Published in final edited form as:

Ann Rheum Dis. 2009 November ; 68(11): . doi:10.1136/ard.2008.097642.

STAT4 Associates with SLE Through Two Independent Effects that Correlate with Gene Expression and Act Additively with IRF5 to Increase Risk

Anna-Karin Abelson^{1,§}, Angélica M. Delgado-Vega^{1,§}, Sergey V. Kozyrev^{1,§}, Elena Sánchez^{2,§}, Rafael Velázquez-Cruz³, Niclas Eriksson⁴, Jerome Wojcik⁵, Prasad Linga Reddy¹, Guadalupe Lima⁶, Sandra D'Alfonso⁷, Sergio Migliaresi⁸, Vicente Baca⁹, Lorena Orozco³, Torsten Witte^{10,*}, Norberto Ortego-Centeno¹¹, the AADEA group^{*}, Hadi Abderrahim⁵, Bernardo A. Pons-Estel^{12,**}, Carmen Gutiérrez¹³, Ana Suárez¹³, Maria Francisca González-Escribano¹⁴, Javier Martín^{2,§}, and Marta E. Alarcón-Riquelme^{1,15,§}

¹Department of Genetics and Pathology, Rudbeck Laboratory, University of Uppsala, Uppsala, Sweden ²Instituto de Biomedicina y Parasitología "López-Neyra", Consejo Superior de Investigaciones Científicas, Granada, Spain ³Instituto Nacional de Medicina Genómica, Mexico City, Mexico ⁴Uppsala Clinical Research Center, Uppsala University, Uppsala, Sweden ⁵Merck Serono, Chemin des Mines, Geneva, Switzerland ⁶Department of Immunology and Rheumatology, Instituto Nacional de Ciencias Médicas y Nutrición "Salvador Zubirán" Mexico City, Mexico ⁷Department of Medical Sciences and IRCAD, University of Eastern Piedmont, Novara, Italy ⁸Rheumatology Unit, Second University of Naples, Naples, Italy ⁹Department of Rheumatology, Pediatric Hospital, Centro Medico Nacional Siglo XXI, IMSS, Mexico City, Mexico ¹⁰Medical School Hannover, Hannover, Germany ¹¹Department of Internal Medicine, Hospital Clínico San Cecilio, Granada, Spain ¹²Sanatorio Parque, Rosario, Argentina ¹³Department of Functional Biology, Hospital Universitario Central de Asturias, Universidad de Oviedo, Oviedo, Spain ¹⁴Department of Immunology, Hospital Virgen del Rocío, Sevilla, Spain ¹⁵Oklahoma Medical Research Foundation, Oklahoma City, USA

Abstract

Objectives—To confirm and define the genetic association of *STAT4* and systemic lupus erythematosus, investigate the possibility of correlations with differential splicing and/or expression levels, and genetic interaction with *IRF5*.

Corresponding author: Marta E. Alarcón-Riquelme, MD, PhD., Uppsala University, The Rudbeck Laboratory, Dag Hammarskjölds väg 20, SE-75185, UPPSALA (Sweden). Marta.alarcon@genpat.uu.se, Phone: +46 18 4714805, Fax: +46 18 4714808.

*The AADEA (Andalusian Association of Autoimmune Diseases) group participants are: José L Callejas-Rubio, Servicio de Medicina Interna, Hospital Clínico San Cecilio, Granada; Juan Jiménez-Alonso and Mario Sabio, Servicio de Medicina Interna, Hospital Virgen de las Nieves, Granada; Julio Sánchez-Román and Francisco J. García-Hernández, Servicio de Medicina Interna, Hospital Virgen del Rocío, Sevilla; Enrique De-Ramon and Mayte Camps, Servicio Medicina Interna, Hospital Carlos Haya, Málaga; Rosa García-Portales, Servicio Reumatología, Hospital Virgen de la Victoria, Málaga; Miguel A López-Nevot, Servicio de Inmunología, Hospital Virgen de las Nieves, Granada.

**Lists of clinical collaborators involved in the collection of patients and controls for this study are listed in the acknowledgements. Bernardo Pons-Estel is the coordinator of the Argentine collaborative group.

§These authors, at their respective positions, contributed equally to this work

Conflict of Interest

JW and HA are employees of Merck Serono Inc. and produced the Argentine 100k data on which our investigation and search for *STAT4* variants was first based.

Publisher's Disclaimer: The Corresponding Author has the right to grant on behalf of all authors and does grant on behalf of all authors, an exclusive license (or non-exclusive for government employees) on a worldwide basis to the BMJ Publishing Group Ltd and its Licensees to permit this article to be published in *Annals of the Rheumatic Diseases* editions and any other BMJPL products to exploit all subsidiary rights, as set out in our license <http://ard.bmjournals.com/fora/licence.pdf>

Methods—30 tag SNPs were genotyped in an independent set of Spanish cases and controls. SNPs surviving correction for multiple tests were genotyped in 5 new sets of cases and controls for replication. *STAT4* cDNA was analyzed by 5'-RACE PCR and sequencing. Expression levels were measured by quantitative PCR.

Results—In the fine-mapping, four SNPs were significant after correction for multiple testing, with rs3821236 and rs3024866 as the strongest signals, followed by the previously associated rs7574865, and by rs1467199. Association was replicated in all cohorts. After conditional regression analyses, two major independent signals represented by SNPs rs3821236 and rs7574865, remained significant across the sets. These SNPs belong to separate haplotype blocks. High levels of *STAT4* expression correlated with SNPs rs3821236, rs3024866 (both in the same haplotype block) and rs7574865 but not with other SNPs. We also detected transcription of alternative tissue-specific exons 1, indicating presence of tissue-specific promoters of potential importance in the expression of *STAT4*. No interaction with associated SNPs of *IRF5* was observed using regression analysis.

Conclusions—These data confirm *STAT4* as a susceptibility gene for SLE and suggest the presence of at least two functional variants affecting levels of *STAT4*. Our results also indicate that both genes *STAT4* and *IRF5* act additively to increase risk for SLE.

