HMG Advance Access published March 22, 2012

Human Molecular Genetics, 2012 **1–11**
doi:10.1093/hmg/dds099

Identification of *CSK* as a systemic sclerosis genetic risk factor through Genome Wide Association Study follow-up

Jose-Ezequiel Martin[1,*], Jasper C. Broen[2], F. David Carmona[1], Maria Teruel[1], Carmen P. Simeon[3], Madelon C. Vonk[2], Ruben van 't Slot[4], Luis Rodriguez-Rodriguez[5], Esther Vicente[6], Vicente Fonollosa[3], Norberto Ortego-Centeno[7], Miguel A. González-Gay[8], Francisco J. García-Hernández[9], Paloma García de la Peña[10], Patricia Carreira[11], Spanish Scleroderma Group[†], Alexandre E. Voskuyl[12], Annemie J. Schuerwegh[13], Piet L.C.M. van Riel[2], Alexander Kreuter[14], Torsten Witte[15], Gabriella Riemekasten[16], Paolo Airo[17], Raffaella Scorza[18], Claudio Lunardi[19], Nicolas Hunzelmann[20], Jörg H.W. Distler[21], Lorenzo Beretta[18], Jacob van Laar[22], Meng May Chee[23], Jane Worthington[24], Ariane Herrick[24], Christopher Denton[25], Filemon K. Tan[26], Frank C. Arnett[26], Shervin Assassi[26], Carmen Fonseca[25], Maureen D. Mayes[26], Timothy R.D.J. Radstake[2,‡], Bobby P.C. Koeleman[4,‡] and Javier Martin[1,‡]

[1]Instituto de Parasitologia y Biomedicina Lopez-Neyra, CSIC, Granada, Spain, [2]Department of Rheumatology, Radboud University Nijmegen Medical Center, Nijmegen, The Netherlands, [3]Servicio de Medicina Interna, Hospital Valle de Hebron, Barcelona, Spain, [4]Department of Medical Genetics, University Medical Center Utrecht, Utrecht, The Netherlands, [5]Servicio de Reumatología, Hospital Clínico San Carlos, Madrid, Spain, [6]Servicio de Reumatología, Hospital La Princesa, Madrid, Spain, [7]Servicio de Medicina Interna, Hospital Clínico Universitario, Granada, Spain, [8]Servicio de Reumatología, Hospital Marqués de Valdecilla, Santander, Spain, [9]Servicio de Medicina Interna, Hospital Virgen del Rocio, Sevilla, Spain, [10]Servicio de Reumatología, Hospital Ramón y Cajal, Madrid, Spain, [11]Hospital 12 de Octubre, Madrid, Spain, [12]VU University Medical Center, Amsterdam, The Netherlands, [13]Department of Rheumatology, University of Leiden, Leiden, The Netherlands, [14]Department of Dermatology, Josefs-Hospital, Ruhr University Bochum, Bochum, Germany, [15]Hannover Medical School, Hannover, Germany, [16]Department of Rheumatology and Clinical Immunology, Charité University Hospital, Berlin, Germany, [17]Rheumatology Unitand Chair, Spedali Civili, Università de gli Studi, Brescia, Italy, [18]Referral Center for Systemic Autoimmune Diseases, Fondazione IRCCS Ca'Granda Ospedale MaRepiore Policlinico and University of Milan, Milan, Italy, [19]Department of Medicine, Policlinico GB Rossi, University of Verona, Verona, Italy, [20]Department of Dermatology, University of Cologne, Cologne, Germany, [21]Department of Internal Medicine 3, Institute for Clinical Immunology, University of Erlangen-Nuremberg, Erlangen 91054, Germany, [22]Institute of Cellular Medicine, Newcastle University, Newcastle Upon Tyne, UK, [23]Centre for Rheumatic Diseases, Glasgow Royal Infirmary, Glasgow, UK, [24]Department of Rheumatology and Epidemiology, University of Manchester, Manchester Academic Health Science Centre, Manchester, UK, [25]Centre for Rheumatology, Royal Free and University College School, London, UK and [26]The University of Texas Health Science Center–Houston, Houston, TX, USA

Received November 23, 2011; Revised and Accepted March 5, 2012

[*]To whom correspondence should be addressed. Email: cebercoto@ipb.csic.es
[†]See supplementary note.
[‡]These authors contributed equally to this work.

Systemic sclerosis (SSc) is complex autoimmune disease affecting the connective tissue; influenced by genetic and environmental components. Recently, we performed the first successful genome-wide association study (GWAS) of SSc. Here, we perform a large replication study to better dissect the genetic component of SSc. We selected 768 polymorphisms from the previous GWAS and genotyped them in seven replication cohorts from Europe. Overall significance was calculated for replicated significant SNPs by meta-analysis of the replication cohorts and replication-GWAS cohorts (3237 cases and 6097 controls). Six SNPs in regions not previously associated with SSc were selected for validation in another five independent cohorts, up to a total of 5270 SSc patients and 8326 controls. We found evidence for replication and overall genome-wide significance for one novel SSc genetic risk locus: *CSK* [*P*-value = 5.04×10^{-12}, odds ratio (OR) = 1.20]. Additionally, we found suggestive association in the loci *PSD3* (*P*-value = 3.18×10^{-7}, OR = 1.36) and *NFKB1* (*P*-value = 1.03×10^{-6}, OR = 1.14). Additionally, we strengthened the evidence for previously confirmed associations. This study significantly increases the number of known putative genetic risk factors for SSc, including the genes *CSK*, *PSD3* and *NFKB1*, and further confirms six previously described ones.

INTRODUCTION

Systemic sclerosis (scleroderma, SSc) is an autoimmune disease characterized by vascular damage, altered immune responses and extensive fibrosis of skin and internal organs (1), in which common genetic factors play an essential role, similar to most complex autoimmune diseases (2,3). So far, only a limited number of genes explaining little of the genetic variance present have been found in SSc (4,5).

In other autoimmune complex diseases for which extensive follow-up studies have been performed, such as inflammatory bowel disease and systemic lupus erythematosus (SLE), ~90 and 35 associated genes, respectively, have been identified (6,7). Therefore, it is expected that a number of risk factors for SSc are still to be defined. To date, only two genome-wide association studies (GWAS) have been published in populations of European ancestry in SSc (8,9). This calls for further replication studies and meta-analysis of SSc.

Due to the high rate of type 1 errors inherent to the GWAS technique, a number of strategies can be followed to discern truly associated genes from false positives in the tier 2 range of associations (ranging *P*-values from 10^{-3} to 5×10^{-8}), e.g. biological pathway analysis, meta-GWAS analysis or genetic interaction analysis. Another approach is to select the most strongly associated genetic variants from GWAS, where most true associations are harbored, by just accepting a small proportion of false negatives left out of the replication study. In this study, we performed the latter approach.

SSc is a clinically heterogeneous disease with a wide range of clinical manifestations (3); patients can be classified according to the severity of skin or organ involvement of the disease (10), or according to the presence or absence of several highly disease-specific auto-antibodies which are almost mutually exclusive in individual patients (1). Each of these disease phenotypes has proven to have specific genetic associations (11–17). In order to determine more of the genetic component of this profoundly disabling disease, such considerations need to be taken into account when selecting a battery of genetic variants to further test in other populations.

