Award Number: W81XWH-11-1-0525

TITLE: Identification of the Gene for Scleroderma in the Tsk/2 Mouse Strain: Implications for Human Scleroderma Pathogenesis and Subset Distinctions

PRINCIPAL INVESTIGATOR: Michael Whitfield Ph.D

CONTRACTING ORGANIZATION: Dartmouth College
Hanover, NH 03755-1404

REPORT DATE: September 2015

TYPE OF REPORT: Final

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

1. REPORT DATE September 2015	2. REPORT TYPE Final	3. DATES COVERED 1Jul2011 - 30Jun2014

4. TITLE AND SUBTITLE

Identification of the Gene for Scleroderma in the Tsk/2 Mouse Strain: Implications for Human Scleroderma Pathogenesis and Subset Distinctions

5a. CONTRACT NUMBER
W81XWH-11-1-0525

5b. GRANT NUMBER
10668888

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S) **Michael L. Whitfield, PhD**

Elizabeth P. Blankenhorn, PhD
Carol Artlett, PhD
E-Mail: eblanken@drexelmed.edu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Dartmouth College
Hanover, NH 03755-1404

8. PERFORMING ORGANIZATION REPORT

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT
This project is focused on an animal model of the human disease, systemic sclerosis (SSc), called Tsk2/+. The SSc-like traits in Tsk2/+ heterozygotes are highly penetrant. With their readily apparent skin fibrosis resulting from ECM anomalies, Tsk2/+ mice have signs that resemble human SSc features, making it useful as a pre-clinical model. In this report, we show a clear time dependence on the gene expression in the skin of the Tsk2/+ mice. We have pinpointed a mutation in *Col3a1* that is the Tsk2 gene, and have confirmed the sequence difference between Tsk2/+ and the parent strain, 101/H. We present results on the expression of TGFβ mRNA from cells cultured from Tsk2/+ and WT littermates that suggest a mechanism for the up-regulation of TGFβ seen in the mutant strain. We show that elastin content in the skin, known to be controlled by TGFβ and possibly up-regulated in SSc, is the earliest indicator of tight-skin in the tissue. Finally, we show that Tsk2/+ mice, and mouse fibroblasts transfected with Col3a1 from Tsk2/+, share a substantial fibrotic gene expression program compared to WT mice or transfectants, indicating that expression of the *Col3a1*^{Tsk2} gene alone accounts for the trait in Tsk2/+ mice.

15. SUBJECT TERMS **Animal model, systemic sclerosis, scleroderma, Tsk2/+, fibrosis, gene, genetics, TGFβ**

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. # OF PAGES	19a. NAME OF RESPONSIBLE PERSON
					USAMRMC
a. REPORT	b. ABSTRACT	c. THIS PAGE			19b. TELEPHONE NUMBER *(include area code)*
U	U	U	UU	72	

Table of Contents

1. INTRODUCTION

Scleroderma and systemic sclerosis (SSc) is a heterogenous disease of fibrosis and inflammation, concomitant with significant autoimmunity. SSc often presents with skin manifestations and Raynaud's phenomenon; the extent and location of fibrotic lesions in people with SSc contributes to the diagnoses of disease subtypes and prognosis. Several preclinical animal models for SSc exist. Tsk2/+ mice were discovered more than two decades ago when progeny of a 101/H male in an ENU mutagenesis experiment were noted with very tight skin. *Tsk2* is homozygous lethal, similar to the *Tsk1* mouse model of SSc which results from duplication of the fibrillin gene. Tsk1 has been one of the most commonly used models for SSc and therefore has been extensively characterized. The SSc-like traits in Tsk2/+ heterozygotes are highly penetrant. In addition to a readily apparent skin fibrosis resulting from ECM anomalies, Tsk2/+ mice show more autoimmune and inflammatory features than Tsk1/+, and their longer lifespan and immune features that closely resemble human SSc features are ideal for use as a pre-clinical model. The Tsk2 mutation has been bred onto a homogeneous inbred (C57Bl/6, or B6) background in Dr. Blankenhorn's laboratory. B6.Tsk2/+ mice have many features of the human disease, including tight skin, dysregulated extracellular matrix deposition, and significant autoimmunity. We have found that Tsk2-mediated autoimmune and fibrotic signs develop *progressively* with age and manifest differently in females than males, a phenomenon also observed in human SSc. These SSc phenotypes in B6.Tsk2 mice are all likely due to a single genetic mutation, which we have now unidentified. We proposed to identify the *Tsk2* gene and understand its mechanism of action as outlined in our statement of work. This mouse affords a unique opportunity to examine the pathways leading to the multiple clinical parameters of fibrotic disease from birth onward, and we describe our studies from the second year of the grant below. This work was accomplished by researchers at Drexel University College of Medicine and Dartmouth Geisel School of Medicine under the partnering PI option.

2. KEYWORDS:

Scleroderma and systemic sclerosis (SSc); animal models; mice; Tsk2/+ mice; genetic mapping; fibroproliferative subset; TGFβ; TWEAK/FN14 signaling; genetic mutation; *Col3a1;* COL3A1; genomic DNA sequencing; RNA-Seq; microarray; *in vitro culture;* transfection; plasmid; skin and lung fibrosis.

3. ACCOMPLISHMENTS

Milestones were assigned to this proposal, with tasks to be accomplished by each investigator. The overall **summary** of our progress relative to these tasks is given below, followed by a complete discussion of our work the past four years.

Milestone 1 Identify Tsk2/+ gene - **This is now DONE.**

Task 1 was for the Blankenhorn laboratory to collect DNA for sequencing (Months 1-6), which we have done. In year 1, we collected the Tsk2/+ and 101/H (parental strain) DNAs for sequencing on the 454.

Task 2 (Months 6-12) was to select anchor sequences for Nimblegen chip design, so that chromosome 1 DNA in the Tsk2/+ interval could be sequenced. This was done in year 1 by our subcontractors at ASRI, Dr. Fen Hu and Dr. Garth Ehrlich.

Task 3 (Month 6-12): Dr. Hu and her colleagues have hybridized the mouse genomic DNA to the chips and collected Tsk2/+ interval DNA, meeting this target. They have sequenced both Tsk2/+ and 101/H interval DNA.

Task 4: We assembled all the sequence data in years 1 and 2, and have aligned the sequences to compare and report all observed polymorphisms. At Drexel, we completed amplification and re-sequencing of the target gene this year (year 2, on schedule), Blankenhorn lab.

Milestone 2 Determination of mechanism of action of Tsk2/+ gene: two hypotheses on this are discussed.

Task 1 (Months 18-32): At Drexel, we were to breed Tsk2/+ mice to a knockout mouse with a deficiency in the newly-identified *Tsk2* gene, to determine if either Tsk2 or wild-type allele can complement the genetic deficiency. Last year, we reported that *Col3A1* is the gene underlying *Tsk2,* and so the Blankenhorn laboratory purchased and bred three Col3A1 KO/WT heterozygous male mice to Tsk2/+ dams in July 2012 (month 12). **This is now DONE.**

Task 2 (Months 1-36). Correlate the known actions of the Tsk2 gene at Drexel with gene expression data at Dartmouth (Aim 2) and with the presence of proliferating cells (Aim 3). In this Task, largely accomplished at Dartmouth with the microarray studies (months 4-12), we are establishing the timeline for the gene signatures in male and female Tsk2/+ mice. We will then examine the corresponding Tsk2/KO mice for these phenotypes to detect alterations in the TGFβ1-driven proliferative, fibrotic signature of the Tsk2/+ gene when it is absent. This year, we have added a task to the project: we have found that mouse skin samples for gene expression studies need to be stratified for hair cycle. This has delayed some of our gene signature analyses until year 2 and 3, as the earlier samples sent to Dartmouth were not all in the same stage of the hair cycle, which dramatically alters the landscape of the skin. **This is now DONE.**

Also part of Milestone 2 was the aim to make fibroblast cultures from the mice. Originally planned for the appropriate conditional TGFβR animals so that alterations to TGFβ1 signaling taking place early in mouse post-natal development can be monitored, we have found this to be a good approach for the analyses of all the genotypes. **This is now DONE.**

Another component of Milestone 2: Based on Dr. Whitfield's preliminary results with whole genome profiling, and on Dr. Blankenhorn's with selected TGFβ1-dependent target gene expression, we expect that TGFβ1 is a necessary component in the disease pathway. Therefore, we are ordering the TGFβR conditional KO mice and the transgenic mice bearing *cre* recombinase under the control of a collagen promoter. We should have them in hand this fall. When bred to Tsk2/+ mice, these constructs will help us to fully understand the timeline of the TGFβ signature. At Drexel, we will breed Tsk2/+ mice to TGFβR conditionally deficient mice (by breeding the Tsk2/+ mutation onto a floxed TGFβR2 KO and then breed the resulting mice to B6.Col-Cre animals) to determine the interaction between TGFβ1 and *Tsk2*. Approximately equal numbers of wild-type (WT) and Tsk2/+ mutant progeny with the dominant TGFβR conditionally deficient trait will be born and used (Months 24-36). We will then correlate the known actions of the *Tsk2* gene AT DREXEL with gene expression data AT DARTMOUTH in Aim 2 and with the presence of proliferating cells in Aim 3, by examining the corresponding Tsk2/KO mice for these phenotypes to detect alterations in the TGFβ1-driven proliferative, fibrotic signature of the Tsk2/+ gene when it is absent. Months 18-36, Blankenhorn laboratory; months 12-18 in the Whitfield laboratory. **We have chosen a different KO to study, based on preliminary results from Year 3.**

Milestone 3 Determine the timing of TGFβ activation in the Tsk2/+ mice, and differences between males and females. Dr. Blankenhorn will send mouse tissues to Dr. Whitfield, who will do the RNA work. **This is now DONE.**

Task 1 (Months 1-36): The Blankenhorn laboratory will breed sufficient numbers of mice to collect skin at postnatal Day 0, day 7, day 14, day 21 as well as 1 month and 4 months. These mice are used by all three investigators, and whenever possible, each individual mouse was studied for the relevant traits in each laboratory, so that histology and RNA transcript analysis will occur on the same animal. We have met our targets in year 1. In Year 2, this task was accomplished, and in year 3, we will add samples where stratification by hair cycle has been done. **This is now DONE.**

Task 2 (Months 4-12) Prepare RNA from skin at Drexel and hybridize DNA microarrays at Dartmouth. Data will be analyzed, processed and stored. In practice, we found it better to send whole skin samples to Dartmouth and prepare the RNA there. **This is now DONE.**

Task 3 (Months 12-36): At Dartmouth perform data analysis for expression of TGFβ as well as other gene signatures, both profibrotic (IL13 and IL4) and those that may not be expected (genome-wide). **This is now DONE.**

Task 4 (12-24 months, if necessary): If the microarray study is unclear, we had proposed a small number of RNASeq runs to validate the gene expression data. This was unnecessary for these goals.

Task 5 (dependent timing): Immunohistochemistry will be performed for the validation of TGFβ signatures found in the microarrays. This will be performed by the Artlett and Blankenhorn laboratories. **This is now done.**

Milestone 4 Characterize how well the Tsk2/+ mouse approximates human SSc at different time points. These tasks are on time, and to be conducted largely in year 3.

Task 1: At Dartmouth, map mouse genes to human orthologs, integrate mouse and human data using Distance Weighted Discrimination to remove systematic biases, and cluster mouse and human data (months 12-36). **This is now DONE.**

Task 2: At Dartmouth, Analyze data-driven groupings, pathways, computational validation and data interpretation (months 12-36). Data analysis for expression of proliferative signatures will give us a way to understand the subset of SSc patients that exhibit diffuse clinical symptoms with signs of cell proliferation. This is a special investigation of proliferative signatures by the Whitfield group to capitalize on their extensive experience with cell cycle and proliferative motifs in gene expression. It was scheduled for months 8-24, and **is now DONE**.

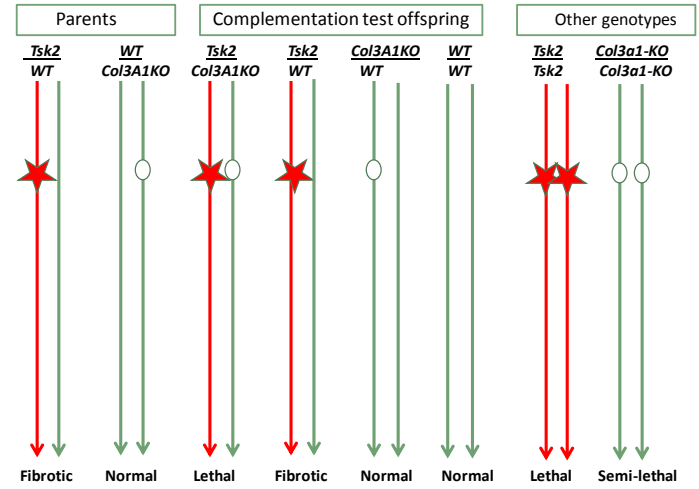


Figure 1. Complementation testing proves Tsk2 is C33S: Red lines indicate the mutated chromosome from Tsk2/+ mice and green lines indicate a normal chromosome 1. Ovals:Col3a1 knocked out in BALB/c mice; red star shows the Tsk2 mutation. Outcomes associated with each genotype are indicated below.

Milestone 5

Task 1: We will perform confirmation qRT-PCR on select genes based on Aim 2 in the Blankenhorn and Whitfield laboratories. We had scheduled this for months 4-24; this work is ongoing due to the confounding hair cycle. We modified this task to extend to year three as well, to ensure full study of interesting gene expression patterns over mouse developmental ages. In the Drexel laboratories, we plan experimentation on the mechanotension of the ECM when it contains Tsk2/+ collagen in comparison to ECM containing WT collagen, after the identification of the Tsk2/+ candidate gene by Aim 1. We have started this; one issue is how slowly the mouse fibroblasts grow. A final goal within this milestone is to characterize the disorder to understand how the *Tsk2* mutation acts to elicit it. **These goals are included in our hypothesis about the mechanism of action of *Col3a1*^{Tsk2} (see below)**

Milestone 6 Cross-breed Tsk2/+ mice to Wsh mast cell knockout mice (at Drexel). This Milestone was ***deleted*** and supplanted by other work, as described in the Year 1 Progress report.

PRELIMINARY RESULTS BY MILESTONE

Milestone 1 : The nucleotide sequence of the Tsk2/+ region was accomplished in year 1 and 2, and the initial report of the sequencing capture and early resequencing at Drexel was made in the year 1 progress report. This work is in our submitted manuscript,

We confirmed the global sequencing result at Drexel and also evaluated the remaining SNPS by phototyping[1, 2]. Of these, only the Col3A1 non-synonymous coding SNP was validated; two intronic SNPs in the GULP1 gene also distinguish Tsk2/+ from all other strains for which chr 1 genotyping is available. The Col3a1 SNP results in a Cys to Ser change in the PIIINP (N-terminal) cleavage product of the Col3a1.

Milestone 2: Breeding to Col3A1 KO:

We began the breeding necessary for the genetic complementation test of *Col3a1* by breeding the Tsk2/+ line to BALB.Col3A1KO mice (heterozygotes as well). Results were collected in year 2. **This study provided the definitive evidence that *Col3a1* is the gene mutated in Tsk2/+ mice**. We hypothesized that combining the Tsk2 mutation with a Col3a1- allele would not generate a viable genotype, and the Tsk2/*Col3a1*KO mice would die in utero. There are four possible genotypes in the cross: (1)Tsk2/+; (2)Tsk2/*Col3A1*KO; (3)Col3A1KO/*Col3A1*KO; and (4)*Col3A1*KO/+. Our analysis of the litters confirmed this prediction: we found no Tsk2/Col3a1- mice among the neonates (i.e., this genotype is nonviable), out of 33 offspring born in this cross (**Figure 1; Table 1**). This implies that the C33S mutation in *Col3a1* is the causative mutation, because the Tsk2-bearing chromosome has no functional *Col3a1* gene. While the compound heterozygotes couldn't survive without a functional allele of *Col3a1* (and C33S is the only mutation within *Col3a1* in the Tsk2/+ mice), there were a few Col3A1KO/Col3A1KO homozygotes born. These mice did not thrive and were sacrificed to provide neonatal skin fibroblast cultures.

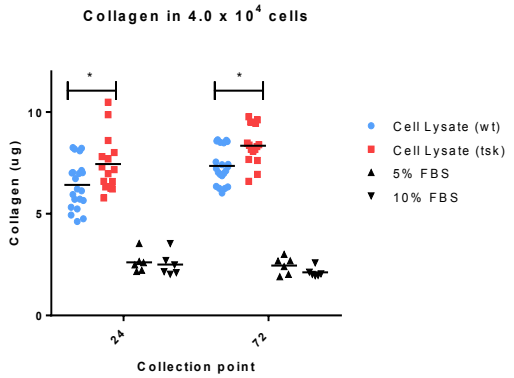
Parents	Tsk2/+ Tight skin	+/+ Normal skin	Tsk2/Tsk2 (lethal)
Tsk2/+ x Tsk2/+	22	21	0
	$Col3a1^{+}$/ $Col3a1^{-}$ Normal skin	$Col3a1^{+}$/ $Col3a1^{+}$ Normal skin	$Col3a1^{-}$/ $Col3a1^{-}$ (Moribund)
$Col3a1^{-/+}$ x $Col3a1^{-/+}$	16	13	3

	WT/$Col3a1^{+}$ Normal skin	Tsk2/$Col3a1^{+}$ Tight skin	WT/$Col3a1^{-}$ Normal skin	Tsk2/$Col3a1^{-}$
Tsk2/+ x $Col3a1^{-/+}$	12	10	11	0

Table 1: Offspring born of Tsk2 and Col3a1-KO mice. The breeding scheme for maintaining the two mutant lines is shown in the first two rows, with the phenotypes in the pups heading each column. The complementation test of (Tsk2/+ x Col3a1-/+) is shown in the bottom panel.

Milestone 2 also had the aim to make fibroblast cultures from the mice.

Fibroblast cultures were established in Blankenhorn and Artlett lab in months 8-24. Western blots of collagen expression were compared to hydroxyproline and Sirius red measurements. A very small but consistent and significant average difference is seen between WT and Tsk2/+ fibroblasts collected at day 1 (**Figure 2**), but by day 28 and older, this difference is difficult to repeat. Ascorbic acid supplementation did not reveal any further differences. One note – we have not observed proliferation differences in Tsk2/= vs. WT fibroblasts in vitro.



Source of Variation	P value summary	Significant?
Interaction	ns	No
Genotype	***	Yes
Hrs	***	Yes

Figure 2. Collagen synthesis in vitro, stratified by Tsk2/+ genotype.

In year 3, we developed a new method using the Col3a1-KO fibroblasts, to prove that Tsk2 is Col3a1(C->S). Although the complementation test proved definitively that lethality is due to *Col3a1*^{Tsk2} in the compound heterozygote, this experiment did not prove that fibrosis was also due to the *Col3a1*^{Tsk2} allele. Because the compound heterozygous animals do not survive to accumulate fibrotic levels of ECM, a direct *in vivo* test is impossible, so we performed an '*in vitro* complementation' test, wherein we transfected mutant or wild-type *Col3a1* cDNA into *Col3a1*-KO fibroblasts, harvested from a *Col3a1*-KO/KO homozygote at birth. Using the production of COL1A1 as a measure of fibrosis (shown to be expressed at high levels in Tsk2/+ skin and used as a marker of fibrosis [3, 4]), we assessed both protein and mRNA levels in fibroblasts that received DNA from a plasmid containing a single allele of a single *Col3a1* gene. In three independent experiments, COL1A1 protein was significantly elevated after 48 hours of transfection with *Col3a1*^{Tsk2} relative to transfection with *Col3a1*^{WT} (**Fig. 3**); mRNA for *Col1a1* was likewise increased in cells transfected with mutant *Col3a1*^{Tsk2} cDNA. Transfection efficiencies were equal in each of the experiments (**Fig. 3**).

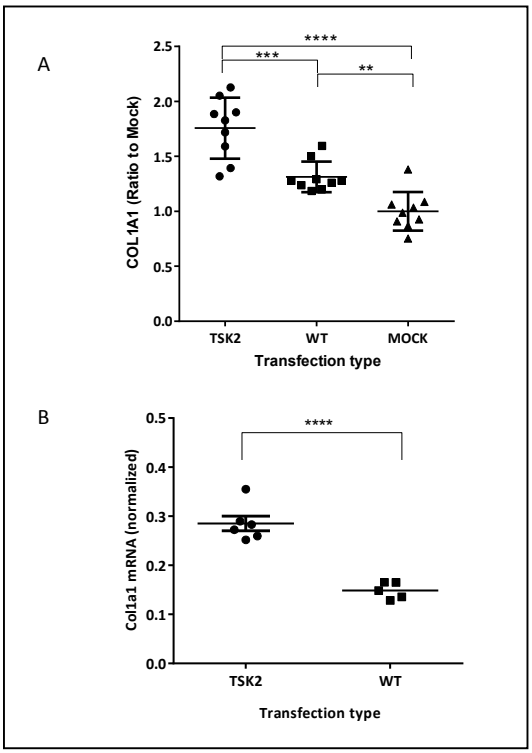


Figure 3. Mouse Col3a1-KO fibroblasts transfected with a plasmid bearing the mutant Col3a1Tsk2 express 34% more COL1A1 protein than Col3a1WT transfectants. Near-confluent dishes of COL3A1-deficient fibroblasts were transfected (three dishes each) with either of the two plasmids for 3 h, glycerol shocked, then washed and incubated for 48 hours. Three independent experiments were performed.

(A) Culture supernatants were collected and assayed by Western blot for COL1A1 secretion. Comparisons by one-way ANOVA show that cells that received Col3a1Tsk2 produced significantly more COL1A1 than cells transfected with Col3a1WT ($p<0.001$) or mock transfectants ($p<0.0001$); cells receiving wild-type Col3a1 also produced more COL1A1 than the mock ($p<0.01$). The mean ratio of COL1A1 in Col3a1Tsk2 -transfected cells to that in the mock transfections is 1.757; mean ratio of COL1A1 in Col3a1WT -transfected cells to mock is 1.313.

(B) Col1a1 mRNA (normalized to housekeeping gene expression) is also more highly expressed in COL3A1-deficient fibroblasts after transfection with Col3a1TSK than with Col3a1WT ($p<0.0001$). Cell lysates from two of the experiments were collected to determine that plasmids were transfected with equal efficiency, not shown).

Milestone 3 and 5: RNA gene expression profiling.

We reported (in our Year 1 Progress Report) the initiation of the planned RNA gene expression profiling by DNA microarray at Dartmouth. To accomplish this milestone we analyzed skin from both wt and Tsk2/+ mice at 4, 8, 12, and 20 weeks of age for both males and females.

We analyzed 4 independent skin samples each for WT and Tsk2/+ mice at each time point for both male (24 microarrays) and female mice (24 microarrays). Each of these was analyzed separately. We found a clear time dependence of the gene expression in Tsk2/+ that also varied by gender. Analysis of the female mice at 4, 8, 12 and 20 weeks of age identified specific gene expression signatures at each time point and it was very clear that some time points had very significant changes in gene expression (4 and 12 weeks), whereas other time points (8 and 20 weeks) show many fewer significant genes with higher False Discovery Rates (FDRs). A similar finding was observed for males except the largest changes were observed at 8 weeks and the fewest changes observed at 4 and 12 weeks, suggesting that the gene expression changes in male Tsk2/+ mice are changing at a different rate, or at different developmental time points, than occurs in female mice (**Figure 4)**.

Based on data from Dr. Blankenhorn's laboratory, the hypothesis was developed that hair cycle varied between Tsk2/+ and WT as well as between males and females. In addition, this process appears somewhat

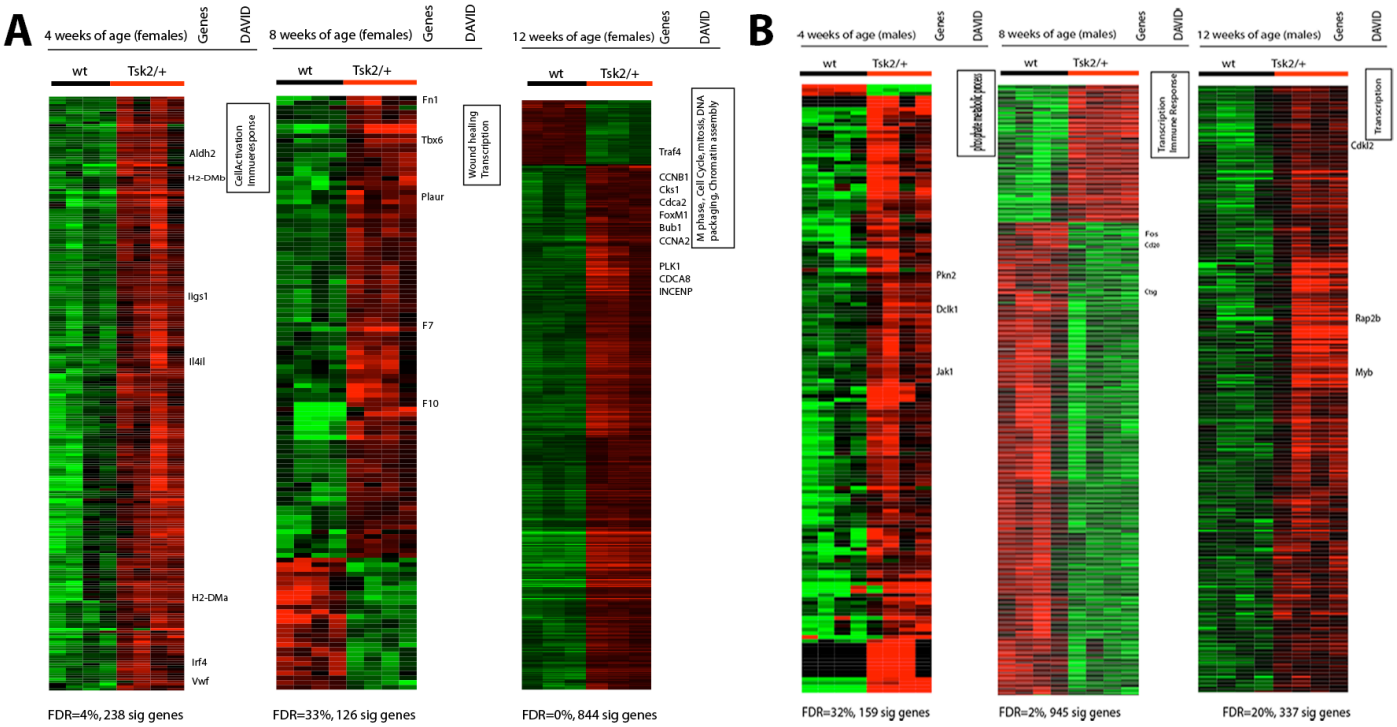


Figure 4. Genes differentially expressed between Tsk2/+ and WT differ by time point and gender. Four-week and 12-week female Tsk2/+ mice (panel A), as well as 8-week male Tsk2/+ mice (panel B) show significantly more differentially genes at a low false discovery rate (FDR). We used Significance Analysis of Microarrays (SAM) to compare WT and Tsk2/+ mice at different time points for each gender. A. 238 significant genes at 4% FDR were identified in 4-week old females, whereas, only 126 significant genes were obtained at 8-week time point (FDR=33%). However, 12-week old females showed the most significant differential gene expression with 844 significant genes (0.1% FDR). B. For males the most significant difference was observed in 8 week old animals. The 8-week old male Tsk2/+ mice showed 945 significant genes (FDR=2). 4-week and 12-week male Tsk2/+ mice showed significant differential gene expression but at very high FDR. Very similar results were seen with confirming RT-PCR of selected transcripts (data not shown) in male and female mice of different ages.

