Università del Piemonte Orientale
"A. Avogadro"

Tesi di dottorato in Medicina Molecolare
XXIII ciclo (2007-2010)

RICERCA DI FATTORI GENETICI DI SUSCETTIBILITA' AL LUPUS ERITEMATOSO SISTEMICO

Tutor: ***Prof.sa. Mara Giordano***
Prof.sa Sandra D'Alfonso
Coordinatore: ***Prof. Umberto Dianzani***
Studente: ***Dott.ssa Nadia Barizzone***

... fatti non foste a viver come bruti,
ma per seguir virtute e canoscenza.
(Inferno, canto XXVI)

INDICE

INTRODUZIONE **1**
LUPUS ERITEMATOSO SISTEMICO 1
EPIDEMIOLOGIA 2
CLINICA 3
QUADRO IMMUNOLOGICO 6
DIAGNOSTICA 11
COMPONENTE AMBIENTALE 12
SUSCETTIBILITA' GENETICA AL LES 13
MODELLI MURINI 30

SCOPO DEL LAVORO **33**

RISULTATI **35**
STUDI DI ASSOCIAZIONE E/O FUNZIONALI SU GENI CANDIDATI 36
ANALISI DI FATTORI GENETICI DI SUSCETTIBILITÀ LOCALIZZATI NELLA REGIONE CROMOSOMICA 1Q32-43 43
ANALISI MUTAZIONALE 62
CREAZIONE DI UN ALGORITMO GENETICO DI RISCHIO 72
META-ANALISI SUI GENI TEC E BMPR1B 81
ALTRI ARTICOLI 87

DISCUSSIONE **89**

MATERIALI E METODI **106**

BIBLIOGRAFIA **115**

RINGRAZIAMENTI 125

ARTICOLI ALLEGATI **126**

INTRODUZIONE

LUPUS ERITEMATOSO SISTEMICO

Il *Lupus Eritematoso Sistemico* (LES) è una malattia autoimmune cronica a eziologia multifattoriale. Il termine fu coniato agli inizi del diciannovesimo secolo per descrivere una malattia cutanea caratterizzata da lesioni eritematose. L'origine della parola Lupus risale al Medio Evo, quando era usata per indicare lesioni sfiguranti della cute del volto. Alla fine del 1800 un famoso medico americano, sir William Osler capì per primo di trovarsi davanti ad una malattia sistemica, che non sempre si accompagnava a manifestazioni cutanee e la denominò *Systemic Lupus Erythematosus*. Il LES venne a lungo considerata una malattia rara, poichè venivano riconosciuti solo i casi più gravi e conclamati. Nel 1948 un gruppo di medici della Mayo Clinic descrisse un test diagnostico per il LES, il test delle cellule LE (globuli bianchi che hanno inglobato il nucleo di altre cellule), che ha facilitato la diagnosi in fase precoce della malattia. Tra il 1950 e il 1992 i casi di LES riportati negli USA sono aumentati da 3 a 7 volte, probabilmente per via della disponibilità di test più sensibili e di criteri diagnostici standardizzati (Borchers et al. 2010).

EPIDEMIOLOGIA

Come si può osservare dalla tabella I1, la prevalenza e l'incidenza del LES variano a seconda della popolazione. Negli Afro-Americani, la popolazione con il rischio più elevato, l'incidenza è di circa 3 volte superiore rispetto alle popolazioni caucasiche (Borchers et al. 2010). Nella popolazione italiana la prevalenza del LES è paragonabile a quella osservata nel resto dell'Europa, ed è di 1/1408 (Benucci et al.2005). La prevalenza su scala mondiale del LES è di circa 1/2.000. Confrontandola con quella di altre malattie autoimmuni si vede che essa è circa uguale a quella della Sclerosi Multipla, ma da 5 a 10 volte minore di quella del Diabete Mellito di tipo 1.

Il LES è molto più diffuso tra le donne, soprattutto durante l'età fertile, quando il rapporto di incidenza donna/uomo è approssimativamente di 9:1, mentre se la malattia si manifesta in età pediatrica o dopo la menopausa tale rapporto è molto più basso (rispettivamente di circa 2:1 e 3:1). Ciò rafforza l'ipotesi di un coinvolgimento degli ormoni sessuali nello sviluppo della malattia. (Kotzin 1996; Simard et al. 2007). La patologia può insorgere a ogni età, ma più frequentemente tra gli 11 e i 50 anni (Cervera et al. 1993).

Negli anni '50 la sopravvivenza a cinque anni dalla diagnosi era approssimativamente del 50%. Oggi è stata migliorata dall'uso di corticosteroidi esogeni e dalla terapia immunosoppressiva, ed è paragonabile a quella del resto della popolazione (Simard et al. 2007).

Tab. I1 Incidenza e prevalenza del LES nei recenti studi epidemiologici (riassunti da Borchers et al. 2010).

Paese	anni	età	etnia	Incidenza totale	Incidenza donne	Incidenza uomini	Prevalenza totale	Prevalenza donne	Prevalenza uomini
USA, rural Wisconsin	1991–2001	tutte	tutte	5.1	8.2	1.9	78.5	131.5	24.8
USA, CA		≥ 18	tutte				107.6	184.2	25.5
			Afro - Americani					406.3	
			Ispanici					138.7	
			Asian/Pacific Islander					92.7	
			Caucasici					164.4	
USA, PA			tutte				149.5	253.0	38.7
			Afro - Americani					693.7	
			Ispanici					244.5	
			Asian/Pacific Islander					103.2	
			Caucasici					203.1	
Canada, Manitoba	1980–1996		Nativi Americani	2.0–7.4			42.3		
			Caucasici	0.9–2.3			20.6		
Canada, Quebec	1994–2003	tutte	tutte	2.8–3.0			44.7		
UK, Birmingham	1991	≥ 18	tutte	3.8	6.8	2.0	27.7		
			Afro-Caraibici	11.9	22.8	0.5	111.8	197.2	6.4
			Asiatici	15.2	29.2		46.7	96.5	4.3
			Caucasici	2.5	4.5		20.7	36.3	3.4
UK, Nottingham	1989–90	tutte	tutte	4.0	6.5	1.5	24.6	45.4	3.7
			Afro-Caraibici	31.9			207.0		
			Asiatici	4.1			48.8		
			Caucasici	3.4			20.3		
UK	1990–99	tutte	tutte	4.7	7.9	1.3			
Australia, centrale	1999		Aborigeni				73.5	122.5	24.5
			Australiani Europei				19.3	32.2	6.4
Danimarca, Funen	1995–2003		tutte	1.0			28.3		
Norvegia, Artico	1996–2006	≥ 16		3.0	5.1	0.9	64.1	108.6	20.0
Norvegia, Artico	1978–96	≥ 16	tutte	2.6	4.6	0.6	49.7	89.3	9.7
Grecia	1982–2001	tutte	tutte	1.9			38.1		
Martinica	1990–99	tutte	tutte (Afro-Caraibici)	4.7	8.5	0.7	64.2	115	9.2
Curacao	1980–89		tutte	4.6			47		
Brasile	2000	≥ 15	tutte	8.7	14.1	2.2			

CLINICA

La sintomatologia del LES è molto complessa, e può variare molto da un paziente ad un altro. Nessun paziente manifesta la gamma completa di manifestazioni cliniche, e i sintomi vanno incontro a fenomeni di remissione ed acutizzazione. E' una malattia sistemica, per cui nelle forme più gravi tutti gli organi possono essere coinvolti e la salute del paziente seriamente compromessa, mentre le forme più leggere sono limitate a manifestazioni cutanee e a lievi dolori articolari, che tendono a regredire spontaneamente. (Kotzin, 1996)

I sintomi più comuni e caratteristici sono l'eritema malare fisso "a farfalla", con aspetto piatto o rilevato che insorge sul volto del paziente nelle pieghe naso-labiali e l'artrite, che coinvolge solitamente 2 o più articolazioni e può provocare dolore molto intenso, spesso accompagnato da

stanchezza e debolezza muscolare. Ma la malattia può causare molte altre manifestazioni cliniche, che vengono qui sotto elencate:

-Lesioni discoidi: placche eritematose rilevate, di forma rotondeggiante, che interessano il volto, i padiglioni auricolari ed il cuoio capelluto. Possono provocare alopecia, che comunque tende a verificarsi durante le fasi di attività della malattia;

-Lesioni cutanee subacute: dermatiti fotosensibili che possono avere un aspetto psoriasiforme o anulare ;

-Fotosensibilità;

-Ulcere al cavo orale o nasofaringeo, di solito non dolorose, rilevate dal medico;

-Sierositi: pleuriti o pericarditi;

-Nefropatie: glomerulonefriti che causano ematuria e proteinuria persistente, a livello ematico si ha azotemia ed un aumento dei livelli di creatina e acido urico. Il coinvolgimento renale è una delle manifestazioni più gravi della malattia;

-Disturbi neurologici: accessi epilettici con convulsioni generalizzate e psicosi non provocati da farmaci o da alterazioni metaboliche note, disturbi della personalità, cefalea, disturbi di tipo ischemico dovuti allo scarso afflusso di sangue al cervello. Fortunatamente i disturbi psicotici veri e propri sono rari, e le convulsioni epilettiche sono meno frequenti che in passato, anche grazie alla soppressione della terapia con clorochina, dotata di proprietà convulsivanti anche a dosi terapeutiche;

-Leucopenia, dovuta a anticorpi diretti contro i leucociti, che ne facilitano la rimozione da parte della milza. Si ha un calo significativo e persistente sia dei neutrofili che dei linfociti.

-Trombocitopenia, dovuta a anticorpi diretti contro le piastrine. Può provocare epistassi, sanguinamennto gengivale o emorragie gravi;

-Anemia: in alcuni casi è dovuta alla presenza di anticorpi contro gli eritrociti (anemia emolitica), in altri da un’alterazione nel trasporto del ferro, che non è quindi disponibile per la sintesi di emoglobina;

-Febbre abbastanza elevata (>38°C) in assenza di infezioni;

-Fenomeno di Raynaud: sbiancamento delle estremità indotto dall’esposizione al freddo o da stress di altra natura;

-Livedo reticularis: decolorazione rossastra o cianotica della pelle;

-Trombosi di vene e/o arterie;

-Miosite: debolezza muscolare;

-Disturbi polmonari: polmoniti acute o croniche (polmoniti lupiche). Si tratta comunque di manifestazioni rare;

-Movimenti coreici;

-Secchezza delle congiuntive;

-Linfoadenopatia: rigonfiamento dei linfonodi nelle regioni cervicale, ascellare e inguinale in assenza di infezione. Meno frequentemente si ha ingrossamento della milza.

In uno studio condotto su 1000 pazienti Cervera et al. hanno raccolto una serie di dati sulla distribuzione percentuale dei vari sintomi (tab.I2), ed hanno notato alcune differenze nella prevalenza delle manifestazioni cliniche in rapporto all'età di insorgenza e al sesso (tab.I3). I pazienti studiati erano quasi tutti appartenenti a popolazioni caucasoidi. Essi osservarono inoltre che nel 6% dei casi il LES si accompagnava ad un'altra malattia autoimmune.

Tab. I2. Manifestazioni cliniche in un gruppo di 1000 pazienti di SLE all'esordio e durante l'evoluzione clinica (da Cervera et al., 1993).

Manifestazione clinica	All'esordio (%)	Durante l'evoluzione (%)
Eritema malare	40	58
Lesioni discoidi	6	10
Lesioni cutanee subacute	3	6
Fotosensibilità	29	45
Ulcere orali	11	24
Artrite	69	84
Sierosite	17	36
Nefropatia	16	39
Manifestazioni neurologiche	12	27
Trombocitopenia	9	22
Anemia emolitica	4	8
Febbre	36	52
Fenomeno di Raynaud	18	34
Livido reticolare	5	14
Trombosi	4	14
Miosite	4	9
Manifestazioni polmonari	3	7
Corea	1	2
Secchezza della congiuntive	5	16
Linfoadenopatia	7	12

Tab. I3. Manifestazioni cliniche in un gruppo di 76 pazienti di SLE con esordio in età infantile (prima dei 14 anni), e in un gruppo di 90 pazienti con esordio dopo i 50 anni di età (da Cervera et al., 1993).

	Esordio in età infantile		**Esordio in età avanzata**	
Manifestazione clinica	**All'esordio**	**Durante l'evoluzi**	**All'esordio**	**Durante l'evoluzione**
Eritema malare	42 (55%) *****	60 (79%) *****	19 (21%) *****	30 (33%) *****
Nefropatie	21 (28%) *****	35 (46%)	3 (3%) *****	20 (22%) *****
Artrite	49 (64%)	66 (87%)	47 (52%) *****	66 (73%) *****
Fotosensibilità	25 (33%)	36 (47%)	18 (20%)	26 (29%) *****
Trombosi	3 (4%)	12 (16%)	1 (1%)	4 (4%) *****
Secchezza delle fauci	2 (3%)	7 (9%)	5 (6%)	30 (33%) *****

I parametri .in tabella si manifestano con frequenze diverse nel LES ad esordio tardivo o precoce rispetto ai pazienti con esordio in età fertile (dati contrassegnati con * = $p<0,05$). Per ogni categoria è indicata la numerosità del campione e, in parentesi, il suo valore percentuale.

Il Lupus nei bambini è una malattia molto grave, con prognosi peggiore che nell'adulto. Come si osserva dai dati riportati le manifestazioni renali sono nel bambino molto più frequenti. Inoltre essa viene spesso diagnosticata con notevole ritardo, poiché spesso all'esordio non vi sono sintomi evidenti, e non ci si accorge della nefropatia finché questa non ha causato danni irreparabili. Anche gli effetti collaterali dei farmaci usati per la terapia possono essere più gravi nel bambino. Invece le differenze tra la clinica del LES negli uomini e nelle donne sono molto più sfumate.

Tra le donne affette da LES esiste una notevole incidenza di aborti spontanei, soprattutto intorno al terzo mese di gravidanza e se la malattia è in fase di attività. E' raro che il bambino erediti dalla madre la suscettibilità alla malattia, ma il neonato talvolta manifesta alcuni sintomi transitori dovuti ad anticorpi materni che attraversano la barriera placentare. Non è ancora chiarito se i pazienti con LES siano abbiano un rischio maggiore di contrarre il cancro, a causa del loro stato immunitario alterato o secondariamente alla terapia immunosoppressiva (Ginzler et al. 1992).

QUADRO IMMUNOLOGICO

Il quadro immunologico del LES è complesso e variato come quello clinico. Esistono diversi meccanismi con i quali può verificarsi l'autoaggressione immunitaria, che possono coesistere nello stesso paziente. Anche in questo caso si sono trovate differenze nella prevalenza degli anticorpi prodotti a seconda dell'età di insorgenza della malattia. Non si sa in modo definitivo se la disregolazione immunologica sia la causa o la conseguenza della malattia.

Contributo della componente umorale Alcune manifestazioni cliniche della malattia sono dovute alla produzione di anticorpi che interagiscono con il complemento e vanno a legarsi alle cellule

ematiche favorendone la distruzione da parte dei fagociti della milza e del fegato. Tra la moltitudine di autoanticorpi prodotti nel LES ricordiamo:

- anticorpi specifici per le componenti nucleari contenenti DNA; (ANA: anti-nuclear antibodyes) sono la categoria prevalente e si distinguono in quelli diretti contro il ssDNA (singola catena), dsDNA (doppia catena), contro la cromatina o contro i complessi acido nucleico-proteina (DNA-istone);
- anticorpi diretti contro antigeni nucleari solubili non contenenti DNA (*anti-ENA*);
- ribonucleo-proteina (U1RNP, le snRNP, Sm, Ro/SSA e p73, La/SSB) o proteine ribosomiali P. Tra questi i più importanti sono quelli diretti contro la ribonucleoproteina U1RNP e quelli diretti contro l'antigene Sm; i primi caratterizzano la connettivite mista (sindrome di Sharp) ma sono anche presenti nel SLE (40% dei casi); i secondi sono specifici del SLE ma presenti solo nel 25% dei casi;
- anticorpi antiribosomiali, diretti contro antigeni ribosomiali (P01, P1, P2) compaiono nel 15% dei pazienti affetti da SLE ed in particolare nei soggetti con forma patologica neuropsichiatrica (sindrome depressiva);
- anticorpi anti-fosfolipidi (*aPL*), definiti anche anti-cardiolipina (*aCL*) in quanto dimostrabili utilizzando la cardiolipina nel metodo ELISA. Essi sono causa di manifestazioni trombotiche vascolari;
- autoanticorpi diretti contro le molecole di superficie delle cellule (anticorpi diretti contro le cellule del sangue costituiti da anti-globuli rossi, anti-globuli bianchi, antilinfociti ed anti-piastrine); essi sono implicati nello sviluppo dell'anemia emolitica, della leucopenia autoimmune, della trombocitopenia, riscontrabili in molti pazienti di LES. Questi anticorpi distruggendo le cellule bersaglio scatenano alcune manifestazioni cliniche.
-

Tab I4 quadro immunologico riscontrato in una coorte di 1.000 pazienti (Cervera et al, 1993)

PARAMETRO	Percentuale (%)
ANA	96
Anti-dsDNA	78
Ro (SSA)	25
La (SSB)	19
U1-snRPN	13
Sm	10
Fattore reumatoide	18
IgG aCL	24
IgM aCL	13
Anticoagulante del Lupus	15

Va ricordato che esiste anche il LES ANA negativo (4% circa dei casi). Questi pazienti presentano almeno 4 segni clinici di malattia, ma sono negativi per gli anticorpi anti-dsDNA. In 2/3 dei casi si evidenzia una positività di anticorpi anti-ENA (Ro/SS-A e La/SSB). Si tratta, con molta probabilità

di forme di passaggio (overlapsyndromes) tra SLE e altre connettiviti con prognosi variabile da caso a caso Nella maggioranza dei casi la causa principale del danno è dovuta alla formazione di immunocomplessi (IC) tra anticorpi IgG ed antigeni solubili che precipitano a livello dei tessuti. Gli autoanticorpi maggiormente candidati a dare questo tipo di manifestazione sono diretti contro il dsDNA, che compaiono nella fase attiva. Una volta formatisi gli IC legano ed attivano il sistema del complemento inducendo la risposta infiammatoria. A questo punto i monociti ed in particolare i neutrofili, attratti nel luogo dell'infiammazione, attuano la rimozione degli IC fagocitandoli e degradando le molecole inglobate. Svolto il loro compito i neutrofili muoiono nel sito dove hanno compiuto l'attacco e rilasciano in questo modo una grande quantità di enzimi che possono provocare danni connettivali.

In condizioni normali la formazione e la rimozione degli IC avvengono fisiologicamente nell'organismo senza provocare conseguenze patologiche, ma nei malati di LES essi si formano soprattutto per interazione con gli autoantigeni, che vengono continuamente prodotti e riconosciuti dagli specifici autoanticorpi. Di conseguenza la concentrazione degli IC aumenta, per cui questi vengono rimossi con maggiore difficoltà e tendono ad accumularsi in circolo e a precipitare in corrispondenza delle principali barriere interposte alla circolazione: a livello renale, della barriera emato-encefalica, della rete capillare. E' proprio l'infiammazione cronica che si produce a questi livelli a rappresentare la causa più frequente di danno tessutale (Cervera et al. 1993; Kotzin et al. 1996; Abbas et al. 1997).

