Supplementary Information

Multiple Polymorphisms in the TNFAIP3 Region are Independently Associated with Systemic Lupus Erythematosus

Stacy L. Musone[1], Kimberly E. Taylor[2], Timothy T. Lu[3], Joanne Nititham[2], Ricardo C. Ferreira[4], Ward Ortmann[4], Nataliya Shifrin[3], Michelle A. Petri[5], M. Ilyas Kamboh[6], Susan Manzi[6], Michael F. Seldin[7], Peter K. Gregersen[8], Timothy W. Behrens[4], Averil Ma[3], Pui-Yan Kwok[1] and Lindsey A. Criswell[2]

[1]Cardiovascular Research Institute, [2]Rosalind Russell Medical Research Center for Arthritis and [3]Colitis and Crohn's Disease Center, Department of Medicine, University of California San Francisco, San Francisco, California; [4]Genentech, South San Francisco, California; [5]Johns Hopkins University School of Medicine, Baltimore, Maryland; [6]University of Pittsburgh Medical Center, Pittsburgh, Pennsylvania; [7]Rowe Program in Molecular Medicine and Human Genetics, University of California Davis, Davis, California; [8]Feinstein Institute for Medical Research, North Shore L.I.J. Health System, Manhasset, New York

S.L.M. and K.E.T. contributed equally to this work.

Supplementary Methods

Subjects

SLE cases were obtained from four sources. Patients from the University of California, San Francisco (UCSF) were participants in the UCSF Lupus Genetics Project and were recruited from UCSF Arthritis Clinics and private rheumatology practices in northern California, as well as by nationwide outreach [1]. SLE patients contributed by the Autoimmune Biomarkers Collaborative Network (ABCoN) [2] were recruited from the Hopkins Lupus cohort[3]. A third case series was part of the Multiple Autoimmune Disease Genetics Consortium (MADGC) collection [4]. Finally, a fourth set of cases recruited from the Pittsburgh Lupus Registry were obtained from the University of Pittsburgh [5]. Unrelated healthy controls were from the New York Health Project (NYHP) [6] (http://www.amdec.org/amdec_initiatives/nycp.html). All cases were

confirmed for SLE diagnosis by documentation of at least four American College of Rheumatology (ACR) criteria[7] in medical record reviews (95%) or by written confirmation from a treating rheumatologist. Cases were typical of SLE case series of European descent, being 93% female and having an average age of onset of 35 years (SD ± 13 years). Twenty-eight percent of subjects meet ACR criteria for renal disease and 79% meet ACR criteria for arthritis, as has been reported previously[8,9]. The Institutional Review Boards of all investigative institutions approved these studies, and all cases and controls gave written informed consent.

Genotyping and SNP selection

All cases and controls were genotyped using the Illumina HumanHap550 array, as reported previously[8]. ABCoN and MADGC cases and a subset of NYHP controls (n = 869) were genotyped on the version 1 Illumina 550K panel. All other subjects were genotyped on the version 3 Illumina550K panel. Additional genotyping for rs2230926 in ABCoN and MADGC cases was performed using a pre-validated TaqMan (Applied Biosystems) assay according to manufacturer's instructions. SNPs were removed from analysis that had a minor allele frequency less than 5% (with the exception of the non-synonymous SNP, rs2230926), greater than 10% missing genotypes, or Hardy-Weinberg equilibrium $p < 0.001$ in controls. Of the 158 SNPs in the extended TNFAIP3 region, 143 passed quality control filters; in the initial 500-kb region (Figure 1), 115 passed quality control filters.

Statistical Analysis

Subjects were first removed for whom there was evidence of duplication or relatedness in the Illumina 550K data, using IBS estimation in PLINK [10] (http://pngu.mgh.harvard.edu/purcell/plink), and who had < 90% of genotypes called. While all

subjects were of self-reported European ancestry, in order to guarantee genetic homogeneity we performed ancestry analysis using STRUCTURE [11] and a set of 235 ancestry-informative markers (AIMs) contained in the Illumina 550K panel. Subjects were removed who had < 90% estimated European ancestry.

We conducted allelic tests of cases and controls using Haploview[12]. Conditional analyses to determine independent effects were performed in Whap[13] (http://pngu.mgh.harvard.edu/purcell/whap), which uses log-ratio testing of alternative models. Stata 9.2 (http://www.stata.com/) was used for multivariate logistic regression of the three independent SNPs. Tagger[14] was used to measure r^2 between SNPs in the HapMap CEU population to determine proxies for SNPs not genotyped in our samples.