Keywords

Association studies; systemic lupus erythematosus; *STAT4* transcription factor; Interferon regulatory factor; genetic predisposition to disease

Introduction

Systemic Lupus Erythematosus (SLE) has a strong genetic component supported by high familial aggregation and twin and family studies [1]. Like most autoimmune diseases, the HLA has an important contribution [2-4] and we recently showed that the HLA is the strongest genetic factor in individuals of European ancestry followed by *IRF5* and *ITGAM* [2-5]. Other somewhat weaker but well established associations have been found to *FCGR1A* [6], *PTPN22* [7], *PDCD1* [8], *TNFSF4* [9], *BLK* [2, 3] and most recently *BANK1* [10]. A genetic association with the signal transducer and activator of transcription 4 (*STAT4*) was identified in rheumatoid arthritis (RA) with SNP rs7574865 and this association was also found in SLE [11]. From the RA studies, the genetic association was defined to the 3rd intron of *STAT4*. As has been shown for *IRF5*, several polymorphisms may contribute to the risk, and the risk might also differ among populations [5]. Thus, our aim with the present study has been to revise the *STAT4* genetic association using independent sets of Europeans and Latin American populations with a dense set of tag SNPs to define if rs7574865 and thus, the 3rd intron signal is the sole genetic contributor to susceptibility in *STAT4*.

STAT4 is a critical regulator of immune responses, primarily induced by the dendritic cell-produced IL-12 leading to the development of Th1 cells, which have the capacity to secrete high levels of IFN- γ . *STAT4* is activated after ligation of IL-12 to its receptor, which associates with the tyrosine kinases Tyk2 and Jak2 [12, 13]. These are expressed in activated T and B cells and particularly NK cells. Activation of *STAT4* leads to the formation of homodimers of *STAT4* that translocate into the nucleus and induce transcription of IFN- γ . In addition, *STAT4* activation is also induced by IFN α/β stimulation. This stimulation does not appear to lead to Th1 development, but only to an acute IFN- γ secretion by CD4⁺ T and NK cells where IL-18 is also required. IFN α/β induces *STAT4* phosphorylation through direct interaction of *STAT4* with the IFN α R2 subunit.

Here, we fine mapped a Spanish set of cases and controls. We found evidence for another peak of association beyond the intron 3 SNP rs7574865. This association was replicated in 4 independent sets of cases and controls. We also find evidence for a correlation between the associated SNPs and expression levels of *STAT4* in PBMCs. When we analyzed the possibility of genetic interaction between *STAT4* and *IRF5*, we found no interaction, but rather an additive increase in the risk for SLE.

Material and Methods

Patients and controls

In total, 1581 cases and 1844 controls were used in this study, all with complete data for the SNPs analyzed: 390 cases and 620 controls from Spain, 247 cases and 220 controls from Germany, 221 patients and 207 controls from Italy, 171 patients and 171 controls from Argentina, and 231 cases and 250 controls from Mexico (adults) along with a set of 321 pediatric patients and 383 adult controls. The patient and control sets studied here have been described previously [10, 14]. All patients fulfilled the 1982 ACR criteria for classification of SLE.

Selection of Tag SNPs

29 Tag SNPs covering the *STAT1* and *STAT4* genes, and the intergenic region, were selected using Haploview version 3.32 from the HapMap-CEU population genotype data. Aggressive tagging mode was used to select tag SNPs with a minor allele frequency $\geq 5\%$, with an r^2 threshold ≥ 0.8 . Rs7574865 was added to the tag list after being reported as associated with RA [11]. SNPs associated in the Spanish fine-mapping, after quality control and correction for multiple testing, were typed in the German, Italian, Argentinean, and both Mexican sets.

Genotyping

Spanish samples were genotyped in Granada using TaqMan® 5' exonuclease assay (ABI, Foster City, CA). German, Italian, and Latin American samples were genotyped at Uppsala University. Mexican pediatric samples were genotyped at Instituto Nacional de Medicina Genómica using the same method. Genotyping consistency between the centers was established to be near to 100% [15].

Statistical Analyses

The Spanish genotype data was processed using Haploview version 4.0 [16], PLINK version 1.02 [17], and R language. Quality control filters were applied to remove SNPs with $>10\%$ missing data in cases or controls (1 SNP excluded), deviations from Hardy-Weinberg equilibrium ($p < 0.001$, 2 SNPs excluded), or minor allele frequency $<5\%$ in controls (2 SNPs excluded). 25 SNPs remained. Subjects with an individual missing genotyping rate $>10\%$ were also removed ($n=42$). Genotyping rate in remaining individuals was 97.4%. Pairwise linkage disequilibrium (LD) measures (D' and r^2) between SNPs and maximum-likelihood haplotype frequencies were estimated with the EM algorithm. Multiple testing was corrected by Bonferroni and false discovery rate (FDR) methods [18].

Statistical analysis of the replication sets included only subjects with 100% individual genotyping calls. DerSimonian-Laird and Cochran-Mantel-Haenszel methods implemented in StatsDirect and PLINK were used to estimate the pooled odds ratio for all populations assuming random and fixed effects models on the allelic association, respectively. These were adjusted by adding the stratification variable "Population" to the logistic regression model containing the genotype as the exposure variable. This test is identical to the 1 degree of freedom Mantel-Haenszel test of the hypothesis that the stratum specific odds ratios are 1

[19]. The heterogeneity test based on partitioning the chi-square statistic implemented in PLINK, was used to test between-population differences. Univariate genotypic odds ratios were estimated by logistic regression [20]. Conditional logistic regression was used to determine independency of the SNPs from rs7574865. All logistic regression analyses were done with R.