Contributo della componente cellulare A livello cellulare sia le cellule B sia le cellule T sembrano essere interessate allo sviluppo della malattia. La maggior parte degli effetti patogeni nel LES sono apparentemente mediati dall'eccessiva produzione di anticorpi IgG da parte dei linfociti B, i quali, sia in alcuni pazienti che nei modelli murini, sono risultati iperresponsivi a citochine Bstimolatorie, come l'IL-5 (da sola o associata con l'IL-4) e l'IL-6. E' stato dimostrato che quest'ultima induce le cellule B a produrre anticorpi IgG anti-DNA. Inoltre, a differenza di quanto avviene negli individui normali, la risposta proliferativa dei linfociti B all'interleuchina-4 non è inibita dall'interferone gamma (Via et al. 1992).

Evidenze sperimentali supportano la conclusione che nel LES la produzione di anticorpi patogenici sia selettiva solo per determinati autoantigeni, responsabili dell'attivazione dei linfociti B autoreattivi (Chen et al. 1995). In uno studio condotto con cellule murine dirette contro anticorpi anti- Sm e contro ssDNA è stato osservato che la maggior parte delle cellule B autoreattive sono immmature (Santulli-Marotto et al. 1998). Anche le cellule T sono ritenute di importanza cruciale nello sviluppo del LES, da ricondurre ad una alterazione della loro azione regolatoria nei riguardi dei linfociti B piuttosto che ad una azione dannosa diretta contro i tessuti.

Friedman et al hanno ipotizzato che prodotti microbici, comportandosi come superantigeni, possano indurre il LES in un modo analogo all'induzione sperimentale della malattia lupus-simile GVHD nel topo. Questi superantigeni si legano con alta affinità alle molecole del MHC di classe II, per cui gli autori hanno ipotizzato che possano mediare l'iperattività dei linfociti B formando un ponte tra cellule B esponenti molecole MHC II ed una ristretta famiglia di cellule T TCR-positive. Ciò potrebbe generare la trasmissione non specifica di segnali dai linfociti T helper ai linfociti B. In questo modo cellule B già attivate dal legame crociato di antigeni autologhi con le Ig di superficie riceverebbero il "secondo segnale" necessario per la produzione di anticorpi.

La frequenza di cellule T g/d nel sangue periferico di pazienti è stata registrata come superiore alla norma (Via et al, 1992).

Ruolo delle citochine Un contributo allo sviluppo della malattia può essere a carico di varianti alleliche particolari nei geni che codificano per le diverse citochine, per i fattori del complemento e per le molecole del MHC.

La vasta gamma di citochine prodotte da cellule del sistema immunitario, come monociti e linfociti, ha un ruolo di fondamentale importanza nella regolazione della risposta immune. Le citochine pro-infiammatorie come le interleuchine 1a e 1b, il fattore di necrosi tumorale alfa (TNFa), le interleuchine 6 e 2 (IL-6, IL-2), e l'interferone gamma (INF-γ) sono in grado di stimolare la risposta infiammatoria attivando la risposta immune cellulare che può portare alla degradazione del tessuto connettivo, tipica del LES e di altre malattie autoimmuni come per esempio l'artrite reumatoide (Crawley et al, 1999).

Nel LES, come già discusso, sono di principale importanza le citochine in grado di attivare i linfociti B, per esempio alcune interleuchine come le IL-4, IL-6, IL-10. Alcuni fratelli sani di malati di SLE mostrano iperattività dei linfociti autoreattivi, sebbene gli anticorpi prodotti presentino una bassa affinità per le molecole del self e non siano patogenici (Miles et al, 1993); una gran parte, invece, mostra una diminuzione dell'immunità cellulo-mediata, un decremento della produzione di IL-12 ed un'iperattività dei linfociti B policlonali (Clark et al, 1996).

Importanza del complemento nella patogenesi del LES Il coinvolgimento del complemento nella patogenesi del LES è originariamente emerso dall'osservazione dei diminuiti livelli delle componenti della via classica di attivazione del complemento nei pazienti con la malattia in fase attiva, e dei depositi di componenti del complemento ed immunoglobuline a livello degli organi interessati dalla malattia, come la pelle ed i reni (Sturfelt & Truedsson 2005). Il legame degli autoanticorpi ai loro antigeni, con conseguente attivazione del complemento, è la causa scatenante del danno tessutale. Il ruolo del complemento è particolarmente importante nel determinare la sintomatologia a livello renale. L'attivazione del complemento e la sua deposizione nei glomeruli è

un fattore determinante anche in molti tipi di glomerulonefriti e di altre patologie renali (Brown et al. 2007). Il paradosso che caratterizza il LES risiede nel fatto che il deficit completo di una delle componenti della via classica del complemento non protegge dal lupus, ma anzi causa una forte predisposizione ad esso (vedi suscettibilità genetica al LES, complemento) Difetti acquisiti della via classica sono più comuni rispetto a quelli ereditari, e sono associati ad un aumento nell'attività della malattia (Boackle 2003).

Sono state avanzate numerose ipotesi per spiegare il meccanismo patogenetico. Una delle principali funzioni del complemento è quella di promuovere la rimozione degli IC, circolanti o depositati a livello dei tessuti, da parte dei fagociti mononucleati. Uno scompenso di questa funzione può causare un ciclo autostimolantesi, in cui gli IC depositati nei tessuti e non rimossi causano infiammazione e rilascio di autoantigeni, i quali a loro volta stimolano la produzione di autoanticorpi con la conseguente formazione di ulteriori IC (Morgan et al. 1991).

Le cellule apoptotiche sono ritenute la principale fonte di autoantigeni nel LES. C1q è in grado di opsonizzare direttamente le cellule in apoptosi, o di legarle attraverso altre opsonine come le IgM e la proteina C-reattiva (CRP). Il legame con C1q porta all'attivazione della via classica del complemento, e di conseguenza all'opsonizzazione della cellula apoptotica ed alla sua fagocitosi. Un difetto nel sistema di eliminazione delle cellule apoptotiche può portare all'esposizione di antigeni self, causando l'attivazione dei linfociti B e T ed alla produzione di autoanticorpi (Flierman et al. 2007). Inoltre i corpi apoptotici opsonizzati da C3b sono fagocitati dai macrofagi ed inducono la produzione di citochine anti-infiammatorie come TGFb, assicurando la rimozione degli IC in assenza di una risposta immune (Boackle 2003). Questa funzione viene a mancare in individui con un deficit del complemento.

Secondo un'altra ipotesi il sistema del complemento, insieme ad altre componenti dell'immunità innata, protegge contro la risposta immune agli auto antigeni partecipando alla presentazione degli antigeni self ai linfociti B immaturi stimolando così l'eliminazione delle cellule B autoreattive. In questo meccanismo interverrebbe anche il legame degli antigeni rivestiti di particelle del complemento con i recettori membranari CR1 e CR2 (vedi proteine regolatorie del complemento). Un deficit nel sistema del complemento potrebbe dunque causare una ridotta tolleranza al self dei

linfociti B (Carrol 2004) Un'altra possibilità è che alcune componenti del complemento siano importanti per la produzione di citochine coinvolte nella patogenesi del LES. Per esempio Yamada et al. 2004 dimostrano che C1q sopprime la produzione di IL-12 nelle cellule dendritiche.

Infine da un deficit del complemento potrebbe risultare una presenza insufficiente di appropriati ligandi per CR2, situazione che potrebbe causare un fenotipo simile a quello caratteristico di un deficit o di una disfunzione di CR2 stesso (Boackle 2003). E' stata analizzata anche l'ipotesi che un

deficit del complemento causi infezioni virali persistenti, le quali potrebbero avere un ruolo nella patogenesi del LES (vedi componente ambientale).

Nel siero di pazienti affetti da LES si trovano spesso autoanticorpi diretti contro le componenti della via classica del complemento. I più frequentemente riscontrati sono gli anticorpi anti-C1q (30-50% dei casi), ma sono stati osservati anche pazienti con autoanticorpi anti-C1s, anti-C1INH, anti-C3 anti-C4 ed anti-CR1. La maggior parte dei pazienti con anticorpi anti-C1q presenta nefrite (Sturfelt & Truedsson 2005).

La figura I1 schematizza un possibile ruolo di una disfunzione del complemento (CD) nella patogenesi del LES. I raggi ultravioletti (UV) e le infezioni virali sono esempi di fattori di natura ambientale che possono contribuire ad innescare il meccanismo patogenetico.

DIAGNOSTICA

La notevole eterogeneità clinica ed immunologica pone non poche difficoltà diagnostiche, per ovviare alle quali l'*American Rheumatism Association* (ARA) ha definito una serie di criteri clinici e di laboratorio che sembrano essere sufficientemente specifici e sensibili per la diagnosi della malattia. Inizialmente furono selezionati 14 criteri (su 74 studiati), ma vennero successivamente rivisti e ridotti ad 11. Essi comprendono molti dei sintomi più comuni elencati in 5.2 ed alcuni criteri immunologici:

-disordini immunologici: cellule LE (globuli bianchi che hanno inglobato il nucleo di altre cellule) o anticorpi anti-dsDNA o anti-Smith (anti Sm) o falsa positività dei tests sierologici per la sifilide;

-positività per anticorpi anti-nucleo, non causata dall'assunzione di farmaci.

La positività di quattro o più criteri, manifestati contemporaneamente o in successione durante il decorso della malattia, permette di diagnosticare correttamente il LES rispetto alle altre malattie connettivali con un margine di errore minore del 5%.

I pazienti che presentano solo due criteri ARA vengono definiti "LES possibili", quelli con tre criteri "LES probabili" (anche definiti "LES incompleti", "Lupus like", "LES latenti"). Solamente i casi con quattro criteri vengono definiti "LES definitivi".

La sensibilità della diagnosi è piuttosto elevata in quanto i criteri ARA fanno diagnosticare la malattia nel 96% dei casi (Wallace et al, 1993; Hochberg et al, 1997).

Nei criteri ARA verranno probabilmente inseriti, in futuro, i bassi livelli di C3, C4 e la positività elevata degli anticorpi anti-fosfolipidi ed in particolare anti-cardiolipina (aPL/aCL) che hanno assunto di recente una notevole importanza nel definire varianti del LES classico. Il dosaggio dell'attività emolitica CH50 del C3 e del C4 sembrano infatti elementi importanti nella diagnosi e nel monitoraggio del LES. La presenza contemporanea di anticorpi anti-nucleo (ANA) positivi e

bassi livelli di complemento sono quasi certamente connessi con un LES in fase acuta o con un pre-LES. Una caduta dei valori del complemento nel decorso della malattia indica in genere una ricaduta clinica.

COMPONENTE AMBIENTALE

Come noto anche la patogenesi del LES, così come quella delle altre malattie multifattoriali, è causata da fattori sia genetici sia ambientali, anche se il ruolo di questi ultimi non è del tutto chiarito. Di certo si sa che l'esposizione ad alcuni di questi fattori può provocare un'esacerbazione dei sintomi preesistenti.

I principali elementi di natura ambientale coinvolti nel LES sono l'esposizione a raggi U.V. e gli agenti infettivi, in particolar modo quelli di natura virale. I raggi ultravioletti esplicano una doppia azione i cui effetti finiscono per interagire in modo sinergico: causano la frammentazione del nucleo e del dsDNA delle cellule epidermiche, liberando così una elevata quantità di antigeni nucleari, ed attivano la risposta immunitaria a livello cutaneo. L'esposizione prolungata al sole in soggetti suscettibili può quindi costituire la causa scatenante, che porta alla rottura della tolleranza verso il *self*. Nei malati provoca di solito la recrudescenza del rash cutaneo.

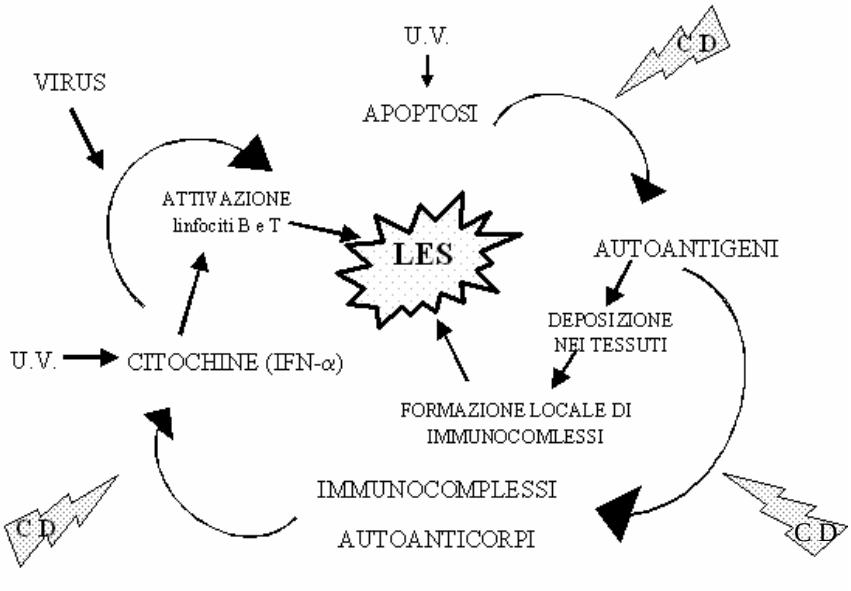
Gli agenti infettivi possono provocare la malattia con un meccanismo simile: uno stimolo eccessivo al sistema immunitario e la presentazione di antigeni che talvolta possono essere molto simili a quelli autologhi, fino a causare la perdita della specificità e comportarsi come autoantigeni. Virus e batteri possono inoltre svolgere le funzioni di "superantigeni", facendo da ponte tra un linfocita T helper ed una cellula B. L'azione iniziante dei microorganismi si manifesta più facilmente in caso di infezioni croniche o comunque protratte nel tempo. Ciò spiega l'aumentata frequenza di malattie autoimmuni nei soggetti con deficit immunitari. (Kotzin, 1996). Gli agenti infettivi che presentano le maggiori evidenze di associazione con il LES sono il virus di Epstein-Barr (EBV) ed il citomegalovirus (CMV) (Hrycek et al. 2005), ma i dati sono ancora di difficile interpretazione (Simard et al. 2007).

Sono stati condotti vari studi riguardanti il possibile effetto del fumo di sigaretta, del consumo di alcool etilico, di fattori nutrizionali e di contaminanti ambientali, ma i dati sono contrastanti (Simard et al. 2007).

Anche molti farmaci (come la procainamide, responsabile della maggior parte dei casi), possono causare una malattia simile al LES in soggetti predisposti, ma il meccanismo con cui questo avviene non è chiaro. Essi possono agire direttamente modificando la risposta immunitaria o comportarsi come "mimetici" di antigeni dell'ospite. I sintomi del LES indotto da farmaci sono normalmente lievi e di solito retrocedono dopo 4-6 settimane dalla sospensione dell'agente causale. I medicinali

capaci di indurre il Lupus non sembrano aggravare il LES primario. E' dato quasi per certo un importante effetto degli ormoni steroidei: iniziante nel caso degli estrogeni e protettivo nel caso del testosterone. Questa ipotesi è confermata anche dai cambiamenti nelle manifestazioni cliniche dopo la somministrazione di ormoni esogeni (Mills, 1994). Sono state avanzate anche numerose ipotesi circa il coinvolgimento di alcuni cosmetici e delle protesi al silicone (Simard et al. 2007).

Fig. I1. Patogenesi del LES



Rappresentazione schematica del ruolo di alcuni fattori ambientali ed immunitari nella patogenesi del LES. Si noti il ruolo centrale che ha il deficit di fattori del complemento**.** CD = Complement Deficit. Adattata da Sturfelt & Truedsson 2005

SUSCETTIBILITA' GENETICA AL LES

Si hanno da tempo considerevoli evidenze che lo sviluppo del LES abbia una forte base genetica, derivanti dai risultati degli studi epidemiologici, di concordanza sui gemelli e di aggregazione famigliare.

Studi di aggregazione famigliare Viene valutata paragonando il rischio per un fratello o un altro parente del malato di contrarre la malattia con quello per il resto della popolazione. Tale valore (denominato λ) per il LES è stimato tra 20 e 80, e dipende dalla popolazione presa in esame. Il valore λ per il Lupus è leggermente più alto di quello stimato per altre malattie autoimmuni, come la Sclerosi Multipla (λ=20) e l'IDDM (λ=15), ma è circa 10 volte inferiore a quello di malattie monogeniche relativamente frequenti, come la Fibrosi Cistica (λ=500) (Lindqvist et al. 1999).

Studi sui gemelli Nel caso del LES i gemelli vengono analizzati in base ai criteri ARA, per ottenere diagnosi certe di positività o negatività. Nonostante la prevalenza abbastanza elevata non sono stati pubblicati molti studi di concordanza sui gemelli relativi al LES.

Vecchi dati, basati su un piccolo numero di casi, attestavano il tasso di concordanza in gemelli MZ al 69%, molto più alto di quello di altre malattie multifattoriali autoimmuni. Ma nel 1992 uno studio più accurato condotto da Deapen et al ha abbassato tale valore al 24%. Sempre nello stesso studio il tasso di concordanza nei dizigotici è risultato del 2%. Dati successivi hanno confermato questi valori. Il tasso di concordanza relativo alla presenza di anticorpi sierici può essere molto superiore(fino al 92% in gemelli monozigotici) (Lindqvist et al. 1999; Deapen et al. 1992 e Grennan et al. 1997).

Negli ultimi anni la comprensione della genetica del LES ha fatto notevoli passi in avanti grazie soprattutto all'impiego di studi di tipo GWAS (Genome Wide Association Studies) su ampie casistiche. Nella suscettibilità alla malattia giocano un ruolo tre tipi diversi di varianti geniche:

1) Variazioni rare ma ad alta penetranza. Includono sostanzialmente il deficit in omozigosi di alcune componenti della via classica di attivazione del complemento ma, più recentemente, sono state descritte mutazioni nella sequenza aminoacidica del gene *TREX1*. Queste forme hanno permesso di fare luce sui meccanismi patogenetici, ma spiegano solo una piccola porzione dell'incidenza complessiva della patologia.

2) Copy Number Variations (CNV). Per i prossimi anni sono attesi studi esaustivi su tutto il genoma delle CNV nel LES. Al momento sono note le associazioni con le CNV nella regione dei recettori per il frammento costante (Fc) delle immunoglobuline e del fattore C4 del complemento. Tuttavia, a causa della presenza in queste regioni di altre varianti di rischio, il ruolo delle CNV non è ancora stato chiarito in modo convincente.

3) Polimorfismi comuni, che contribuiscono alla suscettibilità al lupus ciascuno con valori di Odd Ratio molto bassi (OR<2). La presenza di più fattori comuni di rischio è responsabile della maggior parte dei casi di LES.

Come per la maggior parte delle malattie multifattoriali, anche nel caso del LES per la ricerca di fattori comuni di suscettibilità sono state adoperate tre strategie, cioè, in ordine "storico", studi di Linkage famigliari, studi di associazione di tipo caso-controllo su geni candidati, e, più recentemente, analisi di tipo GWAS.

