AD_____

Award Number: W81XWH-07-1-0277

TITLE: The Integrative Studies of Genetic and Environmental Factors in Systemic Sclerosis

PRINCIPAL INVESTIGATOR: Dr. Xiaodong Zhou

CONTRACTING ORGANIZATION: University of Texas Health Science Center
Houston, TX 77030

REPORT DATE: May 2011

TYPE OF REPORT: Final

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

# REPORT DOCUMENTATION PAGE

*Form Approved*
*OMB No. 0704-0188*



| 1. REPORT DATE *(DD-MM-YYYY)* | 2. REPORT TYPE | 3. DATES COVERED *(From - To)* |
|---|---|---|
| 01-05-2011 | Final | 9 APR 2007 - 8 APR 2011 |

**4. TITLE AND SUBTITLE**
The Integrative Studies of Genetic and Environmental Factors in Systemic Sclerosis

**5a. CONTRACT NUMBER**

**5b. GRANT NUMBER**
W81XWH-07-1-0277

**5c. PROGRAM ELEMENT NUMBER**

**6. AUTHOR(S)**
Dr. Xiaodong Zhou

E-Mail: Xiaodong.Zhou@uth.tmc.edu

**5d. PROJECT NUMBER**

**5e. TASK NUMBER**

**5f. WORK UNIT NUMBER**

**7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)**
University of Texas health Science Center
Houston, TX 77030

**8. PERFORMING ORGANIZATION REPORT NUMBER**

**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**
No abstract provided.

**15. SUBJECT TERMS**
Scleroderma (SSc), fibroblasts, fibrosis, silica, environmental particles, susceptibility

| 16. SECURITY CLASSIFICATION OF: | | | 17. LIMITATION OF ABSTRACT | 18. NUMBER OF PAGES | 19a. NAME OF RESPONSIBLE PERSON |
|---|---|---|---|---|---|
| a. REPORT | b. ABSTRACT | c. THIS PAGE | | | USAMRMC |
| U | U | U | UU | 98 | 19b. TELEPHONE NUMBER *(include area code)* |

## Table of Contents

**Introduction:**

This project aims to study interactions between genetic and environmental factors in a viable system - human fibroblasts. Fibroblasts obtained from scleroderma (SSc) patients have a profibrotic nature that suggests a possibility of a dysregulation of biological function in the cells. On the other hand, SSc occurs in genetically susceptible individuals. A SSc susceptible genetic background may be more vulnerable to environmental triggers, such as silica – a potential SSc trigger. Studies of biological functions of fibroblasts with and without SSc susceptible backgrounds in response to potential environmental triggers will provide a great opportunity to understand etiopathogenesis of SSc.

**Body:**

There are two specific aims proposed in original proposal: 1) To determine whether human fibroblasts with genetic backgrounds for SSc susceptibility are more sensitive to SSc risk particles particularly silica particles. 2) To determine which specific pathways are triggered by risk elements. During this grant funding period, we have been following these aims in the studies.

A. ***Establishment of individual fibroblast strains:*** For specific aim 1, we have established fibroblast strains of 102 SSc patients and 128 normal controls from individual skin biopsies (Table 1). The number of fibroblast strains exceeds the proposed minimum of 100 of each SSc and control fibroblast strains.

**Table 1.** Summary of cultured fibroblasts bank obtained from skin biopsies of SSc patients and controls

| Gender/Ethnicity | Controls | SSc |
|---|---|---|
| Female | 80 | 78 |
| Male | 48 | 24 |
| Caucasian | 64 | 60 |
| African American | 35 | 16 |
| Asian | 8 | 2 |
| Hispanic | 20 | 20 |
| Choctaw | | 3 |
| Unknown | 1 | 1 |
| Total | 128 | 102 |

B. ***Establishment of genetic background of SSc:*** The genetic background of SSc identified at the time of this original proposal included only HLA genes (susceptibility to SSc with HLA-DRB1*11, DQB1*0301, DQA1*0501; protection from SSc with HLADQA*0201). Identifying additional genetic susceptibility genes and loci to SSc will help us in determining selection of genetic background for SSc, although it was not proposed in original grant. Collaborating with Dr. Eun Bong Lee (Seoul National University College of Medicine, Seoul, Korea), we performed and published the first genome-wide association study (GWAS) in SSc that demonstrated specific single-nucleotide-polymorphisms (SNP)s of the HLA-DPB1 and -DPB2 were strongly associated with SSc susceptibility in both Koreans and US Caucasians [published paper 1]. We also identified that the HLA-DPB1*1301 was the most significantly associated loci to SSc in patients with anti-topoisomerase I autoantibodies. Importantly, this study suggested that SSc should not be considered as a single disease. Sub-classification of SSc on the basis of autoantibodies presented in patients appeared better in identifying susceptibility genes and loci [published paper 1].

C. ***Establishment of associations between SSc genetic background and cellular response to environmental trigger*:** To determine whether fibroblasts obtained from SSc patients carrying SSc susceptibility loci or genes are more sensitive and vulnerable for fibrotic responses to environmental trigger, we examined 200 fibroblast strains in cultures with and without adding silica particles (It was proposed in the original proposal as a major experiment). Silica particles were added into the cultures with 5 different doses for 24-hour stimulation, and a low middle dose was used for a time-course stimulation (5 time points). Gene expressions of 6 important extracellular matrix (**ECM**) components including COL1A2, COL3A1, CTGF, MMP1, MMP3 and TIMP3 in the fibroblasts were measured for monitoring fibroblast response to silica stimulations. The real-time RT-PCR was used for absolute quantitation of gene expression.

For genetic information of SSc patients and controls who were skin-biopsied for the fibroblast strains, **we used both SNP typing data from the "Immunochip" and standard HLA allele typing.** It is worth noting that Wellcome Trust Case-Control Consortium-initiated "Immunochip" platform contains a higher density of all significant genome-wide association study (GWAS) loci (about 200,000 SNPs) from a series of studies of immune-mediated diseases. It also contains high density of SNPs in the HLA region and the latest SNP information discovered by sequencing subjects of the 1,000 genomes project.

For data analysis, we use R (2.13.1) to do the data cleansing of 200 subjects Q-RT-PCR gene expression data on six genes. For every time point among 1Day, 2Day, 3Day, 4Day and 5Day, we calculate the log2ratio of absolute quantity(AQ) level between time specific Control and Silica stimulated samples. For Immunochip genotyping data of the same 200 subjects, we use R and PLINK v1.07 to do data cleansing and recode the SNPs genotypes to uniform format: 2 represents minor allele homozygous, 1 represents heterozygous, 0 represents major allele homozygous under genetic additive model assumption. Minor allele type for each SNP locus was recorded as well. Then we use "FDA" this R package by J. O. Ramsay, etc to implement non-parametric functional data analysis on PCR vs. SNPs data. We use cubic B-spline basis to model the longitudinal gene expression data (f(t)), smoothed by optimal penalty term λ selected by GCV process. Through observing the data fit plots, we conclude that the gene expression across five time points are smoothed fitted, which in turn demonstrated that our modeling of the gene expression data is appropriate; on the other hand, we treated the genotype of each SNP as continuous variable (0,1,2) in the functional regression model scalar covariate part ($X_i$). Together with previously modeled functional response variable, we have the function regression model like:

$$f(t) = \beta_0 + \beta_1 X_i$$

If we can test the significance of $\beta_1$, we can find whether or not at this SNP locus the genotype difference from 200 subjects leads to the difference gene expression level and different gene expression change profile across five time points. To fulfill this purpose, we use "Fperm.fd" (Fperm) this function within "FDA" package to do permutation F test on PCR vs. SNPs data, which is a reliable statistical test to render a legitimate significance P-value. Since the covariates contain only intercept and scalar variable, we adopted the "FDA" tutorial to apply constant functional data parameters to the covariates. Since the immune-important regions such as HLA/MHC are on Chromosome 6, at this point we focused our functional permutation F test only on Chromosome 6 SNPs. After filtering out the mis-genotyped SNPs genome-widely, we have in total 178,007 valid SNPs out of 196,517 as total. Then we use the Immunochip annotation file to select those SNPs on Chromosome 6 to be our test candidates. The number of these SNP candidates is 17,042 in total on Chr6. To adjust the multiple test error, for most conservative consideration, if using bonferroni correction method, we will need 1/(0.05/17042) =340,840, rounded up to 350,000 permutation number for one Fperm test on one gene vs. one snp. Due to extremely intensive computation cost, we plan to execute the adaptive permutation test strategy: first we start from a preliminary Fperm test on all the candidate SNPs at the number of

permutation = 200, then we select those with P-value equates 0, which means that the probability of finding an equal or more extreme case than observed is at least less than 1/200 = 0.005 under null hypothesis – this SNP doesn't associate with gene expression level change profile. Once we have the preliminary significant SNPs list for all the seven genes under investigation, we collect their annotation information such as located gene region, coding or intergenic, extron or intron, etc. We did the gene expression level change Table and plots based on three SNPs group (sometimes two because of lacking one group genotyped) for every preliminary significant SNPs. Table 2 displayed the associations between genetic polymorphisms (SNP) and longitudinal gene expressions of 5 ECM genes of fibroblasts in response to silica stimulation. Figure 1 displayed examples of the association plots, in which in response to silica stimulation, the COL1 expression appears higher at early stage and toward end in the fibroblast carrying minor allele ( C ) of rs9275652.

**Table 2.** Associations between genetic polymorphisms of HLA genes and longitudinal gene expression of ECM components of fibroblasts in response to silica stimulation.

| COL1A2 expression | Susceptibility gene | position | SNP ID | p-value |
|---|---|---|---|---|
| | HLA-DQB1 \| HLA-DQA2 | INTERGENIC | rs9275334_G | $< 10^{-7}$ |
| | HLA-DQB1 \| HLA-DQA2 | INTERGENIC | rs9275652_C | $< 10^{-7}$ |
| | HLA-DQB1 \| HLA-DQA2 | INTERGENIC | rs9275660_A | $< 10^{-7}$ |
| | HLA-DQB1 \| HLA-DQA2 | INTERGENIC | rs9275936_A | $< 10^{-7}$ |
| | HLA-DQB1 \| HLA-DQA2 | INTERGENIC | rs9276171_G | $< 10^{-7}$ |

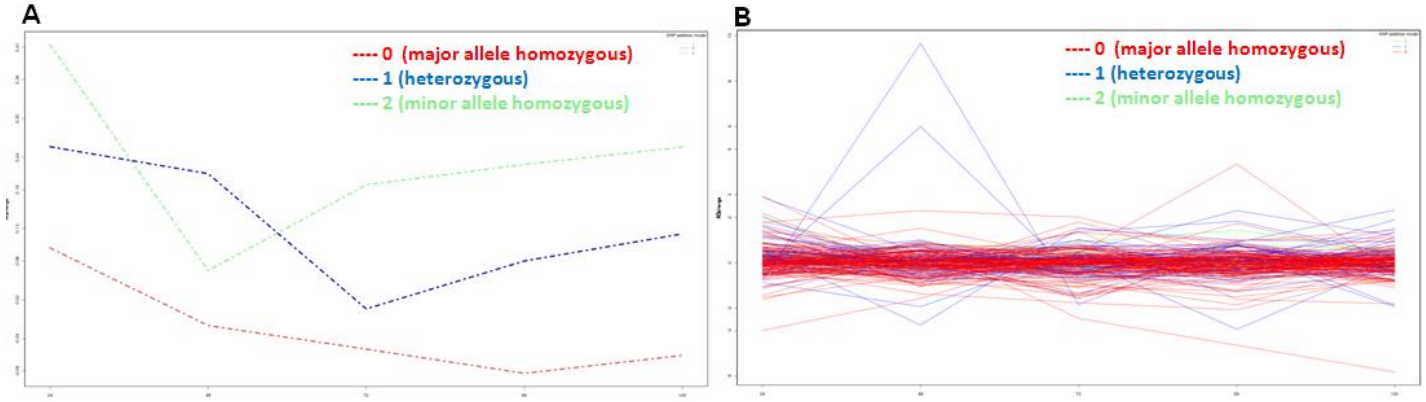
| COL3A1 expression | Susceptibility gene | position | SNP ID | p-value |
|---|---|---|---|---|
| | HLA-DPB1 | INTRON | rs2567279_G | $< 10^{-7}$ |
| | HLA-DPB1 | INTRON | rs2856819_A | $< 10^{-7}$ |
| | HLA-DPB1 \| HLA-DPB2 | INTERGENIC | rs3117230_G | $< 10^{-7}$ |
| | HLA-DPB1 \| HLA-DPB2 | INTERGENIC | rs3130192_A | $< 10^{-7}$ |

| MMP1 expression | Susceptibility gene | position | SNP ID | p-value |
|---|---|---|---|---|
| | HLA-DQB1 \| HLA-DQA2 | INTERGENIC | rs2856705_A | $< 10^{-7}$ |
| | HLA-DQB1 \| HLA-DQA2 | INTERGENIC | rs2858308_A | $< 10^{-7}$ |
| | HLA-DQB1 \| HLA-DQA2 | INTERGENIC | rs3828796_G | $< 10^{-7}$ |
| | HLA-DQB1 \| HLA-DQA2 | INTERGENIC | rs3916765_A | $< 10^{-7}$ |
| | HLA-DQB1 \| HLA-DQA2 | INTERGENIC | rs6936863_C | $< 10^{-7}$ |
| | HLA-DPB1 \| HLA-DPB2 | INTERGENIC | rs9380343_A | $< 10^{-7}$ |

| MMP3 expression | Susceptibility gene | position | SNP ID | p-value |
|---|---|---|---|---|
| | HLA-DQB1 \| HLA-DQA2 | INTERGENIC | rs3828796_G | $< 10^{-7}$ |

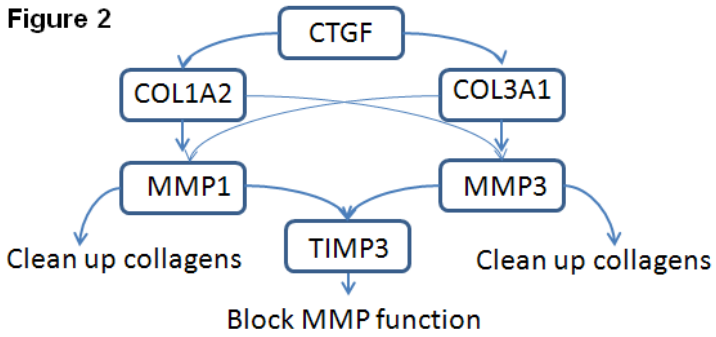
| TIMP3 expression | Susceptibility gene | position | SNP ID | p-value |
|---|---|---|---|---|
| | HLA-DPB1 | INTRON | rs2071351_G | $< 10^{-7}$ |
| | HLA-DOA \| HLA-DPA1 | INTERGENIC | rs435119_G | $< 10^{-7}$ |
| | HLA-DOA \| HLA-DPA1 | INTERGENIC | rs443623_A | $< 10^{-7}$ |
| | HLA-DOA \| HLA-DPA1 | INTERGENIC | rs6457710_G | $< 10^{-7}$ |

**Figure 1.** SNP rs9275652 of HLA-DQB1 vs. expression of COL1A2 of fibroblasts in time-course (y = gene expression levels in log, X = time - 1, 2, 3, 4 and 5 days). A: association with average gene expression level B: association with gene expression data from all assays.




- Explanation for examining six ECM genes: Accumulation of the ECM components is a feature of fibrosis. CTGF is a profibrotic growth factor that is usually activated by fibrotic stimuli. It can up-regulate collagen genes e.g. COL1A2 and COL3A1, which are major ECM structure components. In normal situation, increased collagens can induce matelloproteinases (MMP) expression that function in cleaning up over-expressed collagens. On the other hands, over-production of MMP induces expression of tissue inhibitor of metalloproteinases (TIMP) to block MMP function. Figure 2 illustrates the relationship among six genes.

**Figure 2**




For standard HLA typing, HLA-DRB1, DQA1, DQB1 and DPB1 were typed by standard oligotyping techniques using primers and probes recommended by the 13$^{th}$ International Histocompatibility Testing Workshop (held in Victoria, Canada, May 2002) (Hansen and Dupont, 2004) with high resolution DRB1 typing further achieved by nucleotide sequence analysis of PCR-amplified DRB1 exon 2. Examining association between specific HLA allele and gene expression changes of fibroblasts in response to silica stimulation, we found that SSc susceptibility loci including HLA-DRB1*11, DQB1*03 and DPB1*1301 are associated with specific gene expression patents in the fibroblasts. Figure 3 and 4 displayed examples in time-course and dose response, respectively. In Figure 3, compared to the fibroblast strains of non-carriers of DRB1*11 in SSc patients and controls, the fibroblast strains of SSc patients carrying DRB1*11 (or DRB1*11 positive patients) showed a low response levels of MMP1 and MMP3 genes in the time-course experiment (p-values $< 0.05$). In contrast, the expressions of COL1A2 and COL3A1 were higher after 3-day of cultures ($p < 0.05$). The TIMP3 appeared to be unstable (up and down through 5 days of cultures). In figure 4 (dose-response), both MMP1 and MMP3 showed a low response to silica stimulations in DRB1*11

positive SSc patients ($p < 0.05$), while COL1A2 showed a higher response at dosage over 10 ug. In addition to HLA-DRB1*11, we also examined other SSc susceptibility alleles of HLA genes. We attached the results in appendices.

We data from both SNPs and SSc susceptibility alleles of HLA genes indicated that SSc susceptibility loci of HLA genes are associated with cellular response to silica stimulation. In particular, MMP1 and MMP3 genes appeared to be less responsive while some ECM genes, such as COL1A2 and/or COL3A1, were up-regulated. We are preparing the manuscript from these studies.

**Figure 3.** Association between HLA-DRB1*11 and time-course response of gene expression of fibroblasts to silica stimulation

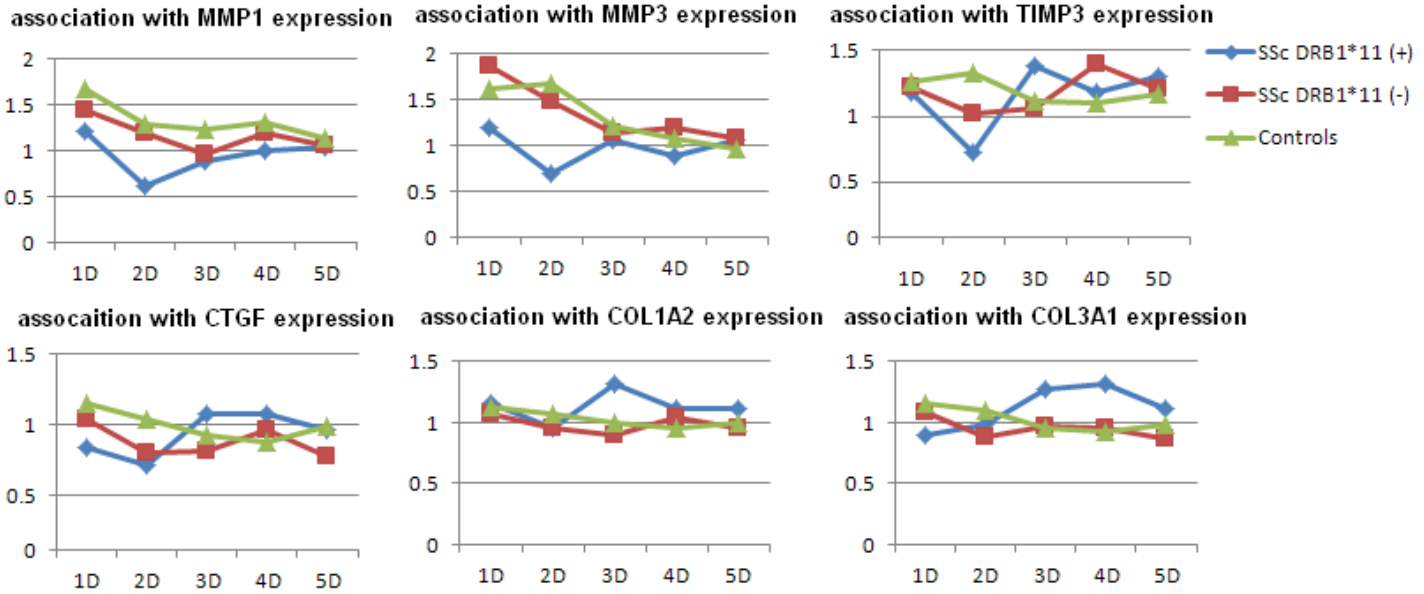
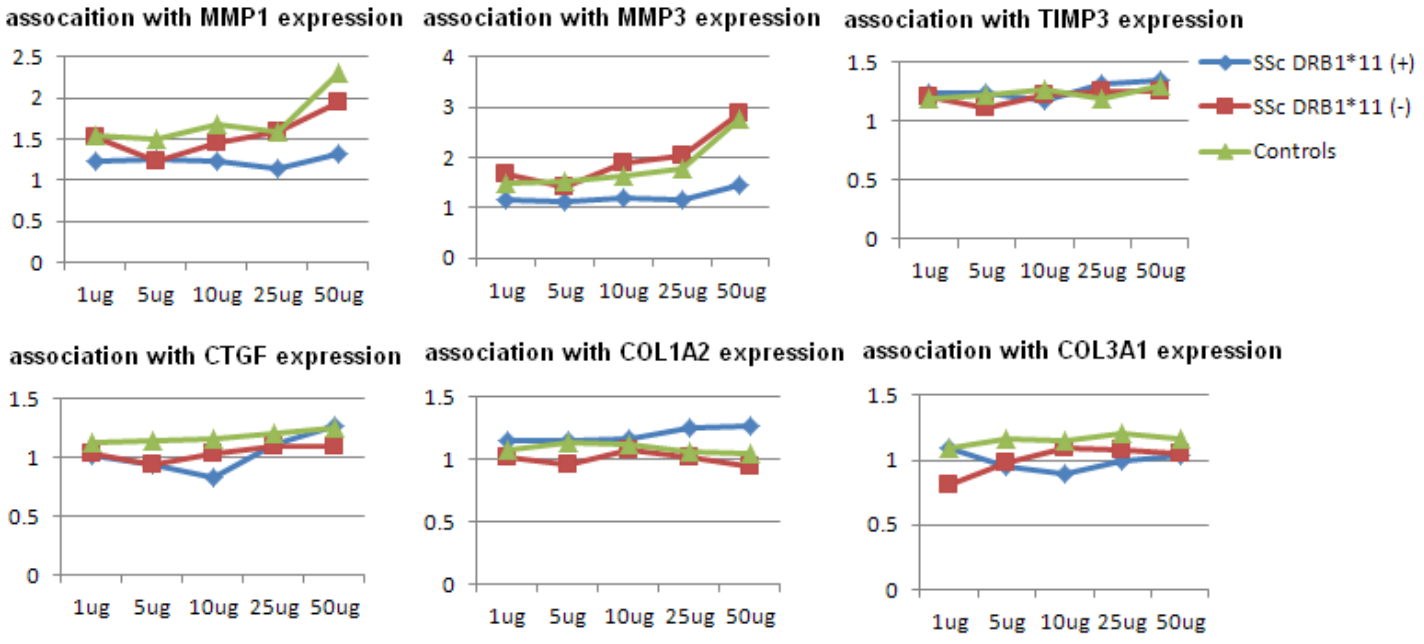



Figure 4. Association between HLA-DRB1*11 and dose response of gene expression of fibroblasts to silica stimulation.



### *D. Examination of potential bio-pathways involved in silica stimulation:*

Previously, we identified that silica stimulation on fibroblasts trigged profibrotic TGF-β signaling. To explore intrinsic dynamic properties of SSc fibroblasts in response to silica stimulation, we applied the developed state-space models in the studies, and performed dynamic analysis of partial TGF-β pathway in both SSc and normal fibroblasts stimulated by silica [published paper 2]. We employed widely used state-space equations in systems science to model biological systems, and use expectation-maximization (EM) algorithms and Kalman filter to estimate the parameters in the models. We found that TGF-β pathway under perturbation of silica showed significant differences in dynamic properties between SSc and normal fibroblasts. Particularly, the gene network of TGF-β in responding to perturbation of silica is relatively unstable in the SSc fibroblasts. In addition, although the TGF-β gene expression network responding to silica in both normal and SSc fibroblasts is controllable, this regulatory network in the SSc fibroblasts showed a low degree of controllability [published paper 2]. These findings may open a new avenue in exploring the functions of cells and mechanism operative in disease development.

In addition to examining direct contact between environmental stimuli and cultured human fibroblasts, we also examined the activation of fibroblasts by silica particles through macrophages/monocytes and T cells. Macrophages are early responding cells in up-taking pathogenic materials to activate lymphocytes. T lymphocytes are central players in cell-mediated immunity, which usually stimulate target cells such as fibroblasts. Therefore, the combination of human macrophages, T cells and fibroblasts provide a live bio-system from the human body for the studies of the potentially hazardous effects of environmental particles. In this study, we first stimulated macrophages/monocyte with silica particles, as well as carbon nanotubes (CNTs) (a potential inflammatory trigger) or titanium particles and PBS for control. After 24-hour culture, stimulated cells were mixed with T cells and then co-cultured with fibroblasts. Through monitoring live cells with a digital camera on a microscopy, we observed the response of macrophages/monocytes to different particles. Within the first 30 minutes of stimulation, macrophages/monocytes started to move toward CNT particles. At time point of 24-hour stimulation, CNTs were heavily surrounded by macrophages/monocytes, while silica particles that appeared smaller than CNTs were surrounding the macrophages/monocytes (Figure 5). In contrast, titanium particles did not such changes. After stimulation, the cultures of macrophages/monocytes were mixed with T cells and fibroblasts. Within the first 30 minutes of co-cultures, the fibroblasts did not show significantly morphological changes. However, after 24 hours, microscopic examination showed deformed fibroblasts around the CNTs (Figure 6).

In both silica and CNTs stimulation, IL1α and IL1β were significantly increased in the culture medium at 1-hour time point after addition of stimulated macrophage/monocytes and T cells into cultured fibroblasts. Increased gene expression of the COL1A2 was followed in cultured fibroblasts at 24-hour time point. IL1α and IL1β are important pro-inflammatory cytokines that may trigger a variety of cellular responses, such as fibrosis, apoptosis, and proliferation. Increased levels of IL1α and IL1β in the culture medium may come from stimulated macrophages/monocytes that are usually the major source of inflammatory cytokines. A down-regulation of the IL1B gene observed in the cultured fibroblasts may be a feedback response. Up-regulated the COL1A2 in cultured fibroblasts is likely triggered at least in part by IL1α and IL1β in the cultures, and which indicates a potential fibrotic response.

Different from silica stimulation, IL8 also was significantly increased in early culture medium (1-hour time point) of fibroblasts with CNTs stimulated macrophages/monocyte and T cells. IL8 is a chemokine that attracts inflammatory cells at the site of inflammation. Concordantly, live-microscope examination showed that macrophages/monocytes aggregated at the site of CNTs in the cultures of either macrophages/monocytes alone or with fibroblasts. This change was not observed in silica

and titanium stimulation. Compared to silica and titanium particles, a bigger size of CNTs may affect cellular responses in cultured cells, which was reported in studies of CNTs.

An early mild upregulation of the CTGF gene in cultured fibroblasts after addition of silica stimulated macrophages/monocytes and T cells at 1-hour time point is distinct from the CNTs stimulation. CTGF is a profibrotic cytokine that may induce collagen expression and deposition in fibrotic diseases. Therefore, a later response of the COL1A2 in the fibroblasts at 24-hour time point also may be triggered by CTGF signaling. Therefore, silica stimulation may induce both IL1 and CTGF signaling in cultured human cells. We have submitted this study for publication [see submitted manuscript].

**Figure 5.** Cultures of macrophages/monocytes with different stimuli at 24-hour time point: A: with PBS; B: with CNTs for 24 hours; C: with silica particles; D: with titanium particles. *Arrows indicate particles.

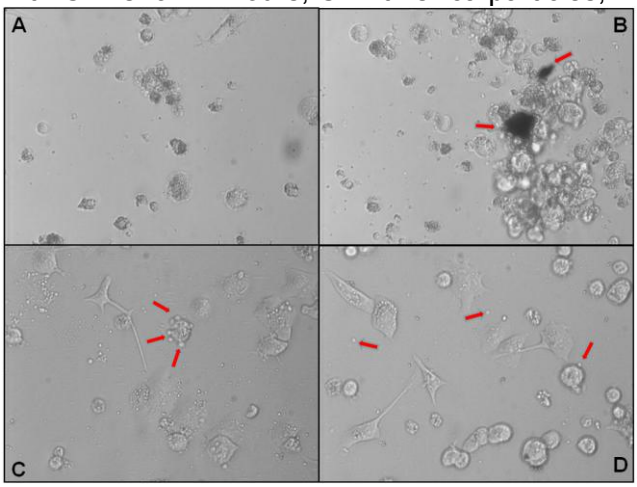
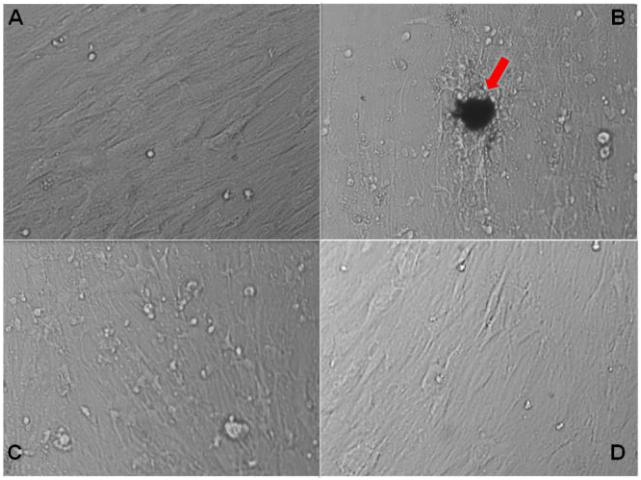



**Figure 6.** Cultured Fibroblasts with stimulated macrophages/monocytes and T cells at 24-hour time point. A: Fibroblasts cultured with PBS stimulated macrophages/monocytes and T cells; B: Fibroblasts cultured with CNT stimulated macrophages/monocytes and T cells; C: Fibroblasts cultured with silica stimulated macrophages/monocytes and T cells; D: Fibroblasts cultured with titanium stimulated macrophages/monocytes and T cells. Arrow indicates a CNT.




Further in the studies of bio-pathways associated with SSc pathogenesis, we continued to explore attenuation of fibrosis through TGF-β pathway which can be induced by silica particles according to our previously reports. We applied gene specific siRNAs of SPARC and/or CTGF to attenuate fibrotic changes in the fibroblasts obtained from either SSc skin or TGF-β transgenic mice. Both SPARC and CTGF are extracellular matrix proteins. We also treated with these siRNAs in fibrotic mouse model induced by bleomycin that is considered as another environmental trigger for scleroderma (stated in original grant). Our results indicated that SPARC siRNA significantly reduced

gene and protein expression of collagen type I, as well as collagen content in the fibroblasts [published paper 3]. Our *in vivo* studies showed that skin and lung fibrosis of the mice induced by bleomycin was markedly reduced by treatment with Sparc siRNA [published paper 3].

In addition to these two fibrotic models, we also examined anti-fibrotic effects of SPARC siRNA and CTGF siRNA in the CTGF transgenic model since the CTGF is a down-stream gene in TGF-β pathway and contribute to persistent signaling for fibrosis. Our results showed that inhibition of Sparc or Ctgf expression by their corresponding siRNA in cultured fibroblasts of *CTGF* transgenic mice down-regulated the expression of collagen type I. Sparc and Ctgf siRNAs also showed a reciprocal inhibition at transcript levels, but Sparc siRNA functioned more efficiently than Ctgf siRNA in reducing the protein level of both Sparc and Ctgf [published paper 4].

Silica exposure has been linked to anti-nuclear autoantibodies (ANA) and other autoantibodies in SSc (*Haustein UF, Ziegler V et al. J Am Acad Dermatol. 1990;22:444-8*). Topo I is an important nuclear protein that catalyzes the breaking and joining of DNA strands and controls DNA replication and transcription. We examined whether and how the catalytic function of topo I is changed in SSc fibroblasts. Our studies indicated that topo I molecules were altered in their function with relocation in the nucleus (from nucleolus to nucleoplasm). In some fibroblasts, especially those obtained from skin biopsies of SSc patients who were positive for anti-topo I or anti-RNA polymerase III autoantibodies, these alterations were associated with increased sumoylation of topo I, which may facilitate relocation of topo I molecules. In contrast, the fibroblasts of anti-centromere positive patients showed unchanged sumoylation of topo I. Inhibition of SUMO1 gene with SUMO1 siRNA improved catalytic function of topo I in SSc fibroblasts. These observations may provide important insights into the nature of SSc fibroblasts that may contribute to pathological processes and/or disease development in SSc [published paper 5].

It is worth noting that we have generated a huge amount of data based on genetic typing information from GWAS and HLA allele typing, as well as cellular functional data on cultures of 200 fibroblast strains in response to silica stimulation in dose and time course. Our biostatistician is still working hard to extract more and more important and novel information from the data. We expect to publish more papers from it very soon.

**Key Research Accomplishments**

- We have obtained a total of 230 human fibroblast strains (102 SSc patients and 128 normal controls). These primary fibroblasts have broad applications in studies of SSc etiopathogenesis and in developing novel strategies for personalized therapies with known genetic backgrounds.
- We have completed silica stimulation in 200 fibroblast strains and obtained RNA and protein extracts from each of the experiments.
- We have completed genotyping of these 200 fibroblast strains for Immunochip containing 200K SNPs of HLA and other genes involved in susceptibility to multiple immune-associated diseases. We also completed standard HLA allele typing.
- We identified HLA-DPB1 and –DPB2 as major genetic factors for SSc patients with anti-topo I autoantibodies.
- We analyzed the data generated from molecular studies of fibroblasts in response to silica stimulations, as well as from genetic studies of SSc patients and controls. We identified the associations between SSc susceptibility loci and fibroblast responses to silica stimulations.

- We applied the developed state-space models in the studies, and performed dynamic analysis of partial TGF-β pathway in both SSc and normal fibroblasts stimulated by silica. We

demonstrated that the gene network of TGF-β in responding to perturbation of silica is relatively unstable in the SSc fibroblasts, and have a low degree of controllability.

- We identify that both CTGF and IL1 signaling may be involved in silica stimulation of human cells, and CNTs also appeared harmful to human cells, in which inflammation may be the major pathologic change.
- We demonstrated that specific inhibition of SPARC and/or CTGF with corresponding siRNA reduced collagen expression in TGF-β activated fibroblasts, and attenuated mouse fibrosis induced by bleomycin (an environmental factor for SSc) *in vivo*.
- We identified that catalytic function of topoisomerase I (topo I) was decreased in SSc fibroblasts, which appeared to be associated with increased sumoylation of topo I. Inhibition of sumoylation of topo I improved the topo I function. This novel finding may facilitate studies of genetic nature of SSc fibroblasts contributing to disease pathogensis.

**Reportable Outcomes**

During this grant period, we published seven papers, presented five abstracts and completed one manuscript, and we are preparing one more manuscript (see the list below).

We established a total of 230 human fibroblast strains (102 SSc and 128 controls), which have broad applications in studies of SSc etiopathogenesis and in developing novel strategies for personalized therapies with known genetic backgrounds.

This project helped to train three postdoctoral fellows including Drs. Jiucun Wang, Wei Lin and Khurshida Begum. Dr. Jiucun Wang (PhD), completed her training and accepted a faculty position in Fudan University, Shanghai, China.

*Published articles (underline for corresponding author):*


1. **Zhou XD**, Lee JU, Arnett FC, Xiong MM, Park MY, Yoo YK, Shin ES, Reveille JD, Mayes MD, Kim JH, Song R, Choi JY, Park JA, Lee YJ, Lee EU, Song YW, Lee EB. HLA-DPB1 and DPB2 are culprit genetic loci for susceptibility to systemic sclerosis, Genome-wide association study in Koreans with Replication in North Americans. Arthritis Rheum. 2009,30;60(12):3807-3814. **PMID: 19950302**
2. Xiong M, Arnett FC, Xiong H and **Zhou XD**. Differential Dynamic Properties of Scleroderma Fibroblasts in Response to Perturbation of Environmental Stimuli -Application of State-Space Model in Studies of human Complex Disease. PLoS ONE, 2008 Feb, 3(2):e1693. **PMID: 19301770**
3. Wang JC, Lai SL, Guo XJ, Zheng XF, de Crembrugghe B, Arnett FC and **Zhou XD**. Attenuation of fibrosis in vitro and in vivo with Sparc siRNA. Arthritis Research & Therapy, 2010 Apr 1;12(2):R60. **PMID: 20359365.**
4. Wang JC, Sonali S, Arnett FC, **de Crombrugghe B, Zhou XD**. Attenuation of gene expression of extracellular matrix elements with siRNAs of Sparc and Ctgf in cultured skin fibroblasts from Ctgf transgenic mice. International Journal of Immunopathology and Pharmacology. 2011 Jul-Sep;24(3):595-601.**PMID:21978691**
5. **Zhou XD**., Lin W., Tan F.K., Assassi S., Fritzler M., Guo XJ., Sharif R., Xia T., Lai S., Arnett, F.C. Altered Sumoylation and Function of DNA Topoisomerase I in Scleroderma Fibroblasts. Arthritis Research & Therapy 2011 Aug 9;13(4):R128. **PMID: 21827649**.
6. Sharif R., Fritzler MF., Mayes M., Gonzalez EB., McNearney TA, Draeger H., Baron M., Furst D. Khanna D., Molitor J., Pope J., Schiopu E., Seibold J., Silver R. Simms R., **Zhou XD**., Perry P., Rojo R., Charles J., Junco DJ., Agarwal SD., Reveille JD., Assassi., Arnett FC. Anti-Fibrillarin Antibody in African American patients with Systemic Sclerosis: Immunogenetics, Clinical Features, and Survival Analysis**.** J. Rheumatology 2011 August**.** PMID:21572159.

7. Momiao Xiong, Gang Peng , Li Luo , Yun Zhu , Pengfei Hu , Shengjun Hong , Jinying Zhao , **Zhou XD**, John Reveille, Li Jin , Christopher Amos. Gene and Pathway-Based Analysis - Second Wave of Genome-wide Association Studies. European Journal of Human Genetics. 2010 Jan;18(1):111-7. **PMID:19584899**.

*Manuscript submitted (see attachment)*

1. Guo XJ, Janannath C, Espitia MG and **Zhou XD**. Up-take of silica and carbon nanotubes by human macrophages/monocytes induces activation of fibroblasts in vitro – potential implication for pathogenesis of inflammation and fibrotic diseases

*Manuscript in preparation*

1. Wei P, Yang Y, Tan FK, Arnett FC, **Zhou XD**. Studies of genetic and environmental factors in human fibroblasts showed gene-environmental interactions.

*Abstract:*

1. **Zhou XD**, Tan FK, Guo XJ, Tejpal N, Kahan BD, Stepkowski SN, Arnett FC. Up-take Of Silica And Carbon Nanotubes By Human Macrophages Induces Activation Of T Cells And Fibroblasts *In Vitro* - Potential Implication For Pathogenesis Of Fibrosing Diseases. Arthritis & Rheuma 2007;56 (9);s245. supplement for ACR annual meeting.
2. Xiong MM, Guo X, Xiong H, Arnett FC and **Zhou XD**. Differential Dynamic Properties of the TGFB Pathway between Normal Fibroblasts and Scleroderma Fibroblasts in Response to Perturbation by Environmental Stimuli. Arthritis & Rheuma 2007;56 (9);S74; supplement for ACR annual meeting.
3. Wang J, Lai S, Sonnylal S, Arnett F.C., Crombrugghe B and **Zhou XD**. Application of Sparc and Ctgf siRNAs in fibrotic murine models of scleroderma (SSc) *in vitro* and *in vivo.* Arthritis & Rheuma 2008;58 (9); supplement for ACR annual meeting.
4. Momiao Xiong, Li Luo, Yu-li Lin, Frank Arnett, Wenyaw Chan and **Xiaodong Zhou.** Single and multi-level multivariate longitudinal models in the genetic studies of time course gene expressions of human fibroblasts in response to environmental stimulation. The American Society of Human Genetics, 59th Annual Meeting 2009.
5. Jiu-Cun Wang, Xinjian Guo, Liming Li, Wensheng Guo, Momiao Xiong, Filemon.k.Tan, Frank C. Arnett, and **Xiaodong Zhou**. Association studies of genetic variants to the functions of human fibroblasts in response to environmental stimulation. The American Society of Human Genetics, 60th Annual Meeting 2010.

**Conclusion**

During this grant period, we established a large number of primary fibroblast strains from normal controls and SSc patients. We performed stimulation assays with silica in 200 primary fibroblast strains. Our current results showed that different fibroblast strains obtained from different individuals responded differently to silica stimulation in terms of the gene expression of the ECM components that are involved in activation of fibrosis.

Using both single and multi-level multivariate longitudinal linear models in analysis of association between specific genotypes and dynamic changes of gene expression of the fibroblasts in responding to silica stimulation, we identified associations between multiple genotypes (SNPs) of HLA genes that confer susceptibility to SSc associated and the expressions of collagen genes and other ECM genes (e.g. COL1A2, COL3A1, MMP1, MMP3 and TIMP3). Fibroblast strains with SSc susceptibility loci (e.g. HLA-DRB1*11) showed less MMP1 and MMP3 responses, while collagen gene expressions were up-regulated. These observations supported our original proposal that genetic elements within SSc fibroblasts might contribute to susceptibility to fibrotic process. Integrative studies of genetic and environmental factors with human fibroblasts may facilitate the discovery of potential pathogenesis of SSc.

In studies of bi-pathway associated with SSc, in addition to previously identified TGF-signaling, we identified that silica stimulation may induce both CTGF and IL1 signaling, while CNTs, another common environmental particles may trigger inflammation through IL1 and IL8 signaling.

We also demonstrated that silencing SPARC and/or CTGF attenuated fibrotic changes *in vivo* and *in vitro* induced by TGF-β signaling (can be induced by environmental element - silica) and/or bleomycin (another environmental trigger for SSc). We also identified that catalytic function of topo I was decreased in SSc fibroblasts, and which appeared to be associated with increased sumoylation of the topo I. Inhibition of SUMO expression improved the topo I function in SSc fibroblasts. These novel observations provided us a potential mechanism underlying dysfunction of SSc fibroblasts. Therefore, our studies are fulfilled with original proposal in the grant.

It is worth noting that our data generated from the studies contain huge amount information that may be explored further. Our biostatistician is still working hard to extract more and more important and novel information from the data. We expect to publish more papers from it.

**Appendices**

1. **Submitted manuscript**
2. **Association between SSc susceptibility HLA alleles and gene expression pattern of fibroblasts in response to silica stimulations.**

# Up-take of silica and carbon nanotubes by human macrophages/monocytes induces activation of fibroblasts in vitro – potential implication for pathogenesis of inflammation and fibrotic diseases

Xinjian Guo[1], Chinnaswamy Jagannath[2], Maribel G Espitia[1], Xiaodong Zhou[1*]

[1]University of Texas Medical School at Houston, Internal Medicine, Houston, TX; [2]University of Texas Medical School at Houston, Pathology and Laboratory Medicine, Houston, TX

[*]Corresponding: Xiaodong Zhou (Xiaodong.zhou@uth.tmc.edu)

University of Texas Health Science Center at Houston
Department of Internal Medicine/Division of Rheumatology and Clinical Immunogenetics
6431 Fannin St.
Office: MSB 5.270
Houston, TX 77030
Phone: 713-500-6088
Fax: 713-500-0580

Key words: Silica, Carbon Nanotubes, Cytokines, CTGF, inflammation, fibrosis

**Abstract**

The potential pathogenic effects of silica and carbon nanotubes (CNTs) on fibroblasts, macrophages/monocytes, and T cells were investigated. Human macrophage/monocytes were cultured and stimulated with silica, CNTs, or titanium particles. After adding human T cells to the stimulated macrophages/monocytes, the cells were added to cultured human fibroblasts. Upon microscopic examination, CNT stimulation after 24 hours showed centralization of macrophages/monocytes around the CNTs. Silica stimulation showed a significant increase of IL1α and IL1β in cultured medium, and an increased gene expression of CTGF in cultured fibroblasts at 1 hour, as well as an up-regulation of the COL1A2 gene at 24-hour time point. In addition to the same changes of IL1α, IL1β and the COL1A2 by silica, CNT stimulation showed an increase of IL8 in cultured medium at 1-hour time point. Titanium stimulation yielded no significant changes. The results indicate a proinflammatory and/or profibrotic effect of silica and CNTs to cultured human cells including macrophages/monocyte, T cells and fibroblasts.

**Introduction**

Environmental factors influence human life to different extents. Silica and carbon nanotubes (CNTs) are two common particles that widely exist in our living environment. Both particles may be small enough to enter the human body through the respiratory system or by directly contacting human skin. Exposure to silica frequently occurs in people whose occupations are involved in stone quarries, mineral and coal mining, glass, ceramic and metal manufactures. Silica exposure has been associated with fibrosis and autoimmune diseases. CNTs are nanoscale structures that have been extensively used in application of recent nanotechnology engineering, such as in electronics, computer, aerospace and biomedical science. CNTs are commonly formed in ordinary flames (1), or by burning methane (2), ethylene

(3) and benzene (4). They exist in soot from both indoor and outdoor air (5). Although specific CNTs- associated human diseases have not been clearly identified, the potentially hazardous effects of CNTs have been observed in animal studies. Typically, CNTs acted as a trigger inducing inflammation, granulomas and fibrosis in mice (6).

This study aimed to examine the potentially pathogenic effects of silica and CNTs in human cell system containing fibroblasts, macrophages and T-cells. Fibroblasts are the most common cells in human connective tissues. They continuously synthesize and maintain the extracellular matrix (ECM) that determines the physical properties and biological functions of tissues (7). In many diseases, fibroblasts of connective tissues are the major target of pathogens and/or primary sites of dysfunction. A typical pathological consequence of dysfunctional fibroblasts is fibrosis, in which fibroblasts synthesize excessive amounts of the ECM components such as collagen (7,8). Fibroblasts obtained from typical fibrotic diseases, such as scleroderma, characteristically show fibrotic features including high levels of the ECM expression (8). While fibroblasts may be important endpoint tissue cells that directly affect disease phenotype, macrophages and T cells induce activation of fibroblasts in the body. Macrophages are early responding cells in up-taking pathogenic materials to activate lymphocytes. T lymphocytes are central players in cell-mediated immunity, which usually stimulate target cells such as fibroblasts. Therefore, the combination of human macrophages, T cells and fibroblasts provide a live bio-system from the human body for the studies of the potentially hazardous effects of environmental particles.

## Materials and Methods

*T cell isolation from PBMC and culture*

Whole blood (20 ml) obtained from a healthy donor were collected by the Vacutainer CPT tube system (Becton Dickenson, Heidelberg, Germany). The cell layer containing peripheral blood mononuclear cells (PBMC) were obtained through centrifugation (1500 *g*, 15 min, 24°C) within 1 hour of blood collection. Cells were recovered and washed twice in phosphate-buffered saline (PBS).

T cells were isolated from PBMC using CD3 MicroBeads and the magnetic cell-sorting system (Miltenyi Biotec, Auburn CA). Briefly, about $10^7$ PBMC were re-suspended in 80 µl PBS supplemented with 2 mM EDTA and 0.5% bovine serum albumin (PBS/E/B). Then 20 µl CD3 MicroBeads were added and incubated for 15 min at 4°C. After the mixture were washed with PBS/E/B, and re-suspended in PBS/E/B, they were added into the LS+ column on the magnetic field. The positive cells were eluted from the column with 5 ml PBS/E/B using the plunger supplied. Flow cytometry confirmed that this population was 95–98% pure ($CD3^+$ vs. $CD3^-$). Cells were grown in a T25 tissue culture flask (Greiner Labortechnik, Frickenhausen, Germany) in complete Roswell Park Memorial Institute medium-1640 (RPMI-1640) with 10% fetal bovine serum (FBS) and 1x phytohemagglutinin (PHA). After two days, the cells were cultured in fresh RPMI-1640 with recombinant human IL-2 at 50U/ml (R&D Systeims, Inc. Minneapolis, MN, USA). The culture medium was replaced with only RPMI-1640 for 48 hours before particle stimulation assays.

*Macrophage culture and stimulation*

THP-1 cells (Human monocytes) were obtained from American Type of Culture Collection (ATCC). After growing confluent in complete RPMI with 10% FBS, cells were treated with $10^{-8}$ M phorbol 12-myristate 13-acetate (PMA) for 36-48 hours to induce differentiation into macrophage like cells (macrophages/monocytes). The macrophage-like cells were cultured in RPMI medium

for 24 hours without serum supplementation, and were then stimulated with addition of either silica, CNTs or titanium particles (10 μg/ml for each) (purchased from Sigma-Aldrich, St. Louis, MO) for 24 hours. In addition, phosphate buffered saline (PBS) stimulation was used as the control. A digital camera equipped on a light microscopy (Nikon, ECLIPSE TE2000-U) was used to monitor cellular response to the particles.

*Normal human fibroblasts culture and stimulation*

A fibroblast strain was obtained from skin biopsy of a healthy donor. Briefly, cultured fibroblast strain was established by mincing tissues and placing them into 60 mm culture dishes secured by glass cover-slips. The primary cultures were maintained in Dulbecco's Modified Essential Media (DMEM) with 10% FBS and supplemented with antibiotic and antimycotic. The 5th passage of fibroblast cell strains were plated at a density of $2.5 \times 10^5$ cells in a 35 mm dish and grown until confluency. The culture medium was replaced with DMEM without FBS before stimulation assays.

Cultured normal human T cells were mixed with stimulated macrophages/monocytes for 10 minutes then added to cultured normal human fibroblasts. The stimulation process was monitored under a light microscope. Cytokine arrays were used to examine cytokines in cultured media after one-hour co-culture. Real-time RT-PCR was used to examine gene expression of fibroblasts responding to the stimulated macrophages/monocytes and T cells at 1 and 24 hours. Before extraction of total RNA from cultured fibroblasts, the fibroblasts were washed two times with culture medium to eliminate macrophages/monocytes and T cells. Whole experiments were performed in triplicates.

*Cytokine arrays*

The cytokine levels in cultured fibroblasts containing stimulated macrophages/monocytes and T cells were examined with Quansys Human Cytokine Array (Quansys Biosciences, Logan Utah). The array contains 12 different cytokines (IL1α, IL1β, IL2, IL4, IL5, IL6, IL8, IL10, IL13, IFNγ, TNFα and TNFβ). The assays were performed in triplicates following manufacture's protocol, and were imaged by a CCD camera. The data was analyzed with Quansys Array Software by Quansys Biosciences.

*Real-time RT PCR for measurement of transcript level*:

Total RNA from each sample was extracted from cultured fibroblasts described above using TRIzol reagent (Invitrogen Life Technology) and treated with Dnase I (Ambion, Austin, TX). The transcript levels of the genes including COL1A2, CTGF, IL1B and IL8 were selected for measurement. Both COL1A2 and CTGF are extracellular matrix genes that are commonly up-regulated in fibrotic process of fibroblasts, while IL1B and IL8 are inflammatory genes that are commonly expressed in inflammation by monocytes.

Quantitative real time RT-PCR was performed using an ABI 7900 sequence detector (Applied Biosystems, Foster City, CA). The specific primers and probes for each gene were purchased through Assays-on-Demand from Applied Biosystems. As described previously (9), cDNAs were synthesized from total RNA (same RNA used in microarrays) using Superscript II reverse transcriptase (Invitrogen Life Technology). Synthesized cDNAs were mixed with primers/probes in the 2x Taqman universal PCR buffer, and then assayed on an ABI 7900. Each subject was measured in triplicates. The data obtained from assays were analyzed with SDS 2.1 software (Applied Biosystems). The amount of total RNA in each sample was normalized with 18 S rRNA and GAPDH transcript levels.

## Results

*Morphological changes*

Through monitoring live cells with a digital camera on a microscopy, we observed the response of macrophages/monocytes to different particles. Within the first 30 minutes of stimulation, macrophages/monocytes started to move toward CNT particles. At time point of 24-hour stimulation, CNTs were heavily surrounded by macrophages/monocytes, while silica particles that appeared smaller than CNTs were surrounding the macrophages/monocytes (Figure 1). In contrast, titanium particles did not such changes. After stimulation, the cultures of macrophages/monocytes were mixed with T cells and fibroblasts. Within the first 30 minutes of co-cultures, the fibroblasts did not show significantly morphological changes. However, after 24 hours, microscopic examination showed deformed fibroblasts around the CNTs (Figure 2). The fibroblasts in the presence of silica- and titanium-stimulated macrophages/T cells did not show clear morphological changes (Figure 1 and 2).

*Cytokines secretion into cell culture medium*

A total of 12 cytokines (IL1α, IL2, IL4, IL5, IL6, IL8, IL10, IL13, IL18, IFNγ, TNFα and TNFβ) were examined with the ELISA assays. The changes of cytokine levels were displayed in Figure 3. In particular, compared to PBS control, one-hour silica stimulation in macrophage/monocyte cultures showed 3.3-fold and 2.5-fold increases of IL1α and IL1β (average changes measured in three assays, $p < 0.05$ by T test), respectively. CNTs stimulation showed 3.5-fold, 5.7-fold and 4.9-fold increases of IL1α, IL1β and IL8, respectively ($p < 0.05$ by T test). In addition, IL10 and IL13 also showed a mild increase in CNT stimulation, but appeared no significant ($p > 0.05$). Titanium stimulation did not show significant changes of the cytokines in the cultures.

*Gene expression changes of the fibroblasts*

At one-hour time point of co-culture of fibroblasts with silica stimulated macrophages/monocytes and T cells, a mild increase of transcript level of CTGF (1.61-fold, $P < 0.05$), but a decrease in IL1B and IL8 (0.56- and 0.49-fold, respectively, $P < 0.05$), were observed in the fibroblasts (Figure 4A). At 24-hour time point, increased expressions of COL1A2 were observed in both silica and CTN stimulation (1.96-fold and 2.77-fold, respectively, $P < 0.05$). In addition, a mild increase of CTGF also was observed in both (1.53-fold and 1.50-fold, respectively, $P < 0.05$) (Figure 4B). No other significant changes were observed.

**Discussion**

Both silica particles and CNTs have been reported as profibrogenic in a number of studies (6,10,11). Inflammation appeared to be a common process induced by nanoparticle (10-16). Our studies in a human cell system containing fibroblasts, macrophages/monocytes and T cells supported this notion. Interestingly, these two particles shared some features in the processes of cellular responses, but also seemed to be distinctive in each.

In both silica and CNTs stimulation, IL1α and IL1β were significantly increased in the culture medium at 1-hour time point after addition of stimulated macrophage/monocytes and T cells into cultured fibroblasts. Increased gene expression of the COL1A2 was followed in cultured fibroblasts at 24-hour time point. IL1α and IL1β are important pro-inflammatory cytokines that may trigger a variety of cellular responses, such as fibrosis, apoptosis, and proliferation. Increased levels of IL1α and IL1β in the culture medium may come from stimulated macrophages/monocytes that are usually the major source of inflammatory cytokines. A down-regulation of the IL1B gene observed in the cultured fibroblasts may be a feedback response.

Up-regulated the COL1A2 in cultured fibroblasts is likely triggered at least in part by IL1α and IL1β in the cultures, and which indicates a potential fibrotic response.