---

stochastic, occurring in different animals of the same genotype and gender at different times. We therefore have recently revisited the above analysis controlling for hair cycle stage for each sample and only using those samples that are at the same stae. We repeated the gene expression analysis and show that by eliminating those samples from mice in anagen, we increased the number of genes we are able to select, and correspondingly reduced our overall FDR due to decrease sample variation (**Figure 5**). The number of genes that can be selected at a given FDR is shown for each gender and age group of mice in **Figure 5.**

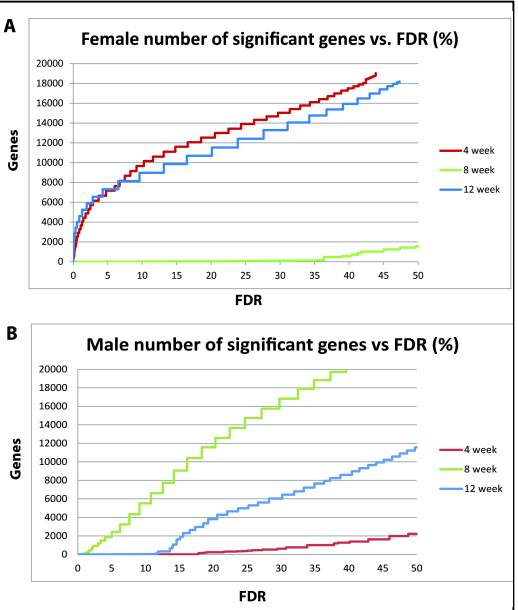


Figure 5. Number of differentially expressed genes by FDR. A. Females at each age were analyzed for genes differentially expressed between WT and Tsk2/+. The most significant changes are observed at 4 and 8 weeks. B. Males show a different distribution with the largest changes occurring at 8 and 12 weeks, and relatively few changes occurring at 4 weeks.

A TGFβ-responsive gene signature is activated in Tsk2/+

Using samples from 4-week old female mice that show the most significant change in gene expression when we compare WT and tsk2/+ mice (**Figure 6**), we asked if there was change in the expression of TGFβ-regulated genes in these mice. Our analysis of

the Gene Ontology (GO) annotations shows that the majority of genes up-regulated in Tsk2/+ mouse skin at 4 weeks of age map to the GO Biological processes of Cell adhesion and Cell morphogenesis (DAVID, Benjamini-corrected p < 0.05). Genes that show increased expression include *Col6a1, Col6a2, Col5a1, Sparc* and *Thy1*. Many of these genes are known targets of the profibrotic cytokine TGFβ.

In order to formally identify the TGFβ responsive genes in an unbiased fashion, we used a unique 674-gene TGFbeta signature that we previously developed from human dermal fibroblasts [5]. Those human genes were then mapped to 526 mouse gene orthologs using the Mouse Genome Informatics (MGI) Mouse/Human Orthology Phenotype Annotations table (The Jackson Laboratory, Bar Harbor, Maine). Out of those 526 orthologs, 459 genes were annotated on the Agilent 4X44K Mouse Whole Genome Microarray. We then removed those genes that were likely to be affected by the hair growth cycle, or background noise, and in the end used 134 genes as our final, mouse TGFβ-responsive gene signature. These 134 genes represent the core TGFβ signature and includes the well-characterized TGFβ target genes. Activation of the TGFb responsive genes in Tsk2/+ has also been analyzed in a manuscript that is currently submitted for publication (Sargent et al. *Submitted*)

Enrichment of the TGFβ responsive gene signature was analyzed in skin samples from 4-week old female mice that show the largest changes in gene expression (see above) and in samples that have been controlled for anagen (e.g. any sample that was in anagen was removed and only samples not in active hair cycle growth were considered). We then calculated enrichment of the TGFβ gene signature in each mouse sample by calculating the Pearson correlation coefficients between the centroid and the gene expression for each 4-week female mouse skin biopsy. These data are shown in Figure 6 and clearly show enrichment for TGFβ-responsive gene expression in Tsk2/+ female mice at 4 weeks of age. Overall, our results clearly show enrichment of the TGFβ-responsive gene signature in Tsk2/+ mice but not in WT mice, and this is dependent on age and hair cycle. We have collected skin samples to study even younger mice (2 weeks of age), in which we have seen a signature difference in males in preliminary analyses (Fig 6, next page)

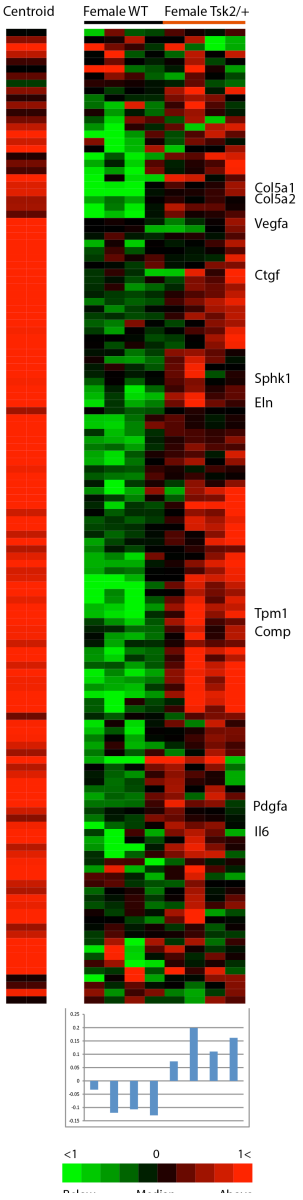


Figure 6. A TGFβ- responsive gene expression signature is activated in 4-week old female Tsk2/+ mice not in WT. A TGFβ responsive gene signature centroid was generated from TGFβ-treated human primary fibroblast [5]. The centroid representing the average of the TGFβ responsive genes is shown to the left of the heat map. Pearson correlations between the centroid and each array were calculated and are plotted directly beneath each microarray analyzing WT or Tsk2/+ mice from 4-week old female mice. Several known Tgfβ target genes are highlighted.

Milestone 4. Characterize how well the Tsk2/+ mouse approximates human SSc at different time points.

These tasks were completed in the no-cost extension year and are now submitted for publication. We have used a genome-wide interspecies comparative analysis of mouse models and SSc patients, to determine which animal models of SSc best reflect the disease on a molecular level, providing models for study and pre-clinical testing[6]. Using gene expression as a readily quantifiable phenotype in skin, we have shown that three SSc models (Tsk2/+, the sclGVHD model, and bleomycin-induced fibrosis) are similar to their respective human subsets, not simply by gross morphology approximations but by robust molecular measures [6]. This is a significant contribution to the SSc field as it provides new, urgently needed insight into the genome-wide molecular changes that are common to mouse models **and**

human SSc patients. It pinpoints which animal model should be used for developing specific therapeutics targeted at each of the different molecular intrinsic subsets of SSc, which are the **inflammatory, fibroproliferative, normal-like and limited subsets**[7-10]. We found that skin of young Tsk2/+ mice share gene expression features with the ***fibroproliferative*** SSc subset, which represents 30 – 40% of the diffuse SSc (dSSc) patients. These features include enrichment of a TGFβ-responsive signature (**Fig. 6)** and a signature characteristic of proliferating cells. Other models examined include the sclGVHD model[11], the bleomycin-induced (Bleo) fibrosis model and the TβRIIΔk-fib model (unpublished); these mimic the *inflammatory* subset of SSc patients[6].

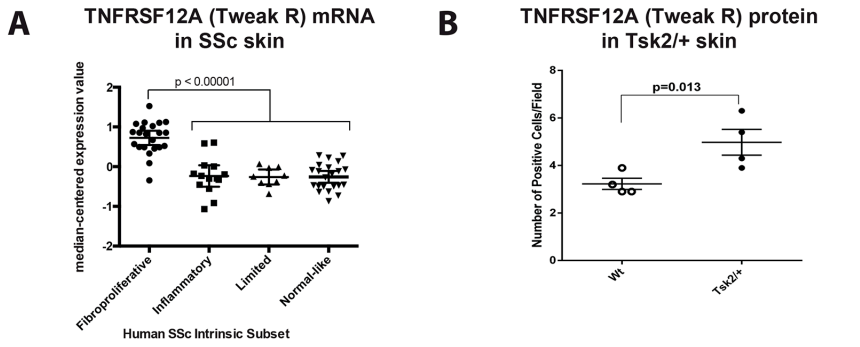


Figure 7. Tweak R (TNFRSF12A) is highly differentially expressed in both the fibroproliferative SSc patients and the Tsk2/+ mouse model. (A) SSc patients in the fibroproliferative subset express significant more tnfrsf12a (Tweak-R) than normal controls or patients in the inflammatory, proliferative or normal-like subsets (p < 0.00001). (B) Tsk2/+ mice expressed 1.6-fold more TNFRSF12A (TWEAK-R) than wild type littermates (p=0.013). Four-week-old female mice were scored for the number of TWEAK-R (Fn14)+ cells per field of view (400X magnification). Significance was calculated with a paired t-test using at least 8 fields of view per mouse and 4 mice per genotype.

Given the observation that the production of a major indicator of fibrosis, COL1A1, is increased by the transfection of the *Col3a1*^{Tsk2} gene, we assessed the impact of the mutant gene genome-wide. RNA from the *Col3a1*^{Tsk2} and *Col3a1*^{WT} transfected *Col3a1*-KO fibroblasts and from four week-old Tsk2/+ and WT littermate skin was analyzed by DNA microarray. Differentially expressed pathways between the two transfections was determined by Gene Set Enrichment Analysis (GSEA). Transfection of *Col3a1*^{Tsk2} results in significant enrichment of genes associated with fibrotic Gene Ontology (GO) terms including *basement membrane*, *extracellular matrix*, *integrin binding*, and *transmembrane receptor protein kinase activity* (Figure 3D; GSEA FDR < 5%). The biological processes observed in the skin of four 4-week old female Tsk2/+ mice relative to WT littermates also shows increases in genes associated with GO terms *extracellular matrix*, *integrin binding* and *basal lamina* (ZL, CB, KBL, CA, EPB, MLW, manuscript *in preparation*).

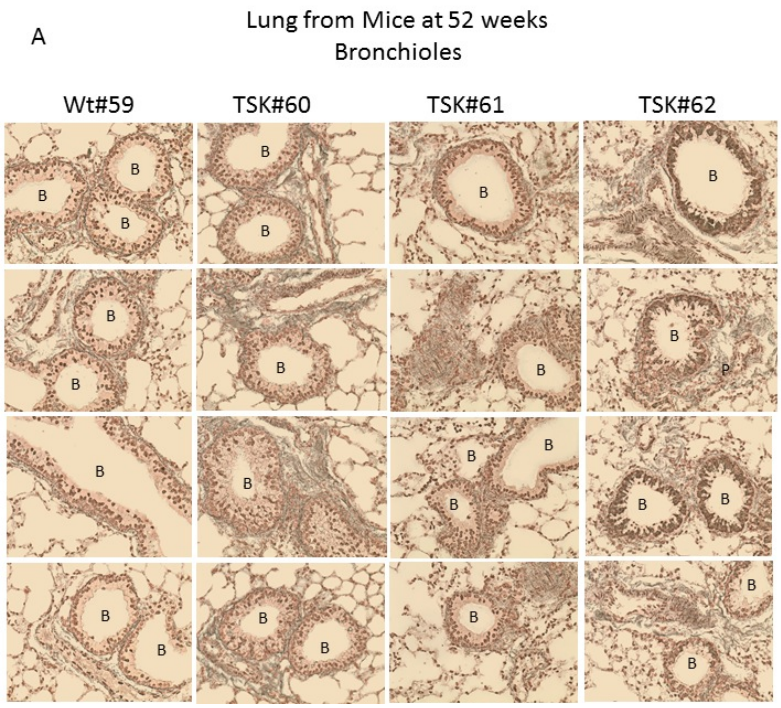


Figure 8. In the wild-type (WT#59), the reticular fibers (gray) line the bronchioles (B) tightly and are well delineated. In TSK2/+ (Tsk2#60, Tsk2#61, and Tsk2#62), the reticular fibers appear to be disorganized around the base of the bronchioles and there is more COL3A1 in the surrounding adventitia.

The genes that significantly contributed to the GSEA pathway enrichment in the transfected fibroblasts were extracted from microarray data of the transfections, as well as from female Tsk2/+ and WT skin at 4 weeks of age (Figure 3E-F), and were elevated both in the fibroblasts transfected with *Col3a1*^{Tsk2} and in Tsk2/+ mouse skin. These include those

genes typically associated with fibrosis including CTGF, THY1, FBN1, the collagens, laminins, TGFB1, TGFBR1, ADAMTS family genes and MMP11. In addition, there was up-regulation in *Col3a1*^{Tsk2}-transfected fibroblasts and Tsk2/+ skin RNA of the VEGF-Receptors *FLT1* and *FLT4*, as well as genes associated with PDGF signaling (PDGFRB and PDGFRL). These data indicate that expression of the *Col3a1*^{Tsk2} gene alone can induce a fibrotic gene expression program.

Milestone 4, new data

Of interest, there is a new pathway in fibrosis that we discovered in this study. To confirm the molecular similarities between the patients in the *fibroproliferative* subset and Tsk2/+ mice we analyzed the expression of the TGFβ-regulated gene tumor necrosis factor receptor superfamily, member 12a (*tnfrsf12a*; also designated *Tweak-R*, *fibroblast growth factor-inducible-14*). *Tweak-R* is highly expressed in the *fibroproliferative* subset and was among a set of genes highly correlated to worse skin disease. Gene expression differences there were confirmed by qRT-PCR [7]. *Tweak-R* mRNA levels were also found increased in the fibroproliferative subsets of Milano *et al.* (**Figure 7a**; $p < 0.00001$). We performed IHC on sections of skin taken from SSc samples, human normal-controls, Tsk2/+ and their wild-type littermates to detect the levels of TWEAK-R. The *Tweak-R* gene is induced by TGFβ [12] and there is an ~1.5-fold increase in numbers of TWEAK-R positive cells at in skin of 4-week-old female Tsk2/+ mice compared to their littermates (Figure 7B; $p = 0.013$). Taken together with the findings of enrichment of TGFβ-signaling in *fibroproliferative* skin biopsies [13], it seems that Tsk2/+ skin at one month reflects the biology of this subset of SSc. These data are now part of Sargent et al. *Submitted*

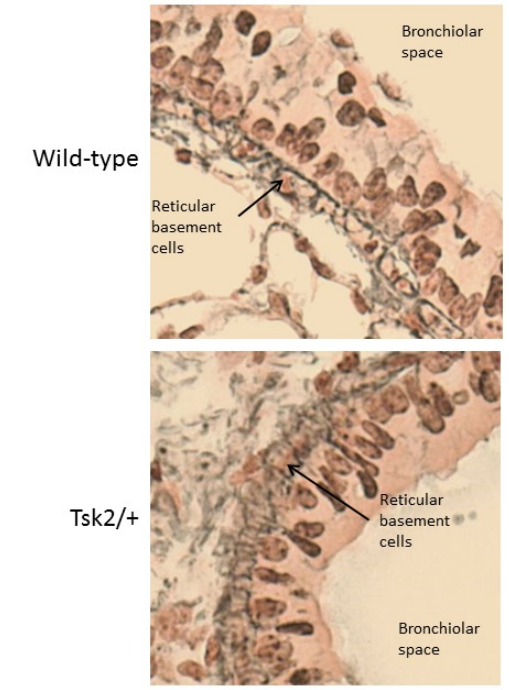


Figure 9. Reticular fibers are disorganized surrounding reticular basement cells in Tsk2/+ mice.

Milestone 5 : Reticular fibers staining of lung tissue

We further pursued the finding of abnormal reticular fibers (type III collagen) in the tissues and investigated lung in Tsk2/+ and wild-type littermate mice. We found that Tsk2/+ mice in addition to having abnormal reticular fibers in the skin (2012 report) we also demonstrate reticular fiber abnormalities in the lung of Tsk2/+ mice at 52 weeks of age (**Figures 8-11).** We are currently investigating earlier time points to determine whether this observation holds true in younger animals. **Figure 8** gives an overview of the overall pathological differences that we observed between Tsk2/+ and their wild-type littermates. These differences are described in more detail in **Figures 9, 10 and 11**. We note that all aspects of the lung tissue in the Tsk2/+ mouse show an overt increase in reticular fibers.

On closer examination of the reticular fibers surrounding the bronchioles, we observed that in the wild-type lung there was a single layer of reticular basement cells surrounded by reticular fibers that was densely stained (arrow). However in the Tsk2/+ mouse the even though the bronchioles were

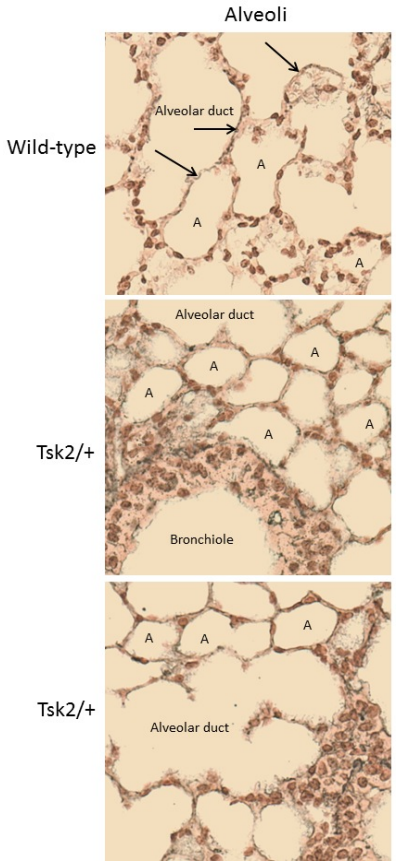


Figure 10. Tsk2/+ alveoli have diffuse and disorganized reticular fibers (gray stain), whereas there was well defined collagen in the wild-type mouse (arrows).

surrounded by a single layer of reticular basement cells (arrow), the reticular fibers were diffuse and not well defined (**Figure 8**). We noted additional lung abnormalities in the Tsk2/+ mouse. We observed that there were increased reticular fibers surrounding the alveoli. Alveoli are normally supported by a fine network of fibers (arrows in wild-type image) however in the Tsk2/+ mouse, there was increased staining for type III collagen and we also noted that this was also disorganized and not well delineated as that observed in the wild-type mouse at the same age (**Figure 9)**.

Finally we note that the smooth muscle in the lung had heavy staining in the Tsk2/+ mouse and that there were thickened areas of stain between the columnar epithelial cells (**Figure 10**).

Previously it has been reported that Tsk2/+ mice had an emphysema type pathology in the lung; however we did not observe this phenotype. We found that the lung parenchyma overall was normal with the exception of the increased reticular fiber staining as shown above.

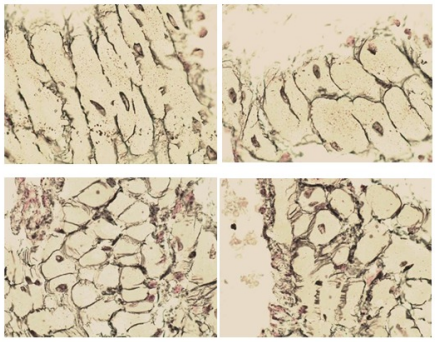


Figure 11. Tsk2/+ mice demonstrate increased reticular fibers surrounding epithelial cells of the lung. Note the increased deposition of stain between the cells in the Tsk2/+ lung.

Based on our finding that there is a mutation in the PIIINP region of the *Col3a1* gene, and that it has been reported that the PIIINP fragment is involved in fibrillogenesis of type I collagen, we speculated that we would not be able to extract protein from the skin of these mice as easily as from the wild-type mice using 1M NaCl or 0.5M acetic acid. Indeed, we found that less protein was extracted with 1M NaCl from the Tsk2/+ skin but the amount of extracted protein with 0.5M acetic acid was not significantly different. Acid extractable protein was found to be 0.33 mg/ml in the wild-type vs 0.32 mg/ml in Tsk2, p=0.47; whereas for salt extracted protein there was 0.86 mg/ml from the wild-type vs. 0.68 mg/ml from Tsk2, p=0.045. We are extending these findings to additional animals and want to determine why there is less extractable protein in Tsk skin. Our goal now is to establish how this mutation alters collagen fibrillogenesis leading to thickened skin and altered reticular fiber pathology.

We submitted a grant application to the Scleroderma Research Foundation on this extensive study of the lung phenotypes in Tsk2/+ mice, which had not been appreciated prior to our study, but it was not funded.

We began the investigation altered protein synthesis of specific genes in the skin of the Tsk2/+ mouse as determined by the increased gene expression that was identified in collaboration with Dr. Michael Whitfield at Dartmouth. This is the subject of a grant application to the NIH (reviewed, Oct 2014).

CONCLUSION:

A non-synonymous mutation in COL3A1 is the causative mutation in Tsk2/+. This mutation results in a time-dependent activation of numerous gene expression pathways, including TGFβ-1 and TWEAK/FN14 signaling, that varies by gender and age. The mechanism by which the COL3A1 mutation causes these changes is yet unknown but we have developed a new hypothesis that the N-terminal propeptide, where the mutation lies, cannot effectively bind TGFβ-1, something we will be examining carefully in future studies.

KEY RESEARCH ACCOMPLISHMENTS Summary (Jul 1, 2011-June 30, 2015)

- We completed the nucleotide sequencing of the mouse Tsk2/+ interval. A number of single nucleotide polymorphisms were found and verified by SNP allele-specific PCR analysis [1] for these previously unsequenced strains (101/H, the parental strain; and B6.tsk2, the B6 congenic strain bearing mutated 101/H DNA on chromosome 1 that contains the Tsk2/+ mutation).

- *Col3A1* is the Tsk2 gene. This SNP is the only one in the interval that shows a coding mutation; it also is interesting that it often shows an expression difference, unlike any of the other genes with non-coding SNPs.
- The mutation lies in the PIIINP fragment of COL3A1^{TSK2}. This peptide is cleaved off the pro-col3a1 pro-protein, and circulates in the body, and is also active locally. We have formed a new hypothesis that the PIIINP^{Tsk2} form cannot bind TGFβ or a related protein, bone morphogenetic protein-2 (BMP-2) as well as PIIINP^{WT} and that this leads to fibrosis and tight skin in mice. This proposition has considerable support from studies of PINP and PIINP[14, 15], and we have recently submitted a grant application to the Scleroderma Foundation to study this intriguing hypothesis. If proven true, for example, it would mean that the excess circulating PIIINP in SSc patients and other fibrotic disorders is functional and is part of a regulatory mechanism to control fibrosis.
- The genetic complementation test of *Col3a1* (mice bred from the Tsk2/+ line to BALB.Col3A1KO mice) was unequivocal and proved *Col3A1* is the *Tsk2* gene
- We proved that the excess deposition of collagen matrix does not occur until well after the Tsk2/+ tight skin phenotype is evident. However, other TGFβ-target genes are elevated as early as two weeks of age.
- Deletion of the NLRP3 inflammasome, or the fibulin-5 gene (Fbln5-KO) had no effect on tight skin or fibrosis, even though the former had a small effect on wound healing, and the latter reduced elastic fibers in the skin. Both these results have been published (see appendix).
- We have prepared a large number of fibroblast cell lines in preparation for *in vitro* studies of the Tsk2 trait under selective conditions.
- Microarray profiling of Tsk2/+ mice at 4, 8, 12 and 20 weeks have now been performed for both male and female mice totaling > 48 samples (we have previously only reported results at two time points for a single gender that totaled 8 samples). We have identified genes that are differentially regulated at each time point. Our prior data was suggestive of increased TGFβ signaling but did not show this directly using an experimentally derived TGFβ gene signature. We have now demonstrated that this is indeed the case. Tsk2/+ mice have a TGFβ1 signature that is seen in a global assessment of mRNA from skin in a carefully controlled study using littermates (Tsk2/+ and WT) at timed stages and stratified by sex. Controlling for hair cycle dramatically improves our ability to detect differently expressed genes and we are now mining these carefully controlled data for differentially expressed pathways.
- We have identified increased activity of members of a new pathway (TWEAK/FN14) in human SSc patients (of the fibroproliferative subset), in Tsk2 mice, and in fibroblasts transfected with *Col3a1*^{Tsk2} cf. *Col3a1*^{WT}. This led us to order FN14-KO mice from Dr. Linda Burkly (Biogen Idec) to breed to Tsk2/+ dams, eventually to create the Tsk2 mutation on a FN14 background. We are excited by this finding, which could be translated into a new therapeutic target for scleroderma.
- **We are in the final stages of publishing the SSc human mouse comparison which was completed in the no-cost extension year.**

The next reporting period:

As this is the final report, "Nothing to Report."

4. IMPACT

What was the impact on the development of the principal discipline(s) of the project?

We have shown that Tsk2/+ is a good model for the fibroprolifertive subset of SSc patients. There were several surprises from our three years of work on fibrosis and scleroderma in the model mouse, Tsk2/+: one is

that a single amino acid change in the N-terminal propeptide of COL3A1 (PIIINP) can cause such a profound change in gene expression, dermal health status, tightness of skin, inhibition of wound healing, and fibrosis. Given this, we have had to revise our hypothesis of the function of this PIIINP fragment that we now believe is not a benign cleavage product but instead a key modulator of skin (and perhaps lung [16]) health. This has led us to further hypothesize that the circulating PIIINP that is elevated in patients with fibrotic disease [17, 18] is not simply a benign biomarker of unfavorable outcomes, but rather may be produced as a consequence of excess COL3A1 production, and that it has a separate function as a compensatory factor, acting to inhibit TGFβ signaling and shut down ECM deposition. The Tsk2 mutation likely inactivates this function, allowing TGFβ production and fibrosis. This notion is highly novel, as there are no publications on the role of PIIINP in scleroderma or any other fibrotic disease.