Studi di linkage Sono stati prodotti almeno 11 studi di linkage su tutto il genoma (Whole Genome Linkage Studies) sul LES, più diversi studi incentrati su di una regione specifica e volti a confermarne il linkage con la malattia. Poiché i geni che conferiscono suscettibilità alla malattia

sono probabilmente diversi a seconda della popolazione, gruppi di ricercatori diversi hanno riportato prove di linkage in regioni cromosomiche diverse. Tra tutte le regioni con linkage suggestivo segnalate, solo 9 superano la soglia di significatività proposta da Lander e Kruglyak (lod score >3,3) e sono state confermate in almeno uno studio indipendente. Per alcune di queste regioni è stata riportata anche un'associazione significativa con almeno uno dei geni che vi mappano. Solo tre di queste associazioni paiono però convincenti (*PDCD1, HLA-DR* e *FCGR*). Le evidenze più significative si hanno per una vasta regione del cromosoma 1 (1q22-1q42),che ha evidenziato un linkage significativo in più popolazioni geneticamente diverse. In questa regione mappano il gruppo dei geni per i recettori del frammento costante delle Ig (FCGR) ed altri geni che, come esposto nel capitolo sugli studi di associazione su geni candidati, sono risultati associati con il LES in almeno uno studio, ma per i quali non esistono ancora evidenze certe di associazione. Le altre regioni cromosomiche per le quali è stato riscontrato un linkage significativo da più di un gruppo di ricerca sono i loci 2q37 (dove mappa *PDCD1*), 6p21-p12 (contenente la regione HLA), 4p16, 11p13, 12q24 e 16q12-13 (tabella I5) (dati riassunti da Harley et al. 2006, Tsao 2003, Tsao & Wong 2006, Kelly et al. 2002). Altre regioni cromosomiche sono risultate in linkage con una particolare caratteristica clinica del LES, come la nefrite (2q34-35) o la vitiligine (17p13), con la predisposizione a sviluppare anche altre malattie associate al LES, quali la tiroidite autoimmune (5q14) o l'anemia emolitica (11q14), con la presenza di determinate classi auto-anticorpali (19p13.2 e 3p21) o anche con l'insorgenza del LES nei fumatori (3p21).

Tab. I5 Regioni confermate di Linkage per la suscettibilità al LES

LINKAGE					ASSOCIAZIONE	
STUDIO ORIGINALE			CONFERMA			
pos. citogenetica	**LOD score (ref.)**	**popolazione**	**LOD o p (ref)**	**popolazione**	**gene (ref)**	**confermato (ref)**
1q23	4,0 (1)	Afro-Americana, Europea	2,2 (2)	Europea	FCGR (12)	Si (14)
1q31-32	3,8 (2)	Europea, Ispanica	2,4 (10)	Europea	RCA?	No
1q41-42	3,3 (3)	Europea, Afro-Americana, est-Asiatica	p=0,004 (11) 1,3 (12)	Americana	PARP TLR5	No
2q37	4,2 (4)	Europea, Ispanica	4,5 (13) p=0,01 (11)	Europea, Americana	PDCD1 (13)	Si (15)
4p16	3,8 (5)	Europea	1,5 (5)	Americana		
6p21-p12	4,2 (6)	Europea	p=0,007 (11)	Americana	HLA-DR	Si (molti)
11p13	3,4 (7)	Afro-Americana	2,3 (7)	Americana		
12q24	3,3 (8)	Ispanica	2,8 (8)	Americana		
16q12-13	3,4 (9)	Europeo-Americana	3,0 (8, 14)	Americana		

(Harley et al. 2006)
1: Kelly et al. 2002a, 2: Johanneson et al. 2001, 3: Tsao et al. 1999, 4: Lindqvist et al. 2000, 5: Gray-McGuire et al. 2000, 6: Gaffney et al. 1998, 7: Kelly et al. 2002b, 8: Nath et al. 2004, 9: Gaffney et al. 2000, 10: Johanneson et al. 2004, 11: Harley et al. 2006, 12: Moser et al. 1998, 13: Prokunina et al.2002, 14: Cowland et al. 1994, 15: Prokunina et al. 2004.

Studi di associazione e GWAS. Al momento in letteratura sono presenti 8 GWAS nelle popolazioni caucasoidi (Cervino et al. 2007, SLEGEN 2008, Hom et al. 2008, Kozyrev et al. 2008, Graham et al. 2008, Gateva et al. 2009) e asiatiche (Han et al. 2009, Yang et al. 2010). Inoltre sono stati

prodotti numerosissimi studi di associazione, e studi di fine mapping su geni o regioni candidate. Tuttavia molti di questi studi sono stati condotti su casistiche troppo piccole, o i risultati si sono rivelati non riproducibili. Al momento mettendo insieme i dati ottenuti da GWAS e studi di associazione su grosse casistiche si può ricavare una lista di almeno 30 geni di suscettibilità confermati (tabella R16, nella sezione risultati).

La patogenesi del LES è generata dall'interazione di più processi e vie molecolari differenti. La maggior parte dei geni coinvolti nella suscettibilità al LES identificati sino ad ora è implicata in tre processi biologici principali: a) processazione degli immuno complessi, b) funzione dei recettori toll-like e produzione di interferone di tipo I, c) trasduzione del segnale nei linfociti.

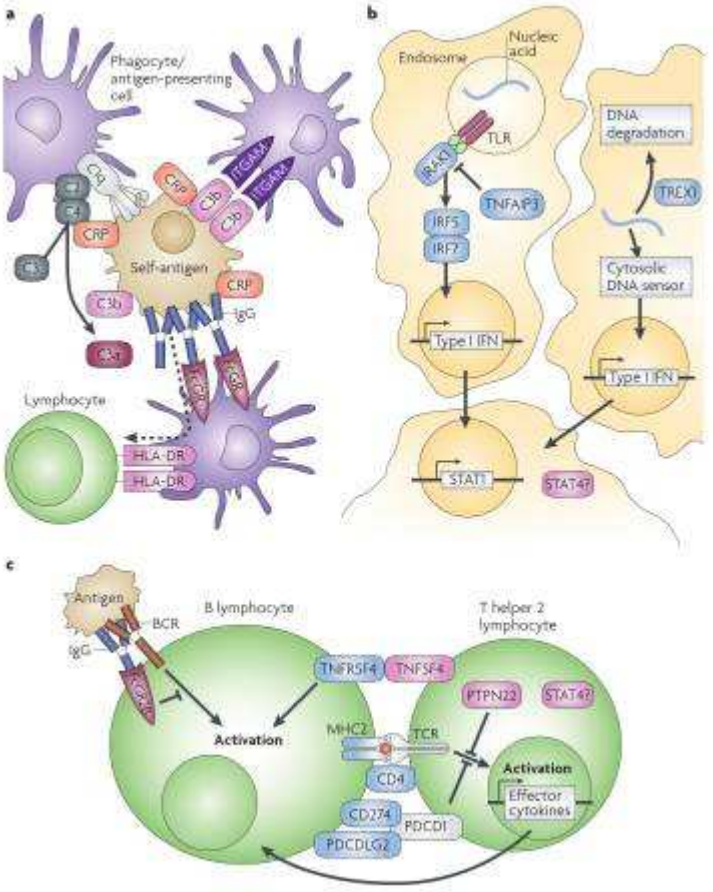
Pathway a: fagocitosi e processazione degli immunocomplessi. Fattori d'innesco di natura ambientale, come per esempio l'esposizione a raggi UV o a infezioni virali, e difetti nella clearance dei corpi apoptotici, causano l'attivazione delle cellule presentanti l'antigene (APC). Le APC fagocitano gli antigeni self rivestiti di opsonine (come C3b, C4b, Fc) che vengono riconosciute dai loro specifici recettori (per esempio ITGAM, FcGR, CRP). Questo porta all'attivazione delle APC e conseguentemente alla presentazione di antigeni self ai linfociti (figura I2a). Varianti alleliche a carico di alcuni loci per i quali è stata identificata un'associazione con il LES (HLA-DR, proteina C reattiva, recettori per Fc) potrebbero influenzare la reattività di queste proteine con gli immunocomplessi. Questa ipotesi è rafforzata dai bassi livelli sierici di fattori del complemento nella forma attiva della malattia. Questo meccanismo gioca potenzialmente un ruolo sia nell'innesco sia nella progressione della malattia. Infatti rappresenta un primo passo nella perdita della tolleranza al self e causa la produzione di autoanticorpi in caso di malattia già conclamata.

Pathway b: produzione di interferone di tipo I (IFNI). Il ruolo dell'interferone nel LES è stato suggerito già nel 1970, e confermato in studi più recenti. La produzione di interferone di tipo I è indotta tramite i recettori Toll-like TLR7 e TLR9. Alcuni dati suggeriscono che TREX1 digerisca il DNA presente nel citosol prevenendo così l'attivazione della via di segnalazione dell'interferon-α (figura I2b). . Parecchi geni di suscettibilità al LES codificano per proteine facenti parte della via di segnalazione indotta da IFNI o di quella che porta alla produzione di IFNI. Tra questi ricordiamo IRAK1, TREX1, IRF5, TNFAIP3. L'iperproduzione di IFNI può indurre l'espressione di citochine e chemochine pro infiammatorie, la maturazione delle cellule dendritiche, l'attivazione di linfociti B e T autoreattivi, la produzione di autoanticorpi e la perdita di tolleranza al self.

Pathway c: trasduzione del segnale nei linfociti. Il riconoscimento di antigeni self ad opera dei linfociti B avviene tramite i B-cell receptors (BCR), che formano legami crociati in risposta al legame con l'antigene, generando così un segnale di attivazione che viene trasdotto da chinasi

intracellulari come BLK e BANK1. L'attivazione e la conseguente interazione dei linfociti B autoreattivi con i linfociti Th2 porta alla perdita della tolleranza al self e di conseguenza all'autoimmunità (figura I2c).

Figura I2 pathways in cui operano i geni noti di suscettibilità al LES (Harley et al. 2009).



Rappresentazione schematica dei principali pathways molecolari a cui partecipano i geni per i quali è stata provato il coinvolgimento nella suscettibilità al LES: a): fagocitosi e processazione degli immunocomplessi; b) produzione di interferone di tipo I, c) trasduzione del segnale nei linfociti.

VARIANTI RARE

Complemento Il sistema del complemento comprende oltre 30 proteine, alcune delle quali codificate nella regione HLA, altre sul cromosoma 1q32. Ha numerose importanti funzioni immunitarie, che comprendono: la citolisi di organismi estranei, l'opsonizzazione di particelle non-self per renderle più attaccabili dai macrofagi, l'attivazione del processo infiammatorio mediata da frammenti proteolitici di componenti complementari (anafilotossine). Inoltre contribuisce alla solubilizzazione degli immunocomplessi favorendone l'eliminazione. Per queste ragioni è stato ritenuto un importante gene candidato nel conferire suscettibilità al LES. Deficit nel sistema del complemento sono stati associati a svariate malattie autoimmuni.

La totale mancanza o la completa perdita di funzione delle componenti C1q, C2, C4A e C4B della via classica del complemento è associata con un rischio elevato di contrarre la malattia.

Per quanto riguarda il C1q, il 90% degli individui che portano una sua mancanza allo stato omozigote (deficit completo) o che portano una mutazione omozigote che lo rende inattivo (deficit funzionale completo) sviluppano il LES, e il rischio di ammalarsi è uguale per uomini e donne. Trattamenti con il C1q portano al ripristino dell'attivazione della via classica del complemento e dell'attività emolitica (Trendelenburg et al. 1999; Kirschfink et al. 1993).

Al contrario, gli individui che presentano un deficit di C1q allo stato eterozigote non sembrano sviluppare un fenotipo patologico (Kirschfink et al. 1993). Topi knock out per il gene C1q sviluppano un fenotipo somigliante a quello dei malati di LES umani. In generale, rispetto ai topi normali, essi presentano un più alto titolo anticorpale, aumentata mortalità e prevalenza di glomerulonefriti, ed un aumento in sede glomerulare di corpi cellulari apoptotici, di immunocomplessi e di C3, che causa danni irreversibili alla funzionalità renale (Carrol et al. 1998; Botto et al. 1996). Una base molecolare per il deficit del gene *C1q* è stata osservata per la prima volta nel 2001 da Dragon-Durey et al.. e consisteva in una mutazione puntiforme introducente uno codone di stop (CGA per arginina ® TGA per lo stop codon).

In più di uno studio si sono osservate percentuali di pazienti C4-difettivi molto alte, variabili di popolazione in popolazione. Il gene C4 è molto polimorfico: nell'intera popolazione sono stati osservati più di 40 alleli di *C4A* e *C4B*, includenti anche alleli "null". Solo il 60% della popolazione ha quattro geni *C4* funzionali, e la parziale deficienza del gene *C4* è probabilmente il deficit immunitario umano più frequente.

In contrasto con l'alta prevalenza degli eterozigoti, la percentuale di omozigoti "null" nella popolazione sana è molto bassa, mentre aumenta negli individui affetti da LES. Dragon-Durey et al. (2001b) hanno riscontrato una percentuale del 79,2% di pazienti con anomalie a livello dei geni C4 (delezione, mancata espressione, conversione genica e duplicazione), con una prevalenza del 52,8% di casi di deficienza completa del gene. Diversi autori hanno riportato che il 75-96% degli individui totalmente difettivi per i geni *C4* sono affetti da Lupus Eritematoso Sistemico. Tali percentuali non sembrano influenzate dall'etnia.

Le basi molecolari di tale difetto sono molto eterogenee, ma l'allele C4Anull più frequentemente osservato in popolazioni caucasiche è una delezione di 28 kbasi che rimuove i geni *C4A* e *21-OHA* ed è associato con l'aplotipo *HLA B8-DR3*. Altre delezioni, duplicazioni e casi di mancata espressione del gene sono abbastanza comuni. Mutazioni frame-shift sono state osservate più raramente. Dragon-Durey et al. (2001) hanno tentato di definire un ruolo specifico per le mutazioni null dei geni *C4A* nel determinare suscettibilità alla malattia, ma l'eterogeneità delle mutazioni

riscontrate li ha portati ad escludere questa ipotesi (Linqvist et al. 1999; Dragon- Durey et al. 2001; Rupert et al. 2002; Fan et al. 1993). Seguardo et al. (1992) hanno osservato una paziente che presentava un deficit combinato delle componenti *C4B* e *C7*.

Difetti nel gene C2 del complemento sono associati con una frequenza più bassa di casi di LES e con una sintomatologia più mite. Rispetto ai geni *C4*, il gene *C2* è caratterizzato da una percentuale di eterozigosità inferiore: il 96% dei cromosomi umani porta la forma wilde-type. Circa 1/10.000 individui di popolazione caucasica è omozigote per il deficit totale del gene. Circa la metà dei soggetti che presentano una deficienza del gene *C2* sono asintomatici, ma tale deficit è stato comunque associato a diverse malattie, la più frequente delle quali è il LES. Il 5,9% dei pazienti affetti da LES è risultata eterozigote per la deficienza del gene *C2* (Colten 1992).

Un'associazione più debole con il LES è stata descritta anche per altri elementi della cascata del complemento: la componente C3 ed i fattori I e H e raramente i fattori C5, C6, C7 (Morgan et al. 1991).

Nonostante le mutazioni null allo stato omozigote siano un importante fattore di rischio per il LES, esse sono comunque una causa piuttosto rara di sviluppo della malattia, poichè sono poco frequenti nella popolazione (Slingsby et al. 1996).

TREX1 TREX1 (Dnase III, Three Prime Repair EXonuclease) è una 3'-5' esonucleasi che ha come bersaglio preferenziale gli ssDNA. Il gene *TREX1* mappa sul cromosoma 3p21 ed è costituito da 16 esoni, di cui solo uno (esone 16) è codificante. Le DNA nucleasi hanno la funzione di mantenere la stabilità del genoma e di controllare i processi di replicazione, di riparo e di ricombinazione del DNA. Si è ipotizzato, di conseguenza, un ruolo di TREX1 nella replicazione e nel riparo del DNA. Tuttavia tale funzione non è stata confermata, in quanto nei topi knockout (Trex1 -/-) non si osserva né un incremento di mutazioni spontanee né un aumentato rischio nello sviluppo del cancro. Questi topi manifestano miocarditi infiammatorie simil-autoimmuni e ridotta sopravvivenza. Il fenotipo di infiammazione cronica che si manifesta nei topi Trex1 -/-, risulta in una alterazione nello sviluppo dei linfociti T con la conseguenza perdita nelle cellule T della tolleranza al self (Morita et al., 2004). Questi fatti hanno suggerito un possibile ruolo di *TREX1* nello sviluppo delle autoimmunità, in seguito confermato dall'osservazione di mutazioni a carico di *TREX1* in alcune malattie autoimmuni umane (LES e SS) (Lee-Kirsh et al., 2007). TREX1 normalmente è presente principalmente in associazione col reticolo endoplasmatico, come parte di un complesso proteico che viene attivato in seguito a stress ossidativi (SET complex). TREX1 viene traslocato nel nucleo in seguito a apoptosi o stress genotossici. Durante l'apoptosi mediata da PRF1 il Granzima A proteolizza i componenti del SET complex, tra cui TREX1, causandone l'attivazione e la traslocazione in sede nucleare. La risposta apoptotica mediata dal Granzima A provoca la

formazione di ssDNA (generati da un'endonucleasi anch'essa facente parte del SET complex), i quali diventeranno il substrato di TREX1. Inoltre TREX1 previene l'accumulo di cDNA derivante da retroelementi endogeni, che potrebbero venire riconosciuti come acidi nucleici virali, provocando così una produzione incontrollata di IFNs (interferoni). In modelli murini è stato verificato che le cellule con deficit di TREX1 accumulano ssDNA nel citosol, causando l'attivazione del pathway di IRF-3 (Vijay et al 2008, Stetson et al 2008, Yang et al 2007).

Sono state osservate mutazioni in eterozigosi di *TREX1* in 9/417 pazienti con LES e 1/169 con SS (Lee-Kirsh, 2007). Mutazioni bialleliche di *TREX1* causano la Sindrome di Aicardi-Goutieres (AGS), una malattia autosomica recessiva che si manifesta principalmente in età pediatrica e che coinvolge il sistema neurologico e cognitivo (Rice et al. 2007). Inoltre mutazioni in eterozigosi di *TREX1* provocano due rare patologie autosomiche dominanti: il Lupus Chilblain Familiare (FCL), una rara forma familiare cutanea di lupus (Lee-Kirsh et al. 2006 e Rice et al. 2007) e la Retinal Vasculopathy and Cerebral Leukodistrophy (RVLC) (Richards et al 2007), una condizione recentemente descritta in tre famiglie, caratterizzata dalla presenza di disfunzioni cerebrali e a carico della retina (Ophoff et al. 2001).

PROCESSAZIONE DEGLI IMMUNO COMPLESSI

CRP (C reactive protein) Il gene *CRP* mappa sul cromosoma 1q23.2, in una regione cromosomica di suscettibilità a l LES, ed è costituito da soli due esoni. CRP è una proteina della classe delle pentraxine, coinvolta nella fagocitosi dei detriti derivati dall'apoptosi e degli immuno complessi. E' espressa in modo ubiquitario ed è filogeneticamente conservata. È' una delle proteine plasmatiche che vengono definite proteine di fase acuta, poiché vanno incontro ad un ingente aumento della loro concentrazione in caso di danno tessutale o di infiammazione. Durante la fase attiva della malattia, nonostante la marcata infiammazione tissutale, i livelli di CRP sono anormalmente bassi a causa della ridotta sintesi. L'introne di *CRP* contiene un microsatellite (GT)n che è stato correlato con i livelli basali di CRP. Russell et al. (2004) osservano una debole associazione tra questo dinucleotidee la suscettibilità alla malattia, e tra un polimorfismo del 3' UTR ed il LES e la produzione di autoanticorpi anti-nucleo. Più recentemente è stata descritta un'associazione tra un polimorfismo nel promotore di questo gene ed il LES (Edberg 2008).

Recettori per il Fc delle Ig I recettori per la regione costante delle immunoglobuline sono espressi in un'ampia varietà di cellule ed hanno un ruolo cruciale nelle risposte umorali e cellulari del sistema immunitario. Ognuno delle cinque classi di immunoglobuline ha un suo specifico gruppo di recettori. Per quanto riguarda i recettori per le IgG (FcγR): essi possono essere divisi geneticamente e strutturalmente in due gruppi distinti :

1) Cellule endoteliali, del sinciziotrofoblasto placentare e cellule epiteliali dell'intestino del neonato esprimono un recettore (FcRB) costituito da una catena MHC class I like (codificata sul cromosoma 19) complessata ad una catena di tipo β2-microglobulinico. I recettori FcRB sono coinvolti nel trasporto delle IgG materne al feto ed hanno un ruolo critico nel proteggere le IgG dal catabolismo.