Different from silica stimulation, IL8 also was significantly increased in early culture medium (1-hour time point) of fibroblasts with CNTs stimulated macrophages/monocyte and T cells. IL8 is a chemokine that attracts inflammatory cells at the site of inflammation. Concordantly, live-microscope examination showed that macrophages/monocytes aggregated at the site of CNTs in the cultures of either macrophages/monocytes alone or with fibroblasts. This change was not observed in silica and titanium stimulation. Compared to silica and titanium particles, a bigger size of CNTs may affect cellular responses in cultured cells, which was reported in studies of CNTs (14,17).

An early mild upregulation of the CTGF gene in cultured fibroblasts after addition of silica stimulated macrophages/monocytes and T cells at 1-hour time point is distinct from the CNTs stimulation. CTGF is a profibrotic cytokine that may induce collagen expression and deposition in fibrotic diseases. Therefore, a later response of the COL1A2 in the fibroblasts at 24-hour time point also may be triggered by CTGF signaling. This observation appeared important since high levels of CTGF have been associated with fibrotic diseases such as systemic sclerosis, hepatic fibrosis and idiopathic pulmonary fibrosis (18-20). Up-regulation of CTGF expression was reported in cultures of fibroblasts from patients with SSc (21). Silica exposure has been widely discussed in the pathogenesis of SSc. Bramwell in 1914 reported five cases of SSc who were stonemasons exposed to silica (22). The incidence of SSc in black South African gold miners who were exposed to silica was reported to be 81.8 per million compared to 3.4 per million in general black South Africans (23). The relative risk for developing SSc was 25-39 times higher in

patients with radiologically documented silicosis (24). Our results supported that silica may be a potential environmental trigger to SSc.

In summary, our studies indicated both silica and CNTs might activate fibroblasts through IL1signaling in the cultures of macrophages/monocytes, T cells and fibroblasts. In addition, silica stimulation also triggered CTGF over-expression in the fibroblasts, while CNTs induced IL8 production from the cultures. These potential proinflammatory changes were followed by an up-regulation of collagen gene. In contrast, titanium stimulation did not show significant changes in the cultures. These findings have relevance for understanding environmental contributions to inflammation as well as fibrosing diseases such as scleroderma, and may suggest a potential health hazard in current application of nano-particles including silica and CTNs.

**Acknowledgements:** The studies were supported by grants from the Department of the Army, Medical Research Acquisition Activity: PR064803 (X.Z)

**References**

1. Singer JM, Grumer J. Carbon formation in very rich hydrocarbon-air flames. I. Studies of chemical content, temperature, ionization and particulate matter. Seventh Symposium (International) on Combustion 1959.
2. Yuan L, Saito K, Pan C, Williams FA, Gordon AS. Nanotubes from methane flames. Chemical physics letters 2001;340: 237–41.
3. Yuan L, Saito K, Hu W, Chen Z. Ethylene flame synthesis of well-aligned multi-walled carbon nanotubes. Chemical physics letters 2001;346: 23–8.

4. Duan HM, McKinnon JT. Nanoclusters Produced in Flames. Journal of Physical Chemistry 98 1994;49: 12815–8.
5. Murr LE, Bang JJ, Esquivel EV, Guerrero PA, Lopez DA. Carbon nanotubes, nanocrystal forms, and complex nanoparticle aggregates in common fuel-gas combustion sources and the ambient air. Journal of Nanoparticle Research 6 2004: 241–251.
6. Lam CW, James JT, McCluskey R, Hunter RL. Pulmonary toxicity of single-wall carbon nanotubes in mice 7 and 90 days after intratracheal instillation. Toxicol Sci 2004;77:126-34.
7. Kulonen E, Pikkarainen J. Biology of fibroblast. Academic Press. London and New York 1973.
8. Claman HN, Giorno RC, Seibold JR. Endothelial and fibroblastic activation in scleroderma. The myth of the "uninvolved skin". Arthritis Rheum 1991;34:1495-501.
9. Wang JC, Sonnylal S, Arnett FC, De Crombrugghe B, Zhou X. Attenuation of expression of extracellular matrix genes with siRNAs to Sparc and Ctgf in skin fibroblasts of CTGF transgenic mice. Int J Immunopathol Pharmacol 2011;24:595-601
10. Fubini B, Hubbard A. Reactive oxygen species (ROS) and reactive nitrogen species (RNS) generation by silica in inflammation and fibrosis. Free Radic Biol Med. 2003;15;34:1507-16.
11. Shvedova AA, Kisin ER, Mercer R, Murray AR, Johnson VJ, Potapovich AI, Tyurina YY, Gorelik O, Arepalli S, Schwegler-Berry D, Hubbs AF, Antonini J, Evans DE, Ku BK, Ramsey D, Maynard A, Kagan VE, Castranova V, Baron P. Unusual inflammatory and fibrogenic pulmonary responses to single-walled carbon nanotubes in mice. Am J Physiol Lung Cell Mol Physiol 2005;289:L698-708.
12. Park EJ, Park K. Oxidative stress and pro-inflammatory responses induced by silica nanoparticles in vivo and in vitro. Toxicol Lett 2009;184:18-25.

13. Sacks M, Gordon J, Bylander J, Porter D, Shi XL, Castranova V, Kaczmarczyk W, Van Dyke K, Reasor MJ. Silica-induced pulmonary inflammation in rats: activation of NF-kappa B and its suppression by dexamethasone.Biochem Biophys Res Commun 1998;253:181-4
14. Qu C, Wang L, He J, Tan J, Liu W, Zhang S, Zhang C, Wang Z, Jiao S, Liu S, Jiang G. Carbon nanotubes provoke inflammation by inducing the pro-inflammatory genes IL-1β and IL-6. Gene 2012;493:9-12.
15. Huizar I, Malur A, Midgette YA, Kukoly C, Chen P, Ke PC, Podila R, Rao AM, Wingard CJ, Dobbs L, Barna BP, Kavuru MS, Thomassen MJ. Novel murine model of chronic granulomatous lung inflammation elicited by carbon nanotubes. Am J Respir Cell Mol Biol. 2011;45:858-66.
16. Yazdi AS, Guarda G, Riteau N, Drexler SK, Tardivel A, Couillin I, Tschopp J. Nanoparticles activate the NLR pyrin domain containing 3 (Nlrp3) inflammasome and cause pulmonary inflammation through release of IL-1α and IL-1β. Proc Natl Acad Sci U S A. 2010;107:19449-54.
17. Yamashita K, Yoshioka Y, Higashisaka K, Morishita Y, Yoshida T, Fujimura M, Kayamuro H, Nabeshi H, Yamashita T, Nagano K, Abe Y, Kamada H, Kawai Y, Mayumi T, Yoshikawa T, Itoh N, Tsunoda S, Tsutsumi Y. Carbon nanotubes elicit DNA damage and inflammatory response relative to their size and shape.Inflammation. 2010;33(4):276-80.
18. Shi-wen X, Pennington D, Holmes A, Leask A, Bradham D, Beauchamp JR, Fonseca C, du Bois RM, Martin GR, Black CM, Abraham DJ. Autocrine overexpression of CTGF maintains fibrosis: RDA analysis of fibrosis genes in systemic sclerosis. Exp Cell Res. 2000;259(1):213-24.
19. Gressner AM, Yagmur E, Lahme B, Gressner O, Stanzel S. Connective tissue growth factor

in serum as a new candidate test for assessment of hepatic fibrosis. Clin Chem. 2006;52(9):1815-7.

20. Pan LH, Yamauchi K, Uzuki M, Nakanishi T, Takigawa M, Inoue H, Sawai T. Type II alveolar epithelial cells and interstitial fibroblasts express connective tissue growth factor in IPF. Eur Respir J. 2001;17(6):1220-7.
21. Zhou X, Tan FK, Xiong M, Arnett FC, Feghali-Bostwick CA. Monozygotic twins clinically discordant for scleroderma show concordance for fibroblast gene expression profiles. Arthritis Rheum. 2005;52(10):3305-14.
22. Bramwell, B. Diffuse scleroderma: its occurrence in stonemasons; its treatment by fibrolysin-elevations in temperature due to fibrinolysin injections. Edinburgh Med. J. 1914;12,387-401.
23. Cowie RL. Silica-dust-exposed mine workers with scleroderma (systemic sclerosis). Chest. 1987;92:260-2.
24. Haustein HU and Zeigler V. Environmentally induced systemic sclerosis-like disorders. Int J Dermatol. 1985;24:147-51.

**Figure 1.** Cultures of macrophages/monocytes with different stimuli at 24-hour time point: A: with PBS; B: with CNTs for 24 hours; C: with silica particles; D: with titanium particles. *Arrows indicate particles.

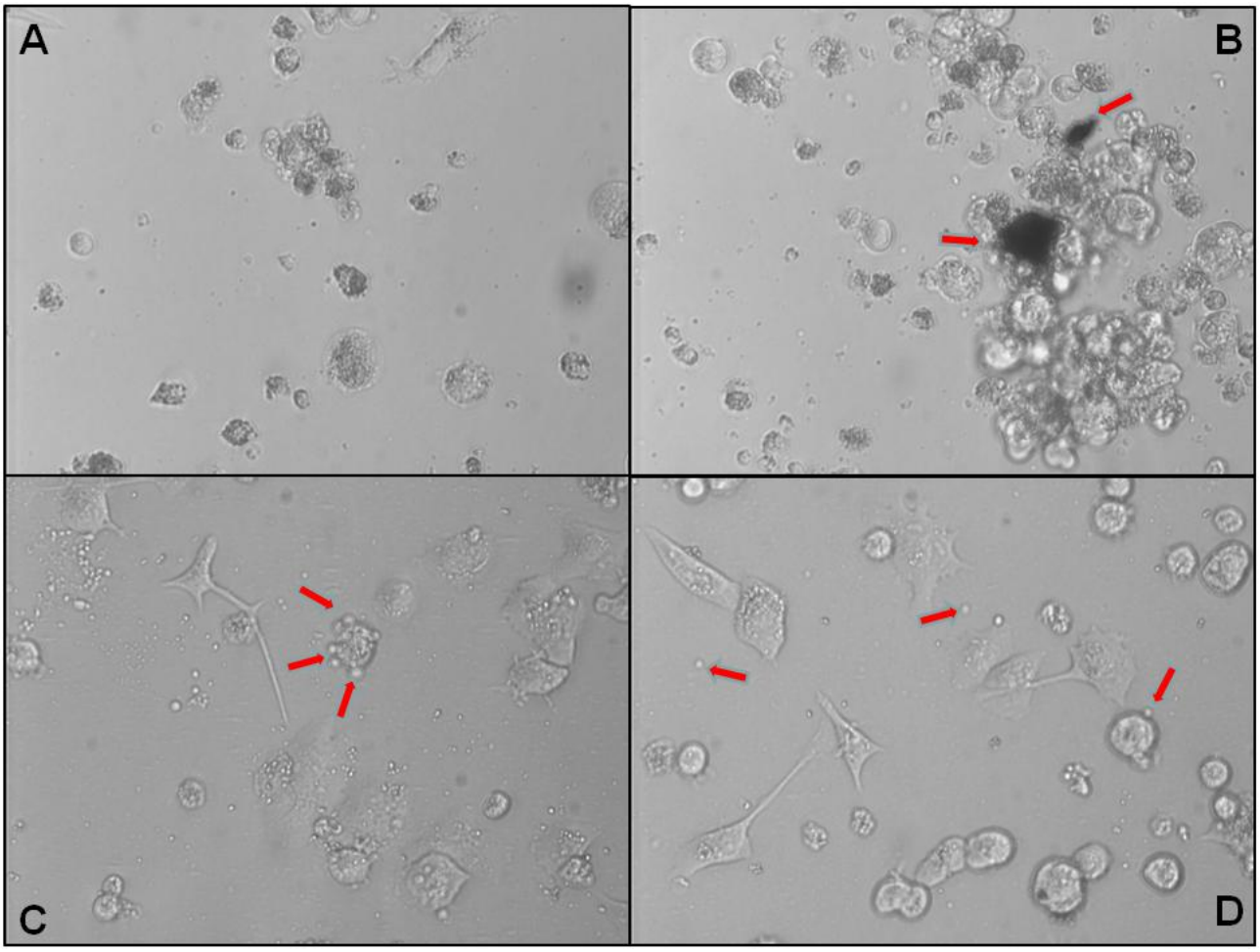


**Figure 2.** Cultured Fibroblasts with stimulated macrophages/monocytes and T cells at 24-hour time point. A: Fibroblasts cultured with PBS stimulated macrophages/monocytes and T cells; B: Fibroblasts cultured with CNT stimulated macrophages/monocytes and T cells; C: Fibroblasts cultured with silica stimulated macrophages/monocytes and T cells; D: Fibroblasts cultured with titanium stimulated macrophages/monocytes and T cells. Arrow indicates a CNT.

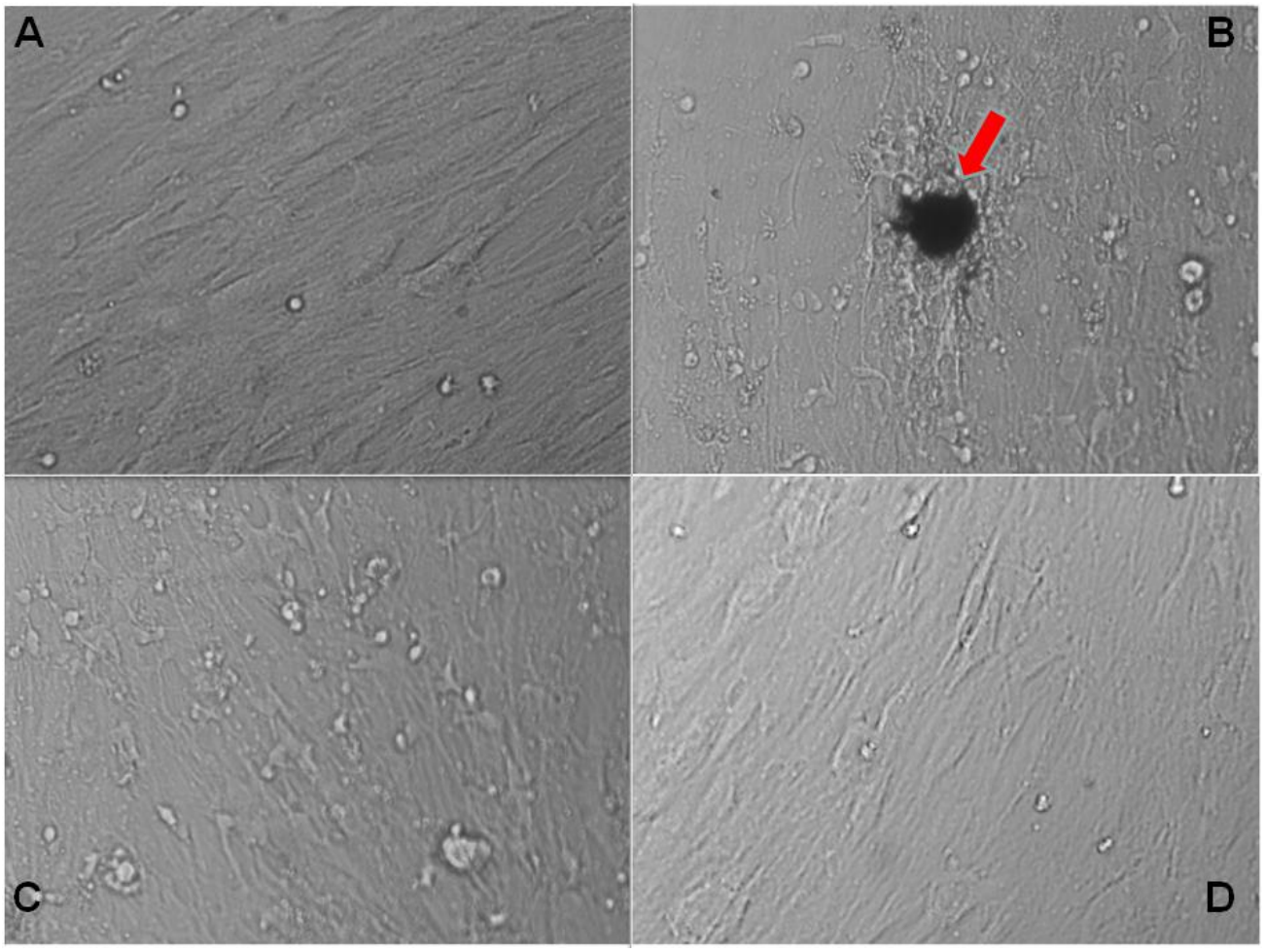


**Figure 3.** Cytokine levels in 1-hour cultured medium of fibroblasts with differently stimulated macrophages/monocytes and T cells. A: Fibroblasts cultured with PBS stimulated macrophages/monocytes and T cells; B: Fibroblasts cultured with CNT stimulated macrophages/monocytes and T cells; C: Fibroblasts cultured with silica stimulated macrophages/monocytes and T cells; D: Fibroblasts cultured with titanium stimulated macrophages/monocytes and T cells. *p<0.05

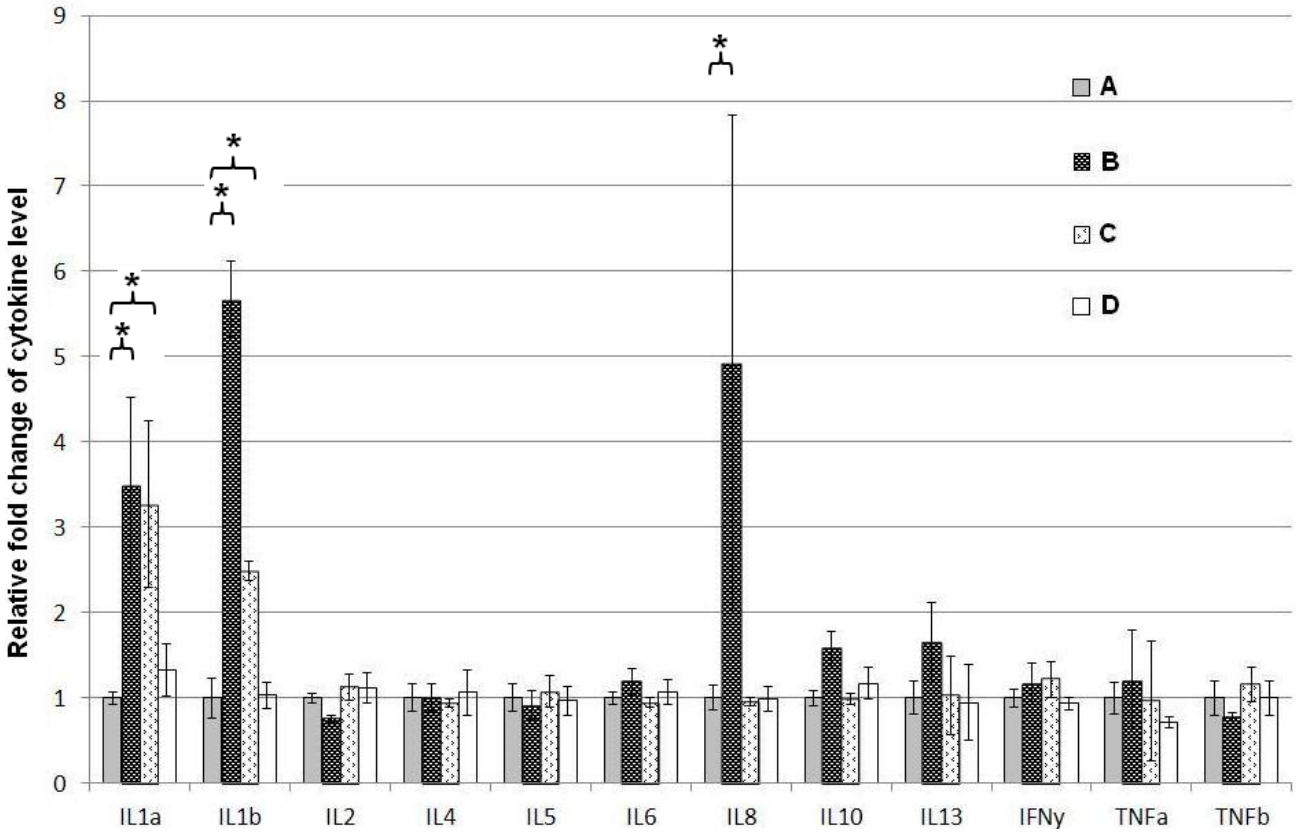


**Figure 4A.** Changes of transcript levels of the COL1A2, CTGF, IL1B and IL8 in the fibroblasts after 1-hour co-culture with stimulated macrophages/monocytes and T cells. A: Fibroblasts cultured with PBS stimulated macrophages/monocytes and T cells; B: Fibroblasts cultured with CNT stimulated macrophages/monocytes and T cells; C: Fibroblasts cultured with silica stimulated macrophages/monocytes and T cells; D: Fibroblasts cultured with titanium stimulated macrophages/monocytes and T cells. *p<0.05

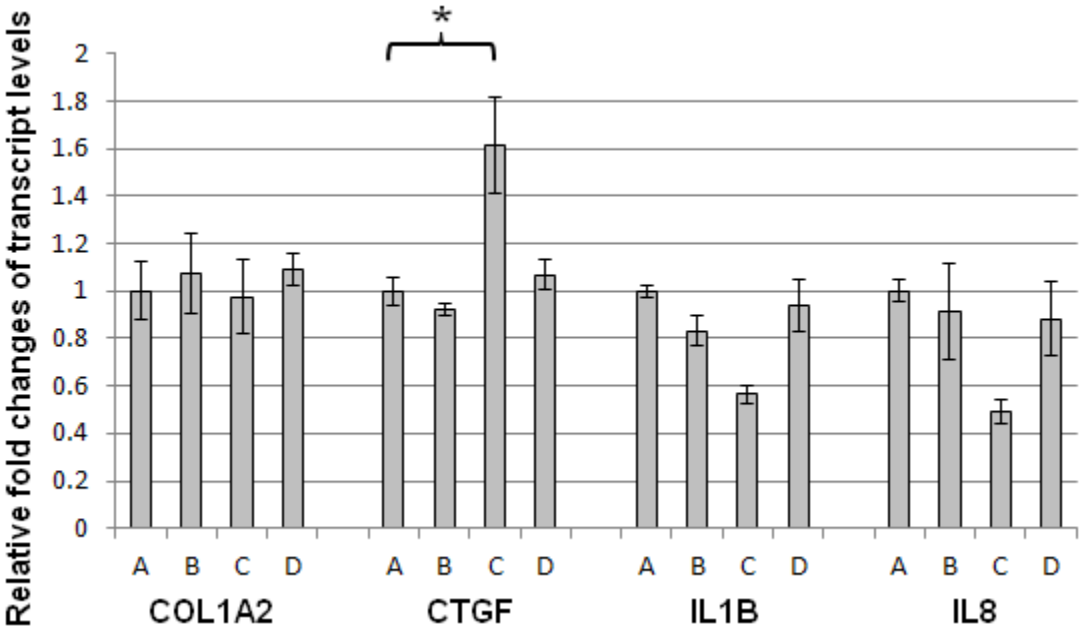
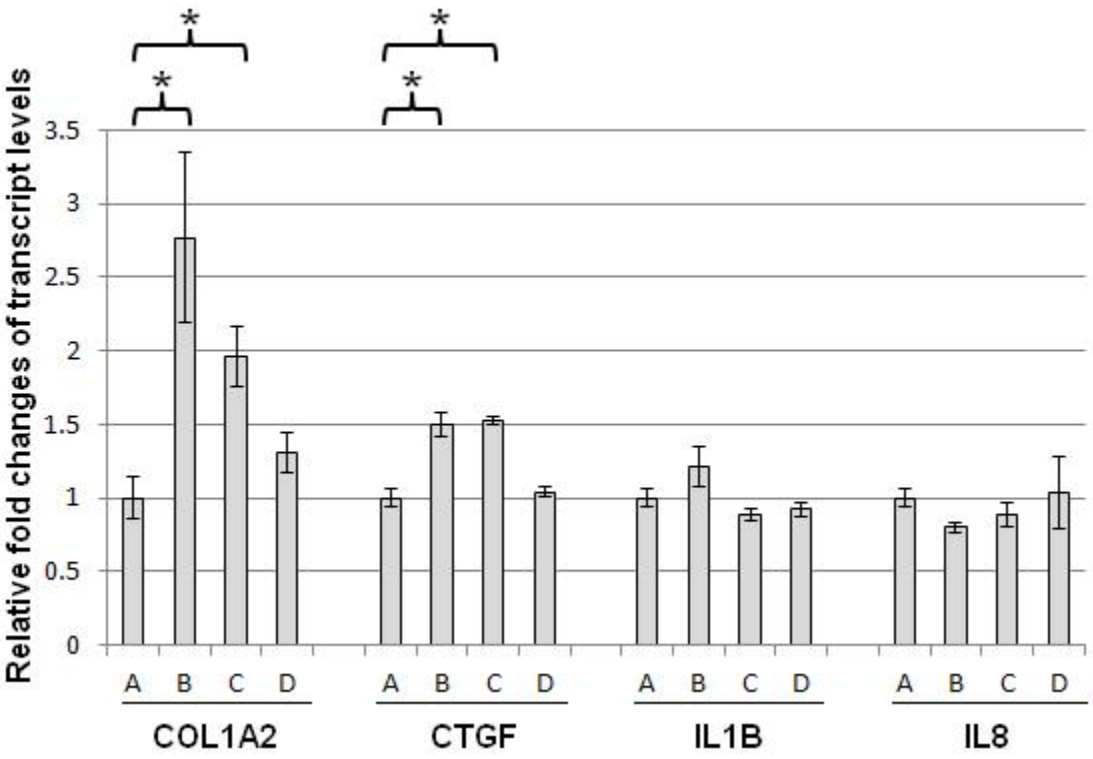


**Figure 4B.** Changes of transcript levels of the COL1A2, CTGF, IL1B and IL8 in the fibroblasts after 24-hour co-culture with stimulated macrophages/monocytes and T cells. A: Fibroblasts cultured with PBS stimulated macrophages/monocytes and T cells; B: Fibroblasts cultured with CNT stimulated macrophages/monocytes and T cells; C: Fibroblasts cultured with silica stimulated macrophages/monocytes and T cells; D: Fibroblasts cultured with titanium stimulated macrophages/monocytes and T cells. *p<0.05



Associations between SSc ssuceptibility HLA alleles and gene expression of fibroblasts in response to sili

| | MMP3 A | MMP3 B | C | p-value | p-value | | MMP1 A | MMP1 B |
|---|---|---|---|---|---|---|---|---|
| MMP3 | SSc DRB1*: | SSc DRB1*: | Controls | A vs B | A vs C | | SSc DRB1*: | SSc DRB1*: |
| 1ug | 1.176251 | 1.682235 | 1.501977 | 0.046825 | 0.05297 | 1ug | 1.225165 | 1.529217 |
| 5ug | 1.112162 | 1.432328 | 1.521283 | 0.14466 | 0.046547 | 5ug | 1.249622 | 1.225964 |
| 10ug | 1.201614 | 1.879568 | 1.628776 | **0.033835** | **0.027518** | 10ug | 1.22216 | 1.452832 |
| 25ug | 1.17485 | 2.042464 | 1.765051 | 0.015318 | 0.012015 | 25ug | 1.136214 | 1.59735 |
| 50ug | 1.445138 | 2.883384 | 2.768349 | 0.004929 | 0.000422 | 50ug | 1.322438 | 1.946496 |
| MMP3 | SSc DRB1*: | SSc DRB1*: | Controls | | | | SSc DRB1*: | SSc DRB1*: |
| 1D | 1.201614 | 1.879568 | 1.628776 | **0.033835** | **0.027518** | 1D | 1.22216 | 1.452832 |
| 2D | 0.69237 | 1.482681 | 1.680742 | 0.011075 | 0.004068 | 2D | 0.614389 | 1.187871 |
| 3D | **1.070011** | 1.143818 | 1.207753 | 0.394995 | 0.31643 | 3D | **0.887964** | 0.95948 |
| 4D | 0.891276 | 1.202977 | 1.074574 | 0.079167 | 0.168075 | 4D | 0.994564 | 1.196035 |
| 5D | 1.063142 | 1.08748 | 0.960065 | 0.470492 | 0.377187 | 5D | 1.034514 | 1.058581 |

| | MMP3 A | MMP3 B | C | p-value | p-value | | MMP1 A | MMP1 B |
|---|---|---|---|---|---|---|---|---|
| MMP3 | SSc DQB1*: | SSc DQB1*: | Controls | A vs B | A vs C | | SSc DQB1*: | SSc DQB1*: |
| 1ug | 1.495261 | 1.656548 | 1.501977 | 0.368759 | 0.492496 | 1ug | 1.506879 | 1.433853 |
| 5ug | 1.358584 | 1.376794 | 1.521283 | 0.482996 | 0.314266 | 5ug | 1.423225 | 1.072563 |
| 10ug | 1.529602 | 1.948499 | 1.628776 | 0.247668 | 0.400802 | 10ug | 1.445452 | 1.373395 |
| 25ug | 1.625374 | 2.083309 | 1.765051 | 0.248379 | 0.380554 | 25ug | 1.474484 | 1.527218 |
| 50ug | 2.284327 | 2.813306 | 2.768349 | 0.294805 | 0.262413 | 50ug | 1.944827 | 1.699625 |
| MMP3 | SSc DQB1*: | SSc DQB1*: | Controls | | | | SSc DQB1*: | SSc DQB1*: |
| 1D | 1.529602 | 1.948499 | 1.628776 | 0.247668 | 0.400802 | 1D | 1.445452 | 1.373395 |
| 2D | 1.110042 | 1.488128 | 1.680742 | 0.158095 | 0.074867 | 2D | 0.960594 | 1.149521 |
| 3D | 1.173221 | 1.090834 | 1.207753 | 0.342466 | 0.434017 | 3D | 1.000123 | 0.897852 |
| 4D | 0.968204 | 1.327494 | 1.074574 | 0.075561 | 0.246571 | 4D | 0.980768 | 1.350759 |
| 5D | 1.099996 | 1.067271 | 0.960065 | 0.429637 | 0.233516 | 5D | 1.020127 | 1.087236 |

| | MMP3 A | MMP3 B | C | p-value | p-value | | MMP1 A | MMP1 B |
|---|---|---|---|---|---|---|---|---|
| MMP3 | SSc DPB1*: | SSc DPB1*: | Controls | A vs B | A vs C | | SSc DPB1*: | SSc DPB1*: |
| 1ug | 0.960156 | 1.478898 | 1.501977 | 0.06008 | 0.050316 | 1ug | 1.201938 | 1.436362 |
| 5ug | 0.451893 | 1.34151 | 1.521283 | 0.070514 | 0.054867 | 5ug | 0.331685 | 1.308107 |
| 10ug | 0.612721 | 1.593817 | 1.628776 | 0.026796 | 0.025429 | 10ug | 0.719796 | 1.365787 |
| 25ug | 1.107415 | 1.613981 | 1.765051 | 0.103961 | 0.055053 | 25ug | 1.168487 | 1.388065 |
| 50ug | 1.491244 | 2.243097 | 2.768349 | 0.201573 | 0.107312 | 50ug | 1.404146 | 1.704504 |
| MMP3 | SSc DPB1*: | SSc DPB1*: | Controls | | | | SSc DPB1*: | SSc DPB1*: |
| 1D | 0.612721 | 1.593817 | 1.628776 | 0.026796 | 0.025429 | 1D | 0.719796 | 1.365787 |
| 2D | 0.949089 | 1.195966 | 1.680742 | 0.304583 | 0.118096 | 2D | 0.801246 | 1.056166 |
| 3D | 0.889698 | 1.073037 | 1.207753 | 0.263446 | 0.161291 | 3D | 0.693467 | 0.977114 |
| 4D | 3.309883 | 0.904677 | 1.074574 | 0.15747 | 0.187489 | 4D | 5.868647 | 1.05478 |
| 5D | 0.868433 | 1.133034 | 0.960065 | 0.305356 | 0.424469 | 5D | 0.957938 | 1.126487 |

ica stimulation

| C Controls | p-value A vs B | p-value A vs C | | CTGF A SSc DRB1*1 | CTGF B SSc DRB1*1 | CTGF C Controls | p-value A vs B | p-value A vs C |
|---|---|---|---|---|---|---|---|---|
| 1.556252 | 0.106904 | 0.080341 | 1ug | 1.012642 | 1.041165 | 1.134968 | 0.433694 | 0.241307 |
| 1.492408 | 0.462154 | 0.16911 | 5ug | 0.944152 | 0.946103 | 1.1435 | 0.495741 | 0.15112 |
| 1.66885 | 0.111336 | 0.004772 | 10ug | 0.833975 | 1.034686 | 1.155553 | 0.133618 | 0.036748 |
| 1.579543 | 0.036923 | 0.036812 | 25ug | 1.110327 | 1.095479 | 1.21418 | 0.472889 | 0.316383 |
| 2.315631 | 0.016739 | 0.000301 | 50ug | 1.272775 | 1.09197 | 1.253234 | 0.223351 | 0.466914 |
| Controls | | | | SSc DRB1*1 | SSc DRB1*1 | Controls | | |
| 1.66885 | 0.111336 | 0.004772 | 1D | 0.833975 | 1.034686 | 1.155553 | 0.133618 | 0.036748 |
| 1.287936 | 0.019656 | 0.014064 | 2D | 0.710369 | 0.79647 | 1.04294 | 0.372388 | 0.144642 |
| 1.228084 | **0.310261** | **0.035663** | 3D | **1.077344** | 0.813015 | 0.923544 | **0.002333** | **0.04048** |
| 1.311402 | 0.203849 | 0.046955 | 4D | 1.080384 | 0.964739 | 0.88094 | 0.241869 | 0.081154 |
| 1.131739 | 0.459427 | 0.342589 | 5D | 0.97011 | 0.778021 | 0.994713 | 0.035738 | 0.405619 |

| C Controls | p-value A vs B | p-value A vs C | | CTGF A SSc DQB1* | CTGF B SSc DQB1* | CTGF C Controls | p-value A vs B | p-value A vs C |
|---|---|---|---|---|---|---|---|---|
| 1.556252 | 0.398765 | 0.429628 | 1ug | 0.797731 | 0.917117 | 1.08781 | 0.061605 | 0.022129 |
| 1.492408 | 0.114332 | 0.395486 | 5ug | 0.894649 | 0.998144 | 1.1435 | 0.230293 | 0.008231 |
| 1.66885 | 0.403997 | 0.224274 | 10ug | 0.894656 | 1.104202 | 1.155553 | 0.116891 | 0.011784 |
| 1.579543 | 0.434505 | 0.367239 | 25ug | 1.072897 | 1.123175 | 1.21418 | 0.383529 | 0.141476 |
| 2.315631 | 0.294961 | 0.230408 | 50ug | 1.128648 | 1.116272 | 1.253234 | 0.473763 | 0.197697 |
| Controls | | | | SSc DQB1* | SSc DQB1* | Controls | | |
| 1.66885 | 0.403997 | 0.224274 | 1D | 0.894656 | 1.104202 | 1.155553 | 0.116891 | 0.011784 |
| 1.287936 | 0.183789 | 0.103462 | 2D | 0.763845 | 0.796642 | 1.04294 | 0.403445 | 0.028275 |
| 1.228084 | 0.169195 | 0.056889 | 3D | 1.009278 | 0.727943 | 0.923544 | 0.00094 | 0.150198 |
| 1.311402 | 0.151009 | 0.015441 | 4D | 1.08952 | 0.940413 | 0.931691 | 0.170573 | 0.09325 |
| 1.131739 | 0.335015 | 0.203443 | 5D | 0.827177 | 0.792482 | 0.994713 | 0.364059 | 0.016189 |

| C Controls | p-value A vs B | p-value A vs C | | CTGF A SSc DPB1*1 | CTGF B SSc DPB1*1 | CTGF C Controls | p-value A vs B | p-value A vs C |
|---|---|---|---|---|---|---|---|---|
| 1.556252 | 0.210582 | 0.126808 | 1ug | 0.979444 | 1.016844 | 1.134968 | 0.419211 | 0.213044 |
| 1.492408 | 0.124024 | 0.106439 | 5ug | 0.564054 | 0.94871 | 1.1435 | 0.238581 | 0.173167 |
| 1.66885 | 0.039031 | 0.016684 | 10ug | 0.581098 | 0.996163 | 1.155553 | 0.231329 | 0.178954 |
| 1.579543 | 0.281924 | 0.163931 | 25ug | 1.000498 | 1.022111 | 1.21418 | 0.470727 | 0.254525 |
| 2.315631 | 0.291931 | 0.092603 | 50ug | 1.150575 | 1.047988 | 1.253234 | 0.3977 | 0.40471 |
| Controls | | | | SSc DPB1*1 | SSc DPB1*1 | Controls | | |
| 1.66885 | 0.039031 | 0.016684 | 1D | 0.581098 | 0.996163 | 1.155553 | 0.231329 | 0.178954 |
| 1.287936 | 0.202245 | 0.093604 | 2D | 0.755634 | 0.788193 | 1.04294 | 0.458555 | 0.221346 |
| 1.228084 | 0.110807 | 0.03482 | 3D | 0.613426 | 0.855105 | 0.923544 | 0.042663 | 0.023036 |
| 1.311402 | 0.222546 | 0.250681 | 4D | 1.896915 | 0.904103 | 0.88094 | 0.193165 | 0.185646 |
| 1.131739 | 0.393361 | 0.390362 | 5D | 0.595095 | 0.838064 | 0.994713 | 0.200899 | 0.11521 |

| | Col3 A1 A SSc DRB1*: | Col3 A1 B SSc DRB1*: | C Controls | p-value A vs B | p-value A vs C | | Col1A2 A SSc DRB1*: | Col1A2 B SSc DRB1*: |
|---|---|---|---|---|---|---|---|---|
| 1ug | 1.087404 | 0.812729 | 1.08781 | 0.061624 | 0.498852 | 1ug | 1.144436 | 1.0199 |
| 5ug | 0.947414 | 0.978715 | 1.160968 | 0.4293 | 0.117719 | 5ug | 1.156097 | 0.967036 |
| 10ug | 0.899443 | 1.086975 | 1.156306 | 0.171566 | 0.092406 | 10ug | 1.160252 | 1.077849 |
| 25ug | 0.996951 | 1.079484 | 1.202863 | 0.374231 | 0.218206 | 25ug | 1.251691 | 1.013961 |
| 50ug | 1.042107 | 1.049363 | 1.161034 | 0.487607 | 0.305412 | 50ug | 1.269985 | 0.938785 |
| | SSc DRB1*: | SSc DRB1*: | Controls | | | | SSc DRB1*: | SSc DRB1*: |
| 1D | 0.899443 | 1.086975 | 1.156306 | 0.171566 | 0.092406 | 1D | 1.160252 | 1.077849 |
| 2D | 0.98313 | 0.883393 | 1.100188 | 0.310225 | 0.302376 | 2D | 0.955578 | 0.949528 |
| 3D | **1.269219** | 0.973959 | 0.959313 | **0.081028** | **0.059621** | 3D | **1.312786** | 0.901605 |
| 4D | 1.321638 | 0.953724 | 0.931691 | 0.092554 | 0.07369 | 4D | 1.115216 | 1.039104 |
| 5D | 1.107165 | 0.870565 | 0.984486 | 0.025717 | 0.13694 | 5D | 1.118087 | 0.949118 |

| | Col3 A1 A SSc DQB1*: | Col3 A1 B SSc DQB1*: | C Controls | p-value A vs B | p-value A vs C | | Col1A2 A SSc DQB1*: | Col1A2 B SSc DQB1*: |
|---|---|---|---|---|---|---|---|---|
| 1ug | 0.9473 | 1.130022 | 1.134968 | 0.283569 | 0.08602 | 1ug | 1.139688 | 0.953689 |
| 5ug | 0.984231 | 0.962288 | 1.160968 | 0.444171 | 0.061851 | 5ug | 1.06181 | 0.940597 |
| 10ug | 1.022372 | 1.078995 | 1.156306 | 0.370082 | 0.152743 | 10ug | 1.145659 | 1.042635 |
| 25ug | 1.202131 | 0.945504 | 1.202863 | 0.074961 | 0.49799 | 25ug | 1.189279 | 0.935581 |
| 50ug | 1.132678 | 0.97198 | 1.161034 | 0.190543 | 0.415639 | 50ug | 1.109589 | 0.888019 |
| | SSc DQB1*: | SSc DQB1*: | Controls | | | | SSc DQB1*: | SSc DQB1*: |
| 1D | 1.022372 | 1.078995 | 1.156306 | 0.370082 | 0.152743 | 1D | 1.145659 | 1.042635 |
| 2D | 0.957404 | 0.848471 | 1.100188 | 0.198262 | 0.145427 | 2D | 0.932252 | 0.968771 |
| 3D | 1.220722 | 0.86461 | 0.959313 | 0.010494 | 0.026665 | 3D | 1.118788 | 0.834899 |
| 4D | 1.008879 | 0.961403 | 0.88094 | 0.378614 | 0.063134 | 4D | 1.100974 | 1.007198 |
| 5D | 0.926092 | 0.892461 | 0.984486 | 0.388779 | 0.301502 | 5D | 0.995021 | 0.960562 |

| | Col3 A1 A SSc DPB1*: | Col3 A1 B SSc DPB1*: | C Controls | p-value A vs B | p-value A vs C | | Col1A2 A SSc DPB1*: | Col1A2 B SSc DPB1*: |
|---|---|---|---|---|---|---|---|---|
| 1ug | 0.967183 | 0.91474 | 1.08781 | 0.385406 | 0.252702 | 1ug | 0.852428 | 1.031647 |
| 5ug | 0.997714 | 0.970756 | 1.160968 | 0.465536 | 0.316065 | 5ug | 0.876663 | 0.958973 |
| 10ug | 0.753296 | 1.043844 | 1.156306 | 0.186831 | 0.127532 | 10ug | 0.603061 | 1.069709 |
| 25ug | 1.074389 | 0.971529 | 1.202863 | 0.277581 | 0.242517 | 25ug | 0.930497 | 0.965812 |
| 50ug | 1.137049 | 0.999147 | 1.161034 | 0.275598 | 0.459717 | 50ug | 0.933026 | 0.930078 |
| | SSc DPB1*: | SSc DPB1*: | Controls | | | | SSc DPB1*: | SSc DPB1*: |
| 1D | 0.753296 | 1.043844 | 1.156306 | 0.186831 | 0.127532 | 1D | 0.603061 | 1.069709 |
| 2D | 0.788119 | 0.920503 | 1.100188 | 0.236037 | 0.076532 | 2D | 0.879814 | 0.964392 |
| 3D | 0.639087 | 1.048278 | 0.959313 | 0.114509 | 0.154269 | 3D | 0.834905 | 0.967678 |
| 4D | 0.924602 | 0.97977 | 0.931691 | 0.462249 | 0.495009 | 4D | 1.415933 | 1.020562 |
| 5D | 0.750265 | 0.930625 | 0.984486 | 0.193248 | 0.141275 | 5D | 0.826117 | 0.996799 |

| C Controls | p-value A vs B | p-value A vs C | | TIMP3 A SSc DRB1*1 | TIMP3 B SSc DRB1*1 | C Controls | p-value A vs B | p-value A vs C |
|---|---|---|---|---|---|---|---|---|
| 1.080674 | 0.212503 | 0.339035 | 1ug | 1.246214 | 1.211986 | 1.197412 | 0.448045 | 0.425245 |
| 1.133823 | 0.136352 | 0.450341 | 5ug | 1.234518 | 1.108624 | 1.228162 | 0.293497 | 0.489137 |
| 1.126126 | 0.344046 | 0.432129 | 10ug | 1.181101 | 1.218565 | 1.267403 | 0.44556 | 0.371765 |
| 1.067974 | 0.155046 | 0.218548 | 25ug | 1.316987 | 1.24865 | 1.19955 | 0.394075 | 0.314899 |
| 1.050754 | 0.058595 | 0.150556 | 50ug | 1.34269 | 1.259179 | 1.309581 | 0.387816 | 0.455096 |
| Controls | | | | SSc DRB1*1 | SSc DRB1*1 | Controls | | |
| 1.126126 | 0.344046 | 0.432129 | 1D | 1.181101 | 1.218565 | 1.267403 | 0.44556 | 0.371765 |
| 1.065991 | 0.489589 | 0.33087 | 2D | 0.730984 | 1.021981 | 1.333873 | 0.200558 | 0.074346 |
| 0.992487 | **0.009815** | **0.03458** | 3D | **1.382897** | 1.066853 | 1.114372 | **0.091993** | **0.129786** |
| 0.954616 | 0.294935 | 0.099538 | 4D | 1.183584 | 1.391675 | 1.096238 | 0.210425 | 0.293791 |
| 1.003405 | 0.028321 | 0.096196 | 5D | 1.298775 | 1.210094 | 1.162863 | 0.341575 | 0.25682 |

| C Controls | p-value A vs B | p-value A vs C | | TIMP3 A SSc DQB1* | TIMP3 B SSc DQB1* | C Controls | p-value A vs B | p-value A vs C |
|---|---|---|---|---|---|---|---|---|
| 1.080674 | 0.047088 | 0.243423 | 1ug | 1.270619 | 1.169102 | 1.197412 | 0.226379 | 0.281712 |
| 1.133823 | 0.151206 | 0.210993 | 5ug | 1.208813 | 1.058771 | 1.228162 | 0.163041 | 0.432883 |
| 1.126126 | 0.240372 | 0.417785 | 10ug | 1.251426 | 1.17448 | 1.267403 | 0.353282 | 0.454095 |
| 1.067974 | 0.023685 | 0.138857 | 25ug | 1.323437 | 1.202509 | 1.19955 | 0.228768 | 0.187412 |
| 1.050754 | 0.050368 | 0.288962 | 50ug | 1.390471 | 1.16983 | 1.309581 | 0.094117 | 0.289044 |
| Controls | | | | SSc DQB1* | SSc DQB1* | Controls | | |
| 1.126126 | 0.240372 | 0.417785 | 1D | 1.251426 | 1.17448 | 1.267403 | 0.353282 | 0.454095 |
| 1.065991 | 0.372398 | 0.130655 | 2D | 0.901454 | 1.02559 | 1.333873 | 0.230996 | 0.017663 |
| 0.992487 | 0.001225 | 0.064901 | 3D | 1.276971 | 0.981965 | 1.114372 | 0.00854 | 0.086061 |
| 0.954616 | 0.250306 | 0.026202 | 4D | 1.209289 | 1.50635 | 1.096238 | 0.199337 | 0.13164 |
| 1.003405 | 0.348651 | 0.458178 | 5D | 1.20318 | 1.247599 | 1.162863 | 0.376692 | 0.375746 |

| C Controls | p-value A vs B | p-value A vs C | | TIMP3 A SSc DPB1*1 | TIMP3 B SSc DPB1*1 | C Controls | p-value A vs B | p-value A vs C |
|---|---|---|---|---|---|---|---|---|
| 1.080674 | 0.133289 | 0.088324 | 1ug | 1.105412 | 1.204366 | 1.197412 | 0.313666 | 0.321832 |
| 1.133823 | 0.397169 | 0.233948 | 5ug | 0.496414 | 1.123663 | 1.228162 | 0.083588 | 0.067371 |
| 1.126126 | 0.110997 | 0.094867 | 10ug | 0.484787 | 1.166526 | 1.267403 | 0.12621 | 0.108486 |
| 1.067974 | 0.355782 | 0.084357 | 25ug | 0.923573 | 1.159702 | 1.19955 | 0.097649 | 0.069494 |
| 1.050754 | 0.489182 | 0.14437 | 50ug | 0.977905 | 1.19108 | 1.309581 | 0.109728 | 0.040623 |
| Controls | | | | SSc DPB1*1 | SSc DPB1*1 | Controls | | |
| 1.126126 | 0.110997 | 0.094867 | 1D | 0.484787 | 1.166526 | 1.267403 | 0.12621 | 0.108486 |
| 1.065991 | 0.37044 | 0.250854 | 2D | 0.687829 | 1.006748 | 1.333873 | 0.169051 | 0.063773 |
| 0.992487 | 0.215827 | 0.179714 | 3D | 0.969629 | 1.08517 | 1.114372 | 0.319787 | 0.281156 |
| 0.954616 | 0.33534 | 0.305345 | 4D | 6.212974 | 1.177179 | 1.096238 | 0.183798 | 0.174838 |
| 1.003405 | 0.10405 | 0.099112 | 5D | 1.274103 | 1.264498 | 1.162863 | 0.4889 | 0.368287 |

ARTHRITIS & RHEUMATISM
Vol. 60, No. 12, December 2009, pp 3807–3814
DOI 10.1002/art.24982


# HLA–DPB1 and DPB2 Are Genetic Loci for Systemic Sclerosis

## A Genome-Wide Association Study in Koreans With Replication in North Americans

Xiaodong Zhou,[1] Jong Eun Lee,[2] Frank C. Arnett,[1] Momiao Xiong,[3] Min Young Park,[2] Yeon Kyeong Yoo,[2] Eun Soon Shin,[2] John D. Reveille,[1] Maureen D. Mayes,[1] Jin Hyun Kim,[4] Ran Song,[4] Ji Yong Choi,[4] Ji Ah Park,[4] Yun Jong Lee,[4] Eun Young Lee,[4] Yeong Wook Song,[4] and Eun Bong Lee[4]

***Objective.*** **To identify systemic sclerosis (SSc) susceptibility loci via a genome-wide association study.**

***Methods.*** **A genome-wide association study was performed in 137 patients with SSc and 564 controls from Korea using the Affymetrix Human SNP Array 5.0. After fine-mapping studies, the results were replicated in 1,107 SSc patients and 2,747 controls from a US Caucasian population.**

***Results.*** **The single-nucleotide polymorphisms (SNPs) (rs3128930, rs7763822, rs7764491, rs3117230, and rs3128965) of HLA–DPB1 and DPB2 on chromosome 6 formed a distinctive peak with log *P* values for association with SSc susceptibility ($P = 8.16 \times 10^{-13}$). Subtyping analysis of HLA–DPB1 showed that DPB1*1301 ($P = 7.61 \times 10^{-8}$) and DPB1*0901 ($P = 2.55 \times 10^{-5}$) were the subtypes most susceptible to SSc in Korean subjects. In US Caucasians, 2 pairs of SNPs, rs7763822/rs7764491 and rs3117230/rs3128965, showed strong association with SSc patients who had either circulating anti–DNA topoisomerase I ($P = 7.58 \times 10^{-17}/4.84 \times 10^{-16}$) or anticentromere autoantibodies ($P = 1.12 \times 10^{-3}/3.2 \times 10^{-5}$), respectively.**

***Conclusion.*** **The results of our genome-wide association study in Korean subjects indicate that the region of HLA–DPB1 and DPB2 contains the loci most susceptible to SSc in a Korean population. The confirmatory studies in US Caucasians indicate that specific SNPs of HLA–DPB1 and/or DPB2 are strongly associated with US Caucasian patients with SSc who are positive for anti–DNA topoisomerase I or anticentromere autoantibodies.**

Systemic sclerosis (SSc) is a rare and complex connective tissue disease of unknown etiology characterized by fibrosis and vasculopathy of the skin and internal organs as well as several mutually exclusive, disease-specific circulating autoantibodies (1,2). SSc can be clinically subclassified based on patterns of skin fibrosis as limited cutaneous SSc (lcSSc) and diffuse cutaneous SSc (dcSSc) (3). In addition, the majority of SSc patients (90%) have circulating antinuclear autoantibodies (2). The 3 most common autoantibodies are anti–DNA topoisomerase I (anti–topo I), anti–RNA polymerase III, and anticentromere antibodies. The first 2 autoantibodies tend to be associated with dcSSc (2,4), and the last one is strongly correlated with lcSSc, although these associations are not complete (2,5).

Genetic predisposition is widely believed to con-

Supported by the Korean Health 21 Research & Development Project, the Ministry of Health, Welfare, and Family Affairs, Republic of Korea (grant A030001); the NIH (grants P50-AR-054144, NO1-AR-02251, and UL1-RR-024148 from the National Institute of Arthritis and Musculoskeletal and Skin Diseases), and the US Department of the Army, Medical Research Acquisition Activity (grants PR-064803 and PR-064851).

[1]Xiaodong Zhou, MD, Frank C. Arnett, MD, John D. Reveille, MD, Maureen D. Mayes, MD: University of Texas Health Science Center at Houston; [2]Jong Eun Lee, PhD, Min Young Park, Yeon Kyeong Yoo, Eun Soon Shin, PhD: DNA Link, Seoul, Republic of Korea; [3]Momiao Xiong, PhD: University of Texas, Houston; [4]Jin Hyun Kim, MD, Ran Song, MD, Ji Yong Choi, MS, Ji Ah Park, Yun Jong Lee, MD, PhD, Eun Young Lee, MD, PhD, Yeong Wook Song, MD, PhD, Eun Bong Lee, MD, PhD: Seoul National University, Seoul, Republic of Korea.

Drs. Zhou and Jong Eun Lee contributed equally to this work.

Address correspondence and reprint requests to Eun Bong Lee, MD, PhD, Department of Internal Medicine, Seoul National University College of Medicine, Seoul, Republic of Korea. E-mail: leb7616@snu.ac.kr.

Submitted for publication March 4, 2009; accepted in revised form August 18, 2009.

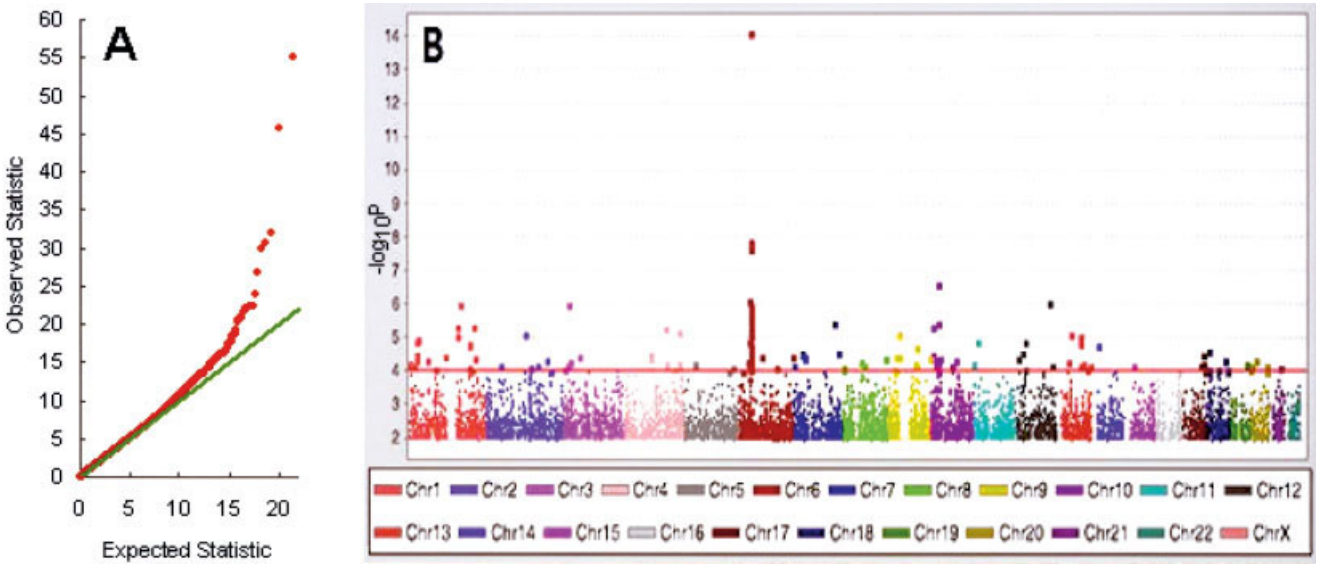



**Figure 1.** Identification of the major locus associated with systemic sclerosis (SSc) in a genome-wide scan of Korean subjects. A total of 440,734 single-nucleotide polymorphisms were evaluated in 133 patients with SSc and 557 healthy controls. **A**, Q–Q plot comparing the distribution of the Cochran-Armitage trend statistic with the expected null statistic. **B**, Distribution of $-\log_{10} P$ values plotted against chromosome 1 (Chr1) through chromosome 22 and chromosome X.

tribute to SSc. However, the low prevalence of SSc (~0.0007–0.049%) (6,7) and clinical/serologic heterogeneity make genetic studies of SSc difficult, with differing results reported for the same genes in different ethnic groups. Examples of such discrepancies are the reports of an association of the genes for connective tissue growth factor (8,9), protein tyrosine phosphatase N22 (10–13), and transforming growth factor $\beta$ (14–16) with SSc. Although some of these genes might have susceptibility markers for SSc in specific ethnic populations, the candidate gene approach that was used in those studies might miss other genes that could be more important to SSc susceptibility. In the present study, we used the genome-wide association scan approach to conduct a 2-step genetic association study in 4 independent populations to identify susceptibility markers for SSc.