We performed stratified analyses designed to determine whether population substructure within our European subjects explained the associations of TNFAIP3 region SNPs with SLE. We first used a set of 1409 EUROSTRUCTURE AIMS[15] to estimate percent northern versus southern European ancestry. We also used the first 4 principal components determined by EIGENSTRAT[16] using whole-genome Illumina 550K data, as in Taylor et al.,[9] to determine a subset of genetically homogeneous subjects and therefore account for more subtle substructure than simply north-south. Greater than or equal to 90% membership in the northern population and membership in the homogeneous subset were each then used as stratifiers in allelic analyses of the top 3 SNPs. Strata were analyzed separately and then combined using the Mantel-Haenszel method; tests of heterogeneity and combined ORs were performed with Stata 9.2.

NFκB response assay

Human A20 cDNAs corresponding to the major and minor alleles at rs2230926 were generated by RT-PCR and Quik-change mutagenesis (Stratagene). These cDNAs were verified by sequencing and transiently transfected into 293T cells along with NFκB-luciferase and CMV-renilla reporter constructs, stimulated with 10 ng/ml TNF for 6 hours and then lysed for renilla and luciferase assays using a dual luciferase reporter assay (Promega). A20 and actin protein expression levels were determined by immunoblotting of whole cell lysates and densitometric quantification. Relative A20 expression levels between samples were determined after quantitating and normalizing A20 expression to actin expression for each sample. All assays were performed at least three times and p-values were determined by unpaired Student's T test.

References

1. Thorburn, C.M. et al. Association of PDCD1 genetic variation with risk and clinical manifestations of systemic lupus erythematosus in a multiethnic cohort. *Genes Immun* **8**, 279-287 (2007).
2. Bauer, J.W. et al. Elevated Serum Levels of Interferon-Regulated Chemokines Are Biomarkers for Active Human Systemic Lupus Erythematosus. *PLoS Med* **3**, e491 (2006).
3. Petri, M. Hopkins Lupus Cohort. 1999 update. *Rheum Dis Clin North Am* **26**, 199-213, v (2000).
4. Criswell, L.A. et al. Analysis of families in the multiple autoimmune disease genetics consortium (MADGC) collection: the PTPN22 620W allele associates with multiple autoimmune phenotypes. *Am J Hum Genet* **76**, 561-71 (2005).
5. Demirci, F.Y.K. et al. Association of a common interferon regulatory factor 5 (IRF5) variant with increased risk of systemic lupus erythematosus (SLE). *Ann Hum Genet* **71**, 308-311 (2006).
6. Mitchell, M.K., Gregersen, P.K., Johnson, S., Parsons, R. & Vlahov, D. The New York Cancer Project: rationale, organization, design, and baseline characteristics. *J Urban Health* **81**, 301-10 (2004).
7. Hochberg, M.C. Updating the American College of Rheumatology revised criteria for the classification of systemic lupus erythematosus. *Arthritis Rheum* **40**, 1725 (1997).
8. Hom, G. et al. Association of Systemic Lupus Erythematosus with C8orf13-BLK and ITGAM-ITGAX. *N Engl J Med* (2008).
9. Taylor, K.E. et al. Specificity of the STAT4 genetic association for severe disease manifestations of systemic lupus erythematosus. *PLoS Genet* **4**, e1000084 (2008).
10. Purcell S, Neale B & Todd-Brown K, T.L., Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, Sham PC. PLINK: a toolset for whole-genome association and population-based linkage analysis. *American Journal of Human Genetics* **81**(2007).
11. Pritchard, J.K., Stephens, M. & Donnelly, P. Inference of population structure using multilocus genotype data. *Genetics* **155**, 945-59. (2000).
12. Barrett, J.C., Fry, B., Maller, J. & Daly, M.J. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* **21**, 263-5 (2005).
13. Purcell, S., Daly, M. & Sham, P. WHAP: haplotype-based association analysis. *Bioinformatics* **23**, 255-256 (2007).
14. de Bakker, P.I. et al. Efficiency and power in genetic association studies. *Nat Genet* **37**, 1217-23 (2005).
15. Seldin, M.F. et al. European population substructure: clustering of northern and southern populations. *PLoS Genet* **2**, e143 (2006).
16. Price, A.L. et al. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* **38**, 904-9 (2006).

Supplementary Table 1. Summary of genotypes by source before and after quality-control filters		
	Illumina 550K genotyped*	**Post-QC****
Cohort 1 (ABCoN and MADGC) cases	446	394
Cohort 2 (U. C. San Francisco) cases	611	564
Cohort 3 (U. Pittsburgh)cases	319	281
Total cases	**1376**	**1239**
NYHP controls	**1762**	**1629**

*After removal of duplicate samples and first-degree relatives. **After removal of subjects with < 90% genotyping or < 90% European ancestry by STRUCTURE[11] analysis.