2) I FcγR espressi dai leucociti appartengono alla superfamiglia delle immunoglobuline e possono a loro volta essere divisi in tre classi principali: FcγRI (CD64), FcγRII (CD32), FcγRIII (CD16). I loro geni si trovano sul braccio lungo del cromosoma 1 (1q21-24). Le tre classi differiscono per la loro distribuzione in differenti tipi cellulari, per la forza con cui legano le immunoglobuline e per la loro capacità di legare specificamente diverse sottoclassi immunoglobuliniche. I FcγRI sono espressi soprattutto sui neutrofili attivati, e legano le IgG monomeriche con alta affinità, mentre i FcγRII e i FcγRIII sono recettori a bassa affinità, che interagiscono con IgG complessate o aggregate. Ognuna di queste famiglie è composta da complessi di geni distinti e da varianti generate da splicing alternativo (Van der Pol et al. 1998; Salmon et al.1996 e Botto et al. 1996).

La struttura base dei recettori consiste in una catena α formata da una componente extracellulare contenente 2-3 domini immunoglobulinici, un dominio trans membrana ed una coda citoplasmatica. Perchè avvenga la trasduzione del segnale è necessario che il recettore in membrana sia associato a proteine segnalatorie (catene γ,βo ξ) (Van der Pol et al. 1998). Gli effetti funzionali che possono essere regolati dai recettori per il frammento cristallizzabile delle IgG sono molteplici e includono la fagocitosi, la citotossicità cellulare anticorpo dipendente (ADCC), la presentazione dell'antigene, il rilascio di citochine, la degranulazione e la regolazione della sintesi degli anticorpi (Cambier 1995; Van der Pol et al. 1998).

I FcγR presentano numerose varianti alleliche, che differiscono nella capacità di iniziare una specifica risposta biologica. L'attitudine a svolgere una determinata funzione varia anche a seconda della classe di recettore.

-L'FcγRIIa, il recettore più comune, espresso su neutrofili, monociti, macrofagi e piastrine. E' particolarmente efficiente nel trasmettere segnali di inizio della fagocitosi e può indurre la risposta citotossica. Si conoscono due forme alleliche codominanti di questo recettore, che differiscono per un solo amminoacido (arginina/istidina, codone 131) nel secondo dominio immunoglobulinico extracellulare, denominate FcγRIIa-R131 e FcγRIIa-H131. Questo polimorfismo (rs1801274) ha un valore funzionale importante, in quanto l'FcγRIIa-H131 è la sola forma che riconosce efficientemente il dominio Fc delle IgG2 nell'uomo, e il legame con l'immunoglobulina è ottimizzato solo se il gene è presente nello stato omozigote (Salmon et al. 1996; Botto et al. 1998; Yee et al. 1997; Manager et al. 1998; Song et al. 1998 e Duits et al. 1995). L'omozigosi H/H del

recettore ha quindi un ruolo essenziale per il legame con le IgG degli immunocomplessi: di conseguenza, nel contesto della fagocitosi mediata dall'opsonizzazione dell'antigene, le due forme alleliche determinano una diversa attività di macrofagi e neutrofili nella rimozione degli immunocomplessi (Anderson et al. 1990; Indik et al. 1991). Dal momento che la clearence degli immunocomplessi ha un ruolo importante per la patogenesi del LES, soprattutto per la sottocategoria di pazienti con glomerulonefrite, in diversi studi è stata valutata l'eventuale associazione l'allele R131 e la malattia. I risultati degli studi sono però controversi: infatti è stato osservato un aumento significativo dell'allele meno efficiente (R131) in tre pannelli di malati di origine africana, caucasica ed orientale (Salmon et al.1996; Song et al. 1998 e Duits et al. 1995) ma il dato non è stato confermato da altri lavori (Botto et al. 1998; Manger et al. 1998; Smyth et al. 1997). E' stata eseguita una meta-analisi che coinvolge 17 studi (uno dei quali è stato eseguito nel nostro laboratorio) riguardanti il polimorfismo R 131 H. In totale sono stati genotipizzati 1405 pazienti con nefrite lupica, 1709 pazienti senza nefrite e 2580 controlli sani. Non è stata trovata associazione tra il genotipo RR ed il rischio di sviluppare una nefrite lupica. Il genotipo RR è comunque risultato significativamente più frequente nei pazienti, ma con un valore di Odd Ratio piuttosto basso. (OR =1,30) (Karassa et al. 2002).

-L' FcγRIIIa è espresso sulla membrana cellulare delle cellule NK, dei monoliti e dei macrofagi. Lega le IGG1 e le IgG3. Un polimorfismo fenilalanina/valina all'aminoacido 176 influenza la capacità di legame con le IgG. Infatti gli omozigoti F176 legano le Ig con meno efficienza degli omozigoti V176, e quindi sono meno efficienti nell'effettuare la clearence degli immunocomplessi. Tuttavia i risultati degli studi di associazione condotti su questo polimorfismo sono stati discordanti. Una meta-analisi che comprende più di mille soggetti per ogni categoria analizzata (pazienti con coinvolgimento renale, pazienti senza coinvolgimento renale e controlli) ha concluso che l'allele F158 conferisce un rischio 1,2 volte maggiore di contrarre la nefrite ludica, ma non è stata osservata associazione con il LES senza coinvolgimento renale (Karassa et al. 2003).

-I recettori FcγRIIb sono espressi sui linfociti B, sui macrofagi, sui monoliti, sulle cellule detritiche. Contengono un dominio inibitorio che regola la proliferazione e la sopravvivenza dei linfociti B. Un polimorfismo I232T che altera la funzione inibitoria è stato riscontrato in associazione con il LES nella popolazione asiatica, ma non in altre popolazioni (Li et al. 2003).

-*FCGR3B* e *FCGR3C*: la porzione cromosomica che incorpora questi due geni è soggetta a variazioni nel numero di copie (il numero diploide varia da nessuna ad un massimo di sei copie). E' stato dimostrato che un basso numero di copie sembra costituisce un fattore predisponente per la glomerulo nefrite, sia nei pazienti affetti da LES che in modelli murini (Aitman et al. 2006). Successivamente uno studio su una casistica più ampia ha dimostrato che gli individui con meno di

due copie presentano un rischio maggiorato sia di sviluppare la nefrite lupica (OR=2,43) che di contrarre il LES senza coinvolgimento renale (OR=2,21) (Fanciulli 2007).

MBL (Mannose-binding lectin) Nella struttura e nelle funzioni MBL è molto simile a C1q. MBL lega il mannosio sulla superficie microbica ed inizia la via di attivazione del complemento mediata dalla lectina. Polimorfismi nella regione del promotore o codificanti causano un'alterata espressione di MBL (ridotta o assente), e potrebbero essere associati con la suscettibilità al LES (Nath et al. 2004). La prevalenza di questo deficit nella popolazione generale è stimata tra il 5% ed il 10% (Turner 1991).

ITGAM -ITGAX (integrin, alpha M - integrin, alpha X). I geni *ITGAM* e *ITGAX* codificano per 2 proteine della classe delle integrine. L'integrina–αM (ITGAM), in combinazione con l'integrina-β2 forma un recettore di membrana di membrana per il complemento (CR3). CR3 è espresso prevalentemente da neutrofili, macrofagi e cellule dendritiche ed è importante nell'aderenza di queste cellule all'epitelio stimolato e nella fagocitosi dei corpi opsonizzati dal complemento. Il legame della componete complementare attivata C3b con l'integrina–αM sulle cellule presentanti l'antigene determina la produzione diTGFβ2 e IL10, un processo essenziale per l'induzione della tolleranza immunitaria.

Conseguentemente all'associazione osservata in due GWAS, è stato effettuato uno studio di fine-mapping sul gene *ITGAM* in una popolazione mista di origine Europeae Africana (Nath et al. 2008). La variante più fortemente candidata è una sostituzione arginina/istidina alla posizione 77 all'interno di un dominio funzionale.

FUNZIONE DEI RECETTORI TOLL-LIKE E PRODUZIONE DI INTERFERONE DI TIPO I

TLR5 (Toll-like receptor 5) Il gene *TRL5* mappa in posizione 1q41-42 ed è espresso nei leucociti, nelle ovaie e nella prostata (Chaudhary et al., 1998). L'ortologo murino di TLR5 mappa nella regione di suscettibilità al LES *Sle1d* (Hawn et al. 2005).

Le proteine TLR costituiscono una sottofamiglia di proteine transmrmbranarie della superfamiglia dei recettori per le interleuchine. Riconoscono gli agenti patogeni ed iniziano la via di traduzione dei segnali infiammatori. Sono espressi su molti tipi cellulari, incluse le cellule presentanti l'antigene, le cellule epiteliali ed endoteliali ed i leucociti (Lin et al. 2007). Dati recenti indicano un ruolo dei TLR nel rigetto dei trapianti (riassunti da Lin et al. 2007). In seguito a stimolazione le TLR reclutano la chinasi associata al recettore per IL-1 (IRAK1) ed innescano una via di segnalazione che porta alla produzione di citochine (Akira et al. 2003) Le proteine della famiglia TLR contengono i caratteristici motivi Toll: un dominio extracellulare ricco in leucina (LRR) coinvolto nel riconoscimento dei patogeni (Bell et al. 2003) ed una regione citoplasmatica simile al

recettore dell'IL-1 (Lin et al. 2007). Il ligando di TRL5 è la flagellina, il principale componente del flagello batterico e noto fattore virulento riconosciuto dal sistema immune nelle piante, negli insetti e nei mammiferi (Bell et al. 2003). L'attivazione di TLR5 innesca la produzione di IL-6, che a sua volta stimola la differenziazione e la proliferazione dei linfociti B e la produzione di anticorpi (Dean et al. 2000).

Un polimorfismo che introduce un codone di stop a livello dell'aminoacido 392 (un'arginina) ha dimostrato un debole effetto protettivo nei confronti del LES (Hawn et al. 2005). La variante TLR5392STTOP non è funzionale ed è associata con una marcata diminuzione nella produzione di citochine proinfiammatorie, in particolare IL-6 (Hawn et al. 2003)), TNF-α e IL-1β(Hawn et al. 2005)

MCP-1 (monocyte chemoactractant protein) MCP-1 codifica per una chemochina che recluta i monociti, gli eosinifili ed i linfociti T della memoria ai siti di infiammazione. Regola l'espressione delle molecole dell'adesione e la funzione dei linfociti T nell'infiammazione acuta e cronica. Nei pazienti affetti da LES è coinvolto nel manifestarsi di glomerulonefrite.

L'aumento dei livelli sierici di MCP-1 correla con l'attività di malattia (Aguilar et al. 2001). Sono stati condotti studi sull'assoociazione tra LES ed i polimorfismi del promotore di MCP-1, ma i dati sono discordanti (Aguilar et al. 2001, Rovin et al. 1999).

Interferone di tipo I (IFN-γ) L'IFN-γ é un importante mediatore dell'immunità virale. I pazienti affetti da LES presentano livelli aumentati di IFN-α e IFN-γ ed é stato osservato che i livelli ematicidi espressione di geni IFN-α e IFN-γ inducibili correlano con l'attività di malattia del LES (Bennet et al. 2003). I topi lpr hanno livelli elevati di IFN-γ (Fan et al. 1997) e topi knockout IFN-γ presentano una regressione della malattia (Haas et al. 1998). Inoltre la somministrazione terapeutica di IFN-γ può causare l'instaurarsi di sintomi simili a quelli del LES sia in modelli umani (Machold & Smolen 1990) che murini (Adam et al. 1980). In uno studio di associazione condotto su una piccola casistica è stato osservata solo una modesta associazione con un polimorfismo (Val14Met) del gene per IFN-γ IFNGR1) ed il LES, non particolarmente convincente (Nakashima et al. 1999). Associazioni ben più significative sono state osservate con i geni che codificano per proteine della via dell'interferone.

TYK2 TYK2 é una tirosina chinasi che trasduce segnali mediati dall'interferone di tipo I. E' associato con la formazione di anticorpi anti –DNA in pazienti caucasici affetti da LES (Namijou et al. 2002). TYK2 e Jak1 si legano rispettivamente ai domini IFNAR1 e IFNAR2 del recettore per IFN-α/β. IFN-α interagisce col recettore coniugato a TYK2 e Jak1 e porta all'iperespressione di IFN-α (Wong & Tsao 2006). Varianti missenso di TYK2 potrebbero ridurne la funzionalità, ed avere quindi un effetto protettivo verso il LES (Sigurdsson et al. 2005).

IRF5 (Fattore di trascrizone dell'interferone, 5) I fattori regolatori dell'interferone (IRFs) costituiscono una famiglia di regolatori trascrizionali che possiedonoun dominio legante il DNA con una struttura elica-giroelica (Taniguchi et al. 2001).

IRF5 è un fattore trascrizionale responsabile della risposta immunitaria innata durante le infezioni virali. E' espresso nei linfociti B e nelle cellule dendritiche. Attiva i geni delle citochine infiammatorie, tra cui TNF-α, IL-12, IL-6, IFN e i geni responsivi a IFN (Takaoka et al. 2005).

E' stata osservata una forte associazione tra polimorfismi di IRF5 ed il LES in popolazioni scandinave, successivamente confermata in altre popolazioni (Graham et al. 2006, Graham et al 2007, Ferreiro-Neira et al. 2007). Il modello genetico più probabile propone un'associazione con un aplotipo contenente variazioni funzionali, alcune delle quali sono state caratterizzate funzionalmente. Il polimorfismo rs2004640 crea un nuovo sito di splicing per un esone alternativo (Graham et al. 2006, Sigurdsson et al. 2005), mentre lo SNP rs10954213 localizzato nella regione 3' non tradotta crea un nuovo sito di poliadenilazione, e porta di conseguenza alla sintesi di un trascritto più corto e più stabile (Cunninghame et al. 2007).

IRF3 (Fattore di trascrizone dell'interferone, 3) L'induzione della trascrizione dei geni IFNA è dovuta principalmente a fattori trascrizionali IRF3 e IRF7. IRF3 è costitutivamente espresso ad alti livelli nella maggior parte delle cellule. Durante l'infezione virale IRF3 viene traslocato nel nucleo ed innesca l'espressione di IFN-β, IRF9 e IRF7 (Akahoshi et al. 2006).

Nella popolazione giapponese è stata osservata un'associazione significativamente protettiva tra il LES e due polimorfismi del promotore di *IRF3*. Questi polimorfismi sono correlati con la ridotta attività trascrizionale del gene, bassa espressione costitutiva di IRF3 e di conseguenza con la bassa produzione di IFN-γ (Akahoshi et al. 2006).

IRF1 (Fattore di trascrizone dell'interferone, 1) Il gene *IRF1* mappa in 5q31.1 e codifica per un fattore di trascrizione che regola la produzione di IFN-β e di altri geni sensibili all'IFN (Harada et al.1998).Lo splicing alternativo del gene *IRF1* produce una proteina non funzionale che è presente ad alti livelli in pazienti con sindrome mielodisplastica e leucemia mielogena acuta (Harada et al. 1994).La sostituzione di una valina con una metionina al codone 114 rende il recettore meno responsivo all'IFN. E' stata riscontrata un'associazione tra questo polimorfismo ed il LES in almeno una popolazione (Tanaka et al. 1999).

IRAK1 (Interleukin-1 receptor associated kinase 1). E' una serina/treonina chinasi coinvolta nella cascata di trasduzione del segnale dei recettori Toll/IL-1 (TIR). Questa famiglia di recettori include i recettori Toll-like (tra cui TLR5) e la sottofamiglia dei recettori per IL-1, che riconoscono le citochine endogene proinfiammatorie IL-1 ed IL-18. Ha un ruolo nell'induzione di IFN-α e IFN-γ. E' stata recentemente descritta un'associazione nella popolazione nord-americana tra il LES ed

alcuni polimorfismi di *IRAK1*, compresi in un unico blocco di Linkage Disequilibrium (LD) che contiene gli esoni 11-13. Questa porzione del gene codifica per un dominio proteico responsabile dell'interazione con proteine traduttrici del segnale come TRAF6.

TRASDUZIONE DEL SEGNALE NEI LINFOCITI

CTLA4 (CD152) Il gene *CTLA-4* (2q33) presenta due isoforme di splicing, la prima codifica per un recettore di superficie omologo strutturale di CD28, mentre la seconda per una proteina solubile priva del dominio transmembranario che viene espressa e secreta dai linfociti T non attivati (Magistrelli et al. 1999). E' un regolatore negativo dell'attivazione dei linfociti T ed ha un ruolo importante nella prevenzione delle malattie autoimmuni in quanto promuove l'anergia. E' espressa ad elevati livelli in seguito ad attivazione dei linfociti T. Esercita un ruolo determinante nel controllo dell'attivazione dei linfociti T inibendo la proliferazione cellulare e la produzione di citochine. Topi deficitari per CTLA-4 manifestano disordini linfoproliferativi, processi autoimmunitari e livelli elevati di immunoglobuline sieriche che conducono gli animali a morte in 4-5 settimane (Waterhouse et al. 1995, Tivol et al. 1995). CTLA-4 contribuisce allo spegnimento della risposta immunitaria attraverso due meccanismi: 1) la competizione con CD28 per il legame con le molecole attivanti B7-1 e B7-2 e 2) segnali di trasduzione inibitori, forse attraverso l'attivazione delle serina/treonina fosfatasi (Krishnan et al. 2006).

E' stata osservata un'associazione significativa tra i polimorfismi di CTLA-4 e più di una malattia autoimmune (riassunte da Yamada et al. 2005). Le associazioni più convincenti si osservano con il diabete mellito di tipo 1 (IDDM) e con la tiroidite autoimmune e I suoi sottotipi (tiroidite di Hashimoto e morbo di Graves). Il gene *CTLA-4* è associato positivamente all'IDDM nella popolazione italiana ed in altre popolazioni caucasoidi. L'associazione è stata dimostrata sia con il microsatellite $(AT)_n$ nel 3'UTR ($P=0.002$) sia col polimorfismo A/G (Ala17Thr) nel primo esone ($P=0.00002$). Quest'ultima variazione è responsabile anche della suscettibilità al morbo di Graves ($P<0.0002$). La variazione Ala17Thr è coinvolta anche nella suscettibilità genetica alla celiachia (Djilali-Saiah et al. 1998), alla tiroidite di Hashimoto e alla malattia di Addison. Al momento in letteratura sono riportati 14 studi di associazione indipendenti tra i polimorfismi di CTLA-4 e la suscettibilità al LES, con risultati spesso contrastanti. Lee et al. (2005) hanno effettuato una meta-analisi tra questi studi prendendo in esame i polimorfismi Ala17Thr e $(AT)_n$ già descritti ed alcuni polimorfismi del promotore, ed osservano un'associazione statisticamente significativa solo con la variazione A/G Ala17Thr. I linfociti T attivati di individui omozigoti G per questo polimorfismo presentano una ridotta attivazione di CTLA-4, ed un'aumentata risposta proliferativa (Maurer et al. 2002).

