Online Methods

This study reports the results of a meta-analysis for LN based on three GWAS of women of European descent with SLE. The three GWAS used include: (1) Set I: samples genotyped using the Illumina HumanHap550v1 as described by Hom et al.; [1] 2) Set II: samples genotyped using the Illumina HumanHap300 as described by Harley et al.; [2] and 3) Set III: samples genotyped using the Affymetrix 5.0 as described by Graham et al. [3] The primary inference of this study is based on a female-only case-case (i.e., LN cases versus SLE cases without LN) analysis to minimize the potential influence of SLE susceptibility variants in our tests of association with LN. The classification of SLE and LN in all three GWAS samples was based on the American College of Rheumatology (ACR) criteria, including proteinuria and/or cellular casts in the urine. [4; 5]

The individual SLE GWAS designs attempted to maximize the number of SLE multiplex cases (i.e., SLE cases with family members with confirmed SLE) but only one case was selected per pedigree (93/1163 in Set I; 591/720 in Set II; and 122/282 in Set III). Details of the individual calling algorithms, call rates and study-specific quality control were previously reported.[1-3] Briefly, each sample was tested for cryptic relatedness (duplicates and first-degree relatives) via the proportion of alleles shared identical by descent, autosomal heterozygosity, gender inconsistencies and departures from Hardy-Weinberg Equilibrium expectations. Four hundred highly polymorphic SNPs common to both the Illumina and Affymetrix chips were used to identify duplicate or first-degree relatives across Sets I, II and III. In total across these three studies, 282 duplicates and 71 first-degree relatives were identified. The data from female samples with the most complete LN data and then the highest call rate was used in the meta-analysis. Thus, the difference in the number of SLE individuals in the previously published

GWAS (1163 in Set I; 720 in Set II; and 282 in Set III) and the numbers in Table 1 reflect the removal of males, related individuals (first degree relatives and identical individuals) and individuals who did not have information on LN status (presence/absence).

Admixture, SNP statistical quality control and association testing.

To account for potential population substructure, principal component (PC) analyses were computed using SNPs that passed quality control and were not in genomic regions with extended linkage disequilibrium (LD). [6] For Set I, a PC analysis was computed using all SNPs that remained after removal of markers in regions with extensive LD on chromosomes 5 (44-51.5 Mb), 6 (25-33.5 Mb), 8 (8-12 Mb), 11 (45-57 Mb), and 17 (40-43 Mb). An independent PC analysis was computed on the combined Set II and Set III samples using 38,092 SNPs common to the two genotyping platforms (i.e., Infinium HumanHap 300 and Affymetrix 5.0), again after removing the regions of extended LD. All tests of association were computed using a logistic regression model and adjusted for five PCs in the Set I samples and four PCs in the Set II and Set III scans.

It is well established that polymorphisms in the extended HLA region are strongly associated with SLE. In an effort to increase the statistical power of the tests of association, the analyses were repeated adjusting for rs9271366 and rs2187668, two SNPs that are reported to tag *HLA-DR2* and *HLA-DR3*, respectively. [7] These two HLA serotypes are the most strongly associated with SLE. [8-10] These two SNPs and the PCs were added as covariates to the logistic regression model and the results were combined as described in the meta-analysis. Separate models were computed for dominant, additive and recessive genetic models.

Imputation

Within each GWAS set, autosomal SNPs that passed quality control filters and exhibited a minor allele frequency greater than 0.01 were used to impute the ~2.5 million SNPs in HapMap2 (Build 35) using IMPUTE. [11] The reference set for imputation was the CEU (i.e., United States samples of northern and western European origin available via the HapMap). Standard post imputation quality control was completed for each SNP, yielding an imputation set of 1,621,689 SNPs in 2000 individuals. Within each GWAS set, logistic regression models included the above stated PCs and accounted for the imputation uncertainty using the software SNPTEST. [11] For each SNP, dominant, additive and recessive genetic models were computed. To increase the robustness of the tests of association, the additive and recessive genetic models required at least 10 and 30 individuals homozygous for the minor allele, respectively.

Meta-Analysis

The evidence of association from each of the three GWAS sets was combined using the weighted inverse normal method as implemented in METAL [12] and weighted by sample size. Assuming appropriate numbers of homozygotes for the model, meta-analysis p-values were calculated for additive, dominant and recessive models and the minimum p-value was chosen from the three. A test for heterogeneity of the odds ratios across the three studies was also computed using METAL. It is important to recognize that because of the small number of samples in Set III that are independent of the Set I and Set II scans, a higher proportion of the SNPs from Set III failed to meet the additive and recessive model count criteria for homozygotes for the minor allele.

