCL studies are previously unpublished. The samples from the LSSc dataset were
542 obtained at Boston University Medical Center (BUMC)/Boston Medical Center (BMC); the BUMC/BMC
543 Institution Review Board approved this study. The samples from the UCL dataset were obtained at
544 University College of London; the London-Hampstead NRES Committee approved this study. The
545 Dartmouth College CPHS approved this work. All subjects gave informed consent. All research
546 conformed to the principles expressed in the Declaration of Helinski.

547

548

549 **Microarray dataset processing**

550 This work contains 10 datasets on multiple microarray platforms. Agilent datasets (Pendergrass,
551 PBMC, Milano, Hinchcliff, ESO, UCL, LSSc) used either Agilent Whole Human Genome (4x44K)
552 Microarrays (G4112F)(Pendergrass, PBMC, Milano, Hinchcliff, ESO, UCL) or 8x60K (LSSc). Data were

553 Log_2-transformed and lowess normalized and filtered for probes with intensity 2-fold over local
554 background in Cy3 or Cy5 channels. Data were multiplied by -1 to convert to Log_2(Cy3/Cy5) ratios.
555 Probes with >20% missing data were excluded. The Illumina dataset (Bostwick, HumanRef-8 v3.0
556 BeadChips) was processed using variance-stabilizing transformation and robust spline normalization
557 using the lumi R package. Dr. Christmann provided the raw data in the form of .CEL files. Dr. Feghali-
558 Bostwick provided Illumina BeadSummary files. Affymetrix datasets (Risbano, HGU133plus2;
559 Christmann, HGU133A_2) were processed using the RMA method as implemented in the affy R
560 package. Batch bias was detected in the ESO dataset. To adjust these data, missing values were
561 imputed via *k*-nearest neighbor algorithm using a GenePattern (64) module with default parameters and
562 the data were adjusted using ComBat (65) run as a GenePattern module to eliminate the batch effect.
563 To compare datasets in our downstream analysis, duplicate genes must not be present in the
564 dataset and must be summarized in some way. First, we annotated each probe with its Entrez gene ID.
565 Agilent 4x44K arrays were annotated using the hgug4112a.db Bioconductor package. LSSc was
566 annotated using UNC Microarray Database with annotations from the manufacturer. Probes annotated
567 to lincRNAs (A19) were removed from the analysis. The Illumina dataset was annotated by converting
568 the gene symbols (provided as part of the BeadSummary file) to Entrez IDs using the org.Hs.eg.db
569 package. The Risbano PBMC dataset was annotated using the hgu133plus2.db package. The
570 Christmann dataset was annotated using an annotation file from the manufacturer. NAs and probes that
571 mapped to multiple Entrez IDs were removed in all cases. Probes that mapped to the same Entrez ID
572 were collapsed to the gene mean using the aggregate function in R, followed by gene median
573 centering.
574
575 **Clustering of microarray data and statistical tests for phenotype association**
576 The collapsed datasets were used to find coherent coexpression modules. We used Weighted
577 Gene Co-expression Network Analysis (WGCNA), a strong clustering method, which allows us to
578 automatically detect the number of coexpression modules and remove outliers (66). Each dataset was
579 clustered using the blockwiseModules function in WGCNA R package using the signed network option
580 and power = 12; all other parameters were set to default. The number of arrays and resulting co-

581 expression modules are summarized in Table 2. Using the WGCNA coexpression modules also
582 reduces the dimensionality of the dataset, as it allows us to test for genes' association with, or
583 differential expression in, a particular pathophenotype of interest on the order of tens, rather than
584 thousands using the module eigengene. The module eigengene is the first principal component, and
585 represents the expression of all genes in a module and an idealized hub of the coexpression module.
586 We used the moduleEigengenes function in the WGCNA R package to extract the eigengenes. A
587 module was considered to be pathophenotype-associated if the module eigengene was significantly
588 differentially expressed in or significantly correlated with a pathophenotype of interest. Only 2-class
589 categorical variables were considered using a Mann-Whitney U test (i.e., all pulmonary fibrosis and
590 pulmonary arterial hypertension patients were grouped together regardless of underlying etiology). We
591 used Spearman correlation for continuous values. P-values were Bonferroni-corrected on a per
592 phenotype basis. See S1 File for complete output. In the main text, we discuss categorical
593 pathophenotypes, as these were enriched at the consensus cluster level. We do find instances
594 coexpression modules that are associated with continuous pathophenotypes, such as pulmonary
595 function test measurements, but these were not apparent at the consensus cluster level of abstraction.
596

597 **Module overlap network construction and community detection**

598 The 10-partite 'module overlap network' was constructed as in Mahoney et al. (23), where it was
599 called the 'information graph' due to its relationship to information theory. We describe the method here
600 in brief and refer to (11) for motivating details. The modules from different datasets have no *a priori*
601 relationship to each other. The module overlap network encodes the pairs of modules that significantly
602 overlap. Specifically, for each pair of modules (C_i and C_j) we compute an overlap score

603 $$W_{ij} = \frac{|C_i \cap C_j|}{N} \log \frac{|C_i \cap C_j| N}{|C_i||C_j|} \quad (1)$$

604 where N is the total number of genes shared between the two datasets. The overlap scores can be
605 positive, negative, or zero, indicating that the modules overlap more, less, or the same as expected at
606 random, respectively. As shown in Mahoney, et al. (11), the overlap scores can be naturally

607 thresholded using information theory to yield a sparse network of significant overlaps. This is the
608 module overlap network.

609 The module overlap network is highly structured. For example, a module representing an
610 inflammatory process in skin often significantly overlaps inflammatory modules in other tissues. Thus,
611 the structure of the module overlap network corresponds to the biological processes that are common
612 to multiple datasets. We can identify these processes by clustering the module overlap network itself.
613 To detect clusters in the module overlap network, we used two methods of community detection in the
614 iGraph R package (67). First, we used fast-greedy modularity optimization (68), which yielded large,
615 diffuse communities. We call these 'top-level' communities. To find smaller, more densely connected
616 sub-communities, we used spin-glass community detection (igraph R package implementation,
617 gamma.minus = 0.125, all other parameters were set to default) (67,69). We call these 'bottom-level'
618 communities. The community/sub-community structure of the module overlap network demonstrates
619 that there is a hierarchy of biological processes that are common across datasets, where large
620 communities contain smaller ones (Fig. 2). To display this hierarchical community structure, we first
621 sorted by top-level community label, and then within each community we sorted by bottom-level label.
622 The adjacency matrix of the module overlap network and its node attributes (including dataset of origin
623 and community labels) are supplied in S2 File.

624 We also tested each top-level community in the module overlap network for enrichment of
625 pathophenotype-associated modules for each phenotype of interest using a Fisher's exact test followed
626 by Bonferroni correction (Table 5). This test takes into account both modules that had increased and
627 decreased in pathophenotypes under study.

628

629 **Functional and pathphenotype annotation of the module overlap network**

630 The module overlap network contains rich information about the biological processes that are
631 active in each tissue under study. We functionally annotated the module overlap network by finding
632 pathways that strongly correlate to each community. Because an edge in the module overlap network
633 corresponds to a significant overlap between coexpression modules from different datasets, we can

634 think of an edge 'encoding' that overlap as a gene set. For each pair of coexpression modules C_i and
635 C_j, we define an 'edge gene set', E_{ij}, as the overlap between the in two datasets

636 $$E_{ij} = C_i \cap C_j \quad (2)$$

637 To annotate this edge gene set with biological pathways, we computed the Jaccard similarity of an
638 edge gene set E and a pathway P

639 $$J(E,P) = \frac{|E \cap P|}{|E \cup P|} \quad (3)$$

640 We used biological pathways from the Kyoto Encyclopedia of Genes and Genomes (70), BioCarta, and
641 Reactome (71) obtained from Molecular Signatures Database from the Broad Institute
642 (software.broadinstitute.org/gsea/msigdb). The Jaccard similarity between the edge and pathway will
643 be equal to one, if all of the genes shared between two modules are exactly the same set of genes
644 annotated to the pathway, or zero if no genes are shared between the two sets. To functionally
645 annotate a community in the information graph, we compared the Jaccard similarities of the edges
646 within the community to edges outside of the community using a Mann-Whitney U test (with Bonferroni
647 adjustment). The full results of this analysis are included as S3 File.