What was the impact on other disciplines?

Recently, general scientific opinion articles have suggested that animal models and studies therein are dispensable; that the mouse does not model human disease sufficiently well to deserve much effort to study it. In the specific case of Tsk2/+, some reviewers have said that because the mutation in *Col3a1*^{Tsk2} has never been observed in human patients (nor has it been searched for), the Tsk2/+ model is irrelevant. In fact, what we found is that Tsk2/+ mice model the fibroproliferative subset of SSc patients at the level of expressed genes and pathways, including TWEAK/FN14 (aka TWEAK-R). Thus, Tsk2/+ is the *only* animal model that resembles this subset of patients, making it particularly important for SSc research. Our careful dissection of this mouse model of SSc at the levels of gene expression, protein expression, tissue injury and whole body traits has given us more information on its relevance to a particularly difficult-to-treat subset of SSc patients, a link that we argue would be true for other disciplines as well.

What was the impact on technology transfer?

We have submitted a proposal to NIH and Falk Foundation for Medical Research to explore the possibilities of inhibiting TWEAK-R in animal models of SSc, with an eye toward doing so in human subjects. We are planning a submission of an SBIR-like grant this fall to the Scleroderma Research Foundation that will, we hope, involve Biogen Idec, a company with extensive experience, reagents and interest in this signaling pathway.

What was the impact on society beyond science and technology?

Scleroderma is an incurable disease that often has a very poor prognosis. If our results are confirmed and found to translate to the human disease, we could propose at least two ways to ameliorate the tightness of skin of these patients, especially for the fibroproliferative intrinsic molecular subset of patients: manipulation of TWEAK/FN14 and exploration of the potential therapeutic role of PIIINP.

5. CHANGES/PROBLEMS

None to report.

6. PRODUCTS:

REPORTABLE OUTCOMES (July 1, 2011 to June 30, 2014)

Oral Presentations: (Chronological Order)

Fall/11- Michael L. Whitfield, PhD. “Capturing the heterogeneity in systemic sclerosis with high-throughput gene expression profiling” Drexel University Department of Immunology and Microbiology Seminar Series.

3/12 Kristen Long Ph.D. “The Tsk2/+ animal model of Scleroderma”, PhD thesis defense.

6/12 Michael L. Whitfield, PhD. “Capturing the heterogeneity in systemic sclerosis with high-throughput gene expression profiling”. Synergy Translation Research Program, Geisel School of Medicine at Dartmouth. Lebanon, NH

6/12 Michael Whitfield, PhD. “Intrinsic Gene Expression Subsets of Systemic Sclerosis Are Stable in Serial Skin Biopsies and Show Differential Therapeutic Response” Division of Rheumatology, Brigham and Women’s Hospital, Harvard Medical School, Boston MA.

6/12 Michael Whitfield, PhD. “Capturing the heterogeneity in systemic sclerosis with high-throughput gene expression profiling” Scleroderma NIAMS Center of Research Translation Advisory Committee and Investigators Meeting, Boston University School of Medicine, Boston MA

8/12 Elizabeth Blankenhorn, " Genetic control of fibrosis in the Tsk2 Model of Scleroderma". Whitfield Laboratory Annual Retreat. New Hampshire, August 2012

9/12 Chelsea Burgwin, year 3 graduate student supported by DOD. "Mapping the Mutation in the Tight Skin 2 model of systemic sclerosis and Development of an *in vitro* Model." Molecular Cell Biology and Genetics Seminar Series, Drexel University College of Medicine ,September 2012.

10/12 Michael Whitfield, PhD. “Subsetting systemic sclerosis by high-throughput gene expression”. Scleroderma SCOT clinical trial investigators meeting (NIH sponsored), Potomac MD

11/12 Michael Whitfield, PhD. “NIAMS p30 Microarray Core for Scleroderma”, P30 Investigators and Users National presentation. Boston, MA

3/13 Michael Whitfield, PhD. “Integrative Genomics in Systemic Sclerosis”, Stanford Epithelial Biology Program. Stanford, CA

3/13 Michael Whitfield, PhD. “Plasticity and interconnectivity of the SSc intrinsic Subsets” Scleroderma Research Foundation Workshop, San Francisco, CA

4/13 Michael L. Whitfield, PhD "P50 Core: A focus on Next Generation Generation sequencing methods in SSc-PAH". Scleroderma NIAMS Center of Research Translation Advisory Committee and Investigators Meeting, Boston University School of Medicine, Boston MA.

3/13 Zhenghui Li (Whitfield Lab). “Identifying Genetic Variation in the Tsk2/+ Model of Systemic Sclerosis” Molecular and Cellular Biology Program, Research in Progress.

5/13 Michael L. Whitfield, PhD "Personalized Medicine and Disease Pathogenesis in Systemic Sclerosis by Integrative Genomics" Geisel School of Medicine, Department of Micro-Immuno Annual Retreat.

6/13 Zhenghui Li (Whitfield lab), "RNA-seq and miR-seq analysis of SSc skin across intrinsic gene expression subsets shows differential expression of non-coding RNAs regulating SSc gene expression" FGED Meeting, Seattle WA.

6/13 Michael L. Whitfield, PhD "P30 SSc Microarray and Bioinformatics Core: Services to the national SSc community". Boston MA.

8/13 Elizabeth Blankenhorn PhD, Plenary talk, “The genetic basis for Tsk2/+ fibrosis” 13th International Workshop on Scleroderma Research, Boston, Massachusetts, August 2013.

8/13 Zhenghui Li(Whitfield lab), “RNA-Seq and miRNA-seq analysis of SSc skin across intrinsic gene expression subsets…” 13th International Workshop on Scleroderma Research, Boston, Massachusetts, August 2013.

1/14 Chelsea Burgwin, year 4 graduate student supported by DOD. “Tsk2/+ Wound Healing and *in vitro* Model” Molecular Cell Biology and Genetics Seminar Series, Drexel University College of Medicine, September 2012.

3/14 Chelsea Burgwin, year 4 graduate student supported by DOD. “Tsk2/+ Wound Healing and *in vitro* Model” University of Pennsylvania’s Cell Mechanics Journal Club and Presentation Series, March 2014.

6/14 Chelsea Burgwin, year 4 graduate student supported by DOD. “Signaling in Skin and Fibroblasts from Tsk2/+ Mice: Elucidating the Mechanism in the Tsk2/+ Mouse Model of Scleroderma” 2014 International Symposium on Molecular Medicine and Infectious Disease, Drexel University, June 2014.

6/14 Elizabeth Blankenhorn PhD, “Genetics of scleroderma in animal models” 2014 International Symposium on Molecular Medicine and Infectious Disease, Drexel University (June 2014).

6/14 Kristen Long PhD, “The role of elastin in wound healing and tight skin trait in Tsk2 mice”2014 International Symposium on Molecular Medicine and Infectious Disease, Drexel University (June 2014).

6/14 Michael L. Whitfield PhD, “Intrinsic molecular subsets of scleroderma”2014 International Symposium on Molecular Medicine and Infectious Disease, Drexel University (June 2014).

6/14 Carol Artlett PhD, “Role of the Inflammasome and miRNA in fibrosis” 2014 International Symposium on Molecular Medicine and Infectious Disease, Drexel University (June 2014).

6/14 Michael L. Whitfield PhD, “*Scleroderma NIAMS Microarray Core Center for Scleroderma*” NIAMS P30 Core Centers for Scleroderma Research, Advisory Committee Meeting

6/14 Michael L. Whitfield PhD, “*Biomarkers of SSc-PAH in SSc skin and blood*”. Boston University NIAMS Scleroderma Center of Resesarch Translation (CORT), Advisory Committee Meeting, Boston MA

11/14 Michael L. Whitfield PhD, “*Identification of the Microbiome As a Potential Trigger of Systemic Sclerosis By Metagenomic RNA-Sequencing of Skin Biopsies*” ACR Basic Research Conference Boston, MA.

12/14 Michael L. Whitfield PhD, “*Untangling molecular changes in SSc clinical trials: Gene expression subsets, response signatures and pathway changes*” ASSET Investigator Meeting. University of Michigan, MI

2/15 Michael L. Whitfield PhD, “*Enabling Precision Medicine in SSc Clinical Trials*” Discussion leader and presenter, NIAMS roundtable discussion on Scleroderma: Advancing Potential Drugs to Patient Care

3/15 Michael L. Whitfield PhD, “*Molecular stratification and drug response for SSc clinical trials*” Pfizer, Cambridge, MA.

3/15 Michael L. Whitfield PhD, “*A macrophage-associated inflammatory signature is found in all SSc tissues and associated with more severe disease*” Scleroderma Research Foundation Workshop on Scleroderma, San Francisco, CA

4/15 Michael L. Whitfield PhD, “*Genomics, Bioinformatics and Systems Biology for Precision Medicine in Systemic Sclerosis*”. SScores (NIH P30) Advisory Committee meeting. Boston University Medical Center, Boston MA.

6/15 Michael L. Whitfield PhD, “*Genomics, Bioinformatics and Systems Biology for Precision Medicine in Systemic Sclerosis*”. NIH CORT (P50) Advisory Committee meeting. Boston University Medical Center, Boston MA

8/15 Michael L. Whitfield PhD, *“Systems Biology in Systemic Sclerosis.”* Session Chair and topic introduction. Scleroderma Basic Science Workshop, Cambridge UK.

Abstracts and Presentations: (Chronological Order)

1. Long, K.B., C.M. Burgwin, C.M. Artlett and E.P. Blankenhorn. “Elastin and collagen anomalies in pre-fibrotic skin in the Tsk2/+ mouse, a model for Scleroderma.” 20th Annual Infection and Immunity Forum, Eastern PA Branch of ASM, Drexel University College of Medicine, Philadelphia, Pennsylvania, June 2011.

2. Long, K.B., C.M. Burgwin, C.M. Artlett and E.P. Blankenhorn. “Examination of prefibrotic and fibrotic disease in Tsk2/+ mice shows early ECM changes are attributable to elastin dysregulation.” 12th International Workshop on Scleroderma Research, Cambridge, UK, July 2011.

3. Long*, K.B., C.M. Burgwin, C.M. Artlett and E.P. Blankenhorn. “Examination of prefibrotic and fibrotic disease in Tsk2/+ mice shows early ECM changes are attributable to elastin dysregulation.” Drexel University College of Medicine, Discovery Day Research Symposium, Philadelphia, Pennsylvania, October 2011. *Outstanding Platform Presentation award winner

4. John, A.K., K.B. Long, L. Cort and E.P. Blankenhorn. “Fibroblast investigations suggest that increased collagen production is cell-autonomous in the Tsk2/+ mouse model of scleroderma.” Drexel University College of Medicine, Discovery Day Research Symposium, Philadelphia, Pennsylvania, October 2011.

5. C.M. Burgwin, K.B. Long, Zhenghui Li*, C.M. Artlett, M. Whitfield*, and E.P. Blankenhorn. “Mapping of the Mutation in the Tight Skin 2 model of systemic sclerosis” Department of Microbiology and Immunology, Drexel Univ. College of Medicine, Philadelphia, PA. *Dartmouth Medical School, Hanover, NH, Institute for Molecular Medicine and Infectious Disease International Symposium, Philadelphia, PA, June 2012

6. Zhenghui L. "Identifying Genetic Variations in the Tsk2 Model of Scleroderma". Dartmouth MCB Program Annual Retreat. Whitefield, New Hampshire, August 2012

7. Burgwin, C.M., K.B. Long, Z. Li., C.M. Artlett, M.L. Whitfield and E.P. Blankenhorn. “Complementation Experiment Examining the Role Collagen3α1 plays in Disease Development of the Tight Skin 2 Mouse Model of Systemic Sclerosis” Discovery Day, Drexel University College of Medicine, Philadelphia, Pennsylvania, October 2012.

8. Zhenghui Li, Eleni Marmarelis, Qu Kun, Lionel Brooks, Gavin D. Grant, Patricia A. Pioli, Howard Chang, Robert Lafyatis, and Michael L. Whitfield. "RNA-seq and miR-seq analysis of SSc skin across intrinsic gene expression subsets shows differential expression of non-coding RNAs regulating SSc gene expression". Functional Genomics Data Society 15th Annual Meeting, Seattle, Washington, June 2013

9. Zhenghui Li, Kristen Long, Carol Artlett, Elizabeth Blankenhorn and Michael L. Whitfield. "Identifying Genetic Variations in the Tsk2 Model of Scleroderma". Dartmouth MCB Program Annual Retreat. Whitefield, New Hampshire, August 2013

10. Burgwin, C.M., K.B. Long, Zhenghui Li, M. Whitfield, C.M. Artlett, E.P Blankenhorn “TGF-β’s and Elastin’s Roles in Disease Progression in the Tsk2/+ Mouse Model of Systemic Sclerosis.”

a. 13th International Workshop on Scleroderma Research, Boston, Massachusetts, August 2013. *Won Drexel GSA Travel Award

b. Discovery Day, Drexel University College of Medicine, Philadelphia, Pennsylvania, October 2013. *Poster won Honorable Mention.

11. Burgwin, C.M., S. Sassi-Gaha, K.B. Long, Zhenghui Li, M. Whitfield, C.M. Artlett, E.P Blankenhorn “Elucidating the Mechanism of Fibrosis in the Tight skin 2 (Tsk2/+) Mouse Model of Scleroderma” 2014 International Symposium on Molecular Medicine and Infectious Disease, Drexel University, Philadelphia, Pennsylvania, June 2014.

12. Zhenghui Li, Kristen Long, Carol Artlett, Elizabeth Blankenhorn and Michael L. Whitfield. "Identifying Genetic Variations in the Tsk2 Model of Scleroderma". Dartmouth MCB Program Annual Retreat. Whitefield, New Hampshire, August 2014.

13. Zhenghui Li, Eleni Marmarelis, Kun Qu, Lionel Brooks, Patricia A. Pioli, Howard Y. Chang, Robert Lafyatis, and **Michael L. Whitfield**. RNA-seq and miR-seq analysis of SSc skin across intrinsic gene expression subsets shows differential expression of non-coding RNAs regulating SSc gene expression. American College of Rheumatology Annual Meeting, 2014

Manuscripts:

1. Long, K.B., Li, Z.., Burgwin, C., Choe, S.G., Martyanov, V., Sassi-Gaha, S., Earl, J., Eutsey, R., Ahmed, A., Ehrlich, G.D., Artlett, C.M., Whitfield, M.L., and Blankenhorn, E. P. The Tsk2/+ fibrotic phenotype is due to a gain-of-function mutation in the PIIINP segment of the Col3a1 gene. *Journal of Investigative Dermatology*, 2014 (in press)
2. Long, K.B., Artlett, C.M., and Blankenhorn E. P. Tight Skin 2 Mice exhibit a novel time line of events leading to increased extracellular matrix deposition and dermal fibrosis, *Matrix Biology*, doi: 10.1016/j.matbio.2014.05.002. PMID: 24820199, 2014
3. Long*, K.B., Burgwin*, C.M., Huneke, R., Artlett, C.M., and Blankenhorn, E. P. The wound healing deficit in Tsk2/+ is due to excess elastin. Advances in Wound Care, *in press,* 2014. PMID: 25207200. *Equal contributions. Invited submission, and Dr. Blankenhorn was a guest editor of this journal issue.
4. Sargent, J.L., Zhenghui Li, A.O. Aliprantis, M. Greenblatt, R. Lemaire, M.H. Wu, J. Wei, A. Harris, K. Long, C.M. Burgwin, C.M. Artlett, E.P. Blankenhorn, R. Lafyatis, J. Varga, S. H. Clark, M.L. Whitfield. "Interspecies Comparative Genomics Identifies Optimal Murine Models of Scleroderma Subsets" Submitted, under revision for Arthritis and Rheumatology
5. Zhenghui Li, Kristen B. Long, Chelsea Burgwin, Carol M. Artlett and Elizabeth P. Blankenhorn, Michael L. Whitfield. The Tsk2/+ mouse model of Systemic Sclerosis shows variable age and gender dependent gene expression in skin. Manuscript *in preparation*.

Degrees obtained that are supported by this award

Chelsea M. Burgwin is working on her PhD thesis, expected defense in Spring, 2016. Her work is directly supported by this grant.

Zhenghui Li has defended his thesis and will receive his PhD thesis in December 2015. His work is directly supported by this grant.

Kristen Long successfully defended her PhD thesis, entitled "Examination of a model of systemic sclerosis, the Tight Skin 2 mouse: before, during and after fibrotic disease" in April, 2012. She was awarded her PhD by Drexel University College of Medicine in May 2012. Her work was directly supported by this grant.

Development of cell lines, tissue or serum repositories

We have developed a large number of Tsk2/+ and WT littermate fibroblast cell lines from mice of various ages and both sexes.

7. PARTICIPANTS & OTHER COLLABORATING ORGANIZATIONS

There were no changes to the personnel at Drexel and no new changes in the laboratory of the partnering PI, Dr. Michael Whitfield at Geisel School of Medicine at Dartmouth College, NH.

8. SPECIAL REPORTING REQUIREMENTS

COLLABORATIVE AWARDS: For collaborative awards, independent reports are required from BOTH the Initiating PI and the Collaborating/Partnering PI. A duplicative report is acceptable; however, tasks shall be clearly marked with the responsible PI and research site. A report shall be submitted to https://ers.amedd.army.mil for each unique award.

An identical final progress report will be sent from Dr. Whitfield.

9. REFERENCES

1. Bunce M, O'Neill CM, Barnardo MC, Krausa P, Browning MJ, Morris PJ, Welsh KI: **Phototyping: comprehensive DNA typing for HLA-A, B, C, DRB1, DRB3, DRB4, DRB5 & DQB1 by PCR with 144 primer mixes utilizing sequence-specific primers (PCR-SSP).** *Tissue Antigens* 1995, **46:**355-367.
2. Ishigame H, Mosaheb MM, Sanjabi S, Flavell RA: **Truncated Form of TGF-betaRII, But Not Its Absence, Induces Memory CD8+ T Cell Expansion and Lymphoproliferative Disorder in Mice.** *J Immunol* 2013, **190:**6340-6350.
3. Barisic-Dujmovic T, Boban I, Clark SH: **Regulation of collagen gene expression in the Tsk2 mouse.** *J Cell Physiol* 2008, **215:**464-471.
4. Christner PJ, Hitraya EG, Peters J, McGrath R, Jimenez SA: **Transcriptional activation of the alpha1(I) procollagen gene and up-regulation of alpha1(I) and alpha1(III) procollagen messenger RNA in dermal fibroblasts from tight skin 2 mice.** *Arthritis Rheum* 1998, **41:**2132-2142.
5. Sargent JL, Milano A, Bhattacharyya S, Varga J, Connolly MK, Chang HY, Whitfield ML: **A TGF beta-Responsive Gene Signature Is Associated with a Subset of Diffuse Scleroderma with Increased Disease Severity.** *Journal of Investigative Dermatology* 2010, **130:**694-705.
6. Sargent JL, Li Z, Aliprantis AO, Greenblatt M, Lemaire R, Wu M-h, Wei J, Harris A, Long K, Burgwin C, et al: **Interspecies Comparative Genomics Identifies Optimal Mouse Models of Scleroderma Subsets.** *Submitted* 2014.
7. Milano A, Pendergrass SA, Sargent JL, George LK, McCalmont TH, Connolly MK, Whitfield ML: **Molecular subsets in the gene expression signatures of scleroderma skin.** *PLoS One* 2008, **3:**e2696.
8. Pendergrass SA, Lemaire R, Francis IP, Mahoney JM, Lafyatis R, Whitfield ML: **Intrinsic Gene Expression Subsets of Diffuse Cutaneous Systemic Sclerosis Are Stable in Serial Skin Biopsies.** *Journal of Investigative Dermatology* 2012, **132:**1363-1373.
9. Hinchcliff ME, Huang CC, Wood TA, Mahoney JM, Martyanov V, Bhattacharya S, Tamaki Z, Carns M, Podlusky S, Sirajuddin A, et al: **Molecular Signatures in Skin Associated with Clinical Improvement During Mycophenolate Treatment in Systemic Sclerosis.** *J Invest Dermatol* 2013, **In Press**.
10. Hinchcliff M, Huang CC, Wood TA, Matthew Mahoney J, Martyanov V, Bhattacharyya S, Tamaki Z, Lee J, Carns M, Podlusky S, et al: **Molecular signatures in skin associated with clinical improvement during mycophenolate treatment in systemic sclerosis.** *J Invest Dermatol* 2013, **133:**1979-1989.
11. Greenblatt MB, Sargent JL, Farina G, Tsang K, Lafyatis R, Glimcher LH, Whitfield ML, Aliprantis AO: **Interspecies comparison of human and murine scleroderma reveals IL-13 and CCL2 as disease subset-specific targets.** *Am J Pathol* 2012, **180:**1080-1094.
12. Milks MW, Cripps JG, Lin H, Wang J, Robinson RT, Sargent JL, Whitfield ML, Gorham JD: **The role of Ifng in alterations in liver gene expression in a mouse model of fulminant autoimmune hepatitis.** *Liver Int* 2009, **29:**1307-1315.
13. Sargent JL, Milano A, Bhattacharyya S, Varga J, Connolly MK, Chang HY, Whitfield ML: **A TGFbeta-responsive gene signature is associated with a subset of diffuse scleroderma with increased disease severity.** *J Invest Dermatol* 2009, **130:**694-705.
14. Oganesian A, Au S, Horst JA, Holzhausen LC, Macy AJ, Pace JM, Bornstein P: **The NH2-terminal propeptide of type I procollagen acts intracellularly to modulate cell function.** *J Biol Chem* 2006, **281:**38507-38518.
15. Zhu Y, Oganesian A, Keene DR, Sandell LJ: **Type IIA procollagen containing the cysteine-rich amino propeptide is deposited in the extracellular matrix of prechondrogenic tissue and binds to TGF-beta1 and BMP-2.** *J Cell Biol* 1999, **144:**1069-1080.
16. Lee YJ, Shin KC, Kang SW, Lee EB, Kim HA, Song YW: **Type III procollagen N-terminal propeptide, soluble interleukin-2 receptor, and von Willebrand factor in systemic sclerosis.** *Clin Exp Rheumatol* 2001, **19:**69-74.

17. Vettori S, Maresca L, Cuomo G, Abbadessa S, Leonardo G, Valentini G: **Clinical and subclinical atherosclerosis in systemic sclerosis: consequences of previous corticosteroid treatment.** *Scand J Rheumatol* 2010, **39:**485-489.
18. Nagy Z, Czirják L: **Increased levels of amino terminal propeptide of type III procollagen are an unfavourable predictor of survival in systemic sclerosis.** *Clin Exp Rheumatol* 2005, **23:**165-172.

10. APPENDIX

1. Manuscript by Long, Li, Burgwin, Choe, Martyanov, Sassi-Gaha, Earl, Eutsey, Ahmed, Ehrlich, Artlett, Whitfield, and Blankenhorn, (Journal of Investigative Dermatology, in press, 2014).
2. Manuscript by Sargent, Li, Aliprantis, Greenblatt, Lemaire, Wu, Wei, Harris, Long, Burgwin, Artlett, Blankenhorn, Lafyatis, Varga, Clark, and Whitfield, submitted 2014.

The Tsk2/+ Mouse Fibrotic Phenotype Is Due to a Gain-of-Function Mutation in the PIIINP Segment of the *Col3a1* Gene

Kristen B. Long[1], Zhenghui Li[2], Chelsea M. Burgwin[1], Susanna G. Choe[2], Viktor Martyanov[2], Sihem Sassi-Gaha[1], Josh P. Earl[3], Rory A. Eutsey[3], Azad Ahmed[3], Garth D. Ehrlich[3], Carol M. Artlett[1], Michael L. Whitfield[2] and Elizabeth P. Blankenhorn[1]

Systemic sclerosis (SSc) is a polygenic, autoimmune disorder of unknown etiology, characterized by the excessive accumulation of extracellular matrix (ECM) proteins, vascular alterations, and autoantibodies. The tight skin (Tsk)2/+ mouse model of SSc demonstrates signs similar to SSc including tight skin and excessive deposition of dermal ECM proteins. By linkage analysis, we mapped the *Tsk2* gene mutation to <3 megabases on chromosome 1. We performed both RNA sequencing of skin transcripts and genome capture DNA sequencing of the region spanning this interval in Tsk2/+ and wild-type littermates. A missense point mutation in the procollagen III amino terminal propeptide segment (PIIINP) of collagen, type III, alpha 1 (*Col3a1*) was found to be the best candidate for *Tsk2*; hence, both *in vivo* and *in vitro* genetic complementation tests were used to prove that this *Col3a1* mutation is the *Tsk2* gene. All previously documented mutations in the human *Col3a1* gene are associated with the Ehlers–Danlos syndrome, a connective tissue disorder that leads to a defect in type III collagen synthesis. To our knowledge, the *Tsk2* point mutation is the first documented gain-of-function mutation associated with *Col3a1*, which leads instead to fibrosis. This discovery provides insight into the mechanism of skin fibrosis manifested by Tsk2/+ mice.

Journal of Investigative Dermatology (2015) **135,** 718–727; doi:10.1038/jid.2014.455; published online 20 November 2014

INTRODUCTION

There are multiple animal models of systemic sclerosis (SSc) (Artlett, 2010); yet, none mimics all facets of SSc disease. Of the genetic models, the cause of disease in tight-skin 1 (Tsk1/+) mice is known to be a tandem duplication in the fibrillin-1 (*Fbn1*) gene (Siracusa *et al.*, 1996). Other models of SSc have employed mice with individual gene deficiencies or overexpression including Fos-related antigen-2 (*Fra2;* Maurer *et al.*, 2009), endothelin-1 (*Edn1;* Hocher *et al.*, 2000; Richard *et al.*, 2008), and Friend leukemia integration 1 transcription factor (*Fli1;* Asano *et al.*, 2010), which have proven useful for understanding the contribution of these proteins to the vasculopathy and/or lung fibrosis seen in SSc. Nongenetic models of SSc include the bleomycin-induced scleroderma model (Yamamoto *et al.*, 1999), which has been used to study many of the initiating events involved in fibrosis.