Considering the above, we followed the strategy of replication and validation of 768 genetic associations selected from our previous GWAS of SSc under different criteria in seven independent cohorts of European ancestry from five Europeans countries, and subsequently performed a meta-analysis including a total of 5270 SSc patients and 8326 healthy controls.

RESULTS

After quality filtering the replication stage data, 720 out of the 768 selected SNPs were analyzed. Mantel–Haenszel meta-analysis was performed for the three replication cohorts and for those cohorts together with the four GWAS cohorts. Six SNPs were selected for the validation stage in five independent cohorts. The overall process followed in the present study is summarized in Figure 1.

The genotyping success call rate was 99.88 and 97.99% in the replication and validation stages, respectively. Table 1 shows the statistics for the six SNPs selected for validation, while Table 2 shows previously described associations with SSc. Figure 2 shows meta-analysis association results of GWAS and replication cohorts for all 720 SNPs. Figure 3 shows relative risk for each of the three novel associations found in this study, for each population analyzed separately and for the meta-analysis. Detailed analysis and results for each associated region can be found below.

Novel SSc genetic associations

After the replication stage (genotyping of 768 SNPs selected from GWAS data), six SNPs were selected for validation with GWAS and replication combined Mantel–Haenszel *P*-value (P_{MH}) lower than 1×10^{-5} (in either the complete set of patients or its subphenotypes) and located in regions not previously associated with SSc. The SNPs in *NFKB1*, *PSD3*, *ACADS* and *CSK* were selected for their association with overall SSc, while those in *IPO5*/*FARP1* and *ADAMTS17* were selected for their association with the anti-topoisomerase autoantibody (ATA)-positive subgroup of patients. Out of the six SNPs, we were able to validate the association in one of them at the GWAS level ($P < 5 \times 10^{-8}$) and in another two at a suggestive level of association ($5 \times 10^{-8} < P < 5 \times 10^{-6}$).

The GWAS level association in the meta-analysis of all 11 analyzed cohorts was observed for single nucleotide

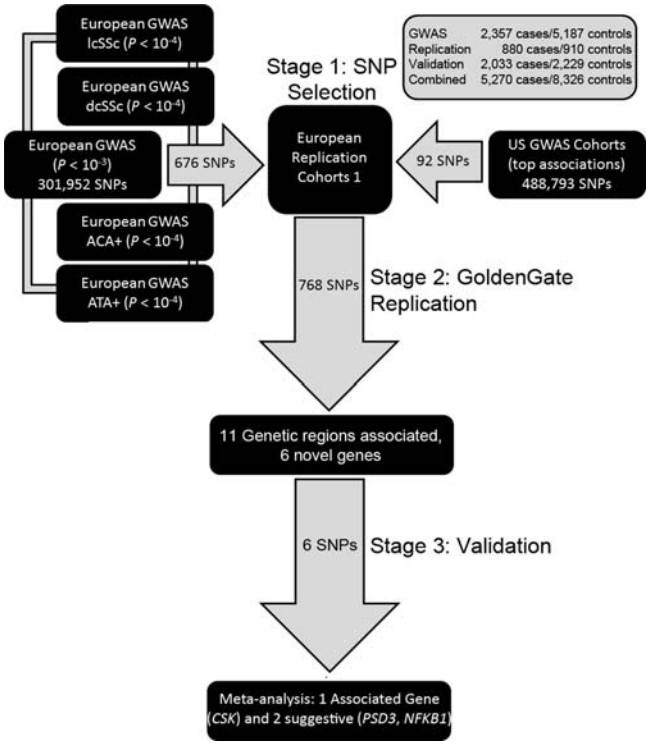

Figure 1. Schematic showing the overall process followed during the present study, along with the number of SNPs associated and considered in each step. lcSSc, limited cutaneous scleroderma; dcSSc, diffuse cutaneous scleroderma; ACA, anti-centromere autoantibody; ATA, anti-topoisomerase autoantibody.

polymorphisms (SNP) rs1378942 located in an intron of the *CSK* gene in chromosome 15 [P_{MH} value $= 5.04 \times 10^{-12}$, odds ratio (OR) = 1.202 (1.14–1.27)] with SSc (Table 1). Furthermore, we observed a suggestive overall level of association in two of the selected SNPs in all GWAS, replication and validation cohorts. One of them was located in the *PSD3* gene, rs10096702 [P_{MH} value $= 3.18 \times 10^{-7}$, OR = 1.363 (1.21–1.54)], while the other was in the *NFKB1* gene, rs1598859 [P_{MH} value $= 1.03 \times 10^{-6}$, OR = 1.140 (1.08–1.20)] (Table 1). Interestingly, we were only able to validate associations with the overall disease, but none of the subphenotype specific ones.

Previously reported SSc genetic associations

Associations with SSc previously reported were found for *STAT4* (rs10168266, P_{MH} value $= 1.81 \times 10^{-7}$, OR = 1.234), *CD247* (rs2056626, P_{MH} value $= 1.14 \times 10^{-8}$, OR = 0.832), *TNFSF4* (rs4916334, P_{MH} value $= 1.00 \times 10^{-5}$, OR = 0.861), *TNFAIP3* (rs2230926, P_{MH} value $= 2.29 \times 10^{-6}$, OR = 1.463) and *TNPO3/IRF5* (rs4728142, P_{MH} value $= 4.74 \times 10^{-10}$, OR = 1.216; and rs10488631, P_{MH} value $= 1.58 \times 10^{-18}$, OR = 1.513), further confirming their role in SSc (Table 2). Since several SNPs within both *TNPO3/IRF5* and *IRF8* genomic regions were selected, a conditional logistic regression analysis was performed to determine the independent signals of both loci (Table 3). While these associations found within each previously reported region spammed across the whole disease, SNP rs11642873 in *IRF8* was significantly associated with lcSSc (P_{MH} value $= 2.30 \times 10^{-9}$, OR = 0.730) and only marginally with dcSSc (P_{MH} value $= 3.77 \times 10^{-2}$, OR = 0.874).

DISCUSSION

Through a large replication study, designed using previously published GWAS data, we have been able to identify novel genetic associations with SSc (Table 1), and to provide further evidence for previously reported associations (Table 2). In the present study, a genetic variant within the *CSK* region, selected for replication because it was associated with SSc in the US cohort, has been identified as susceptibility factor for this disease. Furthermore, two other suggestive associations have been found in *PSD3* and *NFKB1* genetic variants, which were included in this study because they reached statistical significance in the European cohort. Further investigation will be required to further clarify the potential role of these two signals in SSc genetic predisposition.