648

649 **Tissue consensus gene sets**

650 To understand how the immune and fibrotic responses in these phenotypes are functionally
651 related, we found the consensus genes in the combined 4A and 4B clusters. Tissue consensus gene
652 sets were derived by considering all modules within 4A and 4B, finding their unions within their dataset,
653 and then computing their intersection across datasets from the same tissue of origin. For example, the
654 lung consensus gene set (CC_{lung}) was derived by computing the union of the Christmann (denoted *c*)
655 and Bostwick (denoted *b*) modules in 4AB separately, and then computing the intersection across these
656 two datasets:

657 $$CC_{lung} = \left(\bigcup_{c \in C_{4AB}} c \right) \cap \left(\bigcap_{b \in B_{4AB}} b \right) \quad (4)$$

658

659 As each tissue was considered separately (limited skin and diffuse skin were considered
660 separately), 5 tissue consensus gene sets were generated; the union of these tissue consensus
661 datasets was used to query the functional genomic networks and is referred to as the 'immune-fibrotic
662 axis consensus' gene set or genes throughout the text. For all genes in modules in clusters 4A and 4B,
663 we calculated the Pearson correlation to their respective module eigengene (kME). We compared the
664 kME of consensus genes to that of non-consensus genes using a Mann-Whitney U test. S3 Table
665 contains the tissue consensus genes from 4AB or the 'IMMUNE-FIBROTIC AXIS consensus genes.'
666

667 **Querying GIANT functional networks, single tissue network analysis, and network visualization**

668 The GIANT functional genomic networks were obtained as binary (.dab) files and processed
669 using the Sleipnir library for computational functional genomics (72). We queried all networks (lung,
670 skin, 'all tissue') using the immune-fibrotic axis consensus gene sets (as Entrez IDs) and pruned all low
671 probability (< 0.5) edges. All networks are available for download from the GIANT webserver
672 (giant.princeton.edu) (1). For each single tissue analysis (consensus lung and consensus skin
673 networks), we considered only the largest connected component of each network and performed spin-
674 glass community detection as implemented in the igraph R package (67) to obtain the functional
675 modules. We annotated functional modules using g:Profiler (73) using all genes in a module as a query.
676 All networks in this work were visualized using Gephi (74). The network layout was determined by
677 community membership, the strength of connections between communities, and finally the interactions
678 between individual genes.
679

680 **Differential network analysis**

681 The tissue-specific networks from GIANT allow for the analysis of the differing functional connectivity
682 between genes in different microenvironments. In order to understand the specific immune-fibrotic
683 connectivity in lung relative to skin, we performed a differential network analysis (Fig 4). To compare
684 networks we retained only nodes common to consensus skin network and consensus lung largest
685 connected components (see above). We define the 'differential lung network' as the network with
686 adjacency matrix:

687 $A_{diff} = \max(A_{lung} - \max(A_{skin}, A_{global}), 0)$ (5)

688 where A_{lung}, A_{skin}, and A_{global} are the lung, skin, and global (all tissues) adjacency matrices from GIANT.
689 The differential lung network is thus the lung network minus the maximum edge weight from the skin
690 and lung networks, where all edges that are stronger in skin or the global network are set to zero. Thus,
691 the differential lung network contains only highly lung-specific interactions. Functional modules in the
692 lung differential network were found using spin-glass community detection (see above) within the
693 largest connected component of the network.

694

695 **Differential expression and MØ gene set analysis**

696 To identify genes that were differentially expressed in SSc-PF, SSc-PF samples were compared
697 to normal controls in both datasets using SAM (23) (1000 permutations, implemented in samr R
698 package). Genes with an FDR < 5% were considered further. The MØ gene sets used in this study are
699 WGCNA modules taken from a study of human MØ transcriptomes (12). The z-score of each genes'
700 expression (Eqn. 6) was computed in the collapsed Christmann and Hinchcliff datasets (as described in
701 'Microarray dataset processing' section of Methods). The z-score *z* of gene *g* in the *i*th array/sample is
702 computed as:

703 $$z_{gi} = \frac{x_{gi} - \mu_g}{\sigma_g} \quad (6)$$

704 where x_{gi} is the gene expression value in array/sample *i*, is the μ_g gene mean, and σ_g is the gene
705 standard deviation. The average z-score of genes in a set (module from Xue, et al. (12)). computed for
706 an array/sample to summarize gene set expression. Mann-Whitney U tests were used to compare
707 average z-scores between groups (Fig 5).

708

709

710

711

712 **Acknowledgements**

713 JNT would like to thank members of the Whitfield Laboratory for thoughtful discussion.

714

715 *Funding:*

716 This work has been supported by grants from: the Scleroderma Research Foundation (SRF
717 www.srfcure.org) to MLW and MH; grants from the Scleroderma Foundation to PAP and MLW; the Dr.
718 Ralph and Marian Falk Medical Research Trust Catalyst Award to MLW; Dartmouth SYNERGY award
719 to MLW; the Gordon and Betty Moore Foundation (GBMF4552) to CSG; National Institutes of Health
720 (NIH; www.nih.gov) grants 2R01AR051089 to RL, P50 AR060780 and P30AR061271 to RL and MLW,
721 K23 AR059763 to MH. JNT received support from NIH-NIGMS T32GM008704 and the John H.
722 Copenhaver, Jr. and William H. Thomas, MD 1952 Junior Fellowship from Dartmouth Graduate
723 Studies. CPD received support from the EULAR ODP programme. The funders had no role in study
724 design, data collection and analysis, decision to publish, or preparation of the manuscript.

725

726 **Author contribution**

727 JNT, JMM, and MLW conceived of the study. JNT, CSG, VM, JMM, and MLW designed data analyses,
728 performed analyses, and interpreted the results. TAW performed the microarray experiments. RBC,
729 HWF, RAL, and CPD designed study cohorts included in this work and contributed samples and/or
730 data. MEH provided clinical expertise and interpreted the results. PAP provided macrophage biology
731 expertise and interpreted the results. JNT, PAP, JMM, and MLW wrote the paper. All authors read,
732 revised, and approved the manuscript.

733

734 **Competing interests**

735 CPD has been a consultant to Roche, GlaxoSmithKline, Actelion, Inventiva, CSL Behring, Takeda,
736 Merck-Serono, MedImmune and Biogen. MLW and MH have filed patents for gene expression
737 biomarkers in systemic sclerosis. MLW is a scientific founder of Celdara Medical LLC. MLW has served
738 as consultant to GlaxoSmithKline, Bristol Myers Squib, EMD Serono, Biogen and Quintiles. RL has

739 received both grants and consulting fees from Genzyme/Sanofi, Shire, Regeneron, Biogen, BMS,
740 Inception, Precision Dermatology, PRISM, UCB, Pfizer and Roche/Genentech; he received consulting
741 fees from Lycera, Novartis, Celgene, Amira, Celdara, Celltex, Dart Therapeutics, Idera, Intermune,
742 Medimmune, Promedior, Zwitter, Actelion, EMD Serono, Akros, Extera, Reneo, Scholar Rock, and
743 HGS.
744
745

746 # References
747
748 1. Greene CS, Krishnan A, Wong AK, Ricciotti E, Zelaya RA, Himmelstein DS, et al. Understanding multicellular function and
749 disease with human tissue-specific networks. Nat Genet 2015, Apr 27. 2. Gross AM, Ideker T. Molecular networks in context.
750 Nat Biotechnol 2015;33(7):720-1.
751 3. Hofree M, Shen JP, Carter H, Gross A, Ideker T. Network-based stratification of tumor mutations. Nat Methods 2013,
752 Nov;10(11):1108-15.
753 4. Chen S, Wang Q, Wu Z, Li Y, Li P, Sun F, et al. Genetic association study of TNFAIP3, IFIH1, IRF5 polymorphisms with
754 polymyositis/dermatomyositis in chinese han population. PLoS One 2014;9(10):e110044.
755 5. Chen JC, Cerise JE, Jabbari A, Clynes R, Christiano AM. Master regulators of infiltrate recruitment in autoimmune disease
756 identified through network-based molecular deconvolution. Cell Syst 2015, Nov 25;1(5):326-37.
757 6. Sharma A, Menche J, Huang C, Ort T, Zhou X, Kitsak M, et al. A disease module in the interactome explains disease
758 heterogeneity, drug response and captures novel pathways and genes. Hum Mol Genet 2015:ddv001.
759 7. Assassi S, Wu M, Tan FK, Chang J, Graham TA, Furst DE, et al. Skin gene expression correlates of severity of interstitial
760 lung disease in systemic sclerosis. Arthritis Rheum 2013, Nov;65(11):2917-27.
761 8. Joseph CG, Darrah E, Shah AA, Skora AD, Casciola-Rosen LA, Wigley FM, et al. Association of the autoimmune disease
762 scleroderma with an immunologic response to cancer. Science 2014, Jan 10;343(6167):152-7.
763 9. Farina A, Cirone M, York M, Lenna S, Padilla C, McLaughlin S, et al. Epstein-Barr virus infection induces aberrant TLR
764 activation pathway and fibroblast-myofibroblast conversion in scleroderma. J Invest Dermatol 2014, Apr;134(4):954-64.
765 10. Arron ST, Dimon MT, Li Z, Johnson ME, A Wood T, Feeney L, et al. High rhodotorula sequences in skin transcriptome of
766 patients with diffuse systemic sclerosis. J Invest Dermatol 2014, Aug;134(8):2138-45.
767 11. Mahoney JM, Taroni J, Martyanov V, Wood TA, Greene CS, Pioli PA, et al. Systems level analysis of systemic sclerosis
768 shows a network of immune and profibrotic pathways connected with genetic polymorphisms. PLoS Comput Biol
769 2015;11(1):e1004005.
770 12. Xue J, Schmidt SV, Sander J, Draffehn A, Krebs W, Quester I, et al. Transcriptome-based network analysis reveals a
771 spectrum model of human macrophage activation. Immunity 2014, Feb 20;40(2):274-88.
772 13. Amit I, Winter DR, Jung S. The role of the local environment and epigenetics in shaping macrophage identity and their
773 effect on tissue homeostasis. Nat Immunol 2015, Dec 17;17(1):18-25.
774 14. Okabe Y, Medzhitov R. Tissue biology perspective on macrophages. Nat Immunol 2015, Dec 17;17(1):9-17.
775 15. Milano A, Pendergrass SA, Sargent JL, George LK, McCalmont TH, Connolly MK, Whitfield ML. Molecular subsets in the
776 gene expression signatures of scleroderma skin. PLoS One 2008;3(7):e2696.
777 16. Pendergrass SA, Lemaire R, Francis IP, Mahoney JM, Lafyatis R, Whitfield ML. Intrinsic gene expression subsets of
778 diffuse cutaneous systemic sclerosis are stable in serial skin biopsies. J Invest Dermatol 2012, May;132(5):1363-73.