The Tsk2/+ mouse was first described in 1986, when an offspring of a 101/H mouse exposed to the mutagenic agent ethylnitrosourea was noted to have tight skin in the interscapular region (Peters and Ball, 1986). The mutagenized gene causing SSc-like signs in Tsk2/+ mice was reported to be located on chromosome 1 between 42.5 and 52.5 megabases (Mb; Christner *et al.*, 1996); however, the genetic defect was never identified. Similar to *Tsk1*, *Tsk2* SSc-like traits are highly penetrant in Tsk2/+ heterozygotes and it is homozygous embryonic lethal. Tsk2/+ mice have many features of human disease including tight skin, dysregulated dermal extracellular matrix (ECM) deposition, and evidence of an autoimmune response (Christner *et al.*, 1995; Gentiletti *et al.*, 2005).

Herein, we report the positional cloning and identity of the *Tsk2* gene. We have discovered that Tsk2/+ mice carry a deleterious gain-of-function missense mutation in *Col3a1* (collagen, type III, alpha 1), which exchanges a cysteine for serine in the N-terminal propeptide, procollagen III amino terminal propeptide segment (PIIINP). The Tsk2/+ mouse affords a unique opportunity to examine the pathways leading to the multiple clinical parameters of fibrotic disease from birth onward.

[1]*Department of Microbiology and Immunology, Drexel University College of Medicine, Philadelphia, Pennsylvania, USA;* [2]*Department of Genetics, Geisel School of Medicine at Dartmouth, Hanover, New Hampshire, USA and* [3]*Center for Genomic Sciences, Pittsburgh, Pennsylvania, USA*

Correspondence: Elizabeth P. Blankenhorn, Department of Microbiology and Immunology, Drexel University College of Medicine, 2900 Queen Lane, Philadelphia, Pennsylvania 19129, USA.
E-mail: Elizabeth.Blankenhorn@drexelmed.edu

Abbreviations: B6, C57Bl/6; Col3a1, collagen, type III, alpha 1; ECM, extracellular matrix; KO, knockout; Mb, megabase; PIIINP, procollagen III amino terminal propeptide segment; SNP, single-nucleotide polymorphism; SSc, systemic sclerosis; Tsk, tight skin; UTR, untranslated region; WT, wild type

Received 13 June 2014; revised 9 September 2014; accepted 22 September 2014; accepted article preview online 20 October 2014; published online 20 November 2014

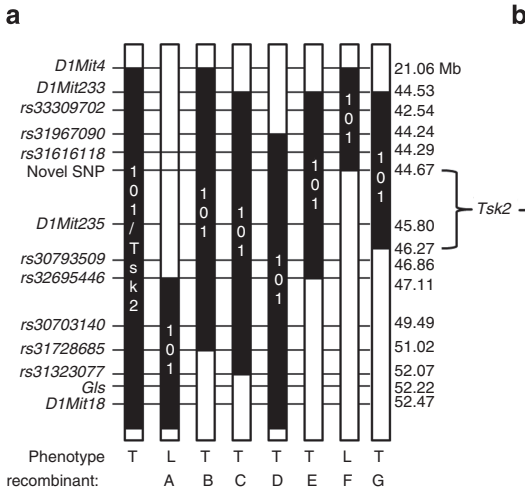
RESULTS

Linkage and sequencing studies reveal a SNP mutation in Col3a1

Identification of the *Tsk2* gene was initiated with further mapping of the *Tsk2* interval by genotyping backcross progeny of Tsk2/+ mice bred to C57Bl/6 (B6) mice. Littermate mice were genotyped for informative microsatellites (*D1Mit233*, *D1Mit235*, a microsatellite in *Gls*, and *D1Mit18*) and single-nucleotide polymorphism (SNP) genotyping assays used for additional markers. Multiple recombinants were recovered that mapped the interval to between 42.53 and 52.22 Mb on chromosome 1. Recombinants were bred and then back-crossed to a consomic B6.chr 1-A/J mouse to fine map the region by SNP typing, as A/J mice bear many known SNPs compared with B6 mice. Additional recombinants were recovered and new SNPs from the sequencing projects (see below) were used to narrow the *Tsk2* interval to between 44.67 and 46.27 Mb (Figure 1a), representing a >3-fold reduction in the size of the interval bearing 101/H genomic DNA and *Tsk2*. There are six known genes in this interval (Figure 1b).

To identify the mutation underlying *Tsk2*, we employed both RNA sequencing (RNA-Seq) and genome capture sequencing of the reduced genomic interval. Sequence reads were aligned to the MM9 reference genome (B6) and analyzed for polymorphisms in the *Tsk2* interval. There were 265 SNPs found in both wild type (WT) and Tsk2/+ littermates that represent differences between the reference B6 genome and the 101/H background; these were excluded from further study. Thirteen SNPs were found in all four Tsk2/+ mice analyzed; 10 of these SNPs were also found to be in liver RNA from 101/H strain or in other non-fibrotic mouse strains (http://phenome.jax.org/) and were also ruled out as candidates for *Tsk2* (Table 1). The remaining three SNPs were heterozygous and confirmed to be only in Tsk2/+ mice. One of these, in a *Gulp1* intron, proved useful as an additional marker that resides outside the supported linkage interval for Tsk2/+ on the proximal end in an informative recombinant mouse (Figure 1a). A second SNP was also found in an intron of *Gulp1*. The RNA-Seq data did not identify any splicing defects in *Gulp1* mRNA in the Tsk2/+ mice (Supplementary Figure S1 online), indicating that this SNP does not change *Gulp1* mRNA splicing, and its gene expression in the skin is unchanged (Figure 2). Thus, the intronic SNP in *Gulp1* is unlikely to have a role in the tight skin phenotype. The remaining mutation was in *Col3a1* that results in a T-to-A transversion at Chr1:45,378,353, causing a Cys→Ser amino acid change in the PIIINP, a natural cleavage product of COL3A1. The mutant protein is designated COL3A1^{Tsk2} (C33S).

We calculated the reads per kilobase per million mapped reads for each gene and found that of the genes in the reduced genomic interval, *Col3a1* shows the highest absolute expression level with all other genes showing negligible expression levels. RNA-Seq results indicate that there is a trend toward higher *Col3a1* mRNA abundance in 4-week-old Tsk2/+ skin samples compared with WT littermates (Figure 2a and b). The *Col3a1*^{Tsk2} (C33S) mutation is unlikely to change the expression levels of the *Col3a1* mRNA directly but will result in a mutated protein that is deposited in the ECM along with the WT protein in mixed heterotrimers, and could result in activation of pathways that impinge on *Col3a1*, such as transforming growth factor-β (Sargent, *et al.*, submitted). Because Tsk2/+ (affected) mice are heterozygous, the *Col3a1*^{Tsk2} (C33S) mutation should account for 50% of the reads assuming equal expression from each allele. We calculated the read count from the RNA-Seq data for the reference and alternate alleles for *Col3a1* at Chr1:45,378,353. In WT mice, we find that all reads (492 total) contain the



b

Gene start site relative to centromere (Mb)	Gene name or gene symbol
43.01	G protein-coupled receptor 45
43.10	TGF beta receptor associated protein-1
43.81	UDP-glucuronate decarboxylase 1
43.99	Tripeptidyl peptidase II
44.20	ERCC5
44.61	Gulp1
45.37	Procollagen, type III, alpha 1
45.43	Procollagen, type V, alpha 2
45.85	WD repeat domain 75
45.97	Solute carrier family 40 (iron-regulated)
46.12	Dynein, axonemal, heavy chain 7B
46.87	Solute carrier family 39 (zinc transporter)
51.53	10 Days neonate skin cDNA, hypothetical
51.56	Cortex cDNA
51.81	Myosin IB
52.07	STAT4
52.20	STAT1
52.24	Glutaminase
52.73	Macrophage MHC receptor 2 isoform 2

Figure 1. *Tsk2* lies between and not including 44.67–46.27 Mb on chromosome 1. (**a**) The *Tsk2* interval was narrowed by genotyping backcrossed mice on the B6 and B6.chr 1-A/J backgrounds. Black bars (101/H) depict the original parental strain, bearing *Tsk2*. White bars depict the B6 genome. Recombinants A–G bear additional recombination sites. The phenotypes are tight (T—Tsk2/+) or loose (L—WT). (**b**) With the use of additional markers (arrows, see text), the current interval comprises *Col3a1, Col5a2, Wdr75, Slc40a1,* part of *Gulp1*, and part of *Dnahc7b*; the five latter genes do not have coding region mutations. The elements of the *Gulp1* gene above 44.67 Mb are excluded by the recombination in mouse F, and *Dnahc7b* below 46.27 is excluded by mouse G. B6, C57Bl/6; Col3a1, collagen, type III, alpha 1; Mb, megabases; SNP, single-nucleotide polymorphism; Tsk, tight skin; WT, wild type.

Table 1. Nucleotide changes between Tsk2/+ mice and 101/H or B6 mice

Nucleotide position on Chr 1 (MM9)	Genotype of Tsk2/+	Genotype of B6	Genotype of 101/H	Present in other strains?	Potential candidate for *Tsk2*?	Gene or mRNA containing substitution
SNP found by RNA-Seq						
44,675,490	A	T	T	No	No, outside interval	*Gulp1* intron
44,833,682*	C	T	T	No	Yes	*Gulp1* Intron
45,378,353*	A	T	T	No	Yes	*Col3a1* exon (C33S)
45,432,389	C	G	ND	Yes	No	*Col5a2* 3′UTR
45,441,243	C	A	C	No	No, in 101/H	*Col5a2* intron
45,860,529	G	A	G	Yes	No	*Wdr75* intron
45,874,790	T	C	T	Yes	No	*Wdr75* intron
45,875,728	C	T	C	Yes	No	*Wdr75* exon
45,880,257	CG	AC	CG	No	No, in 101/H	*Wdr75* exon
46,872,610	T	G	ND	Yes	No	*Slc39a10* intron
46,874,711	C	T	C	Yes	No	*Slc39a10* intron
46,939,340	T	C	T	Yes	No	BC040767 intron
46,939,624	A	G	ND	Yes	No	BC040767 intron
SNP found by Genome Capture Sequencing (454)						
44,833,682*	C	T	T	No	Yes	*Gulp1* intron
45,378,353*	A	T	T	No	YES	*Col3a1* exon (C33S)
45,465,923	A	T	T	No	YES	*Col5a2* intron
46,124,856	A	G	A	Yes	No	*Dnahc76* intron
46,124,857	A	C	T	Yes	No	*Dnahc76* intron
46,268,651	C	T	T	No	No, outside interval	*Dnahc76* intron

Abbreviations: B6, C57Bl/6; Chr, chromosome; Col3a1, collagen, type III, alpha 1; ND,not determined; RNA-Seq, RNA sequencing; SNP, single-nucleotide polymorphism; Tsk, tight skin.
All single-copy nucleotide changes found by RNA-Seq or 454 sequencing were checked for their presence in other non-fibrotic strains (http://phenome.jax.org/) or individually verified by a phototyping assay (Bunce *et al.*, 1995) and/or resequencing to confirm the single-nucleotide change. SNPs that were ruled out by one of these assays are considered not to be potential candidates for *Tsk2*. When known, genotypes shown for 101/H are from RNA-Seq, 454 sequencing, or phototyping. *, Seen in both assays.

reference T allele, whereas in Tsk2/+, we find that 48% of reads (273/564 total reads) contain the WT (T) allele and 52% (291/564 total reads) contain the $Col3a1^{Tsk2}$ (C33S) allele (T -> A; Figure 2c). As a comparison, we show that the intronic *Gulp1* SNP at Chr1:44,833,682 has significantly lower read coverage consistent with its intronic location (11-fold coverage in Tsk2/+ and 2-fold coverage in WT). The intronic *Gulp1* SNP also shows a distribution of reads consistent with heterozygosity in Tsk2/+ and with homozygosity in WT (Figure 2d). These findings show that the $Col3a1^{Tsk2}$ (C33S) locus is heterozygous as expected for the *Tsk2* mutation in these animals, and expression occurs equally from each of the alleles.

Because RNA-Seq only captures variation in the transcribed regions of the genome, and thus might miss an important genomic feature that is unique to *Tsk2*, we sequenced captured genomic DNA samples corresponding to the minimal linkage region from B6.Tsk2/+ heterozygotes and 101/H homozygous parental strain mice. Multiple DNA differences between the Tsk2/+ mouse and its parental 101/H strain were detected. A majority of the differences observed were accounted for by non-chromosome 1 repetitive DNA sequences such as LINE, SINE, and retroviral elements contained within the *Tsk2* interval on chromosome 1. After filtering repetitive elements from the comparison, there were six single-copy DNA sequence differences, of which three were confirmed to be Tsk2/+ specific (Table 1). Among these, there is a SNP that proved useful in demarcating the distal end of the *Tsk2* linkage interval (Chr1:46,268,651; Table 1 and Figure 1), as it was outside the linkage interval. This allowed us to eliminate the only other gene expressed at an appreciable level in the broader interval, *Slc39a10*. In addition, the GULP1 intronic SNP was confirmed and another SNP in an intron of *Col5a2* was observed. Both these latter SNPs are deemed unrelated to the phenotype, again because of their low overall expression and the lack of any influence on splicing or expression in the RNA-Seq results (Figure 2a and b; Supplementary Figure S1 online). Most importantly, however, the heterozygous T-to-A transversion in *Col3A1* at Chr1:45,378,353 was observed in the genomic sequence comparison and was identical to the mutation identified by RNA-Seq. There were no additional variants that could be

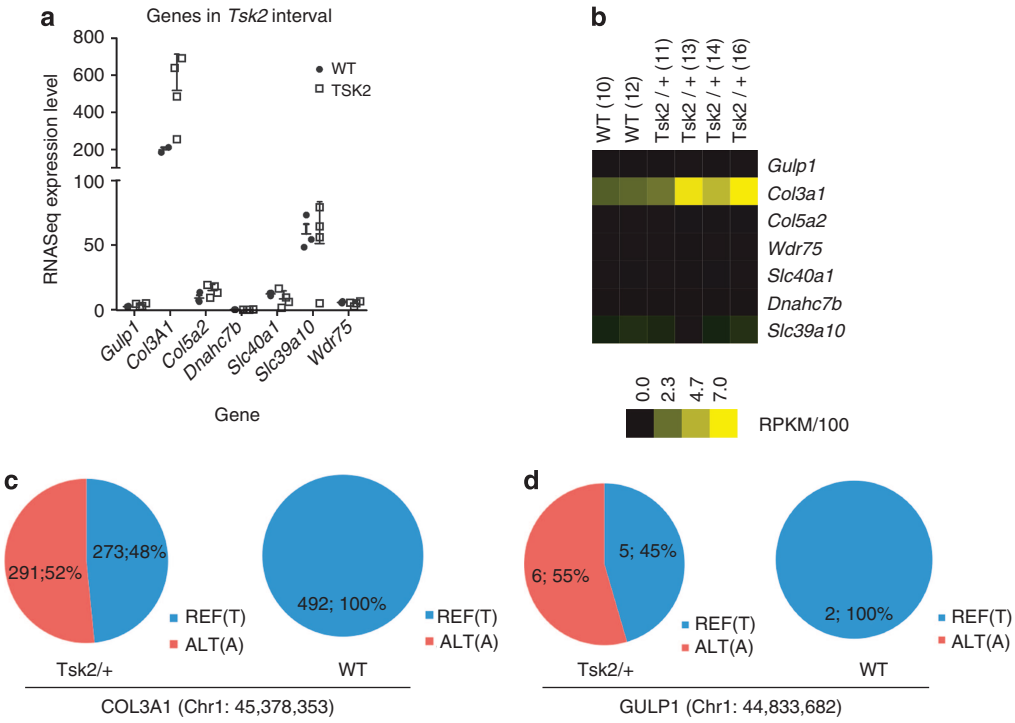


Figure 2. *Col3a1* is the only interval gene expressed at high levels in the skin of Tsk2/+ mice. (**a**) This graph shows gene expression for the seven *Tsk2* interval genes, as determined from the RNA-Seq abundance results. (**b**) Heat map for seven *Tsk2* interval genes detected as transcripts in RNA-Seq. (**c**, **d**) Distribution of nucleotide calls in heterozygous Tsk2/+ and homozygous WT mice for *Col3a1* and *Gulp1*. Col3a1, collagen, type III, alpha 1; RNA-Seq, RNA sequencing; Tsk, tight skin; WT, wild type.

validated on the *Tsk2* chromosome within ~535,000 nucleotides proximal to the transcription start site of *Col3a1* gene or closer than 59,732 nucleotides distal of the end of the *Col3a1* 3′ untranslated region (UTR). Selective resequencing of the 3′UTR likewise revealed no differences between Tsk2 and 101/H (not shown). Thus, this non-synonymous coding mutation is most likely to be *Tsk2* by genomic assessment, as well as by RNA-Seq.

Mice bearing Col3a1^{Tsk2} and Col3a1^{KO} are not viable

To prove that *Tsk2* is a single-nucleotide change in the *Col3A1* coding region required a separate genetic test. Both *Tsk2/Tsk2* (Peters and Ball, 1986) and *Col3a1*-knockout (KO) (Liu *et al.*, 1997) homozygotes exhibit embryonic lethality, which is also seen in our mouse colony (Table 2). We therefore designed a genetic complementation test to determine whether *Col3a1*^{Tsk2} (from Tsk2 mice) could complement and rescue the null allele for *Col3a1*. Conversely, this same cross would determine whether any other gene in the *Col3a1*-homozygous KO could serve to complement the *Tsk2* mutation.

Tsk2/+ × *Col3a1*-/+ mice were bred together, and 37 progeny mice (Table 2) were genotyped. If *Col3a1*^{Tsk2} (C33S) can complement the *Col3a1*-KO, then we would expect to find 9 or 10 *Col3a1*^{Tsk2}/*Col3a1*-KO compound heterozygotes. In fact, no viable compound heterozygotes were born (Table 2, Supplementary Figure S2 online). The hybrid bearing *Tsk2/Col3a1*-null chromosomes was not viable because the *Tsk2* gene on the *Tsk2*-bearing chromosome cannot ''complement'' (rescue) the loss of the *Col3a1* gene on the *Col3a1*-KO chromosome. It bears only the allele of *Col3a1*^{Tsk2} at the *Col3a1* locus, which is insufficient to provide a functional COL3A1 protein that is missing in the *Col3a1*-KO. The *Col3a1*-null chromosome likewise cannot complement the *Tsk2* mutation: the remaining genes on the *Col3a1*-KO chromosome cannot prevent the death of (cannot ''complement'') mice bearing the *Tsk2* chromosome, whereas hybrids carrying *Tsk2/Col3a1*-WT alleles are alive but fibrotic. In fact, having the *Tsk2* mutation is more damaging than not expressing COL3A1 at all, because, although a few *Col3a1*-KO homozygotes make it to birth, *Tsk2/Tsk2* homozygotes (and *Tsk2/Col3a1*-KO) never do, and, whereas *Col3a1/Tsk2* mice are viable but small in stature and fibrotic, *Col3a1* −/+ heterozygotes are normal. Therefore, the mutation in Tsk2/+ mice lies within *Col3a1* and, when homozygous, is substantially more deleterious compared with a complete genetic deficiency of COL3A1.

Col3a1^{Tsk2} induces increased COL1A1 and ECM production *in vitro*

Because the compound heterozygous animals do not survive to accumulate fibrotic levels of ECM, a direct *in vivo* test for fibrosis is impossible; thus, we performed an ''*in vitro* complementation'' test, wherein we transfected mutant or WT *Col3a1* complementary DNA (cDNA) into *Col3a1*-KO fibroblasts, harvested from a *Col3a1*-KO/KO homozygote at birth. Using the production of COL1A1 as a measure of fibrosis

Table 2. Progeny born from *Col3a1*-deficient, *Col3a1*-sufficient, and Tsk2/+ mice

(A) Parents	Genotype and phenotype of progeny		
	Tsk2/+ (tight skin)	WT/WT (normal skin)	Tsk2/Tsk2 (lethal)
Tsk2/+ × Tsk2/+	22	21	0
	$Col3a1^{+}/Col3a1^{-}$ (normal skin)	$Col3a1^{+}/Col3a1^{+}$ (normal skin)	$Col3a1^{-}/Col3a1^{-}$ (moribund)
$Col3a1^{-/+} \times Col3a1^{-/+}$	16	13	3

(B) Parents	Genotype and phenotype of progeny			
	$WT/Col3a1^{+}$ (normal skin)	$Tsk2/Col3a1^{+}$ (tight skin)	$WT/Col3a1^{-}$ (normal skin)	$Tsk2/Col3a1^{-}$
$Tsk2/+ \times Col3a1^{-/+}$	12	10	15	0

Abbreviations: Col3a1, collagen, type III, alpha 1; SNP, single-nucleotide polymorphism; Tsk, tight skin; WT, wild type.
All progenies were assessed for chromosome 1 markers (SNPs and microsatellites) that characterize the origin of the tested allele (*Tsk2* or *Col3a1*).
(A, top) shows the number of mice born of each genotype and phenotype from Tsk2/+ × Tsk2/+ or $Col3a1^{-/+} \times Col3a1^{-/+}$ parents.
(B, bottom) shows the number of mice born of each genotype and phenotype from Tsk2/+ × $Col3a1^{-/+}$ parents; note: there are no compound heterozygotes (*Tsk2/Col3a1*−) born from this mating.

(shown to be expressed at high levels in Tsk2/+ skin and used as a marker of fibrosis (Barisic-Dujmovic *et al.*, 2008; Christner *et al.*, 1998)), we assessed both protein and mRNA levels in fibroblasts that received DNA from a plasmid containing a single allele of a single *Col3a1* gene. In three independent experiments, COL1A1 protein was significantly elevated after 48 hours of transfection with $Col3a1^{Tsk2}$ relative to transfection with $Col3a1^{WT}$ (Figure 3a); mRNA for *Col1a1* was likewise increased in cells transfected with mutant $Col3a1^{Tsk2}$ cDNA (Figure 3b). Transfection efficiencies were equal in each of the experiments (Figure 3c).

Given the observation that the production of a major indicator of fibrosis, COL1A1, is increased by the transfection of the $Col3a1^{Tsk2}$ gene, we assessed the impact of the mutant gene genome-wide. RNA from the $Col3a1^{Tsk2}$ and $Col3a1^{WT}$ transfected *Col3a1*-KO fibroblasts and from 4-week-old Tsk2/+ and WT littermate skin was analyzed by DNA microarray. Differentially expressed pathways between the two transfections were determined by Gene Set Enrichment Analysis (GSEA). Transfection of $Col3a1^{Tsk2}$ results in significant enrichment of genes associated with fibrotic Gene Ontology terms including basement membrane, extracellular matrix, integrin binding, and transmembrane receptor protein kinase activity (Figure 3d; GSEA FDR < 5%). The biological processes observed in the skin of four 4-week-old female Tsk2/+ mice relative to WT littermates also show increases in genes associated with Gene Ontology terms extracellular matrix, integrin binding, and basal lamina (ZL, CB, KBL, CMA, EPB, MLW, manuscript in preparation). The genes that significantly contributed to the GSEA pathway enrichment in the transfected fibroblasts were extracted from microarray data of the transfections, as well as from female Tsk2/+ and WT skin at 4 weeks of age (Figure 3e and f), and were elevated both in the fibroblasts transfected with $Col3a1^{Tsk2}$ and in Tsk2/+ mouse skin. These include those genes typically associated with fibrosis including *CTGF*, *THY1*, *FBN1*, the collagens, laminins, *TGFB1*, *TGFBR1*, *ADAMTS* family genes, and *MMP11*. In addition, there was upregulation in $Col3a1^{Tsk2}$-transfected fibroblasts and Tsk2/+ skin RNA of the vascular endothelial growth factor receptors *FLT1* and *FLT4*, as well as genes associated with platelet-derived growth factor signaling (PDGFRB and PDGFRL; Figure 3f). These data indicate that expression of the $Col3a1^{Tsk2}$ gene alone can induce a substantial fibrotic gene expression program.

Taken together, this means that *Col3a1* and *Tsk2* are almost certainly one and the same gene. $Col3a1^{Tsk2}$ (C33S) is therefore deemed a deleterious gain-of-function allele of *Col3a1*, and the *Col3a1*-KO is a classical loss-of-function allele. Mice thus need at least one copy of a functional, normal *Col3a1* gene.