Supplementary Table 2. Conditional tests for all SNPs with single-marker allelic p<0.005

SNP*	SNP	Location	p-value conditional on rs2230926 and rs13192841	p-value conditional on rs2230926, rs13192841, and rs6922466
2	rs6933404	138000928	0.025	0.086
4	rs600469	138003365	0.74	0.53
5	**rs13192841**	138008907	N/A	N/A
6	**rs12527282**	138008945	(collinear)	(collinear)
8	rs2327832	138014761	0.013	0.054
10	rs686851	138021664	0.79	0.47
11	rs1002658	138023277	0.31	0.59
12	rs525977	138027345	0.79	0.47
13	rs6904167	138029601	0.72	0.027
17	rs636393	138049223	0.37	0.77
18	rs602414	138053358	0.56	0.94
58	**rs2230926**	138237759	N/A	N/A
105	rs2484066	138317462	0.048	0.77
106	rs9494941	138473046	0.0019	0.15
108	rs1931867	138482531	0.0035	0.16
110	**rs6922466**	138486623	**0.00037**	N/A
111	rs12660547	138489755	0.00079	0.13
112	rs12661926	138489803	0.00078	0.13
113	rs7773257	138491248	0.022	0.72
114	rs6920846	138491762	0.0026	0.33
115	rs4896318	138492967	0.0089	0.20

Conditioned p-values obtained from Whap[13]. SNPs rs13192841 and rs12527282 are collinear, i.e. one allele determines the other in > 99% of haplotypes. *SNP number refers to order in Figure 1, containing 115 SNPs passing QC in the initial 500-kb region.

Supplementary Table 3. Multivariate logistic regression for rs13192841, rs2230926, and rs6922466 using additive model					
		Minor allele		**Risk allele**	
	p-value	**OR**	**95% CI**	**OR**	**95% CI**
rs13192841	7.9e-6	0.72	0.62 – 0.83	1.39	1.20 – 1.61
rs2230926	0.0016	1.88	1.27 – 2.79	1.88	1.27 - 2.79
rs6922466	0.00039	0.76	0.65 – 0.88	1.32	1.13 – 1.54

Interaction terms were insignificant by log ratio testing (not shown)

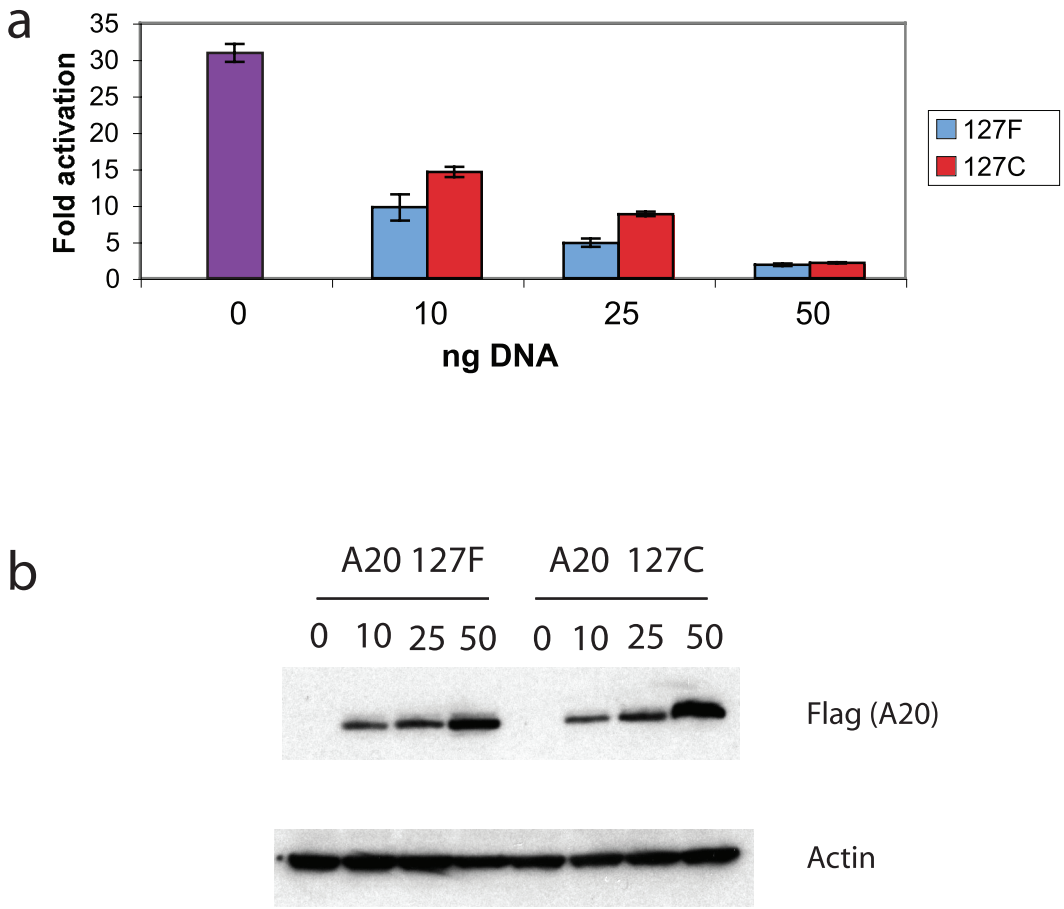
Supplementary Table 4. Associations between TNFAIP3 SNPs and SLE by ancestry strata and combined using allelic model