CSK (c-src tyrosine kinase) is known to phosphorilate a tyrosine at the C-terminus of src kinases leading to their inactivation (18,19). In turn, src kinases are involved in fibrosis through their regulation of *FAK* (19,20), which is necessary for transmission of integrin signaling upon adhesion of fibroblasts to the extracellular matrix [and thus, their differentiation into myofibroblasts (21)] and has been involved in experimental pulmonary fibrosis (20), a major hallmark of SSc. Indeed, it has been demonstrated that either incubation of fibroblasts with Csk inhibitors or overexpression of Csk lead to a decreased expression of *COL1A1*, *COL1A2* and *FN1*, which are key components of the fibrotic process (22). Thus, genetics variants in *CSK* can be affecting its expression or functionality in a way that src kinanses are not inhibited, which in turn will contribute to the fibrosis in SSc. Furthermore, the *CSK* variant rs1378938, which is in relatively high linkage disequilibrium (LD) with the SSc-associated rs1378942 reported here ($r^2 = 0.72$ in the CEU population of the HapMap project), has been recently associated with celiac disease (23). Consequently, it is likely that *CSK* may represent another common autoimmunity risk factor. Further studies in related autoimmune disorders, such as SLE and rheumatoid arthritis (RA), may be performed to draw firm conclusions about this hypothesis.

NFKB has been extensively described to participate in and control the inflammatory process and thus, its role in the development of autoinflammatory disorders is widely accepted (24,25). The gene *NFKB1* (nuclear factor of kappa light polypeptide gene enhancer in B-cell 1) encodes a 105 kDa protein which can undergo cotranslational processing by the 26S proteasome to produce a 50 kDa protein. The 105 kDa protein is a Rel protein-specific transcription inhibitor and the 50 kDa protein is a DNA-binding subunit of the NFκB protein complex. SNP rs1598859, located in an intron of *NFKB1*, has been identified in this study as a risk genetic factor for SSc. This variant, or any other in the same haplotypic block, could be affecting the expression or the function of *NFKB1*, altering the inflammatory response, and thus participating in the development and course of the disease. Indeed, an ATTG in/del in the promoter of *NFKB1* has been recently

Table 1. Novel genetic associations found in this study with SSc

Phenotype	Chr.	Gene	SNP	Base pair[a]	Location	Change[b]	Stage	Sample size Cases	Sample size Controls	MAF	*P*-value[c]	OR[d]	CI (95%)	Breslow–Day *P*-value
SSc	4q24	*NFKB1*	rs1598859	103 725 482	Intron	C/T	GWAS	2357	5187	0.358	3.12×10^{-6}	1.188	1.11–1.28	0.151
							GWAS + Rep	3237	6097	0.354	6.51×10^{-6}	1.159	1.09–1.24	0.043
							GWAS + Rep + Val	5270	8326	0.356	1.03×10^{-6}	1.140	1.08–1.20	0.032
SSc	8p22	*PSD3*	rs10096702	18 642 877	Intron	A/G	GWAS	2357	5187	0.044	1.05×10^{-5}	1.435	1.22–1.69	0.001
							GWAS + Rep	3237	6097	0.045	8.62×10^{-9}	1.523	1.32–1.76	0.002
							GWAS + Rep + Val	5270	8326	0.045	3.18×10^{-7}	1.363	1.21–1.54	0.001
SSc	12q24	*ACADS*	rs558275	119 681 274	Intergenic	A/G	GWAS	2357	5187	0.398	6.46×10^{-6}	0.848	0.79–0.91	0.457
							GWAS + Rep	3237	6097	0.395	4.99×10^{-6}	0.863	0.81–0.92	0.234
							GWAS + Rep + Val	5270	8326	0.392	8.08×10^{-5}	0.901	0.86–0.95	0.180
ATA+	13q32	*IPO5/FARP1*	rs586851	97 542 564	Intergenic	C/A	GWAS	462	5187	0.103	1.36×10^{-3}	1.391	1.14–1.70	0.086
							GWAS + Rep	727	6097	0.108	3.89×10^{-6}	1.454	1.24–1.71	0.023
							GWAS + Rep + Val	1161	8326	0.112	4.59×10^{-5}	1.311	1.15–1.49	0.013
SSc	15q24	*CSK*	rs1378942	72 864 420	Intron	C/A	GWAS	2357	5187	0.362	7.19×10^{-7}	1.199	1.12–1.29	0.573
							GWAS + Rep	3237	6097	0.367	4.42×10^{-8}	1.194	1.12–1.27	0.407
							GWAS + Rep + Val	5270	8326	0.370	5.04×10^{-12}	1.202	1.14–1.27	0.114
ATA+	15q26	*ADAMTS17*	rs2289584	98 707 019	Intergenic	T/G	GWAS	462	5187	0.175	9.04×10^{-5}	1.391	1.18–1.64	0.383
							GWAS + Rep	727	6097	0.175	6.07×10^{-7}	1.416	1.23–1.62	0.457
							GWAS + Rep + Val	1161	8326	0.178	5.15×10^{-5}	1.257	1.13–1.41	0.099

Chr., chromosome; MAF, minor allele frequency; OR, odds ratio; CI, confidence intervals. GWAS, genome-wide association study; Rep, replication stage (768 SNPs); Val, validation stage (6 SNPs).
[a]All genomic positions are referent to genome build 36.
[b]Minor allele first.
[c]Mantel–Haenszel meta-analysis of the cohorts involved in the corresponding stage of the analysis (see Materials and Methods).
[d]All ORs are for the minor allele.

Table 2. Previously reported associations with SSc found in this study

Chr.	Gene	SNP	Base pair[a]	Location	Change[b]	Sample size		MAF	*P*-value[c]	OR[d]	CI (95%)	Breslow–Day *P*-value
						Cases	Controls					
1q24	*CD247*	rs2056626	165 687 049	Intron	G/T	3237	6097	0.396	1.14×10^{-8}	0.832	0.78–0.89	5.14×10^{-1}
1q25	*TNFSF4*	rs10798269	171 576 336	Intergenic	A/G	3237	6097	0.336	1.29×10^{-5}	0.864	0.81–0.92	7.78×10^{-1}
		rs4916334	171 600 452	Intergenic	G/T	3237	6097	0.334	1.00 × 10–5	0.861	0.81–0.92	8.18×10^{-1}
2q32	*STAT4*	rs10168266[e]	191 644 049	Intron	T/C	2372	4395	0.214	1.81×10^{-7}	1.234	1.14–1.34	1.11×10^{-1}
6q23	*TNFAIP3*	rs2230926	138 237 759	Exon	G/T	3237	6097	0.036	2.29×10^{-6}	1.463	1.25–1.71	4.80×10^{-2}
7q32	*TNPO3/IRF5*	rs4728142	128 361 203	Intergenic	A/G	3237	6097	0.463	4.74×10^{-10}	1.216	1.14–1.29	9.12×10^{-2}
		rs7808907	128 371 320	Intron	T/C	3237	6097	0.483	1.47×10^{-7}	0.848	0.80–0.90	2.35×10^{-2}
		rs10488631	128 381 419	Intergenic	C/T	3237	6097	0.116	1.58×10^{-18}	1.513	1.38–1.66	6.94×10^{-1}
		rs12531711	128 404 702	Intergenic	G/A	3237	6097	0.115	3.31×10^{-19}	1.527	1.39–1.68	7.11×10^{-1}
		rs12537284	128 505 142	Intergenic	A/G	3237	6097	0.141	1.85×10^{-9}	1.301	1.19–1.42	8.77×10^{-1}
		rs2084654	128 516 364	Intergenic	G/A	3237	6097	0.339	1.64×10^{-6}	1.171	1.10–1.25	2.58×10^{-1}
16q24	*IRF8*	rs11117425	84 529 772	Intergenic	T/C	3237	6097	0.296	1.43×10^{-6}	0.845	0.79–0.91	1.96×10^{-1}
		rs11644034	84 530 113	Intergenic	A/G	3237	6097	0.196	8.89×10^{-8}	0.804	0.74–0.87	4.03×10^{-1}
		rs12711490	84 530 529	Intergenic	C/T	3237	6097	0.196	6.43×10^{-8}	0.802	0.74–0.87	3.56×10^{-1}
		rs7202472	84 535 003	Intergenic	T/G	3237	6097	0.182	1.58×10^{-7}	0.802	0.74–0.87	3.65×10^{-1}
		rs11642873[f]	84 549 206	Intergenic	C/A	3237	6097	0.176	2.30×10^{-9}	0.730	0.72–0.86	8.28×10^{-1}