Multiple logistic regression was used to evaluate if additive or interaction effects were present between SNPs within *STAT4* and *IRF5*. To measure the ability to discriminate between SLE cases and controls, the area under the receiver-operating characteristic (ROC) curve (C-statistic) was calculated. To statistically compare the C-statistics we applied the method by DeLong et al. [21]

5'-RACE and 3'-RACE PCR

Marathon-Ready cDNA from different tissues (Clontech) was used as template for amplification of tissue-specific 5'- and 3'-UTRs. The following pairs of nested gene-specific primers were used for 5'-RACE PCR: 5'-GAAATTCTACTGAGAGACTCCCATTG-3' and 5'-GAATCGTTGCCATGGTTTCATTGTTAG-3' and for 3'-RACE PCR: 5'-CTAAACTATCAGGTAAAGGTTAAGGCATC-3' and 5'-GGTAAACACTACAGCTCTCAGCCTTG-3'. Adapter primers Ap1 and Ap2 were provided with cDNA. Nested PCR was carried out using 1/30 of the first round PCR products. 35 cycles of PCR (95°C for 20 s, 60°C for 15 s and 72°C for 3 min) were performed after initial denaturation at 95°C for 5 min in buffer containing 1.5 mM MgCl₂, 200 μM of each of dNTPs, 0.4 μM of each of the corresponding primers, and 0.5 U of Platinum Taq high fidelity enzyme (Invitrogen). PCR products were analysed by sequencing.

RNA purification and STAT4 expression analysis

Total RNA was purified as described elsewhere [10] from peripheral blood mononuclear cells (PBMCs) obtained with agreed consent from 73 healthy volunteers. 2 μg of RNA was reverse transcribed using oligo-dT primers according to the manufacturer's protocol (Applied Biosystems).

STAT4 expression was determined by real-time PCR using SYBR Green detection. Cycling conditions were as follows: 95°C for 5 min, 45 cycles of PCR (95°C for 15 s, 60°C for 10 s and 72°C for 20 s). α-isoform was detected with the following primers: forward 5'-CATCTCAACAATCCGAAGTGATTCA-3' and common reverse primer 5'-GTCAGAGTTTATCTGTTCATTGAGCAG-3'. β-isoform-specific forward primer was 5'-TGACCTTGTTATCTCTTTAAGCCGA-3'. Expression levels were normalized against TATA-binding protein (TBP) gene expression amplified with commercial reagents (Applied Biosystems). All experiments were run in triplicate.

Statistical analysis of gene expression

ANOVA and F-test were used to determine the difference in the mRNA expression level in relation to each of the SNPs, taking the three possible genotypes as factor levels. Significance of gene expression was also tested with linear regression using ΔT values as a continuous trait and WHAP for regression with individual SNP alleles [22].

Results

Association was detected for several SNPs across *STAT4* and the *STAT1-STAT4* intergenic region, with the strongest association with rs3821236 ($p=7.07\times10^{-8}$), followed by rs3024866 ($p=3.83\times10^{-7}$), rs7574865 ($p=9.37\times10^{-6}$) and rs1467199 ($p=7\times10^{-5}$), all of which remained

significant after correction for multiple tests (Table 1). The region LD structure defined six haplotype blocks, two located in *STAT1* (blocks 1-2), one in the intergenic region (block 3) and three in *STAT4* (blocks 4-6) (Figure 1 and Supplementary Figure 1). Our results indicate association of two additional haplotype blocks than that containing rs7574865 (block 6): block 3, which contains rs1467199, and block 4 harboring rs3821236 and rs3024866 (Figure 1).

To replicate the genetic associations and increase statistical power, we genotyped the associated SNPs in 5 independent sets from Italy, Germany, Argentina, and Mexico (one adult and one pediatric set). Homogeneity test showed combinability of the odds ratios for the SNPs except rs3024866 and rs7574865 that had some heterogeneity across the strata. None of the SNPs provided association with the German population except for a borderline association with rs7574865 (Tables 2 and Supplementary Table 1). The Mexican pediatric set showed association only with rs1467199 ($p=0.008$, Supplementary Table 2). Despite heterogeneity ($p=0.01$) the German set was included in the meta-analysis. The Mexican pediatric set was not included due to lack of statistical combinability (compare Supplementary Table 2 with Table 2). In the meta-analysis, SNPs rs3821236 ($p=5.96 \times 10^{-20}$) and rs7574865 ($p=4.44 \times 10^{-23}$) showed the strongest association across all strata in the allelic and genotypic tests, but rs3024866 was also associated ($p=2.31 \times 10^{-12}$) (Table 2). Although rs1467199 reaches significant association in the meta-analysis, at the individual population level it was only replicated in the Argentine set (rs1467199-CG $p=1.63 \times 10^{-2}$, rs1467199-GG $p=0.053$) (Supplementary Table 1).

We tested whether the associated SNPs constituted independent effects. Rs7574865 is not in strong LD with rs3024866 ($R^2=0.29$) or rs3821236 ($R^2=0.42$), which are located 62 kb and 42 kb from rs7574865, respectively, and therefore they are not proxies. Rs1467199 has $R^2=0.30$ with rs3821236, $R^2=0.18$ with rs3024866 and $R^2=0.13$ with rs7574865 (Figure 1). The low pairwise correlation coefficients (R^2) suggest that the individual SNP associations reflect independent effects, except for rs3024866, which has relatively high R^2 with rs3821236 ($R^2=0.64$) (Figure 1). This was confirmed by conditional logistic regression analysis: rs3821236 remained significant when conditioning on rs7574865, and vice versa (Supplementary Table 3). Thus, rs3821236 and rs7574865 represent two independent genetic effects within *STAT4*. rs3821236 is located in intron 16th, ~60 kb downstream from the 3rd intron where the association has been confined in previous studies [11]. SNPs rs3821236 and rs3024866 are tagging a 26kb haplotype block (block 4) covering part of the gene between intron 8 to 16 (Figure 1 and Supplementary Figure 1) that contain 3 of the 6 markers associated with SLE in the Spanish fine-mapping after multiple test correction (Table 1). After conditional haplotype-based association test the haplotype block 4 (SNPs rs3821236-rs932169) remain significant ($p=0.002277$) after controlling for rs7574865 confirming its independency.

Using the ROC curve and c-statistic, we observed that rs7574865 was a somewhat better risk predictor ($c=0.590$) than rs3821236 ($c=0.577$), but the difference did not reach statistical significance (Supplementary Table 4). Thus, each SNP alone provides the same level of prediction for risk for SLE, independently.

