Online Supplementary Text:

Determination of *IRF5* transcript levels in monocytes

Purified monocyte cell lines were obtained from a subgroup of SSc patients enrolled in the GENISOS cohort and unaffected age, gender and ethnicity matched controls. First, the whole blood samples, collected in heparin tubes were processed within two hours of collection. The samples were diluted and layered on top of Ficoll and centrifuged for collection of mononuclear cells. The cells were incubated with magnetic bead-bound antibodies to remove non-monocyte cells (negative selection for monocytes). The purity of cell populations was examined by flow cytometry. The total RNA from enriched monocytes was immediately extracted using RNAeasy Mini kits from Qiagen (Chatsworth, CA). The RNA quality was assessed by an Agilent 2100 Bioanalyzer (Agilent Technologies, Inc., Palo Alto, CA). The RNA was amplified and hybridized on Illumina HumanHT-12 arrays (San Diego, CA). The gene expression data were quantile-normalized utilizing GenomeStudio software (Illumina Inc., SanDiego, CA). Two IRF5 probes, representing two transcript variants of this gene were present on the array. The expression of both IRF5 transcript variants was significantly above the background level in all samples. The association of IRF5 expression with the three investigated *IRF5* polymorphisms was separately examined in SSc patients and unaffected controls.
The two investigated IRF5 transcript variants correspond to refSeq IDs NM_001098629 and NM_032643. These two transcript variants were the only registered IRF5 isoforms as the HumanHT-12 arrays were developed in 2009 and are represented by the probes ILMN-2312606 (=NM_001098629) and ILMN-1670576 (=NM_032643). Since then, three other variants have been registered (NCBI→gene→ IRF5 interferon regulatory factor 5 [*Homo sapiens*]). However, the two other transcript variants NM_001098630.1 and NM_001098627.2 code for the same protein like NM_032643, these three transcript variants have only minor differences in the 5' and 3' untranslated regions. The third newly registered gene variant NM_001242452.1 lacks an alternate in-frame exon compared to the gene variant NM_001098629. In summary, there are three different protein isoforms of *IRF5*. The protein isoform coded by NM_001098629 is the longest while the other two protein isoforms are missing segments of protein. We have investigated transcripts translated into two protein isoforms of *IRF5*. The transcript corresponding to the third protein isoform of IRF5 coded by NM_001242452.1 was not investigated in the current study.

Subgroup analysis of GWAS data for association of *IRF5* rs4728142 with limited/diffuse and ACA/ATA subtypes of SSc:

We investigated whether *IRF5* rs4728142 is preferentially associated with limited or ACA subsets of systemic sclerosis (SSc). For this purpose, subgroup analyses were performed on the data obtained from the first SSc-GWAS[1]. In these analyses, rs4728142A did not show a stronger association in comparison of limited and ACA positive subsets to controls versus the comparison of diffuse and ATA positive subsets to controls. These are the results of the above mentioned analyses:

- Limited subsets versus controls: **OR=1.256** (1.154, 1.366), *p*= 1.23E-07
- Diffuse subset versus controls: **OR=1.196** (1.07, 1.335), *p*=0.001524
- ACA positive subset versus controls: **OR=1.304** (1.17, 1.453), *p*= 1.47E-06

- ATA positive subset versus controls: **OR=1.299** (1.131, 1.492), *p*= 0.000203

Association of *IRF5 rs20044640 with survival or severity of ILD:*

The *IRF5* SNP, rs2004640 has been identified as a susceptibility locus for SSc in a case control study.[2] The *IRF5* rs2004640 SNP was not present on the platform used in the SSc-GWA study. We genotyped the SSc samples enrolled in the discovery cohort by Taqman Assay for the *IRF5* rs2004640 SNP, as previously described.[3] This SNP is in linkage disequilibrium with rs4728142 (r2=0.67 - based on the genotype information in the discovery cohort). However, the *IRF5* rs2004640 was not significantly associated with survival (HR=0.82, 95% CI: 0.65 - 1.05, *p*=0.115). In agreement with previously published data,[2] the *IRF5* rs2004640 also did not correlate with severity of ILD (b=0.195, 95% CI: -0.79 – 3.87, *p*=0.195).

Association of *IRF5 risk haplotype with survival or severity of ILD:*

The *IRF5* haplotype rs3757385*C, rs2004640*T, and rs10954213*A has been identified as a risk haplotype for susceptibility to SSc in a case control study. [4] This *IRF5* haplotype tagged the insertion variant of *IRF5* CGGGG insertion /deletion polymorphism while the other five occurring genotype combinations in this haplotype tag the deletion variant in unaffected controls [5]. We have recently completed the genotyping of *IRF5* SNPs rs3757385 and rs10954213 SNPs by Immunochip in the discovery cohort. Building on the above mentioned findings, we investigated whether the *IRF5* haplotype rs3757385*C, rs2004640*T, and rs10954213*A (and indirectly *IRF5* CGGGG insertion /deletion polymorphism) was associated with severity of SSc in our study. In this analysis, the *IRF5* haplotype tagging the insertion variant of the *IRF5* CGGGG ins/del was neither significantly associated with survival (HR= 0.82, 95% CI: 0.63-1.05, *p*=0.122) nor with severity of ILD (b=1.37, 95% CI:-1.07 - 3.81, p=0.271).

Reference List

(1) Radstake TR, Gorlova O, Rueda B et al. Genome-wide association study of systemic sclerosis identifies CD247 as a new susceptibility locus. *Nat Genet* 2010;42:426-429.

(2) Dieude P, Guedj M, Wipff J et al. Association between the IRF5 rs2004640 functional polymorphism and systemic sclerosis: A new perspective for pulmonary fibrosis. *Arthritis Rheum* 2009;60:225-233.

(3) Gourh P, Agarwal SK, Martin E et al. Association of the C8orf13-BLK region with systemic sclerosis in North-American and European populations. *J Autoimmun* 2009.

(4) Dieude P, Dawidowicz K, Guedj M et al. Phenotype-haplotype correlation of IRF5 in systemic sclerosis: role of 2 haplotypes in disease severity. *J Rheumatol* 2010;37:987-992.

(5) Kristjansdottir G, Sandling JK, Bonetti A et al. Interferon regulatory factor 5 (IRF5) gene variants are associated with multiple sclerosis in three distinct populations. *J Med Genet* 2008;45:362-369.