779 17. Hinchcliff M, Huang CC, Wood TA, Matthew Mahoney J, Martyanov V, Bhattacharyya S, et al. Molecular signatures in skin
780 associated with clinical improvement during mycophenolate treatment in systemic sclerosis. J Invest Dermatol 2013,
781 Aug;133(8):1979-89.
782 18. Pendergrass SA, Hayes E, Farina G, Lemaire R, Farber HW, Whitfield ML, Lafyatis R. Limited systemic sclerosis patients
783 with pulmonary arterial hypertension show biomarkers of inflammation and vascular injury. PLoS One 2010;5(8):e12106.
784 19. Risbano MG, Meadows CA, Coldren CD, Jenkins TJ, Edwards MG, Collier D, et al. Altered immune phenotype in
785 peripheral blood cells of patients with scleroderma-associated pulmonary hypertension. Clin Transl Sci 2010, Oct;3(5):210-8.
786 20. Hsu E, Shi H, Jordan RM, Lyons-Weiler J, Pilewski JM, Feghali-Bostwick CA. Lung tissues in patients with systemic
787 sclerosis have gene expression patterns unique to pulmonary fibrosis and pulmonary hypertension. Arthritis Rheum 2011,
788 Mar;63(3):783-94.
789 21. Christmann RB, Sampaio-Barros P, Stifano G, Borges CL, de Carvalho CR, Kairalla R, et al. Association of interferon- and
790 transforming growth factor β-regulated genes and macrophage activation with systemic sclerosis-related progressive lung
791 fibrosis. Arthritis Rheumatol 2014, Mar;66(3):714-25.
792 22. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: A
793 knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 2005, Oct
794 25;102(43):15545-50.
795 23. Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. Proc Natl
796 Acad Sci U S A 2001, Apr 24;98(9):5116-21.
797 24. Cho JH, Gelinas R, Wang K, Etheridge A, Piper MG, Batte K, et al. Systems biology of interstitial lung diseases:
798 Integration of mrna and microrna expression changes. BMC Med Genomics 2011;4:8.
799 25. Barabási A-L, Gulbahce N, Loscalzo J. Network medicine: A network-based approach to human disease. Nature Reviews
800 Genetics 2011;12(1):56-68.
801 26. Li Y, Jiang D, Liang J, Meltzer EB, Gray A, Miura R, et al. Severe lung fibrosis requires an invasive fibroblast phenotype
802 regulated by hyaluronan and CD44. J Exp Med 2011, Jul 4;208(7):1459-71.
803 27. Grove LM, Southern BD, Jin TH, White KE, Paruchuri S, Harel E, et al. Urokinase-type plasminogen activator receptor
804 (upar) ligation induces a raft-localized integrin signaling switch that mediates the hypermotile phenotype of fibrotic fibroblasts.
805 J Biol Chem 2014, May;289(18):12791-804.
806 28. Grove LM, Southern BD, Jin TH, White KE, Paruchuri S, Harel E, et al. Urokinase-type plasminogen activator receptor
807 (upar) ligation induces a raft-localized integrin signaling switch that mediates the hypermotile phenotype of fibrotic fibroblasts.
808 J Biol Chem 2014, May;289(18):12791-804.
809 29. Martinez FO, Helming L, Milde R, Varin A, Melgert BN, Draijer C, et al. Genetic programs expressed in resting and IL-4
810 alternatively activated mouse and human macrophages: Similarities and differences. Blood 2013, Feb 28;121(9):e57-69.
811 30. Uhlen M, Oksvold P, Fagerberg L, Lundberg E, Jonasson K, Forsberg M, et al. Towards a knowledge-based human
812 protein atlas. Nat Biotechnol 2010;28(12):1248-50.

813 31. Koslowski R, Knoch K, Kuhlisch E, Seidel D, Kasper M. Cathepsins in bleomycin-induced lung injury in rat. Eur Respir J
814 2003, Sep;22(3):427-35.
815 32. Nacu N, Luzina IG, Highsmith K, Lockatell V, Pochetuhen K, Cooper ZA, et al. Macrophages produce tgf-beta-induced
816 (beta-ig-h3) following ingestion of apoptotic cells and regulate MMP14 levels and collagen turnover in fibroblasts. J Immunol
817 2008, Apr 1;180(7):5036-44.
818 33. Seyrantepe V, Iannello A, Liang F, Kanshin E, Jayanth P, Samarani S, et al. Regulation of phagocytosis in macrophages
819 by neuraminidase 1. J Biol Chem 2010, Jan 1;285(1):206-15.
820 34. Jackman HL, Tan F, Schraufnagel D, Dragović T, Dezsö B, Becker RP, Erdös EG. Plasma membrane-bound and
821 lysosomal peptidases in human alveolar macrophages. Am J Respir Cell Mol Biol 1995, Aug;13(2):196-204.
822 35. Bell-Temin H, Culver-Cochran AE, Chaput D, Carlson CM, Kuehl M, Burkhardt BR, et al. Novel molecular insights into
823 classical and alternative activation states of microglia as revealed by stable isotope labeling by amino acids in cell culture
824 (SILAC)-based proteomics. Mol Cell Proteomics 2015, Dec;14(12):3173-84.
825 36. Gautier EL, Shay T, Miller J, Greter M, Jakubzick C, Ivanov S, et al. Gene-expression profiles and transcriptional
826 regulatory pathways that underlie the identity and diversity of mouse tissue macrophages. Nat Immunol 2012,
827 Nov;13(11):1118-28.
828 37. Wang J, Nikrad MP, Travanty EA, Zhou B, Phang T, Gao B, et al. Innate immune response of human alveolar
829 macrophages during influenza A infection. PLoS One 2012;7(3):e29879.
830 38. Higashi-Kuwata N, Makino T, Inoue Y, Takeya M, Ihn H. Alternatively activated macrophages (M2 macrophages) in the
831 skin of patient with localized scleroderma. Exp Dermatol 2009;18(8):727-9.
832 39. Hannaford J, Guo H, Chen X. Involvement of cathepsins B and L in inflammation and cholesterol trafficking protein NPC2
833 secretion in macrophages. Obesity (Silver Spring) 2013, Aug;21(8):1586-95.
834 40. Romero F, Shah D, Duong M, Penn RB, Fessler MB, Madenspacher J, et al. A pneumocyte-macrophage paracrine lipid
835 axis drives the lung toward fibrosis. Am J Respir Cell Mol Biol 2014, Nov 19.
836 41. Clavel C, Ceccato L, Anquetil F, Serre G, Sebbag M. Among human macrophages polarised to different phenotypes, the
837 m-csf-oriented cells present the highest pro-inflammatory response to the rheumatoid arthritis-specific immune complexes
838 containing ACPA. Ann Rheum Dis 2016, Mar 23.
839 42. Stubbs VE, Power C, Patel KD. Regulation of eotaxin-3/CCL26 expression in human monocytic cells. Immunology 2010,
840 May;130(1):74-82.
841 43. O'Neill LA, Pearce EJ. Immunometabolism governs dendritic cell and macrophage function. J Exp Med 2016, Jan
842 11;213(1):15-23.
843 44. Christmann RB, Hayes E, Pendergrass S, Padilla C, Farina G, Affandi AJ, et al. Interferon and alternative activation of
844 monocyte/macrophages in systemic sclerosis-associated pulmonary arterial hypertension. Arthritis Rheum 2011,
845 Jun;63(6):1718-28.

846 45. Assassi S, Swindell WR, Wu M, Tan FD, Khanna D, Furst DE, et al. Dissecting the heterogeneity of skin gene expression
847 patterns in systemic sclerosis. Arthritis Rheumatol 2015, Aug 3.
848 46. Okabe Y, Medzhitov R. Tissue-specific signals control reversible program of localization and functional polarization of
849 macrophages. Cell 2014, May 8;157(4):832-44.
850 47. Gosselin D, Link VM, Romanoski CE, Fonseca GJ, Eichenfield DZ, Spann NJ, et al. Environment drives selection and
851 function of enhancers controlling tissue-specific macrophage identities. Cell 2014, Dec 4;159(6):1327-40.
852 48. Khanna D, Denton CP, Jahreis A, van Laar JM, Cheng S, Spotswood H, et al. OP0054 safety and efficacy of
853 subcutaneous tocilizumab in adults with systemic sclerosis: Week 48 data from the fasscinate trial. Ann Rheum Dis
854 2015;74(Suppl 2):87-8.
855 49. Greenblatt MB, Sargent JL, Farina G, Tsang K, Lafyatis R, Glimcher LH, et al. Interspecies comparison of human and
856 murine scleroderma reveals IL-13 and CCL2 as disease subset-specific targets. Am J Pathol 2012, Mar;180(3):1080-94.
857 50. Gerber EE, Gallo EM, Fontana SC, Davis EC, Wigley FM, Huso DL, Dietz HC. Integrin-modulating therapy prevents
858 fibrosis and autoimmunity in mouse models of scleroderma. Nature 2013, Nov 7;503(7474):126-30.
859 51. McGovern N, Schlitzer A, Gunawan M, Jardine L, Shin A, Poyner E, et al. Human dermal CD14 cells are a transient
860 population of monocyte-derived macrophages. Immunity 2014, Sep 18;41(3):465-77.
861 52. Thomas AC, Eijgelaar WJ, Daemen MJ, Newby AC. Foam cell formation in vivo converts macrophages to a pro-fibrotic
862 phenotype. PLoS One 2015;10(7):e0128163.
863 53. Thomas AC, Eijgelaar WJ, Daemen MJ, Newby AC. The pro-fibrotic and anti-inflammatory foam cell macrophage paradox.
864 Genom Data 2015, Dec;6:136-8.
865 54. Eitzman DT, McCoy RD, Zheng X, Fay WP, Shen T, Ginsburg D, Simon RH. Bleomycin-induced pulmonary fibrosis in
866 transgenic mice that either lack or overexpress the murine plasminogen activator inhibitor-1 gene. Journal of Clinical
867 Investigation 1996;97(1):232.
868 55. Gu nther A, Lu bke N, Ermert M, Schermuly RT, Weissmann N, Breithecker A, et al. Prevention of bleomycin-induced lung
869 fibrosis by aerosolization of heparin or urokinase in rabbits. Am J Respir Crit Care Med 2003;168(11):1358-65.
870 56. Chambers RC. Abnormal wound healing responses in pulmonary fibrosis: Focus on coagulation signalling. Eur Respir Rev
871 2008, Dec;17(109):130-7.
872 57. Horowitz JC, Rogers DS, Simon RH, Sisson TH, Thannickal VJ. Plasminogen activation induced pericellular fibronectin
873 proteolysis promotes fibroblast apoptosis. Am J Respir Cell Mol Biol 2008, Jan;38(1):78-87.
874 58. Moodley YP, Caterina P, Scaffidi AK, Misso NL, Papadimitriou JM, McAnulty RJ, et al. Comparison of the morphological
875 and biochemical changes in normal human lung fibroblasts and fibroblasts derived from lungs of patients with idiopathic
876 pulmonary fibrosis during fasl-induced apoptosis. J Pathol 2004, Apr;202(4):486-95.
877 59. Uhal BD. The role of apoptosis in pulmonary fibrosis. European Respiratory Review 2008, Dec 1;17(109):138-44.
878 60. Liu F, Mih JD, Shea BS, Kho AT, Sharif AS, Tager AM, Tschumperlin DJ. Feedback amplification of fibrosis through matrix
879 stiffening and COX-2 suppression. J Cell Biol 2010, Aug 23;190(4):693-706.