Tsk2/+ mice have increased dermal COL3A1 protein accumulation

The behavior of *Col3a1* in Tsk2/+ mice could reveal the mechanism by which this mutation causes very substantial ECM fibrosis and very tight skin. We measured the level of COL3A1 protein by histological examinations of Tsk2/+ and WT littermate skin. Reticular fibers are composed primarily of COL3A1 and are a structural element in the skin, found in the panniculus carnosus and in the dermis. COL3A1 expression in the skin from 2-week-old mice is high and declines after birth in WT littermates but does not decline in the Tsk2/+ mice (Figure 4a). As Tsk2/+ mice age, the reticular fibers thicken and become more pronounced compared with their WT littermates reflecting the accumulation of COL3A1. This finding was confirmed in the skin from 4-week-old mice by western blots, which revealed that there is significantly more COL3A1 in the skin of Tsk2/+ mice compared with age- and sex-matched WT littermates (Figure 4b and c). We propose that the excess COL3A1 protein we observe by several measures in Tsk2/+ mice is due to a trend for excess production of *Col3a1* mRNA (Figure 2a) rather than reduced degradation of the Col3 protein. Because the PIIINP fragment is removed from the majority of Col3 molecules before natural Col3 turnover degradation takes place in the tissue, mature COL3A1 from Tsk2 is identical to mature COL3A1 from WT mice, and its natural degradation is unlikely to be affected by any changes in PIIINP. These data show that there is an overall

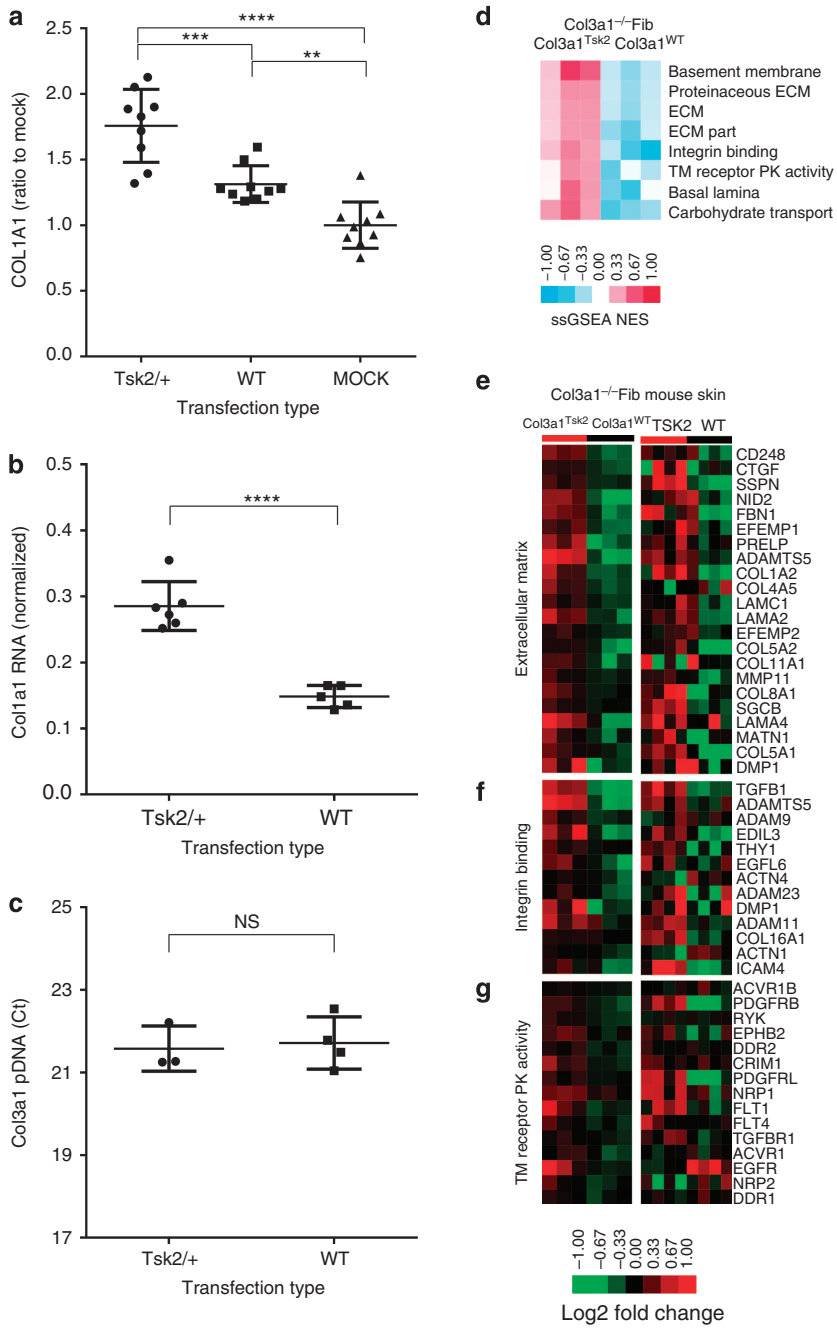


Figure 3. Mouse Col3a1-KO fibroblasts transfected with mutant *Col3a1*^{Tsk2} express a more fibrotic protein profile compared with *Col3a1*^{WT} transfectants. (**a**) Culture supernatants assayed by western blot for COL1A1. *Col3a1*^{Tsk2} transfectants produced 34% more COL1A1 compared with *Col3a1*^{WT} ($P<0.001$) or mock transfectants ($P<0.0001$). (**b**) *Col1a1* mRNA is more highly expressed in *Col3a1*-KO fibroblasts transfected with *Col3a1*^{Tsk2} than with *Col3a1*^{WT} ($P<0.0001$). (**c**) There was no significant difference in efficiency of plasmid transfection between *Col3a1*^{Tsk2} and *Col3a1*^{WT}. (**d**) *Col3a1*$^{-/-}$ fibroblasts transfected with *Col3a1*^{Tsk2} show a significant increase in Gene Ontology terms associated with fibrosis. (**e**) Expression of the genes that contributed most to the ECM enrichment results in *Col3a1*^{Tsk2} versus *Col3a1*^{WT}-transfected mouse fibroblasts or in 4-week-old female Tsk2/+ versus WT mice. (**f**) Expression of genes that contributed to integrin binding term. (**g**) Expression of genes that contributed to transmembrane receptor protein kinase activity term. Col3a1, collagen, type III, alpha 1; ECM, extracellular matrix; KO, knockout; NS, not significant; pDNA, plasmid DNA; Tsk, tight skin; WT, wild type.

increased accumulation of mature COL3A1 protein in the Tsk2/+ mice; in addition, at least half of the type III procollagen and PIIINP trimers produced likely contain one or more strands bearing the *Tsk2* (C33S) mutation.

DISCUSSION

Sequencing of both expressed RNAs and the genomic region in the Tsk2/+ interval, coupled with the genetic complementation study, prove that Tsk2/+ mice harbor a deleterious

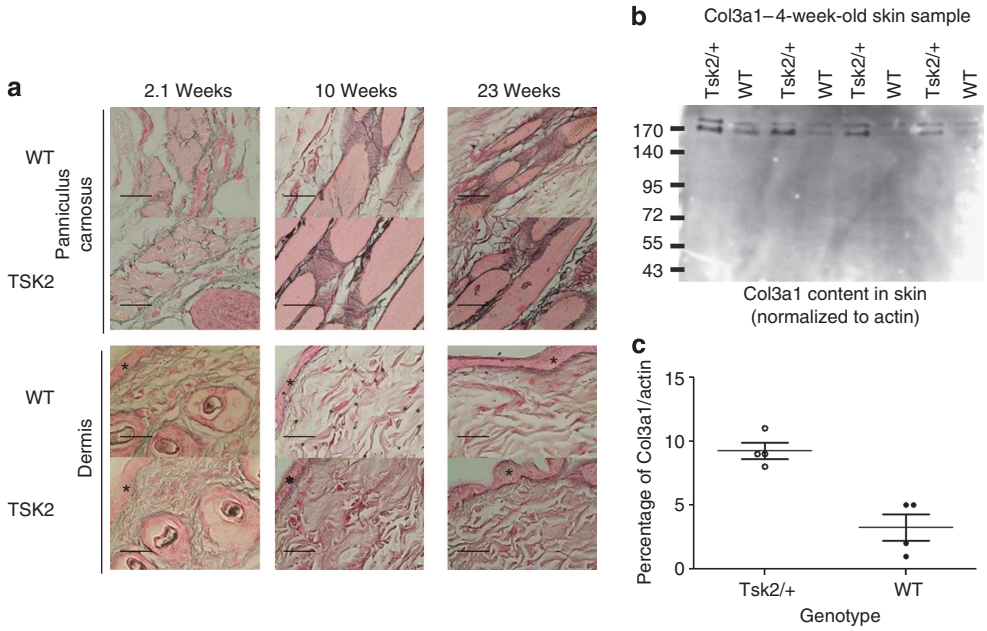


Figure 4. Tsk2/+ mice have increased reticular fiber accumulation and COL3A1 in the skin compared with WT littermates. (**a**) Reticular fiber staining was performed on mice of the indicated ages (2–23 weeks). Stars mark the location of the epidermis. COL3A1 fibers (black staining) are much thicker and more abundant at each life stage in Tsk2/+ than in WT. Fibers were found to be especially pronounced in the panniculus carnosus region of the tissue; increased staining of COL3A1 in the dermis was also noted. The dermal reticular fibers are composed entirely of COL3A1 protein, as this protein is receptive to silver impregnation, and they are increased in Tsk2/+ mice. All images were taken at 200× magnification. Bar size = 100 μm. (**b**, **c**) Skin lysates were analyzed for COL3A1 content (both bands) relative to β-actin (not shown) by western blot analysis. Tsk2/+ mouse skin has significantly more COL3A1 protein than WT mouse skin ($P = 0.0025$, ANOVA). ANOVA, analysis of variance; Col3a1, collagen, type III, alpha 1; Tsk, tight skin; WT, wild type.

coding mutation in *Col3a1*, leading to an amino acid change (C33S) in the N-terminal region of the protein (PIIINP). This point mutation is consistent with those expected from ethylnitrosourea-induced mutagenesis, which generates random single-base-pair point mutations by direct alkylation of nucleic acids. The most common mutations are AT-to-TA and AT-to-GC changes (Noveroske *et al.*, 2000; Cordes, 2005); all three Tsk2-specific mutations identified here were T-to-A or T-to-C mutations. The Tsk2/+ allele is expressed in a 1:1 ratio with the WT by RNA-Seq indicating equal transcription and making a duplication event unlikely.

Effects of the *Tsk2* mutation include the following: (1) accumulation of COL3A1 protein *in vivo* over time; (2) induction and accumulation of COL1A1 protein *in vivo* and in *in vitro* expression models; (3) a more lethal phenotype compared with the homozygous genetic loss of *Col3a1*; and (4) a more lethal compound heterozygous phenotype compared with that of the homozygous gene KO. The latter two characteristics indicate that COL3A1^{Tsk2} (C33S) has a dominant prenatal lethal effect, although our *in vitro* complementation results suggest that the presence of COL3A1-C33S (or its mRNA) is not lethal to skin fibroblasts *per se*. A major function of the *Col3a1* gene is promoting blood vessel development (Liu *et al.*, 1997), which likely led to the lethality observed in the complementation experiment. In the *Col3a1*-KO, a few mice are born with the homozygous deficiency, and these mice die of rupture of the major blood vessels (Liu *et al.*, 1997). The possibility that *Col3a1*^{Tsk2} mutation could directly induce a deleterious vascular phenotype in Tsk2/+ mice is intriguing; it is notable that genes encoding vascular features (*Flt1* and *Flt4*, genes for vascular endothelial growth factor receptors) are significantly upregulated in both *Col3a1*^{Tsk2}-transfected skin fibroblasts and in Tsk2/+ skin relative to WT (Figure 3g). It is possible that a complete *Col3a1* deficiency could be compensated by other collagens, but the *Col3a1*^{Tsk2} mutation is a deleterious gain-of-function, and the deposition of COL3A1-C33S may actively prevent other more benign collagen alternatives from functioning in the vasculature. Thus, our theory is that two doses of a damaging protein are worse than no expression of a normal one.

To our knowledge, this is the first mutation in *Col3a1* that results in a gain-of-function phenotype instead of Ehlers–Danlos-like syndromes that are due to loss-of-function or antimorphic collagen-poor phenotypes. Ehlers–Danlos is a group of connective tissue disorders characterized by highly elastic, fragile but not fibrotic skin due to a defect in collagen synthesis (Nishiyama *et al.*, 2001). In addition, these patients have a significant risk for aneurism. The Ehlers–Danlos syndrome has been associated with 337 mutations in COL3A1 (http://www.le.ac.uk/ge/collagen/), as well as mutations on COL1A1 and COL5A2. These mutations result in amino acid substitutions in the C terminus of the protein, RNA splicing alterations, deletions, or null alleles. Interestingly, in the Ehlers–Danlos syndrome type IV (a very different disease than that observed in Tsk2/+ mice), studies have shown that patients bearing a mutated COL3A1 (compared with a null COL3A1) develop more severe disease and succumb to disease prematurely, whereas those with null COL3A1 were

able to live a relatively normal life with limited disease (Leistritz *et al.*, 2011). Currently, all reported COL3A1 mutations result in decreased collagen protein secretion leading to variably thinner skin and defects in the vasculature that are observed in these patients. In contrast to the mutations observed in Ehlers–Danlos, the Tsk2/+ mouse mutation results in thickened skin with no apparent evidence of aneurism. The mutation reported here occurs in the N-terminal PIIINP fragment of the protein, rather than the C-terminal region associated with Ehlers–Danlos.

The PIIINP molecule is a homotrimer with a molecular weight of ~42,000 daltons and comprises three domains: a cysteine-rich globular domain (Col 1) containing 79 amino acids with five intrachain disulfide bonds, a triple-helical domain (Col 3) with 12 amino acids and three interchain disulfide bonds, and a non-collagenous domain (Col 2) comprising 39 amino acids ending with the N-telopeptide that forms a triple helical structure (Bruckner *et al.*, 1978). The mutation in *Col3a1*^{Tsk2} substitutes a serine for the cysteine in one of the five Col 1-domain cysteines involved in disulfide bonds (Bruckner *et al.*, 1978).

Features shared by Tsk2/+ mice and people with fibrotic diseases (scleroderma, liver fibrosis, and kidney fibrosis) include the dysregulation of PIIINP (Sondergaard *et al.*, 1997; Majewski *et al.*, 1999; Abignano and Del Galdo, 2014; Del Galdo and Matucci-Cerinic, 2014; Quillinan *et al.*, 2014). The PIIINP fragment is a clinically validated biomarker of liver fibrosis (Leroy *et al.*, 2004; Rosenberg *et al.*, 2004) and scleroderma (Sondergaard *et al.*, 1997; Majewski *et al.*, 1999), and it has been used as a surrogate marker of fibrosis in clinical trials of potential SSc therapies (Majewski *et al.*, 1999; Denton *et al.*, 2009). Our finding of a point mutation in the protein that likely has a deleterious effect on PIIINP function is consistent with these clinical results and the fibrotic phenotype in the Tsk2/+ mouse.

Its high level in the sera of such patients may not merely be a benign biomarker. Support for this hypothesis derives from our *in vitro* complementation results showing that the presence of COL3A1-C33S is sufficient to upregulate the synthesis and secretion of COL1A1, consistent with the increased activity of the *Col1a1* promoter and excess production of COL1A1 in Tsk2/+ mice (Christner *et al.*, 1998; Barisic-Dujmovic *et al.*, 2008). It is likely that higher levels of or altered COL3A1 protein or PIIINP fragment also directly influence the composition and size of COL1A1/A2- and COL3A1-containing fibers, and that these features indirectly upregulate transforming growth factor-β1 signaling, an important mediator of collagen production. A previous report from our laboratory has demonstrated increased dermal elastic fibers and transforming growth factor-β1 accumulation in the skin of Tsk2/+ mice beginning at 2 weeks of age, lending further support to our hypothesis (Long *et al.*, 2014). In addition, our gene expression analyses show that similar global impact of the *Col3a1*^{Tsk2} gene occurs both *in vitro* and *in vivo*, and in both settings there are fundamental changes in the ECM and in fibroblasts due to the presence of this mutation. The hypothesis that *Col3a1*^{Tsk2} (or PIIINP^{Tsk2}) directly causes dermal fibrosis and scleroderma-like characteristics is attractive: it would likely be dominant within the heterozygote, as collagen III is a homotrimeric triple helix (Ramachandran **a**nd Kartha, 1955), and the gene product of the mutant chromosome could be expected to contribute to alteration of a majority of collagen helices even in the presence of 50% normal collagen (Strachan and Read, 1999).

MATERIALS AND METHODS

All studies and procedures were approved by the Institutional Animal Care and Use Committee at Drexel University College of Medicine and conducted in accord with recommendations in the ''Guide for the Care and Use of Laboratory Animals'' (Institute of Laboratory Animal Resources, National Research Council, National Academy of Sciences). Detailed methods are provided in the Supplementary Materials online.

Animals

Tsk2/+ mice were serially backcrossed to the C57Bl/6J (B6) background. Recombinant B6.Tsk2/+ mice were also bred to B6.chr 1-A/J mice (Jackson Laboratory, Bar Harbor, ME) and the resulting B6.Tsk2/+ F1 mice were backcrossed to B6.chr 1-A/J mice. Wild-type littermates were used as controls.

DNA isolation from tail snips, microsatellite, and SNP typing

These were performed as in our previous publications (Bunce *et al.*, 1995; Butterfield *et al.*, 1998). Specific locations of SNP polymorphisms between B6 (which is very similar to 101/H) and A/J were determined using Mouse Genome Informatics (www.informatics.jax.org) and Mouse Phenome Database (http://phenome.jax.org/).

Complementation analysis

Tsk2/+ mice were crossed to Col3a1 −/+ mice and their progeny mated to verify that the SNP in *Col3a1* is Tsk2. The resulting generations of the cross were genotyped by PCR for Tsk2/+ using microsatellites and primers specific to Col3a1 or the inserted neomycin cassette (see Supplementary Material online).

In vitro assessment of fibrogenesis by COL3A1^{Tsk2}

We constructed a plasmid harboring the *Col3a1*^{Tsk2} allele by introducing the Tsk2 T-to-A mutation into a wild-type Col3a1 clone (pCMV6-Kan/Neo; OriGene, Rockville, MD). A *Col3a1*-KO line was transfected with either plasmid as described (Artlett *et al.*, 1998). Supernatants were retained and cell lysates were harvested directly from the dish at 48 hours.

RNA isolation and real-time PCR

RNA was isolated from the skin or fibroblasts using a RNA isolation kit from Clontech (Mountain View, CA), and cDNAs synthesized from 2.0 μg of total RNA using an High Capacity cDNA Reverse Transcription kit (Applied Biosystems, Foster City, CA). Relative quantification of all products was measured using SYBR Green chemistry (Applied Biosystems).

RNA sequencing

Total RNA was prepared from three WT and four Tsk2/+ mice skin biopsies using the Qiagen RNeasy Fibrous Tissue Mini Kit (Qiagen Sciences, Germantown, MD). RNA-seq sequencing libraries were prepared for the seven samples using a NuGEN Ovation RNA-Seq

System (NuGEN Technologies, San Carlos, CA). Libraries were multiplexed and sequenced on an Illumina HiSeq 2000 platform to obtain 16.7–50.9 million 50 bp paired-end reads per sample. The raw reads were aligned to the reference mouse genome (MM9 assembly) using Tophat software with default parameters (Trapnell *et al.*, 2009; Trapnell *et al.*, 2012). Supplementary Figure S1 online shows RNA-Seq read coverage for three interval genes. RNA-seq data from this study are available from NCBI Bioproject at accession number PRJNA262679.

454 Sequencing

Samples were captured and amplified as described in the Roche Nimblegen sequence capture manual (version 1.0; Madison, WI). Titanium general libraries were prepared from the captured DNAs from two 101/H mice and two Tsk2/+ mice using 5,000 ng of DNA. Enriched captured fragments were sequenced as described in GS FLX Titanium emPCR and Sequencing Protocols, October 2008. Sequence capture array probes were designed by Roche Nimblegen using the mouse genome sequence between 44,241,286 and 47,116,890 on chromosome 1 of mouse genome (MM9). Multiplexed 454 sequenced reads were assembled using Newbler v2.6 (454 Life Sciences, Branford, CT) with scaffolding against the same chromosome region that the probes were derived from.

DNA microarray hybridization and data analysis

This was performed as in our previous publications (Pendergrass *et al.*, 2012). RNA samples were amplified and labeled using the Agilent Low Input Linear Amplification kit (Agilent Technologies, Santa Clara, CA) and were hybridized against Universal Mouse Reference (Strategene, La Jolla, CA) to Agilent Whole Mouse Genome arrays (G4122F; Agilent Technologies) in a common reference-based design. Microarrays were hybridized and washed in accordance with the manufacturer's protocols and scanned using a dual laser GenePix 4000B scanner (Axon Instruments, Foster City, CA). The pixel intensities of the acquired images were then quantified using GenePix Pro 5.1 software (Axon Instruments). Raw microarray data from this study are available from NCBI GEO at accession number GSE61728.

Western blot analyses

Culture supernatants were collected or the skin was homogenized in RIPA buffer (Sigma-Aldrich, St Louis, MO) using a glass homogenizer. Total protein was measured with a Bradford assay (Sigma-Aldrich), and western blots were performed as in our publications (Sassi-Gaha *et al.*, 2010). Antibodies used included goat anti-COL3A1 (#sc-8781), goat anti-COL1A1 (#sc-28657) from Santa Cruz Biotechnology, Santa Cruz, CA, rabbit anti-β-Actin (#4967, Cell Signaling Technologies, Boston, MA), donkey anti-goat (#705-035-003, Jackson ImmunoResearch Laboratories, West Grove, PA), or goat anti-rabbit (#111-035-003, Jackson ImmunoResearch), and signals were developed using SuperSignal West Dura ECL reagent (Thermo Scientific, Rockford, IL). Band intensities were measured using ImageQuant TL Software (GE Healthcare Life Sciences, Pittsburgh, PA).

Reticular fiber staining

Reticular fibers were stained using the Chandler's Precision Reticular Fiber Stain kit (American Master*Tech, Lodi, CA) according to the manufacturer's protocol.

Statistics

A two-tailed Student *t*-test or a one-way analysis of variance was used to determine statistical significance of collagen protein expression, as noted.

CONFLICT OF INTEREST

The authors state no conflict of interest.

ACKNOWLEDGMENTS

We thank Paul Christner for providing the breeding pairs of the original Tsk2/+ mice and Xianhua Piao at Harvard University for the *Col3a1*-KO mice. This work was supported by a Scleroderma Foundation Grant and awards from the National Institutes of Health (AR061384) and the Department of Defense (PR100338).

Author contributions

KBL and CMB bred and genotyped the B6.Tsk2 mice and all the derivative animals in this report; KBL, CMA, CMB, and SS-G conducted the histology on the skin and transfections on fibroblasts; EPB was responsible for the design and interpretation of the research including the genetic analyses; ZL and MLW conducted the expression analyses and interpreted the results; VM conducted GSEA analysis, SGC constructed the plasmids containing the mutant Col3a1 cDNA; GDE, JE, RE, and AA performed the genomic DNA capture and sequencing and interpreted these results; KBL, EPB, CMA, and MLW wrote the paper.

SUPPLEMENTARY MATERIAL

Supplementary Material is linked to the online version of the paper at http://www.nature.com/jid

REFERENCES

Abignano G, Del Galdo F (2014) Quantitating skin fibrosis: innovative strategies and their clinical implications. *Curr Rheumatol Rep* 16:404

Artlett CM, Chen SJ, Varga J *et al.* (1998) Modulation of basal expression of the human alpha1(I) procollagen gene (COL1A1) by tandem NF-1/Sp1 promoter elements in normal human dermal fibroblasts. *Matrix Biol* 17:425–34

Artlett CM (2010) Animal models of scleroderma: fresh insights. *Curr Opin Rheumatol* 22:677–82

Asano Y, Stawski L, Hant F *et al.* (2010) Endothelial Fli1 deficiency impairs vascular homeostasis: a role in scleroderma vasculopathy. *Am J Pathol* 176:1983–98

Barisic-Dujmovic T, Boban I, Clark SH (2008) Regulation of collagen gene expression in the Tsk2 mouse. *J Cell Physiol* 215:464–71

Bruckner P, Bachinger HP, Timpl R *et al.* (1978) Three conformationally distinct domains in the amino-terminal segment of type III procollagen and its rapid triple helix leads to and comes from coil transition. *Eur J Biochem* 90:595–603

Bunce M, O'Neill CM, Barnardo MC *et al.* (1995) Phototyping: comprehensive DNA typing for HLA-A, B, C, DRB1, DRB3, DRB4, DRB5 & DQB1 by PCR with 144 primer mixes utilizing sequence-specific primers (PCR-SSP). *Tissue Antigens* 46:355–67

Butterfield RJ, Sudweeks JD, Blankenhorn EP *et al.* (1998) New genetic loci that control susceptibility and symptoms of experimental allergic encephalomyelitis in inbred mice. *J Immunol* 161:1860–7

Christner PJ, Hitraya EG, Peters J *et al.* (1998) Transcriptional activation of the alpha1(I) procollagen gene and up-regulation of alpha1(I) and alpha1(III) procollagen messenger RNA in dermal fibroblasts from tight skin 2 mice. *Arthritis Rheum* 41:2132–42

Christner PJ, Peters J, Hawkins D *et al.* (1995) The tight skin 2 mouse. An animal model of scleroderma displaying cutaneous fibrosis and mononuclear cell infiltration. *Arthritis Rheum* 38:1791–8

Christner PJ, Siracusa LD, Hawkins DF *et al.* (1996) A high-resolution linkage map of the tight skin 2 (Tsk2) locus: a mouse model for scleroderma (SSc) and other cutaneous fibrotic diseases. *Mamm Genome* 7:610–2

Cordes SP (2005) N-ethyl-N-nitrosourea mutagenesis: boarding the mouse mutant express. *Microbiol Mol Biol Rev* 69:426–39

Del Galdo F, Matucci-Cerinic M (2014) The search for the perfect animal model discloses the importance of biological targets for the treatment of systemic sclerosis. *Ann Rheum Dis* 73:635–6

Denton CP, Engelhart M, Tvede N *et al.* (2009) An open-label pilot study of infliximab therapy in diffuse cutaneous systemic sclerosis. *Ann Rheum Dis* 68:1433–9

Gentiletti J, McCloskey LJ, Artlett CM *et al.* (2005) Demonstration of autoimmunity in the tight skin-2 mouse: a model for scleroderma. *J Immunol* 175:2418–26

Hocher B, Schwarz A, Fagan KA *et al.* (2000) Pulmonary fibrosis and chronic lung inflammation in ET-1 transgenic mice. *Am J Resp Cell Mol Biol* 23:19–26

Leistritz DF, Pepin MG, Schwarze U *et al.* (2011) COL3A1 haploinsufficiency results in a variety of Ehlers-Danlos syndrome type IV with delayed onset of complications and longer life expectancy. *Genet Med* 13:717–22

Leroy V, Monier F, Bottari S *et al.* (2004) Circulating matrix metalloproteinases 1, 2, 9 and their inhibitors TIMP-1 and TIMP-2 as serum markers of liver fibrosis in patients with chronic hepatitis C: comparison with PIIINP and hyaluronic acid. *Am J Gastroenterol* 99:271–9

Liu X, Wu H, Byrne M *et al.* (1997) Type III collagen is crucial for collagen I fibrillogenesis and for normal cardiovascular development. *Proc Natl Acad Sci USA* 94:1852–6

Long KB, Artlett CM, Blankenhorn EP (2014) Tight skin 2 mice exhibit a novel time line of events leading to increased extracellular matrix deposition and dermal fibrosis. *Matrix Biol* 38:91–100

Majewski S, Wojas-Pelc A, Malejczyk M *et al.* (1999) Serum levels of soluble TNF alpha receptor type I and the severity of systemic sclerosis. *Acta Derm Venereol* 79:207–10

Maurer B, Busch N, Jungel A *et al.* (2009) Transcription factor fos-related antigen-2 induces progressive peripheral vasculopathy in mice closely resembling human systemic sclerosis. *Circulation* 120:2367–76

Nishiyama Y, Nejima J, Watanabe A *et al.* (2001) Ehlers-Danlos syndrome type IV with a unique point mutation in COL3A1 and familial phenotype of myocardial infarction without organic coronary stenosis. *J Intern Med* 249:103–8