Statistical analysis of the Interaction with *IRF5*

IRF5 was found as a well-established non-MHC association with SLE. To study if the effect of *STAT4* was independent of *IRF5*, or if there were epistatic effects between SNPs in these genes we performed multiple logistic regression analysis and c-statistic including the genotype data for 3 SNPs of *IRF5* (rs2004640, rs2070197, and rs10954213) from our previous studies [5, 14]. The analyses revealed no significant interaction effects between

STAT4 and *IRF5* SNPs, but close to complete independent effects on SLE risk as shown by a slight change in estimates when combined in a multivariate model (data not shown).

The three SNPs of *IRF5* had c-statistics of 0.587 0.585 and 0.565, respectively (Supplementary Table 5). Rs2070197 tags the major risk haplotype and has the highest predictive ability [14]. Addition of rs7574865 to the *IRF5* SNPs increased the predictive value of the models significantly (rs7574865 with rs2004640: c-statistic=0.632, $P=1.66\times 10^{-5}$; rs7574865 with rs2070197: C-statistic = 0.636, $P=3.28\times 10^{-11}$ and rs7574865 with rs10954213: C-statistic=0.624, $P=9.62\times 10^{-7}$). The results support an additive effect of SNPs of both genes, particularly rs7574865 and rs2070197 to increase risk for SLE (Figure 2).

Tissue-specific alternative transcripts of *STAT4*

In order to investigate if differential splicing could be related to the genetic association of *STAT4*, functional annotation of gene transcripts expressed in different tissues was performed. Until now, two isoforms have been described for *STAT4*, α and β [23]. Only the known α and β isoforms were detected when we tested spleen testis, kidney, lung, pancreas, intestine and uterus for new isoforms.

Since 5' - and 3' -UTRs may substantially affect gene expression [14], we performed detailed analysis of non-coding 5' -exons and 3' -UTRs. The pattern of 5'UTRs was diverse in different tissues (Table 3), but all of them lead to the known isoforms of *STAT4*.

Analysis of *STAT4* expression levels

Levels of *STAT4* gene expression in human mononuclear cells were assessed by quantitative real-time PCR. Since expression of β transcript was much lower than of the α transcript and its expression followed the trend of the α -transcript in all samples, it was excluded from the analysis. By using ANOVA to test if differences in gene expression correlated with genotypes, we found a modest up regulation of *STAT4* mRNA and the CC risk genotype of rs3024866 ($p = 0.0371$). By regression analysis, risk alleles of SNPs rs3821236 and rs7574865 correlated with higher *STAT4* expression. Importantly, SNP rs1467199 did not correlate with *STAT4* expression (Figure 3).

Discussion

We confirmed the genetic association between *STAT4* and SLE and through a fine-mapping effort we identified a second effect independent of rs7574865 located in the vicinity of intron 16.

The publication of Remmers, et al, [11] identifying *STAT4* as a susceptibility gene for RA and SLE was published while our fine-mapping effort was ongoing, and after having identified a strong signal for this gene in a 100k scan in Argentine individuals (data not shown). The data from the Spanish mapping suggested the presence of several peaks independent of rs7574865. Analyses using a larger set of joined samples confirmed the independent effect of rs3821236 in intron 16. Conditional SNP and haplotype regression analysis supported this result.

We observed differences between North and South European samples: the German set contributed, albeit weakly, to the association only at rs7574865, whereas the association of rs3821236 was contributed particularly by the Spanish, Italian and Latin American sets. In our view it is highly plausible that several independent risk haplotypes are involved in disease susceptibility, with some having stronger effects in some populations than others.

The weak result of the German set could not be due to a lack of power, as the size of the other sets is comparable.

We are approaching a phase in complex disease genetics where identification of the genes involved in disease susceptibility is becoming a reality. Therefore, it is of interest that we understand the relationship between the various genetic effects. Here we examined the possibility of genetic interaction between *IRF5* and *STAT4*. We observed no genetic interaction but we observed a significant increase in the predictive ability for SLE when *STAT4* and *IRF5* SNPs were added. These results suggest that the *IRF5* and *STAT4* SNPs act additively to increase risk for SLE. While this work was under revision, an independent study corroborated this additive effect [24].

Identification of the functional variants for *STAT4* might prove relatively difficult considering the large size of this gene and the location of the associated SNPs. Complete resequencing of the gene and genotyping may be needed to reveal the true functional variants. We observed no differences in splicing of the gene in PBMCs. Instead we found a correlation between expression levels of *STAT4* in PBMCs and the *STAT4* associated SNPs, rs7574865, rs3024866 and rs3821236 but not rs1467199. It should be noted that PBMC samples are expected to have varying numbers of T and NK cells, high variation in gene expression, and a weak correlation. Further, in complex diseases we also expect effects caused by functional variants to be modest and difficult to define [14].

How does a modest increase in *STAT4* expression contribute to the risk for SLE? *STAT4* is a transcription factor through which the functional effects of IL-12 are conducted, leading to IFN- γ production. The *STAT4* pathway has been studied in mice and patients with SLE, with contradictory results. In two studies, mice deficient for *STAT4* show increased development of glomerulonephritis [25, 26], whereas a third study reported opposite findings in line with our observations [27].

We detected only two isoforms of *STAT4*, one of which was expressed at extremely low levels (β), and tissue-specific promoters as observed by the presence of alternative, tissue-specific 5'-UTRs (Table 3). Given that such promoters could adjust transcription of the gene in a particular tissue, this poses an additional obstacle for defining the role of genetic variation. Other authors showed that the *STAT4*'s SLE risk haplotype is over-expressed in mesenchymal cells [24] supporting a tissue-specific component involved in the regulation of *STAT4* expression. Using purified cell populations (e.g. NK cells, which constitute only 5% of the blood leukocytes but have high basal level of *STAT4*, kidney mesangial cells, etc) may be critical for correct assessment of the altered levels of *STAT4* gene expression as well as definition of the splicing isoforms that could be involved.

Major effects of the expression of human *STAT4* may be important in the kidney. Activation of *STAT4* leads to increased expression of IFN- γ . A study has shown that increased expression of IL-12 and IFN- γ in the kidneys of MRL-*lpr/lpr* mice precedes the development of glomerulonephritis [28, 29]. Thus, the localized action of the various genes in specific tissues may be of importance. This is also important to consider in view of the recent results showing a strong correlation between rs7574865 and end-organ disease, in particular kidney disease [30].