Renal biopsies for mRNA analysis

Renal biopsies from 15 healthy pre-transplant living donor controls and 32 patients with LN were collected from an international multi-center repository (European Renal cDNA Bank). [13] All healthy controls had normal kidney function as determined by estimated MDRD-GFR between 80 and 120 ml/min/m^2. The glomerular and tubulointerstitial compartments of kidney biopsies were microdissected and processed as previously described, [13] including the RNA isolation and reverse transcription. [14; 15]

Human microarray data processing and analysis

The glomerular and tubulointerstitial compartments were analyzed using Affymetrix Human Genome U133A Genechips (Affymetrix, Inc, Santa Clara, CA) at the University of Michigan Microarray Core Facility following the manufacturer's instructions. [15] The CEL files were normalized in GenePattern pipeline (www.GenePattern.org) using the RMA (Robust MultiChip Average) method and the Human Entrez Gene custom annotation version 10 (http://brainarray.mbni.med.umich.edu/Brainarray/default.asp). Of the 12,029 gene IDs (corresponding to the 22,283 Affymetrix probe sets), 11,285 and 11,429 genes were expressed above the Poly-A Affymetrix control expression baseline (negative controls) in the glomeruli and tubulointerstitium, respectively, and used for further analyses. Normalized data files were uploaded onto the Gene Omnibus website (http://www.ncbi.nlm.nih.gov/geo/) under the reference GSE32591.

mRNA expression levels generated from the biopsies of the LN cases and from healthy controls (no lupus, no nephritis) were compared using unpaired t-tests and Significance Analysis of Microarrays (SAM) implemented in the MultiExperiment Viewer (MeV) application in the

SAM software. [16; 17] Genes regulated between the two groups with a q-value (False Discovery Rate) less than 0.05 were considered significant and used for further transcriptional analyses.

A transcriptional network from all the genes significantly differentially regulated as defined by our filter criteria was generated using Genomatix Pathway System software (GePS, www.genomatix.de). GePS allows visualization of dependencies among genes in pathways, networks and processes derived from literature based knowledge and genome-wide sequence analysis. As *PDGFRA* was found to be the locus most significantly associated with LN, we queried the software to show the shortest path network from *PDGFRA*. We applied a function-word filter, meaning that to be displayed in the transcriptional network, two genes have to be co-cited in the same sentence with a function word (e.g. gene A activates gene B). Pearson correlations were used to determine the regulated transcripts significantly correlating with *PDGFRA* mRNA expression in LN (p-value<0.05).

Reference List

1. Hom G, Graham RR, Modrek B, Taylor KE, Ortmann W, Garnier S, Lee AT, Chung SA, Ferreira RC, Pant PV, Ballinger DG, Kosoy R, Demirci FY, Kamboh MI, Kao AH, Tian C, Gunnarsson I, Bengtsson AA, Rantapaa-Dahlqvist S, Petri M, Manzi S, Seldin MF, Ronnblom L, Syvanen AC, Criswell LA, Gregersen PK, Behrens TW: Association of systemic lupus erythematosus with C8orf13-BLK and ITGAM-ITGAX. *N Engl J Med* 358:900-909, 2008

2. Harley JB, arcon-Riquelme ME, Criswell LA, Jacob CO, Kimberly RP, Moser KL, Tsao BP, Vyse TJ, Langefeld CD, Nath SK, Guthridge JM, Cobb BL, Mirel DB, Marion MC, Williams AH, Divers J, Wang W, Frank SG, Namjou B, Gabriel SB, Lee AT, Gregersen PK, Behrens TW, Taylor KE, Fernando M, Zidovetzki R, Gaffney PM, Edberg JC, Rioux JD, Ojwang JO, James JA, Merrill JT, Gilkeson GS, Seldin MF, Yin H, Baechler EC, Li QZ, Wakeland EK, Bruner GR, Kaufman KM, Kelly JA: Genome-wide association scan in women with systemic lupus erythematosus identifies susceptibility variants in ITGAM, PXK, KIAA1542 and other loci. *Nat Genet* 40:204-210, 2008

3. Graham RR, Cotsapas C, Davies L, Hackett R, Lessard CJ, Leon JM, Burtt NP, Guiducci C, Parkin M, Gates C, Plenge RM, Behrens TW, Wither JE, Rioux JD, Fortin PR, Graham

DC, Wong AK, Vyse TJ, Daly MJ, Altshuler D, Moser KL, Gaffney PM: Genetic variants near TNFAIP3 on 6q23 are associated with systemic lupus erythematosus. *Nat Genet* 40:1059-1061, 2008

4. Hochberg, M.C. (1997). Updating the American College of Rheumatology revised criteria for the classification of systemic lupus erythematosus. Arthritis Rheum *40*, 1725.