880 61. Morimoto K, Janssen WJ, Terada M. Defective efferocytosis by alveolar macrophages in IPF patients. Respir Med 2012,
881 Dec;106(12):1800-3.
882 62. Vachon E, Martin R, Plumb J, Kwok V, Vandivier RW, Glogauer M, et al. CD44 is a phagocytic receptor. Blood 2006,
883 May;107(10):4149-58.
884 63. Noguera A, Gomez C, Faner R, Cosio B, González-Périz A, Clària J, et al. An investigation of the resolution of
885 inflammation (catabasis) in COPD. Respir Res 2012;13:101.
886 64. Reich M, Liefeld T, Gould J, Lerner J, Tamayo P, Mesirov JP. GenePattern 2.0. Nat Genet 2006;38(5):500-1.
887 65. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical bayes methods.
888 Biostatistics 2007, Jan;8(1):118-27.
889 66. Horvath S, Dong J. Geometric interpretation of gene coexpression network analysis. PLoS Comput Biol
890 2008;4(8):e1000117.
891 67. Csardi G, Nepusz T. The igraph software package for complex network research. InterJournal, Complex Systems
892 2006;1695(5):1-9.
893 68. Newman ME. Modularity and community structure in networks. Proceedings of the National Academy of Sciences
894 2006;103(23):8577-82.
895 69. Reichardt J, Bornholdt S. Statistical mechanics of community detection. Physical Review E 2006;74(1):016110.
896 70. Kanehisa M, Goto S. KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res 2000;28(1):27-30.
897 71. Croft D, Mundo AF, Haw R, Milacic M, Weiser J, Wu G, et al. The reactome pathway knowledgebase. Nucleic Acids Res
898 2014;42(D1):D472-7.
899 72. Huttenhower C, Schroeder M, Chikina MD, Troyanskaya OG. The sleipnir library for computational functional genomics.
900 Bioinformatics 2008, Jul 1;24(13):1559-61.
901 73. Reimand J, Arak T, Vilo J. G: Profilera web server for functional interpretation of gene lists (2011 update). Nucleic Acids
902 Res 2011:gkr378.
903 74. Bastian M, Heymann S, Jacomy M. Gephi: An open source software for exploring and manipulating networks. ICWSM
904 2009;8:361-2.
905 75. Taroni JN, Martyanov V, Huang C, Mahoney JM, Hirano I, Shetuni B, et al. Molecular characterization of systemic
906 sclerosis esophageal pathology identifies inflammatory and proliferative signatures. Arthritis Res Ther 2015, Jul 29;17(1).
907 76. Jung KK, Liu XW, Chirco R, Fridman R, Kim HR. Identification of CD63 as a tissue inhibitor of metalloproteinase-1
908 interacting cell surface protein. EMBO J 2006, Sep 6;25(17):3934-42.
909 77. Matsumoto T, Matsumori H, Taki T, Takagi T, Fukuda Y. Infantile gm1-gangliosidosis with marked manifestation of lungs.
910 Acta Pathol Jpn 1979, Mar;29(2):269-76.
911 78. Roy N, Deveraux QL, Takahashi R, Salvesen GS, Reed JC. The c-iap-1 and c-iap-2 proteins are direct inhibitors of
912 specific caspases. EMBO J 1997, Dec 1;16(23):6914-25.

913 79. Deveraux QL, Roy N, Stennicke HR, Van Arsdale T, Zhou Q, Srinivasula SM, et al. IAPs block apoptotic events induced by
914 caspase-8 and cytochrome c by direct inhibition of distinct caspases. EMBO J 1998, Apr 15;17(8):2215-23.
915 80. Todorovic V, Chen CC, Hay N, Lau LF. The matrix protein CCN1 (CYR61) induces apoptosis in fibroblasts. J Cell Biol
916 2005, Nov 7;171(3):559-68.
917 81. Juric V, Chen C-C, Lau LF. Fas-mediated apoptosis is regulated by the extracellular matrix protein CCN1 (CYR61) in vitro
918 and in vivo. Mol Cell Biol 2009;29(12):3266-79.
919 82. Franzen CA, Chen CC, Todorović V, Juric V, Monzon RI, Lau LF. Matrix protein CCN1 is critical for prostate carcinoma cell
920 proliferation and trail-induced apoptosis. Mol Cancer Res 2009, Jul;7(7):1045-55.
921 83. Rössig L, Haendeler J, Hermann C, Malchow P, Urbich C, Zeiher AM, Dimmeler S. Nitric oxide down-regulates MKP-3
922 mrna levels: Involvement in endothelial cell protection from apoptosis. J Biol Chem 2000, Aug 18;275(33):25502-7.
923 84. Azab NA, Rady HM, Marzouk SA. Elevated serum TRAIL levels in scleroderma patients and its possible association with
924 pulmonary involvement. Clin Rheumatol 2012, Sep;31(9):1359-64.
925 85. Prunier C, Howe PH. Disabled-2 (dab2) is required for transforming growth factor beta-induced epithelial to mesenchymal
926 transition (EMT). J Biol Chem 2005, Apr 29;280(17):17540-8.
927 86. Penheiter SG, Singh RD, Repellin CE, Wilkes MC, Edens M, Howe PH, et al. Type II transforming growth factor-beta
928 receptor recycling is dependent upon the clathrin adaptor protein dab2. Mol Biol Cell 2010, Nov 15;21(22):4009-19.
929 87. Smyth N, Vatansever HS, Murray P, Meyer M, Frie C, Paulsson M, Edgar D. Absence of basement membranes after
930 targeting the LAMC1 gene results in embryonic lethality due to failure of endoderm differentiation. J Cell Biol 1999, Jan
931 11;144(1):151-60.
932 88. Rice LM, Ziemek J, Stratton EA, McLaughlin SR, Padilla CM, Mathes AL, et al. A longitudinal biomarker for the extent of
933 skin disease in patients with diffuse cutaneous systemic sclerosis. Arthritis Rheumatol 2015, Nov;67(11):3004-15.
934

935 # Figure Legends

936 **Fig 1. Schematic overview of analysis pipeline.** Four datasets are shown for simplicity. Each gene expression dataset was
937 partitioned using WGCNA independently to obtain coexpression modules. Module eigengenes were tested for their differential
938 expression in pathophenotypes of interest. Modules were compared across datasets using MICC to form the 'module overlap
939 graph' and community detection algorithms were used to identify communities and subcommunities in the graph. These
940 communities correspond to molecular processes that are conserved across datasets. Each community was examined for
941 enrichment of pathophenotype-associated modules and edge overlap with canonical biological pathways. Gene sets derived
942 from these communities were used to query GIANT functional genomic networks. The resulting networks allow for tissue-
943 specific interrogations of the gene sets. Differential network analysis was performed to compare the lung and skin networks.
944

945 **Fig 2. The multi-tissue module overlap graph demonstrates that severe pathophenotypes have similar underlying**
946 **expression patterns.** (A) The full adjacency matrix of the module overlap graph sorted to reveal hierarchical community
947 structure. A darker cell color is indicative of a higher W score or larger edge weight. Communities (numbered) and sub-
948 communities (lettered) are indicated by the annotation tracks above and on the right side of the matrix, respectively.
949 Coexpression modules with expression that is increased in a phenotype of interest are marked by the annotation bar on the
950 left side of the matrix. If a module was up in SSc as well as another pathophenotype of interest, the other pathophenotype
951 color is displayed. (B) The adjacency matrix of sub-communities 4A and 4B indicates that these clusters contain modules that
952 are up in all pathophenotypes of interest and show that there are many edges between the two sub-communities. Sub-
953 community 4A contains modules from all tissues whereas 4B contains mostly solid tissue modules as indicated by the tissue
954 annotation track to the left of the matrix.