Our results do support an important role of *STAT4* in SLE susceptibility, a role that appears to vary between different populations and deriving from two different and independent risk variants, whose functional nature needs to be addressed.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors would like to express their sincere gratitude to Susanna Lewén for help with purification of PBMCs and total RNA, and Hong Yin for help with the preparation of DNA samples. We also thank Adriana I. Scollo, Armando M. Perichon and Mariano C.R. Tenaglia. CEDIM. Diagnóstico Molecular y Forense SRL. Rosario, Argentina, for their help in DNA preparation of the Argentine samples. The authors would like to thank particularly the Lupus Patient Association of Asturias for their help in the collection of samples as well as to all the patients for their contribution.

The Argentine Collaborative Group Participants are:

Hugo R. Scherbarth MD, Pilar C. Marino MD, Estela L. Motta MD Servicio de Reumatología, Hospital Interzonal General de Agudos “Dr. Oscar Alende”, Mar del Plata, Argentina; Susana Gamron MD, Cristina Drenkard MD, Emilia Menso MD Servicio de Reumatología de la UHMI 1, Hospital Nacional de Clínicas, Universidad Nacional de Córdoba, Córdoba, Argentina; Alberto Allievi MD, Guillermo A. Tate MD Organización Médica de Investigación, Buenos Aires, Argentina; Jose L. Presas MD Hospital General de Agudos Dr. Juan A. Fernandez, Buenos Aires, Argentina; Simon A. Palatnik MD, Marcelo Abdala MD, Mariela Bearzotti PhD Facultad de Ciencias Medicas, Universidad Nacional de Rosario y Hospital Provincial del Centenario, Rosario, Argentina; Alejandro Alvarellos MD, Francisco Caeiro MD, Ana Bertoli MD Servicio de Reumatología, Hospital Privado, Centro Medico de Córdoba, Córdoba, Argentina; Sergio Paira MD, Susana Roverano MD, Hospital José M. Cullen, Santa Fe, Argentina; Cesar E. Graf MD, Estela Bertero PhD Hospital San Martín, Paraná; Cesar Caprarulo MD, Griselda Buchanan PhD Hospital Felipe Heras, Concordia, Entre Ríos, Argentina; Carolina Guillerón MD, Sebastian Grimaudo PhD, Jorge Manni MD Departamento de Inmunología, Instituto de Investigaciones Médicas “Alfredo Lanari”, Buenos Aires, Argentina; Luis J. Catoggio MD, Enrique R. Soriano MD, Carlos D. Santos MD Sección Reumatología, Servicio de Clínica Medica, Hospital Italiano de Buenos Aires y Fundación Dr. Pedro M. Catoggio para el Progreso de la Reumatología, Buenos Aires, Argentina; Cristina Prigione MD, Fernando A. Ramos MD, Sandra M. Navarro MD Servicio de Reumatología, Hospital Provincial de Rosario, Rosario, Argentina; Guillermo A. Berbotto MD, Marisa Jorfen MD, Elisa J. Romero PhD Servicio de Reumatología Hospital Escuela Eva Perón. Granadero Baigorria, Rosario, Argentina; Mercedes A. Garcia MD, Juan C Marcos MD, Ana I. Marcos MD Servicio de Reumatología, Hospital Interzonal General de Agudos General San Martín, La Plata; Carlos E. Perandones MD, Alicia Eimon MD Centro de Educación Médica e Investigaciones Clínicas (CEMIC), Buenos Aires, Argentina; Cristina G. Battagliotti MD Hospital de Niños Dr. Orlando Alassia, Santa Fe, Argentina.

The German Collaborative Group Participants:

K. Armadi-Simab, MD, Wolfgang L. Gross, MD, Abteilung Rheumatologie, University Hospital of Schleswig-Holstein, Campus Luebeck, Rheumaklinik Bad Bramstedt, Luebeck, Germany, Erika Gromnica-Ihle, MD, Rheumaklinik Berlin-Buch, Berlin, Germany, Hans-Hartmut Peter, MD, Medizinische Universitätsklinik, Abteilung Rheumatologie und Klinische Immunologie, Freiburg, Germany, Karin Manger, MD, Medizinische Klinik III derFAU Erlangen-Nuernberg, Erlangen, Germany, Sebastian Schnarr, MD, Henning Zeidler, MD, Abteilung Rheumatologie, Medizinische Hochschule Hannover, Hannover, Germany, Reinhold E. Schmidt, MD, Abteilung Klinische Immunologie, Medizinische Hochschule Hannover, Hannover, Germany.

The Italian collaborative participants are: Gian Domenico Sebastiani (U.O.C. di Reumatologia Ospedale San Camillo, Roma – Italy), Enrica Bozzolo (IRCCS San Raffaele Hospital, Milan, Italy), Mauro Galeazzi, (Department of Clinical Medicine and Immunology Sciences, Section of Rheumatology, Siena University, Siena, Italy), Nadia Barizzone (Department of Medical Sciences, University of Eastern Piedmont, Novara, Italy) and Maria Giovanna Danieli and Prof. Armando Gabrielli (Dipartimento di Scienze Mediche e Chirurgiche, Università Politecnica delle Marche, Ancona, Italy).

Funding

This work has been supported in part by grants from the European CVDIMMUNE project from the European Commission LSHM-CT-2006-037227, the Swedish Research Council (12673), the Torsten and Ragnar Söderbergsstiftelse, the Swedish Association Against Rheumatism, the King Gustaf the Vth 80th-Jubilee Foundation and the Knut and Alice Wallenberg Foundation for supporting MEAR through the Royal Swedish Academy of Sciences. This study was also supported by grant SAF2006-00398 from the Spanish Ministerio de Educacion y Ciencia, grant PI052409 from the Fondo de Investigación Sanitaria (Spain), C2.12 from BMBF Kompetenznetz Rheuma in Germany and FISM, Regione Piemonte (CIPE) and the Consejo Nacional de Ciencia y Tecnología (CONACYT: SALUD-2004-01-153). MEAR is a Greenberg Scholar at the OMRF.