Noveroske JK, Weber JS, Justice MJ (2000) The mutagenic action of N-ethyl-N-nitrosourea in the mouse. *Mamm Genome* 11:478–83

Pendergrass SA, Lemaire R, Francis IP *et al.* (2012) Intrinsic gene expression subsets of diffuse cutaneous systemic sclerosis are stable in serial skin biopsies. *J Invest Dermatol* 132:1363–73

Peters J, Ball ST (1986) Tight Skin 2 (Tsk2). *Mouse News Letters* 74:91–2

Quillinan NP, McIntosh D, Vernes J *et al.* (2014) Treatment of diffuse systemic sclerosis with hyperimmune caprine serum (AIMSPRO): a phase II double-blind placebo-controlled trial. *Ann Rheum Dis* 73:56–61

Ramachandran GN, Kartha G (1955) Structure of collagen. *Nature* 176:593–5

Richard V, Solans V, Favre J *et al.* (2008) Role of endogenous endothelin in endothelial dysfunction in murine model of systemic sclerosis: tight skin mice 1. *Fundam Clin Pharmacol* 22:649–55

Rosenberg WM, Voelker M, Thiel R *et al.* (2004) Serum markers detect the presence of liver fibrosis: a cohort study. *Gastroenterology* 127:1704–13

Sassi-Gaha S, Loughlin DT, Kappler F *et al.* (2010) Two dicarbonyl compounds, 3-deoxyglucosone and methylglyoxal, differentially modulate dermal fibroblasts. *Matrix Biol* 29:127–34

Siracusa LD, McGrath R, Ma Q *et al.* (1996) A tandem duplication within the fibrillin 1 gene is associated with the mouse tight skin mutation. *Genome Res* 6:300–13

Sondergaard K, Heickendorff L, Risteli L *et al.* (1997) Increased levels of type I and III collagen and hyaluronan in scleroderma skin. *Br J Dermatol* 136:47–53

Strachan T, Read AP (1999) *Human Molecular Genetics*. 2nd edn. Wiley-Liss: New York

Trapnell C, Pachter L, Salzberg SL *et al.* (2009) TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 25:1105–11

Trapnell C, Roberts A, Goff L *et al.* (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc* 7:562–78

Yamamoto T, Takagawa S, Katayama I *et al.* (1999) Animal model of sclerotic skin. I: local injections of bleomycin induce sclerotic skin mimicking scleroderma. *J Invest Dermatol* 112:456–62

Interspecies Comparative Genomics Identifies Optimal Mouse Models of Systemic Sclerosis

Journal:	*Arthritis & Rheumatology*
Manuscript ID	ar-14-1579.R1
Wiley - Manuscript type:	Full Length
Date Submitted by the Author:	n/a
Complete List of Authors:	Sargent, Jennifer; Geisel School of Medicine at Dartmouth, Genetics Li, Zhenghui; Geisel School of Medicine at Dartmouth, Genetics Aliprantis, Antonios; Brigham and Women's Hospital, Medicine, Division of Rheumatology Greenblatt, Matthew; Brigham and Women's Hospital, Pathology Lemaire, Raphael; Boston University School of Medicine, Rheumatology Section Wu, Minghua; Northwestern University Feinberg School of Medicine, Division of Rheumatology Wei, Jun; Northwestern University Feinberg School of Medicine, Division of Rheumatology Taroni, Jaclyn; Geisel School of Medicine at Dartmouth, Department of Genetics Harris, Adam; University of Connecticut Health Center, Genetics Long, Kristen; Drexel University College of Medicine, Microbiology and Immunology Burgwin, Chelsea; Drexel University College of Medicine, Microbiology and Immunology Artlett, Carol M.; Drexel University College of Medicine, Microbiology & Immunology Blankenhorn, Elizabeth; Drexel University College of Medicine, Microbiology and Immunology Lafyatis, Robert; Boston University School of Medicine, Rheumatology Section Varga, John; Northwestern University Feinberg School of Medicine, Division of Rheumatology Clark, Stephen; University of Connecticut Health Center, Genetics Whitfield, Michael; Geisel School of Medicine at Dartmouth, Genetics
Keywords:	Systemic sclerosis, Skin, human intrinsic gene expression subsets, murine models, interspecies comparison
<B>Disease Category: Please select the category from the list below that best describes the content of your manuscript.:	Systemic Sclerosis

Title: Interspecies Comparative Genomics Identifies Optimal Mouse Models of Systemic Sclerosis

Authors: Jennifer L. Sargent[1], Zhenghui Li[1], Antonios O. Aliprantis[2], Matthew Greenblatt[3], Raphael Lemaire[4], Ming-hua Wu[5], Jun Wei[5], Jaclyn Taroni[1], Adam Harris[6], Kristen Long[7], Chelsea Burgwin[7], Carol M. Artlett[7], Elizabeth P. Blankenhorn[7], Robert Lafyatis[4], John Varga[5], Stephen H. Clark[6], Michael L. Whitfield[1,§]

Affiliations:
[1] Department of Genetics, Geisel School of Medicine at Dartmouth, Hanover, NH 03755 USA
[2] Department of Medicine, Division of Rheumatology, Allergy and Immunology, [3] Department of Pathology, Brigham and Women's Hospital, Smith Building Rm 650A, 1 Jimmy Fund Way, Boston, MA 02115
[4] Rheumatology Section, Boston University School of Medicine, Boston, MA, 02115, USA
[5] Feinberg School of Medicine, Northwestern University, Division of Rheumatology, Chicago, IL, 60611, USA
[6] Department of Genetics and Developmental Biology, University of Connecticut Health Center, Farmington, CT, 06030, USA
[7] Department of Microbiology and Immunology, Drexel University College of Medicine, Philadelphia, PA, 19129, USA

Running Title: Animal models of SSc molecular subsets

[§]To whom correspondence should be addressed:
Michael L. Whitfield, Ph.D.
Department of Genetics
Dartmouth Medical School
7400 Remsen
Hanover, NH 03755
michael.whitfield@dartmouth.edu
phone 603.650.1109, fax 603.650.1188

Funding Sources
This work has been supported by grants from: the Scleroderma Research Foundation (SRF www.srfcure.org) to MLW, AOA, and JV; National Institutes of Health (NIH; www.nih.gov) grants R01AR061384-01 to MLW, EPB, CA, NIH 2R01AR051089 and P30AR061271 to RL; and PR100338 from the Department of Defense (DoD; www.defense.gov; MLW, EPB, CA). Funding agencies have had no role in the interpretation of the data or in writing the manuscript. The authors report no conflict of interest with regards to this work.

Abstract: (limit 250 words)

Objective

Understanding the pathogenesis of systemic sclerosis (SSc) is confounded by considerable disease heterogeneity. Animal models of SSc that recapitulate distinct subsets of disease at the molecular level have not been delineated. We applied interspecies comparative analysis of genomic data from multiple mouse models of SSc and patients with SSc to determine which animal models best reflect the SSc intrinsic molecular subsets.

Methods

Gene expression measured in skin from mice with sclerodermatous graft-versus-host disease (sclGVHD), bleomycin-induced fibrosis, Tsk1/+ or Tsk2/+ mice was mapped to human orthologs and compared to SSc skin biopsy-derived gene expression. TGFβ activation was assessed using a responsive signature in mouse and *Tnfrsf12a* expression measured in SSc and mouse skin.

Results

Gene expression in skin from mice with sclGVHD and bleomycin-induced fibrosis corresponded to that in SSc patients in the *inflammatory molecular* subset. In contrast, Tsk2/+ mice showed gene expression corresponding to the *fibroproliferative* SSc subset. Bleomycin and Tsk2/+ mice showed enrichment of a TGFβ-responsive signature. Expression of *Tnfrsf12a* (the *Tweak-Receptor / Fn14*) is elevated in skin from fibroproliferative SSc patients and skin of Tsk2/+ mice.

Conclusion

This study reveals similarities in cutaneous gene expression between distinct mouse models of SSc and specific molecular subsets of the disease. Different pathways underlie the intrinsic subsets including TGF-β, IL13 and IL4. We identify a novel target, *Tnfrsf12a* that is elevated in skin from *fibroproliferative* patients and Tsk2/+ mice. The information will serve to inform mechanistic and translational pre-clinical studies in SSc.

Keywords: systemic sclerosis; skin; human intrinsic gene expression subsets; murine models; Tsk2/+; Tsk/+; sclGVHD; Bleomycin; fibrosis; DNA microrrays; interspecies comparison; TGF-beta; IL13; Tnfrsf12a (Tweak-Receptor)

A lack of clear genetic associations and the heterogeneity in clinical presentation of systemic sclerosis (SSc; scleroderma) have limited the development of a single agreed-upon animal model (1). We have demonstrated that molecular subsets of SSc exist (2-4) and that different signaling pathways underlie each subsets (5-7). Given the distinct molecular features of SSc subsets, it is possible that one or more mouse models might most accurately represent each. Identification of appropriate models for each subset is necessary to effectively develop therapeutics that specifically target each of these groups of patients. We conducted an integrated interspecies analysis of gene expression in skin of four commonly used mouse models of SSc to identify those models in which skin gene expression and deregulated pathways most accurately reflects that of the molecular subsets of SSc.

Many different signaling pathways have been implicated in SSc (8, 9). TGFβ and PDGF signatures have been demonstrated to underlie the *fibroproliferative* subset of patients (5, 7), whereas IL4 and IL13 signaling is predominantly enriched in the *inflammatory* subset of patients (6). Recent analyses have demonstrated that TGFβ signaling can span the fibroproliferative and inflammatory subsets (7, 10). Here we have directly compared the similarities and differences in the molecular events underlying pathogenesis of SSc, and in mouse models of the disease. To our knowledge this is the first integrative and comparative analysis of genome-wide expression in animal models for a heterogeneous human autoimmune disease.

Materials and Methods

Study Design and human subjects:

Microarray data from human SSc samples were obtained from Milano et al ((2); GEO accession GSE9285). SSc patients met the ACR criteria for systemic sclerosis and included both diffuse and limited systemic sclerosis patients as well as a subset of patients with morphea. We performed biological rather than technical replicates for the different mouse samples. The study was designed to identify the best mouse models to study each of the human SSc subsets. This is a major unmet need since heterogeneity in the disease has confounded basic science studies as well as clinical trials. Our data show that different mouse models represent different subsets of SSc disease.

Mouse models of SSc:

Mouse skin samples were obtained from different laboratories (Table 1). All animal protocols were institutionally approved by the Animal Care and Use Committee at University of Connecticut Health Center (UCONN), Brigham and Women's Hospital (BWH), Boston University School of Medicine (BUMC), Northwestern Feinberg School of Medicine (NW) and Drexel University College of Medicine (DUCM). Each biopsy included the hypodermal layer and fibrosis was confirmed at the site of biopsy in all models. The sclGVHD model was generated at BWH and data have been described previously (6). Depilated Tsk2/+ (11) mouse back skin samples from UCONN were stored in RNAlater (ThermoFisher Scientific, Waltham MA). These mice are heterozygous for the Tsk2 mutation on chromosome 1 and were maintained in an inbred line developed by backcrossing onto C57BL/6J (>N10). Tsk1/+ back skin samples were obtained from six week old mice (12) at BUMC. C57BL/6-^{Fbn}Tsk+/+Pldn^{pa}

mice were obtained from Jackson Laboratories (Bar Harbor, ME) and maintained by breeding with C57BL/6 mice. Tsk1/+ and Tsk2/+ heterozygous mice were identified by assessment of skin tightness over the back and by genotyping as described (12). The bleomycin-induced skin fibrosis model was generated at the NW. The protocol chosen has been used previously to study skin fibrosis (13). In bleomycin-induced fibrosis, six- to eight-week-old female BALB/cJ mice were given daily subcutaneous injections of bleomycin (1 mg/kg) and sacrificed at either 5 days or 21 days. Skin at the site of injection was biopsied. All skin tissue was stored in RNAlater for shipping and storage.

***In vivo* TGFβ-responsive gene signature:**

Total RNA samples from the back skins of mice treated with subcutaneous pumps containing TGFβ were generated at BUMC. Prior to surgery, mice were injected with buprenorphine. For pump insertion, C57BL/6 mice were placed under general anesthesia using isoflourane by inhalation. After complete anesthesia was achieved, the mice were placed on a warm towel, their backs shaved, the surgical area sterilized with betadine and a 1 cm incision made through the skin over the interscapular region. The sterile pumps containing 50, 250 or 1250ng of TGFβ or PBS were inserted into a subcutaneous pocket made by gently teasing apart the fascia layer. Skin was closed with 1-2 staples. Mice were sacrificed 7 days after surgery and the skin from around the pump insertion site harvested for total RNA preparation.

RNA isolation and microarray hybridization:

Total RNA was isolated from mouse skin samples using standard Trizol (Invitrogen, Carlsbad, CA) procedures. Samples were manually homogenized as previously described (2). Samples were further purified using RNA cleanup with RNeasy mini-columns (Qiagen, Valencia, CA). RNA was quantified on a NanoDrop ND-1000 Spectrophotometer (Agilent Technologies, Santa Clara, CA) and integrity assessed by gel electrophoresis.

Samples were amplified and labeled using the Agilent Low Input Linear Amplification kit (Agilent Technologies, Santa Clara, CA) and were hybridized against Universal Mouse Reference (UMR) (Strategene, La Jolla, CA) to Agilent Whole Mouse Genome arrays (G4122F) (Agilent Technologies, Santa Clara, CA) in a common reference based design as previously described (2, 5).

Microarray data processing and analysis:

Human gene expression data from Milano et al (2) were lowess-normalized $\log_2$ Cy5/Cy3 ratios, filtered for intensity/background ratio ≥1.5 in one or both channels and for which at least 80% of the data of sufficient quality were present. This resulted in 28,495 probes for analysis, which were collapsed to 14,276 gene symbols.

Mouse gene expression data were lowess-normalized $\log_2$ Cy5/Cy3 ratios, filtered for intensity/background ratio ≥1.5 in one or both channels and for which at least 80% of the data was of sufficient quality were present. This resulted in 13,773 probes selected from Agilent Mouse Whole Genome microarray. These probes were then collapsed by gene symbol, resulting in 9,776 genes.

Both human and mouse data tables were multiplied by negative one, thereby converting the $\log_2$ Cy5/Cy3 ratios to $\log_2$ Cy3/Cy5 ratios for all analyses.

For interspecies comparisons mouse genes were matched to their human orthologs using the Mouse Genome Database (MGD) at Jackson Laboratories (14). Mouse genes were matched to their human orthologs based on Mouse/Human Orthology with Phenotype Annotations table (MGI 5.22). 62.53% of mouse genes mapped to a human ortholog. This resulted in 6,113 genes. Missing data values were imputed using KNN-nearest neighbors (k=10) function in the bioinformatics toolbox of Matlab R2008b. Systematic bias was corrected using distance weighted discrimination (DWD) in Matlab (15).

Intrinsic genes were selected using an intrinsic gene identifier algorithm (2, 16) was performed (Table 1). 1,217 genes that demonstrated a weight intrinsic score threshold below 0.30 were selected for further analysis (2, 16). Each column was standardized using the root mean square (RMS) for each array, calculated by taking the sum of the square of the expression values for each gene, divided by the number of genes, and taking the square root of this value. The values of the genes on each array were divided by its RMS. All human and mouse data were median centered and clustered using Cluster 3.0 (17) and visualized in TreeView 1.0.13 (17). The p-values for branches on the dendrogram were calculated using R SigClust package with default settings (* in Figure 2 indicates $p < 0.001$). Consensus Clustering was performed using GenePattern ConsensusClustering with 2,000 iterations and K=3,4 and 5, which clearly shows 4 stable clusters (Figure 2; Supplemental Figure S1).

Pathway module map analysis was performed in Genomica (18). Gene expression data for the merged datasets was matched to the appropriate Entrez Gene Identifier corresponding to the human gene annotation. Statistically significantly enriched GO terms were selected ($p < 0.05$;

FDR 0.05; hypergeometric distribution) and clustered according to their enrichment scores. Samples were organized as per the sample clustering of the 1,217 intrinsic genes (Figure 3).

Gene expression data from this study are available from NCBI GEO at accession number GSE71999 (Tsk1/+ Tsk2/+, and bleomycin-induced skin fibrosis). The data for the sclGVHD mouse model ((6); GSE24410) and human SSc skin data ((2); GSE9285) were published previously.

Immunohistochemistry:

To visualize TWEAK-R positive and Ki-67 positive cells, paraffin-embedded sections were deparaffinized with two washes of xylene and two washes of ethanol. Antigens were unmasked with 10 mM citrate buffer at 100 ^{o}C for 25 min, then blocked with 5% goat serum. For mouse biopsies, rabbit anti-TWEAK-R monoclonal antibody at 1:100 (ab109365, Abcam, Cambridge, MA), or rabbit anti-Ki67 polyclonal antibody at 1:250 (ab9260, Millipore, Darmstadt, Germany) in blocking buffer was applied to the sections overnight at 4˚C. Slides were washed three times with PBS and incubated with Cy3-conjugated goat anti-rabbit IgG at 1:200 (TWEAK-R) or at 1:500 (Ki-67, JacksonImmuno-111-165-006, Jackson ImmunoResearch Laboratories, West Grove) for 40 min at RT. Nuclei were visualized with DAPI (Life Technologies, Carlsbad CA). An IgG isotype negative control was used to control for background staining.

Staining of human SSc skin biopsies followed the procedure above, except the mouse anti-TWEAK-R monoclonal antibody was used (1:20; BioLegend, Clone ITEM-4). Sections were washed to remove unbound antibody and then incubated with goat-anti-mouse-Cy3 (1:400; Jackson ImmunoResearch West Grove PA) for 40 min at room temperature. A mouse IgG2b

isotype negative control (ThermoFisher, Waltham MA) was used and analyzed in parallel. Eleven biopsies from Milano et al. (2) were analyzed using intrinsic subset assignments from (7), which were identical to Milano except that dSSc7 forearm, which was assigned to *fibroproliferative*, and dSSc2 forearm, which was assigned to normal-like. The biopsies stained were *fibroproliferative* (dSSc1, dSSc7, dSSc11), *inflammatory* (dSSc6, lSSc7), *limited*(lSSc4, lSSc5) and *normal-like* (dSSc2, Nor2, Nor3, Nor5). Positive cells in the dermis and epidermal skin appendages were counted by a blinded expert; epidermal staining was not quantified. Statistical significance was calculated using a Mann-Whitney test.

Statistical Analysis:

Pearson correlations were calculated and plotted in Microsoft Office Excel 2007. Positive cells in the sections were counted at 40X magnification from at least 6 fields of view per sample, and four different mice per genotype. The numbers per field were compared for statistical significance by 1-way ANOVA using GraphPad Prism software.

Results

To our knowledge no direct comparisons have been made between the bleomycin-induced SSc mouse model or the Tsk2/+ model with human SSc; although both mouse models share features with human SSc, such as immune cell filtration, ECM deposition, and skin fibrosis (19-22). The datasets and analyses for human SSc skin and the sclGVHD mouse model gene expression have been previously published (2, 6). We used DWD analysis methods (15) to remove platform and species biases, and to integrate the human SSc biopsy and the mouse model gene expression datasets (Figure 1). Groupings in the human and mouse tissue datasets were specified based on intrinsic subset classification (2) or by mouse model (Table 1). 1,217 genes were selected from the merged dataset; samples and genes were organized by hierarchical clustering.

As expected, after hierarchical clustering, SSc skin biopsy samples clustered into the same groupings of the original molecular subsets (2). We were specifically interested in understanding how gene expression in different mouse models resembles specific subsets of human SSc. The sclGVHD samples clustered adjacent to the *inflammatory* patient samples, consistent with prior findings (6), and the Tsk2/+ mice at 4-weeks of age shared a dendrogram branch with the *fibroproliferative* subset, suggesting that the skin of these mice share gene expression features with the skin from patients in this subset (Figure 2A). Bleomycin-injected mice at 5 days and at 21 days clustered together, and shared a branch with the *inflammatory* subset. Samples from Tsk1/+ mice and the Tsk2/+ mice at 16-weeks clustered on a branch with the *limited* subset of patients. Each cluster containing both mouse and human samples is statistically significant (Figure 2; $p < 0.001$). We performed 2,000 iterations of K-means

clustering using consensus cluster, each using 2/3 of the data that showed four stable clusters (Supplemental Figure S1)

We have previously demonstrated enrichment of IL13-driven gene expression patterns in both the *inflammatory* subset and in sclGVHD mice (6). As previously demonstrated, membrane-associated IL13-specific receptor subunit *IL13RA1* has coordinately high relative expression in both groups along with *AIF1* and *CCL2*, both of which are IL13-regulated; all show high expression again in this interspecies comparison (Figure 2C) (23). Expression of many of these genes is similarly increased in the bleomycin-induced fibrosis model, consistent with the characterized role of IL13 signaling in fibrosis in these mice (24, 25).

Despite similarities between sclGVHD and bleomycin-induced skin fibrosis, a prominent cluster of interferon genes distinguishes these two models. This group of genes is not induced by bleomycin, but is highly upregulated in early sclGVHD. Many of the genes in this cluster are expressed at intermediate levels in SSc inflammatory skin and are targets associated with interferon signaling, such as *STAT1, IRF1, CCL5, IFI30*, and *JAK3* (Figure 2D).

A large number of the genes found coordinately regulated in the *fibroproliferative* subset and in Tsk2/+ mice at 4-weeks of age are associated with cell proliferation, including *KI67, CENPE, KIF20A, MCM7*, and *POLE2* (Figure 2F) (26). Skin samples from Tsk2/+ mice and their wild-type littermates were stained for the key signature molecule of the proliferative subset, *KI67*. The results (shown in Figure 5) confirmed the microarray findings that showed up-regulated *KI67* mRNA transcripts, which are expressed in significantly more cells, especially in hair follicle locations, primarily in male Tsk2/+ mice in which the disease is more severe; KI67 was not analyzed in female Tsk2/+ mice. Interestingly, gene expression in skin samples from Tsk2/+ mice at 16-weeks and Tsk1/+ at 6-weeks of age did not show a strong resemblance to the

genes differentially expressed in diffuse SSc, however, they do show some resemblance to limited SSc. In our mechanistic studies of sclGVHD and Tsk2/+ (6, 27), we found a strong correlation between age of the mice, the gene expression in the model and their resemblance of the human SSc subsets.

To examine the shared features of gene expression in the molecular subsets and in the mouse models on a genomic scale, we created a module map of enriched GO terms in the integrated human SSc and mouse model gene expression dataset (Figure 3). Many GO terms associated with proliferation and mitosis such as *mitotic checkpoint, cell division, regulation of mitosis and DNA replication initiation* were similarly positively enriched in the Tsk2/+ mice at 4-weeks and in the *fibroproliferative* biopsies, consistent with the increased expression of genes involved in cell cycle progression. Other GO terms coordinately regulated in this subset and in 4-week Tsk2/+ mice include those associated with RNA metabolism (*RNA splicing, mRNA processing* and *RNA binding*), translation (*protein-RNA complex assembly, ribosome* and *eukaryotic 43S initiation complex*) and DNA repair (*DNA-dependent DNA replication* and *DNA repair*). Proteins involved in these processes such as topoisomerase I and RNA polymerase I are frequent antigenic targets for autoantibodies in SSc (28-32). GO terms associated with lipid biogenesis such as *fatty acid metabolic process*, *lipid metabolic process* and *sterol biosynthetic process*, are enriched and downregulated in a portion of *fibroproliferative* biopsies and similarly downregulated in the Tsk2/+ mouse skin biopsies. Loss of subcutaneous fat is a characteristic feature of both diffuse SSc (33) and Tsk2/+ model (22). Additionally, it has been demonstrated that lipid biogenesis and fibrotic processes in fibroblasts are mutually antagonistic signaling systems. Activation TGFβ-signaling inhibits lipid biogenesis via PPARγ pathways and vice

versa (13, 34). Additionally, PPARγ ligands have been shown to suppress bleomycin-induced fibrosis in lungs (35).

Modules shared in skin from sclGVHD mice at two weeks and the *inflammatory* subset are predominantly associated with inflammatory processes, such as *immune system process, inflammatory response, chemokine activity,* and *response to biotic stimulus*. Increased expression of ECM-associated pathways is also evident in the *inflammatory* and *limited* subsets, as well as in the bleomycin-induced fibrosis and Tsk1/+ models, suggesting that pathways underlying fibrosis in these patients and these models may also be conserved.

Having established that skin from Tsk2/+ mice at 4-weeks shares gene expression features with the *fibroproliferative* subset, we specifically examined a role for TGFβ-signaling associated gene expression in this model. An *in vivo* TGFβ-responsive signature in mouse skin was generated and expression of these genes examined in skin from Tsk2/+ animals. Total RNA was prepared from skin samples of C57Bl/6 mice that had been surgically implanted with subcutaneous pumps containing 50, 250 or 1250ng of TGFβ, or phosphate-buffered saline (PBS) as a control, for 7 days. Gene expression data relative to the PBS treatment control is shown. We have termed the 719 probes that showed a 2-fold or more change in gene expression the mouse TGFβ-responsive signature (Figure 4). Canonical TGFβ targets found induced in this signature include *TIMP1, PAI1*, COL1A1, *SPP1, LOX, THBS* and *SPARC*, demonstrating that this signature is representative of a response to TGFβ in the skin of these animals. The data for the mouse TGFβ-responsive signature was extracted from the compendium of gene expression in skin of the models. Expression of TGFβ-responsive signature genes is enriched in skin from Tsk2/+ animals at 4-weeks (Figure 4B) suggesting that TGFβ-signaling is active in the skin of these mice. Since both the bleomycin-induced skin fibrosis and Tsk1/+ pathogenesis have been

shown to involve TGFβ-signaling, we confirmed the activation of TGFβ-signaling in Tsk1/+ and bleomycin-induced skin fibrosis (Figure 4C). Analysis of the bleomycin-induced skin fibrosis model using GSEA confirms enrichment of the TGFβ-activated gene expression (Supplemental Figure S2).

We recently demonstrated that PDGF signaling is also increased in the fibroproliferative subset of SSc patients (9). We analyzed the PDGF stimulated gene set from Johnson et al. and find PDGF stimulated gene expression increased in Tsk2/+ mice at 4-weeks of age and in the sclGVHD mouse model (data not shown). PDGF gene expression was not enriched in either the bleomycin-induced fibrosis model, Tsk1/+ or Tsk2/+ at 16-weeks of age.

