Supplement 1: Methods

Globin Reduction:

We also conducted globin mRNA reduction on 9 healthy control samples utilizing GLOBINclear-Human kit (Applied Biosystems/Ambion, Austin, TX); subsequently, pre- and post- globin reduction samples were processed and hybridized on HumanRef-8 v2 arrays according to the above mentioned protocol. A comparison of pre- and post-globin samples showed that globin reduction had lead to a decrease in signal intensities for hemoglobin transcripts but not to an increased in the percent present calls; this finding might be explained by the longer transcript probes printed on the Illumina arrays (50 mer probes) compared to Affymetrix platforms. Therefore, the remainder of our experiments were conducted on samples that had not undergone globin reduction because the added advantage of a decrease in hemoglobin transcripts without an increase in percent calls did not justify the substantial changes in gene expression caused by ex vivo handling(1;2).

Ingenuity Pathway Analysis (IPA):

Expression data were analyzed and modeled through the use of Ingenuity Pathways Analysis (Ingenuity Systems®, www.ingenuity.com). The IPA Canonical Pathways Analysis identified the pathways from the Ingenuity Pathways Analysis library of canonical pathways that were most significant in our dataset. Transcripts from our dataset that met criteria for differential expression (p<0.01; FDR<10%) and were associated with a canonical pathway in the Ingenuity Knowledge Base were considered for the analysis. The significance of the association between the dataset and the canonical pathway was measured in 2 ways (Figure 1): a) A ratio of the number of differentially expressed transcripts from the dataset that map to the pathway divided by the total number of molecules that exist in the canonical pathway. The ratio is useful for determining which pathways overlap the most with the differentially expressed transcripts in a given dataset. b) Fischer’s exact test was used to calculate a p-value determining the probability that the association between the genes in the dataset and the canonical pathway is explained by chance alone.

Modular Data Analysis:

We also investigated blood transcriptional phenotypes of our study subjects employing a modular data mining strategy described by Chaussabel et al. (3). Briefly, this dimension reduction approach consists in identifying transcriptional units or modules which are formed by sets of co-expressed genes. This analysis framework was constructed on a collection of datasets obtained for a wide range of diseases, including Systemic Lupus Erythematosus, Systemic Arthritis, Type 1 Diabetes, metastatic melanoma, and acute infections. The transcriptional modules thus obtained were functionally annotated in order to facilitate biological interpretation. This modular analysis framework has been used for the interpretation of public domain datasets (www.biir.net/modules).

Interferon Score:

We calculated an IFN score according to the method outlined by Baechler et al. based on a set of 286 IFN- inducible genes indentified by treatment of healthy control blood cells with IFN α/β and IFN γ (4). The IFN signature score was calculated based on the 43 transcripts that were differentially expressed in SSc and SLE patients when compared to healthy controls in our study. For normalization of transcript
values, the expression data were divided by the top 95th percentile level of each gene. Furthermore, all values in the top 5% category were assigned a value of 1.0. Subsequently, the normalized values for all genes were summed up to obtain an IFN score for each participant.

Real-time qualitative PCR:

qPCR assays for STAT1, IFI6, IFIT3, TLR5 genes were designed to confirm the microarray results. Each sample was assayed in triplicate plus a control without reverse transcriptase to access DNA contamination levels. Samples were reverse-transcribed into cDNA using Superscript II reverse transcriptase (invitrogen, Carlsbad, CA) in 384-well plates PCR master mix, containing JumpStrat Taq Polymerase (sigma, St. Louis, MO) was added to the samples. Each assembled plate was run in a 7900 real-time instrument using the following cycling conditions: 95°C, 1 min; followed by 40 cycles of 95°C, 12 sec and 60°C, 30 sec. The resulting data were analyzed using SDS 2.3 (7900) software (Applied Biosystems, Foster City, CA) with FAM reporter and ROX as the reference dye. The final data were normalized to 18S rRNA levels. The final data are presented as the molecules of unknown transcript/molecules of 18sRNA transcript x 100.

An average gene expression value for IFN-inducible genes was calculated utilizing a normalization method described by Niewold et al. (5). The means and standard deviations (SD’s) of the respective genes were calculated in the healthy controls. The respective means in healthy controls were subtracted from the expression values in each subject and the residues were divided by the SD value for the same gene in healthy controls in order to calculate the relative number of SD’s above that in healthy controls. The average of relative values in the three IFN-inducible genes (STAT1, IFI6, IFIT3) yielded the final score.

Genotyping:

Genomic DNA was collected in addition to PAXgene samples from SSc patients and normal controls. These individuals were genotyped for single nucleotide polymorphisms (SNP’s) of a selected group of IFNIG’s within the framework of a larger candidate gene study. A number of SLE or SSc susceptibility SNP’s (6-8), in addition to few non-synonymous polymorphisms with a minor allele frequency >5% in the HAPMAP CEU population were chosen. The IFNAR1 (rs17875834 and rs2257167), IFNAR2 (rs4986956 and rs7279064), IRF5 (rs2004640, rs729302, rs752637), IRF7 (rs1061502), STAT1 (rs1914408), and STAT4 (rs11889341, rs6752770) SNP’s were examined using an ABI TaqMan 5’ allele discrimination Assay-by-Design method (Applied Biosystems, Foster City, CA). Details of the utilized genotyping method has been previously published (9).

The genotype results were correlated with the respective IFN scores. We adjusted for the effect of ethnicity by building a bivariate linear regression model of ethnicity and genotypes as independent variables.

Reference List