Chr., chromosome; MAF, minor allele frequency; OR, odds ratio; CI, confidence intervals.
[a]All genomic positions are referent to genome build 36.
[b]Minor allele first.
[c]Mantel–Haenszel meta-analysis of GWAS and GoldenGate cohorts.
[d]All ORs are for the minor allele.
[e]*TAT4* SNP rs10168266 was not genotyped in the European GWAS cohorts, nor could its genotype could be imputed.
[f]*RF8* SNP rs11642873 association was confined to lcSSc and these are the statistical shown in the table rather than global SSc as in the rest of the SNPs.

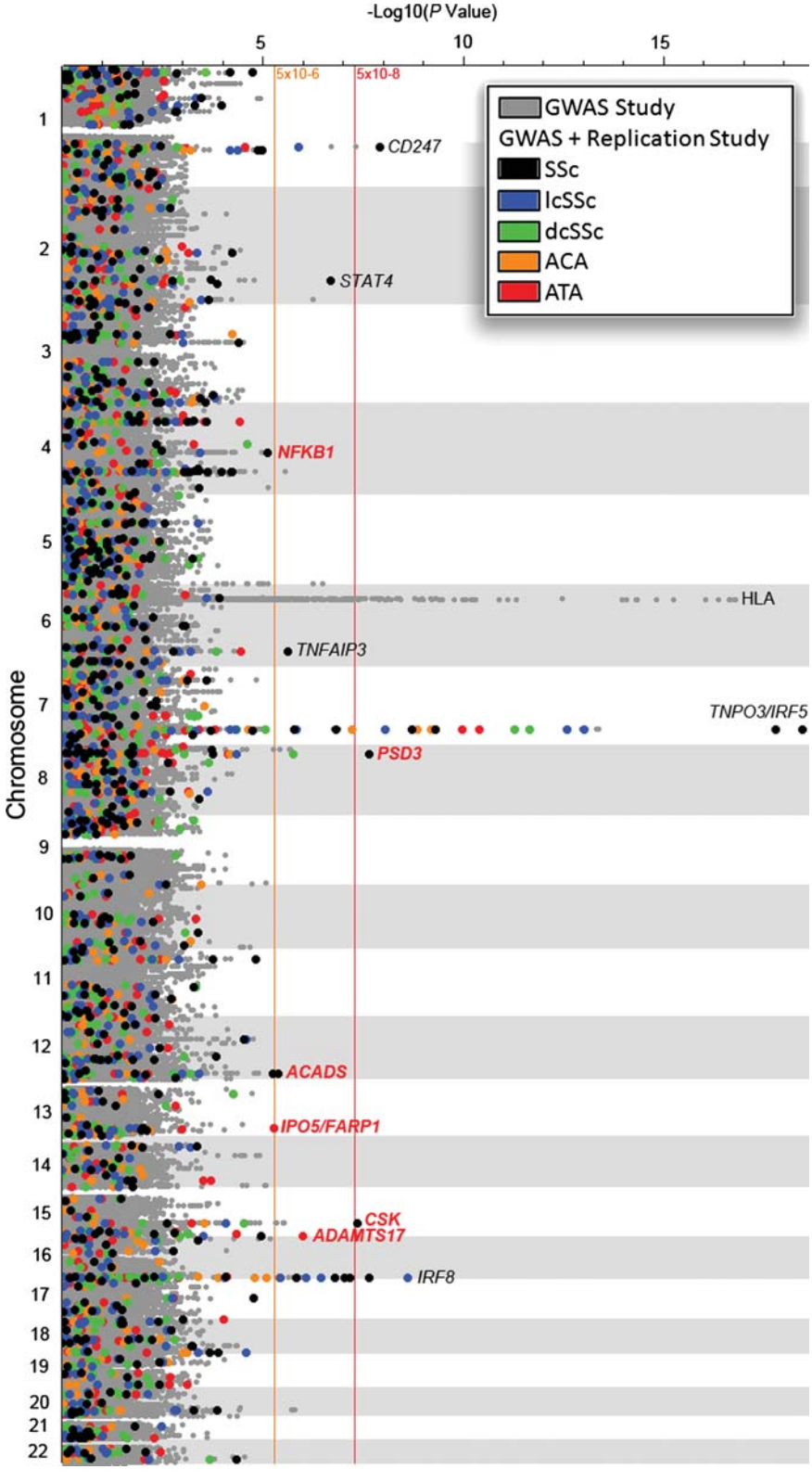

Figure 2. The Manhattan plot of the GWAS and replication cohorts meta-analysis. Grey dots represent *P*-values of SNPs which are genotyped only in the GWAS cohorts in the previous study by Radstake *et al.* (8). Other color dots represent P_{MH} values of the 720 SNPs which were in both the GWAS and the replication cohorts (Mantel–Haenszel meta-analysis). Gene names in black have been previously reported as genetic risk factors for SSc, while gene names in red are novel ones.

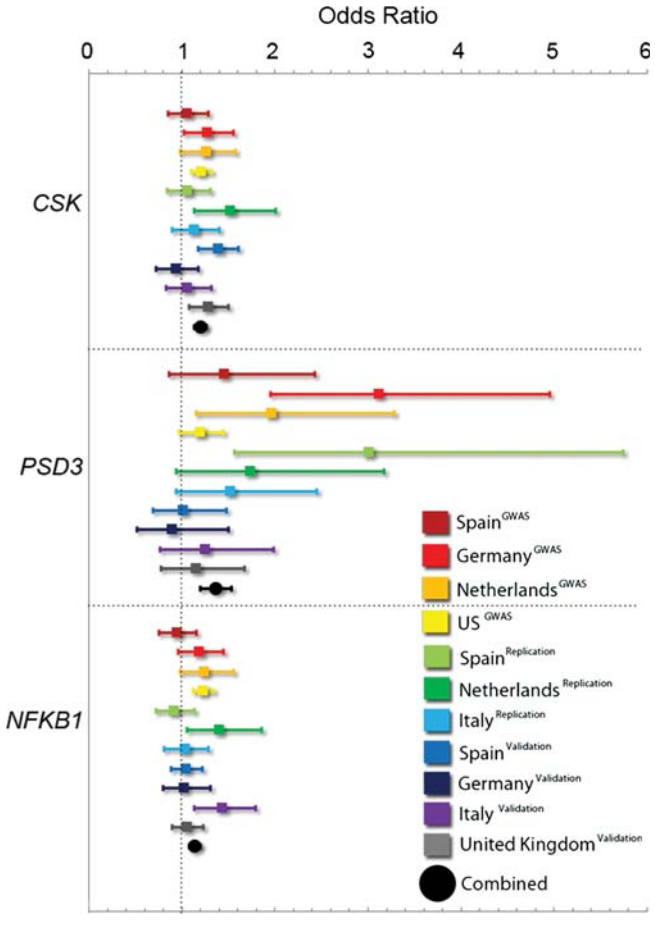

Figure 3. ORs and 95% CIs of each of the novel genetic association with SSc found in this study, in either the meta-analysis of all cohorts and each population separately.

associated with SLE in Asian population (26). These results are nonetheless controversial, since a previous report found no association between this very same in/del and the risk of developing RA or SLE (27). In both the studies, the statistical power was rather limited with cohorts of around 300 patients, hence further fine-mapping studies in the *NFKB1* region with larger study cohorts will be needed to elucidate the role of *NFKB1* genetic variants in autoimmunity in general and SSc specifically.