	rs13192841		
Subgroup (n=called genotypes)	**p-value**	**OR**	**heterogeneity p-value**[†]
All combined raw (n=2731)	6.10E-08	0.71 (0.63 - 0.81)	-
North European* > 90% (n=1456)	0.00069	0.75 (0.63 - 0.89)	0.46
North European* < 90% (n=1273)	3.2E-05	0.68 (0.57 - 0.82)	
Strata MH[†] combined	1.10E-07	0.72 (0.63 - 0.81)	
Homogeneous** subset (n=1191)	2.90E-05	0.67 (0.55 - 0.81)	0.067
Not in homogeneous** subset (n=1540)	0.065	0.85 (0.71 - 1.01)	
Strata MH[†] combined	0.000032	0.76 (0.67 - 0.87)	

	rs2230926		
Subgroup	**p-value**	**OR**	**heterogeneity p-value**[†]
All combined raw (n=1987)	0.00025	2.01 (1.36 - 3.03)	-
North European* > 90% (n=923)	0.025	2.07 (1.07 - 4.37)	0.88
North European* < 90% (n=1063)	0.0073	1.94 (1.16 - 3.30)	
Strata MH[†] combined	0.00048	1.99 (1.34 - 2.94)	
Homogeneous subset** (n=959)	0.15	1.53 (0.84 - 2.96)	0.34
Not in homogeneous**s subset (n=1028)	0.0021	2.23 (1.29 - 3.94)	
Strata MH[†] combined	0.0013	1.87 (1.26 - 2.77)	

	rs6922466		
Subgroup	**p-value**	**OR**	**heterogeneity p-value**[†]
All combined raw (n=2828)	0.00012	0.78 (0.69 - 0.89)	-
North European* > 90% (n=1502)	0.0016	0.76 (0.64 - 0.90)	0.52
North European* < 90% (n=1324)	0.035	0.82 (0.68 - 0.99)	
Strata MH[†] combined	0.00018	0.79 (0.70 - 0.89)	
Homogeneous subset** (n=1233)	0.0008	0.72 (0.60 - 0.88)	0.15
Not in homogeneous** subset (n=1595)	0.13	0.87 (0.73 - 1.05)	
Strata MH[†] combined	0.00076	0.80 (0.70 - 0.91)	

* based on STRUCTURE[11] analysis and 1409 EUROSTRUCTURE AIMs[15]
**based on 4 principal components from EIGENSTRAT[16] analysis with 550K data[9]
[†]Mantel-Haenszel combined odds ratios (OR), p-values, and test of heterogeneity of the stratum-specific associations.



Supplementary Figure 1. Decreased inhibition of TNF induced NFkB response by A20 127C protein corresponding to rs2230926 minor allele. Cells were co-transfected with NFkB-luciferase, CMV-renilla, and the indicated amounts of TNFAIP3 expression constructs encoding either 127F or 127C alleles. (a) NFκB dependent luciferase activity, normalized to CMV-renilla activity in each sample, was measured after stimulation with TNF and the ratio of induced luciferase activity (compared to samples without TNF) is indicated. Cells bearing the minor 127C allele exhibit consistently less inhibition of TNF induced NFκB activity than cells expressing the 127F allele. P values for the difference between 127F and 127C are: 10 ng DNA (p=0.0117); 25 ng (p=0.0005); 50 ng (p=0.067). Error bars represent the standard deviation from three independent experiments. (b) Immunoblots showing levels of A20 and actin proteins. Relative A20 protein expression levels normalized to actin protein expression levels for the 127F vs. 127C alleles were: 10 ng (0.61 vs. 0.62); 25 ng (0.72 vs. 0.87); 50 ng (1.25 vs. 1.26).