955
956 **Fig 3. Genes that are overexpressed in late and early SSc-PF are distributed throughout the consensus lung network.**
957 (A) The lung network shows functional connections between inflammatory and fibrotic processes. Genes in the largest
958 connected component were clustered into functional modules using community detection. Biological processes associated
959 with the functional modules are in boxes next to the modules. Genes are colored by whether they are over-expressed in late
960 SSc-PF (red), early SSc-PF (blue), both ('SSc-PF', purple), or neither if they are grey. Gene symbols in bold have putative SSc
961 risk polymorphisms. Node (gene) size is determined by degree (number of functional interactions) and edge width is
962 determined by the weight (probability of interaction between pairs of genes). The layout is determined by community
963 membership, the strength of connections between communities, and finally the interactions between individual genes in the
964 network. A fully labeled network is supplied as a supplemental figure intended to be viewed digitally (S3). (B) Quantification of
965 differentially expressed genes in each of the five largest functional modules. C-E. Hubs of the consensus lung network; only
966 the first neighbors of the hub that are in the same functional module are shown. (C) *LAMC1* is a hub of the response to TGF-
967 beta module. (D) *NPC2* is a hub of the ECM disassembly, wound healing module. (E) *TNFAIP3* is a hub of the innate immune

968 response, NF-κB signaling, and apoptotic processes module. (F) Bridges of the consensus lung network. First neighbors of
969 *PLAUR*, *CD44*, *TNFSF10*, and *TGFBI* are shown.
970
971 **Fig 4. The lung and skin network structures indicate distinct tissue microenvironments influence fibrosis.** The skin
972 and lung networks were compared by first finding the giant component of the lung network and then collapsing to nodes only
973 found in both the skin and lung networks (which are termed the common skin and common lung networks). (A) A scatterplot of
974 high probability edges (> 0.5 in both networks) illustrates that pairs of genes with a higher probability of interacting in skin than
975 lung exist and vice versa. Edges are colored red if the weight (probability) is 1.25x higher in lung or blue if it is 1.25x higher in
976 skin. (B) The differential adjacency matrix where a cell is colored if the edge weight in a given tissue is over and above the
977 weight in the global average and tissue comparator networks. For instance, a cell is red if the edge weight was positive
978 following the successive subtraction of the global average weight and skin weight. Community detection was performed on the
979 common lung network to identify functional modules; common functional modules largely recapitulate modules from the full
980 lung network. Representative processes that modules are annotated to are above the adjacency matrix. The annotation track
981 indicates a genes functional module membership. Nodes (genes) are ordered within their community by common lung within
982 community degree. A fully labeled heatmap is supplied as a supplemental figure intended to be viewed digitally (S4). (C)
983 Quantification of tissue-specific interactions in each of the 5 largest functional modules. (D) The lung-resident MØ module
984 found in the differential lung network (consists only of edges in red in panel B).
985
986 **Fig 5. Evidence for alternative activation of MØs in SSc-PF lung that is distinct from .** (A) Genes identified by differential
987 network analysis and inferred to be indicative of lung-resident MØs are correlated with canonical markers of alternatively
988 activated MØs such as *CCL18* and *CD163* in the Christmann dataset. (B) Summarized expression values (mean standardized
989 expression value) of gene sets (coexpression modules) upregulated in various MØ states from the Christmann and Hinchcliff
990 datasets - Module CL1: classical activation (IFN-γ); Modules ALT 1 and 2: alternative activation (IL-4, IL-13); Modules FFA 1,
991 2, and 3: treatment with free fatty acids. Taken from (12).
992
993 **Fig 6**. **Overview of SSc-PF disease processes.** (A) Network-centric overview (B) Cell type-centric overview.
994
995

996 **Table 1. Datasets included in this study.**
997

Dataset label	Tissue	Phenotypes of interest	Citation(s)	GEO Accession
Milano	Diffuse skin	Inflammatory subset, Proliferative subset	Milano et al. (15)	GSE9285
Pendergrass	Diffuse skin	Inflammatory subset, Proliferative subset	Pendergrass et al. (16)	GSE32413
Hinchcliff	Diffuse skin	Inflammatory subset, Proliferative subset	Hinchcliff et al. (17) Mahoney et al. (11)	GSE45485, GSE59785
LSSc	Limited skin	N/A	Present study	GSE76806
UCL	Limited skin	N/A	Present study	GSE76807
Christmann	Lung	SSc-PF	Christmann et al. (21)	GSE76808
Bostwick	Lung	SSc-PF, IPF, IPAH, SSc-PAH	Hsu et al. (20)	GSE48149
Esophagus	Esophagus	Inflammatory subset, Proliferative subset, SSc-PAH	Taroni et al. (75)	GSE68698
PBMC	PBMC	SSc-PAH	Pendergrass et al. (18)	GSE19617
Risbano	PBMC	IPAH, SSc-PAH	Risbano et al. (19)	GSE22356
Abbreviations: PBMC – Peripheral blood mononuclear cells, PAH – Pulmonary arterial hypertension, PF – Pulmonary fibrosis				

998
999 **Table 2. Number of arrays and WGCNA coexpression modules in each of the datasets included**
1000 **in this study.**
1001

Datasets	Number of Arrays	Number of Coexpression Modules
Milano	75	39
Pendergrass	89	38
Hinchcliff	165	62
LSSc	24	39
UCL	15	98
Christmann	18	56
Bostwick	62	54
Esophagus	33	71
PBMC	54	38
Risbano	38	54

1002
1003
1004
1005
1006
1007
1008
1009
1010
1011

1012 **Table 3. Selected pathways that are similar to overlapping coexpression patterns in consensus**
1013 **clusters in the information graph.**
1014
1015

Consensus cluster	Summary of selected pathways
1A	DNA repair Cell cycle RNA metabolism Transcription
2	Cell-cell junction organization Aquaporin mediated transport Tight junctions
3A	Endocytosis mRNA processing Metabolism of proteins
4A	T cytotoxic & helper pathway Antigen processing and presentation Allograft rejection
4B	ECM receptor interaction Collagen formation ECM organization TGF-beta signaling Signaling by PDGF
5	G2 M checkpoint Unwinding of DNA Cell cycle
6	Notch signaling Nuclear receptors in lipid metabolism and toxicity
7	Steroid biosynthesis Fatty acid metabolism PPAR signaling pathway
8	Keratin metabolism FGFR ligand binding and activation

1016
1017
1018 Legend: We calculated the Jaccard similarity index between edges in the information graph and
1019 canonical pathways and used a Mann-Whitney U test to assess whether a particular pathway was more
1020 similar to edges within a consensus cluster than outside the consensus cluster.
1021
1022

1023 **Table 4. Selected genes in the consensus lung network.**
1024

Functional Module	Gene symbol	Description	Network Position	Up in	Function/Potential Role in Disease
cell cycle	*BUB3*	BUB3 Mitotic Checkpoint Protein	-	Early SSc-PF	Encodes a mitotic cell cycle checkpoint protein that regulates the onset of anaphase.
	CDC7	Cell Division Cycle 7	-	-	Regulates MCM complex.
	MCM3	Minichromosome Maintenance Complex Component 3	-	Early SSc-PF	Subunit of minichromosome maintenance (MCM) complex
	MSH6	MutS Homolog 6	-	Early SSc-PF	Participates in DNA mismatch repair.
ECM disassembly/wound healing	*CD44*	CD44 Molecule (Indian Blood Group)	Bridge	-	A hyaluronic acid receptor that can interact with many other ligands found in the ECM. Primary idiopathic PF fibroblasts exhibit an invasive phenotype that was abrogated with treatment with anti-CD44 (26).
	CD63	CD63 Molecule	-	-	Has been observed to interact with TIMP1 (76)
	CTSB	Cathepsin B	-	-	Regulates NPC2 secretion, TNF-alpha production, and cholesterol trafficking genes in an animal model of obesity (39)
	CTSL	Cathepsin L	-	-	Regulates NPC2 secretion, TNF-alpha production, and cholesterol trafficking genes in an animal model of obesity (39)
	GLB1	Galactosidase, beta 1	-	Early SSc-PF	Mutations in this gene can lead to GM1-gangliosidosis, a manifestation of which includes foam cell accumulation in the lungs (77).
	NPC2	Niemann-Pick disease, type C2	Hub	Early SSc-PF	Mutations in this gene result in a lipid storage disorder. Functions in the regulation of cholesterol trafficking through the lysosome by binding to cholesterol released from low density lipoproteins taken up by cells.
	TGFBI	Transforming Growth Factor, Beta-Induced	Bridge	Late SSc-PF	Induced by phagocytosis of apoptotic debris in monocyte-derived MØs and regulates collagen turnover (32)
	TIMP1	TIMP Metallopeptidase Inhibitor 1	-	Early SSc-PF	Has been observed to interact with CD63 and overexpression has been noted to inhibit apoptosis in a CD63-dependent manner (76)
innate immune response/NFkB signaling/apoptotic process	*BIRC3*	Baculoviral IAP repeat-containing protein 3	-	Late SSc-PF	Has antiapoptotic activity through interactions with caspases as well as the TNF superfamily members TRAF1 and TRAF2 (78,79).
	CYR61	Cysteine-Rich, Angiogenic Inducer, 61		Late SSc-PF	Also known as CCN1. Implicated in apoptosis in fibroblasts (80). Has been shown to play a role in Fas-mediated and TRAIL-induced apoptosis (81,82).
	DUSP6	Dual Specificity Phosphatase 6	-	Late SSc-PF	Plays a role in the positive regulation of apoptosis (83)
	FAS	Fas Cell Surface Death Receptor	-	Early SSc-PF	Cell surface death receptor.
	NFKBIE	Nuclear Factor Of Kappa Light Polypeptide Gene Enhancer In B-Cells Inhibitor, Epsilon	-	-	Negative regulator of NFkB signaling
	PLAUR	Plasminogen Activator, Urokinase Receptor	Bridge	Late SSc-PF	Also known as uPAR. Contains an SSc risk SNP. Pulmonary fibroblasts from patients with idiopathic PF over express uPAR and that uPAR ligation results in a hypermotile phenotype (28).
	PLSCR 1	Phospholipid Scramblase 1	-	-	Regulates phospholipid membrane asymmetry.
	TNFAIP 3	Tumor Necrosis Factor, Alpha-Induced Protein 3	Hub		Also known as A20. Contains an SSc risk SNP (also associated with other autoimmune conditions). Negative regulator of NFkB signaling.
	TNFSF 10	Tumor Necrosis Factor (Ligand) Superfamily, Member 10	Bridge	-	Also known as TRAIL. Elevated in serum of SSc patients (84)
	TNFRS	Tumor Necrosis	-	Late SSc-PF	Also known as TRAILR2.