To confirm the molecular similarities between the patients in the *fibroproliferative* subset and Tsk2/+ mice we analyzed the expression of the TGFβ-regulated gene tumor necrosis factor receptor superfamily, member 12a (*Tnfrsf12a*; also designated *Tweak-R*, *fibroblast growth factor-inducible-14*). *Tweak-R* is highly expressed in the *fibroproliferative* subset and was among a set of genes highly correlated to worse skin disease (2). *Tweak-R* mRNA levels were also found increased in the fibroproliferative subsets of Milano *et al.* (Figure 5A; $p < 0.00001$); gene expression differences for *Tweak-R* (*Tnfrsf12a*) were confirmed by qRT-PCR (2). Immunohistochemistry analysis of a subset of biopsies from the Milano et al. study (2) showed that TWEAK-R protein levels were increased in the dermis and around epidermal skin appendages of patients of the fibroproliferative and inflammatory subsets (Figure 5B; $p < 0.05$, Mann-Whitney test). The Tsk2/+ animal model has cycles throughout young life while the disease is becoming established. We used this information to select skin samples at peak disease activity for assessment of TWEAK-R expression. We performed immunohistochemistry on sections of skin taken from female Tsk2/+ and their wild-type littermates to detect the levels of

TWEAK-R (Figure 5C - D). *Tweak-R* is induced by TGFβ (36) and there is an ~1.5-fold increase in numbers of TWEAK-R positive cells at in skin of 4-week-old female Tsk2/+ mice compared to their wild-type littermates (Figure 5C - D; p = 0.013) and in male Tsk2/+ mice compared to their wild-type littermates (not shown). Taken together with the findings of enrichment of TGFβ-signaling primarily in the *fibroproliferative* skin biopsies of the Milano cohort (5), we believe that Tsk2/+ skin at 4-weeks reflects the biology of this subset of SSc.

Discussion

The identification of mouse models that share gene expression patterns with the *fibroproliferative* and *inflammatory* subsets of SSc provides the research community with valuable tools in which to closely examine the molecular mechanisms underlying disease in these subsets. The data presented here are consistent with our earlier findings that sclGVHD mice share underlying mechanistic pathways with the *inflammatory* subset of SSc (6). Here we have extended those findings by identifying Tsk2/+ mice at 4-weeks of age as a promising model of disease for the *fibroproliferative* subset. We used multiple approaches to demonstrate parallels in gene expression in this mouse and *fibroproliferative* skin biopsies, including analysis of individual genes and clusters, examination of global similarities based on co-expression of GO terms, and specifically testing for enrichment of TGFβ-responsive gene expression in both species. PDGF stimulated gene expression has also been shown to underlie the *fibroproliferative* subset and we find PDGF gene expression also increased in the Tsk2/+ mouse.

The role of TGFβ signaling in fibrosis and as a driver of disease in SSc is well established. An important new finding is the demonstration that Tsk2/+ mice have activation of TGFβ responsive gene expression. We have shown that Tsk2/+ results from an ENU-induced point mutation in the PIIINP fragment of the *Col3a1* gene (27). We show that the TGFβ target,

Tnfrsf12a (*Tweak-R* or *Fn14*) is highly expressed both in SSc patients of this subset and in Tsk2/+ mice at 4-weeks of age. Interestingly, mice deficient for *Tweak-R* have significantly reduced liver progenitor cell proliferation in response to chemical liver injury (37). Its ligand, *Tweak / Tnfsf12,* has been implicated as a driver of inflammation and proliferation of fibroblasts / epithelial cells and can contribute to kidney fibrosis (38, 39). The kidneys of *Tweak* KO mice show a decreased number of myofibroblasts with lower proliferation, and reduced ECM accumulation; mice over-expressing TWEAK have increased kidney fibrosis (38). We propose that a common mechanism underlying the fibroproliferative subset of SSc and the Tsk2/+ model is due in part to TGFβ signaling leading to activation of the *Tweak / Fn14* signaling axis.

The deregulation of TGFβ signaling in the bleomycin model has been previously demonstrated (40, 41) and we observe activation of TGFβ signaling here (Supplemental Figure S2). The bleomycin-induced fibrosis model has also been used to advance knowledge of the roles for both TGFβ and IL13 in fibrosis and has been particularly informative for revealing the interactions between these signaling pathways in fibrotic processes (24, 25). The inflammatory gene expression signature dominates the Bleomycin fibrosis signature and the clustering results. This is consistent with data showing that mice lacking a functional inflammasome are not susceptible to bleomycin-induced fibrosis (42-45). Our data suggest the bleomycin-induced skin fibrosis model shows similarities to the early inflammatory response observed in patients and shows both TGFβ and IL13 signatures, both of which are implicated in SSc pathogenesis.

The resemblance of the Tsk1/+ mouse to the SSc subsets is not clear or consistent with prior findings (46). We found no significant parallels of gene expression in skin of these animals at 6-weeks of age with the molecular subsets of SSc. Detailed analysis of Tsk1/+ at other time points might show stronger similarities to gene expression in SSc skin. The Genomica module

map analysis did not demonstrate any contribution to gene expression patterns by B-lymphocytes in Tsk1/+ as has been previously reported. It is possible that the reported reliance of the phenotype in these mice on functional IL4 signaling (47) is relevant to the *inflammatory* or *limited* subsets, however the GO term analysis implemented in these studies is not sensitive enough to draw definitive conclusions from these data.

Previous comparisons of human SSc and human cGVHD at the histological level show key similarities as well as distinct differences. Fleming et al. compared human cGVHD to human SSc histologically (48) and showed a similar extent of fibrosis, increased hyaluronan staining, myofibroblasts numbers and intimal hyperplasia of the microvasculature including increased smooth muscle cell layers. SSc and cGVHD showed positive dermal KI67 staining. The major difference was that SSc showed differences in the pattern of fibrotic change and a decrease in staining for endothelial cell markers indicating capillary rarefaction, not found in cGVHD.

This study has several limitations. The bleomycin-induced fibrosis model results are likely specific to the protocol used here. In addition, the TGFβ pump model captures only a single time point and this cytokine likely induces changes in a time dependent-manner. We note that the proliferating cells observed in SSc are mainly located in the epidermis with some proliferating cells around the vasculature and other epidermal skin appendages (2, 48). Since the structure of mouse skin, particularly the epidermis, is quite different in mice, this could limit interpretation of the Tsk2/+ data and mapping of our subsets.

With few tools available for studying disease mechanisms in this poorly understood disease, the benefits of the results of this study are clear. By using gene expression as a readily quantifiable phenotype in skin, we have shown that the Tsk/2+ mouse, the sclGVHD and

bleomycin-induced fibrosis mouse models are similar to their respective human subsets not by gross morphology approximations, but by robust molecular measures. Additionally, identification of the pathways similarly deregulated in human SSc subsets and in mouse models, provides new, urgently needed tools for optimizing development of therapeutics that specifically target each of the intrinsic subsets of scleroderma. The mechanisms in these mouse models can be used to identify drugs for SSc patients in the clinic and target their molecular subsets. The development of a routine diagnostic for SSc patients as well as drugs that target each subset will greatly facilitate treatment of this disease.

References:

1. Christner PJ, Jimenez SA. Animal models of systemic sclerosis: insights into systemic sclerosis pathogenesis and potential therapeutic approaches. Curr Opin Rheumatol. 2004;16(6):746-52.
2. Milano A, Pendergrass SA, Sargent JL, George LK, McCalmont TH, Connolly MK, et al. Molecular subsets in the gene expression signatures of scleroderma skin. PLoS ONE. 2008;3(7):e2696.
3. Pendergrass SA, Lemaire R, Francis IP, Mahoney JM, Lafyatis R, Whitfield ML. Intrinsic gene expression subsets of diffuse cutaneous systemic sclerosis are stable in serial skin biopsies. J Invest Dermatol. 2012;132(5):1363-73.
4. Hinchcliff ME, Huang CC, Wood TA, Mahoney JM, Martyanov V, Bhattacharya S, et al. Molecular Signatures in Skin Associated with Clinical Improvement During Mycophenolate Treatment in Systemic Sclerosis. J Invest Dermatol. 2013;In Press.
5. Sargent JL, Milano A, Bhattacharyya S, Varga J, Connolly MK, Chang HY, et al. A TGFbeta-responsive gene signature is associated with a subset of diffuse scleroderma with increased disease severity. J Invest Dermatol. 2009;130(3):694-705.
6. Greenblatt MB, Sargent JL, Farina G, Tsang K, Lafyatis R, Glimcher LH, et al. Interspecies Comparison of Human and Murine Scleroderma Reveals IL-13 and CCL2 as Disease Subset-Specific Targets. Am J Pathol. 2012.
7. Johnson ME, Mahoney JM, Taroni J, Sargent JL, Marmarelis E, Wu MR, et al. Experimentally-derived fibroblast gene signatures identify molecular pathways associated with distinct subsets of systemic sclerosis patients in three independent cohorts. PLoS One. 2015;10(1):e0114017.
8. Katsumoto TR, Whitfield ML, Connolly MK. The pathogenesis of systemic sclerosis. Annual review of pathology. 2011;6:509-37.
9. Johnson ME, Pioli PA, Whitfield ML. Gene expression profiling offers insights into the role of innate immune signaling in SSc. Seminars in immunopathology. 2015.
10. Mahoney JM, Taroni J, Martyanov V, Wood TA, Greene CS, Pioli PA, et al. Systems level analysis of systemic sclerosis shows a network of immune and profibrotic pathways connected with genetic polymorphisms. PLoS Comput Biol. 2015;11(1):e1004005.
11. Barisic-Dujmovic T, Boban I, Clark SH. Regulation of collagen gene expression in the Tsk2 mouse. J Cell Physiol. 2008;215(2):464-71.
12. Bayle J, Fitch J, Jacobsen K, Kumar R, Lafyatis R, Lemaire R. Increased expression of Wnt2 and SFRP4 in Tsk mouse skin: role of Wnt signaling in altered dermal fibrillin deposition and systemic sclerosis. J Invest Dermatol. 2008;128(4):871-81.
13. Wu M, Melichian DS, Chang E, Warner-Blankenship M, Ghosh AK, Varga J. Rosiglitazone abrogates bleomycin-induced scleroderma and blocks profibrotic responses through peroxisome proliferator-activated receptor-gamma. Am J Pathol. 2009;174(2):519-33.
14. Bult CJ, Eppig JT, Kadin JA, Richardson JE, Blake JA. The Mouse Genome Database (MGD): mouse biology and model systems. Nucleic Acids Res. 2008;36(Database issue):D724-8.
15. Benito M, Parker J, Du Q, Wu J, Xiang D, Perou CM, et al. Adjustment of systematic microarray data biases. Bioinformatics. 2004;20(1):105-14.
16. Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, et al. Molecular portraits of human breast tumours. Nature. 2000;406(6797):747-52.

17. Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci U S A. 1998;95(25):14863-8.
18. Segal E, Shapira M, Regev A, Pe'er D, Botstein D, Koller D, et al. Module networks: identifying regulatory modules and their condition-specific regulators from gene expression data. Nat Genet. 2003;34(2):166-76.
19. Adamson IY, Bowden DH. The pathogenesis of bloemycin-induced pulmonary fibrosis in mice. Am J Pathol. 1974;77(2):185-97.
20. Zhao J, Shi W, Wang YL, Chen H, Bringas P, Jr., Datto MB, et al. Smad3 deficiency attenuates bleomycin-induced pulmonary fibrosis in mice. Am J Physiol Lung Cell Mol Physiol. 2002;282(3):L585-93.
21. Christner PJ, Siracusa LD, Hawkins DF, McGrath R, Betz JK, Ball ST, et al. A high-resolution linkage map of the tight skin 2 (Tsk2) locus: a mouse model for scleroderma (SSc) and other cutaneous fibrotic diseases. Mammalian genome : official journal of the International Mammalian Genome Society. 1996;7(8):610-2.
22. Christner PJ, Peters J, Hawkins D, Siracusa LD, Jimenez SA. The tight skin 2 mouse. An animal model of scleroderma displaying cutaneous fibrosis and mononuclear cell infiltration. Arthritis Rheum. 1995;38(12):1791-8.
23. Fulkerson PC, Fischetti CA, Hassman LM, Nikolaidis NM, Rothenberg ME. Persistent effects induced by IL-13 in the lung. Am J Respir Cell Mol Biol. 2006;35(3):337-46.
24. Matsushita M, Yamamoto T, Nishioka K. Upregulation of interleukin-13 and its receptor in a murine model of bleomycin-induced scleroderma. Int Arch Allergy Immunol. 2004;135(4):348-56.
25. Aliprantis AO, Wang J, Fathman JW, Lemaire R, Dorfman DM, Lafyatis R, et al. Transcription factor T-bet regulates skin sclerosis through its function in innate immunity and via IL-13. Proc Natl Acad Sci U S A. 2007;104(8):2827-30.
26. Whitfield ML, George LK, Grant GD, Perou CM. Common markers of proliferation. Nat Rev Cancer. 2006;6(2):99-106.
27. Long KB, Li Z, Burgwin CM, Choe SG, Martyanov V, Sassi-Gaha S, et al. The Tsk2/+ mouse fibrotic phenotype is due to a gain-of-function mutation in the PIIINP segment of the Col3a1 gene. J Invest Dermatol. 2015;135(3):718-27.
28. Greidinger EL, Flaherty KT, White B, Rosen A, Wigley FM, Wise RA. African-American race and antibodies to topoisomerase I are associated with increased severity of scleroderma lung disease. Chest. 1998;114(3):801-7.
29. Hu PQ, Fertig N, Medsger TA, Jr., Wright TM. Correlation of serum anti-DNA topoisomerase I antibody levels with disease severity and activity in systemic sclerosis. Arthritis Rheum. 2003;48(5):1363-73.
30. Steen VD. Autoantibodies in systemic sclerosis. Semin Arthritis Rheum. 2005;35(1):35-42.
31. Hanke K, Dahnrich C, Bruckner CS, Huscher D, Becker M, Jansen A, et al. Diagnostic value of anti-topoisomerase I antibodies in a large monocentric cohort. Arthritis Res Ther. 2009;11(1):R28.
32. Reimer G, Rose KM, Scheer U, Tan EM. Autoantibody to RNA polymerase I in scleroderma sera. J Clin Invest. 1987;79(1):65-72.
33. Fleischmajer R, Damiano V, Nedwich A. Alteration of subcutaneous tissue in systemic scleroderma. Arch Dermatol. 1972;105(1):59-66.

34. Ghosh AK, Bhattacharyya S, Lakos G, Chen SJ, Mori Y, Varga J. Disruption of transforming growth factor beta signaling and profibrotic responses in normal skin fibroblasts by peroxisome proliferator-activated receptor gamma. Arthritis Rheum. 2004;50(4):1305-18.
35. Aoki Y, Maeno T, Aoyagi K, Ueno M, Aoki F, Aoki N, et al. Pioglitazone, a peroxisome proliferator-activated receptor gamma ligand, suppresses bleomycin-induced acute lung injury and fibrosis. Respiration. 2009;77(3):311-9.
36. Milks MW, Cripps JG, Lin H, Wang J, Robinson RT, Sargent JL, et al. The role of Ifng in alterations in liver gene expression in a mouse model of fulminant autoimmune hepatitis. Liver Int. 2009;29(9):1307-15.
37. Jakubowski A, Ambrose C, Parr M, Lincecum JM, Wang MZ, Zheng TS, et al. TWEAK induces liver progenitor cell proliferation. J Clin Invest. 2005;115(9):2330-40.
38. Ucero AC, Benito-Martin A, Fuentes-Calvo I, Santamaria B, Blanco J, Lopez-Novoa JM, et al. TNF-related weak inducer of apoptosis (TWEAK) promotes kidney fibrosis and Ras-dependent proliferation of cultured renal fibroblast. Biochim Biophys Acta. 2013;1832(10):1744-55.
39. Zhu LX, Zhang HH, Mei YF, Zhao YP, Zhang ZY. Role of tumor necrosis factor-like weak inducer of apoptosis (TWEAK)/fibroblast growth factor-inducible 14 (Fn14) axis in rheumatic diseases. Chinese medical journal. 2012;125(21):3898-904.
40. Lakos G, Takagawa S, Chen SJ, Ferreira AM, Han G, Masuda K, et al. Targeted disruption of TGF-beta/Smad3 signaling modulates skin fibrosis in a mouse model of scleroderma. Am J Pathol. 2004;165(1):203-17.
41. Santiago B, Gutierrez-Canas I, Dotor J, Palao G, Lasarte JJ, Ruiz J, et al. Topical application of a peptide inhibitor of transforming growth factor-beta1 ameliorates bleomycin-induced skin fibrosis. J Invest Dermatol. 2005;125(3):450-5.
42. Gasse P, Mary C, Guenon I, Noulin N, Charron S, Schnyder-Candrian S, et al. IL-1R1/MyD88 signaling and the inflammasome are essential in pulmonary inflammation and fibrosis in mice. J Clin Invest. 2007;117(12):3786-99.
43. Artlett CM, Sassi-Gaha S, Rieger JL, Boesteanu AC, Feghali-Bostwick CA, Katsikis PD. The inflammasome activating caspase 1 mediates fibrosis and myofibroblast differentiation in systemic sclerosis. Arthritis Rheum. 2011;63(11):3563-74.
44. Artlett CM. The Role of the NLRP3 Inflammasome in Fibrosis. The open rheumatology journal. 2012;6:80-6.
45. Derk CT, Artlett CM, Jimenez SA. Morbidity and mortality of patients diagnosed with systemic sclerosis after the age of 75: a nested case-control study. Clin Rheumatol. 2006:1-4.
46. Baxter RM, Crowell TP, McCrann ME, Frew EM, Gardner H. Analysis of the tight skin (Tsk1/+) mouse as a model for testing antifibrotic agents. Lab Invest. 2005;85(10):1199-209.
47. Ong CJ, Ip S, Teh SJ, Wong C, Jirik FR, Grusby MJ, et al. A role for T helper 2 cells in mediating skin fibrosis in tight-skin mice. Cell Immunol. 1999;196(1):60-8.
48. Fleming JN, Shulman HM, Nash RA, Johnson PY, Wight TN, Gown A, et al. Cutaneous chronic graft-versus-host disease does not have the abnormal endothelial phenotype or vascular rarefaction characteristic of systemic sclerosis. PLoS One. 2009;4(7):e6203.
49. Herschkowitz JI, Simin K, Weigman VJ, Mikaelian I, Usary J, Hu Z, et al. Identification of conserved gene expression features between murine mammary carcinoma models and human breast tumors. Genome Biol. 2007;8(5):R76.

Intrinsic Group	Description	Species	# of Microarrays
1	Fibroproliferative (previously *diffuse-proliferation*)	human	27
2	inflammatory	human	17
3	limited	human	9
4	normal-like	human	22
5	cGVHD 2 weeks	mouse	9
6	cGVHD 5 weeks	mouse	4
7	s.c. bleomycin 5 days	mouse	3
8	s.c bleomycin 21 days	mouse	4
9	Tsk2/+ 4-weeks	mouse	4
10	Tsk2/+ 16-weeks	mouse	5
11	Tsk1/+ 6-weeks	mouse	4

Table 1. Eleven groups were specified for intrinsic gene analysis.

Four goups of human samples were specified based on the intrinsic subsets of SSc (2) and seven groups of mouse models were defined based on the model type and different timepoints were specified. (s.c. subcutaneous).

Figure legends

Fig. 1. Implementation of integrated interspecies analysis for systemic sclerosis. The schematic of the analysis strategy based on that of Herschkowitz *et al.* (49) is shown. Human and mouse datasets were merged as described in the text, the biases removed using DWD, and the intrinsic genes selected and clustered for further analysis.

Fig. 2. Clusters of gene expression patterns are shared in human SSc skin and in the mouse models. (A) Integrated human mouse expression data clustering**.** The upper dendrogram shows the organization of the molecular subsets of SSc. The shading indicates the placement of these subsets in the integrated dataset cluster analysis. Microarrays from the mouse models are interspersed among the human subsets. Notably, Tsk2/+ 1 month samples cluster on the same branch as the *fibroproliferative* subset and sclGVHD mice at 2 weeks and at 5 weeks are found clustered with the *inflammatory* subset. Samples from the bleomycin mouse model show are most closely associated with inflammatory subset but show aspects of inflammatory and fibroproliferative gene expression. **(B)** 1217 intrinsic genes were clustered in the gene and array dimensions. The sample dendrogram is that from Figure 1. Selected clusters of interest are shown. Gene expression features shared by the sclGVHD mice and the *inflammatory* subset include those induced by IL13 **(C)** and IFN-signaling **(D)**. Genes associated with proliferation were up-regulated in Tsk2/+ mice at 1 month and in the *fibroproliferative* subset **(F)**. A cluster of genes was found up-regulated in the *inflammatory* and *limited* subsets as well as the sclGVHD and other models **(E)**.

Fig. 3. Module map of coordinately regulated GO terms in human SSc and in mouse models. A module map of enriched GO terms was created using gene expression from the integrated interspecies microarray dataset. Modules significantly enriched ($p < 0.05$, FDR 0.05, hypergeometric distribution) in at least 15 of the 109 arrays were selected and are displayed. Clusters of select GO terms are shown to the right of the module map. Each column represents a microarray and each row is a GO term. The arrays have been ordered as per the intrinsic clustering shown in Figure 1. Positively enriched and negatively enriched modules are shown by red and green squares respectively.

Fig. 4. TGFβ-responsive signature gene expression in SSc mouse models.

(A) Pumps containing PBS or 50, 250 or 1250ng of TGFβ were surgically inserted subcutaneously in B57/B6 mice for 7 days and skin analyzed by DNA microarray. 719 genes changed in expression >2-fold from the PBS control in at two doses of TGFβ. Data were T0 transformed against the PBS control and the blue wedge is indicative of increasing TGFβ concentrations. Data for the 719 TGFβ-responsive genes were extracted from the SSc mouse models and clustered in the array and gene dimensions. Pearson correlations of the 1250ng TGFβ dose and each microarray were calculated and are plotted directly beneath the heatmap. The TGFβ dose response is shown to the left of the heatmap. The highest TGFβ gene expression is observed in 5 week samples from the sclGVHD mouse and 4-week old Tsk2/+ mice. (B) Dendrogram of mouse samples analyzed colored-coded by model. (C) Canonical TGFβ targets COL1A1, WISP1, SPARC, and TIMP1 were found TGFβ-responsive. (D) TGFβ-induced genes highly expressed in the sclGVHD model. (E) Proliferation genes induced by the TGFβ treatment.

Fig. 5. Tweak R (TNFRSF12A) is highly differentially expressed in both the fibroproliferative SSc patients and the Tsk2/+ mouse model. (A) SSc patients in the fibroproliferative subset express significant more *tnfrsf12a* (*Tweak-R*) than normal controls or patients in the inflammatory, proliferative or normal-like subsets ($p < 0.00001$). (B) SSc patients in the fibroproliferative and inflammatory subsets (n=5) have higher TNFRSF12A protein in the dermis and surrounding epidermal skin appendages than patients in the limited intrinsic subset (n=2), or the normal/normal-like subset (n=3 healthy controls; n=1 dSSc normal-like) as determined by staining of paraffin embedded SSc skin sections from a subset of patients of Milano et al. (2)($p < 0.05$; Mann-Whitney test). Variable increased staining was observed in both *fibroproliferative* and *inflammatory* patients, and due to the low (n=2) patients analyzed in the *inflammatory* subset these data were graphed together with those of the *fibroproliferative* subset. (C) Tsk2/+ mice expressed 1.6-fold more TNFRSF12A (TWEAK-R) than wild type littermates ($p=0.013$). Four-week-old female mice were scored for the number of TWEAK-R (Fn14)+ cells per field of view (400X magnification). Significance was calculated with a paired t-test using at least 8 fields of view per mouse and 4 mice per genotype. (D) Representative immunofluorescent images of two wild type mice (Left), two Tsk2/+ mice (center) and the isotype control staining (right). Skin samples were obtained from the lower dorsal back, paraffin embedded, and evaluated by immunofluorescence for the presence of TWEAK-R(red) and DAPI(blue). (E) and (F), KI67 staining for proliferating cells in the hair follicles and dermis of Tsk2 mice at 2, 4, 10 and 23 weeks of age. We find a significant increase only in the 4 and 10 week samples.