5. Tan EM, Cohen AS, Fries JF, Masi AT, McShane DJ, Rothfield NF, Schaller JG, Talal N, Winchester RJ: The 1982 revised criteria for the classification of systemic lupus erythematosus. *Arthritis Rheum* 25:1271-1277, 1982

6. Price AL, Weale ME, Patterson N, Myers SR, Need AC, Shianna KV, Ge D, Rotter JI, Torres E, Taylor KD, Goldstein DB, Reich D: Long-range LD can confound genome scans in admixed populations. *Am J Hum Genet* 83:132-135, 2008

7. de Bakker PI, McVean G, Sabeti PC, Miretti MM, Green T, Marchini J, Ke X, Monsuur AJ, Whittaker P, Delgado M, Morrison J, Richardson A, Walsh EC, Gao X, Galver L, Hart J, Hafler DA, Pericak-Vance M, Todd JA, Daly MJ, Trowsdale J, Wijmenga C, Vyse TJ, Beck S, Murray SS, Carrington M, Gregory S, Deloukas P, Rioux JD: A high-resolution HLA and SNP haplotype map for disease association studies in the extended human MHC. *Nat Genet* 38:1166-1172, 2006

8. Barcellos LF, May SL, Ramsay PP, Quach HL, Lane JA, Nititham J, Noble JA, Taylor KE, Quach DL, Chung SA, Kelly JA, Moser KL, Behrens TW, Seldin MF, Thomson G, Harley JB, Gaffney PM, Criswell LA: High-density SNP screening of the major histocompatibility complex in systemic lupus erythematosus demonstrates strong evidence for independent susceptibility regions. *PLoS Genet* 5:e1000696, 2009

9. Graham RR, Ortmann W, Rodine P, Espe K, Langefeld C, Lange E, Williams A, Beck S, Kyogoku C, Moser K, Gaffney P, Gregersen PK, Criswell LA, Harley JB, Behrens TW: Specific combinations of HLA-DR2 and DR3 class II haplotypes contribute graded risk for disease susceptibility and autoantibodies in human SLE. *Eur J Hum Genet* 15:823-830, 2007

10. Graham RR, Ortmann WA, Langefeld CD, Jawaheer D, Selby SA, Rodine PR, Baechler EC, Rohlf KE, Shark KB, Espe KJ, Green LE, Nair RP, Stuart PE, Elder JT, King RA, Moser KL, Gaffney PM, Bugawan TL, Erlich HA, Rich SS, Gregersen PK, Behrens TW: Visualizing human leukocyte antigen class II risk haplotypes in human systemic lupus erythematosus. *Am J Hum Genet* 71:543-553, 2002

11. Marchini J, Howie B, Myers S, McVean G, Donnelly P: A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet* 39:906-913, 2007

12. Willer CJ, Li Y, Abecasis GR: METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* 26:2190-2191, 2010

13. Cohen CD, Frach K, Schlondorff D, Kretzler M: Quantitative gene expression analysis in renal biopsies: a novel protocol for a high-throughput multicenter application. *Kidney Int* 61:133-140, 2002

14. Barrett T, Troup DB, Wilhite SE, Ledoux P, Rudnev D, Evangelista C, Kim IF, Soboleva A, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM, Muertter RN, Edgar R: NCBI GEO: archive for high-throughput functional genomic data. *Nucleic Acids Res* 37:D885-D890, 2009

15. Schmid H, Boucherot A, Yasuda Y, Henger A, Brunner B, Eichinger F, Nitsche A, Kiss E, Bleich M, Grone HJ, Nelson PJ, Schlondorff D, Cohen CD, Kretzler M: Modular activation of nuclear factor-kappaB transcriptional programs in human diabetic nephropathy. *Diabetes* 55:2993-3003, 2006

16. Saeed AI, Bhagabati NK, Braisted JC, Liang W, Sharov V, Howe EA, Li J, Thiagarajan M, White JA, Quackenbush J: TM4 microarray software suite. *Methods Enzymol* 411:134-193, 2006

17. Saeed AI, Sharov V, White J, Li J, Liang W, Bhagabati N, Braisted J, Klapa M, Currier T, Thiagarajan M, Sturn A, Snuffin M, Rezantsev A, Popov D, Ryltsov A, Kostukovich E, Borisovsky I, Liu Z, Vinsavich A, Trush V, Quackenbush J: TM4: a free, open-source system for microarray data management and analysis. *Biotechniques* 34:374-378, 2003