## PATIENTS AND METHODS

**Study subjects.** We examined 4 different ethnic populations (Koreans, Caucasians, African Americans, and Hispanics). The Korean study population was composed of 151 patients diagnosed as having SSc according to the American College of Rheumatology (ACR; formerly, the American Rheumatism Association) preliminary criteria (17). All Korean patients were enrolled from Seoul National University Hospital between January 1998 and 2007. Genomic DNA was extracted from whole blood using standard methods. A total of 137 patients who passed the DNA quality check were entered into the genome-wide association study using Affymetrix Genome-Wide Human SNP Array 5.0. (Affymetrix, Santa Clara, CA). A total of 133 cases with call rates of >95% were finally entered into the case–control analysis. The mean age at diagnosis was 42 years (range 4–74 years). Mean disease duration was 10 years, and the mean time from diagnosis to blood sampling was 5 years. Anti–topo I antibodies were measured using enzyme-linked immunosorbent assay, and anticentromere antibodies were determined by passive immunodiffusion using the HEp-2 cell line. Of the 133 cases, 79 were positive and 48 were negative for anti–topo I. (Anti–topo I status was not determined in 6 cases.) Sixteen were positive and 117 were negative for anticentromere antibodies. When patients were classified by type of SSc, 66 had dcSSc and 67 had lcSSc (3).

The 564 healthy controls were randomly selected from 10,000 healthy Koreans belonging to the Korean Association Resource Project and were matched for sex with the cases. The mean age of the controls was 52.5 years. The same platform (Affymetrix Genome-Wide Human SNP Array 5.0) was used for the whole-genome scan of the controls. After excluding controls with call rates <95%, mismatched sex, and subjects who were potential relatives, a total of 557 controls were finally entered into the case–control analysis. The institutional review board of Seoul National University Hospital approved the study, and all patients and controls provided written consent.

Our study included 1,107 Caucasian, 70 African American, and 61 Hispanic patients who met the ACR criteria for SSc and 447 Caucasian, 90 African American, and 90 Hispanic controls, all of whom were enrolled in the Division of Rheumatology, University of Texas Health Science Center at Hous-

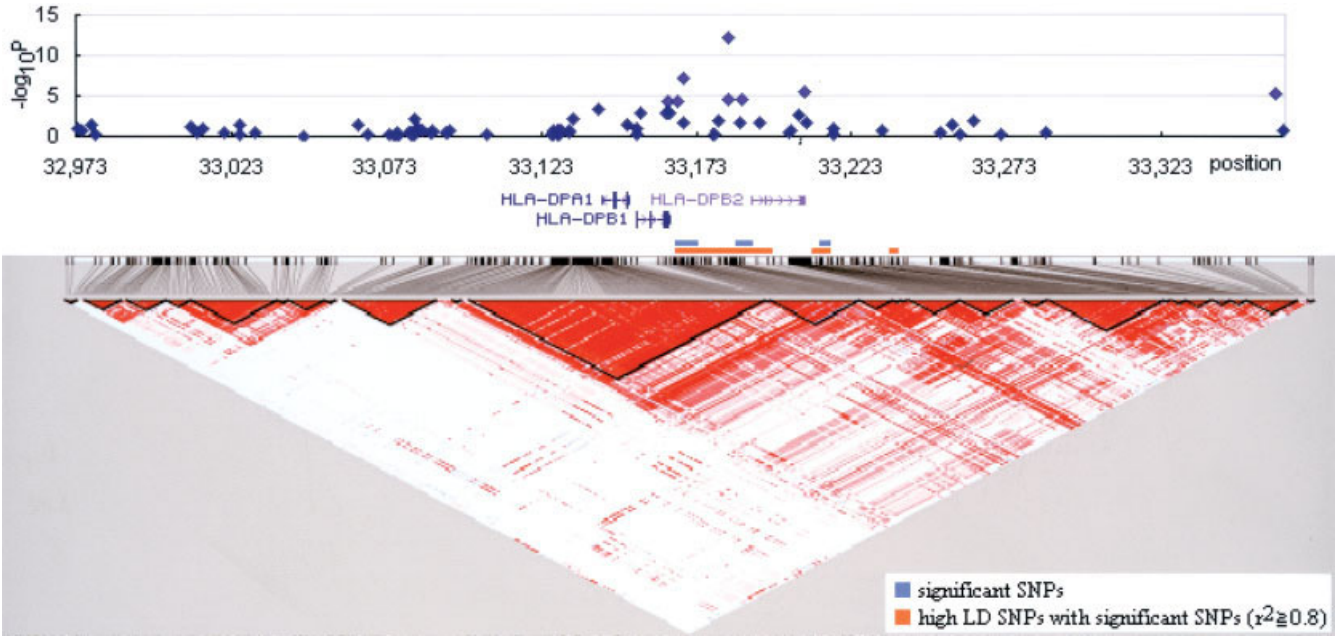



**Figure 2.** HLA–DPB as a candidate region for susceptibility to systemic sclerosis. The $-\log_{10} P$ values of individual single-nucleotide polymorphisms (SNPs) (diamonds) around the HLA–DPB region are shown. All genes in the region are also displayed above the linkage disequilibrium (LD) map. LD between pairs of SNPs is depicted with linkage blocks. Red blocks represent D′ of 1.0, and white blocks represent D′ of 0.0. Orange bars represent high LD SNPs that show association ($r^2 \geq 0.8$) with significant SNPs.

ton. In addition, 2,300 Caucasian controls were selected from the National Institutes of Health (NIH) database of Genotypes and Phenotypes (dbGaP; online at http://www.ncbi.nlm.nih.gov/gap). Enrolled SSc patients were routinely examined for circulating anti–topo I and anticentromere protein autoantibodies using passive immunodiffusion against calf thymus extracts or indirect immunofluorescence patterns on HEp-2 cells, respectively. Of the Caucasian patients, 183 were positive and 917 were negative for anti–topo I, and 316 were positive and 784 were negative for anticentromere antibodies. (Antibody status was not determined in 7 cases.) Diffuse cutaneous SSc was diagnosed in 419 patients, and lcSSc was diagnosed in 654 patients (3). The study was approved by the institutional review boards of the University of Texas Health Science Center at Houston. All patients and controls provided written consent.

**Genome-wide association analysis.** Among the 500,568 single-nucleotide polymorphisms (SNPs) present in Affymetrix Whole-Genome Human SNP Array 5.0, 440,734 SNPs were accessible after excluding hidden SNPs. A Q–Q plot was obtained when $P$ was >0.0001 for Hardy-Weinberg equilibrium and the call rate was >0.95 (Figure 1). The most significant SNPs were determined to be rs3128930, rs7763822, rs7764491, rs3117230, and rs3128965. These SNPs were included in the fine-mapping process.

**Fine-mapping studies.** For the Korean subjects, we performed a fine-mapping study focusing on HLA–DPB1 and DPB2 regions in 137 SSc cases and 548 healthy controls for whom DNA was available. For HLA–DPB1, the highly variable exon 2 of the gene was DNA-sequenced to determine the subtype of HLA–DPB1. For the other region including HLA–DPB2, a total of 22 tag SNPs were selected with $r^2$ threshold of 0.8 and minor allele frequency >5% in HapMap Japanese panel data (release 22) using Haploview version 4.1 (18,19). Of the tag SNPs, 17 SNPs that are included in Affymetrix SNP chip were forced to be included.

For replication studies, TaqMan assays were used for genotyping SNPs rs3128930, rs7763822, rs7764491, rs3117230, and rs3128965 with an ABI 7900HT Fast real-time polymerase chain reaction system in Caucasian, African American, and Hispanic populations. Genotyping results for all 5 SNPs passed quality tests for Hardy-Weinberg equilibrium ($P > 0.001$) and call rate (>95%). The 2 groups of Caucasian controls, from the NIH database and from the University of Texas Health Science Center at Houston, showed concordant association with Caucasian patients with SSc.

**Statistical analysis.** The association of specific SNPs with the disease or a subset of the disease was analyzed by comparing minor allele frequencies in the cases and controls, with significance determined by chi-square tests. The odds ratio (OR) of cases having a selected SNP compared with the controls and its relevant 95% confidence interval (95% CI) were also determined. For the genome-wide association study, the threshold for declaring significance after Bonferroni correction for multiple tests was $P < 1.43 \times 10^{-7}$. It was $P < 2.2 \times 10^{-3}$ for the fine-mapping study (n = 22) and $P < 0.01$ for the replication study (n = 5). All association tests were based on the comparison of alleles. Plink (20) and SAS, version 9.1.3 (SAS Institute, Cary, NC), were used for statistical analysis.

**Table 1.** Association of SNPs rs3128930, rs7764491, rs7763822, rs3128965, and rs3117230 with SSc in Korean patients with and without autoantibodies to DNA topoisomerase I and centromere protein*

| Allele (minor allele) | Patients with SSc | | | | | Controls (n = 557) |
|---|---|---|---|---|---|---|
| | Anti–topo I antibody positive (n = 79) | Anti–topo I antibody negative (n = 48) | Anticentromere antibody positive (n = 16) | Anticentromere antibody negative (n = 117) | All patients (n = 133) | |
| rs3128930 (A) | | | | | | |
| Frequency | 0.47 | 0.12 | 0.13 | 0.37 | 0.34 | 0.15 |
| *P*† | $8.26 \times 10^{-22}$ | 0.47 | 0.73 | $4.38 \times 10^{-15}$ | $8.16 \times 10^{-13}$ | – |
| OR (95% CI) | 5.21 (3.63–7.49) | 0.79 (0.41–1.51) | 0.83 (0.29–2.39) | 3.45 (2.50–4.75) | 3.00 (2.20–4.10) | – |
| rs7764491 (C) | | | | | | |
| Frequency | 0.21 | 0.065 | 0.063 | 0.17 | 0.15 | 0.056 |
| *P*† | $7.29 \times 10^{-12}$ | 0.71 | 0.87 | $8.24 \times 10^{-9}$ | $7.64 \times 10^{-8}$ | – |
| OR (95% CI) | 4.58 (2.87–7.33) | 1.18 (0.50–2.81) | 1.13 (0.26–4.83) | 3.42 (2.20–5.30) | 3.09 (2.01–4.75) | – |
| rs7763822 (T) | | | | | | |
| Frequency | 0.21 | 0.065 | 0.063 | 0.17 | 0.15 | 0.056 |
| *P*† | $7.29 \times 10^{-12}$ | 0.71 | 0.87 | $8.24 \times 10^{-9}$ | $7.64 \times 10^{-8}$ | – |
| OR (95% CI) | 4.58 (2.87–7.33) | 1.18 (0.50–2.81) | 1.13 (0.26–4.83) | 3.42 (2.20–5.30) | 3.09 (2.01–4.75) | – |
| rs3128965 (A) | | | | | | |
| Frequency | 0.25 | 0.054 | 0.063 | 0.20 | 0.18 | 0.093 |
| *P*† | $6.28 \times 10^{-9}$ | 0.21 | 0.56 | $4.37 \times 10^{-6}$ | $4.47 \times 10^{-5}$ | – |
| OR (95% CI) | 3.31 (2.17–5.04) | 0.56 (0.22–1.41) | 0.65 (0.15–2.76) | 2.44 (1.65–3.59) | 2.18 (1.49–3.18) | – |
| rs3117230 (G) | | | | | | |
| Frequency | 0.25 | 0.054 | 0.063 | 0.20 | 0.18 | 0.092 |
| *P*† | $4.51 \times 10^{-9}$ | 0.22 | 0.57 | $3.37 \times 10^{-6}$ | $3.52 \times 10^{-5}$ | – |
| OR (95% CI) | 3.34 (2.19–5.10) | 0.57 (0.22–1.43) | 0.66 (0.15–2.79) | 2.46 (1.67–3.64) | 2.20 (1.50–3.22) | – |

* Anti–DNA topoisomerase I (anti–topo I) antibody status was not determined in 6 patients. SNPs = single-nucleotide polymorphisms; SSc = systemic sclerosis; OR = odds ratio; 95% CI = 95% confidence interval.
† Versus controls.

Linkage disequilibrium analysis for HLA–DPB1 and DPB2 regions was performed with Haploview, version 4.1 (19).

## RESULTS

We first examined a Korean population, which is relatively homogenous, in which 62.2% of the patients were positive for anti–topo I. We performed a genome-wide association study in 133 Korean patients with SSc and 557 sex-matched Korean controls using the Affymetrix Human SNP Array 5.0 containing 440,734 accessible human SNPs. The genome-wide association scan showed a distinctive peak of SNP log *P* value ($P = 8.16 \times 10^{-13}$ for association with SSc) (Figure 1). The peak was formed with the SNPs rs3128930, rs7763822, rs7764491, rs3117230, and rs3128965, which were located in the region of HLA–DPB1 and DPB2 (a pseudogene) on chromosome 6p (Figures 1 and 2).

Fine-mapping analysis of this region confirmed that rs3128930, rs7763822, rs7764491, rs3117230, and rs3128965 were the SNPs that were associated with SSc in the Korean subjects (Figure 2 and Table 1). Interestingly, the association was even stronger in patients who were positive for anti–topo I ($P = 8.26 \times 10^{-22}$, OR 5.21 [95% CI 3.63–7.49] for rs3128930) (Table 1). These SNPs were also associated with dcSSc, but not with lcSSc (See Supplementary Table 1, available on the *Arthritis & Rheumatism* Web site at http://www3.interscience.wiley.com/journal/76509746/home.) Subtyping of HLA–DPB1 showed that HLA–DPB1*1301 (21.0% in SSc versus 5.5% in controls), DPB1*0901 (12.0% versus 2.6%), and DPB1*030101 (10.0% versus 4.3%) were present significantly more frequently in anti–topo I–positive patients than in controls (Table 2). SNPs corresponding to genes that have previously been shown to be associated with SSc, such as *PTPN22* (10), *AIF1* (21), *TNF* (22), *CTLA4* (23), and *CTGF* (8), fell within the significance thresholds of $10^{-5}$–$10^{-6}$ advocated for gene-based scans, as well as the Bonferroni correction for multiple comparisons (24).

To confirm these results, we used TaqMan assays to reexamine the 5 SNPs with the strongest association from the Korean genome-wide association scan in 1,107 US Caucasian patients with SSc, of whom 16% were positive for anti–topo I, and 2,747 normal controls, of whom 447 were from our local group (Houston, TX) and 2,300 were from the dbGaP. SNPs rs7763822 and rs7764491 showed highly significant associations with anti–topo I–positive SSc ($P = 7.58 \times 10^{-17}$ and $4.84 \times$

**Table 2.** Association of HLA–DPB1 allelic subtypes with SSc in Koreans*

| | Patients with SSc | | | | | |
|---|---|---|---|---|---|---|
| HLA allele | Anti–topo I antibody positive (n = 82) | Anti–topo I antibody negative (n = 49) | Anticentromere antibody positive (n = 16) | Anticentromere antibody negative (n = 121) | All patients (n = 137) | Controls |
| *1301 | | | | | | |
| Frequency | 0.21 | 0.05 | 0.031 | 0.17 | 0.15 | 0.055 |
| *P*† | $4.05 \times 10^{-12}$ | 0.88 | 0.56 | $3.25 \times 10^{-9}$ | $7.61 \times 10^{-8}$ | – |
| OR (95% CI) | 4.52 (2.86–7.14) | 0.93 (0.36–2.37) | 0.56 (0.074–4.15) | 3.42 (2.23–5.24) | 3.04 (1.99–4.63) | – |
| *0901 | | | | | | |
| Frequency | 0.12 | 0.01 | 0.031 | 0.087 | 0.08 | 0.026 |
| *P*† | $2.44 \times 10^{-8}$ | 0.33 | 0.87 | $7.55 \times 10^{-6}$ | $2.55 \times 10^{-5}$ | – |
| OR (95% CI) | 4.82 (2.64–8.82) | 0.38 (0.05–2.82) | 1.19 (0.16–8.99) | 3.50 (1.96–6.24) | 3.21 (1.82–5.69) | – |
| *030101 | | | | | | |
| Frequency | 0.10 | 0.02 | 0.031 | 0.083 | 0.077 | 0.043 |
| *P*† | $9.47 \times 10^{-4}$ | 0.28 | 0.75 | 0.01 | 0.021 | – |
| OR (95% CI) | 2.58 (1.44–4.61) | 0.47 (0.11–1.94) | 0.72 (0.096–5.39) | 2.01 (1.17–3.46) | 1.85 (1.09–3.16) | – |

* Anti–topo I antibody status was not determined in 6 patients. See Table 1 for definitions.
† Versus controls.

$10^{-16}$, respectively) (Table 3). The HLA–DPB1*1301 allele, which occurs in only 3% of US Caucasians, was found in 25% of anti–topo I–positive patients and conferred the strongest risk by exact logistic regression ($P = 0.0001$, OR 14) of any HLA class II allele (Arnett FC: unpublished observations). The SNPs rs3128965 and rs3117230 showed strong associations with anticentromere autoantibody–positive SSc ($P = 3.20 \times 10^{-5}$ and $1.12 \times 10^{-3}$, respectively) (Table 3). In addition, the pair of anti–topo I–associated SNPs showed a weaker association with dcSSc ($P = 0.0070$ and 0.014 for rs7763822 and rs7764491, respectively) (See Supplementary Table 2, available on the *Arthritis & Rheumatism* Web site at http://www3.interscience.wiley.com/journal/76509746/home.) The genetic concordance of the patients who were positive for anti–topo I and patients with dcSSc

**Table 3.** Association of SNPs of HLA–DPB1 and DPB2 with SSc in Caucasians*

| | Patients with SSc | | | | | |
|---|---|---|---|---|---|---|
| Allele (minor allele) | Anti–topo I antibody positive (n = 183) | Anti–topo I antibody negative (n = 917) | Anticentromere antibody positive (n = 316) | Anticentromere antibody negative (n = 784) | All patients (n = 1,107) | Controls (n = 2,731) |
| rs3128930 (A) | | | | | | |
| Frequency | 0.31 | 0.28 | 0.30 | 0.28 | 0.29 | 0.27 |
| *P*† | 0.12 | 0.30 | 0.056 | 0.47 | 0.14 | – |
| OR (95% CI) | 1.20 (0.95–1.51) | 1.01 (0.95–1.20) | 1.19 (1.00–1.43) | 1.05 (0.92–1.19) | 1.09 (0.97–1.21) | – |
| rs7764491 (C) | | | | | | |
| Frequency | 0.14 | 0.03 | 0.025 | 0.058 | 0.049 | 0.043 |
| *P*† | $4.84 \times 10^{-16}$ | 0.018 | 0.036 | 0.016 | 0.29 | – |
| OR (95% CI) | 3.56 (2.57–4.94) | 0.70 (0.52–0.94) | 0.58 (0.35–0.97) | 1.36 (1.06–1.75) | 1.14 (0.90–1.44) | – |
| rs7763822 (T) | | | | | | |
| Frequency | 0.14 | 0.031 | 0.024 | 0.059 | 0.049 | 0.043 |
| *P*† | $7.58 \times 10^{-17}$ | 0.021 | 0.023 | 0.0076 | 0.24 | – |
| OR (95% CI) | 3.65 (2.64–5.04) | 0.71 (0.53–0.95) | 0.55 (0.32–0.93) | 1.40 (1.09–1.80) | 1.15 (0.91–1.46) | – |
| rs3128965 (A) | | | | | | |
| Frequency | 0.14 | 0.19 | 0.25 | 0.16 | 0.18 | 0.18 |
| *P*† | 0.024 | 0.39 | $3.20 \times 10^{-5}$ | 0.010 | 0.99 | – |
| OR (95% CI) | 0.70 (0.52–0.96) | 1.06 (0.93–1.21) | 1.50 (1.24–1.82) | 0.82 (0.70–0.95) | 1.00 (0.88–1.14) | – |
| rs3117230 (G) | | | | | | |
| Frequency | 0.17 | 0.26 | 0.29 | 0.22 | 0.24 | 0.24 |
| *P*† | 0.0077 | 0.055 | $1.12 \times 10^{-3}$ | 0.32 | 0.42 | – |
| OR (95% CI) | 0.69 (0.52–0.91) | 1.13 (1.00–1.28) | 1.36 (1.13–1.63) | 0.93 (0.81–1.07) | 1.05 (0.93–1.18) | – |

* Autoantibody status was not determined in 7 patients. See Table 1 for definitions.
† Versus controls.

supports clinical observations that these 2 traits within SSc commonly overlap (2,4). However, the anticentromere antibody–associated SNP pairs did not show strong associations with lcSSc, which usually occurs in patients who are positive for anticentromere autoantibodies (2,5). The SNP rs3128930 showed only a marginal $P$ value of 0.041 for patients with lcSSc who were positive for anticentromere autoantibodies (Supplementary Table 2). Further analysis of Caucasian patients with SSc who were negative for anti–topo I or anticentromere autoantibodies indicated that they had marginal or no association with the genotypes of all 5 SNPs (Table 3).

In contrast, highly significant differences were observed in the comparisons of patients with and without anti–topo I or anticentromere autoantibodies using corresponding anti–topo I–associated or anticentromere-associated SNP pairs (rs7763822 and rs7764491 [$P \leq 2.86 \times 10^{-18}$] or rs3128965 and rs3117230 pair [$P \leq 4.77 \times 10^{-4}$], respectively) (See Supplementary Table 3, available on the *Arthritis & Rheumatism* Web site at http://www3.interscience.wiley.com/journal/76509746/home.) Interestingly, marginally significant differences ($0.05 < P \geq 0.01$) were also observed in the comparison of patients with lcSSc and dcSSc using both pairs of SNPs, but the autoantibody associations were the strongest, perhaps because HLA alleles are specific immune response genes (Table 3 and Supplementary Table 2).

Finally, we studied these same SNPs in limited numbers of African American and Hispanic cases and controls (70 cases versus 90 controls and 61 cases versus 90 controls, respectively). Although the numbers of subjects examined in these 2 populations were small, the SNPs rs7764491 and rs7763822 showed a consistently strong association with anti–topo I–positive SSc in both African American subjects ($P = 9.05 \times 10^{-3}$, OR 4.23 [95% CI 1.33–13.48] for both SNPs) and Hispanic subjects ($P = 7.98 \times 10^{-4}$, OR 5.51 [95% CI 1.85–16.43] and $P = 7.21 \times 10^{-4}$, OR 5.57 [95% CI 1.87–16.62], respectively) (See Supplementary Tables 4 and 5, available on the *Arthritis & Rheumatism* Web site at http://www3.interscience.wiley.com/journal/76509746/home.)

## DISCUSSION

Previous studies of SSc, a complex disease, have resulted in inconsistent findings of genetic associations in different study populations (8–16). The rarity and the heterogeneity of SSc may contribute, at least in part, to such discrepancies. In the present study, we applied a 2-step approach, in which a genome-wide association scan was first conducted in a relatively homogenous population of Korean patients with SSc and was then followed by a confirmatory study in other populations. Our findings that SSc is associated with HLA–DPB1 and DPB2 based on autoantibody status represent the first replicable results in Asians, Caucasians, African Americans, and Hispanics. These complementary studies in 4 independent populations provide strong support for the notion that the identified genetic markers confer susceptibility to subtypes of SSc.

HLA–DPB1, located centromeric to other HLA class II molecules, shows relatively low linkage disequilibrium with other extended major histocompatibility complex haplotypes (25). Although it has not been studied as extensively as HLA–A, B, C, or DR, it has been shown to have a similar antigen-presenting function to activate CD4+ T cells (26). Its genetic polymorphisms have been found to be associated with chronic berylliosis (27), graft-versus-host disease (28), juvenile rheumatoid arthritis (29), type 1 diabetes mellitus (30), and sarcoidosis (31). Some studies have suggested the possible roles of HLA–DPB1 in SSc, in the context of HLA–A, B, C, and DR molecules (32–34). Typically, HLA–DPB1*1301 was previously reported to be associated with anti–topo I in SSc patients in several independent studies (34–36), although each of these studies appeared to have less statistical power and a smaller sample size than the present study. The results of our Korean genome-wide association study and US confirmatory study revealed that specific SNPs of HLA–DPB1 and DPB2 are strongly associated with SSc, especially in those patients with autoantibodies to topo I or centromere protein, and that HLA–DPB1 may be the most important SSc susceptibility gene in the Korean population.

Our results indicate that disease subtype and population origin may be 2 critical factors in identifying genetic susceptibility markers for SSc. This is evidenced by the lack of association of SSc in a Caucasian population with any of the 5 SNPs found in the Korean genome-wide association study except when disease subgroup (dcSSc or lcSSc), and especially specific autoantibody status (anti–topo I or anticentromere autoantibody positivity), were taken into account. In addition, this notion may explain previous inconsistent reports of SSc-associated genes in different populations, as well as the fact that these genes were not identified in our genome-wide association study of Korean subjects.

Importantly, the genetic differentiations of SSc according to distinctive serologic and clinical features in our studies suggest that SSc should not be considered a single disease. While SSc, like other complex human

diseases, is currently incurable, subclassification of SSc on the basis of genetic polymorphisms for disease susceptibility may provide a new dimension for exploring the pathogenesis and treatment of this disease. Our results strongly indicate that subclassification of SSc on the basis of autoantibodies against topo I and centromere proteins is important in defining disease susceptibility genes in this heterogeneous disease.

Despite the successful identification of the SSc susceptibility genes HLA–DPB1 and DPB2, our study has several limitations. First, the number of SSc patients included in the genome-wide association study may not be large enough to distinguish some other true disease risk SNPs from the statistical noise of false-positive SNPs by using a stringent statistical threshold. The power of the genome-wide association study, calculated with the assumption of a 10% difference in minor allele frequency, was 75.0% (37). Second, the majority of the Korean patients with SSc were positive for anti–topo I autoantibodies. Accordingly, genetic predisposition factors for SSc with other SSc autoantibodies may be better unveiled by a genome-wide association study in other populations. Third, the numbers of Hispanic and African American subjects may not have been large enough to give the study adequate power.

Nonetheless, the SNPs of HLA–DPB1 and DPB2 showed a distinctively strong association with anti–topo I–positive SSc in our 2-step genetic study. The results confirmed previous reports of the association of this region with SSc susceptibility and raised the importance of subclassification of SSc for the understanding of disease pathogenesis.

## ACKNOWLEDGMENTS

This work could not have been accomplished without the generous provision of genome-wide association scan data and DNA from healthy Koreans from the Korean Center for Disease Control. We thank Hua Dong for valuable help in statistical analysis.

### AUTHOR CONTRIBUTIONS

All authors were involved in drafting the article or revising it critically for important intellectual content, and all authors approved the final version to be published. Dr. Eun Bong Lee had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

**Study conception and design.** Zhou, Jong Eun Lee, Arnett, Eun Bong Lee.

**Acquisition of data.** Zhou, Jong Eun Lee, Arnett, Xiong, Reveille, Mayes, Kim, Ran Song, Choi, Ji Ah Park, Yun Jong Lee, Eun Young Lee, Yeong Wook Song, Eun Bong Lee.

**Analysis and interpretation of data.** Zhou, Jong Eun Lee, Arnett, Xiong, Min Young Park, Yoo, Shin, Eun Bong Lee.

## REFERENCES

1. Korn JH. Pathogenesis of systemic sclerosis. In: Koopman WJ, editor. Arthritis and allied conditions. 15th ed. Philadelphia: Lippincott Williams & Wilkins; 2005. p. 1621–32.
2. Bunn CC, Black CM. Systemic sclerosis: an autoantibody mosaic. Clin Exp Immunol 1999;117:207–8.
3. LeRoy EC, Black C, Fleischmajer R, Jablonska S, Krieg T, Medsger TA Jr, et al. Scleroderma (systemic sclerosis): classification, subsets and pathogenesis. J Rheumatol 1988;15:202–5.
4. Jarzabek-Chorzelska M, Blaszczyk M, Jablonska S, Chorzelski T, Kumar V, Beutner EH. Scl 70 antibody–a specific marker of systemic sclerosis. Br J Dermatol 1986;115:393–401.
5. Moroi Y, Peebles C, Fritzler MJ, Steigerwald J, Tan EM. Autoantibody to centromere (kinetochore) in scleroderma sera. Proc Natl Acad Sci U S A 1980;77:1627–31.
6. Chifflot H, Fautrel B, Sordet C, Chatelus E, Sibilia J. Incidence and prevalence of systemic sclerosis: a systematic literature review. Semin Arthritis Rheum 2008;37:223–35.
7. Mayes MD, Lacey JV Jr, Beebe-Dimmer J, Gillespie BW, Cooper B, Laing TJ, et al. Prevalence, incidence, survival, and disease characteristics of systemic sclerosis in a large US population. Arthritis Rheum 2003;48:2246–55.
8. Fonseca C, Lindahl GE, Ponticos M, Sestini P, Renzoni EA, Holmes AM, et al. A polymorphism in the CTGF promoter region associated with systemic sclerosis. N Engl J Med 2007;357:1210–20.
9. Rueda B, Simeon C, Hesselstrand R, Herrick A, Worthington J, Ortego-Centeno N, et al. A large multicenter analysis of CTGF -945 promoter polymorphism does not confirm association with systemic sclerosis susceptibility or phenotype. Ann Rheum Dis 2008. E-pub ahead of print.
10. Gourh P, Tan FK, Assassi S, Ahn CW, McNearney TA, Fischbach M, et al. Association of the PTPN22 R620W polymorphism with anti–topoisomerase I– and anticentromere antibody—positive systemic sclerosis. Arthritis Rheum 2006;54:3945–53.
11. Kyogoku C, Langefeld CD, Ortmann WA, Lee A, Selby S, Carlton VE, et al. Genetic association of the R620W polymorphism of protein tyrosine phosphatase PTPN22 with human SLE. Am J Hum Genet 2004;75:504–7.
12. Viken MK, Amundsen SS, Kvien TK, Boberg KM, Gilboe IM, Lilleby V, et al. Association analysis of the 1858C>T polymorphism in the PTPN22 gene in juvenile idiopathic arthritis and other autoimmune diseases. Genes Immun 2005;6:271–3.
13. Wipff J, Allanore Y, Kahan A, Meyer O, Mouthon L, Guillevin L, et al. Lack of association between the protein tyrosine phosphatase non-receptor 22 (PTPN22)*620W allele and systemic sclerosis in the French Caucasian population. Ann Rheum Dis 2006;65:1230–2.
14. Crilly A, Hamilton J, Clark CJ, Jardine A, Madhok R. Analysis of transforming growth factor $\beta$1 gene polymorphisms in patients with systemic sclerosis. Ann Rheum Dis 2002;61:678–81.
15. Lee EB, Kim JY, Lee YJ, Abdallah A, Lympany P, Song YW. Transforming growth factor-$\beta$1 polymorphisms in Korean patients with systemic sclerosis. Tissue Antigens 2004;63:491–5.
16. Sugiura Y, Banno S, Matsumoto Y, Niimi T, Yoshinouchi T, Hayami Y, et al. Transforming growth factor $\beta$1 gene polymorphism in patients with systemic sclerosis. J Rheumatol 2003;30:1520–37.
17. Subcommittee for Scleroderma Criteria of the American Rheumatism Association Diagnostic and Therapeutic Criteria Committee. Preliminary criteria for the classification of systemic sclerosis (scleroderma). Arthritis Rheum 1980;23:581–90.
18. Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, Gibbs RA, et al. A second generation human haplotype map of over 3.1 million SNPs. Nature 2007;449:851–61.
19. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and

visualization of LD and haplotype maps. Bioinformatics 2005;21: 263–5.
20. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 2007; 81:559–75.
21. Alkassab F, Gourh P, Tan FK, McNearney T, Fischbach M, Ahn C, et al. An allograft inflammatory factor 1 (AIF1) single nucleotide polymorphism (SNP) is associated with anticentromere antibody positive systemic sclerosis. Rheumatology (Oxford) 2007;46: 1248–51.
22. Pandey JP, Takeuchi F. TNF-$\alpha$ and TNF-$\beta$ gene polymorphisms in systemic sclerosis. Hum Immunol 1999;60:1128–30.
23. Takeuchi F, Kawasugi K, Nabeta H, Mori M, Tanimoto K. Association of CTLA-4 with systemic sclerosis in Japanese patients. Clin Exp Rheumatol 2002;20:823–8.
24. Thomas DC, Clayton DG. Betting odds and genetic associations. J Natl Cancer Inst 2004;96:421–3.
25. Begovich AB, McClure GR, Suraj VC, Helmuth RC, Fildes N, Bugawan TL, et al. Polymorphism, recombination, and linkage disequilibrium within the HLA class II region. J Immunol 1992; 148:249–58.
26. McCanlies EC, Kreiss K, Andrew M, Weston A. HLA-DPB1 and chronic beryllium disease: a HuGE review. Am J Epidemiol 2003;157:388–98.
27. Richeldi L, Sorrentino R, Saltini C. HLA-DPB1 glutamate 69: a genetic marker of beryllium disease. Science 1993;262:242–4.
28. Shaw BE, Potter MN, Mayor NP, Pay AL, Smith C, Goldman JM, et al. The degree of matching at HLA-DPB1 predicts for acute graft-versus-host disease and disease relapse following haematopoietic stem cell transplantation. Bone Marrow Transplant 2003; 31:1001–8.
29. Begovich AB, Bugawan TL, Nepom BS, Klitz W, Nepom GT, Erlich HA. A specific HLA-DP $\beta$ allele is associated with pauciarticular juvenile rheumatoid arthritis but not adult rheumatoid arthritis. Proc Natl Acad Sci U S A 1989;86:9489–93.
30. Cruz TD, Valdes AM, Santiago A, Frazer de Llado T, Raffel LJ, Zeidler A, et al. DPB1 alleles are associated with type 1 diabetes susceptibility in multiple ethnic groups. Diabetes 2004;53:2158–63.
31. Lympany PA, Petrek M, Southcott AM, Newman Taylor AJ, Welsh KI, du Bois RM. HLA-DPB polymorphisms: Glu 69 association with sarcoidosis. Eur J Immunogenet 1996;23:353–9.
32. Kuwana M, Inoko H, Kameda H, Nojima T, Sato S, Nakamura K, et al. Association of human leukocyte antigen class II genes with autoantibody profiles, but not with disease susceptibility in Japanese patients with systemic sclerosis. Intern Med 1999;38:336–44.
33. Tikly M, Rands A, McHugh N, Wordsworth P, Welsh K. Human leukocyte antigen class II associations with systemic sclerosis in South Africans. Tissue Antigens 2004;63:487–90.
34. Reveille JD, Brady J, MacLeod-St Clair M, Durban E. HLA-DPB1 alleles and autoantibody subsets in systemic lupus erythematosus, Sjögren's syndrome and progressive systemic sclerosis: a question of disease relevance. Tissue Antigens 1992;40:45–8.
35. Rihs HP, Conrad K, Mehlhorn J, May-Taube K, Welticke B, Frank KH, et al. Molecular analysis of HLA-DPB1 alleles in idiopathic systemic sclerosis patients and uranium miners with systemic sclerosis. Int Arch Allergy Immunol 1996;109:216–22.
36. Gilchrist FC, Bunn C, Foley PJ, Lympany PA, Black CM, Welsh KI, et al. Class II HLA associations with autoantibodies in scleroderma: a highly significant role for HLA-DP. Genes Immun 2001;2:76–81.
37. Lachin JM. Introduction to sample size determination and power analysis for clinical trials. Control Clin Trials 1981;2:93–113.

OPEN ACCESS Freely available online

PLoS one

# Differential Dynamic Properties of Scleroderma Fibroblasts in Response to Perturbation of Environmental Stimuli

**Momiao Xiong[1], Frank C. Arnett[2], Xinjian Guo[2], Hao Xiong[3], Xiaodong Zhou[2]***

**1** Human Genetics Center, School of Public Health, The University of Texas Health Science Center at Houston (UTHSC-H), Houston, Texas, United States of America, **2** Division of Rheumatology and Clinical Immunogenetics, Medical School, The University of Texas Medical School at Houston, The University of Texas Health Science Center at Houston (UTHSC-H), Houston, Texas, United States of America, **3** Department of Computer Science, Texas A&M University, College Station, Texas, United States of America

## Abstract

Diseases are believed to arise from dysregulation of biological systems (pathways) perturbed by environmental triggers. Biological systems as a whole are not just the sum of their components, rather ever-changing, complex and dynamic systems over time in response to internal and external perturbation. In the past, biologists have mainly focused on studying either functions of isolated genes or steady-states of small biological pathways. However, it is systems dynamics that play an essential role in giving rise to cellular function/dysfunction which cause diseases, such as growth, differentiation, division and apoptosis. Biological phenomena of the entire organism are not only determined by steady-state characteristics of the biological systems, but also by intrinsic dynamic properties of biological systems, including stability, transient-response, and controllability, which determine how the systems maintain their functions and performance under a broad range of random internal and external perturbations. As a proof of principle, we examine signal transduction pathways and genetic regulatory pathways as biological systems. We employ widely used state-space equations in systems science to model biological systems, and use expectation-maximization (EM) algorithms and Kalman filter to estimate the parameters in the models. We apply the developed state-space models to human fibroblasts obtained from the autoimmune fibrosing disease, scleroderma, and then perform dynamic analysis of partial TGF-$\beta$ pathway in both normal and scleroderma fibroblasts stimulated by silica. We find that TGF-$\beta$ pathway under perturbation of silica shows significant differences in dynamic properties between normal and scleroderma fibroblasts. Our findings may open a new avenue in exploring the functions of cells and mechanism operative in disease development.

**Citation:** Xiong M, Arnett FC, Guo X, Xiong H, Zhou X (2008) Differential Dynamic Properties of Scleroderma Fibroblasts in Response to Perturbation of Environmental Stimuli. PLoS ONE 3(2): e1693. doi:10.1371/journal.pone.0001693

**Editor:** Raya Khanin, University of Glasgow, Germany

**Received** May 2, 2007; **Accepted** January 31, 2008; **Published** February 27, 2008



**Funding:** Supported by UTHSC-H grants from NIH/NIAMS Center of Research Translation (CORT) in Scleroderma (1 P50 AR054144) (F.C.A) and 5RO3AR050517-02 (X.Z.), NIH/NCRR Center for Clinical and Translational Sciences (Houston CTSA Program) (1UL1 RR 024148), NIH PHS NCRR GCRC grant M01RR002558, University of Texas Health Science Center at Houston CreFF (X.Z.), Department of the Army, Medical Research Acquisition Activity: PR064803 (X.Z) and the Scleroderma Foundation (X.Z.).

**Competing Interests:** The authors have declared that no competing interests exist.

*E-mail: xiaodong.zhou@uth.tmc.edu

## Introduction

Identifying differentially expressed genes across distinct conditions and clustering co-expressed genes into different functional groups have been general approaches for unraveling molecular mechanisms involved in disease pathogenesis [1]. Although these approaches are valuable for looking at isolated events and their correlations, they do not explain the behavior of a bio-system.

Another approach to deciphering pathogenesis of complex diseases is system thinking. Human complex diseases are believed to arise from malfunction of a specific biological system, rather than from isolated events. It is increasingly recognized that biological systems as a whole are not just the sum of their components but, rather, ever-changing, complex, interacted and dynamic systems over time in response to internal events and environmental stimuli [2]. Cellular functions, such as growth, differentiation, division and apoptosis, and biological phenomena of the entire organisms are not only determined by steady-state characteristics of the biological systems, but also determined by inherent dynamic properties of biological systems. Dynamic properties include stability, transient-response, observability and controllability, which determine how the systems maintain their functions and performance under a broad range of random internal and external perturbations. Similar to differential expression of genes between normal and abnormal tissues, we can also observe the differential dynamic properties of the biological systems across different types of tissues and conditions. Dynamic properties are correlated with the health status of individuals and are of central importance for comprehensively understanding human biological systems and ultimately complex diseases.

The dynamic behavior of most complex biological systems emerges from the orchestrated activity of many components (e.g. genes and proteins) that interact with each other to form complicated biological networks involving gene regulation and signal transduction [3]. The nodes and links together are referred to as networks. This report is a study of gene regulatory network that focuses on dynamic properties of a biological system.

Investigation of dynamic properties of gene networks has three major tasks: development of mathematic models, estimation of the parameters in the models, and dynamic analysis. Mathematical modeling is to use mathematical language to describe the dynamic characteristics of a system [4]. In the past decade, various methods have been developed to model gene networks, including Boolean networks [5–7], differential equations and Bayesian networks [8–11], and vector autoregressive model [12]. A very powerful approach in modeling complex systems is the state-space approach [13,14], which is a special subclass of dynamic Bayesian networks. It provides a general framework for the application of dynamic systems theory in the analysis of gene regulation. The state-space approach is the core of modern systems theory. Application of the state-space equations to modeling gene networks allows us to use a large body of methodologies and tools in dynamic systems theory for studying dynamics of gene networks. We use Kalman Filter and Expectation-Maximization (EM) to estimate the parameters in the model [15,16]. After state-space model of the gene networks is established, the next task is to perform dynamic analysis for the model in response to perturbation of internal and external stimuli. Dynamic analysis attempts to extract inherent features of the systems that capture and describe the behaviors of the system over time under different operating conditions. The most important operating principle of a dynamic system is its stability (i. e., the ability to return to the original state or equilibrium state after perturbation). The concept of stability can be easily illustrated by the example of a marble sitting at the bowl. When the marble is in the bottom of the bowl it is stable. No matter where the marble is pushed, up the side of the bowl or from the bottom of the bowl, after it is released, the marble will finally settle to the bottom of the bowl at the original, stable equilibrium point. However, when the marble is on the top of an inverted bowl, it is unstable. The marble can remain on the top of the bowl only when the forces acting on the marble on the top of the bowl is completely balanced. Any slight perturbation in the marble's steady state will destroy the balance of the marble and cause it to move away from the top of the bowl. This indicates that when the system is in unstable state small perturbation can cause the system move away from the steady-state [17]. The biological systems are in constant change under the influences of genetic and environmental differences. The ability of the systems to maintain the stable states after perturbation and to resist diverse disturbance of the internal and external forces is critical to the viability of living organisms and plays a central role in biology [18,19]. Consequently, studying stability of biological systems is of great importance for discovering mechanism of complex diseases. Although there has been long history to investigate the stability of biologic systems, to our knowledge, very few studies have been reported on stability of gene networks. Particularly, the relationship between stability of gene networks and status of diseases has not been explored. One of purpose of this paper is to use gene expression data to show that similar to the example of the marble in the bowl, the gene networks will also have stable and unstable states and that unstable gene networks may be associated with the diseases.

Another important property of the dynamic systems is the transient response to disturbance of internal noises and external environmental forces, which measures how fast the systems respond to the perturbation and characterizes damping and oscillation properties of the process in response to the perturbation [13]. Feedback close loops are the basis for maintaining normal function of cells and organisms in the face of internal and external perturbation [19,20]. The essential feature of the transient response of a feedback closed-loop system largely depends on the location of the closed-loop poles. A simple and popular method for searching the poles of the closed-loop system is the root-locus analysis that plots a curve of the location of the poles of a transfer function of the feedback system over the range of the variable (usually loop gain) to determine whether the system will become unstable or oscillate [13]. The third important property of a dynamic system is controllability. Controllability is defined as the capacity of the system to move from undesired states to certain desired final states in finite time through accessible inputs [21]. Germline or somatic mutations lead to the subsequent transcriptional and translational alterations which will affect the phenotypes of the cells and cause diseases. Therapeutic interventions such as radiation, drug and gene therapy intend to alter gene expressions from an undesired state or abnormal state to a desired or normal state. Theoretic and practical analyses in modern control theory demonstrate that there exist systems which we are not able to change from undesired states to desired states. Now the question arises: are all genetic networks controllable? Can always therapeutic interventions change levels of gene expressions to desired states? Controllability provides answers to these questions. It provides a convenient and sufficient criterion for assessing whether we can change undesired gene expression levels to desired gene expression levels. Controllability describes the ability of biological systems to adapt to the changes of environments and deeply characterizes the internal structure of the system. The controllability of the biological networks may reflect the severity of the disease. Thus, the controllability is a fundamental design principle of biological system.

In summary, stability, transient response, feedback and controllability are basic dynamic properties of the biological systems and are essential to the function of the cells and organisms. As a proof of principle, in this report we investigate the differential dynamic properties of TGF-$\beta$ pathway in response to perturbation of silica between normal and scleroderma fibroblasts. Scleroderma or systemic sclerosis (SSc) is a typical complex disease in which fibrosis occurs in multiple organs. Although etiopathogenesis is unknown, both genetic and environmental factors are believed to play critical roles. The major source of fibrosis in SSc is over production of collagens from fibroblasts. Fibroblasts obtained from SSc patients appeared to be genetically engineered to produce more collagens and cytokines [22]. Silica exposure is an important environmental risk factor in some cases, which has been found in association with the development and perturbation of SSc [23]. Subcutaneous injections of silica have been reported to induce sclerodermatous skin changes and activation of skin fibroblasts [23]. Therefore, interactions between fibroblasts and silica may represent a magnification of biological events occurring in SSc and/or SSc-like disorders. The biological system of fibroblasts reacting to silica exposure must involve complex regulations and coordination of molecules to maintain their desirable status. Although multiple experiments of the *in vivo* and the *in vitro* response to silica particles have revealed that fibroblasts are activated to produce more collagens and other extracellular matrix (ECM) components [24–26], there is a lack of a mathematical model to quantify interactions among the molecules, and to predict dynamic behaviors of this bio-system. The purpose of this report is to use gene expression responses of scleroderma and normal fibroblasts exposed to the stimulus of silica as an example to address the issue of differential dynamic properties of the biological systems in response to perturbation by environments across different conditions. To accomplish this, we first formulate a regulatory network involving TGFBRII, CTGF, SPARC, COL1A2, COL3A1 and TIMP3 as a biological system that is associated with TGF-$\beta$ signaling, and then apply mathematical methods and computational algorithms from engineering and

control theory [13] to perform dynamic analysis of this network for both normal and scleroderma fibroblasts in response to perturbation of environmental Stimuli. Based on the results of dynamic network analysis, we examine the differential dynamic properties of this network between normal and scleroderma fibroblasts and reveal the relationship between the dynamic properties of gene networks and the phenotypes of the cells.

## Results

### State Space Model of a gene network responding to silica

Gene regulation involves a large number of biochemical events. Although kinetic models can be developed for gene regulation [12,27,28], they involve many kinetic parameters that are difficult to be estimated from gene expression data with small number of samples. An alternative model of gene expression is a state-space model. It can effectively deal with time invariant or time varying, linear or nonlinear complex systems with multiple inputs and outputs. A state-space model includes three types of variables: input variables, output variables and state variables. A key idea behind state-space model is the concept of the state. The state of a dynamic regulatory system is the smallest set of variables which are referred to as state variables such that the current knowledge of these variables together with the current and future knowledge of the input variables (environments or controls) will completely determine the behavior of the regulatory system. All state variables are hypothetical variables. State variables represent biological forces to regulate transcription of genes, which describe the behavior of gene transcription. Since the mechanisms of gene regulation in the network have still not been well understood, the state variables that determine the regulation may be unknown and hidden in the regulatory process, the concept of state variables is very suitable for description of the regulatory process. The expression levels of genes are output variables and can be observed. The expression levels of the genes are determined by the state variables, which describe states of regulation of the gene expressions.

Previously, we found that the SPARC (secreted protein, acidic, and rich in cysteine) gene is involved in the regulation of extracellular matrix genes such as *COL1A2*, *COL3A1*, *CTGF* and *TIMP3*, and this regulation is associated with activation of the TGF-$\beta$ pathway [29,30]. We used this partial TGF-$\beta$ pathway as an example to illustrate how to perform dynamic analysis of biological networks. This regulatory network was modeled by linear state-space equations defined as:

$$\begin{aligned} x_{k+1} &= Ax_k + Bu_k + w_k \\ y_k &= Cx_k + Du_k + v_k \end{aligned} \tag{1}$$

where $x_k$ is the vector of state variables that describes the behavior of gene regulation, but are hidden; $y_k$ is the output vector whose elements denote the measured gene expression levels; $u_k$ is the input vector; $w$ and $v$ are noises assumed to be white Gaussian noise with zero means and variance $Q$ and $R$ respectively, and they are independent of each other. The inputs can be any external stimuli that influence gene regulation, things like environmental forces, drugs, proteins, RNAs, or the effects from the genes outside the model. Matrix $A$ is called state transition matrix whose elements denote the regulatory strength of one gene on another, $B$ is input to state matrix whose elements quantify the regulatory effects of the input variables on the genes of the network, $C$ is state to output matrix whose elements quantify the dependence of the measured gene expression levels on the hidden regulatory states, and $D$ is input to output matrix whose elements measures the strength of dependence of the observed gene expression levels on the inputs. Matrices $A$, $B$, $C$, $D$ and variance matrices $Q$ and $R$ together make up the parameters of the dynamical system for gene regulatory networks.

We performed experiments on cultured human fibroblasts. We have 5 pairs each of normal and SSc patients' samples. For each sample, we have two replications were perturbed by Silicon. The transcript levels of six genes: *SPARC*, *CTGF*, *COL1A2*, *COL3A1*, *TIMP3* and *TGFBRII were* measured daily from 1 to 5 days. Let $x_1$, $x_2$, $x_3$, $x_4$ and $x_5$ be the expression levels of the *SPARC*, *TIMP3*, *COL3A1*, *CTGF* and *COL1A2*, respectively. Let $u_1$ and $u_2$ be the expression of the *TGFBRII* [31] and 10 μg silica. The estimated state-space model for the normal cell line and SSc are respectively, given by

$$\begin{aligned} x_1(k+1) &= 0.84603x_1(k) + 0.03207u_1(k) - 0.09050u_2 \\ x_2(k+1) &= 0.32847x_2(k) - 0.14840u_1(k) + 1.41049u_2 \\ x_3(k+1) &= 5.21569x_1(k) + 0.23997x_3(k) - 1.39180x_4(k) - \\ & 0.94591u_2 \\ x_4(k+1) &= 1.27852x_1(k) - 0.24027x_4(k) + 0.19823u_2 \\ x_5(k+1) &= 0.47401x_1(k) + 0.01406x_4(k) + 0.66237x_5(k) - \\ & 0.21877u_2 \end{aligned}$$

and

$$\begin{aligned} x_1(k+1) &= 0.63654x_1(k) + 0.25128u_1(k) + 0.49013u_2 \\ x_2(k+1) &= 0.78374x_2(k) + 0.07741u_1(k) + 0.11173u_2 \\ x_3(k+1) &= -0.61896x_1(k) + 1.106627x_3(k) + 0.44131x_4(k) + \\ & 0.25307u_2 \\ x_4(k+1) &= -0.06644x_1(k) + 1.00586x_4(k) + 0.00700u_2 \\ x_5(k+1) &= -0.82173x_1(k) + 0.28944x_4(k) + 1.46736x_5(k) + \\ & 0.26477u_2 \end{aligned}$$

A graph will be used to represent a genetic network. The nodes in the graph will represent the variables that correspond to the expressions of the genes. The edge between two nodes denotes that two variables are dependent. The number next to edges is the elements of the parameter matrices A, B, C and D in the state-space model. The estimated state-space model is shown in Figure 1 where the numbers next to the edges are the coefficients in the above equation for the normal (black color) and SSc (red color) fibroblasts, respectively. We observe differential regulation of *SPRAC* on *CTGF*, *COL3A1* and *COL1A2*, and *CTGF* on *COL3A1* between the SSc and normal fibroblasts. Figure 1 and above equations also demonstrate that the effects of silica (environmental factor) on *TIMP3*, *COL3A1* and *COL1A2* between the SSc and normal fibroblasts are different. Their coefficients in the state-space equations for the normal fibroblasts are negative, but become positive for the SSc fibroblasts. This implies that perturbation of scleroderma fibroblasts by silica will increase the expressions of *COL1A2* and *COL3A1*. This statement can be supported by early observation that excessive amounts of various collagens mainly type I and type III collagens were generated in the fibroblasts from affected scleroderma skin [32–34].

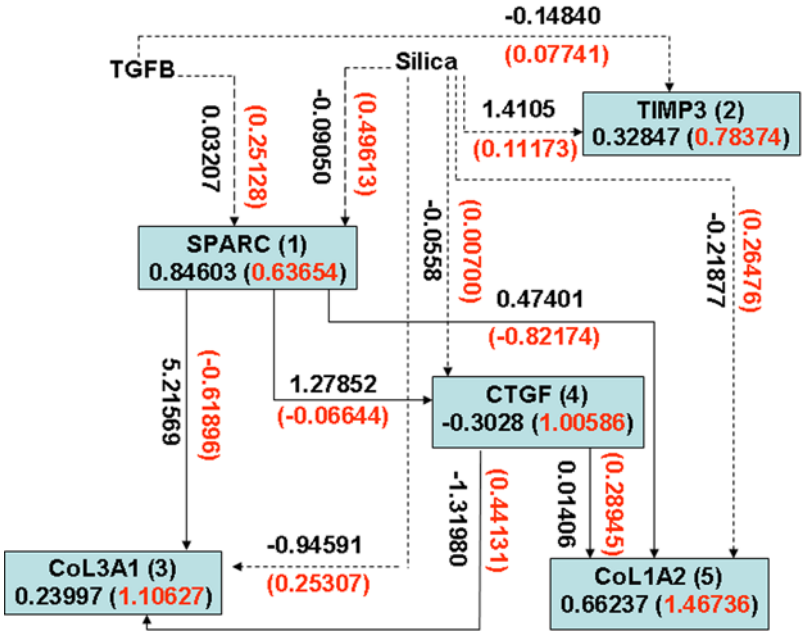



**Figure 1. State-space model for the regulatory gene network responding to silica stimulation in cultured human fibroblasts.** The numbers next to the edges are the coefficients in the state-space equations for the normal (black color) and SSc (red color) fibroblasts, respectively. The numbers in the boxes denote the mean expression values of the genes in normal (black color) and SSc (red color) fibroblasts.
doi:10.1371/journal.pone.0001693.g001

## Stability

The most important dynamic property of gene regulatory networks is concerned with stability. Stability is an organizing principle of gene regulatory networks [35–37]. When gene regulatory networks are perturbed, the expressions of the genes in the network will be changed in response to perturbation of environments. There are two possibilities. One possibility is that although the expressions of the genes will be changed after perturbation of environments, they will finally return to their original values or stay at other equilibrium values forever. In this case, regulatory networks will maintain their steady states under perturbation of environments and hence will function normally. Another possibility is that the expressions of the genes after perturbation of the environments will diverge from their original states and never stay at any steady states, which will finally lead to damage and malfunction of the gene regulatory network. Formally, a dynamic system is called stable if their state variables return to, or towards their original states or equilibrium states following internal and external perturbations [38]. The stability of the system is a property of the system itself. One of the methods for assessing the stability of the linear dynamic systems is to analyze eigenvalues of the state transition matrix $A$ of the linear dynamic systems. For a discrete linear system, if the norm of all eigenvalues of the transition matrix $A$ is less than 1 then the system is stable.