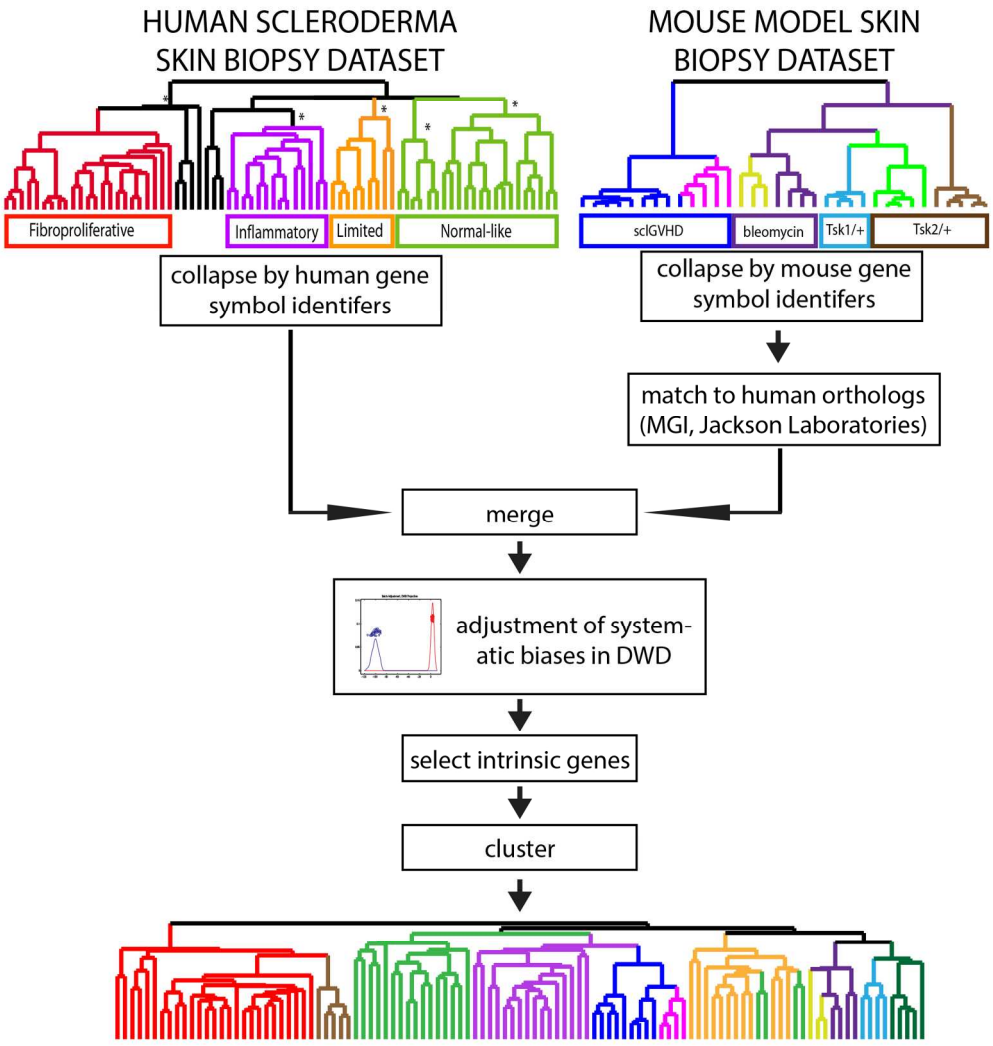


Fig. 1. Implementation of integrated interspecies analysis for systemic sclerosis. The schematic of the analysis strategy based on that of Herschkowitz et al. (49) is shown. Human and mouse datasets were merged as described in the text, the biases removed using DWD, and the intrinsic genes selected and clustered for further analysis.
166x174mm (300 x 300 DPI)

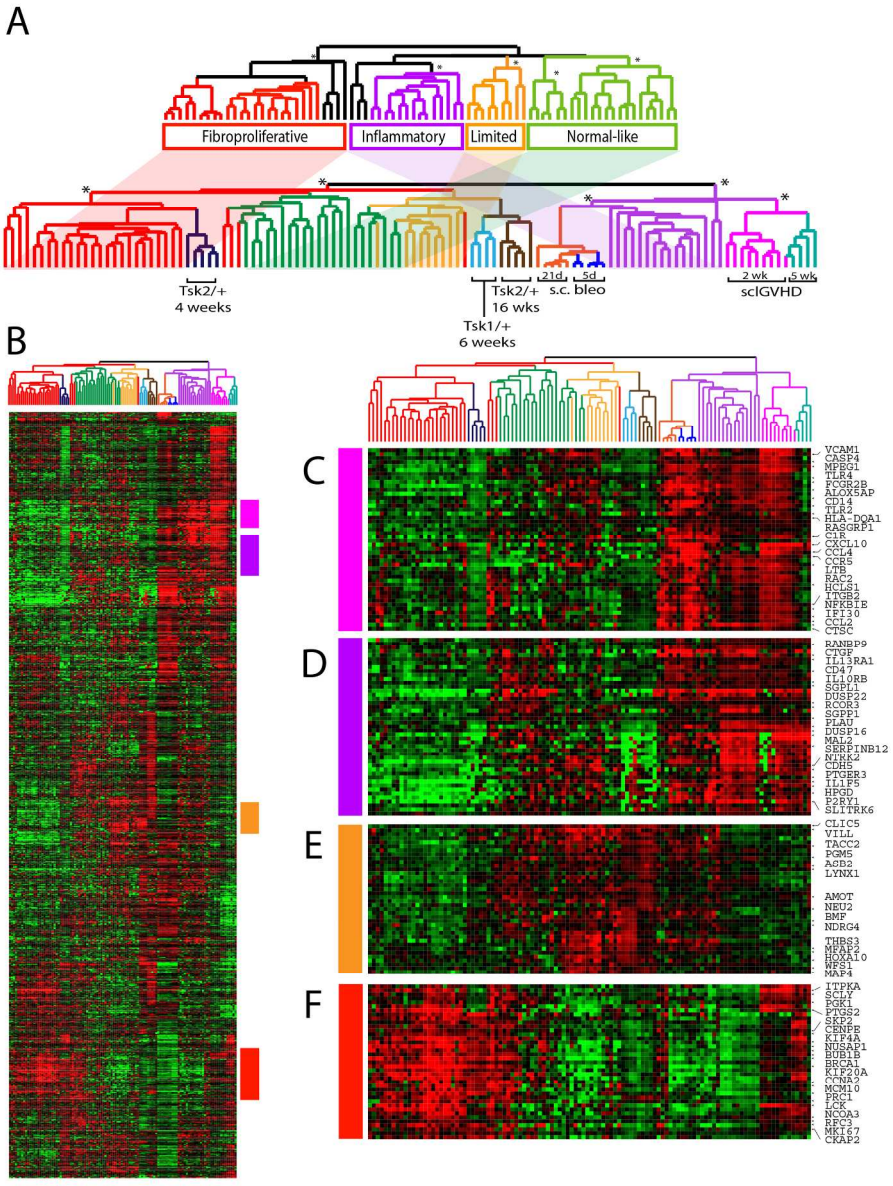


Fig. 2. Clusters of gene expression patterns are shared in human SSc skin and in the mouse models. (A) Integrated human mouse expression data clustering. The upper dendrogram shows the organization of the molecular subsets of SSc. The shading indicates the placement of these subsets in the integrated dataset cluster analysis. Microarrays from the mouse models are interspersed among the human subsets. Notably, Tsk2/+ 1 month samples cluster on the same branch as the fibroproliferative subset and sclGVHD mice at 2 weeks and at 5 weeks are found clustered with the inflammatory subset. Samples from the bleomycin mouse model show are most closely associated with inflammatory subset but show aspects of inflammatory and fibroproliferative gene expression. (B) 1217 intrinsic genes were clustered in the gene and array dimensions. The sample dendrogram is that from Figure 1. Selected clusters of interest are shown. Gene expression features shared by the sclGVHD mice and the inflammatory subset include those induced by IL13 (C) and IFN-signaling (D). Genes associated with proliferation were up-regulated in Tsk2/+ mice at 1 month and in the fibroproliferative subset (F). A cluster of genes was found up-regulated in the inflammatory and limited subsets as well as the sclGVHD and other models (E).

186x250mm (300 x 300 DPI)

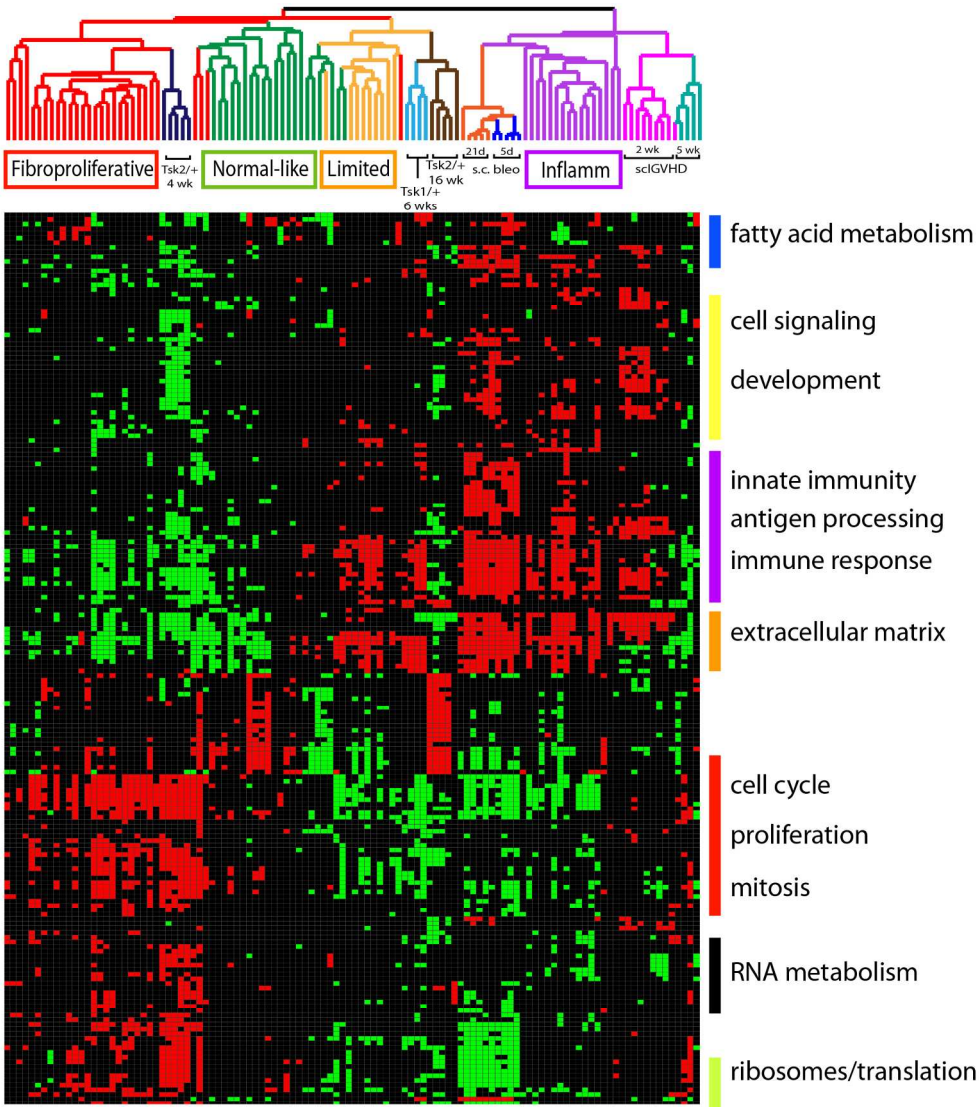


Fig. 3. Module map of coordinately regulated GO terms in human SSc and in mouse models. A module map of enriched GO terms was created using gene expression from the integrated interspecies microarray dataset. Modules significantly enriched ($p < 0.05$, FDR 0.05, hypergeometric distribution) in at least 15 of the 109 arrays were selected and are displayed. Clusters of select GO terms are shown to the right of the module map. Each column represents a microarray and each row is a GO term. The arrays have been ordered as per the intrinsic clustering shown in Figure 1. Positively enriched and negatively enriched modules are shown by red and green squares respectively.
183x201mm (300 x 300 DPI)

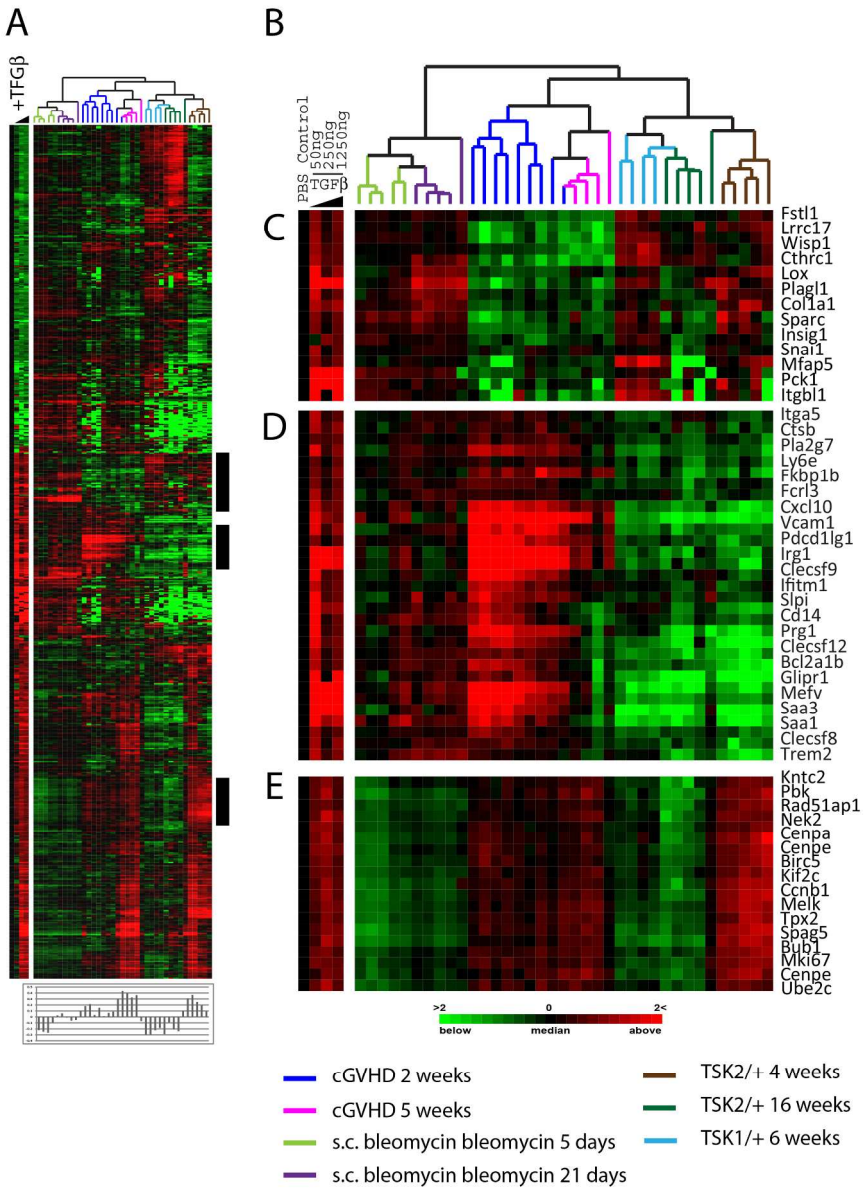


Fig. 4. TGFβ-responsive signature gene expression in SSc mouse models. (A) Pumps containing PBS or 50, 250 or 1250ng of TGFβ were surgically inserted subcutaneously in B57/B6 mice for 7 days and skin analyzed by DNA microarray. 719 genes changed in expression >2-fold from the PBS control in at two doses of TGFβ. Data were T0 transformed against the PBS control and the blue wedge is indicative of increasing TGFβ concentrations. Data for the 719 TGFβ-responsive genes were extracted from the SSc mouse models and clustered in the array and gene dimensions. Pearson correlations of the 1250ng TGFβ dose and each microarray were calculated and are plotted directly beneath the heatmap. The TGFβ dose response is shown to the left of the heatmap. The highest TGFβ gene expression is observed in 5 week samples from the sclGVHD mouse and 4 week old Tsk2/+ mice. (B) Dendrogram of mouse samples analyzed colored-coded by model. (C) Canonical TGFβ targets COL1A1, WISP1, SPARC, and TIMP1 were found TGFβ-responsive. (D) TGFβ-induced genes highly expressed in the sclGVHD model. (E) Proliferation genes induced by the TGFβ treatment.
189x256mm (300 x 300 DPI)

Figure 5

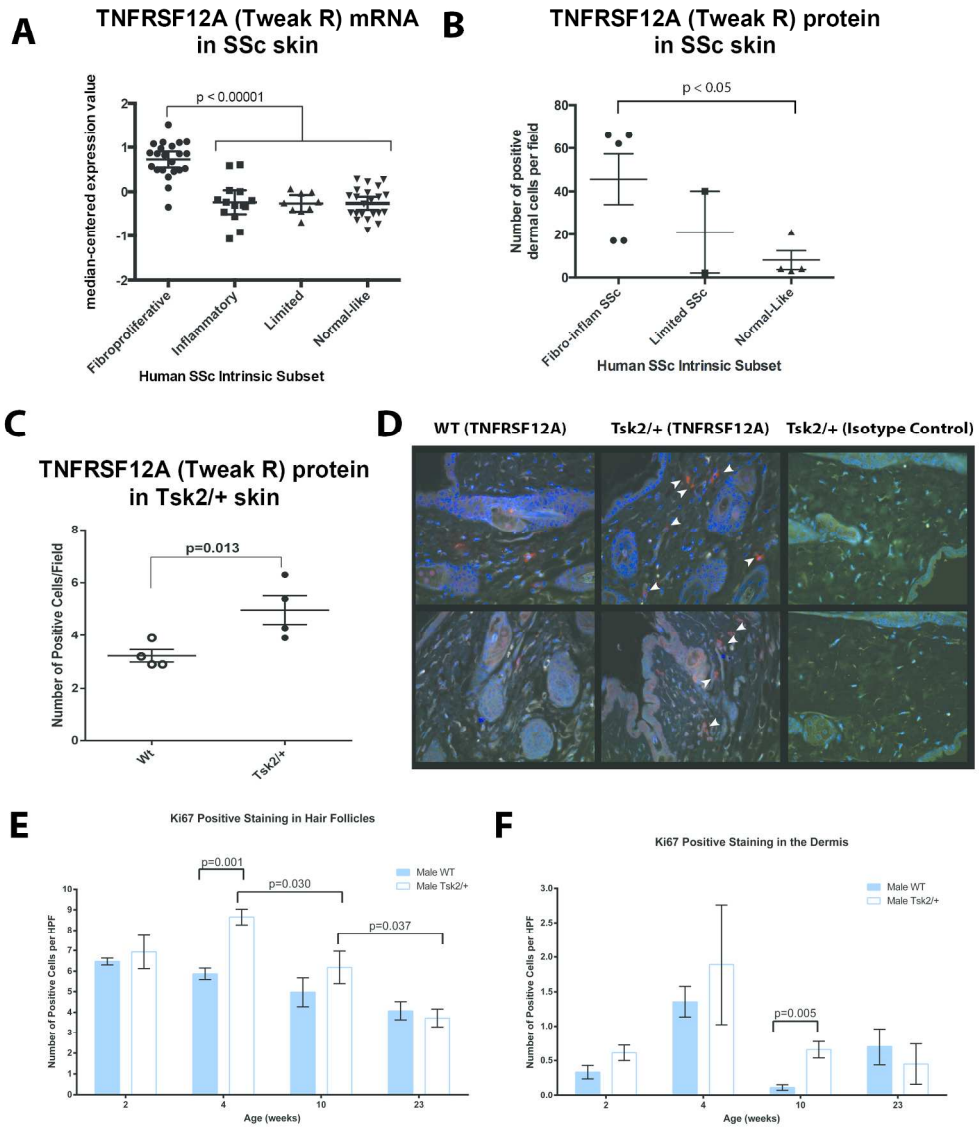
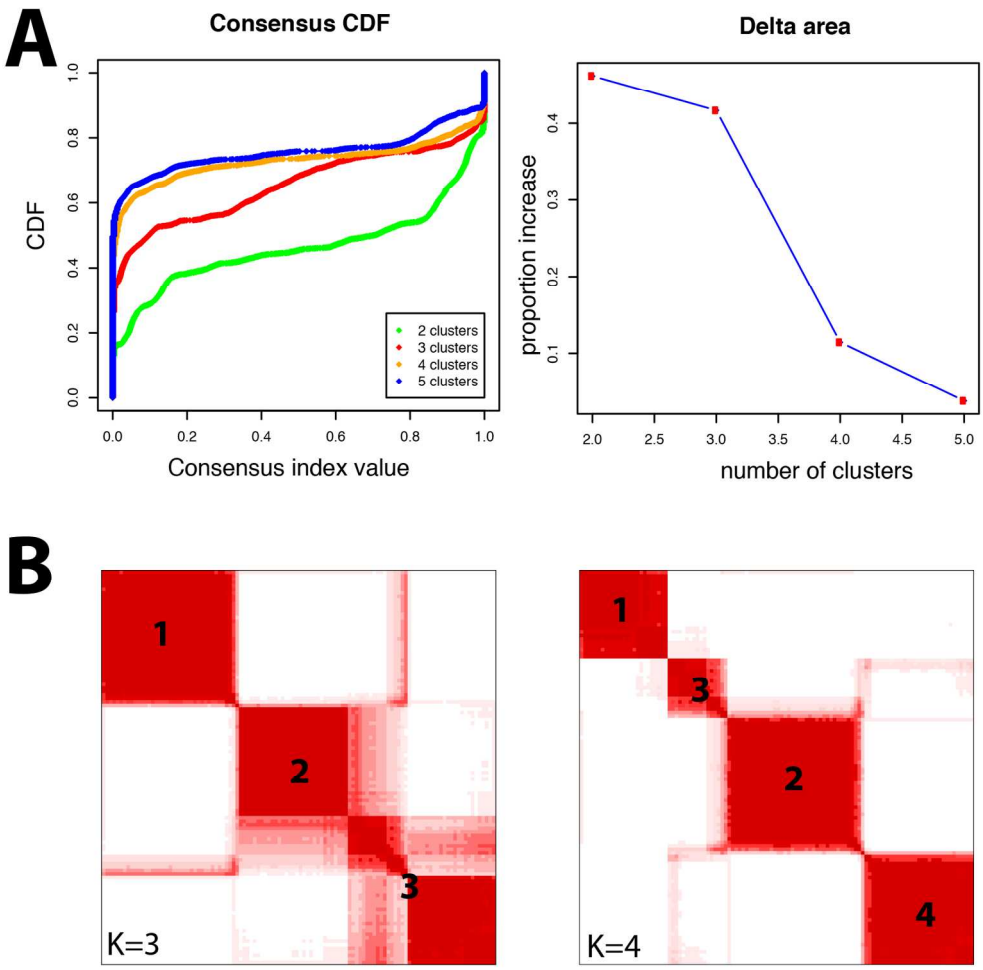
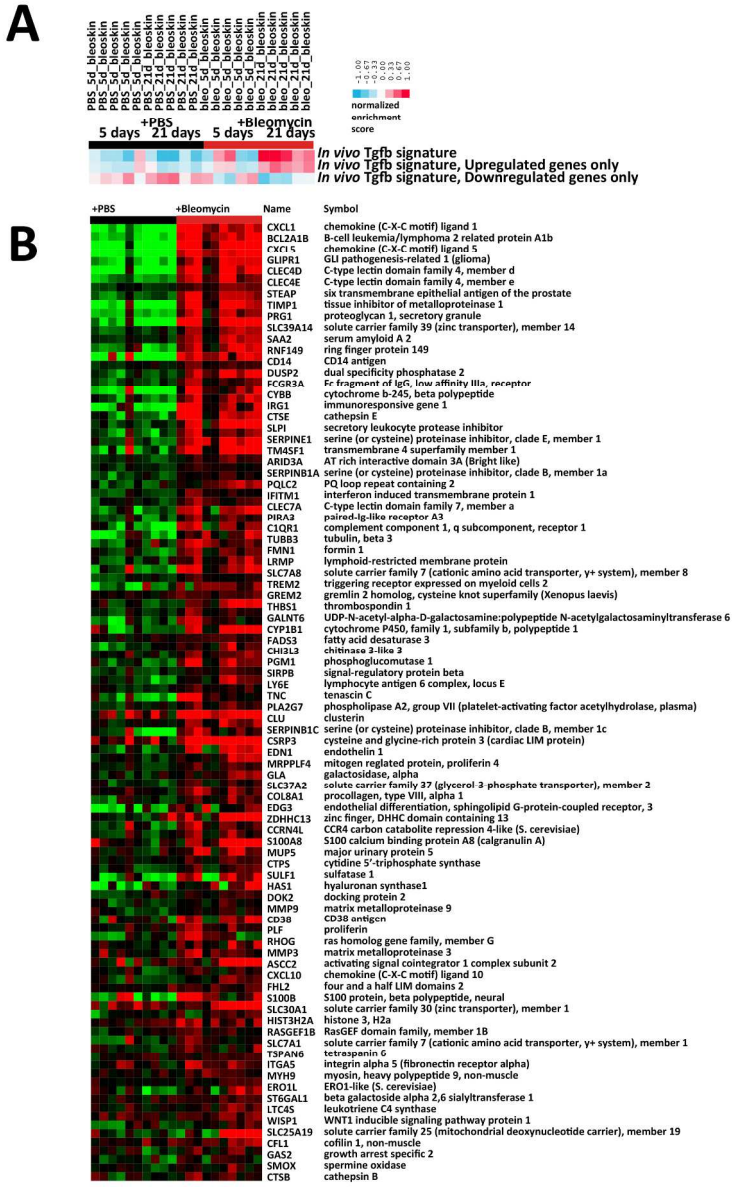


Fig. 5. Tweak R (TNFRSF12A) is highly differentially expressed in both the fibroproliferative SSc patients and the Tsk2/+ mouse model. (A) SSc patients in the fibroproliferative subset express significant more tnfrsf12a (Tweak-R) than normal controls or patients in the inflammatory, proliferative or normal-like subsets ($p < 0.00001$). (B) SSc patients in the fibroproliferative and inflammatory subsets (n=5) have higher TNFRSF12A protein in the dermis and surrounding epidermal skin appendages than patients in the limited intrinsic subset (n=2), or the normal/normal-like subset (n=3 healthy controls; n=1 dSSc normal-like) as determined by staining of paraffin embedded SSc skin sections from a subset of patients of Milano et al. (1)($p < 0.05$; Mann-Whitney test). (C) Tsk2/+ mice expressed 1.6-fold more TNFRSF12A (TWEAK-R) than wild type littermates (p=0.013). Four-week-old female mice were scored for the number of TWEAK-R (Fn14)+ cells per field of view (400X magnification). Significance was calculated with a paired t-test using at least 8 fields of view per mouse and 4 mice per genotype. (D) Representative immunofluorescent images of two wild type mice (Left), two Tsk2/+ mice (center) and the isotype control staining (right). Skin samples were obtained from the lower dorsal back, paraffin embedded, and evaluated by immunofluorescence for the

presence of TWEAK-R(red) and DAPI(blue). (E) and (F), KI67 staining for proliferating cells in the hair follicles and dermis of Tsk2 mice at 2, 4, 10 and 23 weeks of age. We find a significant increase only in the 4 and 10 week samples.
203x243mm (300 x 300 DPI)



Supplemental Figure S1. Stability of clustering results using Consensus Cluster. Shown are the results from 2,000 iterations of clustering using a randomly selected 2/3 of the data. A. The consensus CDF function suggest 3 – 4 clusters in the data. B. K-means clustering using K=3 or K=4 (K=5 performed but not shown) shows stable clusters with K=4 giving the most robust clustering results. The clusters shown in the K=4 K-means clustering correspond to the major clusters in Figure 2. Cluster 1 contains Tsk2/+ (16 weeks), Tsk1/+, and the limited SSc samples. Cluster 2, contains the inflammatory subset, sclGVHD mouse and bleomycin samples. Cluster 3 contains the normal-like group and no mouse samples. Cluster 4 contains the fibroproliferative samples and Tsk2/+ at 4 weeks of age.
150x147mm (300 x 300 DPI)



Supplemental Figure S2. TGF-β responsive gene expression in the bleomycin-induced skin fibrosis model. A. GSEA normalized enrichment score (NES) for the TGF-b signature derived from the pump model shown in Figure 4 using either all genes, or only genes with increased or decrease expression. B. Detailed list of genes from the TGF-b pump model and their expression in the bleomycin-induced skin fibrosis model.
175x270mm (300 x 300 DPI)