PDCD1 Il gene *PDCD1* mappa sul cromosoma 2q37 (Shinohara et al. 1994) e codifica per la proteina PD-1, un recettore di membrana membro della superfamiglia di B7/CD28. PD-1 agisce come inibitore dei linfociti T attraverso l'interazione con i suoi ligandi PDL1 e PDL2 (Kroner et al. 2005). Il legame con PDL1 o PDL2 porta alla fosforilazione di PD-1 ed al legame con la tirosina fosfatasi SHP-2. Conseguentemente si ha la defosforilazione di altre molecole segnalatorie e l'attivazione della via di segnale di Ras, che a seconda della forza e della durata del segnale può causare sia l'arresto del ciclo cellulare in fasa G1 che una risposta proliferativa (Prokunina & Alarcon-Riquelme 2004). Normalmente é espresso sulla membrana dei linfociti B e T attivati e regola il meccanismo della tolleranza immunitaria in entrambi i tipi cellulari (Nishimura & Honjo 2001). Topi BALB/c knockout per *PDCD1* sviluppano una cardiomiopatia autoimmune fatale con deposito di IgG sulla superficie dei cardiomiociti (Nishimura et al. 2001). In un altro modello murino il deficit di PDCD1 causa un fenotipo simile al LES con glomerulonefrite ed artrite (Nishimura et al. 1999).

Un polimorfismo nel quarto introne (chiamato PD1.3) é risultato significativamente associato con il LES in più di una popolazione (Prokunina et al. 2002). Inoltre la frequenza dell'allele A di PD1.3 é significativamente più elevata nei pazienti con nefrite (Prokunina et al. 2004). Questo polimorfismo cade in un sito di legame per il fattore trascrizionale RUNX1, e potrebbe modificare l'espressione di PDCD1.

PTPN22 (Protein Tyrosine Phosphatase Non-receptor 22) Il gene *PTPN22* (Protein Tyrosine Phosphatase Non-receptor 22) si trova sul cromosoma 1p13.2 e codifica per la proteina Lyp che svolge un importante ruolo regolatorio nella via di attivazione del recettore delle cellule T. Lyp è una proteina tirosina fosfatasi intracellulare di peso molecolare di 110Kb, è costituita da un dominio N-terminale catalitico e un C-terminale non catalitico con quattro domini ricchi di prolina. Lyp si lega mediante il primo dominio prolinico (P1) al dominio SH3 della proteina Csk. Il complesso Lyp-Csk inibisce l'attivazione del recettore delle cellule T.

L'SNP R620W influenza il legame di Lyp a Csk, causando un aumento dell'attività fosfatasica con una conseguente ulteriore riduzione della via di segnalazione del TCR, che predispone alle malattie autoimmuni (Orozco et al. 2005). E' stata osservata un'associazione significativa di R620W con più di una malattia autoimmune: LES, artrite reumatoide, diabete di tipo 1, morbo di Graves, artrite giovanile idiopatica (riassunte da Lee et al. 2007). E' stato proposto che l'inibizione terapeutica di Lyp possa aiutare nella prevenzione e nella cura delle malattie autoimmuni. Più recentemente è stata riportata un'associazione significativa tra il LES ed una secondo polimorfismo missense di Lyp (R263Q), una variante con perdita di funzione che rende Lyp meno efficiente nell'inibire il segnale di TCR (Orrù et al. 2009). Gli autori riportano un effetto protettivo di questo

polimorfismo nella suscettibilità al LES, ma ritengono che i pazienti portatori di questa variante potrebbero non beneficiare dell'inibizione terapeutica di Lyp, per cui la conferma di questo dato potrebbe avere importanti implicazioni nella terapia personalizzata.

STAT4 (Signal transducer and activator of transcription 4). STAT 4 è un fattore trascrizionale che media l'espressione di molti geni critici nella risposta immunitaria, in particolare IRF5, IFN-1, IL-12 e IL-23. Tramite queste citochine gioca un ruolo importante nell'induzione della risposta Th1 mediata. La sua attivazione porta alla secrezione acuta di IFN-γ nei linfociti T CD4+ e nelle cellule NK. Sono state descritte associazioni statisticamente significative con polimorfismi situati nella regione genomica contenente *STAT4* sia nell'Artrite Reumatoide che nel LES (Remmers 2007), successivamente confermate dai risultati dei GWAS.

BANK1 (B-cell scaffold protein with ankyrin repeats 1). BANK1 è una proteina specifica dei linfociti B che regola la traduzione di alcune vie di segnale. La sua fosforilazione in seguito all'attivazione dei linfociti B porta all'attivazione dei recettori dell'inosotolo trifosfato e alla mobilitazione dei depositi di calcio intracellulari. Il gene viene espresso in due diverse isoforme di splicing che differiscono per la presenza/assenza dell'esone 2. E' stato osservato un polimorfismo di *BANK1* che influenza il rapporto con il quale le due diverse isoforme vengono prodotte. Questa variante, ed altri due polimorfismi che conducono a variazioni aminoacidiche, sono risultati associati al LES in 5 diverse popolazioni, tra cui quella italiana.

BLK (B lymphoid tyrosine kinase) e LYN (v-yes-1 Yamaguchi sarcoma viral related oncogene homolog). Sono entrambe tirosine chinasi non recettoriali che regolano l'attivazione dei linfociti B in seguito a segnali mediate dai recettori BCR. Il loro ruolo nella determinazione degli eventi cellulari non è ancora stato completamente determinato. Tra le due proteine, LYN è quella che è stata meglio caratterizzata: media l'attivazione dei linfociti B attrverso le Igα/β e la tirosina chinasi SYK, e trasduce segnali inibitori tramite la fosforilazione di CD22 e FcγRIIb. I topi knockout per *Lyn* sviluppano una patologia autoimmune simile al lupus (Hibbs et al.1995). L'associazione con questi geni è fortemente emersa dai recenti GWAS.

ALTRI GENI

SELP (selectin P (granule membrane protein 140kDa, antigen CD62)). La proteina selectina viene accumulata nei granuli alfa delle piastrine e nei corpi di Weibel-Palade delle cellule endoteliali. Viene ridistribuita alla membrana cellulare durante l'attivazione e la degranulazione piastrinica e media l'interazione tra le cellule endoteliali attivate ed i leucociti.

NMNAT2 (Nicotinamide nucleotide adenylyltransferase 2). Appartiene alla famiglia delle nicotinamide mononucleotide adenosiltransferasi, che catalizzano un passaggio essenziale nella

biosintesi del NAD-NADP. A differenza degli altri membri della famiglia, che sono enzimi nucleari espressi ubiquitariamente, NMNAT2 è una proteina citosolica, espressa prevalentemente nel tessuto cerebrale.

Callicreine. Le callicreine costituiscono una famiglia multigenica di serina esterase con un ampio spettro di funzioni biologiche, includenti la regolazione dell'infiammazione, dell'apoptosi, del bilancio redox, e dello sviluppo di fibrosi renale. In un'analisi di trascrittomica eseguita sulla cortex renale di modelli murini sperimentali di glomerulonefriti è emerso che le callicreine costituiscono una frazione significativa delle proteine che evidenziano livelli di espressione diversi rispetto ai modelli di controllo.

UBE2L3 (ubiquitin-conjugating enzyme E2L 3). La modificazione delle proteine tramite le ubiquitine è un importante meccanismo cellulare che ha lo scopo di indirizzare le proteine senescenti o anomale verso la degradazione. L'ubiquitina UBE2L3 partecipa in vitro all'ubiquitinazione di p53, c-Fos, e del precursore di NF-kB.

SCUBE1 (signal peptide, CUB domain, EGF-like 1). SCUBE-1 è una glicoproteina di secrezione, che è stata osservata in molti tessuti altamente vascolarizzati come fegato, reni, polmoni, milza e nel tessuto cerebrale. Presenta omologie strutturali con alcune proteine anti-coagulanti come la trombomodulina e la proteina C, e con le fibrilline. Contiene 10 domini simili a quello del fattore di crescita epiteliale (EGF-like). La sua espressione è inibita in vitro dalle citochine proinfiammatorie IL-1β e TNFα. Di conseguenza SCUBE1 potrebbe giocare un ruolo importante nella risposta infiammatoria e in quella trombotica.

CITOCHINE

TNFSF4 (tumor necrosis factor (ligand) superfamily, member 4). Noto anche come OX40L (OX40 Ligand), è una proteina di membrana espressa sulla superficie delle cellule presentanti l'antigene. Il legame con OX40 sulla superficie dei linfociti T CD4+ innesca un forte segnale attivatorio, che media la proliferazione e l'espansione delle cellule T durante la risposta immunitaria primaria, l'adesione dei linfociti T attivati con le cellule vascolari endoteliali e la generazione di linfociti memoria (Gramaglia et al. 2000). Graham et al. (2008) hanno descritto un aplotipo di rischio di *TNFSF4*, associato con un'aumentata espressione di OX40L, sia a livello di trascritto, che di presenza in membrana della proteina.

Cluster interleuchine sul cromosoma 1 Questo cluster comprende le interkeuchine: IL*10, IL19, IL20* e *IL24*. Originariamente era stata descritta un'associazione con una variante nel promotore di *IL10* (D'Alfonso et al 2000, D'Alfonso et al. 2002) che non è però stata confermata in tutti gli studi. Recentemente uno studio genome-wide ha identificato un'associazione significativa con un

polimorfismo nella regione genomica di *IL10*, non in Linkage Disequilibrium con la prima variante associata (Hawn et al. 2005) .

IL18 (Interleukin 18). L'interleuchina 18 è una importante citochina pro-infiammatoria appartenente alla famiglia dell'interleuchina 1. Può stimolare la risposta immunitaria Th1 e Th2 mediata, ed in combinazione con IL-12 induce la produzione di IFN-γ nei linfociti Th1 e B e nelle cellule Natural Killer. Inoltre può accelerare lo sviluppo del fenotipo autoimmune nei topi MRL/lpr. I livelli sierici di IL18 sono più alti nei pazienti con LES rispetto ai controlli, e correlano positivamente con l'indice di attività della malattia. Recentemente è stato osservato che un polimorfismo funzionale nel promotore di *IL18* ne influenza la trascrizione ed è associato con la suscettibilità al LES nella popolazione spagnola.

HLA

Il primo rapporto di un'associazione tra geni *HLA* ed il LES risale al 1971, ed era relativo a geni HLA di classe I (allele HLA-B8). Studi successivi hanno mostrato un'associazione più forte con gli alleli *HLA-DR2* e *DR3* (HLA di classe II). Da ulteriori analisi è risultato che l'associazione di *HLA-B8* con la malattia in realtà riflette semplicemente un linkage disequilibrium con gli alleli *HLA-DR3* e DQw2. Invece l'associazione con i geni di classe II è stata riconfermata in più di uno studio, nei quali si osserva un aumento significativo della frequenza degli alleli DR2 e DR3 nei pazienti appartenenti alla popolazione caucasica provenienti dall'Europa occidentale. Negli ultimi anni gli studi di associazione Genome-Wide hanno dimostrato l'importanza di questa regione nella suscettibilità al LES. Tuttavia la complessità e l'elevato LD che caratterizzano la regione HLA rendono difficile identificare la variante primariamente associata con la malattia. Barcellos et al. (2009) hanno analizzato 1974 SNP situati nel complessi HLA in 1610 pazienti e nei loro genitori. L'associazione più forte è stata riscontrata con l'allele *HLA-DRB1*0301* (OR=2,21; $p=2,53X10^{-12}$).

MODELLI MURINI

Esistono molti modelli murini pertinenti per lo studio del LES. Nel 1959 M. Bielschowsky della University of Otago Medical School, in Nuova Zelanda, creò un primo modello murino di malattia autoimmune, e lo chiamarono New Zealand Black (NZB). Questo topo sviluppa spontaneamente un'anemia emolitica autoimmune associata a anticorpi anti-eritrociti, splenomegalia, trombocitopenia e reticolocitosi. Essi producono frequentemente anticorpi IgG contro il ssDNA, ma non contro il dsDNA o contro gli istoni, e solo raramente sviluppano nefriti nel primo anno di vita.

Poiché la stirpe murina NZB è stata ottenuta tramite accoppiamento tra consanguinei ed il fenotipo patologico è trasmesso di generazione in generazione, la malattia autoimmune di questo ceppo è apparentemente sotto controllo genetico.

Un altro ricercatore dello steso istituto (W.H. Hall) ha sviluppato indipendentemente da Bielschowsky un secondo modello murino, chiamato New Zealand White per il colore del mantello, ed il suo background genetico è sconosciuto. Questi topi sono per lo più fenotipicamente normali, e la produzione di IgG anti-dsDNA è rara. Gli studi su questi topi hanno subito un ulteriore sviluppo in seguito alla scoperta che i topi appartenenti alla prima generazione derivante da un incrocio NZBxNZW sviluppano un pattern clinico ed immunologico molto simile al LES umano. Nelle femmine (NZBxNZW)F1 sono infatti frequentemente riscontrabili deposizioni di immunocomplessi che portano alla compromissione della funzionalità renale, con conseguente sviluppo in età precoce di nefriti. I sintomi dell'anemia emolitica sono invece notevolmente ridotti. Altre complicazioni frequenti coinvolgono la cute, i polmoni, i vasi sanguigni e il SNC. Inoltre i topi sviluppano anticorpi contro un gran numero di componenti nucleari, incluso il DNA a singola e doppia elica e gli istoni. Il modello (NZBxNZW)F1 è sicuramente quello più studiato, ed anche quello che più si avvicina ai modelli umani. Dal momento che né i topi NZB, né i NZW sviluppano il LES, la predisposizione alla malattia deve essere determinata dal corredo genico di entrambi i genitori. Questi modelli sono estremamente utili per identificare i geni di suscettibilità per ogni singola manifestazione del Lupus, e per determinare le varie interazioni epistatiche tra questi geni. Questo ceppo è stato utilizzato in molti studi per identificare i meccanismi genetici che conducono ad un'ereditarietà poligenica e complessa dei diversi fenotipi del LES (Shirai et al. 2002).

Lo studio del LES è stato ulteriormente facilitato dall'avvento di ulteriori modelli murini, come i ceppi MRL-*lpr/lpr* e BXSB. I ceppi MRL- *lpr/lpr* sono omozigoti per la mutazione *lpr*, consistente in un'inserzione trasposonica nel gene *Fas*, coinvolto nei meccanismi di apoptosi cellulare. I topi omozigoti per questa mutazione sviluppano una patologia linfoproliferativa, e manifestazioni autoimmunitarie simili a quelle del LES. Una malattia simile si sviluppa anche in topi (*gld/gld*) omozigoti per una mutazione che coinvolge il gene del ligando del *Fas*. Topi MRL privi della mutazione *lpr* (MRL+/+)presentano anch'essi predisposizione ad un fenotipo lupus-like, ma con esordio più tardivo rispetto al ceppo MRL*lpr/lpr* (Mountz et al. 1992; Vyse et al. 1998; Shirai et al. 2002). Sono state descritte anche nell'uomo mutazioni a carico del gene *Fas* in bambini affetti da una malattia simile al fenotipo murino *lpr/lpr*, denominata "sindrome linfoproliferativa autoimmune" (Rieux-Laucat et al. 1995; Fisher et al. 1995, Shirai et al. 2002).

I topi BXSB portano un gene mutante legato al cromosoma Y, *Yaa*, la cui funzione non è nota, ma che sembra implicato con lo sviluppo precoce ed accelerato della malattia (Izui et al. 1995). Se

confrontati con topi congenici che portano gli alleli wild-type, i ceppi BXSB e MRL-*lpr/lpr* mostrano una sintomatologia molto più seria, indice che i geni coinvolti sono modificatori epistatici positivi. Nei topi maschi appartenenti alla prima generazione di un'ibridazione tra il ceppo NZW e il ceppo BXSB si sviluppa spontaneamente un disordine simile alla sindrome anti-fosfolipidica associata al LES.

SCOPO DEL LAVORO

Lo scopo della ricerca intrapresa durante il mio periodo di dottorato è stato quello di ricercare fattori genetici di suscettibilità al LES. A tal fine abbiamo seguito diverse linee di ricerca:

a)*Studi di associazione e/o funzionali su geni candidati*.
Il nostro laboratorio ha partecipato a progetti internazionali, in collaborazione con altri gruppi di ricerca, volti a chiarire, sia da un punto di vista genetico che funzionale il ruolo nella suscettibilità genetica al LES di due geni, *BANK1* e *IL18*. Il primo è stato scelto sulla base dei dati di un precedente GWAS, mentre il secondo è stato analizzato perché funzionalmente candidato e perché mappa in una regione cromosomica di suscettibilità al LES, individuata da precedenti studi di linkare su tutto il genoma. Sempre in questo ambito abbiamo partecipato ad un progetto, in collaborazione con un gruppo americano, che si è posto lo scopo di analizzare i geni del cluster delle callicreine, sia in pazienti affetti da LES che in modelli murini. Tutti e tre i progetti hanno prodotto lavori pubblicati su riviste internazionali.

b)*Analisi di fattori genetici di suscettibilità localizzati nelle regione cromosomica 1q32-43.*
Questo ramo del progetto origina da uno studio da me condotto negli anni precedenti, avente il fine di ricercare geni di suscettibilità al LES nella regione cromosomica candidata 1q32-1q43 tramite un fine-mapping. Durante il primo anno di dottorato ho approfondito l'analisi su uno dei geni mappanti in questa regione e che aveva precedentemente evidenziato un'associazione con il LES (*CR1L*). Il fine che ci eravamo proposti era di confermare l'associazione osservata, sia analizzando gli stessi polimorfismi su altre casistiche indipendenti, sia analizzando altri marcatori nelle regioni centromeriche e telomeriche rispetto a *CR1L*.

c)*Studio di tipo mutazionale*
Abbiamo condotto uno studio di tipo mutazionale su due geni funzionalmente candidati, *TREX1* e *PRF1*, allo scopo di cercare varianti genetiche rare di suscettibilità alle malattie autoimmuni. Questa parte del progetto è stata anche su pazienti affetti da altre due malattie autoimmuni clinicamente correlate con il LES, la *Sclerosi Sistemica* (SSc) e la *Sindrome di Sjogren* (SS).

d)*Creazione di un algoritmo genetico di suscettibilità al LES.*
Durante il III anno di dottorato ho disegnato e iniziato un nuovo progetto, ancora in corso d'opera. Lo studio si propone di analizzare nella popolazione italiana una serie di marcatori che sono già

risultati associati al LES in precedenti studi genome-wide, condotti su popolazioni prevalentemente di origine nord-europea. In tal modo ci siamo proposti molteplici scopi:

1) Confermare le associazioni descritte in letteratura nella popolazione Italiana.

2) Valutare l'interazione e l'effetto combinato di queste varianti genetiche nella suscettibilità al LES.

3)Stimare il valore predittivo della combinazione di queste varianti genetiche.

4) Integrare le informazioni genetiche con informazioni di carattere clinico ed immunologico.

A tal scopo abbiamo selezionato una lista di marcatori da includere nell'algoritmo, e abbiamo già raccolto dati preliminari per una parte di essi. I dati sono stati utilizzati per creare un primo algoritmo, includendo anche caratteristiche clinico-demografiche.

e)*Meta-analisi sui geni TEC e BMPR1B*

Due dei geni inclusi nella tabella R15 (*TEC* e *BMPR1B*) sono stati analizzati su una casistica più ampia (1268 casi e 1597 controlli), non solo sui campioni italiani al fine di:

1) Eseguire un'analisi d'imputazione utilizzando software bioinformatici per l'analisi dei GWAS allo scopo di ottenere un dato di genotipizzazione per polimorfismi situati nella stessa regione genomica ma non tipizzati fisicamente.

2) Unire, tramite meta analisi, i dati ottenuti dallo studio di replicazione con quelli derivanti da due precedenti GWAS.

3) Svolgere sui dati così ottenuti un'analisi di associazione di tipo caso-controllo.

RISULTATI

Studi di associazione e/o funzionali su geni candidati

ARTICOLO 1

Functional variants in the B-cell gene BANK1 are associated with systemic lupus erythematosus.