The eigenvalues of the transition matrix $A$ of the state-space model for silica responding gene network for normal and SSc fibroblasts are given in Table 1. It indicates that all eigenvalues of the transition matrix $A$ for the normal fibroblasts are less than 1, but for SSc fibroblasts, three eigenvalues whose absolute values are larger than 1. Therefore, the examined network for normal fibroblasts after perturbation of silica stimulation are relatively stable, but for SSc fibroblasts are unstable. Unstable gene regulatory or signal transduction networks will lead to erratic changes and malfunction of the whole biological system, which may be the case in the SSc fibroblasts that are associated with dramatic and irregular changes of *COL1A2* and *COL3A1*.

## Transient-Response Analysis of Genetic Networks

The dynamic behavior of a system is encoded in the temporal evolution of its states. Cell functions are essentially temporary processes and largely determined by the dynamic properties of the biological systems in the cells. Transient and steady state responses are two steps of the response of a gene network to perturbation of external environments. The transient response of the gene network to perturbation of environments is defined as rapid changes of the expressions of the genes in the network over time which go from their initial states to final states after perturbation of external input [13]. Steady-state response studies the system behavior at infinite time. The transient response of the dynamic systems is also a property of the system itself. The transient response of the gene network to environmental changes characterizes the dynamical process of the gene regulation networks in response to perturbation of environments. It can be used to study damped vibration behavior of the gene network and reveal how fast the gene networks respond to perturbation of environments and how accurately the networks can finally achieve the designed steady-

**Table 1**. Eigenvalues of the transition matrix A of the state-space model for the genes in a regulatory network responding to silica in cultured human normal and SSc fibroblasts.

| Normal fibroblasts | SSc fibroblasts |
|---|---|
| 0.23997 | 1.10627 |
| 0.66237 | 1.46736 |
| −0.24207 | 1.00586 |
| 0.84603 | 0.63654 |
| 0.32847 | 0.78374 |

doi:10.1371/journal.pone.0001693.t001

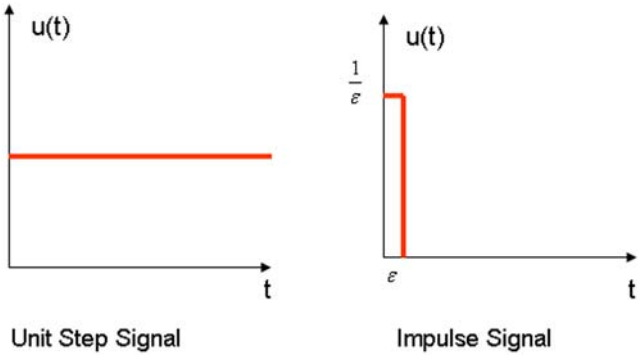



**Figure 2. Unit step signal and impulse signal.**
doi:10.1371/journal.pone.0001693.g002

state values. In the previous section we studied stability of the whole gene network, but we did not investigate the stability of the expression of the individual gene in the network. Since the transient response analysis of the gene network will study the dynamic process of the expression of the individual gene, it can be used to assess whether the expression of individual gene after perturbation of environment is stable or unstable. Although many transient response analyses is concerned with delay time, rise time, peak time, maximum overshoot and setting time, in this report, our transient response analysis mainly focuses on investigation of the stability, divergence or oscillation of individual gene expression.

Popular methods for investigation of the transient responses of the dynamic systems are to study the time domain characteristics of the system under perturbation of the external signals. The transient response of the dynamic system depends on the input signal. Different input signal will lead to the different transient response of the system. There are numerous types of signal in practice. For the convenience of analysis and comparison, we consider two types of signals: (i) unit-step signal and (ii) unit-impulse signal as shown in the Figure 2.

For discrete dynamic systems, the transient response of the system is obtained by using the inverse $z$ transform method [13]. To investigate the transient response of the genes in the network responding to silica, the silica was taken as input signal. Figures 3A and 3B show transient response of genes to a unit step perturbation of silica for normal and SSc fibroblasts, respectively. Figures 3C and 3D show transient response of genes to an impulse perturbation of silica for normal and SSc fibroblasts, respectively. Figures 3A, 3B, 3C and 3D demonstrate that the transient response of *SPARC*, *TIMP3*, *CTGF* to the perturbation of silica between the normal and SSc fibroblasts are similar, but the transient response of *COL1A2* and *COL3A1* to the perturbation of the silica between the normal and the SSc fibroblasts were dramatically different. The expressions of *COL1A2* and *COL3A1* after perturbation of the silica in normal fibroblasts quickly reach the steady states. However, the expressions of *COL1A2* and *COL3A1* in the SSc fibroblasts after perturbation of silica were unstable and will never reach the steady-state values. This phenomenon suggests that dynamic responses of the expressions

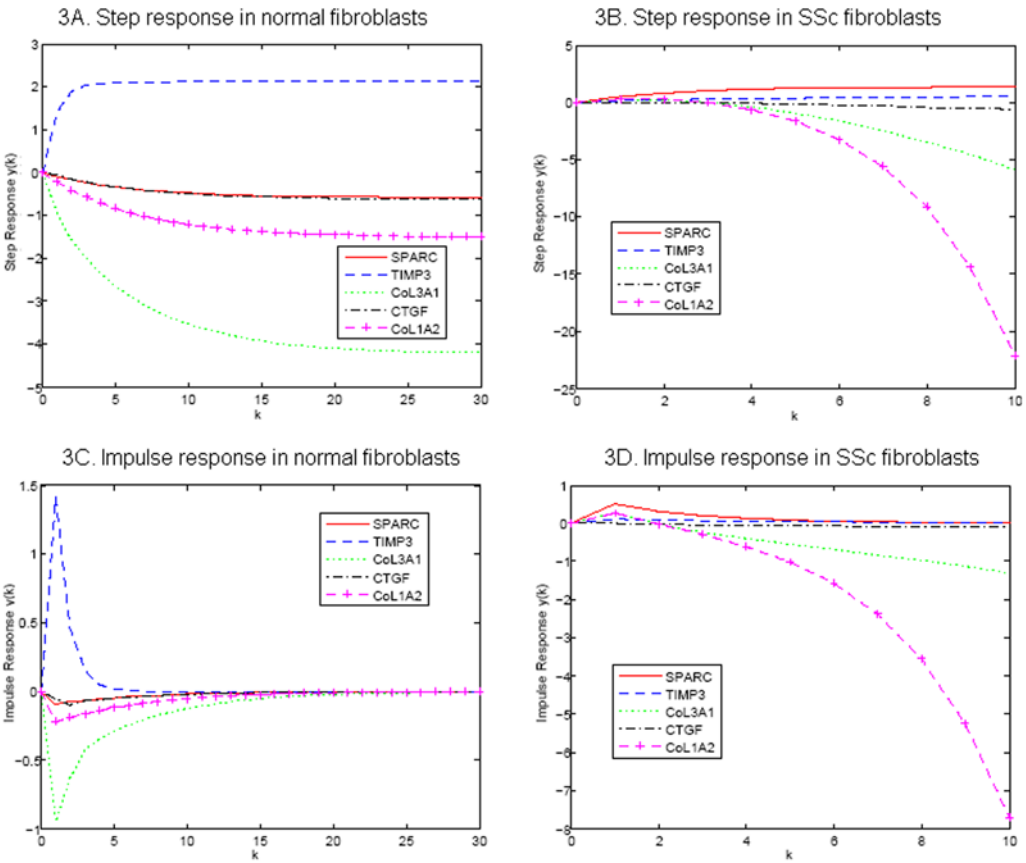



**Figure 3. Step response and impulse response of the genes to perturbation of silica in cultured fibroblasts.**
doi:10.1371/journal.pone.0001693.g003

of *COL1A2* and *COL3A1* in the SSc fibroblasts to environmental stimuli are irregular.

## Root-Locus Analysis

The performance of the gene networks under the design for stability, time response and reliability can be studied by analysis of their corresponding closed-loop system. The basic features of the stability and transient response of the closed-loop system are largely determined by the location of the closed-loop poles, which in turn is related to the value of the loop gain [13]. The roots of the characteristic equation of a system which is derived from the denominator of transfer function of the closed-loop system are the system's closed-loop poles. In general, the poles are complex variable and can be represented on the complex plane which is called s-plane. (Negative real) poles on the left hand side of the complex plane cause the response to decrease, while poles on the right hand side cause it to increase. Consequently, if the poles of the closed-loop system lie in the left half s plane, the system is stable. If any of these poles lies in the right-hand side of the s-plane, then the system is unstable. In this case, with increasing time, the transient response of the system will increase monotonically or oscillate with increasing magnitude [13]. As the loop gain changes the location of the closed-loop, poles will also changes. A root locus is defined as the locus of the poles of a transfer function of a closed-loop as a specific parameter (generally, loop gain) is varied from zero to infinity. The locus of the poles will be plotted on the complex plane (s-plane) as the system gain is varied on some interval. Since the location of the poles will change as the gain changes a system that is stable for gain $K_1$ may become unstable for a different gain $K_2$. We often observe that the root-locus will move from the left-hand of the s-plane to the right-hand of the s-plane, which implies that stable system becomes unstable as the system gain changes from one region to another region.

The root-locus plots the locations of the poles of the closed-loop single input and single output system (SISO) as the system gain varies. We use the symbol "x" to denote the poles of the closed-loop SISO and the symbol circle "o" to denote the zeros of the open-loop SISO. If the pole and zero coincide then the symbol : will be used to represent this situation. To study the dynamic behavior of the five genes to respond to the perturbation of silica, we consider the SISO system which takes one of the five genes as the output and silica as the input.

Figures 4A, 4B, 4C, 4D and 4E, and Figures 5A, 5B, 5C, 5D and 5E show the root-locus plot of *SPARC*, *TIMP3*, *COL3A1*, *CTGF*, and *COL1A2* with silica as the input in the normal and SSc fibroblasts, respectively. We noted that three remarkable features emerged from two panels of the Figures. First, all poles of the closed-loop SISO systems for five genes in the normal cell lines lie in the left hand side of the s-plane, but their corresponding poles in the SSc fibroblasts lie in the right hand side of the s-plane. This indicates that the expressions of all five genes to respond to the disturbance of the environmental silica in normal fibroblasts are stable, but become unstable in the SSc fibroblasts, which confirms the previous stability assessment. Second, although all poles of the closed-loop SISO for five genes in the normal fibroblasts are on the left hand side of the s-plane, the *SPARC*, *COL3A1*, *CTGF and COL1A2*, each has at least one branch of the root-locus plot which will enter the right-hand side of the s-plane. This indicates that the system becomes unstable as the increasing system gain reaches the some range. This may imply that the regulations of these four genes are sensitive to the changes of the system. Third, the poles and zeros of the SISO on the right hand sides of the s-plane for the

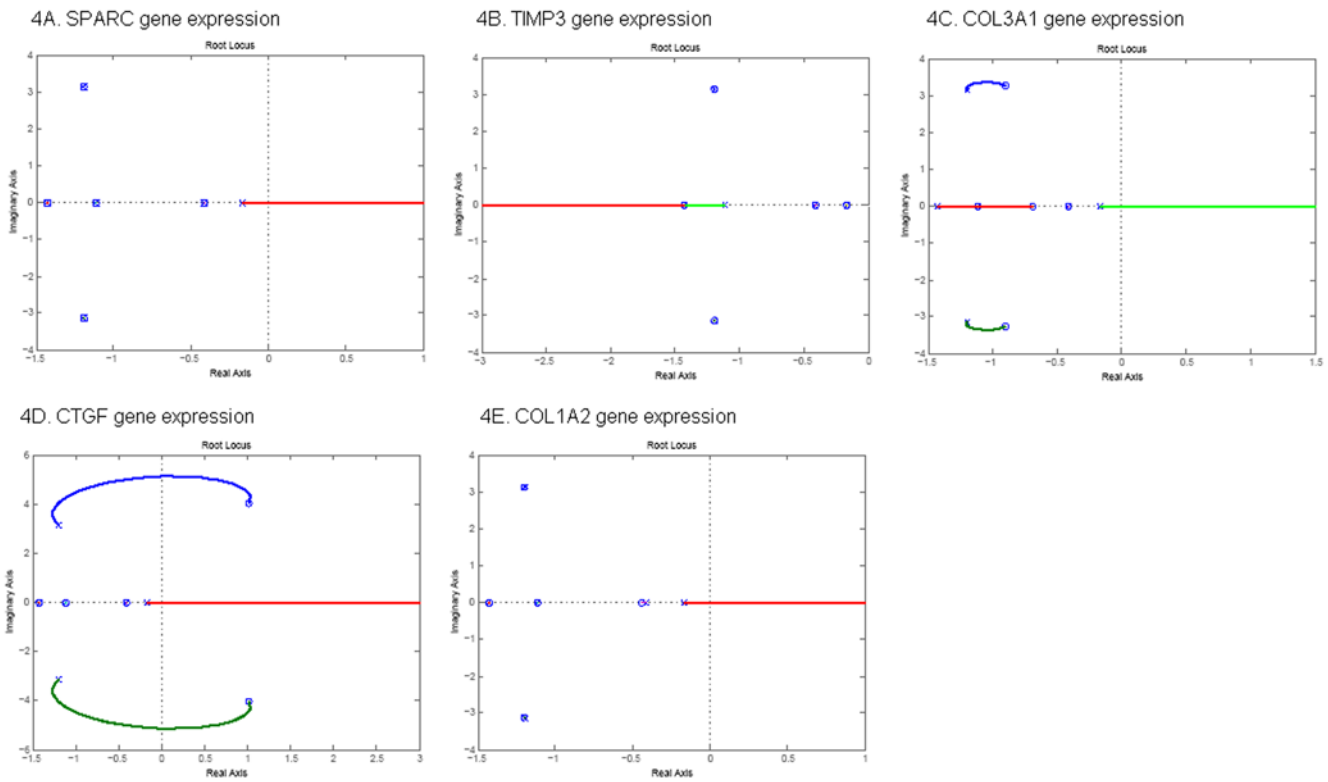



**Figure 4. Root-locus of gene expression in normal fibroblasts.**
doi:10.1371/journal.pone.0001693.g004

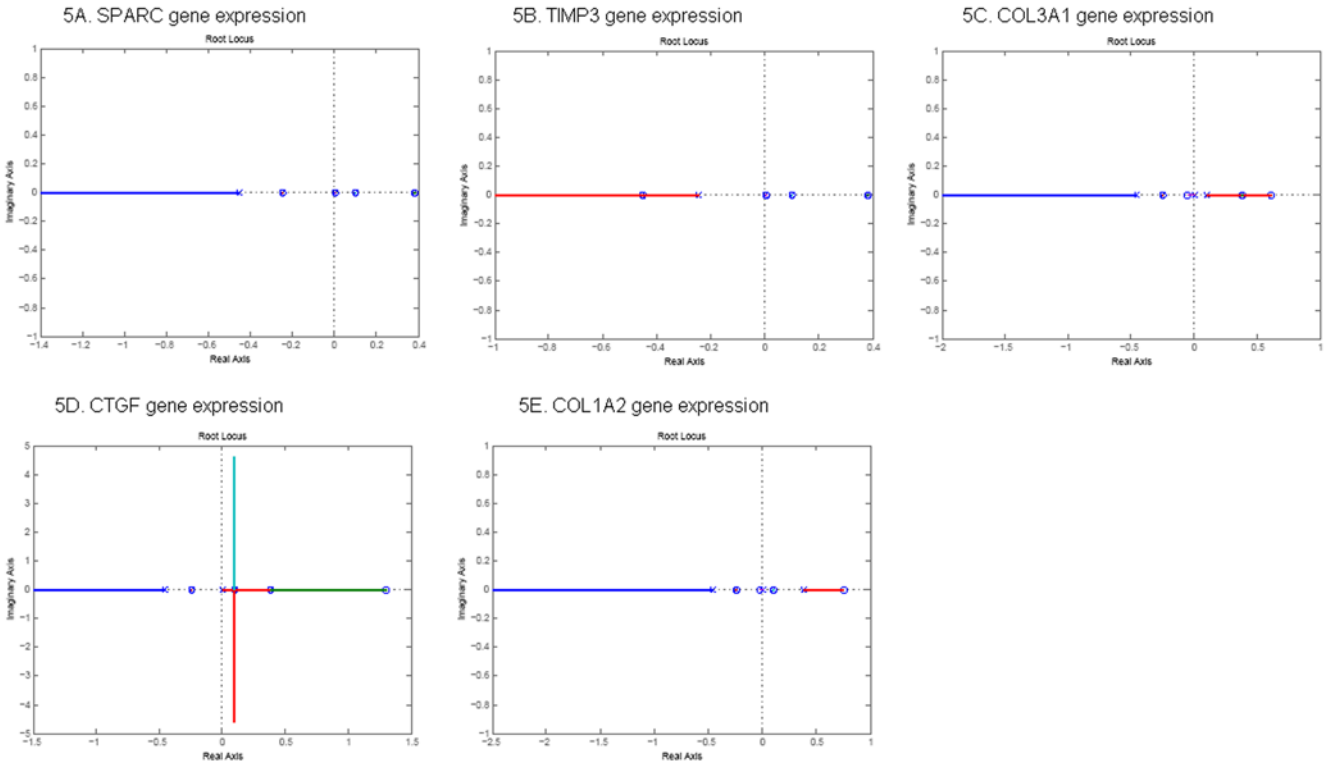



**Figure 5. Root-locus of gene expression in SSc fibroblasts.**
doi:10.1371/journal.pone.0001693.g005

*SPARC* and *TIMP3* in the SSc fibroblasts have the same location, i.e., the poles and zeros are cancelled out. This shows that the expressions of *CTGF*, *COL1A2* and *COL3A1* in response to the perturbation of the silica are more unstable than that of *SPARC* and *TIMP3*. Differential regulations of *CTGF*, *COL1A2* and *COL3A1* in response to the perturbation of silica between the normal and SSc fibroblasts may imply that the interactions of these three genes with the silica may be involved in the pathogenesis of the SSc. Forth, these Figures also demonstrate that when normal fibroblasts were changed to SSc fibroblasts, the root-locus will be moved toward the right-half s-plane. Classical control theory indicates that moving of the root-locus toward the right-half s-plane will reduce stability and increase response time of the system.

## Controllability

Changes in expression levels of genes and proteins in the regulatory networks will lead to status transition of the cells from normal cells to abnormal cells. One way to correct molecular changes is to transform cells from an undesirable state to a desirable one by altering gene or protein expressions. Now the question is whether we can use potentially therapeutic interventions to change gene or protein expressions from undesired states to desired states. This important issue can be addressed by examining the controllability of gene regulatory networks.

The fundamental controllability in gene regulation is associated with two questions. The first question is whether an input (therapy) can be found such that the system states can be driven from the undesired initial value to the desired values in a given time interval. The second question is how difficult it may be to change the system from undesired states to the desired states if the system is controllable.

The system (regulatory network) is called controllable, if for any state of the system, there exists a finite time and an admissible control function such that the system can achieve the desired state transition. In other words, the state controllability indicates that we can find an input to change the states from any initial value to any final value within some finite time. The controllability provides a binary information about whether the system is controllable or not, but it does not provide a quantitative measure to quantify the amount of control effort needed to accomplish the control task. It has been recognized that to get into insides of controllability of the system it is indispensible to define a quantity to measure how the system is controllable. In other words, we need to develop a measure to evaluate the amount of control efforts required to change the system from the initial state to the desired state [39]. The test for controllability is that the controllability matrix (see methods) has full rank, i.e., the rank of the controllability matrix is equal to the number of the state variables of the system. To assess how difficult to achieve control goal, we calculated the conditional number of the controllability matrix which measures the degree of difficulty to change the state of the system (or gene expression in our problem) by the external forces such as treatments. The larger the conditional number of controllability matrix, the more control efforts required to accomplish control task.

The rank of controllability matrix of the system in the state-space model of this partial TGF-$\beta$ pathway under perturbation of silica in both normal and SSc fibroblasts is equal to 5 which is the number of the state variables in the model. Thus, TGF-$\beta$ pathway is controllable in both normal and SSc fibroblasts. However, the conditional numbers of the controllability matrix of the system for normal and SSc fibroblasts were 80 and 398, respectively, which showed that the conditional number of the controllability matrix

for the SSc fibroblasts is much larger than that for normal fibroblasts. Therefore, much more control efforts are required to change gene expressions to desired levels for the SSc fibroblasts than that for the normal fibroblasts. This implies that the controllability of this network between the normal and SSc fibroblasts are differentiable.

## Discussion

In the past, large efforts have been devoted to studying the function of individual genes and static properties of biological pathways. However, the molecular concentrations and activities in living organisms are in constant change as a result of their interactions. The pathogenesis of disease involves evolution and temporal process. The functions of the cells, tissues and entire organisms are not only due to the steady states of the biological pathways, but also due to the dynamic interactions of biological pathways with the external environments. Although investigation of the function of individual genes, proteins and steady-states of the biological pathways is still valuable, it is time to devote more efforts and resources to study dynamic behaviors and properties of the biological pathways. It is dynamic properties that play a central role in giving rise to the function of cells and organisms [40].

To exemplify this principle, we studied the differential dynamic properties of a partial TGF-$\beta$ signaling network under perturbation of silica between normal and SSc fibroblasts. We took this network as a dynamic system and performed dynamic system analysis. Investigation of differential dynamics of this network between the normal and SSc fibroblasts consisted of three steps. The first step was to use the EM algorithm and Kalman filter to estimate the parameters in the state-space model of this TGF-$\beta$ signaling network. The second step was to study stability, the transient response and controllability of the system, and to perform root-locus analysis based on the identified state-space model of the gene network. The third step was to assess whether the dynamic properties of this network between the normal and SSc fibroblasts were different.

Our study in dynamic analysis of these gene regulations addressed several remarkable issues. First, the stability analysis may be used as a powerful tool for identifying biological pathways that are associated with the diseases. The stability is one of systems-level principles underlying complex biological pathways [41]. The stability of the system is the ability of the system to return to the equilibrium states after perturbation of the internal and external stimuli. The requirements for stable biological pathways are necessary conditions for the normal operations of the cells and organisms. The unstable biological pathway will inevitably lead to the malfunction of cells or even death of the living organism. Our results showed that a gene network in responding to perturbation of silica is relatively stable in the normal fibroblasts, but unstable in the SSc fibroblasts. This assessment of differential stability of biological pathway between normal and abnormal cells represents a novel approach in study associations of biological pathways with human diseases.

Second, root-locus analysis can provide valuable information for finding genes that show strong differential dynamics between normal and abnormal cells. Not all genes in the unstable pathway show unstable dynamics. Expressions of some genes in the unstable pathway may be stable themselves. Our task is to distinguish the genes that show stable expressions from those show unstable expressions in the unstable pathway. The state transition matrix of the state-space model of the studied gene network in the SSc fibroblasts has three poles that were in the right hand sides of the complex s-plane (Figures 3 and 4), which implies that this network in the SSc fibroblasts is unstable. The zeros of the genes of SPARC and TIMP3 in SISO system coincided with three poles. Therefore, although this gene network was unstable in the SSc fibroblasts, the expressions of the genes of SPARC and TIMP3 were stable. At least one branch of the root locus plots of other three genes (*CTGF*, *COL1A2* and *COL3A1*) were on the right hand sides of the s-plane. This indicates that the responses of the genes of CTGF, COL1A2 and COL3A1 to the perturbation of silica in the SSc fibroblasts were unstable no matter how the system gains were changed. These findings can be confirmed by the transient response analysis of the genes. The poles and zeros of characteristic equations of the SISO systems of the genes in response to the perturbation of internal and external signals are intrinsic properties of the gene regulations and are largely not affected by the expressions of other genes. Unlike the concept of differential expressions of the genes where the differentially expressed genes may be just consequences of differential expressions of other genes lying up in the pathway, the differential stability of the response of the genes to the perturbation of the signal between normal and abnormal tissues may be involved in the pathogenesis of the diseases. Therefore, the genes showing differential stability are supposed to be associated with the diseases. The root-locus analysis and the transient response will provide new tools for identifying the genes that are associated with the diseases. The differential stability and the transient response of the gene in the response to perturbation of the environment between the normal and abnormal cells characterize the interaction between the genes and environments. Therefore, the root-locus analysis and the transient response analysis also provide a powerful tool for detection of the gene-environment interaction.

Third, the controllability of biological pathway is an important property of the system. Germline or somatic mutations lead to the subsequent transcriptional and translational alterations which will finally cause diseases. Therapeutic interventions such as radiation, drug and gene therapy are intended to alter gene expressions from an undesired state to a desired or normal state. Gene regulation is a complex biological system. Theoretic and practical analyses in modern control theory demonstrate that there exist systems which we are not able (or find difficult) to change from undesired states to desired states of gene regulation. Now the question arises: are all biological pathways controllable? Are degrees of controllability of the biological pathways different between normal and abnormal Cells? The controllability measures the ability to move a system around in its entire state space using certain admissible intervention. In this report, we developed a conditional number of controllability matrix, to measure the degree of controllability of the system. Our results show that although a gene expression network responding to silica in both normal and SSc fibroblasts is controllable, the degree of controllability of this regulatory network between the normal and SSc fibroblasts is different. This regulatory network in the SSc fibroblasts has a low degree of controllability. In other words, adjustment of regulation of genes in the network by external intervention in the SSc fibroblasts is more difficult than that in the normal fibroblasts. We suspect that the degree of controllability is correlated with the severity of the diseases. When the diseases are at the initial stages, the biological systems are easy to move from abnormal states to the normal states. The degree of controllability of the system will provide valuable information on the curability of the diseases. Although in the past a number of authors have studied dynamic properties of biological networks, their studies have mainly used kinetic data or artificial data and nonlinear models [28,35–37]. Due to limitation of experiments, many kinetic parameters in the genetic regulation have not been available in practice. Large-scale kinetic analysis of

biological networks is infeasible. Here we use gene expressions and linear models to study dynamic properties of genetic networks. The results of this report showed that the dynamic properties of genetic network between normal and abnormal cells were differential.

In summary, dynamic properties of the biological systems are intrinsic system properties. The gene expressions are the phenotype of the cells. Their changes are governed by the hidden dynamic properties of the gene regulatory systems. It is dynamic properties that determine the phenotypes of the cells. This report represents a paradigm shift from the studies of individual components and static properties of the system to the studies of dynamic properties of the system consisting of individual components.

Although the preliminary results are appealing, they suffer from several limitations. First, sample sizes were small. Small sample size will limit the accuracy of the state-space models for biological pathways, which in turn affect estimation of dynamic properties of the systems. No much attention in control theory has been paid to investigation of impact of uncertainty inherent in dynamic systems on dynamic properties of the system. We will treat biological networks as stochastic dynamic system and study dynamic properties of stochastic dynamic systems in the future. Second, the quantities to characterize the dynamic properties are essentially random variables. Their distributions are unknown. We have not developed statistical methods to test significant differences in the dynamic properties of the pathways between the normal and abnormal cells. Third, the relations between the dynamic properties of the genes and their genotypes have not been investigated. Fourth, we have not performed large-scale dynamic analysis of the biological pathways. More theoretical development and large-scale real data analysis for the dynamic properties of the biological pathways are urgently needed.

## Methods

### Dermal fibroblast cultures

Skin biopsies of clinically uninvolved skin (3 mm punch) were obtained from 5 patients with SSc and 5 normal controls after informed consent was granted. All five patients fulfilled American College of Rheumatology criteria for SSc [42]. All five had diffuse skin involvement as defined by LeRoy et al [43], and disease duration of less than five years. Skin biopsies from five normal controls with no history of autoimmune diseases undergoing dermatologic surgery were matched for age (+/− 5 years) and sex. The study was approved by the Committee for the Protection of Human Subjects at University of Texas Health Science Center at Houston.

The skin sample was transported in Dulbecco's Modified Essential Media (DMEM) with 10% fetal calf serum (FCS) (supplemented with an antibiotic and antimycotic) for processing the same day. The tissue sample was washed in 70% ethanol, PBS, and DMEM with 10% FCS. Cultured fibroblast cell strains were established by mincing tissues and placing them into 60 mm culture dishes secured by glass cover slips. The primary cultures were maintained in DMEM with 10% FCS and supplemented with antibiotic and antimycotic.

### Silica stimulation on fibroblasts

The 5th passage of fibroblast strains were plated at a density of $2.5\times10^5$ cells in a 35 mm dish and grown until 80% confluence. Culture media then were replaced with FCS–free DMEM containing different doses (1, 5, 10, 25 and 50 μM) of silica particles obtained from Sigma-Aldrich, St Louis, MO. After 24-hour culture at this condition, the fibroblasts were harvested for extraction of RNA. The RNAs were examined with RT-PCR for gene expression of *COL1A2, COL3A1, TGFBRII, CTGF, SPARC* and *TIMP3*. The results from this dose-response assay provided an optimal dose (10 μM) in a time-dependent exposure for fibroblasts, in which 24-, 48-, 72-, 96- and 120-hour exposure of silica were assayed in cultured fibroblasts.

### Quantitative RT-PCR

Quantitative real time RT-PCR was performed using an ABI 7900 sequence detector (Applied Biosystems, Foster City, CA). The specific primers and probes for each gene were purchased through Assays-on-Demand from Applied Biosystems. As described previously (19), total RNA from each sample was extracted from cultured fibroblasts described above using TRIzol reagent (Invitrogen Life Technology) and treated with Dnase I (Ambion, Austin, TX). cDNA was synthesized using Superscript II reverse transcriptase (Invitrogen Life Technology). Synthesized cDNAs were mixed with primers/probes in the 2× Taqman universal PCR buffer, and then assayed on an ABI 7900. The data obtained from assays were analyzed with SDS 2.1 software (Applied Biosystems). The amount of total RNA in each sample was normalized with 18 S rRNA transcript levels.

### State-Space Model and Parameter Estimation

A biological pathway is taken as a dynamic system. The biological system is modeled by linear state-space equations defined as

$$\begin{aligned} x_{k+1} &= Ax_k + Bu_k + w_k \\ y_k &= Cx_k + Du_k + v_k \end{aligned} \tag{2}$$

where $x_k$ is a vector of state variables at the time $k$ that determine the dynamics of the regulation and unobserved, $u_k$ is a vector of input variables at the time $k$ such as drugs, environmental forces, and other state variables that lie outside the model, $y_k$ are observed variables at the time , for example, the gene expressions, $A$, $B$, $C$, and $D$ are matrices called state matrix, input matrix, output matrix and direct transmission matrix, respectively, $w$ and $v$ are noises assumed to be white Gaussian noise with zero means and variance $Q$ and $R$ respectively, and they are independent of each other.

A fundamental and widely applicable parameter estimation method is Maximum Likelihood (ML) method that maximizes the likelihood of the observed data with respect to parameters. However, the state-space models involve unobserved state variables that are unavailable. It makes calculation of the likelihood in the setting of state-space models very difficult. To solve this problem, we use expectation-maximization (EM) Algorithm that is widely used iterative parameter estimation method [15]. Specifically, we first assume that the state variables are available and then calculate the likelihood of both the observed data and hidden state variables which will be maximized with respect to the parameters in the models. After the estimated parameters are in our hands we then specify new state space models using the estimated parameters.

For the convenience of presentation, equation (2) can be rewritten as

$$\xi_k = \Gamma z_k + \eta_k \tag{3}$$

where $Z_k = \begin{bmatrix} x_k \\ u_k \end{bmatrix}$, $\xi_k = \begin{bmatrix} x_{k+1} \\ y_k \end{bmatrix}$, $\Gamma = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$, and $\eta_k \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} Q & 0 \\ 0 & R \end{bmatrix}\right)$. Then, the conditional density function of $\zeta_k$, given $\mathcal{Z}_k$ is given by $\mathcal{N}(\Gamma z_k, \Pi)$, where $\Pi = \left(\begin{bmatrix} Q & 0 \\ 0 & R \end{bmatrix}\right)$.

Assume that the distribution of the initial state is given by $x_1 \sim \mathcal{N}(\mu_1, P_1)$. Let a sequence of input-output data samples and the state be denoted by

$$U_N = \{u_1, \ldots, u_N\}, Y_N = \{y_1, \ldots y_N\}, X_{N+1} = \{x_1, \ldots, x_{N+1}\}.$$

The E-M algorithm for estimation of the parameters in the state-space model of discrete dynamic systems consists of two iterated steps: E-step and M-step. They are summarized as follows [16].

## E-Step

Calculate the expectation of the augmented log-likelihood function of both the observed data and hidden state variables defined as follows:

$$Q(\theta,\theta') = E_{\theta'}[\log P_\theta(X, Y|U)|Y, U].$$

To calculate $Q(\theta, \theta')$, we **first** need to calculate the conditional likelihood function $P_\theta(X, Y|U)$. From the model (3), we have

$$P_\theta(Y_N, X_{N+1}|U_N) = P_\theta(x_1)\prod_{k=1}^{N} P_\theta(x_{k+1}, y_k|x_k, u_k), \quad (4)$$

where

$$P_\theta(x_1) \sim N(\mu, P_1) \text{ and } P_\theta\left(\begin{bmatrix} x_{k+1} \\ y_k \end{bmatrix}\middle| x_k, u_k\right) \sim N(\Gamma z_k, \Pi) \quad (5)$$

Combining equations (4) and (5), we have

$$\begin{aligned} &-2\log P_\theta(Y_N, X_{N+1}|U_N) = \log|P_1| + (x_1-\mu)^T P_1^{-1}(x_1-\mu) + \\ &N\log|\Pi| + \sum_{k=1}^{N}(\xi_k - \Gamma z_k)^T \Pi^{-1}(\xi_k - \Gamma z_k) \end{aligned} \quad (6)$$

Let

$$\Phi = \frac{1}{N}\sum_{k=1}^{N} E_{\theta'}\{\xi_k \xi_k^T | Y_N, U_N\},\ \psi = \frac{1}{N}\sum_{k=1}^{N} E_{\theta'}\{\xi_k z_k^T | Y_N, U_N\},$$
$$\Sigma = \frac{1}{N}\sum_{k=1}^{N} E_{\theta'}\{z_k z_k^T | Y_N, U_N\}.$$

Taking expectation $E_\theta\{.|Y_N, U_N\}$ on both sides of equation (6), we obtain

$$\begin{aligned} &-2Q(\theta,\theta') = \log|P_1| + Tr\left\{P_1^{-1}E_{\theta'}\left\{(x_1-\mu)^T(x_1-\mu)\right\}\right\} + \\ &N\log|\Pi| + NTr\{\Pi^{-1}[\Phi - \Psi\Gamma^T - \Gamma\Psi^T + \Gamma\Sigma\Gamma^T\} \end{aligned} \quad (7)$$

To calculate the matrices $\Phi$ and $\Psi$, we use the following quantities

$$\begin{aligned} E_{\theta'}\{y_k x_k^T | Y_N, U_N\} &= y_k \hat{x}_{k|N}^T \\ E_{\theta'}\{x_k x_k^T | Y_N, U_N\} &= \hat{x}_{k|N}\hat{x}_{k|N}^T + P_{k|N} \\ E_{\theta'}\{x_k x_{k-1}^T | Y_N, U_N\} &= \hat{x}_{k|N}\hat{x}_{k-1|N}^T + M_{k|N}, \end{aligned} \quad (8)$$

and they can be calculated using Kalman smoother [44]:

$$\begin{aligned} J_k &= P_{k|k}A^T P_{k+1|k}^{-1} \\ \hat{x}_{k|N} &= \hat{x}_{k|k} + J_k\left[\hat{x}_{k+1|N} - A\hat{x}_{k|k} - Bu_k - R^{-1}y_k\right] \\ P_{k|N} &= P_{k|k} + J_k\left[P_{k+1|N} - P_{k+1|k}\right]J_k^N \\ M_{k|N} &= P_{k|k}J_{k-1}^T + J_k\left[M_{k+1|N} - AP_{k|k}\right]J_{k-1}^T \end{aligned} \quad (9)$$

The quantities $\hat{x}_{k|k}$, $P_{k|k}$, $P_{k|k-1}$ are calculated by the Kalman filter equations as follows:

$$\begin{aligned} P_{k|k-1} &= AP_{k-1|k-1}A^T + Q \\ G_k &= P_{k|k-1}C^T\left(CP_{k|k-1}C^T + R\right)^{-1} \\ P_{k|k} &= P_{k|k-1} - G_k C P_{k|k-1} \\ \hat{x}_{k|k-1} &= A\hat{x}_{k-1|k-1} + Bu_{k-1} \\ \hat{x}_{k|k} &= \hat{x}_{k|k-1} + G_k\left(y_k - C\hat{x}_{k|k-1} - Du_k\right), k = 1, \ldots, N \\ M_{N|N} &= (I - G_N C)AP_{N-1|N-1} \end{aligned} \quad (10)$$

## M-step

Maximizing the likelihood function defined in equation (7) with respect to parameters yields

$$\begin{aligned} \mu &= \hat{x}_{1|N} \\ P_1 &= P_{1|N} \\ \Gamma &= \begin{bmatrix} A & B \\ C & D \end{bmatrix} = \Psi\Sigma^{-1},\ \Pi = \Phi - \Psi\Sigma^{-1}\Psi^T. \end{aligned} \quad (11)$$

Since the network has structure, which enforces certain parameters in the model to be zeros and leaves others to free to change, we develop constrained EM algorithms.

First we define a matrix product operation of two matrices called Hadamard product, denoted by $\circ$, as element wise product, i.e.

$$[AB]_{ij} = [A]_{ij}[B]_{ij}.$$

Then, we define a modification of the vector $V$, denoted by $[V]_{\text{mod}}$, as the vector in which all elements corresponding to the zeros elements in the matrix $\Gamma$ are deleted. We define a modification of the matrix as the matrix in which if intersection of the row and column corresponds to the zeros elements in the matrix $\Gamma$ then such row and column are deleted. The equation for estimation of parameters (18) is reduced to

$$\begin{bmatrix} \Psi_{i1} \\ \Psi_{i2} \\ \vdots \end{bmatrix}_{\text{mod}} - \Sigma_{\text{mod}}[\Gamma_i]_{mod}^T = 0$$

$$\Pi = \left\{\Phi - \Psi\Gamma^T - \Gamma\Psi^T + \Gamma\Psi\Gamma^T\right\} \circ I$$

## Transient-Response Analysis

Response of a biological pathway to perturbation of internal and external stimuli has two parts: the transient and the steady state response. The time varying process generated in going from the initial state to the final state in the response to the perturbation of the

internal and external stimuli is called transient response. Steady-state response studies the system behavior at infinite time. Transient-response analysis of biological pathways can be used to quantify their dynamics. It can reveal how fast the biological pathways respond to perturbation of environments and how accurately the pathways can finally achieve the desired steady-state values. It can also be used to study damped vibration behavior and stability of the biological pathways.

The transient response of the dynamic systems depends on the input signals. Different signal will cause different response. There are numerous types of signal in practice. For the convenience of comparison, we consider two types of signals: (i) unit-step signal and (ii) unit-impulse signal as shown in Figure 2.

The transient response of dynamic systems can be studied by transfer function that is used to characterize the input-output relationships of a linear, time-invariant, differential equation system. The transfer function is defined as the ratio of the Laplace transform of the output to the Laplace transform of the input under the assumption of zero initial conditions. The transfer function of the response of the biological pathway to unit-step and unit-impulse input signals are given by $Y(s) = \frac{G(s)}{S}$ and $Y(s) = G(s)$ respectively, where $G(s)$ is the transfer function of the biological pathway. The transient-response analysis of the biological pathway can be performed by inverse Laplace transformation. We performed the transient-response analysis with MATLAB [13].

## Stability Analysis

The most important dynamic property of biological pathways is concerned with stability. Dynamic systems are called stable if their variables such as gene expressions return to, or towards, their original or equilibrium states following internal and external perturbations. For any practical purpose, the biological pathways must be stable. Unstable gene regulations will lead to the malfunction or even the death of the cells. A biological pathway will remain at steady state until occurrence of external perturbation. Depending on dynamic behavior of the system after perturbation of environments, the steady-states of the system are either stable (the system returns to the initial state or changes to other steady-states) or unstable (the system leaves the initial equilibrium state).

One of the methods for assessing the stability of the linear dynamic systems is to analyze eigenvalues of the state transition matrix $A$ of the linear dynamic systems. For a continuous linear dynamic system, if real parts of all eigenvalues of the transition matrix $A$ are strictly negative then the system is stable. For a discrete linear system, if the norm of all eigenvalues of the state transition matrix $A$ is less than 1 then the system is stable.

## Root-Locus Analysis

Open and close loop poles and zeros largely determine the stability and performance of the open and close systems. They provide valuable information on how to improve stability and transient response of the systems. Root-locus analysis, in which the roots of the characteristic equation of the closed-loop system are plotted for all values of a system parameter, is a powerful tool for study and design of dynamic pathway. The loop gain is often chosen to be the parameter. Varying the gain value will change the location of the closed-loop poles.

Consider a SISO system that consists of a gene regulator and an input to the gene regulator shown in Figure 6. The transfer function of the closed-loop system is given by

$$\frac{C(s)}{R(s)} = \frac{G(s)}{1 + G(s)H(s)},$$

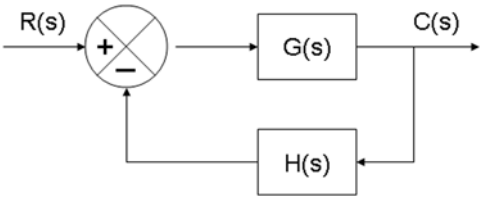



**Figure 6. Scheme of a SISO system**
doi:10.1371/journal.pone.0001693.g006

which implies the following characteristic equation of this closed-loop system:

$$1 + G(s)H(s) = 0. \tag{12}$$

In general, $G(s)$ $H(s)$ involves a gain parameter $K$. A plot of the points in the complex plane satisfying the characteristic equation (12) is the root locus. As the gain parameter changes the root locus will plot a curve in the complex s-plane. A simple method for plotting root-locus has been developed by W. R. Evans [29]. In this report, we use MATLAB to generate root-locus plots [13].

## Controllability

A dynamic system is called controllable, if there is an admissible control function, which can change the system from any given initial state to any finite state or to the origin of the state space in the finite time. Define the controllability matrix of the system as $H = [B, AB, ..., A^{n-1}B]$, where $A$ and $B$ are the state transition matrix and input to the state matrix in the linear dynamic system, respectively. If rank $(H) = n$, i.e., the rank of the controllability matrix is equal to the number of the state variables of the system, then the genetic network is completely controllable.

We use the condition number of the controllability matrix to measure the degree of the controllability of the system. The condition number of the controllability matrix is defined as [45]

$$\kappa(H) = \|H^{-}\|\|H\|.$$

where $H^{-}$ is a generalized inverse of the matrix $H$ and $||.||$ denotes a matrix norm. This can be justified by the following arguments. The general solution of the discrete linear system is given by [13]

$$x_k = A^k x(0) + \sum_{j=0}^{k-1} A^{k-j-1} B u_j, \tag{13}$$

By definition, if the system is controllable, then at some time $t_k$, we have $x_k = 0$, which implies that

$$0 = A^k x(0) + \sum_{j=0}^{k-1} A^{k-j-1} B u_j \tag{14}$$

Equation (14) can be reduced to

$$\begin{bmatrix} B & AB & A^2B \ldots A^{k-1}B \end{bmatrix} \begin{bmatrix} u_{k-1} \\ . \\ . \\ . \\ u_0 \end{bmatrix} = -A^k x(0)$$

Or

$$Hu = -A^k x(0) \tag{15}$$

where $u$ is a control vector. Solving the equation (15) yields

$$u = H^{-} A^k x(0) \tag{16}$$

The norm of the control vector represents the amount of control efforts required to change the states from initial value to the desired value and hence measures the degree of the controllability. From equation (16), we note that the norm of the control vector $||u||$ is proportional to the condition number: $\kappa(H) = ||H^{-}|| ||H||$. Therefore, we can use the condition number of the controllability matrix to measure the degree of controllability.

## Author Contributions

Conceived and designed the experiments: XZ MX. Performed the experiments: XG. Analyzed the data: MX HX. Contributed reagents/materials/analysis tools: XZ. Wrote the paper: XZ MX FA.

## References


1. Herrero J, Vaquerizas JM, Al-Shahrour F, Conde L, Mateos A, et al. (2004) New challenges in gene expression data analysis and the extended GEPAS. Nucleic Acids Res 32: W485–91.
2. Assmus HE, Herwig R, Kwang-Hyun Cho K-H, Wolkenhauer O (2006) Dynamics of biological systems: role of systems biology in medical research. Expert Review of Molecular Diagnostics 6: 891–902.
3. Barabási A-L, Oltvai ZN (2004) Network biology: understanding the cell's functional organization. Nature Reviews Genetics 5: 101–113.
4. Wolkenhauer O, Ullah M, Wellstead P, Cho KH (2005) The dynamic systems approach to control and regulation of intracellular networks. FEBS Lett 579: 1846–53.
5. Akutsu T, Miyano S, Kuhara S (1999) Identification of genetic networks from a small number of gene expression patterns under the Boolean network model. Pacific Symposium on Biocomputing. pp 17–28.
6. Kauffman S, Peterson C, Samuelsson B, Troein C (2003) Random Boolean network models and the yeast transcriptional network. Proc Natl Acad Sci U S A 100: 14796–14799.
7. Martin S, Zhang Z, Martino A, Faulon JL (2007) Boolean Dynamics of Genetic Regulatory Networks Inferred from Microarray Time Series Data. Bioinformatics 23: 866–874.
8. de Jong H (2002) Modeling and simulation of genetic regulatory systems: a literature review. J Comput Biol 9: 67–103.
9. Beal MJ, Falciani F, Ghahramani Z, Rangel C, Wild DL (2005) A Bayesian approach to reconstructing genetic regulatory networks with hidden factors. Bioinformatics 21: 349–356.
10. Dojer N, Gambin A, Mizera A, Wilczynski B, Tiuryn J (2006) Applying dynamic Bayesian networks to perturbed gene expression data. BMC Bioinformatics 7: 249.
11. Werhli A, Husmeier D (2007) Reconstructing gene regulatory networks with bayesian networks by combining expression data with multiple sources of prior knowledge. Stat Appl Genet Mol Biol 6: 15.
12. Nachman I, Regeve A, Friedman N (2004) Inferring quantitative models of regulatory networks from expression data. Bioinformatics 20: Suppl. 1248–1256.
13. Ogata K (1998) System dynamics. Third Edition. Upper Saddle Rive, New Jersey: Prentice Hall.
14. Rangel C, Wild DL, Falciani F, Ghahramani Z, Gaiba A (2001) Modeling biological responses using gene expression profiling and linear dynamical systems. Proceedings of the 2nd International Conference on Systems Biology. Omipress. pp 248–256.
15. Dempster A, Laird N, Rubin D (1977) maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society, Seres B 39: 1–38.
16. Gibson S, Ninness B (2005) Robust Maximum-Likelihood Estimation of Multivariable Dynamic Systems. Automatica 41: 1667–1682.
17. Voit EO, Schwacke JH (2007) Understanding through modeling – A historical perspective and review of biochemical system theory as a powerful tool for systems biology. Konpopka AK, ed. Systems biology: principles, methods and concepts CRC Press.
18. Prill RJ, Iglesias PA, Levchenko A (2005) Dynamic properties of network motifs contribute to biological network organization. PLoS Biol 3: e343.
19. Stelling J, Sauer U, Szallasi Z, Doyle FJ 3rd, Doyle J (2004) Robustness of cellular functions. Cell 118: 675–85.
20. Martins NC, Dahleh MA, Elia N (2006) Feedback stabilization of uncertain systems in the presence of a direct link. IEEE Transactions on Automatic Control 51: 438–447.
21. Kremling A, Saez-Rodriguez J (2007) Systems biology-An engineering perspective. J Biotechnol 129: 329–351.
22. Claman HN, Giorno RC, Seibold JR (1991) Endothelial and fibroblastic activation in scleroderma. The myth of the "uninvolved skin". Arthritis Rheum 34: 1495–1501.
23. Cabral AR, Alcocer-Varela J, Orozco-Topete R, Reyes E, Fernandez-Dominguez L, et al. (1994) Clinical, histopathological, immunological and fibroblast studies in 30 patients with subcutaneous injections of modelants including silicone and mineral oils. Rev Invest Clin 46: 257–66.
24. Gerriets JE, Reiser KM, Last JA (1996) Lung collagen cross-links in rats with experimentally induced pulmonary fibrosis. Biochim Biophys Acta 1316: 121–31.
25. Kim KA, Kim YH, Seok Seo M, Kyu Lee W, Won Kim S, et al. (2002) Mechanism of silica-induced ROS generation in Rat2 fibroblast cells. Toxicol Lett 135: 185–91.
26. Arcangeli G, Cupelli V, Giuliano G (2001) Effects of silica on human lung fibroblast in culture. Sci Total Environ 270: 135–9.
27. Pan Y, Durfee T, Bockhorst J, Craven M (2007) Connecting quantitative regulatory-network models to the genome. Bioinformatics 23: 1367–1376.
28. Steuer R, Gross T, Selbig J, Blasius B (2006) Structural kinetic modeling of metabolic networks. *Proc. Natl Acad. Sci.* USA 103: 11868–11873.
29. Zhou XD, Xiong MM, Tan FK, Guo XJ, Arnett FC (2006) SPARC, an upstream regulator of CTGF in response to TGF-$\beta$ stimulation. Arthritis & Rheum 54: 3885–3889.
30. Xiong M, Feghali-Bostwick CA, Arnett FC, Zhou X (2005) A systems biology approach to genetic studies of complex diseases. FEBS Lett 579: 5325–5332.
31. Derynck R, Zhang YE (2003) Smad-dependent and Smad-independent pathways in TGF-beta family signalling. Nature 425: 577–584.
32. Jinnin M, Ihn H, Mimura Y, Asano Y, Tamaki K (2006) Potential regulatory elements of the constitutive up-regulated $\alpha$2(I) collagen gene in scleroderma dermal fibroblasts. Biochem Biophys Res Commun. 343: 904–909.
33. Jimenez SA, Feldman G, Bashey RL, Bienkowski R, Rosenbloom J (1986) Co-ordinate increase in the expression of type I and type III collagen genes in progressive systemic sclerosis. Biochem J 237: 837–843.
34. LeRoy EC (1974) Increased collagen synthesis by scleroderma skin fibroblasts in vitro: a possible defect in the regulation or activation of the scleroderma fibroblast. J Clin Invest 54: 880–889.
35. Chen L, Aihara K (2002) Stability of genetic regulatory networks with time delay. IEEE Transactions on Circuits and Systems-I: Fundamental Theory and Applications 49: 602–608.
36. Prill RJ, Iglesias PA, Levchenko A (2005) Dynamic Properties of Network Motifs Contribute to Biological Network Organization. PLoS Biol 3(11): e343.
37. Binder B, Heinrich R (2002) Dynamic stability of signal transduction networks depending on downstream and upstream specificity of protein kinases. Mol Biol Rep 29: 51–55.
38. Kremling A, Saez-Rodriguez J (2007) Systems biology-An engineering perspective. J Biotechnol 129: 329–351.
39. Vinson DR, Georgaki C (2000) A new measure of process output controllability. Journal of Process Control 10: 185–194.
40. Wolkenhauer O, Mesarovic M (2005) Feedback dynamics and cell function: Why systems biology is called Systems Biology. Mol. Biosyst 1: 14–16.
41. Aldana M, Balleza E, Kauffman S, Resendiz O (2007) Robustness and evolvability in genetic regulatory networks. J Theor Biol 245: 433–448.
42. Subcommittee for Scleroderma Criteria of the American Rheumatism Association Diagnostic and Therapeutic Criteria Committee (1980) Preliminary criteria for the classification of systemic sclerosis (scleroderma). Arthritis Rheum 23: 581–590.
43. LeRoy EC, Black C, Fleischmajer R, Jablonska S, Krieg T, et al. (1988) Scleroderma (systemic sclerosis): classification, subsets and pathogenesis. J Rheumatol 15: 202–205.
44. Kailath T, Sayed AH, Hassibi B (1999) Linear Estimation. Upper Saddle River, New Jersey: Prentice Hall.
45. Wei Y, Wang D (2003) Condition numbers and perturbation of the weighted Moore-Penrose inverse and weighted linear squares problem. Applied mathematics and Computation 145: 45–58.

# ATTENUATION OF EXPRESSION OF EXTRACELLULAR MATRIX GENES WITH SIRNAS TO SPARC AND CTGF IN SKIN FIBROBLASTS OF CTGF TRANSGENIC MICE

J.C. WANG[1,2], S. SONNYLAL[3], F.C. ARNETT[2], B. DE CROMBRUGGHE[3] and X. ZHOU[2]

*[1]MOE Key Laboratory of Contemporary Anthropology and State Key Laboratory of Genetic Engineering, School of Life Sciences, Fudan University, Shanghai, China; [2]University of Texas Health Science Center at Houston, U.S.A.; [3]University of Texas M.D. Anderson Cancer Center, U.S.A.*

*Received February 15, 2011 – Accepted June 22, 2011*

The first two authors contributed equally to this manuscript

**Transgenic mice that over-express connective tissue growth factor (CTGF) in fibroblasts under the control of an enhancer/promoter element of the Col1a2 gene (*Col1a2-CTGF*) recapitulate multiorgan fibrosis similar to fibrosis observed in Scleroderma (SSc). In this study we investigate the regulation of secreted protein acidic and rich in cysteine (Sparc) and Ctgf siRNAs on the expression of several extracellular matrix components in the fibroblasts derived from *Col1a2-CTGF* transgenic mice. Three fibroblast lines were obtained from each of wide type C57BL/6 and CTGF transgenic C57BL/6, and were transfected with Sparc siRNA or Ctgf siRNA. Real-time quantitative RT-PCR and Western blotting were used to examine the transcription and protein levels of type I collagen, CTGF and SPARC. Student's *t*-tests were used to determine the significance of the results. Our results showed that Col1a2 and Ctgf increased expression at both transcriptional and translational levels in the fibroblasts from the *Col1a2-CTGF* transgenic mice compared with those in the fibroblasts from their normal wild-type littermate. The treatment with Sparc siRNA or Ctgf siRNA attenuated the mRNA and/or protein expression of the Col1a2, Ctgf and Sparc in these fibroblasts. Sparc and Ctgf siRNAs also showed a reciprocal inhibition at transcript levels. Therefore, our results indicated that both SPARC and CTGF appeared to be involved in the same biological pathway, and they have the potential to serve as a therapeutic target for fibrotic diseases such as SSc.**

Systemic sclerosis (SSc), also known as scleroderma, is a complex autoimmune disease characterized by skin and internal organ fibrosis. Currently, there is neither effective therapy nor effective prevention for this disease. Although the etiology of SSc is still unknown, both *in vitro* and *in vivo* studies have indicated that the extensive deposition of collagens, and other extracellular matrix (ECM) proteins by activated fibroblasts is a major pathologic property of SSc (1-2).