	F10B	Factor Receptor Superfamily, Member 10b			
IFN/antigen presentation	*HLA-E*	Major Histocompatibility Complex, Class I, E	-	-	Class I MHC molecule.
	HLA-F	Major Histocompatibility Complex, Class I, F	-	-	Class I MHC molecule.
	IFITM1	IFN Induced Transmembrane Protein 1	-	SSc-PF	IFN signaling.
	IFITM2	IFN Induced Transmembrane Protein 2	-	Early SSc-PF	IFN signaling.
	IFITM3	IFN Induced Transmembrane Protein 3	-	Early SSc-PF	IFN signaling.
	IRF1	IFN Regulatory Factor 1	-	Late SSc-PF	Activator of type I IFN signaling.
	OAS1	2'-5'-Oligoadenylate Synthetase 1, 40/46kDa	-	Early SSc-PF	Involved in innate immune response to viral infection.
response to TGF-beta	*CAV1*	Caveolin 1	-	-	Contains an SSc risk SNP.
	CTGF	Connective tissue growth factor	-	-	Also known as CCN2. Has been shown to play a role in Fas-mediated and TRAIL-induced apoptosis (81,82).
	DAB2	Dab, Mitogen-Responsive Phosphoprotein, Homolog 2 (Drosophila)	-	SSc-PF	Required for the epithelial to mesenchymal transition induced by TGF-beta in mouse and for type II TGFbR recycling (85,86)
	FN1	Fibronectin 1	-	-	Extracellular matrix protein.
	LAMC1	Laminin gamma1 chain	Hub	Early SSc-PF	Expression of this gene is essential for the development of basement membranes (87).
	THBS1	Thrombospondin 1	-	-	Mediates cell-to-cell and cell-to-matrix interactions. Putative biomarker of modified Rodnan skin score (88).

1025
1026
1027 **Table 5. Bonferroni-corrected p-values, Fisher's exact test pathophenotype-associated modules**
1028 **in top-level communities in the module overlap graph.**
1029

Top-level community	'In SSc' p-value	'In Inflammatory' p-value	'In Proliferative' p-value	'In PAH' p-value	'In PF' p-value
1	1	0.02	1	1	1
2	0.71	0.07	1	1	1
3	0.09	0.27	1	0.77	0.29
4	8.56E-07	6.30E-12	1	0.30	1
5	1	1	0.03	1	1
6	1	1	1	1	1
7	1	0.64	1	0.03	1
8	1	1	1	1	1

1030
1031
1032

Supporting Information Captions

S1 Fig. Network view of consensus clusters 4A and 4B in the information graph.

S2 Fig. Density plot of correlation to respective module eigengene (kME). Tissue consensus genes have significantly higher kME and are therefore more 'hub-like' than non-consensus genes. Mann Whitney U, $p < 2.2x10^{-16}$ reported by R

S3 Fig. Fully labeled version of the consensus lung network (Fig 3A). This file is intended to be viewed digitally.

S4 Fig. Fully labeled version of the differential adjacency matrix in Fig 4B. This file is intended to be viewed digitally.

S5 Fig. The differential lung network. The highly lung-specific network (minus global network and skin network) contains functional modules.

S1 Table. Table describing clinical characteristics of cohorts included in this study.

S2 Table. Consensus gene set sizes.

S3 Table. Immune-fibrotic axis consensus genes.

S4 Table. Mapping of Xue, et al. module numbers to our module names (Figure 5B).

S5 Table. P-values of all Xue, et al. modules tested.

S1 File. Tables - pathophenotype associations with WGCNA co-expression modules.

S2 File. Information graph adjacency matrix and module consensus cluster membership.

S3 File. Full output of edge-pathway (Jaccard) similarity Mann-Whitney U tests.

S4 File. Functional network edge lists and node attribute files (networks from Figures 3 and 4). The "common lung network" tab provides the module membership information for Fig 4B.

S1 Text. Glossary of terms.

S2 Text. Additional results about pathophenotype-associated consensus clusters in the information graph.

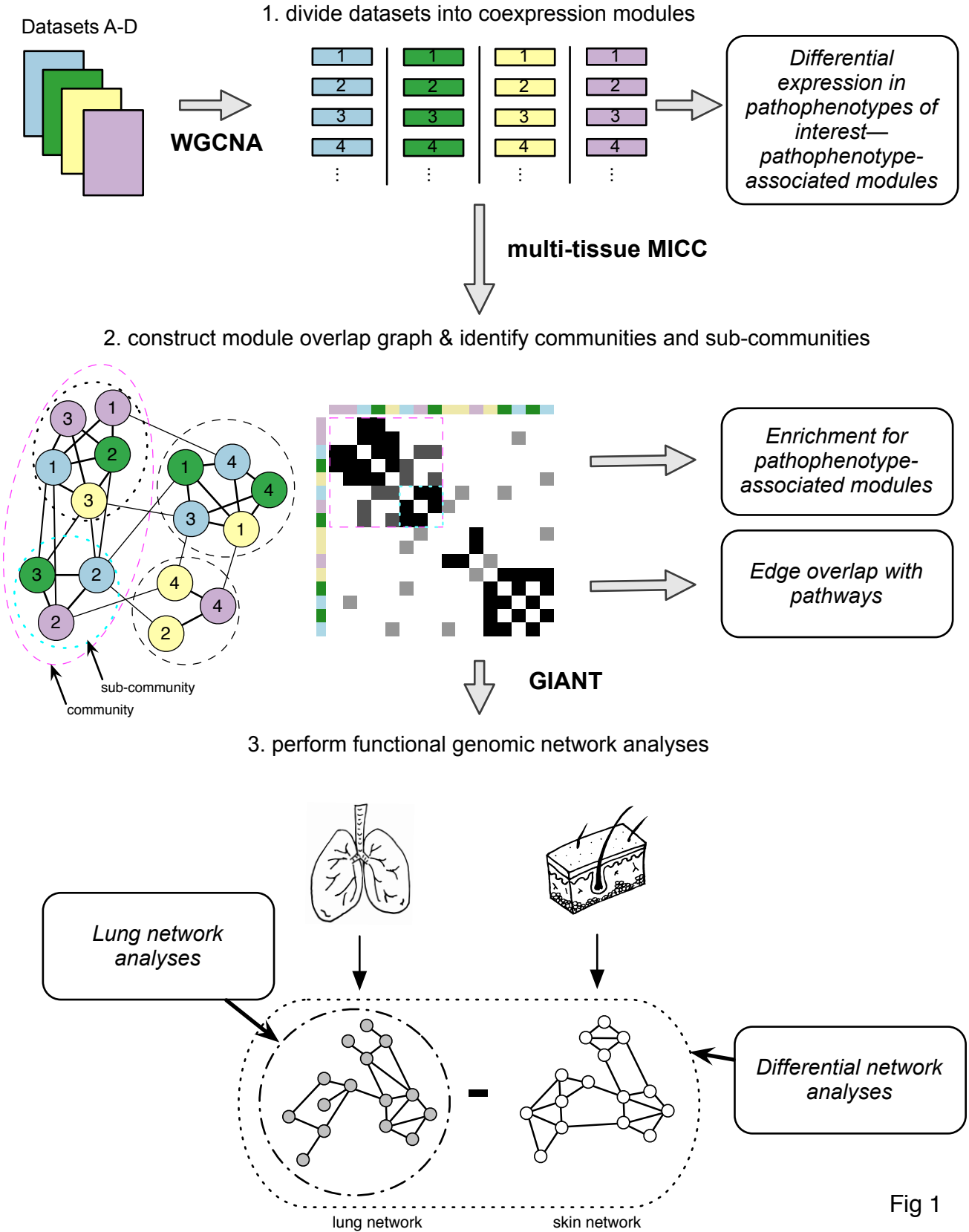
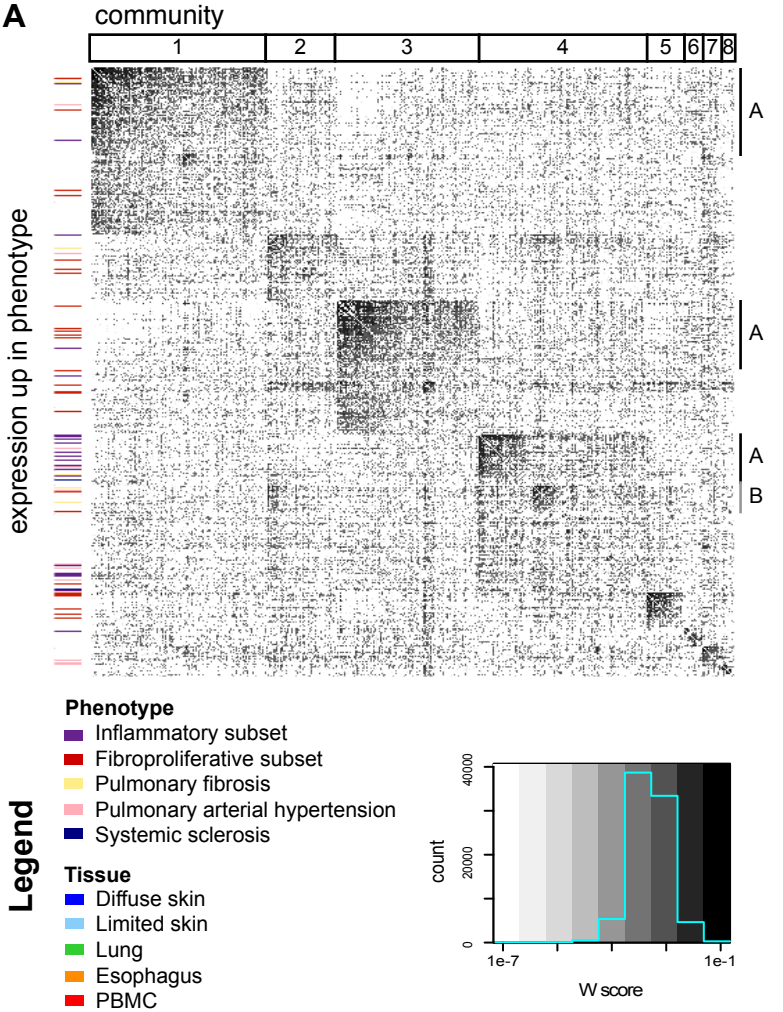