The *PSD3* gene encodes a protein of unknown function, which has a pleckstrin domain and a Sec7 domain. The pleckstrin domain is found in a wide range of proteins, and is capable of binding phosphatidylinositol, G proteins and protein kinase C, thus acting as a scaffold protein in signal transduction pathways, while the Sec7 domain is a guanine nucleotide exchange factor, which is a component of intracellular signaling networks. Further research will be needed to determine the role of *PSD3*, but these data point to a role in the immune system and the pathogenesis of SSc.

Previously described associations in *CD247* (8), *STAT4* (15,28), *TNFAIP3* (29), *IRF8* (30), *TNFSF4* (31,32) and *TNPO3*/*IRF5* (33,34) have been firmly replicated in the present study, confirming their role in the pathogenesis of SSc.

Dieude *et al.* (29) found the SNP rs5029939 in *TNFAIP3* associated with SSc, which showed stronger association in the dcSSc and ATA+ subgroups of the disease, although the association remained significant in the opposing lcSSc and ACA+ subgroups. In the present study, rs2230926 (in total LD with rs5029939 in the HapMap CEU population) has shown GWAS level association with the global SSc, and no sign of heterogeneity in the SSc subgroups was detected (Table 2). Taking both studies together, the association of *TNFAIP3* is most likely due to an effect on SSc susceptibility overall rather than any of its sub-phenotypes. The previously reported fluctuating level of significance is probably due to different population composition and cohort size (1656 cases and 1311 controls in the study by Dieude *et al.* (29) and 3237 cases and 6097 controls in the present study). SNP rs2230926 is located in exon 3 of *TNFAIP3*, and encodes for a phenylalanine to cysteine change at residue 127 which has been demonstrated to alter the function of the protein, which makes it a functionally associated genetic variant with SLE (35).

Association in the *TNPO3*/*IRF5* region has been detected previously in SSc (8,33,34) and SLE (36,37). The three independently associated polymorphisms have been detected in SLE: an in/del in exon 6 which change the expression level of *IRF5* (tagged by rs10488631 in our study), a variant in exon 1B which affects splicing of this exon (not tagged in our study) and a variant which disrupts a polyA signal site (tagged by rs4728142 in our study). Our findings in the present study are consistent with associations found in SLE. We detect two of the three independent variants (Table 3): the in/del and the polyA signal, yet the variant in exon 1B was not captured in this study. This observation adds more evidence for a similar genetic component of SSc and SLE. Nevertheless, further research will be needed in order to determine whether the variant in exon 1B, affecting splicing of the gene, also has a major role in SSc.

IRF8 has been recently described as a risk factor for lcSSc (30) and has also been associated with the risk of multiple sclerosis (38). Of all the SNPs selected for replication in the present study, only rs11642873 remained independently significant (Table 3), and its association was found, as in the previous study, exclusively confined to lcSSc. Further fine-mapping and functional studies are necessary in order to determine the role of genetic variants in *IRF8* in the pathogenesis of SSc and autoimmunity.

In summary, by the analysis of 720 genetic variants, selected from GWAS data's 'grey zone' of association, we describe three new genetic risk factors for SSc (*CSK*, *PSD3* and *NFKB1*) and confirm five previously reported associations (*CD247*, *STAT4*, *TNFAIP3*, *TNPO3*/*IRF5*, *TNFSF4* and *IRF8*). Also in the case of *TNPO3*/*IRF5*, we clarify the nature of the association with SSc which is similar to that found in SLE, thus highlighting the similarities of the genetic component for both diseases.

MATERIALS AND METHODS

Populations

The four populations used in the previous SSc GWAS have been previously described (8). Replication and validation

Table 3. Conditional analysis in the TNPO3/IRF5 and IRF8 genetic regions

Gene	Chr.	SNP	Base pair[a]	Location	Change[b]	MAF	P_{MH}	OR[c]	CI (95%)	Breslow–Day *P*-value	Conditioned *P*-value
TNPO3/IRF5	7q32	rs4728142	128 361 203	Intergenic	A/G	0.463	4.74E − 10	1.216	1.14–1.29	9.12E − 02	$\mathbf{9.86 \times 10^{-4}}$
		rs7808907	128 371 320	Intron	T/C	0.483	1.47E − 07	0.848	0.80–0.90	2.35E − 02	9.41×10^{-1}
		rs10488631	128 381 419	Intergenic	C/T	0.116	1.58E − 18	1.513	1.38–1.66	6.94E − 01	$\mathbf{1.96 \times 10^{-7}}$
		rs12531711	128 404 702	Intergenic	G/A	0.115	3.31E − 19	1.527	1.39–1.68	7.11E − 01	$\mathbf{2.46 \times 10^{-7}}$
		rs12537284	128 505 142	Intergenic	A/G	0.141	1.85E − 09	1.301	1.19–1.42	8.77E − 01	2.43×10^{-1}
		rs2084654	128 516 364	Intergenic	G/A	0.339	1.64E − 06	1.171	1.10–1.25	2.58E − 01	1.36×10^{-1}
IRF8	16q24	rs8056420	83 969 142	Intergenic	G/A	0.162	5.49×10^{-1}	1.029	0.94–1.13	3.13×10^{-3}	8.40×10^{-1}
		rs7186021	84 043 560	Intergenic	C/A	0.473	2.31×10^{-2}	0.922	0.86–0.99	2.48×10^{-3}	5.32×10^{-1}
		rs8053194	84 072 623	Intergenic	T/G	0.404	3.01×10^{-2}	0.925	0.86–0.99	6.69×10^{-3}	6.92×10^{-1}
		rs11117425	84 529 772	Intergenic	T/C	0.296	1.43E − 06	0.845	0.79–0.90	1.96E − 01	6.85×10^{-2}
		rs11644034	84 530 113	Intergenic	A/G	0.196	8.89E − 08	0.8043	0.74–0.87	4.03E − 01	3.68×10^{-1}
		rs12711490	84 530 529	Intergenic	C/T	0.196	6.43E − 08	0.802	0.74–0.87	3.56E − 01	7.90×10^{-1}
		rs7202472	84 535 003	Intergenic	T/G	0.182	1.58E − 07	0.802	0.74–0.87	3.65E − 01	6.07×10^{-1}
		rs11642873	84 549 206	Intergenic	C/A	0.179	2.30E − 09	0.730	0.66–0.81	9.21E − 01	$\mathbf{2.05 \times 10^{-5}}$
		rs10514613	84 688 032	Intergenic	C/T	0.056	4.81×10^{-2}	1.162	1.00–1.35	6.44×10^{-4}	3.34×10^{-2}