To better understand the pathogenic mechanisms and to find potential therapeutic targets of SSc, several kinds of animal models, including genetically modified mice harboring disruptions or manipulation of pivotal signaling pathways, have been established (2-3). Transforming growth factor

*Key words: Sparc, Ctgf, fibroblasts, fibrosis, siRNA*

*Mailing address:* Dr Xiaodong Zhou,
Division of Rheumatology,
University of Texas-Houston,
Health Science Center, 6431 Fannin, Houston,
TX 77030, USA
Tel: ++713 500 6088 Fax: ++713 500 0580
e-mail: xiaodong.zhou@uth.tmc.edu



0394-6320 (2011)

β (TGFβ) is a fibrotic growth factor. Over-activity of TGFβ signaling has been widely accepted to play important roles in the fibrosis of SSc (4). Connective tissue growth factor (CTGF) is a downstream mediator of TGFβ signaling. Many of the profibrotic properties of TGFβ are induced by the actions of CTGF (5). *Col1a2-CTGF* transgenic mice that over-express *CTGF* in fibroblasts under the collogen type 1 (*Col1a2*) promoter showed an SSc-like fibrotic phenotype (6). The animal models provide a platform for testing potential anti-fibrotic reagents for SSc.

SPARC (secreted protein, acidic and rich in cysteine) is a matricellular component of the ECM. It participates in the modulation of cell-matrix interactions, cell adhesion, wound repair, and angiogenesis (7-9) and possibly plays an important role in fibrosis. Increased expression of SPARC have been found in many fibrotic diseases including SSc, pulmonary fibrosis, renal interstitial fibrosis, hepatic cirrhosis, and atherosclerotic vascular lesions (10-14). SPARC has shown the ability to interact with the TGFβ signaling system through a TGFβ receptor and Smad2/3-dependent pathway (15).

In our previous studies, we observed that SPARC can regulate the expression of type 1 collagen, a major structural protein of the ECM, in normal human fibroblasts (16). Moreover, after exogenous TGFβ stimulation, SPARC siRNA showed a protective role against overexpression of collagen genes (16). Specific inhibition of SPARC expression with siRNA led to a down-regulation of collagen and CTGF gene expression in SSc fibroblasts (17). In order to evaluate the influence of the inhibition of Sparc in the *Col1a2-CTGF* transgenic mouse model and its potential as a therapeutic target of SSc, an *in vitro* study was performed using the fibroblasts derived from the novel *Col1a2-CTGF* transgenic mouse model to investigate the regulation of Sparc siRNA on the expression of several ECM components, and to compare it with that of Ctgf siRNA.

## MATERIALS AND METHODS

### *Cell lines*

Two Three fibroblast lines derived from skin biopsies of *Col1a2-CTGF* transgenic (heterozygous) mice and wild-type littermate controls (wide type C57BL/6) (6) were used in experiments described here. The cultures were maintained in DMEM with 10% FCS and supplemented with antibiotics (50 U/ml penicillin and 50 μg/ml streptomycin). Fifth-passage fibroblast cells were seeded at a density of $5 \times 10^5$ cells in 25-cm$^2$ flasks and grown until confluence.

### *Transient transfection with siRNA*

Double-stranded ON-TARGET*plus* siRNAs of murine *Sparc* and *Ctgf* were purchased from DHARMACON (Lafayette, CO). The corresponding target sequences are 5'- GCACCACACGUUUCUUUG -3' for *Sparc* and 5'- GCACCAGUGUGAAGACAUA -3' for *Ctgf*, respectively. The culture medium in each culture flask with confluent fibroblasts was replaced with Opti-MEM I medium (Invitrogen, Carlsbad, CA) without FCS and antibiotics. The fibroblasts were transfected with Sparc siRNA or Ctgf siRNA, using Metafectene (Biontex, Munich, Germany) in a concentration of 3 μg siRNA per ml medium. Fibroblasts with Non-Targeting siRNA treatment were used as negative control. After 8 hours, the culture medium was replaced with DMEM. The cells transfected with siRNA were examined after 72 hours of transfection and used for RNA expression and protein assays.

### *Determination of gene expression by quantitative RT-PCR*

Quantitative real-time RT-PCR was performed using an ABI 7900 Sequence Detector System (Applied Biosystems, Foster City, CA). The specific primers and probes for each gene (*Col1a2, Ctgf, and Sparc*) were purchased from the Assays-on-Demand product line (Applied Biosystems). Total RNA from each sample was extracted from the cultured fibroblasts using RNeasy Mini Kit (Qiagen, Valencia, CA). Complementary DNA (cDNA) was synthesized using MultiScribe™ Reverse Transcriptase (Applied Biosystems). Synthesized cDNAs were mixed with primers/probes in 2 × TaqMan universal PCR buffer and then assayed on an ABI 7900 sequence detector. The data obtained from the assays were analyzed with SDS 2.2 software (Applied Biosystems). The amount of total RNA in each sample was normalized with *Gapdh* transcript levels.

### *Western blot analysis*

The cellular lysates extracted from the above cultured fibroblasts were used for protein assays. The protein concentration was determined by a spectrophotometer using Bradford protein assay kit (Bio-Rad Laboratories, Hercules, CA). Equal amounts of protein from each sample were subjected to sodium dodecyl sulfate-polyacrylamide gel electrophoresis. Resolved proteins were transferred onto PVDF membrane and incubated

with respective primary antibodies, including anti-type I collagen antibody (CTGF International, Saco, ME), anti-Ctgf antibody (GeneTex Inc, San Antonio, TX), and anti-Sparc antibody (R&D Systems Inc, Minneapolis, MN). Mouse β-actin (Alexis Biochemicals, San Diego, CA) was used as an internal control. The secondary antibody was peroxidase-conjugated anti-rabbit, anti-goat, or anti-mouse IgG. Specific proteins were detected by chemiluminescence using an enhanced chemiluminescence system (Amersham, Piscataway, NJ). The intensity of the bands was quantified using ImageQuant software (Molecular Dynamics, Sunnyvale, CA).

## RESULTS

*Col1a2, Ctgf and Sparc expression in Col1a2-CTGF transgenic mice fibroblasts*

As measured by quantitative real-time RT-PCR, *Col1a2, Ctgf* and *Sparc* showed increased gene expression in the fibroblasts from *Col1a2-CTGF* transgenic mice compared with those from wild-type littermate controls (wild type) (Fig. 1). The fold changes of each gene in transgenic fibroblasts were 2.11±0.01 for *Col1a2*, 5.77±0.36 for *Ctgf*, and 1.66±0.18 for *Sparc*, respectively.

*Transfection efficiency*

Two methods were used for measuring transfection efficiency of siRNA. First, the fibroblasts were transfected with fluorescein-labeled non-silencing siRNA, and then examined under fluorescence microscopy which showed ~80% transfection efficiency by direct cell counting. Second, the fibroblasts from GFP transgenic C57BL/6 mouse (The Jackson Laboratory, Bar Harbor, Maine) were transfected with specific siRNA of GFP, and then examined to see how many cells responded with decreased levels of GFP. A similar efficiency of transfection was seen (Fig. 2).

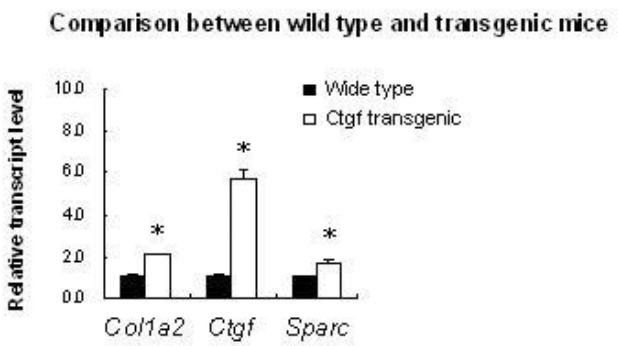



**Fig. 1.** *Comparison of gene expression among the wide type and Col1a2-CTGF transgenic mice fibroblasts. The expression level of each gene in wild type fibroblasts was normalized to 1. Bars show the mean±SD results of analysis of 3 independent experiments performed in triplicate. * P < 0.05.*

*Gene expression of Col1a2, Ctgf and Sparc after transfection of siRNAs in wild type and Col1a2-CTGF transgenic mice fibroblasts*

Seventy-two hours after transfection, the reduction of *Ctgf* (73% and 85% in wide type and transgenic type, respectively) by Ctgf siRNA and *Sparc* (69% and 82% in wide type and transgenic type, respectively) by Sparc siRNA were observed in all the two fibroblast lines from wide type or Ctgf transgenic mice (Table I and Fig. 3). Interestingly, the expression of Ctgf and Sparc showed a reciprocal down-regulation by Sparc siRNA and Ctgf siRNA (26% and 62% down-regulation of Ctgf by Sparc siRNA in wide type and transgenic, respectively; 29% and 53% down-regulation of Sparc by Ctgf siRNA in wide type and transgenic, respectively) (Table I and Fig. 3). In addition to the expression of Ctgf and Sparc, the *Col1a2 also* showed to be consistently down-regulated in all the fibroblasts after treated either by Ctgf siRNA or Sparc siRNA (Table I and Fig. 3).

*Protein expression of Col1a2, CTGF and SPARC in the fibroblasts with or without siRNA treatment*

The expression of the three ECM components at protein level was examined by Western blot analysis. Collagen type I and CTGF showed increased expression in the fibroblasts of *Col1a2-CTGF* transgenic mice compared with those in the cells from their normal littermate (wide type) (Fig. 4), which was consistent with their increased expression at the mRNA level. SPARC protein did not show significant reduction in the fibroblasts of Col1a2-CTGF transgenic mice, although its transcriptional level was lower.

Western blots also showed that the expression of Collagen type I, CTGF and SPARC were decreased after Ctgf siRNA or Sparc siRNA treatment in all fibroblast lines except for Sparc expression in the fibroblasts transfected with Ctgf siRNA (Fig. 4).

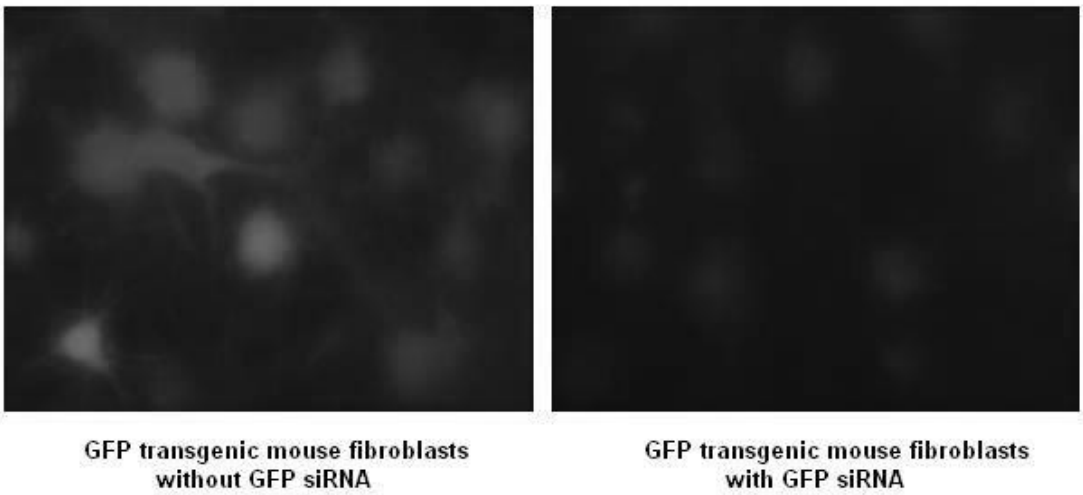



**Fig. 2.** *GFP expression in the fibroblasts from GFP transgenic mouse with or without GFP siRNA treatment. Left, without GFP siRNA treatment; right, with GFP siRNA treatment.*

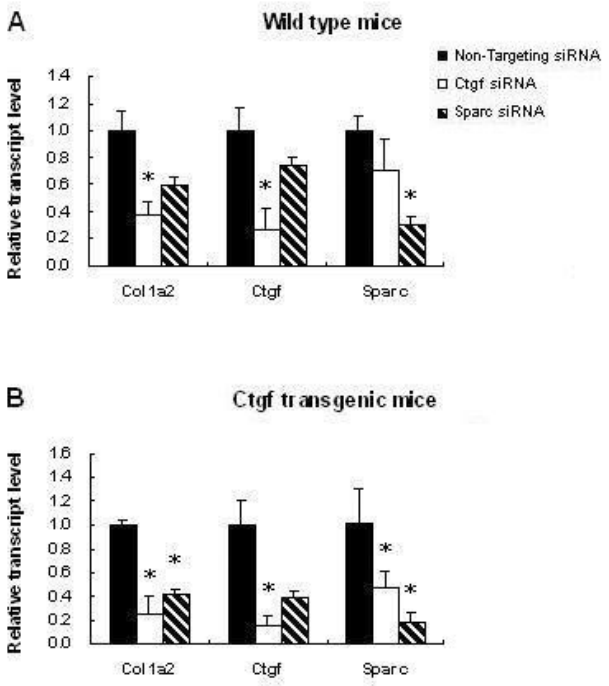



**Fig. 3.** *Gene expression in the fibroblasts with and without siRNA transfection. Comparison of gene expression with Ctgf siRNA, Sparc siRNA, or Non-Targeting siRNA treatment in wide type (**A**) and Col1a2-CTGF transgenic mice fibroblasts (**B**), respectively. The expression level of each gene in each fibroblast line with Non-Targeting siRNA treatment was normalized to 1. Assays were performed in triplicates. * $P < 0.05$.*

The down-regulation of CTGF protein by Ctgf siRNA was obviously less efficient than that of Sparc protein by Sparc siRNA (27% *vs* 86%, and 39% *vs* 92% in wild type and transgenic fibroblasts, respectively), although transcriptional levels of inhibition by their respective siRNAs were similar between Ctgf and Sparc. Sparc siRNA also down-regulated Ctgf, while Ctgf siRNA did not show a reciprocal inhibition to Sparc protein. In addition, the reduction of Col1a2 protein by Ctgf siRNA was less than that by Sparc (37% *vs* 47%, 21% *vs* 35% in wide type and transgenic fibroblasts, respectively) (Fig. 4.).

## DISCUSSION

Scleroderma is a devastating fibrotic disease which confers a high risk of mortality. No optimal therapy for reducing excessive collagen production in this disease is available (17). Animal model studies are crucial in finding therapeutic targets in SSc (18). Recently, *Col1a2-CTGF* transgenic mouse model was generated, that constitutively over-expressed CTGF specifically in fibroblasts (6). These mice displayed features similar to human scleroderma, including dermal fibrosis and lung fibrosis, and thus provided useful tools in the study of fibrogenesis and identification of possible therapeutic targets.

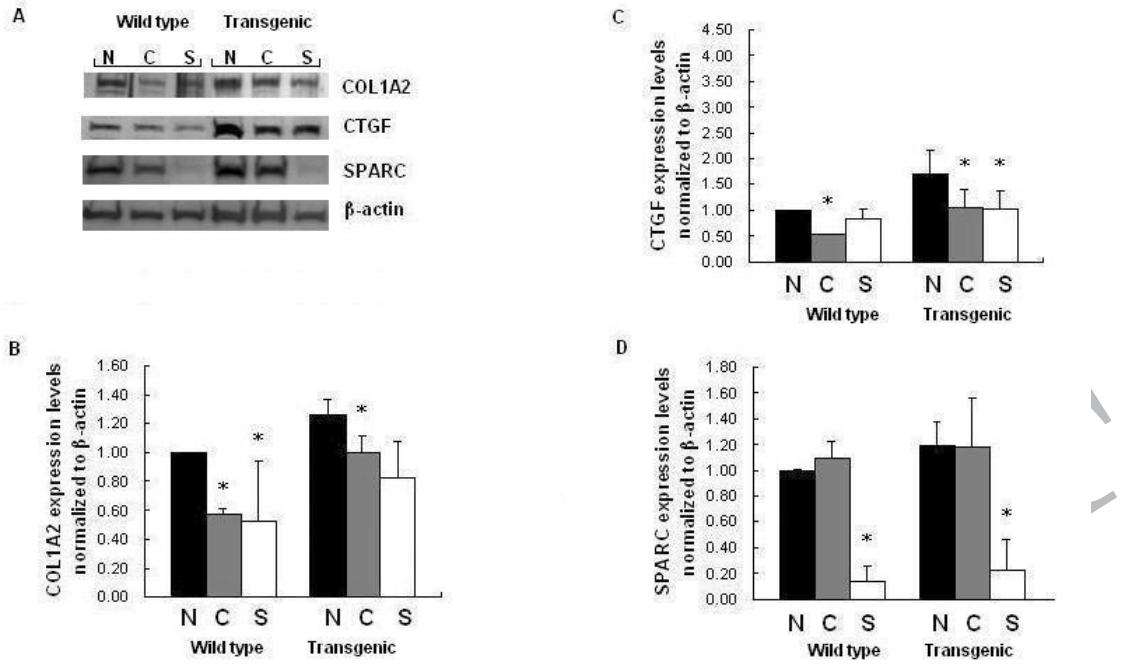



**Fig. 4.** *Western blot analysis of type I collagen, Ctgf, and Sparc in wild type and Col1a2-CTGF transgenic mice fibroblasts with Ctgf siRNA or Sparc siRNA transfection (**A**). N: Non-Targeting siRNA treatment; C: Ctgf siRNA treatment; S: Sparc siRNA treatment. Wild type stands for wild type fibroblasts. Densitometric analysis of Western blots for Col1, Ctgf and Sparc are summarized in **B**, **C**, and **D**. Protein expression levels were compared between transgenic mice fibroblasts and wild type fibroblasts with or without Ctgf or Sparc siRNA treatment. Assays were performed in triplicates. * P < 0.05.*

**Table I.** *Inhibition of gene expression by siRNA (assays were performed in triplicates).*

| Fibroblast line | Gene name | Relative transcription level (mean±SD) | | |
|---|---|---|---|---|
| | | Non-Targeting siRNA | Ctgf siRNA | Sparc siRNA |
| Wide type mice | *Col1a2* | 1 | 0.38±0.09 | 0.59±0.06 |
| | *Ctgf* | 1 | 0.27±0.16 | 0.74±0.07 |
| | *Sparc* | 1 | 0.71±0.22 | 0.31±0.06 |
| *Col1a2-CTGF* transgenic mice | *Col1a2* | 1 | 0.26±0.14 | 0.42±0.03 |
| | *Ctgf* | 1 | 0.15±0.08 | 0.38±0.06 |
| | *Sparc* | 1 | 0.47±0.14 | 0.18±0.08 |

Extensive deposition of collagens and other ECM proteins represent biomarkers for activated fibroblasts of SSc. SPARC and CTGF are two such ECM proteins important in regulating the production of collagen. Several mechanisms of such regulation have been explored, such as through SPARC and/or CTGF directly down-stream regulation or feedback control of TGFβ signaling transduction, and through direct binding to TGFβ receptor or TGFβ itself (15,19-23). In our previous study, it was shown that SPARC siRNA could attenuate the production of some ECM components, such as type 1 and 3 collagens, SPARC and CTGF in human normal and SSc fibroblasts (16-17, 23). We and others also showed that the blockade of CTGF expression either by CTGF siRNA, or its antisense oligonucleotide or

corresponding antibody, can block TGFβ-induced collagen production and/or fibronectin expression (23-26). We hereby further validated the anti-fibrotic effect of CTGF and SPARC inhibitors in profibrotic fibroblasts of *Col1a2-CTGF* transgenic mice that constantly over-express CTGF and collagens.

Our studies indicate that, similar to human fibroblasts, Sparc siRNA and Ctgf siRNA down-regulated the expression of collagen in mice fibroblasts of both wild type and *Col1a2-CTGF* transgenic types at both the transcriptional and translational levels (Table I, Figs. 3 and 4). Since parallel expression of Ctgf, Sparc and collagen in *Col1a2-CTGF* transgenic mice fibroblasts suggested that all three genes are involved in Ctgf-driven biological pathways, attenuation of over-expressed collagen type I in *Col1a2-CTGF* transgenic mouse fibroblasts by Sparc siRNA is likely mediated by Ctgf function.

Although mRNA expression levels of Sparc and Ctgf showed a reciprocal inhibition of these two genes by corresponding siRNA treatment in mouse fibroblasts, protein expression of these two appeared different. Sparc siRNA down-regulated CTGF protein in both wild type and transgenic mouse fibroblasts. In contrast, Ctgf siRNA did not show down-regulatory effect on the SPARC protein expression in the fibroblasts. Discordant expression levels of mRNA and protein of Sparc in both wild type and transgenic fibroblasts treated with Ctgf siRNA may reflect translational efficiency in the cells. On the other hand, concordant regulation of mRNA and protein of Ctgf by Sparc siRNA supports our previous finding in human fibroblasts, in which SPARC showed as an upstream regulator of CTGF in response to TGFβ stimulation (23). It is worth noting that an up-regulated gene expression level of Sparc was observed in the fibroblasts of *Col1a2-CTGF* transgenic mice. Along with the reciprocal inhibition of Sparc and Ctgf genes by corresponding siRNA treatment in mouse fibroblasts, these observations further suggested a mutual regulatory effect between SPARC and CTGF in the ECM compartment.

In conclusion, the fibroblasts from *Col1a2-CTGF* transgenic mice showed profibrotic features, which can be ameliorated by Inhibition of Sparc or Ctgf expression using their corresponding siRNAs. Sparc and Ctgf siRNAs showed a reciprocal inhibition in transcript levels, and Sparc siRNA also reduced the protein level of CTGF. The present *in vitro* study using fibroblasts from *Col1a2-CTGF* transgenic mouse model provides useful and potentially sufficient information for *in vivo* research to proceed.

## ACKNOWLEDGEMENTS

Supported by UTHSC-H grants from NIH/NIAMS Center of Research Translation (CORT) in Scleroderma (1 P50 AR054144) (F.C.A) and 5RO3AR050517-02 (X.Z.), NIH/NCRR Center for Clinical and Translational Sciences (Houston CTSA Program) (1UL1 RR 024148), NIH PHS NCRR GCRC grant M01RR002558, University of Texas Health Science Center at Houston CreFF (X.Z.), Department of the Army, Medical Research Acquisition Activity: PR064803 (X.Z) and the Scleroderma Foundation (X.Z.), the National Science Foundation of China, grant number 30971594 (J.W), the Major National Science and Technology Program of China, grant number 2008ZX10002-002 and the Major Project on Basic Research, Science and Technology Committee of Shanghai Municipality (10JC1402100 and 09XD1400200) (J.W).

## REFERENCES

1. Claman HN, Giorno RC, Seibold JR. Endothelial and fibroblastic activation in scleroderma: the myth of the "uninvolved skin." Arthritis Rheum 1991; 34: 1495-501.
2. Abraham DJ, Varga J. Scleroderma: from cell and molecular mechanisms to disease models. Trends Immunol 2005; 26:587-95.
3. Sonnylal S, Denton CP, Zheng B, Keene DR, He R, Adams HP. Postnatal induction of transforming growth factor β signaling in fibroblasts of mice recapitulates clinical, histologic, and biochemical features of scleroderma. Arthritis Rheum 2007; 56: 334-44.
4. Border WA, Noble NA. Transforming growth factor β in tissue fibrosis. N Engl J Med 1994; 331:1286-92.
5. Grotendorst GR. Connective tissue growth factor: a mediator of TGF-β action on fibroblasts. Cytokine

Growth Factor Rev 1997; 8:171-79.
6. Sonnylal S, Shi-Wen X, Leoni P, et al. Selective expression of connective tissue growth factor in fibroblasts *in vivo* promotes systemic tissue fibrosis. Arthritis Rheum 2010; 62:1523-532.
7. Brekken RA, Sage EH. SPARC, a matricellular protein: at the crossroads of cell-matrix. Matrix Biol 2001; 19:816-27.
8. Lane TF, Sage EH. The biology of SPARC, a protein that modulates cell-matrix interactions. FASEB J 1994; 8:163-73.
9. Raines EW, Lane TF, Iruela-Arispe ML, Ross R, Sage EH. The extracellular glycoprotein SPARC interacts with platelet-derived growth factor (PDGF)-AB and -BB and inhibits the binding of PDGF to its receptors. Proc Natl Acad Sci USA 1992; 89:1281-85.
10. Zhou X, Tan FK, Reveille JD, Wallis D, Milewicz DM, Ahn C. Association of novel polymorphisms with the expression of SPARC in normal fibroblasts and with susceptibility to scleroderma. Arthritis Rheum 2002; 46:2990-99.
11. Kuhn C, Mason RJ. Immunolocalization of SPARC, tenascin, and thrombospondin in pulmonary fibrosis. Am J Pathol 1995; 147:1759-69.
12. Pichler RH, Hugo C, Shankland SJ, Reed MJ, Bassuk JA, Andoh TF. SPARC is expressed in renal interstitial fibrosis and in renal vascular injury. Kidney Int 1996; 50:1978-89.
13. Frizell E, Liu SL, Abraham A, Ozaki I, Eghbali M, Sage EH. Expression of SPARC in normal and fibrotic livers. Hepatology 1995; 21:847-54.
14. Dhore CR, Cleutjens JP, Lutgens E, Cleutjens KB, Geusens PP, Kitslaar PJ. Differential expression of bone matrix regulatory proteins in human atherosclerotic plaques. Arterioscler Thromb Vasc Biol 2001; 21:1998-2003.
15. Schiemann BJ, Neil JR, Schiemann WP. SPARC inhibits epithelial cell proliferation in part through stimulation of the transforming growth factor-β-signaling system. Mol Biol Cell 2003; 14:3977-88.
16. Zhou X, Tan FK, Guo X, Wallis D, Milewicz DM, Xue S, Arnett FC. Small interfering RNA inhibition of SPARC attenuates the profibrotic effect of transforming growth factor β1 in cultured normal human fibroblasts. Arthritis Rheum 2005; 52:257-61.
17. Zhou X, Tan FK, Guo X, Arnett FC. Attenuation of collagen production with small interfering RNA of SPARC in cultured fibroblasts from the skin of patients with scleroderma. Arthritis Rheum 2006; 54:2626-31.
18. Eckes B, Hunzelmann N, Moinzadeh P, Krieg T. Scleroderma -- news to tell. Arch Dermatol Res 2007; 299:139-44.
19. Francki A, Bradshaw AD, Bassuk JA, Howe CC, Couser WG, Sage EH. SPARC regulates the expression of collagen type I and transforming growth factor-β1 in mesangial cells. J Biol Chem 1999; 274:32145-2152.
20. Francki A, McClure TD, Brekken RA, Motamed K, Murri C, Wang T, Sage EH. SPARC regulates TGF-beta1-dependent signaling in primary glomerular mesangial cells. J Cell Biochem 2004; 91:915-25.
21. Chujo S, Shirasaki F, Kawara S, Inagaki Y, Kinbara T, Inaoki M. Connective tissue growth factor causes persistent proalpha2(I) collagen gene expression induced by transforming growth factor-β in a mouse fibrosis model. J Cell Physiol 2005; 203:447-56.
22. Abreu JG, Ketpura NI, Reversade B, De Robertis EM. Connective-tissue growth factor (CTGF) modulates cell signalling by BMP and TGF-β. Nat Cell Biol 2002; 4:599-604.
23. Zhou XD, Xiong MM, Tan FK, Guo XJ, Arnett FC. SPARC, an upstream regulator of connective tissue growth factor in response to transforming growth factor β stimulation. Arthritis Rheum 2006; 54:3885-89.
24. Duncan MR, Frazier KS, Abramson S, Williams S, Klapper H, Huang X, Grotendorst GR. Connective tissue growth factor mediates transforming growth factor beta-induced collagen synthesis: down-regulation by cAMP FASEB J 1999; 13:1774-86.
25. Yokoi H, Mukoyama M, Sugawara A, et al. Role of connective tissue growth factor in fibronectin expression and tubulointerstitial fibrosis. Am J Physiol Renal Physiol 2002; 282:933-42.
26. Arnott JA, Nuglozeh E, Rico MC, Arango-Hisijara I, Odgren PR, Safadi FF, Popoff SN. Connective tissue growth factor (CTGF/CCN2) is a downstream mediator for TGF-beta1-induced extracellular matrix production in osteoblasts. J Cell Physiol 2007; 210: 843-52.

arthritis research&therapy

RESEARCH ARTICLE Open Access

# Attenuation of fibrosis *in vitro* and *in vivo* with *SPARC* siRNA

Jiu-Cun Wang[1,2], Syeling Lai[3], Xinjian Guo[2], Xuefeng Zhang[2], Benoit de Crombrugghe[4], Sonali Sonnylal[4], Frank C Arnett[2] and Xiaodong Zhou*[2]

## Abstract

**Introduction:** *SPARC* is a matricellular protein, which, along with other extracellular matrix components including collagens, is commonly over-expressed in fibrotic diseases. The purpose of this study was to examine whether inhibition of *SPARC* can regulate collagen expression *in vitro* and *in vivo*, and subsequently attenuate fibrotic stimulation by bleomycin in mouse skin and lungs.

**Methods:** In *in vitro* studies, skin fibroblasts obtained from a Tgfbr1 knock-in mouse (TBR1$^{CA}$; Cre-ER) were transfected with *SPARC* siRNA. Gene and protein expressions of the Col1a2 and the Ctgf were examined by real-time RT-PCR and Western blotting, respectively. In *in vivo* studies, C57BL/6 mice were induced for skin and lung fibrosis by bleomycin and followed by *SPARC* siRNA treatment through subcutaneous injection and intratracheal instillation, respectively. The pathological changes of skin and lungs were assessed by hematoxylin and eosin and Masson's trichrome stains. The expression changes of collagen in the tissues were assessed by real-time RT-PCR and non-crosslinked fibrillar collagen content assays.

**Results:** *SPARC* siRNA significantly reduced gene and protein expression of collagen type 1 in fibroblasts obtained from the TBR1$^{CA}$; Cre-ER mouse that was induced for constitutively active TGF-β receptor I. Skin and lung fibrosis induced by bleomycin was markedly reduced by treatment with *SPARC* siRNA. The anti-fibrotic effect of *SPARC* siRNA *in vivo* was accompanied by an inhibition of Ctgf expression in these same tissues.

**Conclusions:** Specific inhibition of *SPARC* effectively reduced fibrotic changes *in vitro* and *in vivo*. SPARC inhibition may represent a potential therapeutic approach to fibrotic diseases.

## Introduction

Fibrosis is a general pathological process in which excessive deposition of extracellular matrix (ECM) occurs in the tissues. It is currently untreatable. Although therapeutic uses of some anti-inflammatory and immunosuppressive agents such as colchicine, interferon-gamma, corticosteroids and cyclophosphamide have been reported, many of these approaches have not proven successful [1-3]. Recently, SPARC (secreted protein, acidic and rich in cysteine), a matricellular component of the ECM, has been reported as a bio-marker for fibrosis in multiple fibrotic diseases, such as interstitial pulmonary fibrosis, renal interstitial fibrosis, cirrhosis, atherosclerotic lesions and scleroderma or systemic sclerosis (SSc) [4-9]. Notably, increased expression of SPARC has been observed in affected skin and circulation of patients with SSc [10,11], a devastating disease of systemic fibrosis, as well as in cultured dermal fibroblasts obtained from SSc skin [8,9].

SPARC, also called osteonectin or BM-40, is an important mediator of cell-matrix interaction [12]. Increasing evidence indicates that SPARC may play an important role in tissue fibrosis. In addition to its higher expression level in the tissues of fibrotic diseases, SPARC has shown a capacity to stimulate the transforming growth factor beta (TGF-β) signaling system [13]. Inhibition of SPARC attenuates the profibrotic effect of exogenous TGF-β in cultured human fibroblasts [14]. Moreover, in animal studies, SPARC-null mice display a diminished amount of pulmonary fibrosis compared with control mice after

* Correspondence: Xiaodong.zhou@uth.tmc.edu
[2] Division of Rheumatology and Clinical Immunogenetics, Department of Internal Medicine, The University of Texas Medical School at Houston, 6431 Fannin St, Houston, Texas 77030, USA
Full list of author information is available at the end of the article

exposure to bleomycin, a chemotherapeutic antibiotic with a profibrotic effect [15]. These observations suggest that SPARC is a potential bio-target for anti-fibrotic therapy.

Recently, application of double-stranded small interfering RNA (siRNA) to induce RNA silencing in cells has been widely accepted in many studies of gene functions and potential therapeutic targets [16]. The selective and robust effect of RNAi on gene expression makes it a valuable research tool, both in cell culture and in living organisms. Unlike a gene knockout method, siRNA-based technology can easily silence the expression of a specific gene and is more feasible in practice, such as in disease therapy. Therefore, tissue-specific administration of the siRNA of candidate genes is currently being developed as a potential therapy in a great number of diseases, such as pulmonary diseases, ocular diseases, and others [17-19]. Our previous studies demonstrated that the overproduction of collagens in the fibroblasts obtained from SSc skin can be attenuated through SPARC silencing with siRNA. It suggested that application of SPARC silencing represents a potential therapeutic approach to fibrosis in SSc and other fibrotic diseases [20]. However, it is still unknown whether SPARC siRNA can improve fibrotic manifestations *in vivo*. The main purpose of the studies herein was to explore the feasibility of inhibition of SPARC with siRNA to counter fibrotic processes in a fibrotic mouse model *in vivo*. As a preliminary experiment in the *in vivo* studies, the fibroblasts cultured from a transgenic fibrotic model were used to assess the possibility and potential mechanisms of SPARC siRNA in attenuating the collagen expression *in vitro*. At the same time, the effects of SPARC siRNA to encounter fibrosis were compared with that of siRNA of CTGF, a well-known fibrotic marker. The fibrotic models used herein were the very popular bleomycin-induced skin and pulmonary fibrosis in mice. Subcutaneous injection and intratracheal instillation of siRNAs were used for tissue-specific treatments of skin and pulmonary fibrosis, respectively.

## Materials and methods

### Fibroblast cell lines from Tgfbr1 knock-in mouse

Constitutively activated Tgfbr1 mice, which recapitulated clinical, histological, and biochemical features of human SSc, have been reported previously [21]. They are termed $TBR1^{CA}$; Cre-ER mice and harbor both the DNA for an inducible constitutively active TGFβ receptor I (TGFβRI) mutation targeted to the *ROSA* locus, and a Cre-ER transgene driven by a *Col1* fibroblast-specific promoter. Fibroblasts were derived from skin biopsy specimens of these mice. The cultures were maintained in DMEM with 10% FCS and supplemented with antibiotics (50 U/ml penicillin and 50 μg/ml streptomycin). Fifth-passage fibroblast cells were seeded at a density of $5 \times 10^5$ cells in 25-$cm^2$ flasks and grown until confluence. Experiments were performed in triplicates.

### Transient transfection with siRNA in fibroblasts

Double-stranded ON-TARGET*plus* siRNAs of murine *SPARC* and *Ctgf* were purchased from Dharmacon, Inc. (Lafayette, CO, USA). The corresponding target sequences are 5'-GCACCACACGUUUCUUUG-3' for *SPARC* and 5'-GCACCAGUGUGAAGACAUA-3' for *Ctgf*, respectively. The culture medium in each culture flask with confluent fibroblasts was replaced with Opti-MEM I medium (Invitrogen, Carlsbad, CA, USA) without FCS and antibiotics. The fibroblasts were incubated for 24 hours and transfected with *SPARC* siRNA or Ctgf siRNA in a concentration of 100 nmol/L, using Dharma-*FECT*™ 1 siRNA Transfection Reagent (Dharmacon). Fibroblasts with Non-Targeting siRNA (Dharmacon) treatment were used as negative controls. The non-targeting siRNA was characterized by genome-wide microarray analysis and found to have minimal off-target signatures to human cells. It targets firefly luciferase (U47296). After 24 hours, the culture medium was replaced with DMEM. The cells transfected with siRNA were examined after 72 hours of transfection and used for RNA and protein expression analysis. The experiments were performed in triplicates.

### Animal models of fibrosis

C57BL/6 mice of about 20 grams were purchased from Jackson Laboratory (Bar Harbor, ME, USA). Bleomycin from Teva Parenteral Medicines Inc. (Irvine, CA, USA) was dissolved in saline and used in the mice at a concentration of 3.5 units/kg. Pulmonary fibrosis was induced in these mice with one time intratracheal instillation of bleomycin. For dermal fibrosis, female C57BL/6 mice at six weeks (weighing about 20 g) were treated daily for four weeks with local subcutaneous injection of 100 μl bleomycin in the shaved lower back. Four mice were used in each group. The animal protocols were approved by the Center for Laboratory Animal Medicine and Care in the University of Texas Health Science Center at Houston, the Institutional Animal Use and Care Committee of M.D. Anderson Cancer Center, and Fudan University, China.

### Administration of siRNAs in vivo

For pulmonary fibrosis, 3 μg of siRNA for *in vivo* use (siSTABLE, Dharmacon), mixed with Dharma*FECT*™ 1 siRNA Transfection Reagent, was administrated intratracheally in 60 μl on Days 2, 5, 12 after bleomycin treatment. In addition, the siGLO Green transfection indicator (Dharmacon), a fluorescent RNA duplex was used for evaluating distribution of intratracheally injected siRNA. Twenty-four hours after injection, lung tissues

were obtained for processing slides using a cryo-microtomy. All the mice were sacrificed on Day 23 after anesthesia, and the lung samples were collected. The left lungs were fixed by 4% formalin and used for further histological analysis. The right lungs were minced to small pieces and divided into two parts, one for RNA extraction and one for collagen content analysis.

For dermal fibrosis, the above siRNAs were injected into the same area as that of bleomycin three hours after bleomycin treatment and continued for four weeks. The mice were sacrificed on Day 29 and the skin samples were collected. Saline was used as a negative control in both fibrosis studies.

#### Determination of gene expression by quantitative RT-PCR

Total RNA from each cell line was extracted from the cultured fibroblasts using RNeasy Mini Kit (Qiagen, Valencia, CA, USA). For mice lung and skin tissues, the minced samples were homogenized in lysis solution (Sigma-Aldrich, St. Louis, MO, USA) with a blender. Then total RNA was extracted using GenElute™ Mammalian Total RNA Miniprep Kit (Sigma-Aldrich). Complementary DNA (cDNA) was synthesized using MultiScribe™ Reverse Transcriptase (Applied Biosystems, Foster city, CA, USA). Quantitative real-time RT-PCR was performed using an ABI 7900 Sequence Detector System (Applied Biosystems). The specific primers and probes for each gene (*Col1a2, Col3A1, Ctgf, SPARC* and *Ccl2*) were purchased from the Assays-on-Demand product line (Applied Biosystems). Synthesized cDNAs were mixed with primers/probes in 2 × TaqMan universal PCR buffer and then assayed on an ABI 7900 sequence detector. The data obtained from the assays were analyzed with SDS 2.2 software (Applied Biosystems). The expression level of each gene in each sample was normalized with *Gapdh* transcript level.

#### Western blot analysis

The lysis buffer for Western blot analysis consisted of 1% Triton X-100, 0.5% Deoxycholate Acid, 0.1% SDS, 1 mM EDTA in PBS and proteinase inhibitor cocktail from Roche (Basel, Switzerland). The cellular lysates extracted from the cultured fibroblasts were used for protein assays. The protein concentration was determined by a spectrophotometer using Bradford protein assay kit (Bio-Rad Laboratories, Hercules, CA, USA). Equal amounts of protein from each sample were subjected to sodium dodecyl sulfate-polyacrylamide gel electrophoresis. Resolved proteins were transferred onto PVDF membranes and incubated with respective primary antibodies, including anti-type I collagen antibody (Biodesign International, Saco, ME, USA), anti-CTGF antibody (GeneTex Inc, San Antonio, TX, USA), and anti-SPARC antibody (R&D Systems Inc, Minneapolis, MN, USA). Mouse β-actin (Alexis Biochemicals, San Diego, CA, USA) was used as an internal control. The secondary antibody was peroxidase-conjugated anti-rabbit, anti-goat, or anti-mouse IgG. Specific proteins were detected by chemiluminescence using an enhanced chemiluminescence system (Amersham, Piscataway, NJ, USA). The intensity of the bands was quantified using ImageQuant software (Molecular Dynamics, Sunnyvale, CA, USA).

#### Determination of collagen content

Non-crosslinked fibrillar collagen in lung samples and skin samples was measured using the Sircol colorimetric assay (Biocolor, Belfast, UK). Minced tissues were homogenized in 0.5 M acetic acid with about 1:10 ratio of pepsin (Sigma-Aldrich). Tissues were weighted, and then incubated overnight at 4°C with vigorous stirring. Digested samples were centrifuged and the supernatant was used for the analysis with the Sircol dye reagent. The protein concentration was determined using Bradford protein assay kits and the collagen content of each sample was normalized to total protein.

#### Histological analysis

The tissue samples of both lung and skin were fixed in 4% formalin and embedded in paraffin. Sections of 5 μm were stained either with hematoxylin and eosin (HE) and Masson's trichrome.

#### Statistical analysis

Results were expressed as mean ± SD). The difference between different conditions or treatments was assessed by Student's t-test. A *P*-value of less than 0.05 was considered statistically significant.

## Results

#### Gene and protein expression of Col1a2, Ctgf and SPARC in the fibroblasts from TBR1$^{CA}$; Cre-ER mice with and without transfection of siRNAs of SPARC or Ctgf

As measured by quantitative real-time RT-PCR, the transcripts of *Col1a2, Ctgf* and *SPARC* showed increased expression in the fibroblasts from TBR1$^{CA}$; Cre-ER mice injected with 4-OHT, in which Tgfbr1 was constitutively active, compared with those in the cells from TBR1$^{CA}$; Cre-ER mice injected with oil (Figure 1). The fold-changes of each gene in 4-OHT-injected TBR1$^{CA}$; Cre-ER mice fibroblasts were 3.06 ± 1.42 for *Col1a2* ($P$ = 0.050), 4.15 ± 1.18 for *Ctgf* ($P$ = 0.049), and 2.49 ± 0.63 for *SPARC* ($P$ = 0.017), respectively. To study whether inhibition of *SPARC* induced a reduction of collagen in the fibroblasts from constitutively active Tgfbr1 mice, we transfected *SPARC* siRNA into cultured fibroblasts obtained from TBR1$^{CA}$; Cre-ER mice injected with 4-OHT. Ctgf is a down-stream gene in the TGF-β pathway [22-25], and inhibition of Ctgf reduced expression of the fibrotic effect

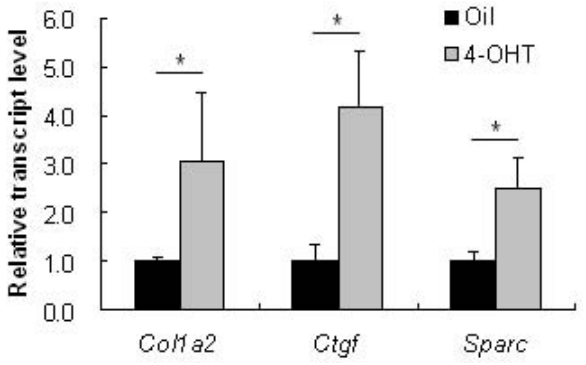



**Figure 1 Comparison of gene expression between the fibroblasts of TBR1$^{CA}$; Cre-ER mice injected with oil and 4-OHT**. The expression level of each gene in the fibroblasts of TBR1$^{CA}$; Cre-ER mice injected with oil was normalized to 1. Bars show the mean ± SD results of analysis of three independent experiments performed in triplicate. *, $P < 0.05$.

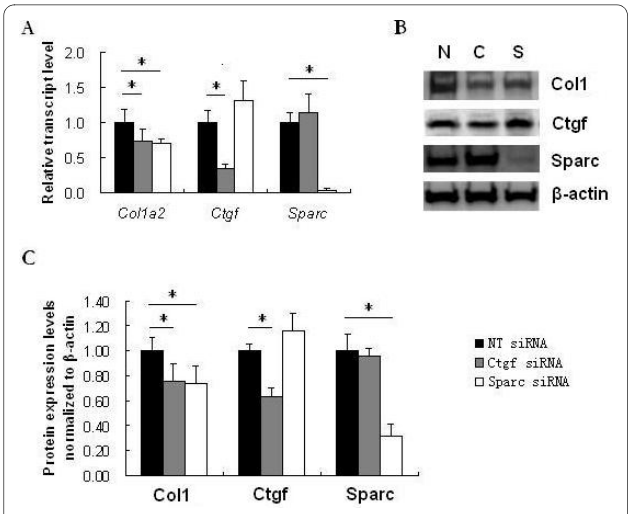



**Figure 2 Gene and protein expression in original and siRNA treated fibroblasts from TBR1CA; Cre-ER mice injected with 4-OHT**. **(A)** Relative transcript levels of Col1a2, Ctgf, and *SPARC* in cultured fibroblasts transfected with non-targeting siRNA (NT siRNA), Ctgf siRNA and *SPARC* siRNA. The expression level of each gene in the fibroblast lines with NT siRNA transfection was normalized to 1. *, $P < 0.05$. **(B)** Western blot analysis of type I collagen (COL1), CTGF, and SPARC in the fibroblasts from constitutively active Tgfbr1 mice transfected with NT siRNA, Ctgf siRNA or *SPARC* siRNA. N, non-targeting siRNA transfected fibroblasts; C, Ctgf siRNA transfected fibroblasts; S, *SPARC* siRNA transfected fibroblasts. **(C)** Densitometric analysis of Western blots for protein level of COL1, CTGF, and SPARC. Compared to non-targeting siRNA treatment, Ctgf siRNA or *SPARC* siRNA transfected fibroblasts showed significant reduction of COL1 ($P = .015$ or 0.009 respectively). Significant reduction of CTGF ($P = .002$) by Ctgf siRNA and SPARC ($P = 0.0004$) by *SPARC* siRNA were also shown. Bars show the mean ± SD results of analysis of three independent experiments performed in triplicate. *, $P < 0.05$.

of TGF-β [26]. We used Ctgf siRNA as a positive control for inhibition of Ctgf and collagen expression. Transfection efficiency of siRNAs into fibroblasts was measured using fluorescent RNA duplex si*GLO* Green transfection indicator (Dharmacon) and was determined to be over 80%. The gene expression levels from the Non-Targeting siRNA treated fibroblasts were compared with those from saline-treatment fibroblasts, and no significant differences were found (1.05 ± 0.18-folds for *Col1a2*, 1.14 ± 0.16-folds for *Ctgf*, and 1.12 ± 0.12-folds for *SPARC*). Therefore, in the following *in vitro* study, fibroblasts with Non-Targeting siRNA treatment were used as negative controls. Seventy-two hours after *SPARC* siRNA or Ctgf siRNA transfection, significant reductions of *SPARC* (95%) by *SPARC* siRNA and *Ctgf* (64%) by Ctgf siRNA were observed in the fibroblasts (Figure 2A). In parallel, *Col1a2* showed decreased expression in both siRNA transfected fibroblasts (27% and 29% decrease with $P < 0.05$ for Ctgf siRNA and *SPARC* siRNA, respectively) (Figure 2A). Western blot analysis showed a similar level of protein reduction of type I collagen by either *SPARC* siRNA or Ctgf siRNA treatment. As illustrated in Figure 2B, C, both *SPARC* siRNA and Ctgf siRNA showed significant attenuation of collagen type I in the fibroblasts ($P = 0.009$ or 0.015, respectively). CTGF and SPARC protein levels also were reduced by their corresponding siRNAs ($P = 0.002$ and 0.0004, respectively).

### siRNAs of SPARC and Ctgf ameliorated fibrosis in skin and reduced inflammation in lungs induced by bleomycin

HE stains of mouse skin tissues (Figure 3-1) showed that four-week injections of bleomycin induced significant fibrosis in skin where the fat cells were replaced by fiber bundles (Figure 3-1B, compared with normal skin injected with saline only (Figure 3-1A). Bleomycin-injected skin treated with *SPARC* siRNA or Ctgf siRNA showed that most of the fat cells still existed in the dermis without prominent fiber bundles (Figure 3-1C, D). Masson's trichrome staining of the samples also showed the same results. Notably, increased hair follicles were inconsistently seen in Ctgf siRNA- and *SPARC* siRNA-treated bleomycin-induced skins.

The lung distribution of intratracheally injected fluorescent siRNA showed that intense fluorescence was distributed within epithelial cells of bronchi and bronchioles, and only weak fluorescence was detected in the parenchyma (Figure 4-1).

HE stain of mouse lung tissues (Figure 4-2) showed a significant disruption of the alveolar units and infiltration of inflammatory cells in the lungs induced by bleomycin (Figure 4-2B), compared with saline injection (Figure 4-2A). However, after treatment with Ctgf siRNA or *SPARC* siRNA, the disruption of the alveoli was improved with less infiltrating inflammatory cells (Figure 4-2C, D). In addition, both siRNA treatments showed a significant reduction of gene expression of *Ccl2*, an active biomaker

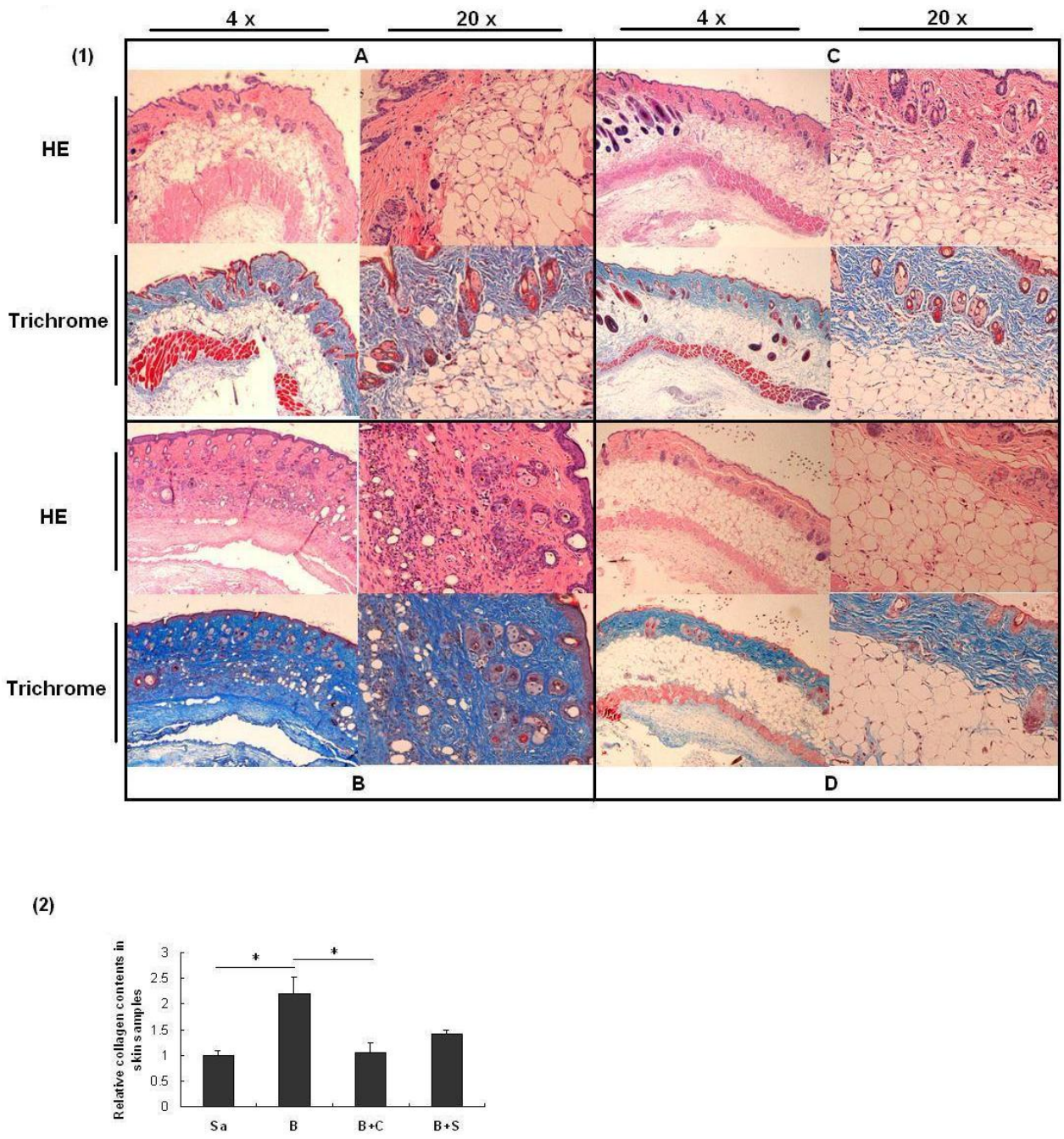



**Figure 3 Examination of skin tissues**. **(1)** Representative histological analysis of HE and Trichrome stain of mouse skin with different treatments for four weeks in low (4 ×) and high magnifications (20 ×). Four mice were used for each group. **A**. Injection with saline (negative control) only; **B**. Injection with bleomycin only; **C**. Injection with bleomycin and treatment with *SPARC* siRNA; **D**. Injection with bleomycin and treatment with Ctgf siRNA. **(2)** Collagen contents in skin samples with different treatments. The collagen content in the skin sample from saline treated mice was normalized to 1. Treatments: Sa, saline; B, bleomycin; B + C, bleomycin and Ctgf siRNAs; B + S, bleomycin and *SPARC* siRNA. $P < 0.05$.

of inflammation, which was up-regulated in bleomycin stimulated mice (Figure 5B).