Kozyrev SV, Abelson AK, Wojcik J, Zaghlool A, Linga Reddy MV, Sanchez E, Gunnarsson I, Svenungsson E, Sturfelt G, Jönsen A, Truedsson L, Pons-Estel BA, Witte T, D'Alfonso S, Barizzone N, Danieli MG, Gutierrez C, Suarez A, Junker P, Laustrup H, González-Escribano MF, Martin J, Abderrahim H, Alarcón-Riquelme ME.

Nat Genet. 2008 Feb;40(2):211-6. Epub 2008 Jan 20

Questo lavoro deriva da una collaborazione internazionale con più centri di ricerca europei. Lo scopo previsto era quello di replicare un'associazione osservata in uno studio di tipo GWAS con un polimorfismo funzionale (rs10516487, causante la sostituzione aminoacidica R61H) del gene *BANK1* (B-cell scaffold protein with ankyrin repeats). Il gruppo di ricerca svedese ha effettuato un fine mapping della regione genomica contenente *BANK1* (circa 284-kb): 30 SNPs sono stati tipizzati in 279 pazienti affetti da LES di origine svedese e 515 controlli concordi per provenienza geografica. Nove delle varianti testate hanno evidenziato frequenze alleliche diverse in modo statisticamente significativo (p>0,5) tra pazienti e controlli.

Successivamente lo stesso gruppo ha condotto un'analisi dettagliata dell'espressione e della struttura di questa proteina. E' stato osservato che BANK1 è prevalentemente espressa nei linfociti B CD19+, come tre isoforme di splicing: due isoforme "full-lenght" (FL) alternative originate dallo splicing alternativo degli esoni 1A e 1B e una isoforma mai identificata precedentemente con la delezione dell'esone 2 (isoforma Δ2). La delezione di questo esone mantiene inalterata la cornice di lettura. Questa isoforma era presente in tutti i campioni esaminati (83 controlli sani e 30 individui con il LES), oltre ad essere espressa nella milza di topo e scimpanzé, per cui è stata ritenuta un'isoforma espressa costitutivamente. Tuttavia l'analisi quantitativa ha dimostrato che il rapporto tra l'isoforma FL e l'isoforma Δ2 non è costante nei diversi individui, e che tale rapporto è influenzato da un polimorfismo (rs17266594) localizzato nel sito di ramificazione dell'esone 2. I soggetti omozigoti per l'allele T (cioè i portatori della struttura classica del sito di ramificazione YNY*TG***A**YYN) esprimono l'isoforma FL in misura sensibilmente maggiore. L'espressione totale di *BANK1* non è al contrario influenzata da tale variante allelica.

La regione contenente il promotore, gli esoni 1A, 1B e 2 e 500 basi delle porzioni introniche fiancheggianti ciascuno di questi esoni sono state sequenziate in 24 pazienti e 8 controlli, ma non sono state osservate nuove varianti con un possibile ruolo funzionale. Inoltre nella stessa casistica

sono state analizzate 5 varianti non sinonime presenti sulle banche dati, una sola delle quali (A383T – rs3733197) risultata presente nella coorte presa in esame.

Le tre variazioni funzionali così identificate (rs10516487, rs17266594 e rs3733197) sono quindi state genotipizzate in altri quattro pannelli di pazienti e controlli, provenienti da: Argentina, Germania, Spagna e Italia. Questi ultimi provengono interamente dal nostro laboratorio. Tutti e tre i polimorfismi hanno mostrato un'associazione significativa con il LES (nella casistica combinata: $p=3,74 \times 10^{-10}$, $p=4,74 \times 10^{-11}$, $p=4,67 \times 10^{-5}$ rispettivamente). Tutte e tre le varianti erano significativamente associate al LES anche nella sottocasistica italiana (rispettivamente: p=0,0078; p=0,0016; p=0,0097).

ARTICOLO 2

Kallikrein genes are associated with lupus and glomerular basement membrane-specific antibody-induced nephritis in mice and humans.

Liu K, Li QZ, Delgado-Vega AM, Abelson AK, Sánchez E, Kelly JA, Li L, Liu Y, Zhou J, Yan M, Ye Q, Liu S, Xie C, Zhou XJ, Chung SA, Pons-Estel B, Witte T, de Ramón E, Bae SC, Barizzone N, Sebastiani GD, Merrill JT, Gregersen PK, Gilkeson GG, Kimberly RP, Vyse TJ, Kim I, D'Alfonso S, Martin J, Harley JB, Criswell LA; Profile Study Group; Italian Collaborative Group; German Collaborative Group; Spanish Collaborative Group; Argentinian Collaborative Group; SLEGEN Consortium, Wakeland EK, Alarcón-Riquelme ME, Mohan C.

J Clin Invest. 2009 Apr;119(4):911-23.

Questo studio è stato eseguito in collaborazione con numerosi centri di ricerca Europei, Americani ed internazionali. La nefrite immuno-mediata è una caratteristica clinica sia del LES che della sindrome di Goodpasture, che è una patologia causata da anticorpi specifici per la membrana basale dei glomeruli (anticorpi anti-GMB). Inoltre gli anticorpi anti-DNA e anti-glomerulo sono stati implicati nella patogenesi della nefrite lupica, sia in modelli umani che murini. Un modello sperimentale utile per la comprensione dei meccanismi che innescano la nefrite immuno-mediata è la glomerunonefrite indotta sperimentalmente con anticorpi anti-GMB (AIGN). Tuttavia non tutti i ceppi murini sono ugualmente sensibili alla stimolazione con anticorpi. Precedenti dati dei gruppi con cui abbiamo collaborato hanno dimostrato che su 20 ceppi murini stimolati con anticorpi anti-GMB solo 5 hanno sviluppato una patologia renale severa. Per capire se la differente reattività alla stimolazione con anticorpi fosse dovuta a fattori genetici e molecolari specifici per i reni lo stesso gruppo aveva condotto un'analisi trascrittomica basata su microarray sulla corteccia renale di 3 ceppi murini AIGN-sensibili (DBA/1, NZW e 129/SvJ) e su 2 ceppi resistenti (B6 e BALB/C), osservando un pannello di 50 geni espressi in misura significativamente inferiore nelle tre linee sensibili.

Dieci di questi geni appartenevano alla famiglia delle callicreine (*Klk*). Le callicreine sono parte di una famiglia multigenica di serina-esterasi con un ampio spettro di funzioni biologiche, includenti la regolazione all'interno dei reni di meccanismi quali infiammazione, apoptosi, bilancio redox e fibrosi. L'intero cluster dei geni delle callicreine nel topo è codificato all'interno di un locus murino di suscettibilità al LES ($Sle3^{z}$) situato sul cromosoma 7.

L'espressione genica delle callicreine a livello renale prima e dopo l'induzione dell'AIGN è stata valutata tramite real-time PCR, confermando quanto già osservato nell'analisi con microarrays. Infatti i livelli di espressione delle callicreine nei reni dei topi appartenenti alle linee AIGN-resistenti aumentano in seguito all'esposizione agli anticorpi anti-GMB, mentre questo non accade

nelle linee AIGN-sensibili. I livelli basali, precedenti l'esposizione con anticorpi, di callicreine, erano al contrario simili in tutti e cinque i ceppi.

Il gruppo con cui collaboriamo ha quidi sequenziato i cinque geni che hanno evidenziato le differenze di espressione maggiori (*Klk1, Klk1b3, Klk1b5, Klk1b26, Klk1b27*) nei cinque ceppi murini esaminati. Sono state osservate variazioni di sequenza ceppo-specifiche nel promotore di tutti e cinque i geni. In particolare le due linee murine AIGN-resistenti (B6 e BALB/C) si sono rivelate pressoché identiche tra loro, così come le linee DBA1 e NZW. La terza linea AIGN-sensibile (129/SvJ) era molto più simile agli altri due ceppi sensibili. Alcuni delle varianti di sequenza ceppo-specifiche cadono in siti di legame per fattori di trascrizione putativi. Non sono state osservate varianti di sequenza nelle regioni codificanti o 3' non tradotte di nessuno dei 5 geni.

Nell'uomo il cluster delle callicreine è codificato in una regione, situata sul cromosoma 19q13, che è già stata implicata nella suscettibilità al LES in precedenti studi di linkage genome-wide. Per verificare se questa classe proteica abbia un ruolo nella suscettibilità alla malattia anche nell'uomo 8 SNP situati nei geni per le callicreine mappanti nellaregione 19q13 ed espressi a livello renale (*KLK1, KLK5, KLK6* e *KLK7*) sono stati genotipizzati in 340 pazienti affetti da LES e 400 controlli appartenenti alla popolazione tedesca. Due polimorfismi di *KLK1* (rs1054713 e rs274052) hanno mostrato una debole associazione nei pazienti con nefrite (rispettivamente p=0,051 e 0,010). Inoltre l'associazione più forte (p=0,007) è stata riscontrata per lo SNP rs274052 confrontando i pazienti LES con nefrite con quelli non affetti da nefrite.

Per confermare questo risultato gli otto polimorfismi sono stati tipizzati in altre 6 coorti di pazienti e controlli di varia origine geografica (Germania, USA, Argentina, Spagna, Italia, Corea) per un totale di 1268 casi e 5378 controlli. L'associazione con il polimorfismo rs274052 di *KLK1* è stata confermata nella popolazione italiana (p=0,008) ed in quella tedesca (p=0,010), ma non negli altri gruppi. Non è stato possibile eseguire una meta-analisi a causa dell'eterogeneità genetica tra le diverse popolazioni. I pazienti italiani provengono interamente dal nostro laboratorio.

Il risultato ottenuto è stato ulteriormente validato utilizzando dati di genotipizzazione per gli SNP dell'intera regione cromosomica contenente il cluster delle callicreine, ottenuti in precedenti studi di tipo GWAS in 689 pazienti LES e 3718 controlli appartenenti all'Intenational Consortium for Systemic Lupus Erythematosus Genetics (SLEGEN) e 595 pazienti LES provenienti dal UCSF Lupus Genetics Project. I due studi hanno utilizzato due set di SNP solo parzialmente sovrapponibili, per cui per l'analisi sono stati utilizzati 56 SNP genotipizzati in entrambi gli studi. L'associazione più forte è stata osservata per polimorfismi situati in prossimità del gene *KLK1*: nella regione intergenica tra *KLK1* e *KLK15* e nella regione intergenica tra *KLK1* e *KLK3*. Inoltre sono stati definiti i blocchi di Linkage Disequilibrium nella regione e gli aplotipi risultanti sono stati

testati per l’associazione con la suscettibilità al LES. L’associazione più forte è stata riscontrata a carico di un aplotipo formato da SNP situati nella regione contenente il promotore del gene *KLK3*.

ARTICOLO 3

Identification of a new putative functional IL18 gene variant through an association study in systemic lupus erythematosus.

Sánchez E, Palomino-Morales RJ, Ortego-Centeno N, Jiménez-Alonso J, González-Gay MA, López-Nevot MA, Sánchez-Román J, de Ramón E, González-Escribano MF, Pons-Estel BA, D'Alfonso S, Sebastiani GD; Italian collaborative group, Alarcón-Riquelme ME, Martín J.

Hum Mol Genet. 2009 Oct 1;18(19):3739-48. Epub 2009 Jul 7.

Questo lavoro è stato condotto in collaborazione con un gruppo di ricerca spagnolo, e si è posto come obbiettivo di investigare il ruolo di un gene candidato (*IL18*) nella suscettibilità al LES.

Il gene per l'interleuchina 18 (*IL18)* mappa sul cromosoma 11q22.2-22.3, in una regione di suscettibilità al LES identificata in studi di linkage condotti su popolazioni europee. Inoltre sono state riportate associazioni tra numerosi polimorfismi della regione regolatoria 5' fiancheggiante di *IL18* e diverse malattie autoimmuni ed infiammatorie (asma, diabete di tipo 1, artrite giovanile idiopatica, artrite reumatoide, morbo di Chron). L'IL18 è una citochina pleiotropica che può mediare la risposta immunitaria sia di tipo Th1 che di tipo Th2. Inoltre nei topi MRL/lpr accelera la patologia autoimmune spontanea simile al lupus. Elevati livelli sierici di IL18 sono stati osservati nei pazienti affetti da LES, e tali livelli correlano con l'attività della malattia.

Per chiarire se il gene *IL18* abbia un ruolo nella suscettibilità alla malattia è stata eseguita un'analisi genetica di genotipizzazione per 9 polimorfismi comuni in 752 pazienti e 595 controlli sani di origine spagnola. Questi polimorfismi sono tagSNP che permettono una copertura di tutti gli aplotipi con frequenza allelica >0,01 (r2>0,8). Inoltre si è ricorsi all'imputazione dei dati mancanti per analizzare anche le varianti non genotipizzate. In tal modo sono state ottenute le tipizzazioni relative ad altri 19 polimorfismi con maf (minor allele frequency) >5%. Il marcatore che ha evidenziato l'associazione statisticamente più significativa (p=7,8E-07) è lo SNP rs360719 nel promotore di *IL18*, in posizione -1297. In particolare l'allele C mostra un aumento significativo di frequenza nei pazienti (OR=1,53).

L'associazione con il polimorfismo sopra citato è stata replicata in due popolazioni indipendenti: quella italiana (348 casi e 364 controlli) e quella argentina (275 casi e 245 controlli). I pazienti ed i controlli italiani sono stati raccolti dal nostro centro, ed il DNA è stato estratto nel nostro laboratorio. In entrambe le popolazioni il genotipo rs360719 CC è leggermente più frequente nei pazienti LES rispetto ai controlli, ma a livello di frequenza allelica tale differenza non raggiunge la significatività statistica. I dati relativi alle tre popolazioni sono stati uniti in una meta-analisi

comprendente 2579 individui, che ha confermato l'associazione con rs360719 (OR=1,37 p=3,8E-07).

Per confermare il ruolo di rs360719 nella patogenesi del LES il gruppo spagnolo ha valutato se vi fossero differenze quantitative genotipo-specifiche nell'espressione dell'mRNA di *IL18* nei leucociti mononucleati periferici di 23 individui con diversa tipizzazione, e ha osservato un aumento di espressione nei soggetti portatori dell'allele C. Il polimorfismo rs360719 cade in un sito di legame per il fattore di trascrizione OCT-1. L'analisi in silico ha predetto che OCT-1 lega l'allele T, ma non l'allele C di rischio. Questo dato è stato confermato anche in vitro tramite un saggio EMSA (electrophoretic mobility shift assay).

Questi risultati suggeriscono che la variante *IL18* rs360719 abbia un ruolo importante nella suscettibilità al LES e nell'espressione di IL18.

Analisi di fattori genetici di suscettibilità localizzati nella regione cromosomica 1q32-43

DATI NON ANCORA PUBBLICATI

La regione cromosomica 1q32-q43 ha in passato evidenziato prove che hanno indotto a ritenere che possa contenere uno o più geni coinvolti nella patogenesi del LES. Ciò è emerso sia da precedenti studi di linkage condotti da diversi gruppi su popolazioni diverse (Johanneson et al. 2001; Graham et al. 2001; Tsao et al. 1999; Johanneson et al. 2004), sia in base a convergenti evidenze su modelli murini (Shirai et al. 2002). Per questa ragione in passato nel nostro laboratorio è stato condotto uno studio di associazione sistematico su pool di DNA in questa regione, nel corso del quale erano stati analizzati tutti i geni in essa compresi, indipendentemente dalla loro rilevanza funzionale. A questo scopo erano stati utilizzati 219 marcatori di tipo htSNPs (haplotype tag SNPs), selezionati tra quelli genotipizzati dal progetto HapMap, situati in 74 geni. L'associazione più significativa era stata osservata a carico dell'elemento genico *CR1L* (*CR1-like*): ben tre marcatori intronici (rs2796239, rs6667500, rs2761424) avevano evidenziato una differenza significativa tra le frequenze alleliche dei pazienti e quelle dei controlli (rispettivamente p = 0,0001531, p =0,0019393, p =0,000364). Questo gene era stato quindi analizzato in modo molto approfondito. Una rappresentazione schematica della regione genomica contenente *CR1L* si può osservare nella figura R1.

L'intera sequenza codificante delle tre predette isoforme di *CR1L*, le regioni 5'UTR e 3'UTR, i confini esone/introne e 870 basi nella sequenza 5' flanking del gene erano quindi state analizzate con DHPLC e sequenziamento per la ricerca di polimorfismi in 31 pazienti italiani affetti da LES ed erano state così riscontrate 24 variazioni di sequenza. Tutte le varianti osservate, oltre ai tag SNPs di *CR1L* selezionati dalle banche dati e precedentemente analizzati su pool, per un totale di 25 marcatori, erano stati tipizzati su di un primo pannello di 180-240 pazienti LES e 185-234 controlli sani appartenenti alla popolazione italiana.

Cinque marcatori (3'UTR +41 T/C, 3'UTR +137 T/C, rs2796239, rs6667500 e rs2761424) avevano evidenziato un'associazione con il LES in questa casistica. Per tutti e 5 i polimorfismi l'allele più raro era sovrarappresentato nella popolazione sana, il che farebbe pensare ad un effetto protettivo: 3'UTR +41 (OR=0,63; p=0,00318), 3'UTR +137 (OR=0,72; p=0,01777), rs2761424 (OR=0,62; p=0,00213), rs6667500 (OR=0,65; p=0,00155), rs2796239 (OR=0,66; p=0,00568). Inoltre, il dato osservato su un polimorfismo relativamente raro nella regione 3'UTR di *CR1L* (3'UTR+163 G/A), che si presenta con una frequenza dello 0,01 nei pazienti e dello 0,03 nei controlli, si avvicinava alla significatività statistica. Tre di queste varianti (rs2796239, rs6667500 e rs2761424) sono tag SNPs intronici, mentre le altre due sono SNPs della regione 3'UTR. Tra questi polimorfismi è stato osservato un forte grado di LD.

Per confermare il dato precedentemente osservato i polimorfismi 3'UTR +41 T/C, 3'UTR +137 T/C, 3'UTR+163 G/A, rs2796239, rs6667500 e rs2761424 sono stati tipizzati su una seconda casistica (203 pazienti LES e 203 controlli, per un totale, considerando entrambe le casistiche, di 378 pazienti e 383 controlli appartenenti alla popolazione italiana). Inoltre i polimorfismi 3'UTR +137 T/C, rs2796239, rs6667500 e rs2761424 sono stati analizzati anche su una seconda popolazione indipendente (671 pazienti affetti da LES e 678 controlli sani di nazionalità spagnola). Il polimorfismo 3'UTR +41 T/C è risultato in LD assoluto con lo SNP rs2761424 sia nella popolazione italiana (D'=1; r^2=0,98), sia in un primo campione di 230 LES e 259 controlli sani della popolazione spagnola, per cui non è stato ulteriormente analizzato.

Nella seconda casistica di pazienti e controlli italiani confermiamo l'associazione per i polimorfismi +41 (LES=0,23; controlli=0,30; p=0,0450; OR=0,71) e rs2761424 (LES=0,23; controlli=0,30; p=0,0191; OR=0,68). Se si considera la casistica italiana totale (casistica 1 + casistica 2), l'associazione si conferma per tutti e cinque i polimorfismi. (vedi tab. R1). Il polimorfismo 3'UTR+163 presentava una lieve differenza non significativa nella popolazione originaria, che però scompare del tutto nella seconda casistica. L'associazione più significativa sulla casistica totale è stata osservata per i 2 polimorfismi tra loro in LD quasi assoluto 3'UTR+41 e rs2761424 (p=0,0003; OR=0,65).

Non è stata osservata alcuna differenza di frequenza statisticamente significativa sulla popolazione spagnola (vedi tab. R2). Se si effettua una meta-analisi tra le popolazioni italiana (casistica complessiva) e spagnola, si osserva ancora un'associazione significativa per tre polimorfismi: 3'UTR +41 (p=0,001817; OR=0,76, 95), rs2761424 (p=0,001739; OR=0,80) ed rs6667500; (p=0,006556; OR=0,85).