1. divide datasets into coexpression modules
Datasets A-D
WGCNA
Differential expression in pathophenotypes of interest—pathophenotype-associated modules
multi-tissue MICC
2. construct module overlap graph & identify communities and sub-communities
sub-community
community
Enrichment for pathophenotype-associated modules
Edge overlap with pathways
GIANT
3. perform functional genomic network analyses
Lung network analyses
Differential network analyses
lung network
skin network

Fig 1


A
community
1
2
3
4
5
6
7
8
A
A
A
B
expression up in phenotype
Legend
Phenotype
Inflammatory subset
Fibroproliferative subset
Pulmonary fibrosis
Pulmonary arterial hypertension
Systemic sclerosis
Tissue
Diffuse skin
Limited skin
Lung
Esophagus
PBMC
count
40000
20000
0
1e-7
1e-1
W score

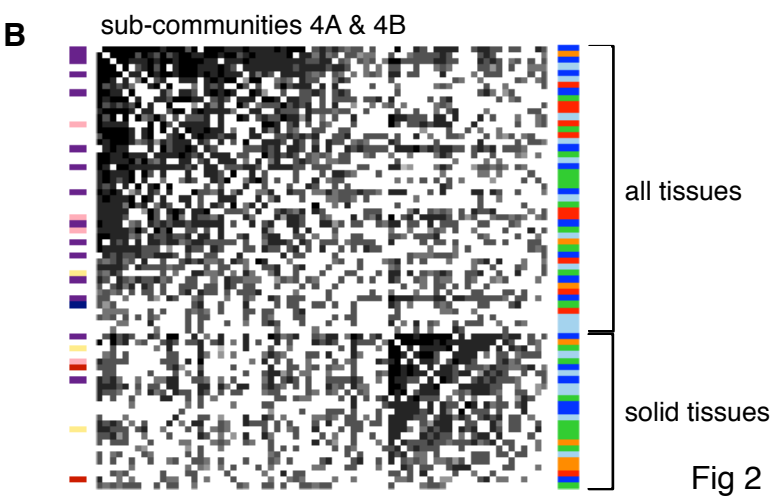

B
sub-communities 4A & 4B
all tissues
solid tissues

Fig 2

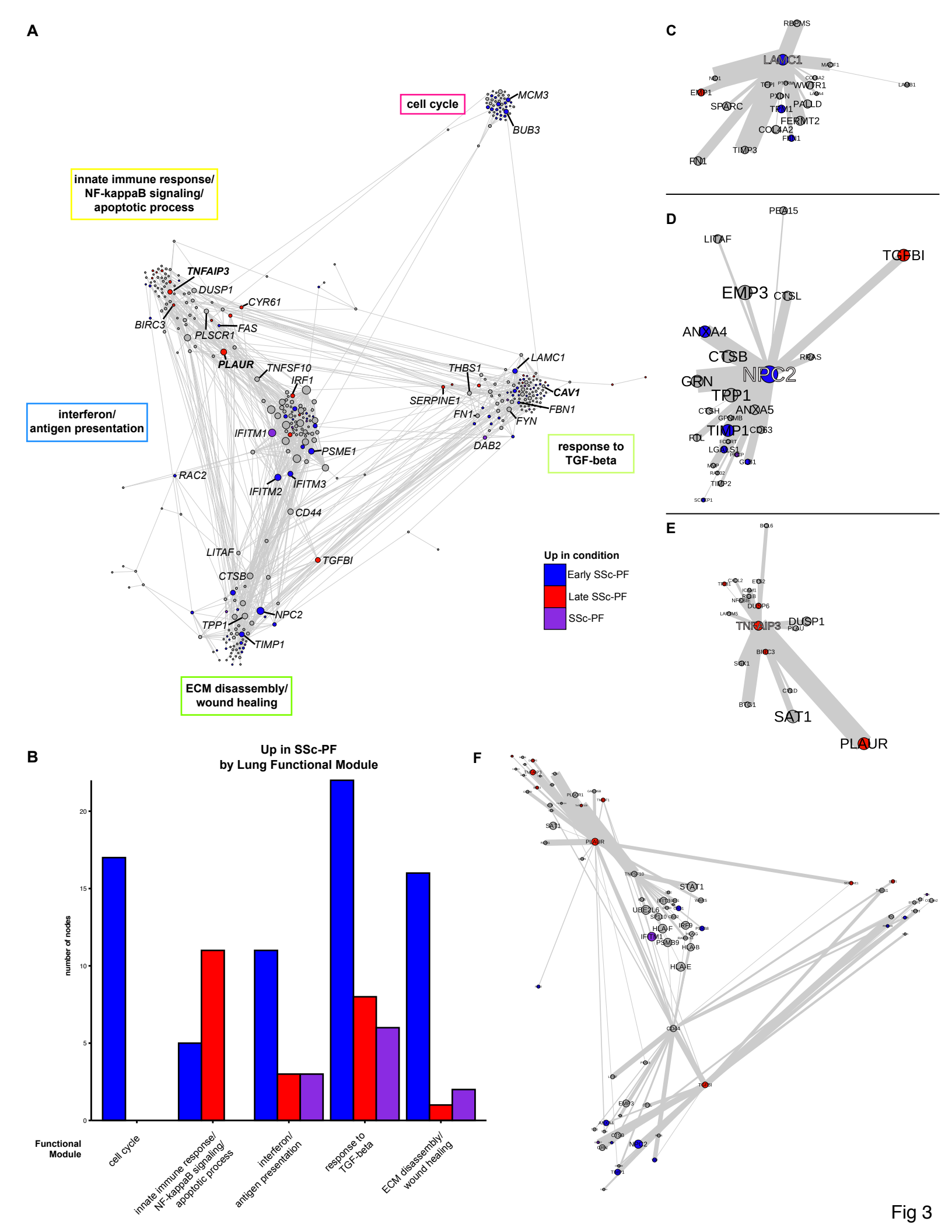

A
cell cycle
MCM3
BUB3
innate immune response/
NF-kappaB signaling/
apoptotic process
TNFAIP3
DUSP1
CYR61
BIRC3
FAS
PLSCR1
PLAUR
TNFSF10
IRF1
interferon/
antigen presentation
IFITM1
PSME1
RAC2
IFITM2
IFITM3
CD44
LITAF
TGFBI
CTSB
NPC2
TPP1
TIMP1
ECM disassembly/
wound healing
THBS1
LAMC1
CAV1
SERPINE1
FN1
FBN1
FYN
DAB2
response to
TGF-beta
Up in condition
Early SSc-PF
Late SSc-PF
SSc-PF
B
Up in SSc-PF
by Lung Functional Module
number of nodes
0
5
10
15
20
Functional
Module
cell cycle
innate immune response/
NF-kappaB signaling/
apoptotic process
interferon/
antigen presentation
response to
TGF-beta
ECM disassembly/
wound healing
C
RBPMS
LAMC1
EMP1
SPARC
WWTR1
PALLD
FERMT2
COL4A2
TIMP3
FN1
D
PEA15
LITAF
TGFBI
EMP3
CTSL
ANXA4
CTSB
RRAS
NPC2
GRN
TPP1
ANXA5
TIMP1
CD63
FTL
LGALS1
TIMP2
E
DUSP6
TNFAIP3
DUSP1
BIRC3
SAT1
PLAUR
F
SAT1
PLAUR
STAT1
UBE2L6
HLA-F
IFITM1
PSMB9
HLA-B
HLA-E
CD44
EMP3
NPC2