### siRNAs of SPARC and Ctgf reduced the collagen contents in bleomycin-induced mouse skin and lung tissues

To further evaluate anti-fibrotic effects of siRNAs on the fibrogenesis of skin and lung, the collagen content was measured in the collected dermal and pulmonary samples. Quantification of total collagen in skin samples with the Sircol assay showed a 2.2-fold increase in bleomycin-induced skin compared with saline-injected skin ($P$ = 0.050). Ctgf siRNA treatment reduced the collagen content significantly to 47.6% ($P$ = 0.028) of that in bleomy-

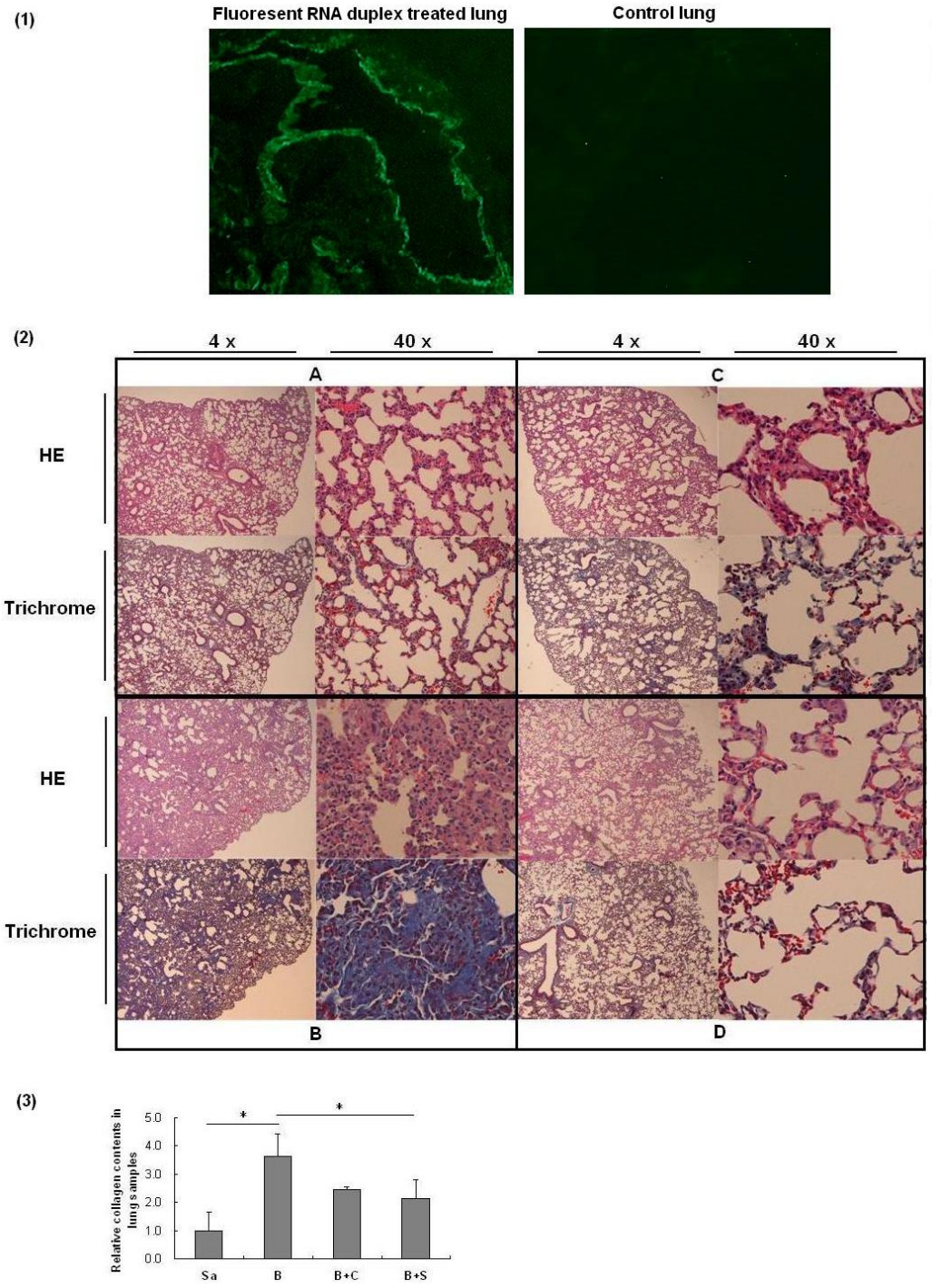



**Figure 4 Examination of lung tissues**. **(1)** The lung tissue staining for intratracheally injected fluorescent siRNA. Intense fluorescence was observed within epithelial cells of bronchi and bronchioles, and weak fluorescence was detected in the parenchyma. **(2)** Representative histological features of HE and Trichrome stain of mouse lung samples with different treatments intratracheally in low (4 ×) and high magnifications (40 ×). Four mice were used for each group. **A**. Injection with saline (negative control) only; **B**. Injection with bleomycin only on Day 0; **C**. Injection with bleomycin on Day 0 and *SPARC* siRNA on Days 2, 5, and 12; **D**. Injection with bleomycin on Day 0 and Ctgf siRNA on Days 2, 5, and 12. **(3)** Collagen contents in lung samples with different treatments. The collagen content in the lung sample from saline treated mice was normalized to 1. Four mice were used for each group. Treatments: Sa, saline; B, bleomycin; B + C, bleomycin and Ctgf siRNAs; B + S, bleomycin and *SPARC* siRNA. *, $P < 0.05$.

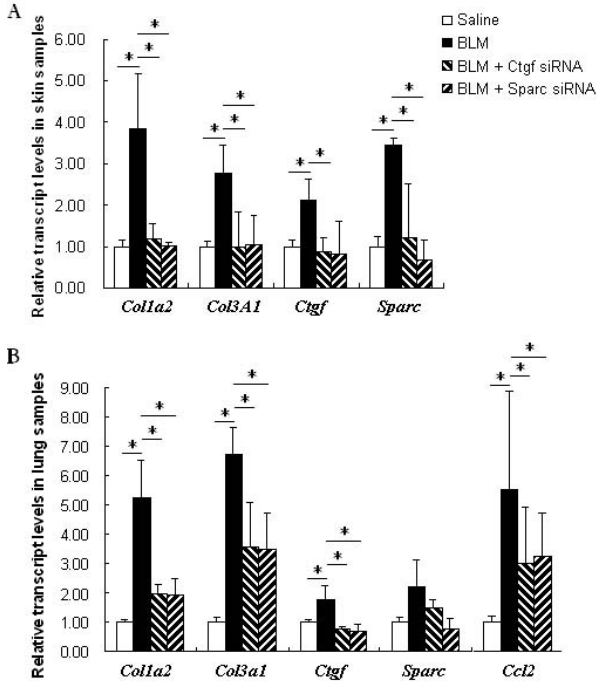



**Figure 5 Gene expression in skin (A) or lung samples (B) with different treatments**. Four mice were used for each treatment. The relative transcript levels of *Col1a2, Col3a1, Ctgf, SPARC* and *Ccl2* in siRNA-treated or untreated bleomycin-induced skins or lungs, respectively. The expression level of each gene in the skin or lung sample from saline treated mice was normalized to 1. Treatments: Saline; BLM (bleomycin); BLM + Ctgf siRNA and BLM + *SPARC* siRNA. *, $P < 0.05$.

cin-induced skin, and *SPARC* siRNA treatment reduced the collagen content to 64.6% ($P = 0.077$) but not very significantly (Figure 3-2). The difference of collagen reduction ($P = 0.076$) between *SPARC* siRNA treatment and Ctgf siRNA treatment was not very significant might due to the small sample size.

The siRNA treatments also showed a reduction of collagen in the lung tissues of bleomycin-induced mice (Figure 4-2). In bleomycin-induced mice, collagen content of lung tissues was 3.6-fold higher than that in saline-injected control mice ($P = 0.014$). In *SPARC* siRNA treated mice that also were bleomycin-induced, the collagen content of lung tissues was significantly reduced to 58% ($P = 0.019$) of that in bleomycin-induced mice without siRNA treatment. Ctgf siRNA also reduced the collagen content to a quite low level (68% of that without siRNA treatment) but without significance ($P = 0.128$). Further, no significant difference of collagen content was found between *SPARC* siRNA treatment and Ctgf siRNA treatment in bleomycin-injured lungs ($P = 0.277$).

### siRNAs of SPARC and Ctgf attenuated over-expression of collagen and other fibrotic ECM genes induced by bleomycin in skin and lung tissues

Bleomycin injection induced an up-regulation of the *Col1a2, Col3a1, Ctgf* and *SPARC* gene in both skin (Figure 5A, $P = 0.028$, 0.016, 0.049 and 0.0005, respectively) and lung tissues (Figure 5B, $P = 0.015$, 0.005, 0.041 and 0.056, respectively) of the mice significantly or marginal significantly. However, in Ctgf siRNA or *SPARC* siRNA treated mice skin that also received bleomycin injection, the expression of the *Col1a2* and *Col3a1* appeared to be normal in skin tissues (Figure 5A, $P = 0.025$ and 0.003 for each gene in Ctgf siRNA treatment, and $P = 0.031$ and 0.010 in *SPARC* siRNA treatment), and were significantly improved in lung tissues (about 2.7-fold reduction for *Col1a2* and 1.9-fold reduction for *Col3a1*, compared to bleomyin-injected mice without siRNA treatment, $P < 0.05$ for both) (Figure 5B). In addition to collagen gene expression, the *Ctgf* and the *SPARC* expression were significantly or marginal significantly reduced by *SPARC* siRNA and Ctgf siRNA treatment, respectively (Figure 5B). In detail, compared to bleomycin-induced skin and lungs, *SPARC* siRNA normalized *Ctgf* expression in both skin and lungs (2.6-fold reduction in both with $P = 0.100$ and 0.039, respectively). Similarly, Ctgf siRNA also reduced *SPARC* expression in skin and lungs (2.9-fold and 1.5-fold reduction with $P = 0.044$ and 0.102, respectively).

## Discussion

Although fibrosis is usually an irreversible pathological condition, targeting underlying molecular effectors may reverse an active status of the fibrotic process, and subsequently inhibit fibrosis. The TGF-β signaling pathway is associated with active fibrosis [22,23]. It begins with the binding of the TGF-β ligand to the TGF-β type II receptor, which catalyses the phosphorylation of the type I receptor on the cell membrane. The type I receptor then induces the phosphorylation of receptor-regulated SMADs (R-SMADs) that bind the coSMAD. The phosphorylated R-SMAD/coSMAD complex enters the nucleus acting as transcription factors to regulate target gene expression [22,23]. CTGF (connective tissue growth factor) is a down-stream gene that can be activated by the TGF-β signaling pathway [23,24]. Activation of CTGF is associated with potent and persistent fibrotic changes in the tissues, which is typically represented as accumulation of the ECM components including collagens [24,25]. SPARC also is involved in TGF-β signaling. It was reported that SPARC stimulated Smad2 phosphorylation and Smad2/3 nuclear translation in lung epithelial cells [27]. Recently, while examining SPARC regulatory role on the ECM components in human fibroblasts using linear structure equations, we demonstrated that SPARC positively controlled the expression of CTGF [26]. Although down-regulation of CTGF has been employed in treating fibrotic conditions [28], application of SPARC inhibition in attenuation of a fibrotic process in a therapeutic animal model has not been reported.

The studies described here first utilized the fibroblasts obtained from the $TBR1^{CA}$; Cre-ER mice that were induced for constitutively active TGF-β receptor I. After transfection of *SPARC* siRNA, the fibroblasts showed a decreased expression of Col1a2 that was originally over-expressed in the $TBR1^{CA}$; Cre-ER mice (Figure 2). This phenomenon suggests that *SPARC* inhibition may interrupt fibrotic TGF-β signaling, which generally induces collagen production. Although the specific mechanism for this suppression is unclear, multiple previous studies have demonstrated a mutual regulatory relationship between SPARC and TGF-β signaling [14,26,29]. This notion also is supported by the observation of an over-expression of *SPARC* in the fibroblasts of the $TBR1^{CA}$; Cre-ER mice (Figure 1). It should be noted that the Ctgf expression in the fibroblasts was not reduced upon *SPARC* inhibition. These results appear to contradict our previous report of parallel inhibition of SPARC and CTGF expression in human fibroblasts by SPARC siRNA [14]. A possible explanation is that over-expressed Ctgf from constitutively activated TGF-β signaling in these fibroblasts may confer resistance to a down-regulatory effect from *SPARC* siRNA. However, such resistance appeared to have limited influence on any down-regulatory effect of *SPARC* siRNA on collagen type 1, which suggests that CTGF is not a sole contributor to TGF-β signaling-associated fibrosis.

Bleomycin induced fibrosis in mice usually occurs after inflammation in which TGF-β is up-regulated [30]. Our *in vivo* application of *SPARC* siRNA demonstrated that inhibition of *SPARC* significantly reduced fibrosis in skin and lungs induced by bleomycin. In the treatment of skin fibrosis, *SPARC* siRNAs reduced fiber bundles accumulated in the dermis with less mononuclear cell infiltrates (Figure 3-1). In addition to histological changes, the thickness of bleomycin-induced skin treated with *SPARC* siRNA showed over 50% reduction compared to that without *SPARC* siRNA treatment (data not shown). The changes of tissue fibrotic level further were confirmed with significantly decreased collagen gene expression (Figure 5A). Non-crosslinked fibrillar collagen in the skin tissues also showed an average of 35.4% reduction after *SPARC* siRNA treatment (Figure 3-2).

In the treatment of lungs, *SPARC* siRNA reduced the disruption and inflammatory cells of the alveoli induced by bleomycin (Figure 4-2), which was accompanied with attenuated gene expression and protein content of collagens as compared to that without siRNA treatment (Figures 5B and 4-3). In addition, a significant reduction of the Ccl2 expression in the siRNA-treated lung tissues also suggests an improvement of inflammation supporting the findings in histological staining. These observations are consistent with previous reports on SPARC-null mice that exhibited attenuation of inflammation and fibrosis in kidneys [31]. While precise mechanism of these changes is still unknown, increased expression of SPARC was reported to correlate with the levels of inflammatory markers [32,33]. It is likely that *SPARC* inhibition altered composition of microenvironment of the tissues that may restrain inflammatory response. On the other hand, much higher levels of gene expression of *Col1A2* and *Col3A1*, and protein content of collagen were observed in bleomycin-induced lung tissues when they were compared to that in skin tissues (5.2-fold, 6.7-fold and 3.6-fold increase vs. 3.8-fold, 2.8-fold and 2.2-fold increase, respectively), which suggested that tissue damage and fibrosis in lung might be more severe than that in skin. In this case, treatment of bleomycin-induced lung damage might present a bigger challenge than that of skin, and the siRNA treatment through intratracheal instillation may be in need of further optimization. These notions were supported by similar findings in the treatment with the Ctgf siRNA, a positive control for anti-fibrotic effects.

Nevertheless, *SPARC* inhibition showed a clear anti-fibrotic effect in bleomycin-induced skin and lung tissues. Notably, these changes were accompanied with a significant down regulation of *Ctgf* that paralleled with *Ctgf* up-regulation in bleomycin-induced tissues. Thus, *SPARC* might regulate the collagen expression through affecting the expression of Ctgf, a TGF-β activity biomarker and down-strain gene, in bleomycin-induced mice. These observations combined with the results of anti-fibrotic effects of *SPARC* siRNA in fibroblasts of the Tgfbr1 knock-in mouse further support a mutually regulatory relationship between SPARC and TGF-β signaling.

## Conclusions

Studies described here consistently demonstrated that inhibition of *SPARC* with siRNA significantly reduced collagen expression in both *in vitro* transgenic Tgfbr1 fibroblast model and *in vivo* bleomycin-induced fibrotic mouse models. This is the first attempt to examine the anti-fibrotic effects of SPARC inhibition using siRNA with tissue-specific administration in skin and lungs *in vivo*. The results obtained from these studies provide favorable evidence that SPARC may be used as a bio-target for application of anti-fibrosis therapies.

#### Abbreviations

Ccl2: Chemokine (C-C motif) ligand 2, also known as monocyte chemotactic protein-1 (MCP-1); Col: collagen; Ctgf: connective growth factor; ECM: extracellular matrix; HE: hematoxylin and eosin; siRNA: small interfering RNA; *SPARC:* secreted protein, acidic and rich in cysteine; SSc: systemic sclerosis; TGF-β: transforming growth factor beta;

#### Competing interests

The authors are preparing a patent application for SPARC inhibition in the treatment of fibrosis. The authors declare that they have no other competing interests.

**Authors' contributions**

WJ carried out the animal studies and most of the molecular studies. LS carried out tissue histological examination. GX and ZX carried out molecular studies. CB and SS provided fibroblasts from $TBR1^{CA}$; Cre-ER mice. FA participated in coordination and helped to draft the manuscript. ZX carried out animal studies and participated in study design and drafting of the manuscript. All authors read and approved the final manuscript.

**Acknowledgements**

This study was supported by grants from the Department of the Army, Medical Research Acquisition Activity, grant number PR064803 to Zhou, the National Institutes of Health, grant number P50 AR054144 to Arnett and the National Science Foundation of China, grant number 30971574 to Wang.

**Author Details**

[1]State Key Laboratory of Genetic Engineering and MOE Key Laboratory of Contemporary Anthropology, School of Life Sciences, Fudan University, 220 Handan Road, Shanghai 200433, PR China, [2]Division of Rheumatology and Clinical Immunogenetics, Department of Internal Medicine, The University of Texas Medical School at Houston, 6431 Fannin St, Houston, Texas 77030, USA, [3]Department of Pathology, Baylor College of Medicine, One Baylor plaza, Houston, Texas 77030, USA and [4]Department of Molecular Genetics, MD. Anderson Cancer Center, University of Texas, 1515 Holcombe Blvd, Houston, Texas 77030, USA

Received: 28 October 2009 Revised: 12 February 2010
Accepted: 1 April 2010 Published: 1 April 2010

**References**

1. Abdelaziz MM, Samman YS, Wali SO, Hamad MM: **Treatment of idiopathic pulmonary fibrosis: is there anything new?** *Respirology* 2005, **10:**284-289.
2. Tome S, Lucey MR: **Review article: Current management of alcoholic liver disease.** *Aliment Pharmacol Ther* 2004, **19:**707-714.
3. Henness S, Wigley FM: **Current drug therapy for scleroderma and secondary Raynaud's phenomenon: evidence-based review.** *Curr Opin Rheumatol* 2007, **19:**611-618.
4. Kuhn C, Mason RJ: **Immunolocalization of SPARC, tenascin, and thrombospondin in pulmonary fibrosis.** *Am J Pathol* 1995, **147:**1759-1769.
5. Pichler RH, Hugo C, Shankland SJ, Reed MJ, Bassuk JA, Andoh TF, Lombardi DM, Schwartz SM, Bennett WM, Alpers CE, Sage EH, Johnson RJ, Couser WG: **SPARC is expressed in renal interstitial fibrosis and in renal vascular injury.** *Kidney Int* 1996, **50:**1978-1989.
6. Frizell E, Liu SL, Abraham A, Ozaki I, Eghbali M, Sage EH, Zern MA: **Expression of SPARC in normal and fibrotic livers.** *Hepatology* 1995, **21:**847-854.
7. Raines EW, Lane TF, Iruela-Arispe ML, Ross R, Sage EH: **The extracellular glycoprotein SPARC interacts with platelet-derived growth factor (PDGF)-AB and -BB and inhibits the binding of PDGF to its receptors.** *Proc Natl Acad Sci USA* 1992, **89:**1281-1285.
8. Zhou X, Tan FK, Reveille JD, Wallis D, Milewicz DM, Ahn C, Wang A, Arnett FC: **Association of novel polymorphisms with the expression of SPARC in normal fibroblasts and with susceptibility to scleroderma.** *Arthritis Rheum* 2002, **46:**2990-2999.
9. Vuorio T, Kähäri VM, Black C, Vuorio E: **Expression of osteonectin, decorin, and transforming growth factor-beta 1 genes in fibroblasts cultured from patients with systemic sclerosis and morphea.** *J Rheumatol* 1991, **18:**247-251.
10. Macko RF, Gelber AC, Young BA, Lowitt MH, White B, Wigley FM, Goldblum SE: **Increased circulating concentrations of the counteradhesive proteins SPARC and thrombospondin-1 in systemic sclerosis (scleroderma). Relationship to platelet and endothelial cell activation.** *J Rheumatol* 2002, **29:**2565-2570.
11. Davies CA, Jeziorska M, Freemont AJ, Herrick AL: **Expression of osteonectin and matrix Gla protein in scleroderma patients with and without calcinosis.** *Rheumatology (Oxford)* 2006, **45:**1349-1355.
12. Brekken RA, Sage EH: **SPARC, a matricellular protein: at the crossroads of cell-matrix.** *Matrix Biol* 2001, **19:**816-827.
13. Schiemann BJ, Neil JR, Schiemann WP: **SPARC inhibits epithelial cell proliferation in part through stimulation of the transforming growth factor-beta-signaling system.** *Mol Biol Cell* 2003, **14:**3977-3988.
14. Zhou X, Tan FK, Guo X, Wallis D, Milewicz DM, Xue S, Arnett FC: **Small interfering RNA inhibition of SPARC attenuates the profibrotic effect of transforming growth factor β1 in cultured normal human fibroblasts.** *Arthritis Rheum* 2005, **52:**257-261.
15. Strandjord TP, Madtes DK, Weiss DJ, Sage EH: **Collagen accumulation is decreased in SPARC-null mice with bleomycin-induced pulmonary fibrosis.** *Am J Physiol* 1999, **277:**L628-635.
16. Dykxhoorn DM, Novina CD, Sharp PA: **Killing the messenger: short RNAs that silence gene expression.** *Nat Rev Mol Cell Biol* 2003, **4:**457-467.
17. Duchaine TF, Slack FJ: **RNA interference and micro RNA-oriented therapy in cancer: rationales, promises, and challenges.** *Curr Oncol* 2009, **16:**61-66.
18. de Fougerolles A, Vornlocher HP, Maraganore J, Lieberman J: **Interfering with disease: a progress report on siRNA-based therapeutics.** *Nat Rev Drug Discov* 2007, **6:**443-453.
19. Whitehead KA, Langer R, Anderson DG: **Knocking down barriers: advances in siRNA delivery.** *Nat Rev Drug Discov* 2009, **8:**129-138.
20. Zhou XD, Tan FK, Guo X, Arnett FC: **Attenuation of collagen synthesis with small interfering RNA of SPARC in cultured fibroblasts from the skin of patients with scleroderma.** *Arthritis Rheum* 2006, **54:**2626-2631.
21. Sonnylal S, Denton CP, Zheng B, Keene DR, He R, Adams HP, Vanpelt CS, Geng YJ, Deng JM, Behringer RR, de Crombrugghe B: **Postnatal induction of transforming growth factor β signaling in fibroblasts of mice recapitulates clinical, histologic, and biochemical features of scleroderma.** *Arthritis Rheum* 2007, **56:**334-344.
22. Border WA, Noble NA: **Transforming growth factor beta in tissue fibrosis.** *N Engl J Med* 1994, **331:**1286-1292.
23. Blobe GC, Schiemann WP, Lodish HF: **Role of transforming growth factor beta in human disease.** *N Engl J Med* 2000, **342:**1350-1358.
24. Grotendorst GR: **Connective tissue growth factor: a mediator of TGF-β action on fibroblasts.** *Cytokine Growth Factor Rev* 1997, **8:**171-179.
25. Chujo S, Shirasaki F, Kawara S, Inagaki Y, Kinbara T, Inaoki M, Takigawa M, Takehara K: **Connective tissue growth factor causes persistent proalpha2(I) collagen gene expression induced by transforming growth factor-β in a mouse fibrosis model.** *J Cell Physiol* 2005, **203:**447-456.
26. Zhou XD, Xiong MM, Tan FK, Guo XJ, Arnett FC: **SPARC. An upstream regulator of connective tissue growth factor in response to transforming growth factor β stimulation.** *Arthritis Rheum* 2006, **54:**3885-3889.
27. Schiemann BJ, Neil JR, Schiemann WP: **SPARC inhibits epithelial cell proliferation in part through stimulation of the transforming growth factor-β-signaling system.** *Mol Biol Cell* 2003, **14:**3977-3988.
28. George J, Tsutsumi M: **siRNA-mediated knockdown of connective tissue growth factor prevents N-nitrosodimethylamine-induced hepatic fibrosis in rats.** *Gene Ther* 2007, **14:**790-803.
29. Francki A, McClure TD, Brekken RA, Motamed K, Murri C, Wang T, Sage EH: **SPARC regulates TGF-beta1-dependent signaling in primary glomerular mesangial cells.** *J Cell Biochem* 2004, **91:**915-925.
30. Cutroneo KR, White SL, Phan SH, Ehrlich HP: **Therapies for bleomycin induced lung fibrosis through regulation of TGF-beta1 induced collagen gene expression.** *J Cell Physiol* 2007, **211:**585-589.
31. Socha MJ, Manhiani M, Said N, Imig JD, Motamed K: **Secreted protein acidic and rich in cysteine deficiency ameliorates renal inflammation and fibrosis in angiotensin hypertension.** *Am J Pathol* 2007, **171:**1104-1112.
32. Kos K, Wong S, Tan B, Gummesson A, Jernas M, Franck N, Kerrigan D, Nystrom FH, Carlsson LM, Randeva HS, Pinkney JH, Wilding JP: **Regulation of the fibrosis and angiogenesis promoter SPARC/osteonectin in human adipose tissue by weight change, leptin, insulin, and glucose.** *Diabetes* 2009, **58:**1780-1788.
33. Reding T, Wagner U, Silva AB, Sun LK, Bain M, Kim SY, Bimmler D, Graf R: **Inflammation-dependent expression of SPARC during development of chronic pancreatitis in WBN/Kob rats and a microarray gene expression analysis.** *Physiol Genomics* 2009, **38:**196-204.

doi: 10.1186/ar2973
**Cite this article as:** Wang *et al.*, Attenuation of fibrosis in vitro and in vivo with SPARC siRNA *Arthritis Research & Therapy* 2010, **12**:R60

arthritis research&therapy

RESEARCH ARTICLE Open Access

# Decreased catalytic function with altered sumoylation of DNA topoisomerase I in the nuclei of scleroderma fibroblasts

Xiaodong Zhou[1*], Wei Lin[1], Filemon K Tan[1], Shervin Assassi[1], Mavin J Fritzler[2], Xinjian Guo[1], Roozbeh Sharif[1], Tom Xia[3], Syeling Lai[4] and Frank C Arnett[1]

## Abstract

**Introduction:** Sumoylation is involved in nucleolus-nucleoplasm transport of DNA topoisomerase I (topo I), which may associate with changes of cellular and topo I functions. Skin fibroblasts of patients with systemic sclerosis (SSc) exhibit profibrotic cellular changes. The aims of this study were to examine the catalytic function and sumoylation of topo I in the nuclei of SSc fibroblasts, a major cell type involved in the fibrotic process.

**Methods:** Eleven pairs of fibroblast strains obtained from nonlesional skin biopsies of SSc patients and age/sex/ethnicity-matched normal controls were examined for catalytic function of nuclear topo I. Immunoprecipitation (IP)-Western blots were used to examine sumoylation of fibroblast topo I. Real-time quantitative RT-PCR was used to measure transcript levels of SUMO1 and COL1A2 in the fibroblasts.

**Results:** Topo I in nuclear extracts of SSc fibroblasts generally showed a significantly lower efficiency than that of normal fibroblasts in relaxing equivalent amounts of supercoiled DNA. Increased sumoylation of topo I was clearly observed in 7 of 11 SSc fibroblast strains. Inhibition of SUMO1 with SUMO1 siRNA improved the catalytic efficiency of topo I in the SSc fibroblasts. In contrast, sumoylation of recombinant topo I proteins reduced their catalytic function.

**Conclusions:** The catalytic function of topo I was decreased in SSc fibroblasts, to which increased sumoylation of topo I may contribute.

## Introduction

Systemic sclerosis (SSc) is a human multi-system fibrotic disease with high morbidity and mortality but the etiology is largely unknown and the pathogenesis has yet to be clearly elucidated. Cutaneous fibrosis is a common clinical presentation and, based on the extent of skin involvement, SSc is classified into limited and diffuse cutaneous forms. The latter subset is characterized by more rapid progression of skin and visceral involvement, as well as poorer prognosis [1,2]. Skin fibroblasts obtained from SSc patients have been found to be profibrotic and to synthesize excessive amounts of ECM proteins, which contribute to tissue fibrosis [3]. It is believed that a possible defect in regulation of biological functions is present in SSc fibroblasts.

The majority of SSc patients (95%) have autoantibodies against various nuclear, nucleolar and cytoplasmic proteins, which include non-specific antinuclear antibodies (ANA) and a number of disease specific autoantibodies. Anti-DNA topoisomerase I (topo I) autoantibody is one of the disease-specific autoantibodies, and it occurs in 15 to 25% of patients [4-6]. A causal contribution of anti-topo I to the SSc phenotype is still unclear. There is no direct evidence indicating pathogenic roles of the antibodies. On the other hand, there is a strong association between anti-topo I autoantibody and the diffuse cutaneous form of SSc [5,6]. Levels of anti-topo I autoantibodies have been reported to correlate with disease severity and activity in SSc, and the lack of these antibodies conveys a better outcome in SSc [7]. In addition to

* Correspondence: xiaodong.zhou@uth.tmc.edu
[1]Division of Rheumatology, Department of Internal Medicine, University of Texas Health Science Center at Houston, Houston, TX 77030, USA
Full list of author information is available at the end of the article

anti-topo I, other SSc specific autoantibodies include those directed against centromeric proteins (ACA) that are associated with limited cutaneous disease, RNA polymerases (I, II and III) (ARA) and fibrillarin that are associated most often with diffuse skin involvement [8].

Topo I is a monomeric 100 kD nuclear protein that catalyzes the breaking and joining of DNA strands prior to transcription [9,10], and is associated with transcription, DNA replication and chromatin condensation. Topo I translocates between the nucleolus and the nucleoplasm, but is enriched in the nucleolus where there is a high level of transcription and replication of the ribosomal DNA [9,10]. Sumoylation is a post-translational modification, in which the substrates covalently attach the small ubiquitin-like modifier (SUMO) to lysine residues. Sumoylation is an important mechanism in regulating functions of target proteins and has been associated with the pathogenesis of autoimmune and inflammatory diseases, such as type I diabetes mellitus and rheumatoid arthritis [11,12]. Sumoylation of topo I was reported to facilitate its movement between the nucleolus and the nucleoplasm [13,14].

The goal of this study was to determine whether there is abnormal function, distribution and/or sumoylation of topo I in fibroblasts obtained from SSc patients that might associate with the presence of anti-nuclear and -nucleolar autoantibodies.

## Material and methods

### Dermal fibroblast cultures

Nonlesional skin biopsies (3 mm punch biopsies) were obtained from the upper arms of 11 SSc patients with disease of less than five years duration and 11 age- and gender-matched normal controls. All SSc patients fulfilled American College of Rheumatology criteria for SSc [15], and were positive for ANA. Two patients were positive for anti-topo I, four for ACA, two for ARA and one for anti-fibrillarin. Six patients had a diffused form of SSc, and five had limited SSc. Normal controls were undergoing dermatologic surgery and had no identified history of autoimmune diseases. All subjects provided informed consent and the study was approved by the Committee for the Protection of Human Subjects at The University of Texas Health Science Center at Houston.

Each skin sample was transported in Dulbecco's Modified Essential Media (DMEM) with 10% fetal calf serum (FCS) supplemented with penicillin and streptomycin for processing the same day. The tissue samples were washed in 70% ethanol, PBS and DMEM supplemented with 10% FCS. Cultured fibroblast cell strains were established by mincing tissues and placing them into 60 mm culture dishes secured by glass coverslips. The primary cultures were maintained in DMEM with 10% FCS and supplemented with penicillin and streptomycin. The early passage (< 5 passages) fibroblast strains were plated at a density of $2.5 \times 10^5$ cells in 35 mm plates and grown for assays accordingly.

### Catalytic function of topo I in SSc fibroblasts

Nuclear proteins were extracted from equal amounts of the cultured fibroblast cells by using nuclear extract kits (Active Motif, Carlsbad, CA, USA). The Topoisomerase I Assay kit (TopoGEN Inc., Port Orange, FL, USA) was used for measuring the catalytic function of topo I. Briefly, supercoiled DNA substrate (0.25 μg) (TopoGen, Inc.) was reacted with nuclear proteins containing topo I at serial dilutions. After 30-minute incubations at 37° C, the reaction was terminated with stop buffer (5% Sarkosyl, 0.125% bromophenol blue and 25% glycerol). The reaction mixtures were loaded and electrophoretically separated on a 1% agarose gel, and then stained with ethidium bromide. The catalytic activity of topo I was determined by measuring the intensity of the supercoiled DNA bands after reactions with a serial dilution of topo I in the nuclear extract of fibroblasts. A Bioimaging system (Gene Genius, Syngene, Frederick, MD, USA) was used to scan the bands in agarose gel. The Gene Snap software (Syngene) was used to quantify the intensity of the bands. A total of 11 pairs of SSc and control fibroblast strains were examined with this assay.

#### *Immunostaining*

SSc and normal fibroblasts were grown in culture media as described above. After 7, 14 and 18 days, the cells were washed with PBS and fixed with 100% methanol at 4°C for two minutes. The cells were washed with PBS again, and incubated with serum from SSc patients (evenly pooled from four SSc patients) who had positive anti-topo I autoantibodies, or monoclonal antibodies of mouse anti-human topo I or mouse anti-human SUMO 1. This was followed by incubation with green fluorescent protein (GFP) tagged secondary antibodies (rabbit anti-human IgG antibodies and anti-mouse antibodies). Nuclei were visualized by counterstaining DNA with 4',6-diamidino-2-phenylindole (DAPI) (Vector Laboratory Inc., Burlingame, CA, USA). The images of fibroblasts with fluorescence labeled proteins were acquired using fluorescence microscopy (Nikon Eclipse TE2000-4., Melville, NY. USA).

### Western blotting

The protein concentration of nuclear extracts from cultured fibroblasts was measured using the standard curve in a TECAN spectrophotometer (Tecan Group Ltd., Switzerland, 8708 Mannedorf). Equal amounts of protein from each sample were subjected to SDS-polyacrylamide gel electrophoresis. Resolved proteins were transferred onto nitrocellulose membranes and

incubated with 1:1,000 diluted primary antibodies including mouse anti-human topo I (ImmunoVision, Springdale, AR, USA), anti-human SUMO1 (ABGENT, San Diego, CA, USA) and anti-collagen type I, individually. The secondary antibody was a peroxidase-conjugated anti-mouse IgG (Amersham, Piscataway, NJ, USA). Specific proteins were detected by chemiluminescence using an Enhanced Chemiluminescence (ECL) system (Amersham). The intensity of the bands was quantified using ImageQuant software (Molecular Dynamics, Sunnyvale, CA, USA).

#### *Immunoprecipitation (IP) Western blotting*

Approximately $3.5 \times 10^7$ fibroblast cells of each subject were harvested by trypsinizing the adherent cells and washed twice with 25 ml ice-cold PBS containing phosphatase inhibitors. Cell pellets were then gently resuspended by 2 ml hypotonic buffer and nuclear extracts prepared and measured for protein concentration by a spectrophotometer as described above. Equal amounts of protein (500 ug) from each sample were subjected to immunoprecipitation (IP) with mouse anti-SUMO-1 (GMP1, Invitrogen, Carlsbad, CA, USA) using nuclear complex co-IP kit (Active Motif, Carlsbad, CA), and then subjected to SDS-polyacrylamide gel electrophoresis. Resolved proteins were transferred onto nitrocellulose membranes and incubated with primary antibodies of mouse anti-human topo I (ImmunoVision) diluted to 1:1,000. The secondary antibody was a horseradish peroxidase-conjugated anti-mouse IgG (eBioscience, San Diego, CA, USA). Specific proteins were detected by chemiluminescence using Supersignal West Pico stable peroxide solution (Thermo Scientific, Rockford, IL, USA). The intensity of the bands was quantified using ImageQuant software (Molecular Dynamics).

### Inhibition of SUMO1 with siRNA transfection in fibroblasts

SUMO1 siRNAs were purchased from Invitrogen. Three SSc fibroblast strains that showed stronger sumoylation of topo I and weaker catalytic topo I function were used for transfection of SUMO1 siRNA. Briefly, the fibroblasts were grown at a density of $1.5 \times 10^5$ cells in 25-$cm^2$ flasks until confluency. The DMEM culture medium in each culture flask was replaced with Opti_MEM 1 (Invitrogen) without FCS. The fibroblasts were transfected with SUMO siRNA using Lipofectamine RNAiMAX (Invitrogen) at a concentration of 15 ug/ml. A fluorescein-labeled non-silencing control siRNA (Qiagen, Valencia, CA, USA) was used for detection of transfection efficiency. After 24 hours, the culture medium was replaced with normal DMEM. The fibroblasts were examined for gene and protein expression, as well as topo I catalytic function after 48- or 72-hour transfection.

### Sumoylation assay of topo I

A mixture containing recombinant topo I protein (TopoGEN Inc.), SUMO-1 protein (Active Motif), activating enzyme E1/conjugating enzyme E2 (Active Motif) and sumoylation buffer (15 mM ATP, 25 mM MgCl2 and 250 mM Tris-HCl) was incubated at 30°C for three hours. A mutant SUMO-1 protein (Active Motif) lacking sumoylation function was used as a negative control. The reaction was stopped with 5 mM EDTA and the recombinant human topo I with and without sumoylation were examined by Western blotting and topo I catalytic assays. The experiments were performed in triplicate.

### Quantitative reverse-transcriptase-polymerase chain reaction (RT-PCR) for measurement of SUMO1 expression, as well as COL1A2 expression after SUMO1 siRNA transfection

The primers and probes of SUMO1, COL1A2, 18S and GAPDH were obtained from Applied Biosystems (Assays-on-Demand product line; Foster City, CA, USA). Total RNA from each sample was extracted from the cultured fibroblasts described above using a total RNA kit from OMEGA Biotek (Norcross, GA, USA) after treatment with DNase I. Complementary DNA (cDNA) was synthesized using SuperScript II reverse transcriptase (Invitrogen). Synthesized cDNAs were mixed with primer/probe of SUMO1 or COL1A2 in 2 × TaqMan universal PCR buffer and then assayed on an ABI Prism 7900 Sequence Detector System (Applied Biosystems). Each sample was assayed in triplicate. The data were analyzed with SDS2.2 (ABI). The amount of each transcript was normalized with 18S and GAPDH levels.

### Measurement of autoantibodies

Patients' sera were tested for antinuclear antibodies by indirect immunofluorescence (IIF) using HEp-2 cells as antigen substrate and fluorescent goat anti-human IgG as a secondary antibody (Antibodies Inc., Davis, CA, USA). Anti-topo I antibodies were detected by passive immunodiffusion kits that employed calf thymus extracts as the antigen source (INOVA Diagnostics, San Diego, CA, USA), anti-RNA polymerase III antibodies were detected by ELISA using commercial kits (MBL, Nagoya, Japan). Anti-centromere antibodies were determined visually by their distinctive IIF patterns on HEp-2 cells. Anti-fibrillarin antibodies were detected by immunoprecipitation as described previously [16].

## Results

### Reduced catalytic function of topo I in SSc fibroblasts

After catalytic reactions with a serial dilution of topo I in the nuclear extracts, the supercoiled DNA band was

gradually diminished following increased amounts of topo I in the nuclear extracts. Based on the intensity of supercoiled DNA bands that were correlated with the amounts of topo I in the nuclear extracts, the efficiency of SSc topo I in relaxing the supercoiled DNA appeared to be less than that of control topo I in each concentration of nuclear extracts (Figure 1A). Comparison of average band intensity of remaining supercoiled DNA in each of six dilutions between all SSc and all control fibroblasts showed a significant *P* value (*P* = 0.0041) (Student's *t*-test) (Figure 1B).

### Altered localization of topo I in SSc fibroblasts

When anti-topo I monoclonal antibodies were used as probes, the majority of SSc fibroblasts from each patient showed strong nucleoplasmic staining (multiple speckles) compared to normal fibroblasts in which topo I staining was enriched in the nucleolus (Figure 2A). A few SSc fibroblasts (less than 1%) showed cytoplasmic (cytosolic) staining which was not observed in normal fibroblasts. However, there were more SSc fibroblasts (approximately 2%) showing cytoplasmic staining of topo I molecules when anti-topo I positive sera from SSc patients were used as probes (Figure 2B). The cytoplasmic staining of topo I appeared to be stronger at 14 or 18 days of culture compared to 7 days.

### Altered sumoylation of topo I in SSc fibroblasts

Western blotting showed that the quantitative levels of topo I proteins were similar between SSc and normal control fibroblasts, while SUMO 1 levels were increased in SSc fibroblasts. To validate this finding, we examined sumoylated topo I in the nuclear proteins using IP Western blotting (Figure 3). Increased sumoylation of topo I (higher intensity of the bands and presence of polysumoylation of topo I) evaluated by IP Western blots was clearly observed in 7 of 11 SSc fibroblast strains (2 anti-topo I positive patients, 4 anti-RNA polymerase III positive patients and 1 anti-fibrillarin positive patient (Figure 3). Interestingly, four SSc fibroblast strains, including two each from patients with anti-centromere and with no detectable SSc specific autoantibodies,

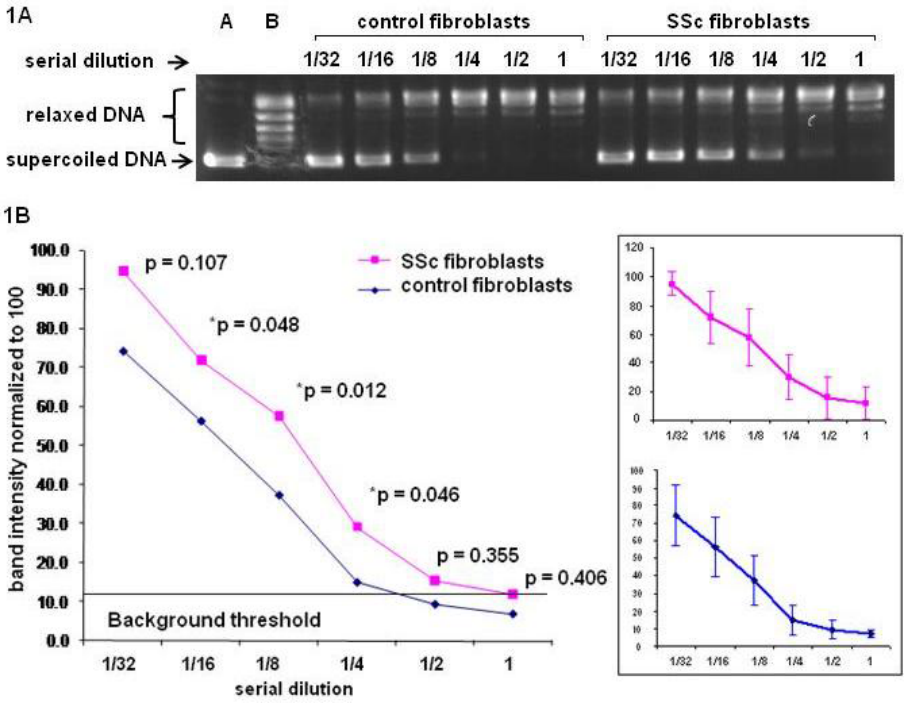



**Figure 1 Measurement of catalytic function of topo I in cultured fibroblasts**. A serial dilution of topo I in the nuclear extracts obtained from SSc and control fibroblasts was used to relax 0.25 μg supercoiled DNA. **A**. The supercoiled DNA band is gradually diminished following increased amounts of topo I in the nuclear extracts in the relaxing assays. The efficiency of SSc topo I in relaxing the supercoiled DNA appeared to be less than that of control topo I in each concentration of nuclear extracts. **B**. Comparison of 11 paired SSc and control fibroblasts for mean values of intensity of supercoiled DNA bands after relaxing assay with different concentrations of topo I in the nuclear extracts. Each *P*-value of comparison at different dilution points is listed in the figure. Comparison of average band intensity of remaining supercoiled DNA in each of six dilutions between all SSc and all control fibroblasts showed a significant *P*-value (*P* = 0.0041) (Student's *t*-test). A = standard supercoiled DNA band; B = standard relaxed DNA bands; the numbers (1/32, 1/16, 1/8, 1/4, 1/2 and 1) indicate serial dilutions of topo I in nuclear extracts used for relaxing supercoiled DNA. The error bars indicate standard deviation (SD).

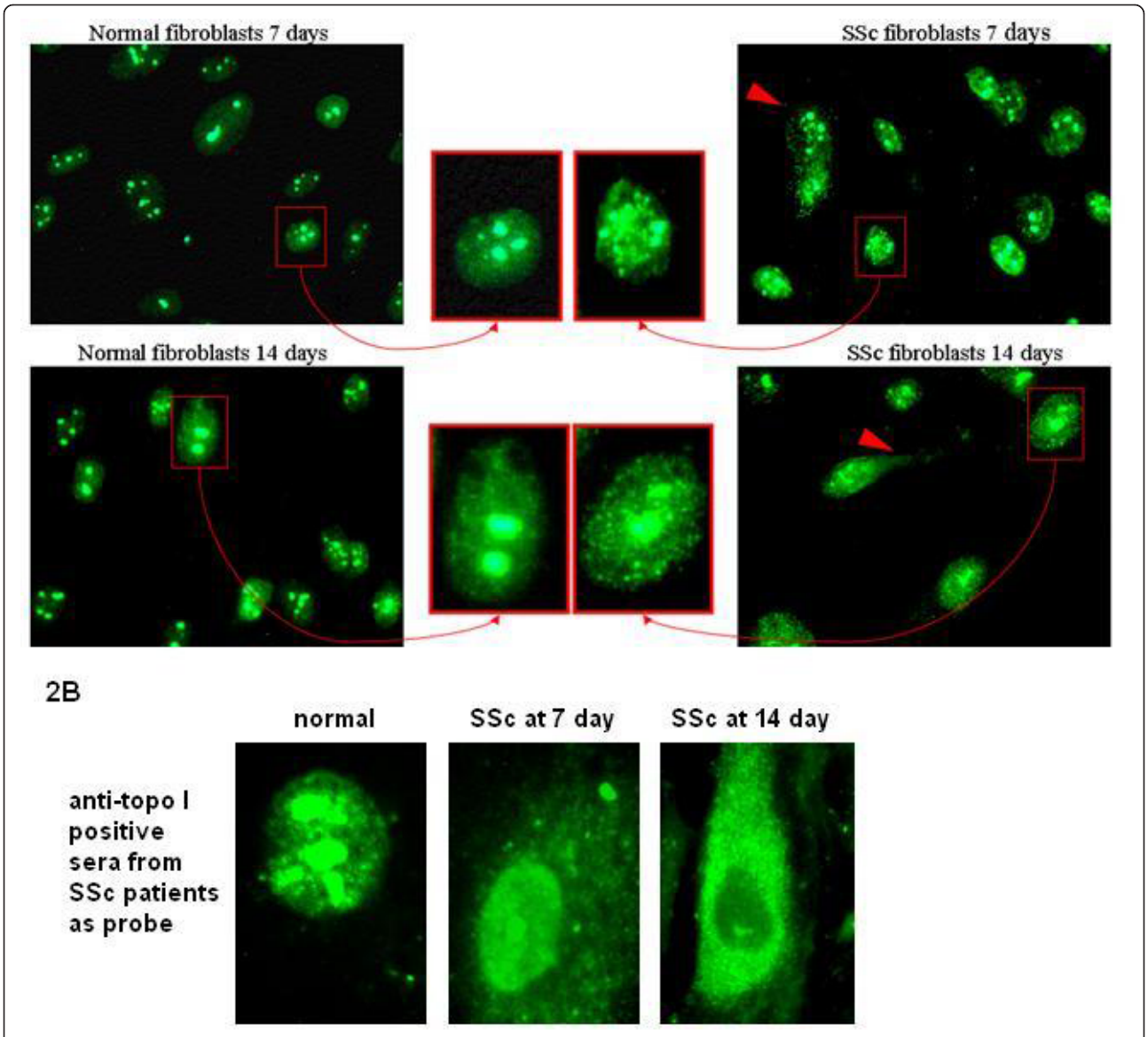



**Figure 2 Comparison of topo I staining in cultured fibroblasts of normal controls and SSc patients**. **A**. Topo I immunostaining with anti-topo I monoclonal antibodies showed multiple speckles in the nucleoplasm of SSc fibroblasts, which is differentiated from that in normal fibroblasts (relatively homogenous stain of topo I) at both 7 and 14 days of cultures. Some SSc fibroblasts show cytoplasmic staining of topo I protein (marked with red arrow heads). **B**. Topo I immunostaining with anti-topo I positive sera from SSc patients show the expected nuclear/nucleolar staining as well as cytoplasmic staining of SSc fibroblasts. At Day 14, the cytoplasmic staining appeared to increase relative to the nucleoplasm and nucleolar staining.

showed similar levels of sumoylation as their normal counterparts.

### Inhibition of SUMO1 in SSc fibroblasts increased catalytic function of topo I

Real-time quantitative RT-PCR showed that inhibition of SUMO1 with siRNA achieved a significant reduction of gene expression of SUMO1 (Figure 4). Compared to non-target siRNA transfected fibroblasts, SUMO1 siRNA transfected fibroblasts showed a 30.97-times reduction of SUMO1 expression ($P < 0.001$, T test) (Figure 4a). Western blots showed a concordant change of the SUMO1 protein (Figure 4b). Importantly, compared to either non-target siRNA transfected or non-siRNA transfected fibroblasts, catalytic function of topo I of sumo1 siRNA transfected SSc fibroblasts showed a marked improvement in all three test fibroblast strains (Figure 5). Measurements of the COL1A2 gene expression with quantitative RT-PCR and collagen type I protein expression with Western blots did not show

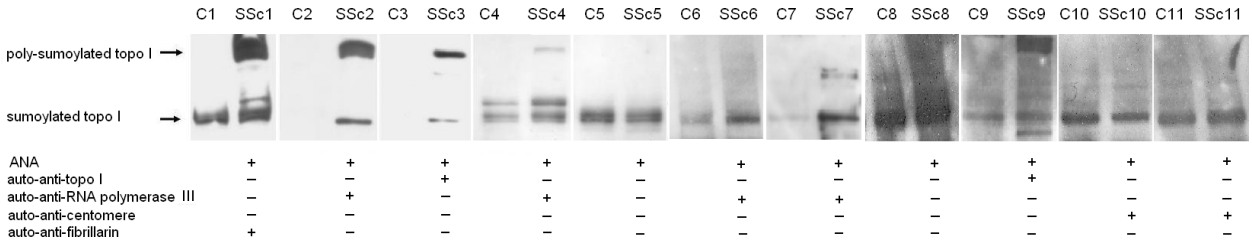



**Figure 3 Immunoprecipitated Western blots and autoantibody profiles for 11 SSc patients**. Each SSc patient (SSc1 to 11) has an age and sex matched normal control (C1 to 11) for comparison of sumoylated topo I expression with IP Western blots. Poly-sumoylated topo I appeared in SSc fibroblast strain number 1, 2, 3, 4, 7 and 9. Increased sumoylation of topo I also is observed in the case number 6 compared to its normal counterpart, but not in the case number 5, 8, 10 and 11. ANA, antinuclear antibodies.

significant changes after SUMO1 siRNA transfection in the fibroblasts.

### Sumoylation of recombinant topo I decreased its catalytic function

Recombinant human topo I proteins were sumoylated with either wild type SUMO1 or mutant SUMO1 or negative control (without sumoylation) and then were examined with Western blot for sumoylated topo I and with topo I catalytic assays for topo I function. Poly-sumoylation of topo I was observed in the topo I proteins sumoylated with wild type SUMO1 (Figure 6). Sumoylation of topo I with wild type SUMO1 showed a reduction of efficiency in catalytic function compared to the topo I protein sumoylated with mutant sumo 1 or negative control (Figure 7). The assays were performed in triplicates, which showed similar results.

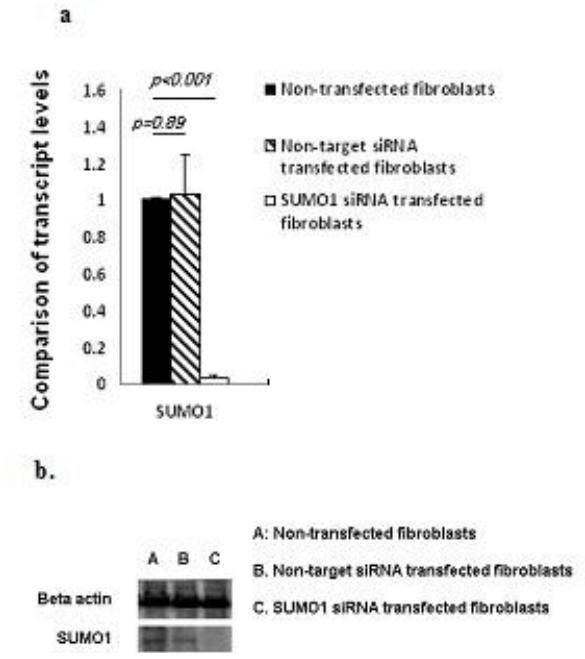



**Figure 4 Real-time RT-PCR and Western blots for SUMO1 with and without SUMO1 siRNA transfection in fibroblasts**. Three SSc fibroblast strains (two with anti-topo I and one with anti-RNA polymerase III positive serum) were transfected with SUMO1 siRNA. After 48-hour transfection, total RNAs were used for measuring SUMO1 transcript levels (Figure 4a), and the nuclear extracts were used for measuring SUMO1 protein (Figure 4b). Error bars indicate standard deviation.

## Discussion

A novel finding of these studies is the observation that human SSc fibroblasts have a decreased catalytic function of topo I. Human topo I plays an important role in DNA metabolic processes, such as transcription and replication, in which it releases topological stress in DNA chains [9,10]. Topo I is generally localized in the nucleolus where a high level of transcription and replication of ribosomal DNA occurs. In response to inhibitory factors to topo I, such as camptothecin, UV irradiation and transcription inhibitors, topo I molecules were usually relocated from the nucleolus to the nucleoplasm due to mechanisms that are not clearly understood. [17-19]. Interestingly, SSc fibroblasts examined herein showed enhanced staining of topo I in the nucleoplasm, which suggests a relocation of topo I, and also supports a reduced function of topo I-associated DNA metabolic processes.

The cytoplasmic staining of topo I observed in some SSc fibroblasts was mainly detected by anti-topo I positive serum from SSc patients and was different from that found using anti-topo I monoclonal antibodies. Considering that the polyclonal human sera may contain mainly a variety of autoantibodies that have non-specific and antigen specific cross-reactions to cytoplasmic proteins is a possible explanation. With respect to possible cross-reactions, it is interesting that mitochondrial topo I has high amino acid homology to nuclear topo I. On the other hand, it is also possible that the cytoplasmic staining of topo I may represent ubiquitinated topo I molecules being processed by cytoplasmic proteasomes. It is worth noting that the topo I autoantigenic component, a 70 kD polypeptide, has been reported to be

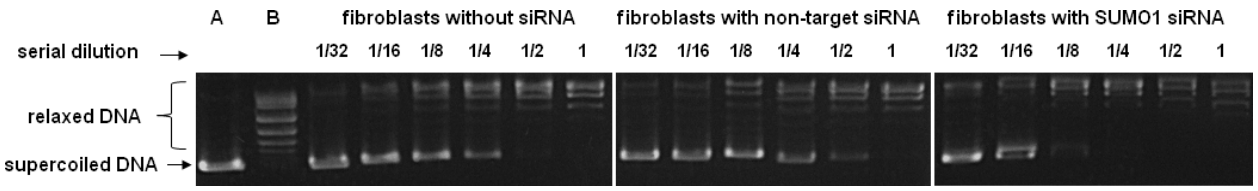



**Figure 5 Catalytic function of topo I in cultured SSc fibroblasts with and without SUMO1 siRNA transfection**. A serial dilution of the nuclear extract containing topo I obtained from SSc fibroblasts was used to relax 0.25 μg supercoiled DNA. In this figure, the supercoiled DNA band was completely transformed to relaxed DNA at dilutions of one half and one in the fibroblasts without siRNA transfection or non-target siRNA transfection. In contrast, this change was observed between the one-eighth and one-fourth dilutions in the fibroblasts with SUMO1 transfection, which indicates a higher efficiency of catalytic function of topo I after SUMO1 inhibition in the fibroblasts. According to the intensity of the bands of remaining supercoiled DNA in serial dilutions in the assays of three fibroblast strains, these changes are significant. The *P*-values are 0.045 and 0.027 at the one-fourth dilution for comparisons between SUMO1 siRNA vs. non-target siRNA, or vs. without siRNA transfected fibroblasts, respectively (Student's *t*-test). This is representative of three SSc fibroblast strains examined in SUMO1 siRNA studies. *__A__, supercoiled DNA; __B__, relaxed DNA.

exported via ectocytosis in SSc fibroblasts [20], and anti-topo I autoantibodies of SSc patients have been shown to bind to SSc fibroblasts [21].