Le frequenze genotipiche nella popolazione italiana per i 5 polimorfismi significativamente associati con il LES sono riportate nella tabella R3. Per gli SNPs 3'UTR+41, 3'UTR+137, rs2761424, rs6667500, rs2796239 la distribuzione dei genotipi differisce significativamente tra pazienti e controlli. Inoltre dall'analisi delle distribuzioni genotipiche per tutti e 5 i polimorfismi sembra emergere un ruolo protettivo per l'allele più raro che è sempre sovrarappresentato in omozigosi nei controlli. Per i due polimorfismi in LD assoluto 3'UTR+41 e rs2761424 l'effetto è probabilmente dominante, poiché si nota nei controlli un aumento delle frequenze genotipiche contenenti l'allele che si suppone protettivo sia in omozigosi che in eterozigosi. Anche nel caso delle frequenze genotipiche, nella popolazione spagnola non si osserva alcuna associazione significativa (dati non presentati).

A causa del LD i 6 SNPs analizzati sulla casistica più estesa generano solo 5 aplotipi, due dei quali presentano una frequenza inferiore al 5% sia nei pazienti che nei controlli (tabelle R7). La frequenza dei due aplotipi più comuni nella popolazione italiana differisce significativamente tra pazienti e controlli. In particolare l'aplotipo contenente tutti gli alleli singolarmente più frequenti nei controlli sani presenta una frequenza significativamente più elevata nella popolazione di controllo (p=0,00036; OR=0,65). Tuttavia la significatività dell'associazione osservata con l'aplotipo non differisce da quella osservata con i singoli polimorfismi in LD assoluto 3'UTR+41 e rs2761424. L'allele T del polimorfismo rs2761424 è l'unico allele protettivo presente solo nell'aplotipo protettivo. Gli alleli più rari degli altri polimorfismi associati, invece, pur presentando singolarmente un effetto protettivo, compaiono anche in aplotipi "di rischio" (più frequenti nei pazienti). Questo dato farebbe pensare che l'associazione osservata sia dovuta esclusivamente all'effetto del polimorfismo rs2761424, non ad una particolare combinazione aplotipica. Per confermare tale ipotesi è stata valutata la distribuzione delle frequenze alleliche e genotipiche dei polimorfismi rs2796239, 3'UTR+137 e rs6667500 nel sottocampione di individui "T-negativi" per il polimorfismo rs2761424. Non si osserva alcuna differenza statisticamente significativa né a carico delle frequenze alleliche (tabella R4) né a carico di quelle genotipiche (tabella R5). In base a questi dati il polimorfismo rs2761424, in LD assoluto con lo SNP 3'UTR+41, sembrerebbe quello primariamente associato con la suscettibilità al LES. L'ipotesi è stata ulteriormente confermata effettuando un'analisi di regressione. Come si può osservare (tabella R6), l'associazione con i due SNP rs2761424 e 3'UTR+41 rimane significativa dopo la correzione per gli altri polimorfismi (rs2796239, 3'UTR+137, rs6667500), indicando che tale associazione è indipendente dal LD con le altre varianti. Al contrario le associazioni con gli SNP rs2796239, 3'UTR+137 e rs6667500 non sono più significative se corrette per l'effetto di rs2761424 o di 3'UTR+41, il che indica che siano dovute prevalentemente al LD con questi due polimorfismi. A causa dell'elevato LD tra rs2761424 e 3'UTR+41 (r^2=0,98) l'analisi non ha il potere statistico di determinare quale tra le due varianti sia quella primariamente associata.

Non si osserva alcuna differenza significativa nella distribuzione aplotipica tra pazienti e controlli spagnoli (tabella R7B).

Non appare non del tutto convincente che l'associazione osservata sia a carico primariamente di *CR1L*Infatti, nonostante alcuni dati lo suggeriscano (Logar et al. 2002, Irshaid et al. dati non pubblicati), non ci sono prove che *CR1L* sia effettivamente un gene funzionale. Centromericamente rispetto a *CR1L* mappa il gene strutturalmente e sequenzialmente omologo *CR1* (complement component (3b/4b) receptor 1), da cui probabilmente *CR1L* è filogeneticamente derivato in seguito a fenomeni di crossing-over disuguale. In posizione telomerica rispetto a *CR1L* mappa il gene *MCP*

(*CD 46*). Sia *CR1* che *MCP* sembrano funzionalmente correlati con il LES, in quanto la loro espressione correla con l'attività della malattia, tuttavia si ritiene che queste alterazioni siano secondarie allo sviluppo della patologia e non determinate geneticamente.

Il linkage disequilibrium (LD) tra i tre geni è stato analizzato utilizzando i polimorfismi genotipizzati dal progetto HapMap sulle famiglie CEPH situati nella regione genomica (301 kb) compresa tra il 5' del gene *CR1* ed il 3' del gene *MCP* (321 SNPs). Per l'elaborazione statistica e grafica è stato adoperato il software Haploview I blocchi di LD sono stati definiti tramite la definizione di blocco aplotipico di Gabriel (2002). Nella regione analizzata sono presenti 9 blocchi di LD, i primi 5 compresi nel gene *CR1*, 3 nel gene *CR1L*, mentre il nono contiene l'intero gene *MCP*. Tra i geni *CR1* e *CR1L* sembra esserci un hot spot di ricombinazione, e lo stesso accade all'interno del gene *CR1L*. Tra i polimorfismi di *CR1L* ed i marcatori del gene *MCP* (situati in due blocchi diversi), si osserva tuttavia un certo grado di LD (0,91<D'<0,67). I polimorfismi di *CR1L* significativamente associati con la malattia ricadono nel terzo blocco aplotipico di questo gene, e sono pertanto in LD con *MCP*.

Tra i 5 tagSNP di *MCP* analizzati nel corso nello studio di associazione sistematico della regione cromosomica 1q32-1q43 solo uno (rs2724377) raggiungeva la significatività sulla popolazione complessiva, ma non sulle due coorti esaminate separatamente. Questo dato potrebbe quindi riflettere un linkage disequilibrium con la più forte associazione osservata per i polimorfismi di *CR1L*. Al contrario, nessun polimorfismo del gene *CR1* era risultato statisticamente associato al LES e questo dato, unito a quelli di LD predetti sulla base dei genotipi osservati in 30 famiglie CEPH, porterebbe ad escludere che l'associazione riscontrata a carico del gene *CR1L* possa riflettere un LD con il gene *CR1*.

Per escludere che l'associazione osservata rifletta quella più forte con un altro polimorfismo in linkage disequilibrium con l'rs2761424 di *CR1L* abbiamo tipizzato, in 181-240 pazienti italiani affetti da LES ed in 198-215 controlli sani appartenenti alla popolazione italiana, 5 tagSNPs nel gene *CR1* e 9 tagSNPs situati nella regione telomerica rispetto a *CR1L*, 4 dei quali contenuti nel gene *MCP* uno intergenico, gli altri in posizione ancora più telomerica (2 nel gene *CD34*, uno nell'elemento genico *LOC148696* e 1 intergenico. In questa sottocasistica le frequenze alleliche dei polimorfismi associati continuano a presentare una differenza significativa tra pazienti e controlli: rs2796239 (p=0,0201; OR=0,70), rs2761424 (p=0,0107; OR=0,65), ex10+137 (p=0,0262; OR=0,72;), rs6667500 (p=0,0043; OR=0,66) (tabella R8). Pertanto il potere statistico dell'analisi è sufficiente per evidenziare almeno un effetto grande quanto quello osservato a carico di questi SNPs.

Non è stata osservata un'associazione significativa con nessuno dei 13 nuovi tagSNPs analizzati (tabella R8). I polimorfismi sono stati successivamente analizzati uno per uno in combinazione aplotipica con il polimorfismo primariamente associato rs2761424. Da questa analisi si è notata un'associazione a carico di 2 aplotipi, contenenti, oltre al polimorfismo rs2761424, rispettivamente lo SNP rs926631 (p=0,0050; OR=1,48) e lo SNP rs2761437 (p=0,0033; OR=1,59). Gli SNP rs926631 e rs2761437 sono situati rispettivamente in posizione intergenica e nel gene *MCP*. Stranamente però non si osserva un aumento nei controlli dell'aplotipo contenente l'allele T protettivo dell' rs2761424 di *CR1L*, ma l'associazione osservata è in entrambi i casi a carico di un aplotipo che contiene l'allele C del suddetto polimorfismo, che presenta una frequenza più elevata nei pazienti sia in singolo che in combinazione aplotipica. Si osserva un dato simile di associazione anche a carico di un aplotipo che comprende tutti e tre i polimorfismi (p=0,0023; OR=1,53d) (tabelle R9).

I due polimorfismi rs926631 ed rs2761437 sono stati testati anche su una sottocasistica della popolazione spagnola (415 pazienti ed 400 controlli). Il dato non si è però riconfermato in questa popolazione (dati non presentati).

Linkage Disequilibrium nei geni CR1 e CR1L

I dati di genotipizzazione di tutti i polimorfismi per le popolazioni italiana sono stati analizzati con il software Haploviev per determinare i valori di LD (D' ed r2) e la distribuzione dei blocchi aplotipici sulla nostra popolazione. Rispetto ai dati predetti sulle famiglie CEPH del progetto HapMap, sul nostro campione si osserva LD tra i polimorfismi di *CR1* e quelli della porzione più 5'-terminale di *CR1L*. Tuttavia non si osserva LD tra i polimorfismi situati nel 3'UTR di *CR1L* (che hanno evidenziato un'associazione significativa sulla prima casistica) e gli altri SNPs di *CR1L* e *CR1*. Questo dato conferma la predizione effettuata sulle genotipizzazioni di HapMap. I polimorfismi della porzione 3' terminale di *CR1L* sono caratterizzati da un forte grado di LD.

I polimorfismi telomerici rispetto a *CR1L* mostrano un certo grado di LD anche nella nostra popolazione, e questo dato conferma quello predetto utilizzando i polimorfismi tipizzati sulle famiglie CEPH dal progetto Hapmap.

Studi di espressione su CR1L

La banca dati ncbi.nlm.nih.gov riporta tre diverse isoforme di RNA di *CR1L* (XM_931256, XM_931252 e XM_114735). Si tratta di isoforme ipotetiche basate su dati in silico, ma alcune di queste hanno trovato una conferma da dati sperimentali. Logar et al. sono stati gli unici ad aver evidenziato l'mRNA dell'isoforma XM_114735 (isoforma 1), da cellule ematopoietiche e da

tessuto linfoide fetale utilizzando primers specifici per la porzioni 3'UTR e per l'SCR1 di *CR1*. Questa isoforma è la prima ad essere stata riportata su banche dati, corrisponde al trascritto più lungo (1826 basi), contiene 14 esoni e codifica per i 7 SCR di *CR1L*. L'isoforma XM_931256 (isoforma 2) è stata evidenziata nel midollo osseo e nel tessuto linfoide fetale ed interamente sequenziata da Irshaid et al. (dati non pubblicati) e deriva da un evento di poliadenilazione alternativa. Al momento abbiamo studiato l'espressione dell'isoforma 1 e dell'isoforma 2. L'isoforma XM_931252 (isoforma 3) è stata segnalata solo più recentemente, pertanto la sua espressione deve ancora essere indagata. Quest'ultima è costituita solo dai primi quattro esoni del gene, più le prime 74 basi del quarto introne, e codifica solo per i primi due SCR.

Abbiamo valutato l'espressione di *CR1L* con la tecnica RT-PCR su tre diversi tessuti: linfociti attivati e posti in coltura, linfociti non attivati e midollo osseo. Tutti i tipi cellulari provenivano da umani adulti non affetti da patologie riconosciute. In ogni esperimento è stato amplificato sempre almeno un campione di DNA genomico, per escludere la possibilità di un risultato falsamente positivo dovuto all'amplificazione aspecifica di sequenze di DNA genomico che potrebbero contaminare il campione di RNA.

Per l'amplificazione dell'isoforma 1 abbiamo deciso di non seguire il protocollo adottato da Logar et al, poiché i primers da essi utilizzati non sono specifici per *CR1L*, ma possono amplificare anche sequenze del gene *CR1*. L'isoforma 1 è stata amplificata in due tempi:

1) con una prima RT-PCR "long-transcript" per la quale sono stati usati un primer senso specifico per l'SCR 5 di *CR1L*, sul quinto esone, (SCR5 sense cDNA F) ed un primer antisenso specifico per la porzione 3'UTR (*CR1L* cDNA ex12-13 R), che amplificano su cDNA un frammento di 918 basi corrispondente alla porzione di gene (dall'esone cinque alla porzione 3' terminale) che non è trascritta nelle altre due isoforme putative. Sono state tentate due differenti condizioni di amplificazione. Non è stata evidenziata alcuna trascrizione a questo livello, con nessuna condizione di PCR, in nessuno dei tre tipi cellulari utilizzati (dati non mostrati).

2) L'RT-PCR "long-transcript" è stata riamplificata tramite una PCR "emi-nested" per la quale sono stati utilizzati lo stesso primer SCR5 sense, ed un primer antisenso interno alla sequenza amplificata dalla prima coppia di primers, che si allinea sulla sequenza del messaggero di *CR1L* in corrispondenza della giunzione tra quinto e sesto esone (*CR1L* rna ex4 R). Questi primer amplificano una sequenza molto più corta, di 91bp, ugualmente specifica per l'isoforma 1. La sequenza completa dell'isoforma 1 come è stata riportata da Logar et al. (2004), ed i primer utilizzati per il nostro studio sono riportati nella fig. R4.

Con la seconda amplificazione (figura R5) abbiamo ottenuto una intensa banda di peso molecolare corrispondente a quello atteso (91bp). Nelle linee corrispondenti ai tre campioni di DNA genomico

utilizzati come controllo non si osserva questo prodotto, ma solo bande probabilmente aspecifiche. Tramite sequenziamento, è stata confermata l'identificazione del prodotto amplificato e ci si è accertati che non si trattasse di un prodotto causato dall'appaiamento aspecifico dei primers con sequenze codificanti del gene *CR1*, espresso in quantità nettamente superiore.

Il dato ottenuto dopo la PCR eminested è compatibile con livelli molto bassi o basali di trascrizione dell'isoforma più lunga di *CR1L* nei linfociti. Da questo esperimento non è possibile quantificare il trascrito. Inoltre occorrerebbe verificare se avviene la traduzione a proteina.

E' stato effettuato anche un esperimento utilizzando primers, in grado di amplificare indistintamente sia l'esone 4 di *CR1L* che l'esone 4 di CR1. Il prodotto ottenuto (181bp) è stato sequenziato. Nonostante l'elevata omologia tra i due geni sarebbe stato possibile in questo modo sequenziare entrambi i prodotti (*CR1* e *CR1L*) e distinguere il contributo apportato dall'uno o dall'altro gene concentrandosi sui punti in cui le sequenze dei due geni differiscono, che, qualora i due geni fossero espressi in quantità stechiometricamente paragonabile, sarebbero evidenti nel sequenziamento. Nell'esperimento effettuato solo il contributo apportato da *CR1* era evidenziabile, il che rafforza la tesi che l'isoforma 1 di *CR1L* sia trascritta solo a livelli molto bassi.

Per indagare sull'eventuale espressione dell'isoforma 2 è stata effettuata una RT-PCR utilizzando un primer antisenso specifico per l'isoforma 2 complementare alla sequenza consenso di poliadenilazione alternativa di questa isoforma ed ed un primer senso specifico per il primo esone (5' UTR) di *CR1L*. Il frammento amplificato è lungo 636 bp e contiene i primi 4 esoni di *CR1L*. Sono stati utilizzati due cDNA di midollo osseo di individui diversi, un cDNA di linfociti non attivati e due cDNA provenienti da linfociti attivati con PHA e posti in coltura di due diversi individui. In entrambi i campioni di midollo osseo si osserva una banda molto netta con un peso molecolare compatibile con quello del frammento cercato. Una banda di ugual peso molecolare ma molto più debole è presente anche in corrispondenza del cDNA di linfociti non attivati (figura R6). La banda non è visibile in corrispondenza del campione di DNA genomico co-amplificato come controllo, per cui si esclude che l'amplificato ottenuto possa essere dovuto ad una contaminazione da DNA genomico. Non è stata osservato alcun segno di amplificazione per il cDNA di linfociti attivati (immagine non inclusa).

Tab. R1 Associazione dei polimorfismi di *CR1L* nella popolazione italiana

SNP (allele)	CASISTICA 1					CASISTICA 2					TOTALE				
	frequenze (n camp)*		statistica			frequenze (n camp)*		statistica			frequenze (n camp)*		statistica		
	LES	CT	X2	p	OR (95% CI)	LES	CT	X2	p	OR	LES	CT	X2	p	OR (95% CI)
rs2796239 (A)	0,37 (217)	0,47 (185)	7,65	0,0057	0,66 (0,50-0,89)	0,40 (244)	0,42 (334)	NS	NS	NS	0,39 (461)	0,44 (519)	3,4	0,0652	0,82 (0,66-1,01)
3’UTR+41 (G)	0,20 (225)	0,29 (232)	8,70	0,0032	0,63 (0,46-0,86)	0,23 (191)	0,30 (190)	4,02	0,0450	0,71 (0,51-0,99)	0,21 (416)	0,30 (422)	12,7	0,0004	0,65 (0,51-0,83)
rs2761424 (T)	0,20 (236)	0,29 (233)	9,44	0,0021	0,62 (0,45-0,84)	0,23 (201)	0,30 (204)	5,50	0,0191	0,68 (0,49-0,94)	0,21 (437)	0,29 (437)	12,9	0,0003	0,65 (0,51-0,82)
3’UTR+137 (C)	0,37 (235)	0,45 (233)	5,62	0,0178	0,72 (0,55-0,95)	0,41 (202)	0,44 (203)	NS	NS	NS	0,38 (437)	0,44 (436)	5,12	0,0237	0,79 (0,64-0,97)
3’UTR+163 (A)	0,01 (235)	0,03 (234)	3,15	0,0758	NS	0,02 (203)	0,02 (203)	NS	NS	NS	0,02 (438)	0,03 (437)	0,96	0,3276	NS
rs6667500 (G)	0,38 (231)	0,49 (227)	10,01	0,0016	0,65 (0,49-0,85)	0,44 (202)	0,47 (203)	NS	NS	NS	0,41 (433)	0,48 (430)	7,14	0,0076	0,75 (0,61-0,93)

* frequenza dell’allele più raro (indicato tra parentesi nella colonna 1) e numero di campioni su cui il polimorfismo è stato analizzato

Tab. R2 Associazione dei polimorfismi di *CR1L* nella popolazione italiana ed in quella spagnola

SNP (allele)	ITALIANI					SPAGNOLI			ITALIANI + SPAGNOLI				
	frequenze (n camp)*		statistica			frequenze (n camp)*		statistica	frequenze (n camp)*		statistica		
	LES	CT	X2	p	OR (95% CI)	LES	CT		LES	CT	X2	p	OR
rs2796239 (A)	0,39 (461)	0,44 (519)	3,4	0,0652	0,82 (0,66-1,01)	0,43 (647)	0,43 (671)	NS	0,41 (1108)	0,43 (1190)	NS	NS	NS
3'UTR+41 (G)	0,21 (416)	0,30 (422)	12,7	0,0004	0,65 (0,51-0,83)	0,26 (230)	0,28 (259)	NS	0,23 (646)	0,29 (681)	9,73	0,001817	0,76 (0,63-0,90)
rs2761424 (T)	0,21 (437)	0,29 (437)	12,9	0,0003	0,65 (0,51-0,82)	0,25 (671)	0,27 (678)	NS	0,24 (1108)	0,28 (1115)	9,81	0,001739	0,80 (0,70-0,92)
3'UTR+137 (C)	0,38 (437)	0,44 (436)	5,12	0,0237	0,79 (0,64-0,97)	0,43 (660)	0,43 (666)	NS	0,41 (1097)	0,44 (1102)	3,00	0,083334	NS
3'UTR+163 (A)	0,02 (438)	0,03 (437)	0,96	0,3276	NS	0,02 (663)	0,03 (659)	NS	0,02 (1101)	0,03 (1096)	NS	NS	NS
rs6667500 (G)	0,41 (433)	0,48 (430)	7,14	0,0076	0,75 (0,61-0,93)	0,46 (658)	0,49 (671)	NS	0,44 (1091)	0,48 (1101)	7,39	0,006556	0,85 (0,75-0,96)