SNPs in bold form a haplotype.
Chr., chromosome; MAF, minor allele frequency; OR, odds ratio; CI, confidence intervals.
[a]All genomic positions are referent to genome build 36.
[b]Minor allele first.
[c]All ORs are for the minor allele.

cohorts from Spain (two independent cohorts), Germany, Netherlands, Italy (two independent cohorts) and the UK for a total of 2913 cases and 3139 controls were recruited from hospitals and blood banks from each of these countries (numbers before quality control filters). Key features of these populations can be found in Supplementary Material, Table S1. This study was approved by the local ethics committees of the participating hospitals. All of the SSc patients participating in this study met American College of Rheumatology criteria (39) and were classified in the disease subtypes according to LeRoy and Medsger (10). All individuals in this study were of European ancestry [either self-reported and/or principal component analysis (PCA) determined, see below] and gave written informed consent.

Samples in this study included those in the initial discovery GWAS cohorts as well as those in the replication cohorts in the work by Radstake *et al.* (8). In the present study, the previous replication cohorts were genotyped for 768 additional SNPs.

Combination of GWAS and replication cohorts genotyped in this study resulted in a total of 5270 cases and 8326 controls (after QC was applied), which gave us a statistical power of 100% to detect an OR of 1.3 with a minor allele frequency of 0.20. Other power calculations can be found in Supplementary Material, Table S2.

Study design

SNP selection. First, we aimed to select the putative SSc-associated genetic variants that did not reach genome-wide significance in our initial GWAS (8). We filtered GWAS data as previously published. We excluded from our analysis the extended HLA region on chromosome six, since association of this region with SSc is well known, and was not the focus of the present study. Since there is growing evidence for the existence of genetic heterogeneity in SSc, we included the clinical subtypes of the disease (i.e. lcSSc and dcSSc) and the presence of the two most common autoantibodies (i.e. ACA and ATA) in our selection criteria (1,3,10). Because of the differences between the US and European samples in the PCA, we decided to select the SNPs for each set separately, lending greater weight to the European cohorts (it should be noted that all our replication cohorts were of European countries).

We selected all SNPs with a Mantel–Haenszel, λ corrected *P*-value lower than 10^{-3} from the European cohorts (i.e. Netherlands, Germany and Spain) meta-analysis. We also selected all SNPs with a Mantel–Haenszel λ-corrected *P*-value lower than 10^{-4} in the lcSSc, dcSSc, ACA+ and ATA+ subgroups. We also considered the possibility that SNPs associated in one of the three populations could have been filtered out in the others due to any criteria, so we included all SNPs with a corrected *P*-value lower than 10^{-4} in any of the three populations before merging the data sets. This gave a total of 676 SNPs selected. In order to use all the genotyping capacity of the chosen platform (in which 768 SNPs could be analyzed), and taking into consideration that the first 676 SNPs were selected only from the European panel, we also included the 92 previously non-described independent signals from the US panel that showed the most significant *P*-values. The overall strategy followed in the present study can be found in Figure 1.

Replication stage. In this stage, we aimed to confirm selected associations from GWAS data on case/control cohorts from Spain, Holland and Italy up to a total of 880 cases and 910 controls. All individuals in these cohorts were genotyped for the 768 selected SNPs and association analyses were performed.

Validation stage. At this point, we selected the top SNPs from the meta-analysis of the pervious stages (those genotyped in

GWAS and GoldenGate platforms) and further tested their association in larger cohorts from Spain, Germany, Italy and the UK up to a total of 2033 cases and 2229 controls. All SNPs with a meta-analysis P-value $<1 \times 10^{-5}$ either in the disease or any of its subphenotypes (i.e. lcSSc, dcSSc, ACA+ or ATA+) and not previously reported to be associated with SSc were selected for validation, comprising a total of six SNPs. Finally, we performed meta-analyses for all associated SNPs combining the GWAS and all replication cohorts.

Genotyping methods

The replication cohort was genotyped for 768 SNPs, selected from the GWAS analysis as described above, using a custom Illumina GoldenGate array run on the Illumina iScan system. The validation stage was genotyped using TaqMan genotyping assays from Applied Biosystems.

Genotype imputation of data

As described in Radstake *et al*. (8), the different cohorts were genotyped using different genotyping arrays (the European cohorts were genotyped mostly with the Illumina HumanHap 370k array, whereas the US cohort was genotyped with the Illumina 550k array). As a consequence, it was possible that some SNPs were present in one but not in the other platform. To prevent this, a genotype imputation was performed in all the GWAS cohorts to obtain a full overlap between platforms. Imputation was performed with IMPUTE software 1.00 as previously described (40), using as reference panels the CEU and TSI HapMap populations.

Data analysis

All data were analyzed using Plink software version 1.07 (http://pngu.mgh.harvard.edu/purcell/plink/) (41). LD patterns among SNPs were also calculated with the r^2 statistic using HaploView software (42). Manhattan plots were generated using HelixTree SNP Variation Suite 7 (http://www.goldenhelix.com/SNP_Variation/HelixTree/index.html). In order to avoid any position discrepancies, all base-pair locations for the analyzed genetic variations correspond to those reported in the genome build 36, as these were the positions used in the original GWAS.

All data were quality filtered using the following criteria: Hardy–Weinberg equilibrium P-value > 0.001 on controls of any of the populations analyzed separately, minor allele frequency > 0.01, success call rate per individual > 0.95 and per SNP >0.95. After applying quality control, remaining SNPs were statistically analyzed using the Chi-squared case/control approach in 1433 cases and 1644 controls. The meta-analysis of the different cohorts was conducted using the Mantel–Haenszel test to calculate a pooled OR, and the 95% confidence interval for the OR was estimated using a random effect model. Heterogeneity between cohorts was tested using the Breslow–Day test (P-values <0.05 were considered statistically significant); nevertheless, we did not rule out any association in the basis of OR heterogeneity. Combined P-values for the Mantel–Haenszel tests were calculated as implemented in Plink. The independence of effects of SNPs in the same genetic region was tested by multiple logistic regression analysis, conditioning each SNP to all others, as implemented in the Plink software. Population stratification was assessed by PCA as previously described (43). All individuals who deviated more than four standard deviations from the centroid of their population were removed as outliers.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at *HMG* online.

ACKNOWLEDGEMENTS

We thank Sofia Vargas, Sonia Garcia and Gema Robledo for their excellent technical assistance, and all the patients and healthy controls for kindly accepting their essential collaboration. We would also like to thank the following organizations: the EULAR Scleroderma Trials and Research (EUSTAR), the German Network of Systemic Sclerosis and Banco Nacional de ADN (University of Salamanca, Spain).

Conflict of Interest statement: none declared.