Sumoylation is an important post-translational modification. Previous studies have indicated that sumoylation of topo I facilitates translocation of topo I protein from the nucleolus to nucleoplasm [13,14]. Increased sumoylation of topo I in certain SSc fibroblasts observed herein supports a potential mechanism that may drive the movement of topo I from the nucleolar compartments to the nucleoplasm where a degradation process may occur in proteasomes. To further investigate the association between altered sumoylation and topo I function in SSc fibroblasts, we inhibited the SUMO1 expression with sequence specific SUMO1 siRNA. Interestingly, SUMO1 inhibition was associated with a favorable improvement of the catalytic function of fibroblast topo I, suggesting that decreased topo I function observed in SSc fibroblasts may be a result of increased sumoylation. This possibility was consistent with the follow-up studies of sumoylation of recombinant human topo I that showed a reduction of catalytic function. However, sumoylation may not fully explain the reduction of topo I function in all SSc fibroblasts, especially in those fibroblasts which did not show the changes of sumoylation of topo I. These fibroblasts include two each from patients with ACA and with non-SSc specific ANAs. In contrast, the fibroblasts from all seven patients with either anti-topo I ARA or anti-fibrillarin showed hyper-sumoylation of topo I. All these three autoantibodies target primary nucleolar proteins. It is worth noting that the presence of any one of these autoantibodies in SSc patients is associated with the diffuse form of SSc and internal organ fibrosis [8], while the anti-centromere positive patients usually have a limited form of SSc with favorable clinical outcomes [8]. Indeed, all SSc patients examined here with hypersumoylation of topo I presented as the diffuse form of SSc, except one, who was positive to ARA, but also clinically had lupus-like disease and anti-ribonucleoprotein (RNP) autoantibodies. All four SSc patients with unchanged sumoylation of topo I presented as the limited form of SSc at the time of skin biopsies. Therefore, sumoylaton of topo I in SSc fibroblasts appeared to be correlated with the status of skin fibrosis, which in some SSc patients changes over time. Recent studies of SSc genetics have indicated that different genetic susceptibility markers may determine the types of autoantibodies presenting in SSc patients [22,23]. The characteristic patterns and specific genetic associations of SSc autoantibodies suggest that distinctive mechanisms contribute to different autoantibody-associated SSc subsets.

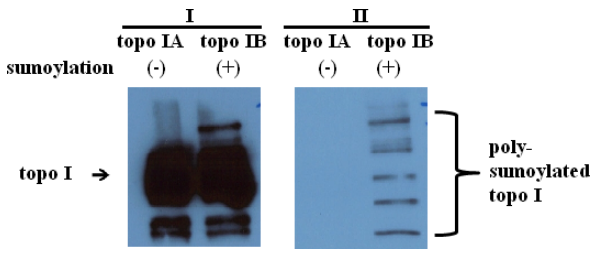



**Figure 6 Western blots show sumoylation of recombinant human topo I**. Recombinant human topo I protein was subjected to the sumoylation reaction and examined by Western blotting using anti-topo I (I) and anti-SUMO1 antibodies (II). Compared to topo I protein without sumoylation reaction (topo I **A**), topo I protein with sumoylation reaction (topo I **B**) showed poly-sumoylation of topo I (II). The assays showed similar results in triplicates.

Topo I is an essential functional component of human cells. Previous reports indicated that knock out of the topo I gene was associated with death at an early stage of embryogenesis [24,25]. Inactivation of the topo I gene *in vitro* was found to induce genomic instability with chromosomal aberrations [26]. Inhibition of topo I function through camptothecin or topotecan (a

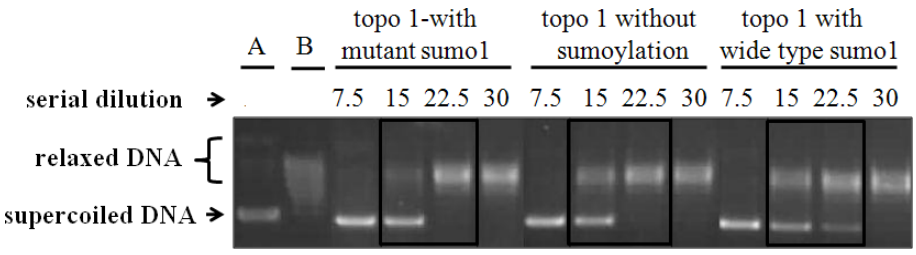



**Figure 7 Measurement of catalytic function of recombinant human topo I with and without sumoylation reaction**. Recombinant human topo I proteins were sumoylated with either mutant sumo1 or wild type sumo1 or negative control (without sumoylation), and then were examined for their catalytic function in a serial dilution. Sumoylation of topo I with wild type sumo1 showed a reduction of efficiency in catalytic function (supercoiled DNA disappeared at the dilution of topo I concentration of 30) compared to the topo I protein sumoylated with mutant sumo1 or negative control (supercoiled DNA disappeared at topo I concentration of 22.5). This is representative of three assays. *__A__, standard supercoiled DNA band; __B__, standard relaxed DNA bands.

camptothecin derivative) in human HEp-2 cells altered nuclear structure and function and targeted topo I for proteasomal degradation [27]. Although, we do not know whether sumoylation of topo I in SSc fibroblasts contributes to any changes of specific antigen binding or autoantibody presentation in SSc patients, decreased catalytic function of topo I may alter the nuclear structure and function of the fibroblasts, which may influence other nuclear proteins including RNA pol III and fibrillarin. Of potential significance to our study, topotecan used therapeutically for cancer has been reported to induce SSc-like disease [28]. Whether decreased catalytic function of topo I in SSc fibroblasts examined herein may result in any consequences associated with pathological changes in SSc is worthy of further investigations.

## Conclusions

In summary, our studies of topo I in SSc fibroblasts indicate that topo I is functionally altered and is relocated to the nucleoplasm. In some fibroblasts, especially those obtained from skin biopsies of SSc patients who were positive for anti-topo I, anti-RNA polymerase III and anti-fibrillarin autoantibodies, these alterations were associated with increased sumoylation of topo I. In contrast, the fibroblasts of anti-centromere positive patients showed unchanged sumoylation of topo I. Inhibition of SUMO1 gene improved catalytic function of topo I in SSc fibroblasts. These observations may provide important insights into the nature of SSc fibroblasts that may contribute to pathological processes, induction of an autoimmune response to topo I, and/or disease development in SSc.

### Abbreviations

ANA: anti-nuclear antibodies; COL1A2: collagen type 1A2; DAPI: 4',6-diamidino-2-phenylindole; DMEM: Dulbecco's Modified Essential Media; ECL: Enhanced Chemiluminescence; ECM: extracellular matrix; FCS: fetal calf serum; GFP: green fluorescent protein; IIF: indirect immunofluorescence; IP: immunoprecipitation; RNP: ribonucleoprotein; SSc: systemic sclerosis; SUMO1: small ubiquitin-like modifier 1; Topo I: DNA topoisomerase I.

### Acknowledgements

This study was supported by grants from the Department of the Army, Medical Research Acquisition Activity, grant number PR064803 to Zhou and the National Institutes of Health, grant number P50 AR054144 to Arnett.

### Author details

[1]Division of Rheumatology, Department of Internal Medicine, University of Texas Health Science Center at Houston, Houston, TX 77030, USA. [2]Department of Medicine, University of Calgary, Calgary, AB T2N 1N4, Canada. [3]Rice University, Houston, Post Office Box 1892, TX 77030, USA. [4]Departmant of Pathology, Baylor College of Medicine, Houston, TX 77030, USA.

### Authors' contributions

ZX carried out research design, experiments and manuscript writing. WL conducted molecular studies and cell cultures. TF and AS conducted skin biopsies and helped with manuscript preparation. FM conducted autoantibody tests and manuscript preparation. GX carried out molecular studies and cell cultures. SR enrolled patients and did skin biopsies. XT conducted molecular studies. LS did skin biopsies for controls. AF carried out research design and manuscript preparation. All authors read and approved the final version of the manuscript.

### Competing interests

The authors declare that they have no competing interests.

Received: 12 November 2010 Revised: 18 June 2011
Accepted: 9 August 2011 Published: 9 August 2011

### References

1. Tamby MC, Chanseaud Y, Guillevin L, Mouthon L: **New insights into the pathogenesis of systemic sclerosis.** *Autoimmun Rev* 2003, **2**:152-157.
2. Steen VD: **The many faces of scleroderma.** *Rheum Dis Clin N Am* 2008, **34**:1-15.
3. Claman HN, Giorno RC, Seibold JR: **Endothelial and fibroblastic activation in scleroderma. The myth of the "uninvolved skin".** *Arthritis Rheum* 1991, **34**:1495-1501.
4. Maul GG, French BT, van Venrooij WJ, Jimenez SA: **Topoisomerase I identified by scleroderma 70 antisera: enrichment of topoisomerase I at the centromere in mouse mitotic cell before anaphase.** *Proc Natl Acad Sci USA* 1986, **83**:5145-5149.
5. Diot E, Giraudeau B, Diot P, Degenne D, Ritz L, Guilmot JL, Lemarie E: **Is anti-topoisomerase I a serum marker of pulmonary involvement in systemic sclerosis?** *Chest* 1999, **116**:715-720.
6. Jarzabek-Chorzelska M, Blaszczyk M, Jablonska S, Chorzelski T, Kumar V, Beutner EH: **Scl 70 antibody–a specific marker of systemic sclerosis.** *Br J Dermatol* 1986, **115**:393-401.

7. Kuwana M, Kaburaki J, Mimori T, Kawakami Y, Tojo T: **Longitudinal analysis of autoantibody response to topoisomerase I in systemic sclerosis.** *Arthritis Rheum* 2000, **43**:1074-1084.
8. Bunn CC, Black CM: **Systemic sclerosis: an autoantibody mosaic.** *Clin Exp Immunol* 1999, **117**:207-208.
9. Leppard JB, Champoux JJ: **Human DNA topoisomerase I: relaxation, roles, and damage control.** *Chromosoma* 2005, **114**:75-85.
10. Wang JC: **Cellular roles of DNA topoisomerases: a molecular perspective.** *Nat Rev Mol Cell Biol* 2002, **3**:430-440.
11. Li M, Guo D, Isales CM, Eizirik DL, Atkinson M, She JX, Wang CY: **SUMO wrestling with type 1 diabetes.** *J Mol Med* 2005, **83**:504-514.
12. Meinecke I, Cinski A, Baier A, Peters MA, Dankbar B, Wille A, Drynda A, Mendoza H, Gay RE, Hay RT, Ink B, Gay S, Pap T: **Modification of nuclear PML protein by SUMO-1 regulates Fas-induced apoptosis in rheumatoid arthritis synovial fibroblasts.** *Proc Natl Acad Sci USA* 2007, **104**:5073-5078.
13. Rallabhandi P, Hashimoto K, Mo YY, Beck WT, Moitra PK, D'Arpa P: **Sumoylation of topoisomerase I is involved in its partitioning between nucleoli and nucleoplasm and its clearing from nucleoli in response to camptothecin.** *J Biol Chem* 2002, **277**:40020-40026.
14. Mo YY, Yu Y, Shen Z, Beck WT: **Nucleolar delocalization of human topoisomerase I in response to topotecan correlates with sumoylation of the protein.** *J Biol Chem* 2002, **277**:2958-2964.
15. Subcommittee for Scleroderma, Criteria of the American Rheumatism Association Diagnostic and Therapeutic Criteria Committee: **Preliminary criteria for the classification of systemic sclerosis (scleroderma).** *Arthritis Rheum* 1980, **23**:581-590.
16. Arnett FC, Reveille JD, Goldstein R, Pollard KM, Leaird K, Smith EA, Leroy EC, Fritzler MJ: **Autoantibodies to fibrillarin in systemic sclerosis (scleroderma). An immunogenetic, serologic, and clinical analysis.** *Arthritis Rheum* 1996, **39**:1151-1160.
17. Buckwalter CA, Lin AH, Tanizawa A, Pommier YG, Cheng YC, Kaufmann SH: **RNA synthesis inhibitors alter the subnuclear distribution of DNA topoisomerase I.** *Cancer Res* 1996, **56**:1674-1681.
18. Christensen MO, Krokowski RM, Barthelmes HU, Hock R, Boege F, Mielke C: **Distinct effects of topoisomerase I and RNA polymerase I inhibitors suggest a dual mechanism of nucleolar/nucleoplasmic partitioning of topoisomerase I.** *J Biol Chem* 2004, **279**:21873-21882.
19. Mao Y, Mehl IR, Muller MT: **Subnuclear distribution of topoisomerase I is linked to ongoing transcription and p53 status.** *Proc Natl Acad Sci USA* 2002, **99**:1235-1240.
20. Hsu TC, Lee TL, Tsay GJ: **Autoantigen components recognizable by scleroderma sera are exported via ectocytosis of fibroblasts.** *Br J Rheumatol* 1997, **36**:1038-1044.
21. Henault J, Tremblay M, Clement I, Raymond Y, Senecal JL: **Direct binding of anti-DNA topoisomerase I autoantibodies to the cell surface of fibroblasts in patients with systemic sclerosis.** *Arthritis Rheum* 2004, **50**:3265-3274.
22. Zhou X, Lee JE, Arnett FC, Xiong M, Park MY, Yoo YK, Shin ES, Reveille JD, Mayes MD, Kim JH, Song R, Choi JY, Park JA, Lee YJ, Lee EY, Song YW, Lee EB: **HLA-DPB1 and DPB2 are genetic loci for systemic sclerosis: a genome-wide association study in Koreans with replication in North Americans.** *Arthritis Rheum* 2009, **60**:3807-3814.
23. Arnett FC, Gourh P, Shete S, Ahn CW, Honey R, Agarwal SK, Tan FK, McNearney T, Fischbach M, Fritzler MJ, Mayes MD, Reveille JD: **Major Histocompatibility Complex (MHC) class II alleles, haplotypes, and epitopes which confer susceptibility or protection in the fibrosing autoimmune disease systemic sclerosis: analyses in 1,300 Caucasian, African-American and Hispanic cases and 1,000 controls.** *Ann Rheum Dis* 2010, **69**:822-827.
24. Lee MP, Brown SD, Chen A, Hsieh TS: **DNA topoisomerase I is essential in Drosophila melanogaster.** *Proc Natl Acad Sci USA* 1993, **90**:6656-6660.
25. Morham SG, Kluckman KD, Voulomanos N, Smithies O: **Targeted disruption of the mouse topoisomerase I gene by camptothecin selection.** *Mol Cell Biol* 1996, **16**:6804-6809.
26. Miao ZH, Player A, Shankavaram U, Wang YH, Zimonjic DB, Lorenzi PL, Liao ZY, Liu H, Shimura T, Zhang HL, Meng LH, Zhang YW, Kawasaki ES, Popescu NC, Aladjem MI, Goldstein DJ, Weinstein JN, Pommier Y: **Nonclassic functions of human topoisomerase I: genome-wide and pharmacologic analyses.** *Cancer Res* 2007, **67**:8752-8761.
27. Chen M, Dittmann A, Kuhn A, Ruzicka T, von Mikecz A: **Recruitment of topoisomerase I (Scl-70) to nucleoplasmic proteasomes in response to xenobiotics suggests a role for altered antigen processing in scleroderma.** *Arthritis Rheum* 2005, **52**:877-884.
28. Ene-Stroescu D, Ellman MH, Peterson CE: **Topotecan and the development of scleroderma or a scleroderma-like illness.** *Arthritis Rheum* 2002, **46**:844-845.

doi:10.1186/ar3435
**Cite this article as:** Zhou *et al.*: Decreased catalytic function with altered sumoylation of DNA topoisomerase I in the nuclei of scleroderma fibroblasts. *Arthritis Research & Therapy* 2011 **13**:R128.

The Journal of Rheumatology

**The Journal of Rheumatology** **Volume 38, no. 8**

# Anti-Fibrillarin Antibody in African American Patients with Systemic Sclerosis: Immunogenetics, Clinical Features, and Survival Analysis

ROOZBEH SHARIF, MARVIN J. FRITZLER, MAUREEN D. MAYES, EMILIO B. GONZALEZ, TERRY A. McNEARNEY, HILDA DRAEGER, MURRAY BARON, the Canadian Scleroderma Research Group, DANIEL E. FURST, DINESH K. KHANNA, DEBORAH J. DEL JUNCO, JERRY A. MOLITOR, ELENA SCHIOPU, KRISTINE PHILLIPS, JAMES R. SEIBOLD, RICHARD M. SILVER, ROBERT W. SIMMS, GENISOS Study Group, MARILYN PERRY, CARLOS ROJO, JULIO CHARLES, XIAODONG ZHOU, SANDEEP K. AGARWAL, JOHN D. REVEILLE, SHERVIN ASSASSI and FRANK C. ARNETT

J Rheumatol 2011;38;1622-1630
http://www.jrheum.org/content/38/8/1622

# Anti-Fibrillarin Antibody in African American Patients with Systemic Sclerosis: Immunogenetics, Clinical Features, and Survival Analysis

ROOZBEH SHARIF, MARVIN J. FRITZLER, MAUREEN D. MAYES, EMILIO B. GONZALEZ, TERRY A. McNEARNEY, HILDA DRAEGER, MURRAY BARON, the Canadian Scleroderma Research Group, DANIEL E. FURST, DINESH K. KHANNA, DEBORAH J. DEL JUNCO, JERRY A. MOLITOR, ELENA SCHIOPU, KRISTINE PHILLIPS, JAMES R. SEIBOLD, RICHARD M. SILVER, ROBERT W. SIMMS, GENISOS Study Group, MARILYN PERRY, CARLOS ROJO, JULIO CHARLES, XIAODONG ZHOU, SANDEEP K. AGARWAL, JOHN D. REVEILLE, SHERVIN ASSASSI, and FRANK C. ARNETT

***ABSTRACT.*** ***Objective.*** Anti-U3-RNP, or anti-fibrillarin antibodies (AFA), are detected more frequently among African American (AA) patients with systemic sclerosis (SSc) compared to other ethnic groups and are associated with distinct clinical features. We examined the immunogenetic, clinical, and survival correlates of AFA in a large group of AA patients with SSc.

***Methods.*** Overall, 278 AA patients with SSc and 328 unaffected AA controls were enrolled from 3 North American cohorts. Clinical features, autoantibody profile, and HLA class II genotyping were determined. To compare clinical manifestations, relevant clinical features were adjusted for disease duration. Cox proportional hazards regression was used to determine the effect of AFA on survival.

***Results.*** Fifty (18.5%) AA patients had AFA. After Bonferroni correction, HLA-DRB1*08:04 was associated with AFA, compared to unaffected AA controls (OR 11.5, $p < 0.0001$) and AFA-negative SSc patients (OR 5.2, $p = 0.0002$). AFA-positive AA patients had younger age of disease onset, higher frequency of digital ulcers, diarrhea, pericarditis, higher Medsger perivascular and lower Medsger lung severity indices ($p = 0.004$, $p = 0.014$, $p = 0.019$, $p = 0.092$, $p = 0.006$, and $p = 0.016$, respectively). After adjustment for age at enrollment, AFA-positive patients did not have different survival compared to patients without AFA ($p = 0.493$).

***Conclusion.*** Our findings demonstrate strong association between AFA and HLA-DRB1*08:04 allele in AA patients with SSc. AA SSc patients with AFA had younger age of onset, higher frequency of digital ulcers, pericarditis and severe lower gastrointestinal involvement, but less severe lung involvement compared to AA patients without AFA. Presence of AFA did not change survival.
(First Release May 15 2011; J Rheumatol 2011;38:1622–30; doi:10.3899/jrheum.110071)

*Key Indexing Terms:*
SCLERODERMA GENISOS ANTI-U3-RNP DIGITAL ULCER
HLA-DRB1 SCLERODERMA FAMILY REGISTRY

*From the Division of Rheumatology and Immunogenetics, University of Texas Health Science Center at Houston, Houston, Texas; Department of Medicine, University of Calgary, Calgary, Alberta, Canada; Division of Rheumatology, University of Texas Medical Branch at Galveston, Galveston, Texas; Division of Rheumatology, University of Texas Health Science Center at San Antonio, San Antonio, Texas; Jewish General Hospital and McGill University, Montreal, Quebec, Canada; Division of Rheumatology, Department of Medicine, David Geffen School of Medicine at UCLA, Los Angeles, California; Division of Rheumatic and Autoimmune Diseases, University of Minnesota, Minneapolis, Minnesota, USA; University of Michigan Scleroderma Program, Ann Arbor, Michigan, USA; Division of Rheumatology and Immunology, Medical University of South Carolina, Charleston, South Carolina, USA; and Rheumatology Section, Department of Medicine, Boston University School of Medicine, Boston, Massachusetts, USA.*

*Supported by the National Institute of Health (NIH) Center of Research Translation P50AR054144 (F.C. Arnett, M.D. Mayes); NIH Training grant 5T32-AR052283-03 (J.D. Reveille); NIH Family Registry and DNA Repository N01-AR0-2251 (M.D. Mayes); NIH-KL2RR024149-04 (S. Assassi); NIH-U01-AI090909-01 Studies of HLA Region Genomics in Systemic Sclerosis and Ankylosing Spondylitis (X. Zhou); the United States Army Medical Research and Material Command PR064251 Candidate Gene Polymorphisms in Scleroderma: Defining Genetic Susceptibility Factors (M.D. Mayes); University Clinic Research Center grants M01-RR00073 (UTMB) and M01-RR01346 (UT-HSC-SA); and NIH Clinical and Translational Sciences Award UL1-RR024148 and TL1 RR024147 from the National Center for Research Resources.*

*R. Sharif, MD, Postdoctoral Fellow, University of Texas Health Science Center at Houston; M.J. Fritzler, MD, PhD, Professor of Medicine, University of Calgary; M.D. Mayes, MD, MPH, Professor of Medicine, University of Texas Health Science Center at Houston; E.B. Gonzalez, MD, Professor of Medicine, University of Texas Medical Branch; T.A. McNearney, MD, Professor of Medicine, University of Texas Medical Branch (currently employed at Eli Lilly and Company, Indianapolis, IN); H. Draeger, MD, Assistant Professor of Medicine, University of Texas Health Science Center at San Antonio; M. Baron, MD, Professor of Medicine, Jewish General Hospital, Director, Canadian Scleroderma Research Group; D.E. Furst, MD, Professor of Medicine; D.K. Khanna, MD, MS, Associate Professor of Medicine, University of California at Los Angeles; D.J. del Junco, PhD, Director of Outcome Research Center for Translational Injury Research (CeTIR), University of Texas Health Science Center at Houston; J.A. Molitor, MD, PhD, Associate Professor of Medicine, University of Minnesota; E. Schiopu, MD, Assistant Professor of Medicine; K. Phillips, MD, PhD, Assistant Professor of Medicine, University of Michigan; J.R. Seibold, MD, Professor of Medicine, University of Michigan (currently at Scleroderma Research*

*Consultants, LLc, Avon, CT); R.M. Silver, MD, Professor of Medicine and Pediatrics, Medical University of South Carolina; R.W. Simms, MD, Professor of Medicine, Boston University; M. Perry, BSc, Senior Research Associate; C. Rojo, BSc; J. Charles, BSc, Senior Research Associate; X. Zhou, MD, Associate Professor of Medicine; S.K. Agarwal, MD, Associate Professor of Medicine; J.D. Reveille, MD, Professor of Medicine; S. Assassi, MD, MS, Assistant Professor of Medicine; F.C. Arnett, MD, Professor of Medicine, University of Texas Health Science Center at Houston.*

*Address correspondence to Dr. R. Sharif, University of Texas Health Science Center, 6431 Fannin Street, MSB 5.261, Houston, TX 77030, USA. E-mail: roozbeh.sharif@uth.tmc.edu*

*Accepted for publication March 11, 2011.*

African American (AA) patients with systemic sclerosis (SSc; scleroderma) are reported to have a worse overall prognosis than Caucasians, which might be explained by a younger age of disease onset, higher frequency of diffuse cutaneous involvement, more severe lung involvement, and younger age at onset of pulmonary artery hypertension (PAH)[1,2,3,4,5].

Anti-U3-RNP, or anti-fibrillarin antibody (AFA), is directed against a 35-kDa protein component of a nucleolar ribonucleoprotein called fibrillarin, which is an early marker for the formation site of nucleolus in dividing cells[6]. The frequency of AFA differs across ethnic groups, ranging from zero in a large cohort of Italian patients with SSc[7] to 50% in an African American SSc population[8]. The higher prevalence of AFA in the sera of AA patients with SSc has been noted in several studies[9,10,11,12,13].

Studies have shown that HLA-DRB1*08 and DQB1*03:01 are associated with AFA in African Americans[10,14]. Clinically, SSc patients with AFA have been reported to have younger ages of disease onset, higher frequency of diffuse cutaneous involvement, PAH, SSc-associated musculoskeletal and cardiac involvement, and lower frequency of arthritis[9,10,11,15,16,17]. However, there is a lack of large robust studies on the immunogenetic associations, clinical manifestations, and survival effect of AFA in AA patients with SSc.

We compared the HLA class II alleles in AA SSc patients with AFA with unaffected controls matched for ethnicity and sex and with SSc patients without AFA. We investigated the clinical features and survival effect of AFA in AA patients with SSc.

## MATERIALS AND METHODS

*Study population*. Between 1985 and 2010, 3033 patients with SSc were enrolled in the following cohorts: (1) the Genetics versus ENvironment In Scleroderma Outcomes Study (GENISOS)[3,5,18]; (2) the NIH/NIAMS Scleroderma Family Registry and DNA Repository[19]; and (3) the Division of Rheumatology, University of Texas Health Science Center at Houston (UTHSC-H)[10]. Patients were included if they met the American College of Rheumatology (formerly American Rheumatism Association) classification criteria for SSc[20] or had at least 3 of the 5 CREST features (calcinosis, Raynaud's phenomenon, esophageal dysmotility, sclerodactyly, telangiectasias)[21]. We included all AA patients from these cohorts (n = 278). Patients enrolled in more than one of the cohorts were identified and duplicate entries were omitted. We enrolled 328 unaffected AA controls to determine any HLA class II allele associations with AFA. The unaffected AA individuals were volunteers with no personal or family history of SSc or other autoimmune disease by screening questionnaire. All study subjects enrolled (SSc patients and unaffected controls) provided written informed consent, and the institutional review board of all participating institutions approved the study.

*Autoantibody profile and HLA class II allele genotyping*. All autoantibody determinations and HLA class II allele typing were conducted in the Division of Rheumatology at UTHSC-H and the Mitogen Advanced Diagnostics Laboratory, University of Calgary, Calgary, Canada. Antinuclear antibodies (ANA) and anticentromere antibodies were determined using indirect immunofluorescence with HEp-2 cells as substrate (Antibodies Inc., Davis, CA, USA). Passive immunodiffusion gels against calf thymus extract were used to examine sera for antitopoisomerase-I (ATA; Scl-70), anti-Ro/SSA, anti-La/SSB, and anti-U1-RNP autoantibodies (Inova Diagnostics, San Diego, CA, USA). Anti-RNA polymerase III (RNAP III) was detected by ELISA kits (MBL Co. Ltd., Nagoya, Japan) and AFA were determined by a line immunoassay at a serum dilution of 1:1000 using purified recombinant fibrillarin protein (Euroline-WB; Euroimmun, Lubeck, Germany) in patients who had a positive ANA in antinucleolar pattern on the indirect immunofluorescence.

As described[5,22], we genotyped HLA class II alleles (DRB1, DQA1, DQB1, and DPB1) on extracted and purified genomic DNA. Further, we examined the HLA class II allele-binding peptide using the ProPred MHC Class II Binding Peptide Prediction Server[23] in order to predict binding peptides of human fibrillarin protein. This prediction is based on quantitative matrices derived from the literature[23,24].

*Clinical manifestation*. Age, sex, disease type (categorized as limited or diffuse cutaneous involvement at time of enrollment[21]), disease duration (calculated from the onset of the first non-Raynaud's phenomenon symptom attributable to SSc), and modified Rodnan skin score (MRSS)[25] were recorded.

To assess the severity of individual organ system involvement, the Medsger severity indices[13,26] of 8 organ systems were measured: peripheral vessels, skin, joints/tendons, skeletal muscle, gastrointestinal (GI) tract, lung, heart, and kidney. However, these data were available only for the patients enrolled in the GENISOS cohort (n = 78). The presence of digital ulcers was determined based on the participating rheumatologist's clinical assessment. Arthritis was defined as presence of joint swelling and tenderness on examination not attributable to osteoarthritis, crystalline arthropathy, or trauma. A decrease in range of motion > 25% in at least one joint axis was defined as joint contracture. Dysphagia, diarrhea attributable to SSc, and history of SSc renal crisis were recorded. Electrocardiography and 2-dimensional echocardiography findings and/or presence of an auscultatory friction rub determined the presence of pericarditis or clinically significant pericardial effusion.

As described[18], pulmonary function tests were obtained at enrollment. Interstitial lung fibrosis, defined as chest radiograph showing fibrosis and/or forced vital capacity (FVC) < 75% of predicted value, was recorded.

For the purpose of our review, PAH was defined if the patient had (1) mean pulmonary artery pressure ≥ 25 mm Hg on right heart catheterization; (2) right ventricular systolic pressure ≥ 40 mm Hg on 2-dimensional echocardiography; or (3) if the ratio of FVC% predicted to diffusion capacity of carbon monoxide (DLCO)% predicted was ≥ 1.6. Serum creatine kinase (CK) levels were recorded and myositis was diagnosed if the patient had proximal muscle weakness with at least one of the following: elevated levels of CK, features of myositis on electromyography, and/or a characteristic muscle biopsy.

*Death search*. The vital status of patients was determined through the National Death Index (NDI) at the US Centers for Disease Control and Prevention, which provided data up until 2007. We then reviewed the US Social Security Death Index (SSDI) to update our results as of August 2010. SSDI is an online death search tool that provides fatality reports

based on death certificates and family confirmation. Patients not found on NDI or SSDI were assumed to be alive.

*Statistical analysis*. Homozygosity for alleles at each of the tested HLA loci was not suggestive of recessive inheritance, regardless of whether the referent comparison group comprised disease-free controls or AFA-negative cases. There were too few homozygous subjects to distinguish additive from dominant modes of inheritance, regardless of the referent. Therefore, a dominant mode of inheritance approach was used to compare the HLA association with AFA. Heterozygosity and homozygosity for a particular allele were both recoded as "1" in a binary (zero or 1) variable created for each specific HLA gene of interest. In other words, subjects negative for the gene on both their alleles for the particular HLA locus were coded "0" for the gene on the new binary variable. Bonferroni correction for multiple comparisons was performed for HLA allelic analyses.

Age, sex, disease type, and disease duration of AFA-positive and AFA-negative patients were evaluated utilizing chi-square and Student t test accordingly. SSc clinical manifestations might have changed over the disease course; therefore logistic regression was used to adjust for disease duration as a possible confounding factor in clinical features and to examine the independent effect of AFA.

We utilized Cox proportional hazards regression analysis to examine the association of AFA with survival. We investigated the potential association of relevant HLA class II with survival of the AA patients with SSc. Survival analysis was corrected for age at enrollment. Survival was calculated from the date of enrollment.

ATA and AFA are the 2 most common antinuclear antibodies among AA patients with SSc. We also compared the clinical features and survival of AA scleroderma patients with AFA (n = 50) to those with ATA (n = 61) for comparative analysis between more homogeneous groups.

All the statistical analyses were performed with SAS Version 9.2 (SAS Institute Inc., Cary, NC, USA) and Stata 11 (StataCorp., College Station, TX, USA). Hypothesis testing was 2-sided with a $p \leq 0.05$ significance level.

## RESULTS

*Study population, disease, and autoantibody characteristics*. All 278 AA scleroderma patients from the 3 cohorts were included in the study. The mean age (± SD) of patients at enrollment was 46.9 (13.9) years, and 237 (85.3%) were female. At enrollment, 171 (61.5%) AA patients with SSc were diagnosed with diffuse cutaneous involvement. Average disease duration (± SD) was 6.0 (6.5) years.

ANA on HEp-2 substrate were detected in 93.1% of AA SSc patients. ATA, RNAP-III, and AFA were present in 21.8%, 15.4%, and 18.5% of patients, respectively (Table 1).

*HLA class II allelic frequencies*. As illustrated in Table 2, comparison of HLA class II allelic frequencies of AFA-positive patients with 329 ethnically matched unaffected controls revealed the HLA-DRB1*08:04 allele more frequently in AFA-positive patients (47.6% vs 6.4%, respectively; OR 11.52, 95% CI 5.43, 24.40; corrected $p < 0.0001$). Two other alleles located on the same haplotype, DQA1*04:01 and DQB1*03:01, had similar patterns. However, the increased frequency of DQA1*04:01 was not statistically significant.

The frequency of HLA-DRB1*08:04 in AFA-positive patients also was higher in comparison to AA patients without AFA, even after correction for multiple comparisons (47.6% vs 14.9%; OR 5.21, 95% CI 2.44, 11.09; corrected $p = 0.0002$). Both HLA-DQA1*04:01 and DQB1*03:01 showed similar trends. However, neither of them remained significant after correction for multiple comparisons.

*Table 1*. Characteristics of the study population (n = 278).

| Characteristics | |
|---|---|
| Age, mean (SD), yrs | 46.9 (13.6) |
| Female, n (%) | 237 (85.3) |
| Diffuse cutaneous involvement, n (%) | 171 (61.5) |
| Disease duration, mean (± SD), yrs | 6.0 (6.5) |
| Modified Rodnan skin score, mean (± SD) | 17.3 (12.3) |
| Deceased patients, n (%) | 83 (29.7) |
| Survival time (from time of enrollment), mean (± SD), yrs | 5.8 (5.0) |
| Autoantibody profile, % | |
| Antinuclear | 93.1 |
| Anticentromere | 6.2 |
| Antitopoisomerase I | 21.8 |
| Antinucleolar | 44.5 |
| Antifibrillarin | 18.5 |
| RNA polymerase III | 15.4 |
| U1-ribonucleoprotein | 12.2 |
| Polymyositis/scleroderma | 3.2 |
| Ro/SSA60 | 9.3 |

HLA-DPB1*01:01 was also seen more frequently among AFA-positive AA patients compared to unaffected controls and SSc patients without AFA, whereas HLA-DRB1*11:01 seemed to be protective. HLA-DPB1*01:01 and HLA-DRB1*11:01 are not in linkage disequilibrium with HLA-DRB1*08:04. However, the association of these 2 alleles with AFA did not withstand correction for multiple comparisons. The frequencies of all relevant HLA class II alleles in AA SSc patients and unaffected individuals are illustrated in Table 3.

*HLA-DRB1*08:04 binding peptides*. Using virtual matrix for HLA-DRB1*08:04, at a threshold of 1% (the percentage best scoring natural peptides), we identified 4 binding peptides (FRSKLAAAI, FRGRGRGGG, IHIKPGAKV, and FVISIKANC) from the human fibrillarin protein that could serve as potential binding sites within the antigen-binding groove.

*Clinical features*. AA SSc patients with AFA were younger at disease onset ($p = 0.004$) but sex, disease type, and duration were not significantly different compared to AA SSc patients with AFA. Table 4 illustrates the comparison of clinical manifestations between AFA-positive and AFA-negative AA patients with SSc.

After adjustment for disease duration, AA SSc patients with AFA were 3.31-times more likely to have digital ulcers ($p = 0.014$). Diarrhea and pericarditis occurred more frequently in AFA-positive AA SSc patients (OR 4.84, $p = 0.019$; OR 2.45, $p = 0.092$, respectively) than AA patients without AFA. However, there were no differences between AFA-positive and AFA-negative AA SSc patients in MRSS, dysphagia, PAH, SSc-associated interstitial lung fibrosis, FVC and DLCO predicted values, SSc renal crisis, myositis

*Table 2*. Frequency of HLA class II alleles in antifibrillarin (AFA)-positive African American (AA) patients with SSc compared to ethnically matched AFA-negative patients and unaffected controls.

| HLA Class II Alleles | AFA-Positive, n = 50, % | AFA-negative, n = 221, % | Unaffected Controls, n = 329, % | AFA-positive AA Patients vs Unaffected AA Controls | | | AFA-positive vs negative AA Patients | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | OR (95% CI) | p | p* | OR (95% CI) | p | p* |
| DRB1*08:04 | 47.6 | 14.9 | 6.4 | 13.20 (6.24, 27.94) | < 0.001 | < 0.001 | 5.21 (2.44, 11.09) | < 0.001 | < 0.001 |
| DQA1*04:01 | 33.3 | 17.5 | 20.6 | 1.95 (0.97, 3.90) | 0.060 | NS | 2.37 (1.10, 5.09) | 0.026 | NS |
| DQB1*03:01 | 69.1 | 56.5 | 39.0 | 3.49 (1.75, 6.95) | < 0.001 | 0.005 | 1.72 (0.83, 3.56) | 0.153 | NS |
| DPB1*01:01 | 80 | 50.8 | 47.8 | 4.38 (1.08, 25.21) | 0.019 | NS | 4.12 (1.07, 15.89) | 0.041 | NS |
| DRB1*11:01 | 0 | 16.9 | 11.8 | NA | 0.019 | NS | NA | 0.004 | NS |

* Corrected p value. NS: not significant; NA: not applicable.

or muscle weakness, serum CK, joint contracture, or sicca symptoms.

AFA-positive patients had higher Medsger peripheral vascular severity index scores (regression coefficient [b] = 0.79, 95% CI 0.27, 1.30; p = 0.003), indicating more severe peripheral vascular involvement, and lower Medsger lung severity index (b = –0.82, 95% CI –1.50, –0.14; p = 0.019), indicating less severe lung involvement. The other Mesdger severity indices were not significantly different (Table 4).

*Survival analysis*. At the time of analysis, 30% of AFA-positive AA SSc patients and 29.5% of AFA-negative patients were deceased (Table 4). After correction for age at enrollment, AFA-positive patients did not have different survival compared to AFA-negative patients (hazard ratio = 0.79, p = 0.493). In addition, none of the relevant HLA class II was a predictor of mortality in AA patients with SSc (Table 5).

*AFA and ATA among AA patients with SSc*. Although age at onset of the first non-Raynaud's symptom was not statistically different between these 2 groups, the AA scleroderma patients with AFA had higher frequency of digital ulcers, and lower GI tract involvement, pericarditis, and Medsger peripheral vascular severity index scores (Table 6). AFA-positive patients had lower Medsger lung severity index, higher FVC and DLCO predicted values, and fewer cases of PAH. Despite less severe lung disease, after adjusting for age of disease onset, AFA-positive patients did not have better or worse survival compared to ATA-positive (Table 5).

## DISCUSSION

At a frequency of 18.5%, AFA is the second most common antinuclear antibody among AA patients with SSc (second to ATA). Our report represents the first study of the genetic associations, clinical manifestations, and infuence of AFA on survival in a large population of AA patients with SSc.

Distinct HLA class II allelic associations of SSc-specific autoantibodies in different ethnic groups have been described in several studies[5,10,14,27,28]. In a large sample of Caucasian patients, we previously reported that HLA-DRB1*13:02, DQB1*06:04/06:05 haplotype correlated with AFA[14]. In the current study, we did not observe a similar pattern among AA patients with AFA. Our results indicated that HLA-DRB1*08:04 is strongly associated with AFA in AA patients with SSc, compared to unaffected individuals or AFA-negative AA patients with SSc.

Previous studies investigated potential association of HLA-DRB1*08:04 with other rheumatic conditions such as systemic lupus erythematosus (SLE)[29] and rheumatoid arthritis (RA)[30]. Reveille, *et al*[29] detected no difference in frequency of HLA-DRB1*08:04 between 88 AA patients with SLE and 88 unaffected AA controls. Hughes, *et al*[30] reported no difference in frequency of HLA-DRB1*08:04 between 321 AA patients with RA and 564 unaffected individuals. Previously, we showed that HLA-DRB1*08:04 might be a susceptibility gene for SSc among AA[14]; whereas the results of the current study demonstrated that the higher frequency of HLA-DRB1*08:04 with SSc in AA patients is mainly driven by its strong association with AFA in this ethnic group. Through the Binding Peptide Prediction Server[23] for HLA-DRB1*0804, we identified 4 potential binding peptides from the human fibrillarin protein that could serve as potential binding sites within the antigen-binding groove. The large effect sizes (Table 2) and predicted binding peptides should prompt more studies to investigate potential causal and/or environmental relationships of these autoantibodies.

An animal model for induction of AFA has been studied extensively and may provide clues to an environmental trigger in humans with AFA-positive SSc. Certain mouse strains possessing specific H2 (the murine counterpart for HLA) haplotypes develop a non-SSc autoimmune disease and high-titer AFA following administration of mercuric chloride or silver nitrate[31,32,33,34]. Of note, one study of urinary

*Table 3*. Frequency of HLA class II alleles in AFA-positive AA patients with SSc compared to ethnically matched AFA-negative patients and unaffected controls. Data are percentages. The results of the prevalence of all HLA class II alleles with frequency ≥ 5% in at least one group are included.

| HLA Class II Alleles | AFA-positive, n = 50 | AFA-negative, n = 221 | Unaffected Controls, n = 329 | AFA-positive AA and Control AA | | AFA-positive and negative Patients | |
|---|---|---|---|---|---|---|---|
| | | | | OR (95% CI) | p | OR (95% CI) | p |
| DRB1 | | | | | | | |
| 01:01 | 0 | 2.7 | 7.7 | 0.14 (0.01, 2.33) | 0.06 | 0.38 (0.02, 7.16) | 0.28 |
| 01:02 | 9.5 | 4.1 | 4.9 | 2.05 (0.64, 6.56) | 0.22 | 2.49 (0.67, 9.28) | 0.16 |
| 03:01 | 9.5 | 16.2 | 15.7 | 0.57 (0.19, 1.66) | 0.30 | 0.54 (0.18, 1.66) | 0.28 |
| 03:02 | 7.1 | 5.4 | 14.3 | 0.46 (0.14, 1.56) | 0.20 | 1.35 (0.34, 5.32) | 0.67 |
| 04:01 | 0 | 4.7 | 5.9 | 0.18 (0.01, 3.08) | 0.11 | 0.22 (0.01, 3.97) | 0.15 |
| 07:01 | 11.9 | 6.8 | 14.6 | 0.79 (0.29, 2.12) | 0.64 | 1.86 (0.60, 5.79) | 0.28 |
| 08:04 | 47.6 | 14.9 | 7.3 | 11.51 (5.43–24.4) | < 0.0001 | 5.21 (2.44–11.09) | < 0.0001 |
| 11:01 | 0 | 16.9 | 11.8 | NA | 0.02 | NA | 0.004 |
| 13:02 | 21.4 | 16.2 | 10.4 | 2.34 (1.02, 5.35) | 0.04 | 1.41 (0.60, 3.32) | 0.43 |
| 14:01 | 0 | 0.7 | 6.3 | NA | 0.10 | NA | 0.59 |
| 15:03 | 14.3 | 27.0 | 16.7 | 0.83 (0.33, 2.08) | 0.69 | 0.45 (0.18, 1.15) | 0.09 |
| DQA1 | | | | | | | |
| 01:01 | 14.3 | 15.6 | 21.2 | 0.62 (0.25, 1.53) | 0.30 | 0.90 (0.34, 2.38) | 0.84 |
| 01:02 | 42.9 | 52.6 | 38.0 | 1.22 (0.64, 2.34) | 0.55 | 0.68 (0.34, 1.35) | 0.26 |
| 01:03 | 14.3 | 7.1 | 17.2 | 0.80 (0.32, 2.00) | 0.64 | 2.17 (0.75, 6.25) | 0.15 |
| 02:01 | 9.5 | 11.7 | 19.6 | 0.43 (0.15, 1.25) | 0.11 | 0.80 (0.25, 2.49) | 0.69 |
| 04:01 | 33.3 | 17.5 | 20.6 | 1.93 (0.96–3.87) | 0.06 | 2.35 (1.09–5.05) | 0.03 |
| 05:01 | 50.0 | 59.1 | 46.3 | 1.16 (0.61, 2.20) | 0.65 | 0.69 (0.35, 1.37) | 0.29 |
| DQB1 | | | | | | | |
| 02:01 | 11.9 | 18.2 | 16.2 | 0.70 (0.26, 1.87) | 0.48 | 0.61 (0.22, 1.69) | 0.34 |
| 02:02 | 11.9 | 13.6 | 20.1 | 0.54 (0.20, 1.42) | 0.20 | 0.86 (0.30, 2.42) | 0.77 |
| 03:01 | 69.1 | 56.5 | 39.0 | 3.49 (1.75–6.95) | 0.0002 | 1.72 (0.83–3.56) | 0.14 |
| 03:02 | 2.4 | 6.5 | 9.8 | 0.22 (0.03, 1.70) | 0.11 | 0.35 (0.04, 2.83) | 0.31 |
| 04:02 | 7.1 | 11.0 | 16.2 | 0.40 (0.12, 1.34) | 0.12 | 0.62 (0.17, 2.22) | 0.46 |
| 05:01 | 16.7 | 18.8 | 22.3 | 0.70 (0.30, 1.64) | 0.41 | 0.86 (0.35, 2.13) | 0.75 |
| 06:02 | 26.2 | 35.1 | 28.7 | 0.88 (0.43, 1.83) | 0.74 | 0.66 (0.31, 1.41) | 0.28 |
| 06:04 | 16.7 | 9.7 | 8.5 | 2.14 (0.87, 5.27) | 0.09 | 1.85 (0.70, 4.89) | 0.21 |
| 06:05 | 0 | 7.1 | 1.8 | NA | 0.38 | NA | 0.07 |
| DPB1 | | | | | | | |
| 01:01 | 80 | 50.8 | 47.8 | 4.38 (1.08, 25.21) | 0.02 | 4.12 (1.07, 15.89) | 0.04 |
| 02:01 | 26.7 | 23.1 | 18.9 | 1.56 (0.32, 5.94) | 0.48 | 1.21 (0.34, 4.37) | 0.77 |
| 03:01 | 6.7 | 16.9 | 14.4 | 0.42 (0.01, 3.20) | 0.41 | 0.35 (0.04, 2.95) | 0.32 |
| 04:01 | 20.0 | 27.7 | 16.2 | 1.29 (0.21, 5.49) | 0.71 | 0.65 (0.16, 2.59) | 0.54 |
| 04:02 | 20.0 | 15.4 | 26.1 | 0.71 (0.11, 2.89) | 0.61 | 1.38 (0.33, 5.76) | 0.66 |
| 17:01 | 6.7 | 20.0 | 18.0 | 0.33 (0.01, 2.40) | 0.27 | 0.29 (0.03, 2.38) | 0.22 |
| 18:01 | 0 | 6.2 | 2.4 | NA | 0.54 | NA | 0.96 |

NA: Not applicable.

mercury levels in SSc patients noted higher levels in those with AFA. However, this observation did not maintain statistical significance following corrections[35]. Interestingly, heavy metals have been noted to be highly concentrated in the nucleolus[36]. It was reported by Pollard, *et al*[37] that most if not all of the SSc-specific autoantigens were at some time during their life cycle localized to the nucleolus. Clearly, larger and more targeted studies of heavy metal and other environmental exposures are warranted in AFA-positive SSc patients, perhaps selected for the associated HLA class II alleles (DRB1*13:02 in Caucasians and DRB1*08:04 in AA) and AFA-negative SSc patients, as well as well matched healthy controls.

Confirming our previous findings[10], we showed that HLA-DQB1*03:01 had a higher frequency among AFA-positive AA patients compared to unaffected AA individuals. However, there was no difference between AFA-positive and negative AA patients with SSc.

AA patients with SSc are younger at SSc onset compared to other ethnic groups[1,2,5,28]. Moreover, other studies show that SSc patients with AFA have younger age of disease onset[2,12,13,16,17]. In support of these findings, we demonstrated that AA SSc patients with AFA had younger age of onset in comparison to AA patients without AFA.

The higher Medsger peripheral vascular severity index and prevalence of digital ulcers in AA patients with AFA compared to those without AFA are novel findings. These findings were also present when we compared AFA-positive and ATA-positive AA patients with SSc. Previous studies have shown higher rates of digital ulcers among AA patients

*Table 4*. Clinical manifestations of African American patients with AFA compared to those without AFA (adjusted for disease duration).

| Characteristic | AFA-positive, n = 50 | AFA-negative, n = 221 | OR (95% CI) | p |
|---|---|---|---|---|
| Age*, mean (SD), yrs | 41.7 (13.31) | 47.9 (13.25) | –6.29 (–10.59, –1.99)** | 0.004 |
| Female*, % | 86.0 | 84.6 | 0.89 (0.31, 2.24) | 0.806 |
| Cutaneous involvement*, diffuse, % | 36.0 | 39.5 | 0.86 (0.43, 1.69) | 0.642 |
| Disease duration at enrollment* | 5.34 (5.14) | 6.28 (6.85) | –0.93 (–3.51, 1.64)** | 0.475 |
| Deceased, % | 30.0 | 29.5 | 0.99 (0.47, 2.03) | 0.892 |
| Modified Rodnan skin score, mean (± SD) | 14.31 (7.62) | 17.83 (13.50) | –2.55 (–9.66, 4.54)** | 0.476 |
| Digital ulcer, % | 79.3 | 53.8 | 3.31 (1.27, 8.62) | 0.014 |
| Dysphagia, % | 60.0 | 54.8 | 1.24 (0.49, 3.19) | 0.619 |
| Diarrhea, % | 53.9 | 19.2 | 4.84 (1.29, 18.13) | 0.019 |
| Pericarditis, % | 28.2 | 12.8 | 2.45 (0.86, 6.93) | 0.092 |
| Pulmonary artery hypertension, % | 10.0 | 22.4 | 0.38 (0.09, 1.19) | 0.081 |
| SSc interstitial lung fibrosis, % | 28.2 | 47.3 | 0.64 (0.28, 1.52) | 0.322 |
| FVC % predicted, mean (± SD) | 77.8 (17.7) | 72.9 (23.2) | 6.08 (–4.54, 16.69)** | 0.259 |
| DLCO % predicted, mean (± SD) | 67.7 (17.2) | 57.9 (24.6) | 9.45 (–2.24, 21.14)** | 0.112 |
| SSc renal crisis, % | 10.5 | 8.4 | 1.28 (0.28, 4.54) | 0.921 |
| Myositis or muscle weakness, % | 30.0 | 40.6 | 1.05 (0.21, 5.28) | 0.953 |
| Elevated serum creatine kinase | 28.6 | 30.9 | 0.88 (0.31, 2.54) | 0.817 |
| Arthritis, % | 21.9 | 31.0 | 0.53 (0.13, 2.13) | 0.370 |
| Joint contracture | 22.9 | 21.4 | 1.44 (0.56, 3.72) | 0.447 |
| Sicca symptoms† | 20.0 | 28.1 | 0.64 (0.11, 3.69) | 0.621 |
| Medsger severity index, mean (± SD) | | | | |
| General | 0.6 | 0.5 | 0.06 (–0.53, 0.64)** | 0.838 |
| Peripheral vascular | 2.2 | 1.4 | 0.79 (0.27, 1.30)** | 0.003 |
| Skin | 1.5 | 1.7 | –0.27 (–0.84, 0.31)** | 0.361 |
| Joint | 0.8 | 0.9 | –0.09 (–0.92, 0.73)** | 0.813 |
| Muscle | 0.2 | 0.3 | –0.10 (–0.41, 0.21)** | 0.521 |
| GI tract | 0.6 | 0.6 | 0.02 (–0.43, 0.39)** | 0.935 |
| Lung | 1.1 | 1.9 | –0.82 (–1.50, –0.14)** | 0.019 |
| Heart | 0.5 | 0.3 | 0.16 (–0.27, 0.58)** | 0.457 |
| Kidney | 0.1 | 0.3 | 0.19 (–0.68, 0.29)** | 0.441 |

* These comparisons were not adjusted for disease duration. Student's t test and chi-square were utilized for comparisons, accordingly. ** Mean differences. † Two of 3 symptoms of dry mouth, dry eye, and/or enlarged parotid.

*Table 5*. Cox proportional hazards regression analysis of African American patients with systemic sclerosis (SSc).

| | Hazard Ratio (95% CI) | p |
|---|---|---|
| AFA | 0.80 (0.41, 1.53) | 0.493 |
| AFA vs ATA | 0.84 (0.42, 1.69) | 0.623 |
| HLA-DRB1*08:04 | 1.00 (0.55, 1.83) | 0.996 |
| HLA-DQA1*04:01 | 1.50 (0.84, 2.68) | 0.170 |
| HLA-DQB1*03:01 | 0.85 (0.51, 1.41) | 0.520 |
| HLA-DQB1*01:01 | 1.13 (0.49, 2.60) | 0.766 |
| HLA-DRB1*11:01 | 1.17 (0.57, 2.39) | 0.671 |

with SSc compared to Caucasians[2,4]. Higher frequencies of AFA in AA might contribute to this finding. Steen[12] reported higher frequency of digital ulcers in AFA-positive patients; however, these findings were not stratified for ethnic background.

In agreement with studies reporting more severe GI involvement in AFA-positive patients (regardless of ethnicity)[10,12], we observed a higher frequency of SSc-associated diarrhea in AFA-positive AA patients. The higher frequency of lower GI tract involvement was more significant when AFA-positive AA patients were compared to ATA-positive patients. It is possible that AFA-positive patients have more severe lower GI tract hypomotility and bacterial overgrowth that contribute to diarrhea.