References

1. Alarcon-Segovia D, Alarcon-Riquelme ME, Cardiel MH, Caeiro F, Massardo L, Villa AR, et al. Familial aggregation of systemic lupus erythematosus, rheumatoid arthritis, and other autoimmune diseases in 1,177 lupus patients from the GLADEL cohort. *Arthritis Rheum.* 2005; 52:1138–47. [PubMed: 15818688]
2. Harley JB, Alarcon-Riquelme ME, Criswell LA, Jacob CO, Kimberly RP, Moser KL, et al. Genome-wide association scan in women with systemic lupus erythematosus identifies susceptibility variants in ITGAM, PXX, KIAA1542 and other loci. *Nat Genet.* 2008; 40:204–10. [PubMed: 18204446]
3. Hom G, Graham RR, Modrek B, Taylor KE, Ortmann W, Garnier S, et al. Association of systemic lupus erythematosus with C8orf13-BLK and ITGAM-ITGAX. *N Engl J Med.* 2008; 358:900–9. [PubMed: 18204098]
4. Fernando MM, Stevens CR, Sabeti PC, Walsh EC, McWhinnie AJ, Shah A, et al. Identification of Two Independent Risk Factors for Lupus within the MHC in United Kingdom Families. *PLoS Genet.* 2007; 3:e192. [PubMed: 17997607]
5. Graham RR, Kozyrev SV, Baechler EC, Reddy MV, Plenge RM, Bauer JW, et al. A common haplotype of interferon regulatory factor 5 (IRF5) regulates splicing and expression and is associated with increased risk of systemic lupus erythematosus. *Nat Genet.* 2006; 38:550–5. [PubMed: 16642019]
6. Magnusson V, Johanneson B, Lima G, Odeberg J, Alarcon-Segovia D, Alarcon-Riquelme ME. Both risk alleles for FcgammaRIIA and FcgammaRIIIA are susceptibility factors for SLE: a unifying hypothesis. *Genes Immun.* 2004; 5:130–7. [PubMed: 14737097]
7. 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
8. Prokunina L, Castillejo-Lopez C, Oberg F, Gunnarsson I, Berg L, Magnusson V, et al. A regulatory polymorphism in PDCD1 is associated with susceptibility to systemic lupus erythematosus in humans. *Nat Genet.* 2002; 32:666–9. [PubMed: 12402038]
9. Graham DS, Graham RR, Manku H, Wong AK, Whittaker JC, Gaffney PM, et al. Polymorphism at the TNF superfamily gene TNFSF4 confers susceptibility to systemic lupus erythematosus. *Nat Genet.* 2008; 40:83–9. [PubMed: 18059267]
10. Kozyrev SV, Abelson AK, Wojcik J, Zaghlool A, Linga Reddy MV, Sanchez E, et al. Functional variants in the B-cell gene BANK1 are associated with systemic lupus erythematosus. *Nat Genet.* 2008; 40:211–6. [PubMed: 18204447]
11. Remmers EF, Plenge RM, Lee AT, Graham RR, Hom G, Behrens TW, et al. STAT4 and the risk of rheumatoid arthritis and systemic lupus erythematosus. *N Engl J Med.* 2007; 357:977–86. [PubMed: 17804842]
12. Kaplan MH. STAT4: A Critical Regulator of Inflammation In Vivo. *Immunol Res.* 2005; 31:231–42. [PubMed: 15888914]
13. Yao BB, Niu P, Surowy CS, Faltynek CR. Direct interaction of STAT4 with the IL-12 receptor. *Arch Biochem Biophys.* 1999; 368:147–55. [PubMed: 10415122]
14. Kozyrev SV, Lewén S, Linga Reddy MVP, Pons-Estel BA, Witte T, Junker P, Laustrop H, Gutiérrez C, Suárez A, González-Escribano MF, Martín J, Alarcón-Riquelme ME. The Argentine Collaborative Group, The German Collaborative Group, The Spanish Collaborative Group. Structural insertion/deletion Variation in IRF5 is Associated with a Risk Haplotype and Defines the Precise Isoforms Expressed in SLE. *Arthritis and Rheumatism.* 2007; 56:1234–41. [PubMed: 17393452]
15. Sanchez E, Abelson AK, Sabio JM, Gonzalez-Gay MA, Ortego-Centeno N, Jimenez-Alonso J, et al. Association of a CD24 gene polymorphism with susceptibility to systemic lupus erythematosus. *Arthritis Rheum.* 2007; 56:3080–6. [PubMed: 17763438]
16. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics.* 2005; 21:263–5. [PubMed: 15297300]