Fig 3

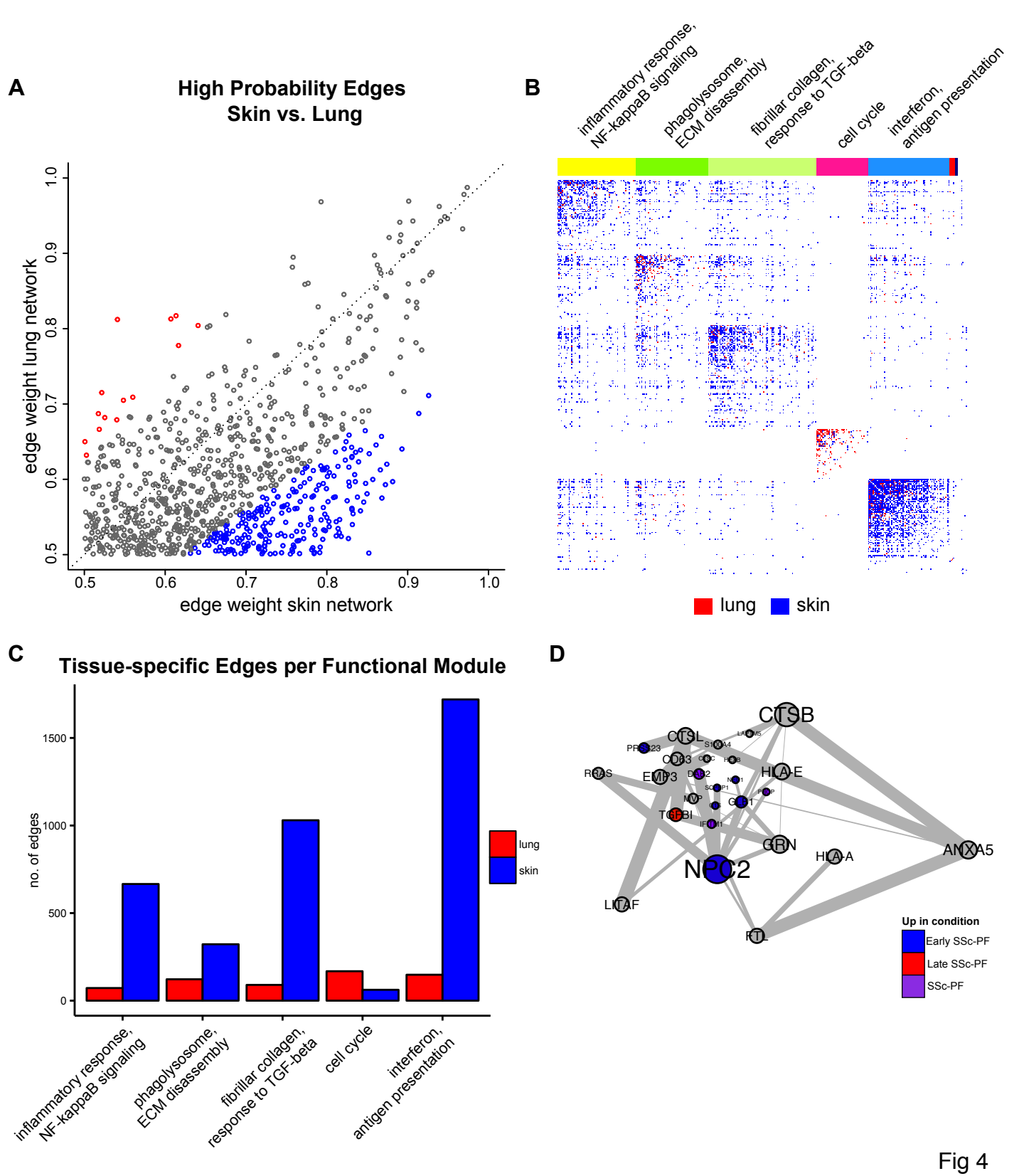

A
High Probability Edges
Skin vs. Lung
edge weight lung network
edge weight skin network
0.5
0.6
0.7
0.8
0.9
1.0
B
inflammatory response, NF-kappaB signaling
phagolysosome, ECM disassembly
fibrillar collagen, response to TGF-beta
cell cycle
interferon, antigen presentation
lung
skin
C
Tissue-specific Edges per Functional Module
no. of edges
0
500
1000
1500
lung
skin
inflammatory response, NF-kappaB signaling
phagolysosome, ECM disassembly
fibrillar collagen, response to TGF-beta
cell cycle
interferon, antigen presentation
D
CTSB
CTSL
CD63
EMP3
RRAS
HLA-E
TGFBI
GRN
HLA-A
ANXA5
NPC2
LITAF
FTL
Up in condition
Early SSc-PF
Late SSc-PF
SSc-PF

Fig 4

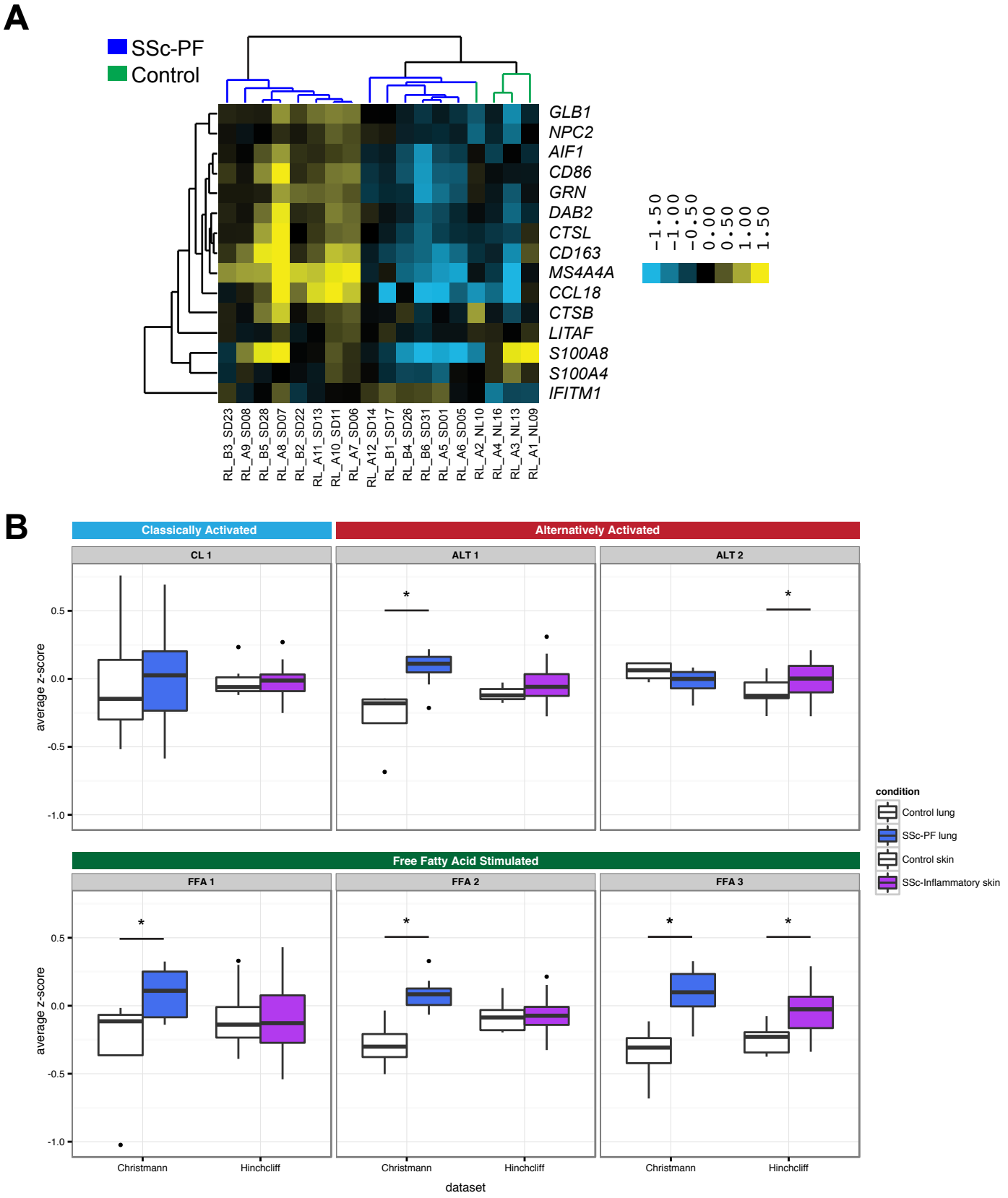

A
SSc-PF
Control
GLB1
NPC2
AIF1
CD86
GRN
DAB2
CTSL
CD163
MS4A4A
CCL18
CTSB
LITAF
S100A8
S100A4
IFITM1
RL_B3_SD23
RL_A9_SD08
RL_B5_SD28
RL_A8_SD07
RL_B2_SD22
RL_A11_SD13
RL_A10_SD11
RL_A7_SD06
RL_A12_SD14
RL_B1_SD17
RL_B4_SD26
RL_B6_SD31
RL_A5_SD01
RL_A6_SD05
RL_A2_NL10
RL_A4_NL16
RL_A3_NL13
RL_A1_NL09
-1.50
-1.00
-0.50
0.00
0.50
1.00
1.50
B
Classically Activated
Alternatively Activated
CL 1
ALT 1
ALT 2
Free Fatty Acid Stimulated
FFA 1
FFA 2
FFA 3
average z-score
0.5
0.0
-0.5
-1.0
Christmann
Hinchcliff
dataset
condition
Control lung
SSc-PF lung
Control skin
SSc-Inflammatory skin

Fig 5

A

NF-kappaB signaling, apoptotic processes

resistance to apoptosis

coagulation / inhibition of fibrin degradation & collagen deposition

TRAIL

efferoctyosis

injury?

IFN, antigen presentation

TGF-beta

response to TGF-beta

lipids

TGFbR recycling

ECM disassembly, wound healing

lung macrophage response

B

initial injury

interferon response

lipids

epithelial cells

lung macrophage

TGF-beta

ECM deposition

resistance to apoptosis

myofibroblasts

Fig 6

Early/NSIP | Common | Late/UIP