FUNDING

This work was supported by the following grants: J.M. was funded by GEN-FER from the Spanish Society of Rheumatology, SAF2009-11110 from the Spanish Ministry of Science, CTS-4977 from Junta de Andalucía (Spain), Redes Temáticas de Investigación Cooperativa Sanitaria Program, RD08/0075 (RIER) from Instituto de Salud Carlos III (ISCIII, Spain) and Fondo Europeo de Desarrollo Regional (FEDER). T.R.D.J.R. was funded by the VIDI laureate from the Dutch Association of Research (NWO) and Dutch Arthritis Foundation (National Reumafonds). J.M. and T.R.D.J.R. were sponsored by the Orphan Disease Program grant from the European League Against Rheumatism (EULAR). B.P.C.K. is supported by the Dutch Diabetes Research Foundation (grant 2008.40.001) and the Dutch Arthritis Foundation (Reumafonds, grant NR 09-1-408). T.W. is supported by the grant DFG WI 1031/6-1 and by the grant DFG WI 1031/6-1. N.O.-C. was funded by PI-0590-2010, Consejeria de Salud, Junta de Andalucia, Spain. F.K.T., F.C.A. and M.D.M. were supported by NIH Scleroderma Family Registry and DNA Repository (N01-AR-0-2251), NIH RO1- AR055258 and NIH Center of Research Translation in Scleroderma (1P50AR054144), and the Department of Defense Congressionally Directed Medical Research Programs (W81XWH-07-01-0111). C.F. was supported by 'The Raynaud's and Scleroderma Association' and 'The Scleroderma Society'.

REFERENCES

1. Gabrielli, A., Avvedimento, E.V. and Krieg, T. (2009) Scleroderma. *N. Engl. J. Med.*, **360**, 1989–2003.
2. Agarwal, S.K., Tan, F.K. and Arnett, F.C. (2008) Genetics and genomic studies in scleroderma (systemic sclerosis). *Rheum. Dis. Clin. North Am.*, **34**, 17–40.