17. 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. [PubMed: 17701901]
18. Benjamini Y, Drai D, Elmer G, Kafkafi N, Golani I. Controlling the false discovery rate in behavior genetics research. *Behav Brain Res.* 2001; 125:279–84. [PubMed: 11682119]
19. Day NE, Byar DP. Testing hypotheses in case-control studies--equivalence of Mantel-Haenszel statistics and logit score tests. *Biometrics.* 1979; 35:623–30. [PubMed: 497345]
20. Sasieni PD. From genotypes to genes: doubling the sample size. *Biometrics.* 1997; 53:1253–61. [PubMed: 9423247]
21. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics.* 1988; 44:837–45. [PubMed: 3203132]
22. Purcell S, Daly MJ, Sham PC. WHAP: haplotype-based association analysis. *Bioinformatics.* 2007; 23:255–6. [PubMed: 17118959]
23. Hoey T, Zhang S, Schmidt N, Yu Q, Ramchandani S, Xu X, et al. Distinct requirements for the naturally occurring splice forms Stat4alpha and Stat4beta in IL-12 responses. *Embo J.* 2003; 22:4237–48. [PubMed: 12912921]
24. Sigurdsson S, Nordmark G, Garnier S, Grundberg E, Kwan T, Nilsson O, et al. A risk haplotype of STAT4 for systemic lupus erythematosus is over-expressed, correlates with anti-dsDNA and shows additive effects with two risk alleles of IRF5. *Hum Mol Genet.* 2008; 17:2868–76. [PubMed: 18579578]
25. Singh RR, Saxena V, Zang S, Li L, Finkelman FD, Witte DP, et al. Differential contribution of IL-4 and STAT6 vs STAT4 to the development of lupus nephritis. *J Immunol.* 2003; 170:4818–25. [PubMed: 12707364]
26. Jacob CO, Zang S, Li L, Ciobanu V, Quismorio F, Mizutani A, et al. Pivotal role of Stat4 and Stat6 in the pathogenesis of the lupus-like disease in the New Zealand mixed 2328 mice. *J Immunol.* 2003; 171:1564–71. [PubMed: 12874250]
27. Xu Z, Duan B, Croker BP, Morel L. STAT4 deficiency reduces autoantibody production and glomerulonephritis in a mouse model of lupus. *Clin Immunol.* 2006; 120:189–98. [PubMed: 16713741]
28. Fan X, Oertli B, Wuthrich RP. Up-regulation of tubular epithelial interleukin-12 in autoimmune MRL-Fas(lpr) mice with renal injury. *Kidney Int.* 1997; 51:79–86. [PubMed: 8995720]
29. Schwarting A, Tesch G, Kinoshita K, Maron R, Weiner HL, Kelley VR. IL-12 drives IFN-gamma-dependent autoimmune kidney disease in MRL-Fas(lpr) mice. *J Immunol.* 1999; 163:6884–91. [PubMed: 10586090]
30. Taylor KE, Remmers EF, Lee AT, Ortmann WA, Plenge RM, Tian C, et al. Specificity of the STAT4 Genetic Association for Severe Disease Manifestations of Systemic Lupus Erythematosus. *PLoS Genetics.* 2008; 4:e1000084. [PubMed: 18516230]

Figure 1. Fine mapping of *STAT1-STAT4* region

The physical position (top panel) of the SNPs typed in 390 patients and 480 controls from Spain covering the ~200kb *STAT1-STAT4* region is shown. The region LD structure defined six haplotype blocks (middle panel) of which three were associated with disease susceptibility (See block structure and R^2 values in Supplementary Figure 1). Risk haplotypes are shown in red and main single-marker hits replicated are underlined (See Tables 2 and Supplementary Table 1). The bottom panel shows the significance of the association data presented as the $-\log_{10}(P\text{-value})$ for 25 tag SNPs passing genotyping quality control. The blocks have been defined using the solid spine of LD method in Haploview v4.0. The p values of the risk and protective haplotypes at each block in the Spanish population are as follows:

- ¹ Block 3 risk haplotype p-value = 0.0037
- ² Block 3 protective haplotype p-value = 7×10^{-4}
- ³ Block 4 risk haplotype p-value = 2.64×10^{-6}
- ⁴ Block 4 protective haplotype p-value = 2.24×10^{-5}
- ⁵ Block 6 risk haplotype p-value = 9.52×10^{-6}
- ⁶ Block 6 protective haplotype p-value = 0.0331

Figure 2. Predictive ability of *STAT4* and *IRF5* SNPs for SLE

The predictive ability of the two genes for SLE was investigated using the c-statistic. For comparison of the c-statistics the test for comparing two dependent ROC curves is used. Within *STAT4*, the SNP rs7574865 is the strongest predictor for SLE and adds a significant fraction to the predictive ability of the *IRF5* SNP rs2070197. Between genes the best combination is rs7574865+rs2070197 having an overall c-statistic of 0.636.

Figure 3. Association between *STAT4* polymorphism and expression levels

ANOVA and F-test were used to determine difference in the mRNA relative expression levels between individuals carrying different genotypes, taking the three possible genotypes as factor levels and the major allele homozygous as reference. Multiple comparisons of means revealed that individuals with rs3024866-CC genotype have higher *STAT4* expression levels compared to the reference level (p-value=0.0281)

Table 1

Results from the fine mapping conducted in Spanish patients with SLE and matched controls.

Chr.	Position	SNP	Cases aa/Aa/AA	Controls aa/Aa/AA	Assoc allele	Frequency Cases	Frequency Controls	Genotypic test p-value	Allelic test p-value	OR	L95	U95	<u>Multiple test correction</u> Bonferroni	FDR
2	191546759	rs13395505	83/189/110	90/222/160	A	0.46	0.43	0.252	0.108	1.17	0.97	1.42	1.000	0.574
2	191553970	rs1547550	70/165/145	52/225/195	G	0.40	0.35	9,55E-03	2,50E-02	1.25	1.03	1.53	0.625	0.184
2	191554382	rs4327257	8/81/291	15/111/346	C	0.13	0.15	0.459	0.198	0.83	0.63	1.10	1.000	0.946
2	191558344	rs2280234	88/169/124	75/224/171	A	0.45	0.40	3,07E-02	2,26E-02	1.25	1.03	1.52	0.566	0.184
2	191558811	rs2280233	68/161/141	92/216/148	C	0.40	0.44	0.243	0.127	0.86	0.70	1.05	1.000	0.640
2	191559011	rs2280232	29/131/219	28/168/240	C	0.25	0.26	0.450	0.727	0.96	0.77	1.20	1.000	1.000
2	191568747	rs12693591	5/102/268	14/143/311	A	0.15	0.18	0.137	0.069	0.79	0.61	1.02	1.000	0.385
2	191574903	rs13029247	56/138/139	48/197/216	C	0.38	0.32	2,76E-02	1,70E-02	1.29	1.05	1.59	0.424	0.180
2	191577408	rs2030171	69/172/133	58/202/194	C	0.41	0.35	2,99E-02	7,35E-03	1.31	1.08	1.60	0.184	0.088
2	191588747	rs1467199	47/146/167	24/183/265	G	0.33	0.24	7,45E-05	7,00E-05	1.54	1.25	1.91	1,75E-03	1,67E-03
2	191604299	rs3024935	3/48/323	3/69/385	T	0.07	0.08	0.647	0.455	0.87	0.60	1.25	1.000	1.000
2	191605785	rs925847	50/171/158	46/196/230	T	0.36	0.31	0.076	2,21E-02	1.27	1.04	1.55	0.552	0.184
2	191611003	rs3821236	42/151/183	18/151/302	A	0.31	0.20	5,90E-07	7,07E-08	1.84	1.47	2.29	1,77E-06	6,74E-06
2	191613134	rs3024877	66/181/132	49/219/203	T	0.41	0.34	3,50E-03	1,19E-03	1.39	1.14	1.69	2,96E-02	2,26E-02
2	191625589	rs16833220	8/77/254	10/134/293	G	0.14	0.18	4,39E-02	3,71E-02	0.74	0.56	0.98	0.927	0.236
2	191631086	rs3024866	60/171/151	31/189/252	C	0.38	0.27	1,95E-06	3,83E-07	1.70	1.38	2.09	9,58E-06	1,83E-05
2	191637523	rs932169	3/55/321	5/59/382	C	0.08	0.08	0.781	0.815	1.04	0.73	1.50	1.000	1.000
2	191639709	rs1517352	102/159/103	75/241/137	A	0.50	0.43	3,12E-04	6,88E-03	1.31	1.08	1.59	0.172	0.088
2	191646845	rs7594501	0/37/335	3/65/403	A	0.05	0.08	0.057	3,28E-02	0.64	0.43	0.97	0.819	0.223
2	191651517	rs3024921	1/37/338	3/43/423	T	0.05	0.05	0.752	0.972	0.99	0.64	1.53	1.000	1.000
2	191662292	rs10931480	14/95/270	21/151/298	G	0.16	0.21	0.054	2,35E-02	0.75	0.58	0.96	0.586	0.184
2	191663097	rs10931481	58/174/150	42/207/221	G	0.38	0.31	6,95E-03	2,43E-03	1.36	1.12	1.67	0.061	3,87E-02
2	191664494	rs13011805	1/63/314	6/96/367	T	0.09	0.12	0.100	4,88E-02	0.72	0.52	1.00	1.000	0.291
2	191672878	rs7574865	41/153/181	18/170/284	T	0.31	0.22	1,74E-05	9,37E-06	1.64	1.31	2.03	2,34E-04	2,98E-04
2	191743288	rs10176621	29/146/205	33/148/262	C	0.27	0.24	0.294	0.212	1.15	0.92	1.44	1.000	0.961