Our results imply a less severe lung involvement among AFA-positive AA patients with SSc, as assessed by lower scores of the Medsger lung severity index. The comparison of AFA-positive and ATA-positive AA scleroderma patients further demonstrated less severe lung involvement (higher FVC and DLCO predicted values and lower Medsger lung severity index). In agreement with our findings, in an ethnically homogenous cohort of Japanese patients with SSc, AFA-positive patients had less severe lung involvement[17]. While data from several multiethnic cohorts suggested a

*Table 6*. Comparison of clinical manifestations of African American SSc patients with AFA to ATA-positive patients, adjusted for disease duration.

| | AFA-positive, n = 50 | ATA-positive, n = 61 | OR (95% CI) | p |
|---|---|---|---|---|
| Age*, mean (SD), yr | 41.7 (13.31) | 45.9 (11.81) | –4.17 (–9.16, 0.81)** | 0.099 |
| Female*, % | 86.0 | 74.6 | 2.09 (0.71, 6.66) | 0.139 |
| Cutaneous involvement*, diffuse, % | 36.0 | 42.4 | 0.77 (0.33, 1.78) | 0.498 |
| Disease duration at enrollment* | 5.34 (5.14) | 5.77 (6.83) | –0.43 (–3.32, 2.47)** | 0.769 |
| Deceased, % | 30.0 | 33.9 | 0.57 (0.19, 1.73) | 0.325 |
| MRSS, mean (± SD) | 14.31 (7.62) | 23.32 (13.19) | –7.38 (–15.35, 0.58)** | 0.078 |
| Digital ulcer, % | 79.3 | 59.1 | 2.68 (0.90, 8.01) | 0.078 |
| Dysphagia, % | 60.0 | 50.0 | 1.50 (0.43, 5.24) | 0.314 |
| Diarrhea, % | 53.9 | 0 | N/A | 0.002 |
| Pericarditis, % | 28.2 | 10.6 | 3.30 (0.92, 13.29) | 0.037 |
| Pulmonary artery hypertension, % | 10.0 | 34.0 | 0.28 (0.07, 1.13) | 0.075 |
| SSc interstitial lung fibrosis, % | 28.2 | 55.1 | 0.48 (0.18, 1.29) | 0.144 |
| FVC % predicted, mean (± SD) | 77.8 (17.7) | 66.2 (18.4) | 11.82 (1.37, 22.18)** | 0.030 |
| DLCO % predicted, mean (± SD) | 67.7 (17.2) | 47.5 (18.8) | 20.14 (0.35, 30.93)** | 0.004 |
| SSc renal crisis, % | 10.5 | 6.9 | 1.01 (0.16, 6.52) | 0.995 |
| Myositis or muscle weakness, % | 30.0 | 22.2 | 2.46 (0.21, 28.69) | 0.471 |
| Elevated serum creatine kinase | 28.6 | 30.4 | 0.94 (0.25, 3.47) | 0.925 |
| Arthritis, % | 21.9 | 13.0 | 1.12 (0.14, 8.68) | 0.914 |
| Joint contracture | 22.9 | 20.4 | 1.82 (0.57, 5.84) | 0.314 |
| Sicca symptoms† | 20.0 | 19.7 | 1.01 (0.06, 17.08) | 0.991 |
| Medsger severity index, mean (± SD) | | | | |
| General | 0.6 | 0.3 | 0.36 (–0.39, 1.12)** | 0.329 |
| Peripheral vascular | 2.2 | 1.1 | 1.15 (0.47, 1.82)** | 0.018 |
| Skin | 1.5 | 1.9 | –0.46 (–1.23, 0.32)** | 0.237 |
| Joint | 0.8 | 0.9 | –0.14 (–1.23, 0.93)** | 0.780 |
| Muscle | 0.2 | 0.3 | –0.08 (–0.44, 0.27)** | 0.633 |
| GI tract | 0.6 | 0.6 | 0.00 (–0.62, 0.62)** | 1.000 |
| Lung | 1.1 | 2.4 | –1.21 (–1.94, 0.48)** | 0.014 |
| Heart | 0.5 | 0.5 | 0.01 (–0.62, 0.63)** | 0.982 |
| Kidney | 0.1 | 0.4 | –0.34 (–1.02, 0.35)** | 0.317 |

* Comparisons were not adjusted for disease duration. Student t test and chi-square were utilized accordingly. ** Mean difference. † Two of 3 symptoms of dry mouth, dry eye, and/or enlarged parotid. MRSS: modified Rodnan skin score; GI: gastrointestinal.

higher frequency of isolated PAH and/or pulmonary fibrosis in the SSc patients with AFA[9,10,12,13,38], these comparisons were not adjusted for ethnicity or for other antibodies, i.e., ATA, as potential confounders. Therefore, the higher frequency of lung fibrosis and PAH might be due to a sizeable AA population in the AFA-positive group and a large number of Caucasian SSc patients in the AFA-negative group. More severe SSc-associated lung involvement in AA patients with SSc compared to other ethnic groups has been reported in several studies[2,18,28,39,40].

Based on our findings, AFA-positive AA patients with SSc have a higher prevalence of pericarditis, compared to AFA-negative as well as ATA-positive patients. This is in agreement with studies indicating higher frequency of cardiac involvement in AFA-positive[10,12] and AA patients with SSc[39].

Our study did not confirm reports of worse[13] or better[11] survival in AFA-positive patients with SSc. The poorer survival of AFA-positive patients in one report[13] might be attributable to the confounding or modifying effects of ethnicity in studies that are not stratified by ethnicity, as AA ethnicity is associated with AFA positivity as well as poorer survival[1,10,12,13].

This study has limitations. Although potentially important, data on heavy metal exposure were not collected. Medsger severity indices were available only for patients from the longitudinal GENISOS cohort. High-resolution computed tomography scans and echocardiography were not performed on all patients, which might have led to underreporting of pulmonary involvement; and despite being the largest genetic study reported to date in AA patients with SSc, the findings might be underpowered to detect more subtle HLA associations with AFA in the AA population.

Anti-fibrillarin antibody was the second most common antinuclear antibody in African Americans with SSc. Presence of AFA was strongly associated with the HLA-DRB1*08:04 in the AA patients with SSc. In addition,

AA SSc patients with AFA had a younger age of disease onset, higher frequency of digital ulcers and pericarditis, more severe lower GI involvement, and less severe pulmonary involvement. Future studies should focus on environmental factors, such as heavy metal exposure, that may influence the B cell response and the immunopathology of the disease.

## APPENDIX

List of study collaborators: The Canadian Scleroderma Research Group: Janet E. Pope, Janet Markland, David Robinson, Niall Jones, Nader Khalidi, Peter Docherty, Maysan Abu-Hakima, Sharon LeClercq, Evelyn Sutton, Douglas Smith, Jean-Pierre Mathieu, Alejandra Masetto, Elzbieta Kaminska, Sophie Ligier.

## REFERENCES


1. Mayes MD, Lacey JV Jr, Beebe-Dimmer J, Gillespie BW, Cooper B, Laing TJ, et al. Prevalence, incidence, survival, and disease characteristics of systemic sclerosis in a large US population. Arthritis Rheum 2003;48:2246-55.
2. Nietert PJ, Mitchell HC, Bolster MB, Shaftman SR, Tilley BC, Silver RM. Racial variation in clinical and immunological manifestations of systemic sclerosis. J Rheumatol 2006;33:263-8.
3. Assassi S, Del Junco D, Sutter K, McNearney TA, Reveille JD, Karnavas A, et al. Clinical and genetic factors predictive of mortality in early systemic sclerosis. Arthritis Rheum 2009;61:1403-11.
4. Beall AD, Nietert PJ, Taylor MH, Mitchell HC, Shaftman SR, Silver RM, et al. Ethnic disparities among patients with pulmonary hypertension associated with systemic sclerosis. J Rheumatol 2007;34:1277-82.
5. Reveille JD, Fischbach M, McNearney T, Friedman AW, Aguilar MB, Lisse J, et al. Systemic sclerosis in 3 US ethnic groups: a comparison of clinical, sociodemographic, serologic, and immunogenetic determinants. Semin Arthritis Rheum 2001; 30:332-46.
6. Ochs RL, Lischwe MA, Spohn WH, Busch H. Fibrillarin: a new protein of the nucleolus identified by autoimmune sera. Biol Cell 1985;54:123-33.
7. Ceribelli A, Cavazzana I, Franceschini F, Airo P, Tincani A, Cattaneo R, et al. Anti-Th/To are common antinucleolar autoantibodies in Italian patients with scleroderma. J Rheumatol 2010;37:2071-5.
8. Kuwana M, Okano Y, Kaburaki J, Tojo T, Medsger TA Jr. Racial differences in the distribution of systemic sclerosis-related serum antinuclear antibodies. Arthritis Rheum 1994;37:902-6.
9. Okano Y, Steen VD, Medsger TA Jr. Autoantibody to U3 nucleolar ribonucleoprotein (fibrillarin) in patients with systemic sclerosis. Arthritis Rheum 1992;35:95-100.
10. Arnett FC, Reveille JD, Goldstein R, Pollard KM, Leaird K, Smith EA, et al. Autoantibodies to fibrillarin in systemic sclerosis (scleroderma). An immunogenetic, serologic, and clinical analysis. Arthritis Rheum 1996;39:1151-60.
11. Tormey VJ, Bunn CC, Denton CP, Black CM. Anti-fibrillarin antibodies in systemic sclerosis. Rheumatology 2001;40:1157-62.
12. Steen VD. Autoantibodies in systemic sclerosis. Semin Arthritis Rheum 2005;35:35-42.
13. Aggarwal R, Lucas M, Fertig N, Oddis CV, Medsger TA Jr. Anti-U3 RNP autoantibodies in systemic sclerosis. Arthritis Rheum 2009;60:1112-8.
14. Arnett FC, Gourh P, Shete S, Ahn CW, Honey RE, Agarwal SK, et al. Major histocompatibility complex (MHC) class II alleles, haplotypes and epitopes which confer susceptibility or protection in systemic sclerosis: analyses in 1300 Caucasian, African-American and Hispanic cases and 1000 controls. Ann Rheum Dis 2010;69:822-7.
15. Satoh M, Akizuki M, Kuwana M, Mimori T, Yamagata H, Yoshida S, et al. Genetic and immunological differences between Japanese patients with diffuse scleroderma and limited scleroderma. J Rheumatol 1994;21:111-4.
16. Reimer G, Steen VD, Penning CA, Medsger TA Jr, Tan EM. Correlates between autoantibodies to nucleolar antigens and clinical features in patients with systemic sclerosis (scleroderma). Arthritis Rheum 1988;31:525-32.
17. Kuwana M, Kaburaki J, Okano Y, Tojo T, Homma M. Clinical and prognostic associations based on serum antinuclear antibodies in Japanese patients with systemic sclerosis. Arthritis Rheum 1994;37:75-83.
18. Assassi S, Sharif R, Lasky RE, McNearney TA, Estrada YMR, Draeger H, et al. Predictors of interstitial lung disease in early systemic sclerosis: a prospective longitudinal study of the GENISOS cohort. Arthritis Res Ther 2010;12:R166.
19. Mayes MD. The establishment and utility of a population-based registry to understand the epidemiology of systemic sclerosis. Curr Rheumatol Rep 2000;2:512-6.
20. Preliminary criteria for the classification of systemic sclerosis (scleroderma). Subcommittee for scleroderma criteria of the American Rheumatism Association Diagnostic and Therapeutic Criteria Committee. Arthritis Rheum 1980;23:581-90.
21. Leroy EC, Black C, Fleischmajer R, Jablonska S, Krieg T, Medsger TA Jr, et al. Scleroderma (systemic sclerosis): classification, subsets and pathogenesis. J Rheumatol 1988;15:202-5.
22. Olsen ML, Arnett FC, Reveille JD. Contrasting molecular patterns of MHC class II alleles associated with the anti-Sm and anti-RNP precipitin autoantibodies in systemic lupus erythematosus. Arthritis Rheum 1993;36:94-104.
23. Singh H, Raghava GP. ProPred: prediction of HLA-DR binding sites. Bioinformatics 2001;17:1236-7.
24. Sturniolo T, Bono E, Ding J, Raddrizzani L, Tuereci O, Sahin U, et al. Generation of tissue-specific and promiscuous HLA ligand databases using DNA microarrays and virtual HLA class II matrices. Nat Biotechnol 1999;17:555-61.
25. Clements PJ, Lachenbruch PA, Seibold JR, Zee B, Steen VD, Brennan P, et al. Skin thickness score in systemic sclerosis: an assessment of interobserver variability in 3 independent studies. J Rheumatol 1993;20:1892-6.
26. Steen VD, Medsger TA Jr, Rodnan GP. D-Penicillamine therapy in progressive systemic sclerosis (scleroderma): a retrospective analysis. Ann Intern Med 1982;97:652-9.
27. Reveille JD, Durban E, Goldstein R, Moreda R, Arnett FC. Racial differences in the frequencies of scleroderma-related autoantibodies. Arthritis Rheum 1992;35:216-8.
28. Reveille JD. Ethnicity and race and systemic sclerosis: how it affects susceptibility, severity, antibody genetics, and clinical manifestations. Curr Rheumatol Rep 2003;5:160-7.
29. Reveille JD, Moulds JM, Ahn C, Friedman AW, Baethge B, Roseman J, et al. Systemic lupus erythematosus in three ethnic groups: I. The effects of HLA class II, C4, and CR1 alleles, socioeconomic factors, and ethnicity at disease onset. LUMINA Study Group. Lupus in Minority Populations, Nature versus Nurture. Arthritis Rheum 1998;41:1161-72.
30. Hughes LB, Morrison D, Kelley JM, Padilla MA, Vaughan LK, Westfall AO, et al. The HLA-DRB1 shared epitope is associated with susceptibility to rheumatoid arthritis in African Americans through European genetic admixture. Arthritis Rheum 2008; 58:349-58.
31. Hultman P, Enestrom S, Pollard KM, Tan EM. Anti-fibrillarin autoantibodies in mercury-treated mice. Clin Exp Immunol

1989;78:470-7.
32. Chen M, Rockel T, Steinweger G, Hemmerich P, Risch J, von Mikecz A. Subcellular recruitment of fibrillarin to nucleoplasmic proteasomes: implications for processing of a nucleolar autoantigen. Mol Biol Cell 2002;13:3576-87.
33. Havarinasab S, Pollard KM, Hultman P. Gold- and silver-induced murine autoimmunity — requirement for cytokines and CD28 in murine heavy metal-induced autoimmunity. Clin Exp Immunol 2009;155:567-76.
34. Pollard KM, Hultman P, Kono DH. Toxicology of autoimmune diseases. Chem Res Toxicol 2010;23:455-66.
35. Arnett FC, Fritzler MJ, Ahn C, Holian A. Urinary mercury levels in patients with autoantibodies to U3-RNP (fibrillarin). J Rheumatol 2000;27:405-10.
36. Rosen A, Casciola-Rosen L, Wigley F. Role of metal-catalyzed oxidation reactions in the early pathogenesis of scleroderma. Curr Opin Rheumatol 1997;9:538-43.
37. Pollard KM, Reimer G, Tan EM. Autoantibodies in scleroderma. Clin Exp Rheumatol 1989;7 Suppl 3:S57-S62.
38. Sacks DG, Okano Y, Steen VD, Curtiss E, Shapiro LS, Medsger TA Jr. Isolated pulmonary hypertension in systemic sclerosis with diffuse cutaneous involvement: association with serum anti-U3RNP antibody. J Rheumatol 1996;23:639-42.
39. Laing TJ, Gillespie BW, Toth MB, Mayes MD, Gallavan RH Jr, Burns CJ, et al. Racial differences in scleroderma among women in Michigan. Arthritis Rheum 1997;40:734-42.
40. McNearney TA, Reveille JD, Fischbach M, Friedman AW, Lisse JR, Goel N, et al. Pulmonary involvement in systemic sclerosis: associations with genetic, serologic, sociodemographic, and behavioral factors. Arthritis Rheum 2007;57:318-26.

European Journal of Human Genetics (2010) 18, 111–117

www.nature.com/ejhg

ARTICLE

# Gene and pathway-based second-wave analysis of genome-wide association studies

Gang Peng[1], Li Luo[2], Hoicheong Siu[1], Yun Zhu[1], Pengfei Hu[1], Shengjun Hong[1], Jinying Zhao[3], Xiaodong Zhou[4], John D Reveille[4], Li Jin[1], Christopher I Amos[5] and Momiao Xiong*[,2]

**Despite the great success of genome-wide association studies (GWAS) in identification of the common genetic variants associated with complex diseases, the current GWAS have focused on single-SNP analysis. However, single-SNP analysis often identifies only a few of the most significant SNPs that account for a small proportion of the genetic variants and offers only a limited understanding of complex diseases. To overcome these limitations, we propose gene and pathway-based association analysis as a new paradigm for GWAS. As a proof of concept, we performed a comprehensive gene and pathway-based association analysis of 13 published GWAS. Our results showed that the proposed new paradigm for GWAS not only identified the genes that include significant SNPs found by single-SNP analysis, but also detected new genes in which each single SNP conferred a small disease risk; however, their joint actions were implicated in the development of diseases. The results also showed that the new paradigm for GWAS was able to identify biologically meaningful pathways associated with the diseases, which were confirmed by a gene-set-rich analysis using gene expression data.**
*European Journal of Human Genetics* (2010) **18,** 111–117; doi:10.1038/ejhg.2009.115; published online 8 July 2009

**Keywords:** genome-wide association studies; gene and pathway-based analysis; complex diseases; combining *P*-values; gene-set enrichment analysis

## INTRODUCTION

Genome-wide association studies (GWAS) are emerging as a major tool to identify disease susceptibility loci and have been successful in detecting the association of a number of SNPs with complex diseases.[1–12] However, testing only for association of a single SNP is insufficient to dissect the complex genetic structure of common diseases. Extracting biological insight from GWAS and understanding the principles underlying the complex phenomena that take place on various biological pathways remain a major challenge. The common approach of GWAS is to select dozens of the most significant SNPs in the list for further investigations. This approach, which takes only SNPs as basic units of association analysis, has a few serious limitations. First, a single SNP showing a significant association with complex diseases typically has only mild effects.[13] The common disease often arises from the joint action of multiple loci within a gene or the joint action of multiple genes within a pathway. If we consider only the most significant SNPs, the genetic variants that jointly have significant risk effects but individually make only a small contribution will be missed. Second, locus heterogeneity, which implies that alleles at different loci cause diseases in different populations, will increase difficulty in the replication of association of a single marker.[14] A gene, particularly a pathway, consists of a group of interacting components that act in concert to perform specific biological tasks. Replication of association finding at the gene level or pathway level is much easier than replication at the SNP level. Third, attempting to understand and interpret a number of significant SNPs without any unifying biological theme can be challenging and demanding. SNPs and genes carry out their functions through intricate pathways of reactions and interactions. The function of many SNPs may not be well characterized, but the function of genes and particular pathways have been much better investigated. Therefore, the gene and pathway-based association analysis allows us to gain insight into the functional basis of the association and facilitates to unravel the mechanisms of complex diseases.

To meet the conceptual and technical challenges raised by GWAS and to take full advantage of the wide opportunities provided by GWAS, the gene and pathway-based association analysis can be used as a complementary approach to the genome-wide search association of a single SNP with a disease . The gene and pathway-based association analysis considers a gene or a pathway as the basic unit of analysis. Gene and pathway-based GWAS aim to study simultaneously the association of a group of genetic variants in the same biological pathway,[14–16] which can help us to holistically unravel the complex genetic structure of common diseases in order to gain insight into the biological processes and disease mechanisms.[17]

Gene and pathway-based GWAS can be performed by extension of a gene-set enrichment analysis for gene expression data,[18] to genome-wide association studies. However, a simple application of gene-set analysis methods for gene expression data to GWAS may not work very well. The key difference between the gene expression data and

[1]School of Life Science, Fudan University, Shanghai, China; [2]Human Genetics Center, University of Texas School of Public Health, Houston, TX, USA; [3]Department of Medicine, Emory University of School of Medicine, Atlanta, GA, USA; [4]Division of Rheumatology, Medical School, University of Texas Health Science Center at Houston, Houston, TX, USA; [5]Department of Epidemiology, University of Texas, MD Anderson Cancer Center, Houston, TX, USA
*Correspondence: Professor M Xiong, Human Genetics Center, University of Texas–Houston, Human Genetics Center, UT Houston School of Public Health, P.O. Box 20334, Houston, TX, 77225, USA, Tel: +1 713 500 9894, Fax: +1 713 500 0900; E-mail: Momiao.Xiong@uth.tmc.edu
Received 14 August 2008; revised 7 April 2009; accepted 26 May 2009; published online 8 July 2009

SNP data is that in expression data analysis each gene is represented by one value of expression level of the gene, but in GWAS each gene is represented by a varied number of SNPs. The challenge facing us is how to represent a gene.[19,20] One promising approach is to combine *P*-values for correlated SNPs into an overall significance level to represent a gene and to combine *P*-values for the genes into an overall significance level to investigate the association of a pathway with the disease.[21]

## MATERIALS AND METHODS

### Gene-based association analysis

Statistical analyses for testing the association of a gene with a disease were conducted on the basis of the combination of *P*-values of the SNPs in the gene[14]. We assume that the *P*-values $P_i$ are independent and uniformly distributed under their null hypotheses although the independence assumption may be violated because of linkage disequilibrium among SNPs in the gene. Several methods were used to combine independent *P*-values. A general framework for combining independent *P*-values is as follows. Let $P_i$ be the *P*-value for the corresponding statistic $T_i$ with G distribution to test the *i*-th marker $M_i$. Let $H$ be a continuous monotonic function. A transformation of the *P*-value is defined as $Z_i = H^{-1}(1 - P_i)$

### Fisher's combination test

The full combination methods are to combine *P*-values of all SNPs within the gene. The statistic for combining $K$ independent *P*-values or for combining information from $K$ SNPs is usually given by

$$Z_F = -2\sum_{i=1}^{K} \log P_i$$

which follows a $\chi^2_{(2K)}$ distribution.[21]

### Sidak's combination test (the best SNP)

If we consider only the best SNP in the gene, then the statistic is defined as $Z_B = P_{(1)}$, which is distributed as $P(Z_B \leq w) = 1-(1-w)^K$. This statistic is often referred to as Sidak's correction.

### Simes' combination test

Let *P*-values be ordered as $P_{(1)} \leqslant P_{(2)} \leqslant \ldots \leqslant P_{(k)}$. The *P*-value is calculated as

$$P_S = \min_i \left\{ \frac{kP_{(i)}}{i} \right\}$$

### The FDR method

Let $\pi$ be the proportion of tests with a true null hypothesis and $F(\alpha)$ be the expected proportion of tests yielding a *P*-value less than or equal to $\alpha$, $V(\alpha)$ be the expected proportion of tests giving a false positive result with significance level $\alpha$.

Suppose that there are $d$ distinct *P*-values among $p = \{p_1, \ldots, p_k\}$. Let $\tilde{p}_1 < \tilde{p}_2 < \ldots < \tilde{p}_d$ . Let $m_j$ be the number of *P*-values among $P$ that are equal to $\tilde{p}_j$.

Then, $\tilde{F}(\alpha) = \frac{1}{k}\sum_{j=1}^{d} I(\tilde{p}_j \leq \alpha) m_j$,where $I$ is an indicator function. For a two-sided test define $\pi = \min(1, 2\bar{p})$, and for a one-sided test ($\chi^2$-test, trend test) define $\pi = \min(1, 2\bar{a})$, where $\bar{p} = \frac{1}{k}\sum_{i=1}^{k} pi$, $\bar{a} = \frac{1}{k}\sum_{i=1}^{k} a_i$, $a_i = 2\min(p_i, 1 - p_i)$

Then, $v(\alpha)$is estimated by $v(\alpha) = \pi\alpha$. Define $t(i) = \frac{v(p_{(i)})}{F(p_{(i)})}$ and $q(i) = \min_{j \geq i}\{t_{(j)}\}$,

$q_{(1)} \leqslant q_{(2)} \leqslant \ldots \leqslant q_{(m)}$ are the ordered false discovery rates. We also take $q_{(1)} = \min\{t_{(j)}\}$ as the false discovery rate for the gene or pathway.[19]

### Pathway-based association analysis

Consider $m$ genes in a pathway. Assume that the *P*-value for each gene is calculated using one of the methods of combining independent *P*-values mentioned in the previous section. The methods for testing the association of a pathway with the disease are given below.

### Hypergeometric test (Fisher's exact test)

Fisher's exact test is performed to search for an overrepresentation of significantly associated genes among all the genes in the pathway. We assume that the total number of genes that are of interest is $N$. Let $S$ be the number of genes that are significantly associated with the disease (*P*-value $\leq 0.05$, calculated by Fisher's combination test) and $m$ be the number of genes in the pathway. Let $k$ be the number of significantly associated genes in the pathway. The *P*-value of observing $k$-significant genes in the pathway is calculated by

$$P = 1 - \sum_{i=0}^{K} \frac{\binom{S}{i}\binom{N-S}{m-i}}{\binom{N}{m}}$$

### Sidak's method

Both *P*-values for testing the association of the gene and the pathway are calculated by Sidak's method, which is described in the previous section.

### Simes' method

Both *P*-values for testing the association of the gene and the pathway are calculated by Simes' method that is described in the previous section.

### Simes/FDR method

The *P*-value for testing the association of the gene is calculated by Simes' method and the *P*-value for testing the association of the pathway is calculated by the FDR method.

## RESULTS

To investigate what should be the basic units for genome-wide association studies and to illustrate how to perform the gene and pathway-based genome-wide association analysis, we examine the 13 published GWAS (Supplementary Table 1), in which WTCCC represents the Wellcome Trust Case Control Consortium, NARAC, the North American Rheumatoid Arthritis Consortium, EIRA, the Swedish Epidemiological Investigation of Rheumatoid Arthritis, DGI, the Diabetes Genetics Initiative, AREDS, The Age-Related Eye Disease Study, CORIELL, Coriell Institute for Medical Research, and 10 diseases: bipolar disorder (BD), coronary artery disease (CAD), Crohn's disease (CD), hypertension (HT), rheumatoid arthritis (RA), type I diabetes (T1D), type II diabetes (T2D), Parkinson's disease (PD), age-related eye disease (AREDS) and Amyotrophic lateral sclerosis (ALS). As only *P*-values for testing the association of a single SNP (but not individual genotypes) were publically accessible, we used the statistical methods for combining independent *P*-values to perform gene and pathway-based GWAS (see Materials and methods). The methods for combining dependent *P*-values require individual genotype information and cannot be applied here. The number of typed cases and controls, the number of typed SNPs and genes, and *P*-values for ensuring genome-wide significance using Bonferroni correction for each study are listed in Supplementary Table 1.

The procedure for gene and pathway-based GWAS consists of two steps. The first step is to combine a set of *P*-values for SNPs in a gene, which is obtained from GWAS of a single SNP, into an overall significance level of the gene. The second step is to combine a set of *P*-values for genes in a pathway into an overall *P*-value for the pathway. To combine *P*-values, one typically assumes that the *P*-values are independent and uniformly distributed under the null hypothesis. In this report, four combination tests: Fisher's combination test, Sidak's combination test, Simes' combination test and a test based on false discovery rate, were used (see Materials and methods). As the SNPs within a gene may be in linkage disequilibrium, *P*-values of SNPs from the same gene are often not independent and hence

**Table 1 Number of replicated or shared SNPs and genes**

| *Study 1* | *Study 2* | *Number of replicated or shared SNPs* | *Number of replicated or shared SNPs which are not located in significant genes* | *Number of replicated or shared genes* |
|---|---|---|---|---|
| *(a) Fisher's method* | | | | |
| RA (WTCCC) | RA(NARAC and EIRA) | 28 | 0 | 42 |
| T2D (WTCCC) | T2D (DGI) | 0 | 0 | 7 |
| PD(CORIELL) | PD(NCBI) | 4 | 4 | 82 |
| WTCCC | | | | |
| CAD+HT+T2D | | 0 | 0 | 6 |
| RA+T1D | | 29 | 0 | 57 |
| CD+RA+T1D | | 0 | 0 | 5 |
| *(b) FDR Method* | | | | |
| RA (WTCCC) | RA(NARAC and EIRA) | 28 | 0 | 36 |
| T2D (WTCCC) | T2D (DGI) | 0 | 0 | 0 |
| PA(CORIELL) | PA(NCBI) | 4 | 2 | 4 |
| WTCCC | | | | |
| CAD+HT+T2D | | 0 | 0 | 0 |
| RA+T1D | | 29 | 0 | 35 |
| CD+RA+T1D | | 0 | 0 | 0 |

independent assumption of combining *P*-values is violated. We used methods for combining independent *P*-values for the following reasons. First, the methods for combining dependent *P*-values require the data of individual genotypes. However, in many cases, individual genotypes cannot be publically accessed. Second, errors that arise from violation of independent assumptions are not very high. (We will present the results of comparison of methods combining independent *P*-values and those combining dependent *P*-values elsewhere.) Third, Q–Q plots for the four combining tests (Supplementary Figure 1) showed that the observed distribution of *P*-values of the combining tests (except for Fisher's combination test) matches that expected for the majority of the data, but begins to depart from the null at $3.15 \times 10^{-6}$ (gene) and $10^{-4}$ (pathway).

We obtained the combined *P*-values for each gene. Supplementary Table 2a and 2b summarizes the total number of significant genes, significant SNPs and significant SNPs that belong to insignificant genes. The numbers of replicated SNPs and genes in the different studies, or the numbers of significant SNPs and genes shared by several diseases, are shown in Table 1. In Supplementary Tables S3–S15 we have listed all significant genes with *P*-values $\leq 3.15\times10^{-6}$, which were calculated by the Fisher's combination test or by the test based on the false discovery rate (FDR) for 13 studies. In these tables we also included the number of typed SNPs within each significant gene and *P*-value of the most significant SNP in the gene. Supplementary Tables S16–S18 list the significant SNPs and genes for PA, RA and T2D diseases shared by two independent studies. Three remarkable features emerge from these tables. First, these tables show that except for the diseases RA and T1D, the number of significant SNPs in each study is very small, but the number of significant genes is quite large. From these tables we can find that the large proportion of significant genes even contains no single significant SNP. For example, in the T2D study (WTCCC), the *P*-values of the best SNPs in the genes PPARG, JAZF1, TSPAN8 and THADA were 0.001205, 0.001681, 0.0000156 , and 0.01080, respectively, but the overall *P*-values of these genes were $2.87\times10^{-5}$, $8.58\times10^{-7}$, $3.17\times10^{-13}$, and $1.80\times10^{-5}$, respectively. Although an initial single SNP analysis did not find any significant SNPs in these genes, a recent meta-analysis[22] showed that the *P*-values of the best SNPs in these genes were $2.00\times10^{-7}$, $5.00\times10^{-14}$, $1.10\times10^{-9}$, and $1.10\times10^{-9}$, respectively. This shows that the results of the gene-based association analysis were consistent with the results of meta-analysis. If we conduct only the single-SNP association analysis, these significant genes might be missed because of the low power of small sample sizes in the initial GWAS. Second, replication of association findings at gene level in additional independent samples is much easier than that at SNP level. We examined association studies of three diseases: T2D, PA, and RA, each with two independent studies. For T2D, no SNPs were replicated in two independent studies (WTCCC and DGI) after correction for multiple tests by the Bonferroni method. However, seven genes, including genes TCF7L2 (transcription factor 7-like 2) and CDKAL1 (CDK5 regulatory subunit associated protein 1-like 1), were replicated (Supplementary Table S17). The gene TCF7L2, which has a marked effect on type II diabetes, had a widely replicated association in several studies [2,23]. In single-SNP association analysis, although a strong association of CDKAL1 was reported from WTCCC ($P=1.02\times10^{-6}$) and WTCCC/UKT2D[2,3] ($P=10^{-8}$), the original scan and follow-up replication samples from DGI only support nominal association ($P=0.0024$). In gene-based analysis, a strong association of CDKAL1 was observed from WTCCC ($P<10^{-20}$) and DGI ($P=1.84\times10^{-6}$) (Supplementary Table S17). To explain why replication of significant genes in independent samples is much easier than replication of significant SNPs, we have listed all SNPs with *P*-values $<0.05$ for the genes in Table 2. Table 2 shows that although a few single SNPs in the genes CDKAL1, TTLL5 and BTBD16 showed significant association in the WTCCC study or DGI study, the joint effects of multiple SNPs with very mild effects led to three genes being strongly associated with the diseases in both studies. Third, gene-based association analysis can more effectively identify the common genes that are shared within a disease group than single-SNP association analysis. Although there is considerable heterogeneity among complex diseases, many diseases share common phenotypes, forming a group of diseases. In the studies that we examined here, CD+RA+T1D are autoimmune diseases, and CAD+HT+T2D have metabolic and cardiovascular phenotypes in common. GWAS offers us an opportunity to reveal the genetic variants that confer a risk of more than one disease. Supplementary Table 19 summarizes the shared genes within the disease group based on the best SNP within the gene. In other words, a gene is shared within a disease group if at least one significant

**Table 2 Overall *P*-values of the genes CDKAL1, TTLL5 and BTBD16 and their SNPs with *P*-values less than 0.05 in WTCCC and DGI studies**

| *WTCCC* | | | | *DGI* | | | |
|---|---|---|---|---|---|---|---|
| *Gene* | *P-value* | *Gene* | *P-value* | *Gene* | *P-value* | *Gene* | *P-value* |
| *CDKAL1* | *< 1.0E-20* | *TTLL5* | *3.0E-15* | *CDKAL1* | *2.0E-6* | *BTBD16* | *1.0E-6* |
| *No of SNPs* | *126* | *No of SNPs* | *25* | *No of SNPs* | *114* | *No of SNPs* | *30* |
| *SNP* | *P-value* | *SNP* | *P-value* | *SNP* | *P-value* | *SNP* | *P-value* |
| rs714831 | 0.0022 | rs760233 | 0.0093 | rs714830 | 0.0135 | rs1885512 | 0.0183 |
| rs2294809 | 0.037 | rs1158282 | 0.0206 | rs736425 | 0.0208 | rs2273796 | 0.0086 |
| rs2328529 | 0.0011 | rs2302592 | 0.0465 | rs1548145 | 0.0117 | rs7078328 | 0.0165 |
| rs2328549 | 0.0001 | rs2303345 | 0.0458 | rs2305955 | 0.0394 | rs7098436 | 0.0098 |
| rs2328573 | 0.0183 | rs2359866 | 0.0267 | rs2820001 | 0.0188 | rs10510107 | 0.0165 |
| rs2819999 | 0.0246 | rs2359983 | 0.0177 | rs6905567 | 0.0354 | rs10788281 | 0.0167 |
| rs4236002 | 0.0054 | rs4903350 | 0.0273 | rs6926388 | 0.0237 | rs11200528 | 0.0132 |
| rs4291090 | 0.0163 | rs4903359 | 0.0089 | rs6927356 | 0.0478 | rs11200537 | 0.0351 |
| rs4413596 | 0.032 | rs6574258 | 0.0092 | rs6938184 | 0.0183 | | |
| rs4527692 | 0.0254 | rs7156551 | 0.0356 | rs7747752 | 0.0468 | | |
| rs6456368 | 2.0E-05 | rs8015242 | 0.0441 | rs7754840 | 0.0075 | | |
| rs6908425 | 0.0074 | rs8020986 | 0.0396 | rs7767391 | 0.0365 | | |
| rs7739578 | 0.0064 | rs9323619 | 0.0178 | rs9460546 | 0.0057 | | |
| rs7739596 | 0.0076 | rs10131117 | 0.0053 | rs9465871 | 0.0445 | | |
| rs7741604 | 0.0198 | rs10143790 | 0.0353 | rs10484632 | 0.0122 | | |
| rs7747752 | 0.0018 | rs11621464 | 0.0394 | rs10946398 | 0.0059 | | |
| rs7752602 | 0.0351 | rs11621718 | 0.0129 | rs11970425 | 0.0375 | | |
| rs7754840 | 4.5E–05 | rs12887886 | 0.0427 | rs16884481 | 0.0073 | | |
| rs7763304 | 0.0067 | Gene | *P*-value | Gene | *P*-value | | |
| rs7766346 | 0.0271 | BTBD16 | 5.0E-08 | TTLL5 | 4.0E-07 | | |
| rs7767391 | 5.5E–06 | No of SNPs | 31 | No of SNPs | 21 | | |
| rs9348440 | 8.5E–05 | SNP | *P*-value | SNP | *P*-value | | |
| rs9350257 | 0.0427 | rs1022782 | 0.0017 | rs760233 | 0.0316 | | |
| rs9358395 | 0.0071 | rs4237539 | 0.0021 | rs4903359 | 0.0268 | | |
| rs9366357 | 0.0057 | rs4317918 | 0.0027 | rs6574258 | 0.0129 | | |
| rs9368283 | 0.0157 | rs7078328 | 0.004 | rs8018962 | 0.0272 | | |
| rs9460546 | 3.7E–05 | rs10510107 | 0.0025 | rs8020986 | 0.0382 | | |
| rs9465871 | 1.0E–06 | rs10887121 | 0.0053 | rs10131117 | 0.0128 | | |
| rs10946398 | 2.5E–05 | rs10887122 | 0.001 | rs11621464 | 0.0231 | | |
| rs16883996 | 0.0469 | rs11200528 | 0.002 | rs17183738 | 0.0454 | | |
| | | rs11200537 | 0.0053 | | | | |

**Table 3 The number of pathways showing a significant association**

| | | *Number of pathways* | | | |
|---|---|---|---|---|---|
| *Sources* | *Disease* | *Exact* | | *Simes/FDR* | |
| WTCCC | BD | 15 | 3.23% | 22 | 4.73% |
| | CAD | 22 | 4.73% | 28 | 6.02% |
| | CD | 26 | 5.59% | 77 | 16.56% |
| | HT | 23 | 4.95% | 21 | 4.52% |
| | RA | 36 | 7.74% | 67 | 14.41% |
| | T1D | 24 | 5.16% | 136 | 29.25% |
| | T2D | 33 | 7.10% | 28 | 6.02% |
| DGI | T2D | 53 | 11.40% | 24 | 5.16% |
| NARAC & EIRA | RA | 40 | 8.60% | 103 | 22.15% |
| CORIELL | PD | 24 | 5.16% | 47 | 10.11% |
| NCBI | PD | 15 | 3.23% | 31 | 6.67% |
| CORIELL | ALS | 35 | 7.53% | 29 | 6.24% |
| NCBI | AREDS | 26 | 5.59% | 104 | 22.37% |

**Table 4 Number of replicated or shared pathways**

| *Study 1* | *Study 2* | *Exact* | *Simes/FDR* |
|---|---|---|---|
| RA (WTCCC) | RA(NARAC & EIRA) | 7 | 45 |
| T2D (WTCCC) | T2D (DGI) | 5 | 10 |
| PD(CORIELL) | PD(NCBI) | 10 | 30 |
| WTCCC | | | |
| | | Number of shared pathways | |
| | | Exact | Simes/FDR |
| CAD+HT+T2D | | 1 | 0 |
| RA+T1D | | 6 | 49 |
| CD+RA+T1D | | 1 | 7 |

SNP in the gene is common within the disease group. As shown in Supplementary Table 19, based on the most significant SNPs in the gene shared within a disease group, we can only find the shared genes in the RA+T1D disease group. However, if we perform gene-based

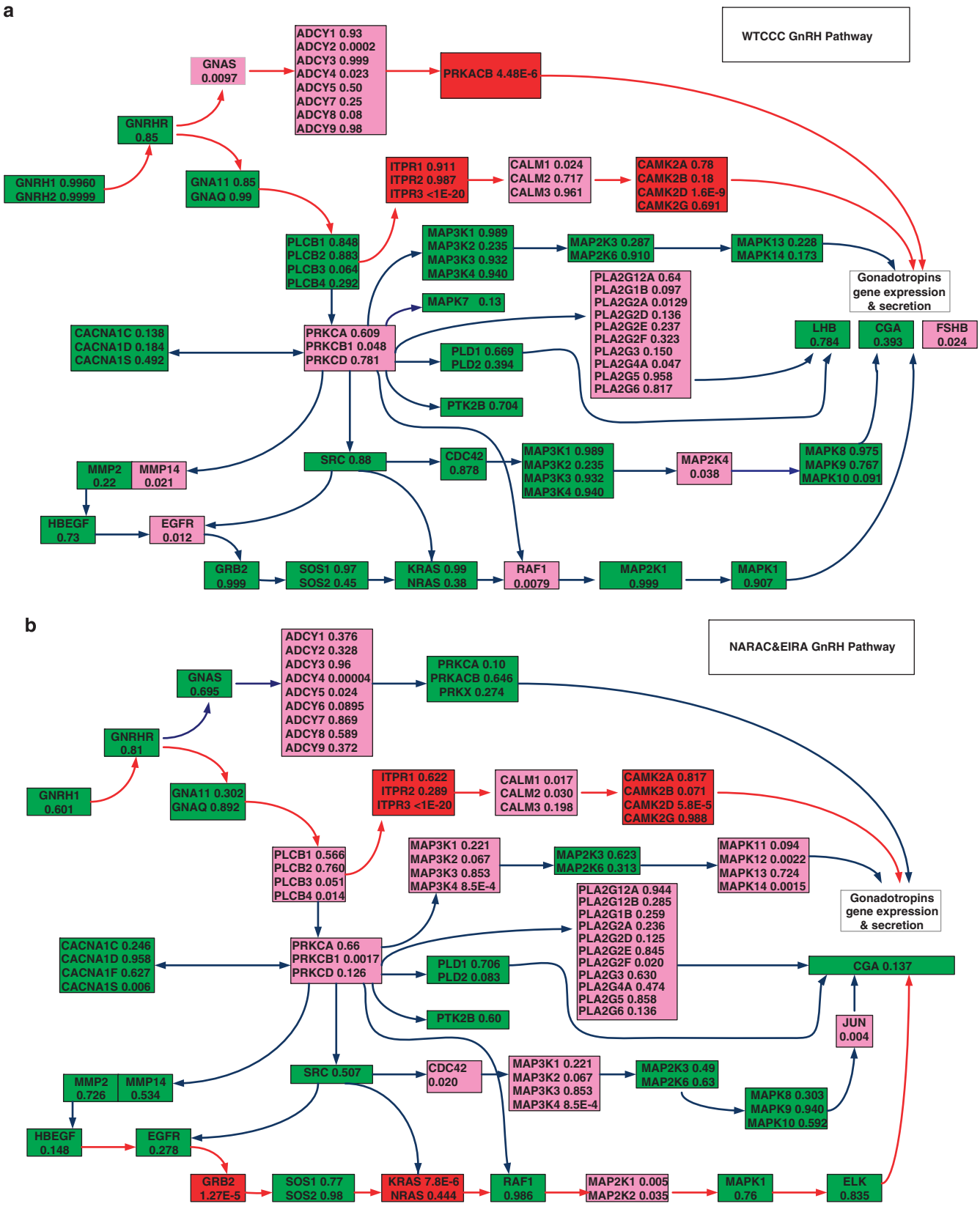



**Figure 1** *P*-values of genes in GnRH pathway for RA. (**a**) *P*-values of genes in GnRH pathway for RA in WTCCC studies. Blocks containing significant genes are in red color, blocks containing mild significant genes are in light red color and blocks containing no significant genes are in green color. (**b**) *P*-values of genes in GnRH pathway for RA in NARAC and EIRA studies. Blocks containing significant genes are in red color, blocks containing mild significant genes are in light red color and blocks containing no significant genes are in green color.

association analysis, as shown in Supplementary Table 20, we can find a number of shared genes within CD+RA+T1D, CAD+HT+T2D and RA + T1D disease groups. Numerous genome-wide gene expression analyses have shown that single-gene analysis can find little similarity between two independent studies, but pathway-based analysis may find a number of pathways in common.[24] A pathway analysis is done to identify pathways that are significantly associated with the disease. In other words, we attempt to test whether the pathway is over-represented by the genes that are significantly associated with the disease. We assembled 465 pathways from KEGG[25] and Biocarta (http://www.biocarta.com). Table 3 summarizes the number of significant pathways and Table 4 summarizes the number of replicated pathways associated with the diseases RA, T2D, and PA in two independent studies, or the number of pathways shared within the diseases CAD+HT+T2D, RA+T1D, and CD+RA+T1D in the WTCCC studies. These significant pathways were identified by an overrepresentation test and the Simes/FDR method. Supplementary Tables 21–33 summarize all significant pathways with *P*-values $\leq 0.01$, which were calculated by Fisher's exact test and by the Simes/FDR method for 13 studies. Supplementary Tables 34–36 list all significant pathways associated with the diseases RA, T2D and PA, which were replicated in two independent studies, and Supplementary Tables 37–39 list the significant pathways shared by the disease groups CAD+HT+T2D, RA+T1D, and CD+RA+T1D. These tables show several remarkable features that should be used to extract biological insight from GWAS. First, As shown in Table 3, a much larger proportion of pathways was significantly associated with the disease than that of genes, let alone SNPs. This implies that pathways have essential roles in causing disease. We note that many identified pathways showing significant association form the core of the pathway definition of complex diseases. For example, the MAPK pathway, JNK pathway, the ubiquitin–proteasome pathway, O-Glycan biosynthesis and Axon guidance, which showed significant association with PD in two studies (COR-IELL and NCBI), have been reported as a set of major pathways implicated in PD.[26,27] Pathway-based association analysis identified NF-kB, p38 MAPK, Angiotensin II-mediated activation of the JNK pathway, activation of PKC through G-protein-coupled receptor pathway, Wnt-signaling pathway, adherens junction, melanogenesis, ECM-receptor interaction and vitamin C in the brain pathway, which form the major pathways defining T2D[28] (Supplementary Table 40). Second, the results of pathway-based GWAS can be verified by functional pathway enrichment analysis of gene expressions. For example, RA is an autoimmune disease. Its major feature is a chronic inflammation of the joints. Our pathway-based association analysis identified cytokine–cytokine receptor interaction, IFN $\alpha$ signaling, Jak-STAT signaling, complement and coagulation cascades, and fatty acid biosynthesis pathways that were confirmed by pathway enrichment analysis of gene expression profiling of the peripheral blood cells of RA[29]. Third, a replication of the association of pathways in independent samples is much easier than a replication of genes or SNPs. Replications can be performed at the level of the SNP, the gene or the pathway. As shown in Table 1, no significant SNPs (using the Bonferroni method for correction of multiple tests) can be replicated in GWAS of T2D, and only seven significant genes can be replicated in the WTCCC and DGI studies. However, 10 (Simes/FDR) or 5 (Fisher's exact test) pathways can be replicated (Table 4). Risk genes may be different for different individuals, but may be in the same pathway. Identification of the pathways associated with a disease allows to easily discover the pathogenesis of the disease. Figures 1a and b plot the GnRH-signaling pathway that was associated with RA in the WTCCC studies with *P*-value $\leqslant 1.48\times10^{-14}$ (Fisher's combination test), $\leqslant 0.025$ (Fisher's exact test) and $\leqslant 0.017$ (Simes/FDR), and in the NARAC and EIRA studies with *P*-value $\leqslant 1.00\times10^{-17}$ (Fisher's combination test), $\leqslant 0.0055$(Fisher's exact test) and $\leqslant 1.39\times10^{-16}$ (Simes/FDR). Although the GnRH pathway was significantly associated with RA in both studies, the genes that showed significant association in the two studies were different. Two paths: Gs $\rightarrow$ AC $\rightarrow$ PKA $\rightarrow$ Gonadotropins gene expression and secretion and MAPK pathway (GRB2 $\rightarrow$ Sos $->$ Ras $\rightarrow$ Raf1 $\rightarrow$ MEK1/2 $\rightarrow$ ERK1/2 $\rightarrow$ Gonadotropins gene expression and secretion) are involved in the GnRH pathway. In the WTCCC studies, genes, such as GNAS (Gs, *P*-value $<0.0097$), ADCY2 (AC, *P*-value $<0.000191$) and PRKACB (PKA, *P*-value $<4.48\times10^{-6}$) in the first path showed a strong or mild association, but did not show any association in the NARAC and EIRA studies. The genes in the second path (MAPK pathway): GRB2 (*P*-value $<1.27\times10^{-5}$), KRAS (Ras, *P*-value $<7.77\times10^{-6}$) and MAP2K1 (ERK, *P*-value $<0.005$), were associated with RA in the NARAC and EIRA studies, but not in the WTCCC studies. It is well known that the endocrine system may have an important role in the pathogenesis of RA. Gonadotropins are hormones secreted by gonadotrope cells of the pituitary gland. The two major gonadotropins are luteinizing hormone and follicle-stimulating hormone. Gonadotropins have marked immunomodulatory properties and may have important roles in the pathogenesis of various immune-regulatory diseases. Sex hormone levels, including estrogen and/or progesterone in women and testosterone in men, are reported as relatively low in most RA patients.[30] These observations are consistent with the disease mechanisms associated with gonadotropin. It is interesting to note that the *P*-values of the best SNP in genes PRKACB, GRB2 and KRAS were 0.013, 0.006 and 0.0012, respectively. This example shows that each SNP may confer a small contribution, but their joint actions may affect the functioning of the pathway, which in turn will cause the disease.

## DISCUSSION

Despite the rapid progress of GWAS, the most widely used approach in GWAS is individual SNP association analysis. In other words, it evaluates the significance of individual SNPs. However, GWAS at only SNP level has serious limitations. It offers only a limited understanding of complex diseases as an integrated whole. What should be the future developments for GWAS? To address this issue, we proposed to take a system biology approach, which considers not only SNP but also gene and pathway as basic units of GWAS, to decipher a complex path from genotype to phenotype. The proposed paradigm for GWAS consists of three components: SN*P*-, gene- and pathway-based association analyses. We performed comprehensive gene and pathway-based GWAS for 11 diseases, assuming that the results of single-SNP association analysis are available. Our results showed that the proposed new paradigm for GWAS not only identified the genes that include significant SNPs found by single-SNP analysis, but also detected new genes in which each single SNP conferred a small disease risk; however, their joint actions were implicated in the development of diseases. We analysed the new genes that were identified by the new paradigm for GWAS from two aspects. First, these new findings were replicated in two independent samples. Second, the SNPs that are located in the newly identified genes were not significant in any of their original studies, but showed strong association in the recently published meta-analysis of genome-wide association data and large-scale replication. Our results also strongly showed that the replication of an association finding at the gene or pathway level is much easier than replication at the individual SNP level. One of the major advantages offered by the new paradigm

for GWAS is that the pathway-based analysis can add structure to genomic data and allows us to gain insight into a deeper understanding of cellular processes as intricate networks of functionally related genes. We further showed that the new paradigm can also offer opportunities for finding the pathways that are common within disease groups. We used RA as an example to show that the pathways identified by the new paradigm for GWAS can be confirmed by a gene-set-rich analysis using gene expression data. This implies that the new paradigm for GWAS will open a new avenue to integrate GWAS with other functional analyses and hence will facilitate to uncover the mechanism of complex diseases.

As the current GWAS only report the *P*-value for a single SNP, and the individual genotype data are not publically available, our methods for a gene and pathway-based GWAS are designed for the *P*-value data. The major tool for gene and pathway-based analyses is to combine independent *P*-values of single SNPs in the gene into an overall *P*-value for the gene and independent *P*-values of a single gene in the pathway into an overall *P*-value for the pathway. As the SNPs in a gene are often dependent, we need methods for combining dependent *P*-values, which in turn require individual genotype information. The limitation of the proposed gene and pathway-based association analysis is that it is based on combining independent *P*-values and is not appropriate to be applied to dependent data. Therefore, the *P*-values for the gene or pathway, which are calculated by Fisher's method of combining independent *P*-values of SNPs, will be inflated if there exist large correlations among SNPs in the gene. A gene and pathway-based analysis that uses methods to combine dependent *P*-values will be needed. Gene and pathway-based GWAS that take correlations among the SNP and genes into account will be carried out in the near future.

## ACKNOWLEDGEMENTS

MM Xiong is supported by a grant from the National Institutes of Health NIAMS P01 AR052915-01A1, NIAMS P50 AR054144-01 CORT, HL74735, and ES09912, and a grant from the Hi-Grant from the National Institutes of Health Tech Research and Development Program of China(863) (2007AA02Z312). CI Amos is supported by a grant from the National Institutes of Health ES09912, JD Reveille is supported by a grant from the National Institutes of Health NIAMS P01 AR052915-01A1, L Jin is supported by a grant from the Shanghai Commission of Science and Technology (04dz14003) and a grant from the Hi-Tech Research and Development Program of China(863) (2007AA02Z312).

1 Saxena R, Voight BF, Lyssenko V *et al*: Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science* 2007; **316**: 1331–1336.
2 The Wellcome Trust Case Control Consortium: genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 2007; **447**: 661–678.
3 Rioux JD, Xavier RJ, Taylor KD *et al*: Genome-wide association study identifies new susceptibility loci for Crohn disease and implicates autophagy in disease pathogenesis. *Nat Genet* 2007; **39**: 596–604.
4 Sladek R, Rocheleau G, Rung J *et al*: A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature* 2007; **445**: 881–885.
5 Zanke BW, Greenwood CM, Rangrej J *et al*: Genome-wide association scan identifies a colorectal cancer susceptibility locus on chromosome 8q24. *Nat Genet* 2007; **39**: 989–994.
6 Haiman CA, Patterson N, Freedman ML *et al*: Multiple regions within 8q24 independently affect risk for prostate cancer. *Nat Genet* 2007; **39**: 638–644.
7 Gudmundsson J, Sulem P, Steinthorsdottir V *et al*: Two variants on chromosome 17 confer prostate cancer risk, and the one in TCF2 protects against type 2 diabetes. *Nat Genet* 2007; **39**: 977–983.
8 Moffatt MF, Kabesch M, Liang L *et al*: Genetic variants regulating ORMDL3 expression contribute to the risk of childhood asthma. *Nature* 2007; **448**: 470–473.
9 Zeggini E, Weedon MN, Lindgren CM *et al*: Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. *Science* 2007; **316**: 1336–1341.
10 Scott LJ, Mohlke KL, Bonnycastle LL *et al*: A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science* 2007; **316**: 1341–1345.
11 Frayling TM, Timpson NJ, Weedon MN *et al*: A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. *Science* 2007; **316**: 889–894.
12 Plenge RM, Seielstad M, Padyukov L *et al*: TRAF1-C5 as a risk locus for rheumatoid arthritis–a genomewide study. *N Engl J Med* 2007; **357**: 1199–1209.
13 Lesnick TG, Papapetropoulos S, Mash DC *et al*: A genomic pathway approach to a complex disease: axon guidance and Parkinson disease. *PLoS Genet* 2007; **3**: e98.
14 Neale BM, Sham PC: The future of association studies: gene-based analysis and replication. *Am J Hum Genet* 2004; **75**: 353–362.
15 Casci T: The best of the rest. *Nat Rev Genet* 2007; **8**: 907.
16 Wang K, Li M, Bucan M: Pathway-Based Approaches for Analysis of Genomewide Association Studies. *Am J Hum Genet* 2007; **81**.
17 Curtis RK, Oresic M, Vidal-Puig A: Pathways to the analysis of microarray data. *Trends Biotechnol* 2005; **23**: 429–435.
18 Subramanian A, Tamayo P, Mootha VK *et al*: Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA* 2005; **102**: 15545–15550.
19 Pounds S, Cheng C: Robust estimation of the false discovery rate. *Bioinformatics* 2006; **22**: 1979–1987.
20 Casci T: The best of the rest. *Nat Rev Genet* 2007; **8**: 907.
21 Zaykin DV, Zhivotovsky LA, Czika W, Shao S, Wolfinger RD: Combining *P*-values in large-scale genomics experiments. *Pharm Stat* 2007; **6**: 217–226.
22 Zeggini E, Scott LJ, Saxena R *et al*: Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. *Nat Genet* 2008; **40**: 638–645.
23 Diabetes Genetics Initiative of Broad Institute of Harvard and MIT, Lund University, and Novartis Institutes of BioMedical Research: genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science* 2007; **316**: 1331–1336.
24 Nam D, Kim SY: Gene-set approach for expression pattern analysis. *Brief Bioinform* 2008; **9**: 189–197.
25 Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, Kanehisa M: KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* 1999; **27**: 29–34.
26 Jankowski M: The role of JNK pathway in familial Parkinson's disease. *Postepy Biochem* 2007; **53**: 297–303.
27 Moran LB, Graeber MB: Towards a pathway definition of Parkinson's disease: a complex disorder with links to cancer, diabetes and inflammation. *Neurogenetics* 2008; **9**: 1–13.
28 Evans JL, Goldfine ID, Maddux BA, Grodsky GM: Oxidative stress and stress-activated signaling pathways: a unifying hypothesis of type 2 diabetes. *Endocr Rev* 2002; **23**: 599–622.
29 van der Pouw Kraan TC, Wijbrandts CA, van Baarsen LG *et al*: Rheumatoid arthritis subtypes identified by genomic profiling of peripheral blood cells: assignment of a type I interferon signature in a subpopulation of patients. *Ann Rheum Dis* 2007; **66**: 1008–1014.
30 Wilder RL: Adrenal and gonadal steroid hormone deficiency in the pathogenesis of rheumatoid arthritis. *J Rheumatol Suppl* 1996; **44**: 10–12.

Supplementary Information accompanies the paper on European Journal of Human Genetics website (http://www.nature.com/ejhg)