3. Jimenez, S.A. and Derk, C.T. (2004) Following the molecular pathways toward an understanding of the pathogenesis of systemic sclerosis. *Ann. Intern. Med.*, **140**, 37–50.
4. Romano, E., Manetti, M., Guiducci, S., Ceccarelli, C., Allanore, Y. and Matucci-Cerinic, M. (2011) The genetics of systemic sclerosis: an update. *Clin. Exp. Rheumatol.*, **29**, S75–S86.
5. Martin, J. and Fonseca, C. (2011) The genetics of scleroderma. *Curr. Rheumatol. Rep.*, **13**, 13–20.
6. Khor, B., Gardet, A. and Xavier, R.J. (2011) Genetics and pathogenesis of inflammatory bowel disease. *Nature*, **474**, 307–317.
7. Sestak, A.L., Furnrohr, B.G., Harley, J.B., Merrill, J.T. and Namjou, B. (2011) The genetics of systemic lupus erythematosus and implications for targeted therapy. *Ann. Rheum. Dis.*, **70**(Suppl. 1), 37–43.
8. Radstake, T.R., Gorlova, O., Rueda, B., Martin, J.E., Alizadeh, B.Z., Palomino-Morales, R., Coenen, M.J., Vonk, M.C., Voskuyl, A.E., Schuerwegh, A.J. *et al.* (2010) Genome-wide association study of systemic sclerosis identifies CD247 as a new susceptibility locus. *Nat. Genet.*, **42**, 426–429.
9. Allanore, Y., Saad, M., Dieude, P., Avouac, J., Distler, J.H., Amouyel, P., Matucci-Cerinic, M., Riemekasten, G., Airo, P., Melchers, I. *et al.* (2011) Genome-wide scan identifies TNIP1, PSORS1C1, and RHOB as novel risk loci for systemic sclerosis. *PLoS Genet.*, **7**, e1002091.
10. LeRoy, E.C. and Medsger, T.A. Jr (2001) Criteria for the classification of early systemic sclerosis. *J. Rheumatol.*, **28**, 1573–1576.
11. Nietert, P.J., Mitchell, H.C., Bolster, M.B., Shaftman, S.R., Tilley, B.C. and Silver, R.M. (2006) Racial variation in clinical and immunological manifestations of systemic sclerosis. *J. Rheumatol.*, **33**, 263–268.
12. Assassi, S., Arnett, F.C., Reveille, J.D., Gourh, P. and Mayes, M.D. (2007) Clinical, immunologic, and genetic features of familial systemic sclerosis. *Arthritis Rheum.*, **56**, 2031–2037.
13. Steen, V.D. (2008) The many faces of scleroderma. *Rheum. Dis. Clin. North Am.*, **34**, 1–15. ; v.
14. Arnett, F.C., Gourh, P., Shete, S., Ahn, C.W., Honey, R.E., Agarwal, S.K., Tan, F.K., McNearney, T., Fischbach, M., Fritzler, M.J. *et al.* (2010) Major histocompatibility complex (MHC) class II alleles, haplotypes and epitopes which confer susceptibility or protection in systemic sclerosis: analyses in 1300 Caucasian, African-American and Hispanic cases and 1000 controls. *Ann. Rheum. Dis.*, **69**, 822–827.
15. Rueda, B., Broen, J., Simeon, C., Hesselstrand, R., Diaz, B., Suarez, H., Ortego-Centeno, N., Riemekasten, G., Fonollosa, V., Vonk, M.C. *et al.* (2009) The STAT4 gene influences the genetic predisposition to systemic sclerosis phenotype. *Hum. Mol. Genet.*, **18**, 2071–2077.
16. Rueda, B., Gourh, P., Broen, J., Agarwal, S.K., Simeon, C., Ortego-Centeno, N., Vonk, M.C., Coenen, M., Riemekasten, G., Hunzelmann, N. *et al.* (2010) BANK1 functional variants are associated with susceptibility to diffuse systemic sclerosis in Caucasians. *Ann. Rheum. Dis.*, **69**, 700–705.
17. Gourh, P., Tan, F.K., Assassi, S., Ahn, C.W., McNearney, T.A., Fischbach, M., Arnett, F.C. and Mayes, M.D. (2006) Association of the PTPN22 R620W polymorphism with anti-topoisomerase I- and anticentromere antibody-positive systemic sclerosis. *Arthritis Rheum.*, **54**, 3945–3953.
18. Lowell, C.A. (2004) Src-family kinases: rheostats of immune cell signaling. *Mol. Immunol.*, **41**, 631–643.
19. Okutani, D., Lodyga, M., Han, B. and Liu, M. (2006) Src protein tyrosine kinase family and acute inflammatory responses. *Am. J. Physiol. Lung Cell Mol. Physiol.*, **291**, 129–141.
20. Vittal, R., Horowitz, J.C., Moore, B.B., Zhang, H., Martinez, F.J., Toews, G.B., Standiford, T.J. and Thannickal, V.J. (2005) Modulation of prosurvival signaling in fibroblasts by a protein kinase inhibitor protects against fibrotic tissue injury. *Am. J. Pathol.*, **166**, 367–375.
21. Thannickal, V.J., Lee, D.Y., White, E.S., Cui, Z., Larios, J.M., Chacon, R., Horowitz, J.C., Day, R.M. and Thomas, P.E. (2003) Myofibroblast differentiation by transforming growth factor-beta1 is dependent on cell adhesion and integrin signaling via focal adhesion kinase. *J. Biol. Chem.*, **278**, 12384–12389.
22. Skhirtladze, C., Distler, O., Dees, C., Akhmetshina, A., Busch, N., Venalis, P., Zwerina, J., Spriewald, B., Pileckyte, M., Schett, G. *et al.* (2008) Src kinases in systemic sclerosis: central roles in fibroblast activation and in skin fibrosis. *Arthritis Rheum.*, **58**, 1475–1484.
23. Trynka, G., Hunt, K.A., Bockett, N.A., Romanos, J., Mistry, V., Szperl, A., Bakker, S.F., Bardella, M.T., Bhaw-Rosun, L., Castillejo, G. *et al.* (2011) Dense genotyping identifies and localizes multiple common and rare variant association signals in celiac disease. *Nat. Genet.*, **43**, 1193–1201.
24. Barnes, P.J. and Karin, M. (1997) Nuclear factor-kappaB: a pivotal transcription factor in chronic inflammatory diseases. *N. Engl. J. Med.*, **336**, 1066–1071.
25. Tak, P.P. and Firestein, G.S. (2001) NF-kappaB: a key role in inflammatory diseases. *J. Clin. Invest.*, **107**, 7–11.
26. Gao, M., Wang, C.H., Sima, X. and Han, X.M. (2011) NFKB1 -94 insertion/deletion ATTG polymorphism contributes to risk of systemic lupus erythematosus. *DNA Cell Biol.* [Epub ahead of print].
27. Orozco, G., Sanchez, E., Collado, M.D., Lopez-Nevot, M.A., Paco, L., Garcia, A., Jimenez-Alonso, J. and Martin, J. (2005) Analysis of the functional NFKB1 promoter polymorphism in rheumatoid arthritis and systemic lupus erythematosus. *Tissue Antigens*, **65**, 183–186.
28. Dieude, P., Guedj, M., Wipff, J., Ruiz, B., Hachulla, E., Diot, E., Granel, B., Sibilia, J., Tiev, K., Mouthon, L. *et al.* (2009) STAT4 is a genetic risk factor for systemic sclerosis having additive effects with IRF5 on disease susceptibility and related pulmonary fibrosis. *Arthritis. Rheum.*, **60**, 2472–2479.
29. Dieude, P., Guedj, M., Wipff, J., Ruiz, B., Riemekasten, G., Matucci-Cerinic, M., Melchers, I., Hachulla, E., Airo, P., Diot, E. *et al.* (2010) Association of the TNFAIP3 rs5029939 variant with systemic sclerosis in the European Caucasian population. *Ann. Rheum. Dis.*, **69**, 1958–1964.
30. Gorlova, O., Martin, J.E., Rueda, B., Koeleman, B.P., Ying, J., Teruel, M., Diaz-Gallo, L.M., Broen, J.C., Vonk, M.C., Simeon, C.P. *et al.* (2011) Identification of novel genetic markers associated with clinical phenotypes of systemic sclerosis through a genome-wide association strategy. *PLoS Genet.*, **7**, e1002178.
31. Bossini-Castillo, L., Broen, J.C., Simeon, C.P., Beretta, L., Vonk, M.C., Ortego-Centeno, N., Espinosa, G., Carreira, P., Camps, M.T., Navarrete, N. *et al.* (2011) A replication study confirms the association of TNFSF4 (OX40L) polymorphisms with systemic sclerosis in a large European cohort. *Ann. Rheum. Dis.*, **70**, 638–641.
32. Gourh, P., Arnett, F.C., Tan, F.K., Assassi, S., Divecha, D., Paz, G., McNearney, T., Draeger, H., Reveille, J.D., Mayes, M.D. *et al.* (2010) Association of TNFSF4 (OX40L) polymorphisms with susceptibility to systemic sclerosis. *Ann. Rheum. Dis.*, **69**, 550–555.
33. Dieude, P., Guedj, M., Wipff, J., Avouac, J., Fajardy, I., Diot, E., Granel, B., Sibilia, J., Cabane, J., Mouthon, L. *et al.* (2009) Association between the IRF5 rs2004640 functional polymorphism and systemic sclerosis: a new perspective for pulmonary fibrosis. *Arthritis. Rheum.*, **60**, 225–233.
34. Ito, I., Kawaguchi, Y., Kawasaki, A., Hasegawa, M., Ohashi, J., Hikami, K., Kawamoto, M., Fujimoto, M., Takehara, K., Sato, S. *et al.* (2009) Association of a functional polymorphism in the IRF5 region with systemic sclerosis in a Japanese population. *Arthritis Rheum.*, **60**, 1845–1850.
35. Musone, S.L., Taylor, K.E., Lu, T.T., Nititham, J., Ferreira, R.C., Ortmann, W., Shifrin, N., Petri, M.A., Kamboh, M.I., Manzi, S. *et al.* (2008) Multiple polymorphisms in the TNFAIP3 region are independently associated with systemic lupus erythematosus. *Nat. Genet.*, **40**, 1062–1064.
36. Sigurdsson, S., Nordmark, G., Goring, H.H., Lindroos, K., Wiman, A.C., Sturfelt, G., Jonsen, A., Rantapaa-Dahlqvist, S., Moller, B., Kere, J. *et al.* (2005) Polymorphisms in the tyrosine kinase 2 and interferon regulatory factor 5 genes are associated with systemic lupus erythematosus. *Am. J. Hum. Genet.*, **76**, 528–537.
37. Graham, R.R., Kyogoku, C., Sigurdsson, S., Vlasova, I.A., Davies, L.R., Baechler, E.C., Plenge, R.M., Koeuth, T., Ortmann, W.A., Hom, G. *et al.* (2007) Three functional variants of IFN regulatory factor 5 (IRF5) define risk and protective haplotypes for human lupus. *Proc. Natl Acad. Sci. USA*, **104**, 6758–6763.
38. De Jager, P.L., Jia, X., Wang, J., de Bakker, P.I., Ottoboni, L., Aggarwal, N.T., Piccio, L., Raychaudhuri, S., Tran, D., Aubin, C. *et al.* (2009) Meta-analysis of genome scans and replication identify CD6, IRF8 and TNFRSF1A as new multiple sclerosis susceptibility loci. *Nat. Genet.*, **41**, 776–782.
39. 1980) Preliminary criteria for the classification of systemic sclerosis (scleroderma). Subcommittee for scleroderma criteria of the American Rheumatism Association Diagnostic and Therapeutic Criteria Committee. *Arthritis Rheum.*, **23**, 581–590.

40. Marchini, J., Howie, B., Myers, S., McVean, G. and Donnelly, P. (2007) A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat. Genet.*, **39**, 906–913.
41. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J. *et al.* (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.*, **81**, 559–575.
42. Barrett, J.C., Fry, B., Maller, J. and Daly, M.J. (2005) Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*, **21**, 263–265.
43. Tian, C., Kosoy, R., Nassir, R., Lee, A., Villoslada, P., Klareskog, L., Hammarstrom, L., Garchon, H.J., Pulver, A.E., Ransom, M. *et al.* (2009) European population genetic substructure: further definition of ancestry informative markers for distinguishing among diverse European ethnic groups. *Mol. Med.*, **15**, 371–383.