The table shows the genotypes counts for the 25 SNPs that passed frequency and genotyping pruning, as well as the results for both allelic and genotypic association tests. The four SNPs highlighted where chosen to be replicated in independent sets of cases and controls since they had a significant p-value < 1.00E-05 and remained associated after multiple tests correction and were independent. OR: odds ratio, L95: lower limit of the 95% confidence interval, U95: upper limit of the 95% confidence interval; FDR: false discovery rate.

Table 2

Population-specific replication and general stratified allelic association analysis of the main associated SNPs in the fine mapping.

SNP	Risk Allele	Population/Test	Frequency Cases	Frequency controls	P-value	OR
rs1467199	G	Germany	0.229	0.231	0.941	0.99
		Italy	0.235	0.229	0.840	1.03
		Spain	0.333	0.250	7.73E-05	1.50
		Argentina	0.345	0.284	0.084	1.33
		Mexico *	0.171	0.134	0.121	1.33
		Pooled (Fixed effects)			1.82E-04	1.27
		Pooled (Random effects)			1.41E-02	1.24
		Homogeneity test			0.138	
		Germany	0.213	0.186	0.302	1.18
		Italy	0.301	0.174	1.36E-05	2.04
rs3821236	A	Spain	0.317	0.197	2.07E-09	1.89
		Argentina	0.506	0.354	5.91E-05	1.87
		Mexico	0.458	0.318	1.30E-05	1.81
		Pooled (Fixed effects)			5.96E-20	1.77
		Pooled (Random effects)			8.98E-11	1.75
		Homogeneity test			0.124	
		Germany	0.251	0.258	0.809	0.96
		Italy	0.355	0.249	7.16E-04	1.66
		Spain	0.383	0.268	9.64E-08	1.70
		Argentina	0.477	0.383	1.34E-02	1.47
rs3024866	C	Mexico	0.520	0.395	1.38E-04	1.66
		Pooled (Fixed effects)			2.31E-12	1.51
		Pooled (Random effects)			1.30E-04	1.48
		Homogeneity test			0.024	
		Germany	0.265	0.210	0.050	1.35
		Italy	0.394	0.193	1.41E-10	2.70
		Spain	0.327	0.221	1.81E-07	1.72
		T				
rs7574865						

SNP	Risk Allele	Population/Test	Frequency Cases	Frequency controls	P-value	OR
		Argentina	0.415	0.319	8.85E-03	1.52
		Mexico	0.547	0.364	2.41E-08	2.10
		Pooled (Fixed effects)			4.44E-23	1.82
		Pooled (Random effects)			7.81E-08	1.82
		Homogeneity test			0.013	

Fixed effects: Cochran-Mantel-Haenszel meta-analysis controlling for strata under fixed effects model; Random effects: DerSimonian-Laird meta-analysis controlling for strata under random effects model;
Homogeneity test: Between-strata homogeneity test. OR: Odds ratio.

* Only Mexican adult samples were used in this analysis

Table 3

Usage and splicing of 5' and 3' UTRs in STAT4

Tissue	5'-terminal exons	Acc Number
Kidney	exon1B-exon2-exon3	NM_003151
	exon1C-exon1D-exon2-exon3	EU304788
Spleen	exon1B-exon2-exon3	
	exon1A-intron-exon1B-exon2-exon3	EU304789
	exon1C-D-E-F-exon2-exon3	EU304790
	exon1D-E-G-H-exon2-exon3	EU304791
Testis	exon1B-exon2-exon3	
	exon1A-exon2-exon3	BC031212
	exon1A-alt.spliced exon1B last 68 bp-exon2-exon3	EU304792
Lung	exon1B-exon2-exon3	
Pancreas	exon1B-exon2-exon3	
	exon1A-intron-exon1B-297nt intron1-exon2-exon3	EU304793
Small intestine	exon1B-exon2-exon3	
Uterus	exon1B-exon2-exon3	
	exon1A-intron-exon1B-297nt intron1-exon2-exon3	EU304793

Accession numbers for previously described α and β isoforms are shown in bold.