* frequenza dell'allele più raro (indicato tra parentesi nella colonna 1) e numero di campioni su cui il polimorfismo è stato analizzato

TabR3 Associazione con le frequenze genotipiche di *CR1L*

		LES		controlli					
SNP		**n**	**freq**	**n**	**freq**	**χ^2**	**p**	**OR**	**95% CI**
rs2796239	GG	166	0,36	163	0,31	NS	NS		
	GA	233	0,51	260	0,50	NS	NS		
	AA	62	0,13	96	0,18	4,23	**0,039600**	0,68	0,48-0,98
	tot	461		519					
				Overall p value			0,06691		
3'UTR+41	AA	250	0,60	203	0,48	11,65	**0,000641**	1,62	1,22-2,16
	GA	150	0,36	188	0,45	5,93	**0,014888**	0,70	0,53-0,94
	GG	16	0,04	31	0,07	4,21	**0,040225**	0,50	0,26-0,97
	tot	416		422					
				Overall p value			**0,000962**		
rs2761424	CC	266	0,61	208	0,48	14,98	**0,000109**	1,71	1,30-2,26
	CT	156	0,36	200	0,46	8,76	**0,003074**	0,66	0,50-0,87
	TT	15	0,03	29	0,07	NS	NS		
	tot	437		437					
				Overall p value			**0,000205**		
3'UTR+137	TT	156	0,36	129	0,30		0,06388		
	CT	225	0,51	228	0,52	NS	NS		
	CC	56	0,13	79	0,18	4,30	**0,038093**	0,66	0,45-0,98
	tot	437		436					
				Overall p value			**0,038868**		
rs6667500	CC	145	0,33	113	0,26	5,01	**0,025197**	1,41	1,04-1,91
	GC	223	0,52	222	0,52	NS	NS		
	GG	65	0,15	95	0,22	6,70	**0,009628**	0,62	0,43-0,90
	tot	433		430					
				Overall p value			**0,008288**		

Tab. R4 Frequenze genotipiche dei polimorfismi rs2796239, 3'UTR+137 e rs6667500 nei campioni rs2761424 T negativi appartenenti alla popolazione italiana

		LES		controlli		statistica
SNP		**n**	**freq**	**n**	**freq**	
rs2796239	GG	145	0,58	118	0,62	NS
	GA	89	0,36	63	0,33	NS
	AA	16	0,06	9	0,05	NS
3'UTR+137	TT	156	0,59	129	0,62	NS
	CT	91	0,34	70	0,34	NS
	CC	17	0,06	8	0,04	NS
rs6667500	CC	138	0,52	109	0,53	NS
	GC	104	0,40	85	0,41	NS
	GG	21	0,08	12	0,06	NS

Tab. R5 Frequenze geniche dei polimorfismi rs2796239, 3'UTR+137 e rs6667500 nei campioni rs2761424 T negativi appartenenti alla popolazione italiana

		LES		controlli		
polimorfismo	**allele**	**N***	**freq**	**N***	**freq**	**statistica**
rs2796239	A	250	0,24	190	0,21	NS
3'UTR+137	C	264	0,24	207	0,21	NS
rs6667500	G	263	0,28	206	0,26	NS

* numero di individui analizzati per ogni polimorfismo

Tab. R6 Analisi di regressione sulla popolazione italiana

Marcatore analizzato per l'associazione

Marcatore condizionante	**rs2796239**	**3'UTR+41**	**rs2761424**	**3'UTR+137**	**rs6667500**
rs2796239	-	0,00084	0,0011	0,043	0,035
3'UTR+41	NS	-	NS	NS	NS
rs2761424	NS	NS	-	NS	NS
3'UTR+137	NS	0,0035	0,0044	-	0,19
rs6667500	NS	0,0079	0,0096	NS	-

Tab. R7 Associazione con gli aplotipi nella popolazione italiana (A) ed in quella spagnola (B)

A

	frequenze		statistica			
aplotipo	**LES (n=378)**	**Controlli (n=381)**	**χ2**	**p**	**OR**	**95% CI**
GCTGC	0,567	0,506	5,31	0,02120	1,27	1,04-1,57
ATCGG	0,202	0,283	12,72	0,00036	0,65	0,51-0,82
ACCGG	0,162	0,138	NS	NS	NS	NS
GCTAG	0,019	0,028	NS	NS	NS	NS
GCTGG	0,022	0,016	NS	NS	NS	NS

B

	frequenze		statistica	
aplotipo	**LES (n=671)**	**Controlli (n=678)**	**χ^2**	**p**
GCTGC	0,525	0,500	NS	NS
ATCGG	0,246	0,253	NS	NS
ACCGG	0,170	0,164	NS	NS
GCTAG	0,023	0,033	NS	NS
GCTGG	0,017	0,027	NS	NS

I polimorfismi che compongono l'aplotipo sono elencati nel seguente ordine: rs2796239 (G/A), rs2761424 (C/T), 3'UTR+137 (T/C), 3'UTR+163 (G/A), rs6667500 (C/G). Il polimorfismo 3'UTR+41 (A/G), che è in LD assoluto con l'rs2761424 non è stato incluso nell'analisi. Analisi effettuata con il software Haploview.

Tab. R8 Frequenze alleliche ed analisi statistica dei tagSNPs nelle regioni centromeriche e telomeriche rispetto a *CR1L*

		LES		controlli		
polimorfismo	**allele**	**n**	**freq**	**n**	**freq**	**statistica**
CR1 rs9429944	T	215	0,25	205	0,25	NS
CR1 rs2274567	G	240	0,26	198	0,24	NS
CR1 rs2296160	T	240	0,20	200	0,21	NS
CR1 rs 3737002	A	237	0,20	202	0,25	NS
CR1 rs 6691117	C	237	0,28	206	0,25	NS
MCP rs2796270	G	216	0,42	214	0,43	NS
rs926631	T	214	0,37	215	0,43	p=0,07903
MCP rs2761437	T	216	0,14	212	0,19	NS
rs2796249	A	215	0,34	212	0,32	NS
MCP rs7144	C	215	0,28	207	0,30	NS
CD34 rs607952	T	181	0,40	202	0,44	NS
LOC148696 rs2724373	T	210	0,39	207	0,38	NS
MCP rs11118555	A	190	0,12	207	0,10	NS
CD34 rs2745955	A	213	0,32	209	0,30	NS
CR1L rs2796239	*A*	*206*	*0,36*	*200*	*0,45*	*p=0,0201 OR=0,70 95% CI =0,52-0,95*
CR1L rs2761424	*T*	*206*	*0,20*	*200*	*0,28*	*p=0,0107 OR=0,65 95% CI=0,46-0,91*
CR1L ex10+137	*C*	*206*	*0,36*	*200*	*0,45*	*p=0,0262 OR=0,72 95% CI=0,54-0,96*
CR1L rs6667500	*C*	*206*	*0,38*	*200*	*0,52*	*p=0,0043 0R=0,66 95% CI=0,49-0,88*

Le ultime 4 righe della tabella contengono le frequenze alleliche e l'analisi statistica relativa ai polimorfismi statisticamente associati di *CR1L* (rs2796239, rs2761424 ex10+137 rs6667500) nella sottocasistica esaminata per i polimorfismi centromerici e telomerici rispetto a *CR1L*.

Tab. R9A Associazione con gli aplotipi rs2761424 + rs926631 + rs2761437

	frequenze		statistica		
aplotipo	**LES (n=206)**	**Controlli (n=200)**	**$\chi 2$**	**p**	**OR (95% CI)**
CCC	0,55	0,45	9,31	0,0023	1,53 (1,15-2,04)
CTC	0,22	0,23	NS	NS	NS
TTT	0,11	0,13	NS	NS	NS
TCC	0,07	0,11	3,63	0,0567	0,61 (0,36-1,03)
CTT	0,03	0,05	NS	NS	NS
TTC	0,02	0,03	NS	NS	NS

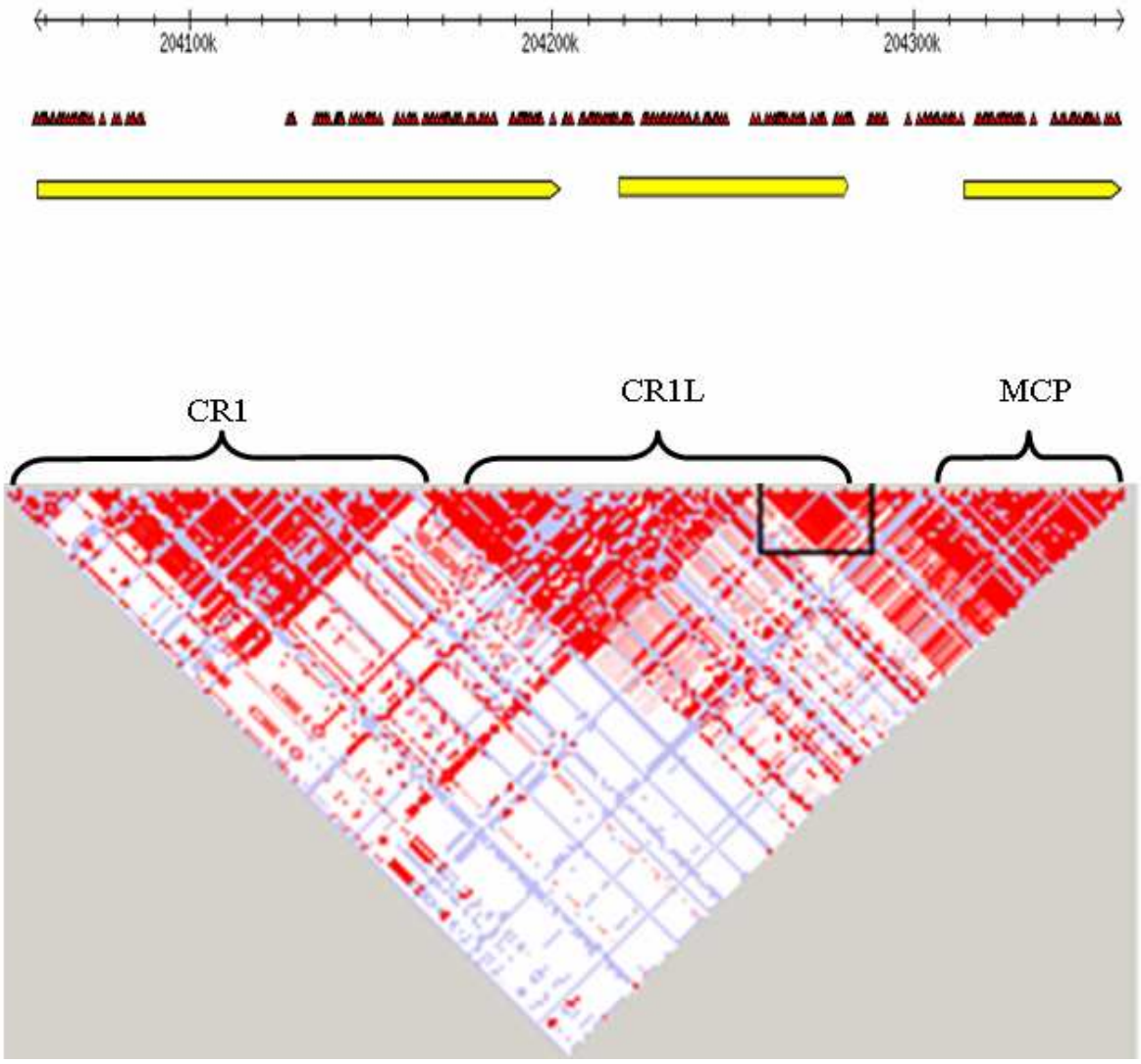
Tab. R9B Associazione con gli aplotipi rs2761424 + rs2761437

	frequenze		statistica		
aplotipo	**LES (n=206)**	**Controlli (n=200)**	**$\chi 2$**	**p**	**OR (95% CI)**
CC	0,77	0,68	8,62	0,0033	1,59 (1,15-2,19)
TT	0,11	0,14	NS	NS	NS
TC	0,09	0,14	5,04	0,0248	0,60 (0,38-0,96)
CT	0,03	0,05	NS	NS	NS

Tab. R9C Associazione con gli aplotipi rs2761424 + rs926631

	frequenze		statistica		
aplotipo	**LES (n=206)**	**Controlli (n=200)**	**$\chi 2$**	**p**	**OR (95% CI)**
CC	0,56	0,46	7,87	0,0050	1,48 (1,11-1,98)
CT	0,25	0,27	NS	NS	NS
TT	0,13	0,17	NS	NS	NS
TC	0,07	0,10	NS	NS	NS

Fig. R1 LD nella regione cromosomica compresa tra *CR1* ed *MCP*



Dati di genotipizzazione di HapMap sulle famiglie CEPH. Nell'elaborazione sono presenti tutti gli SNP al momento genotipizzati. Le aree rosse sono caratterizzate da D'>0,9. Il riquadro contiene i polimorfismi di *CR1L* associati con il LES.

Fig R2 LD nella nostra popolazione tra i polimorfismi di *CR1L* e quelli di *CR1*

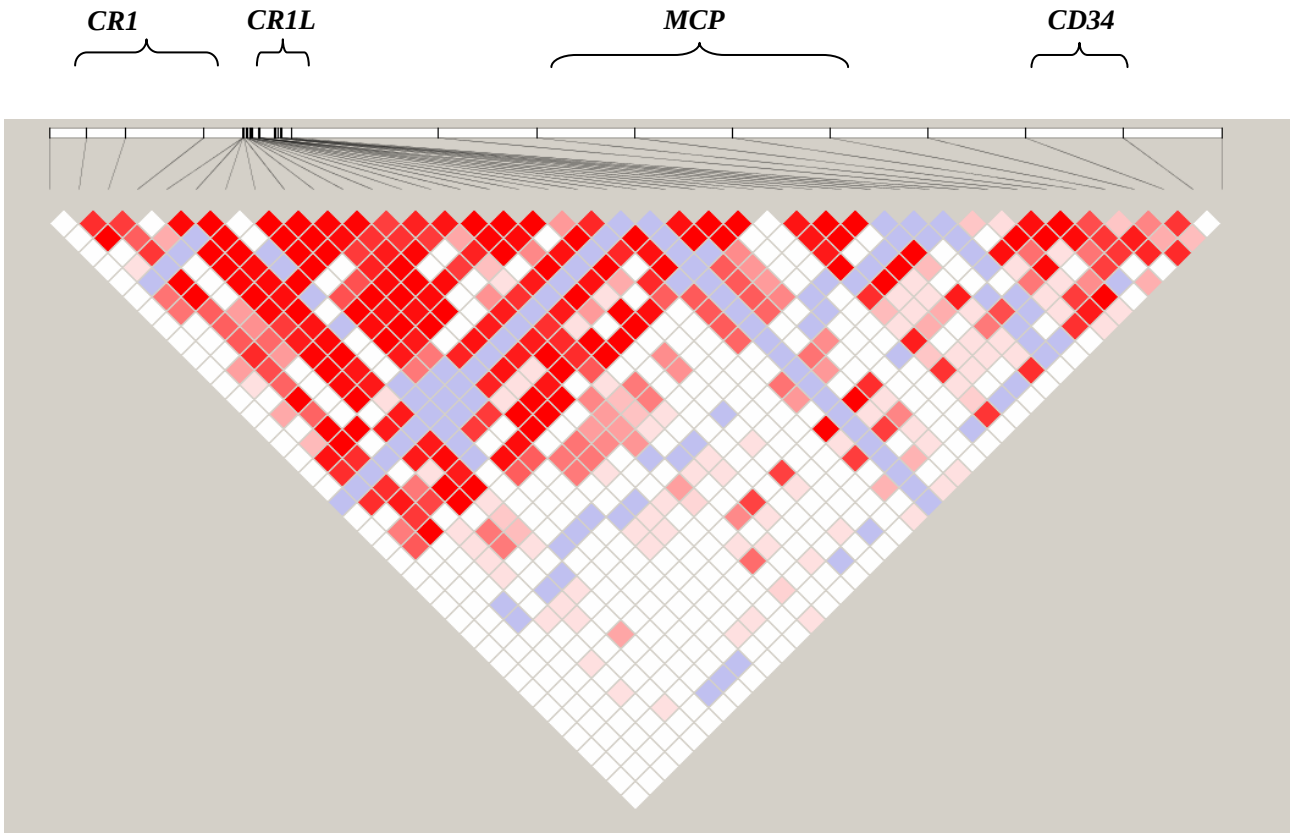
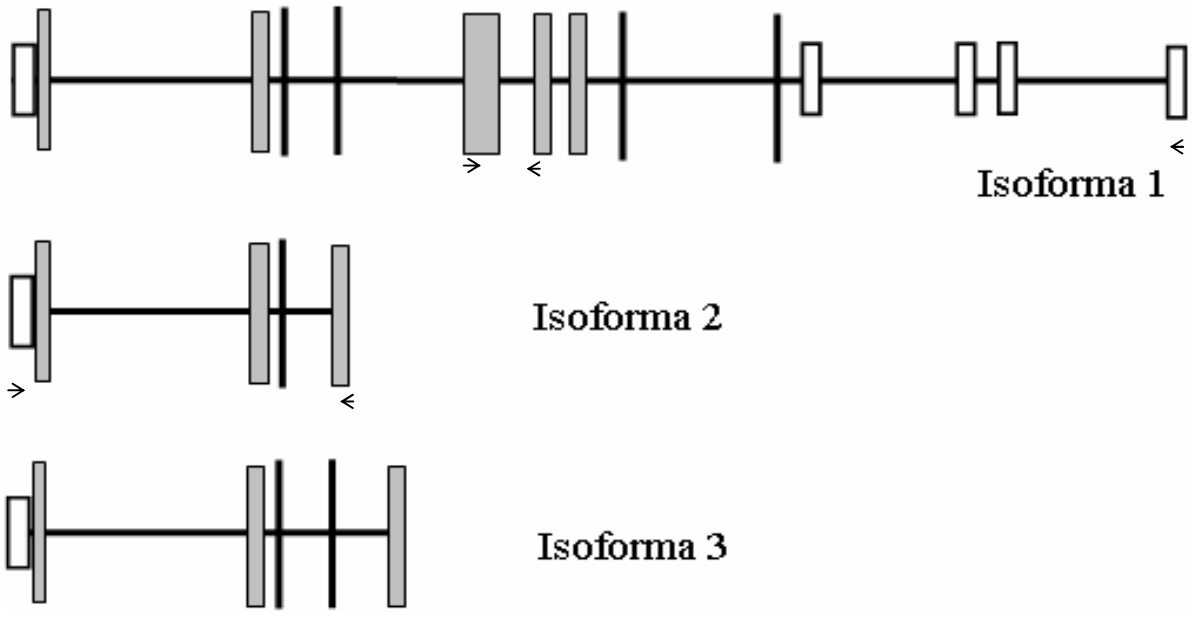
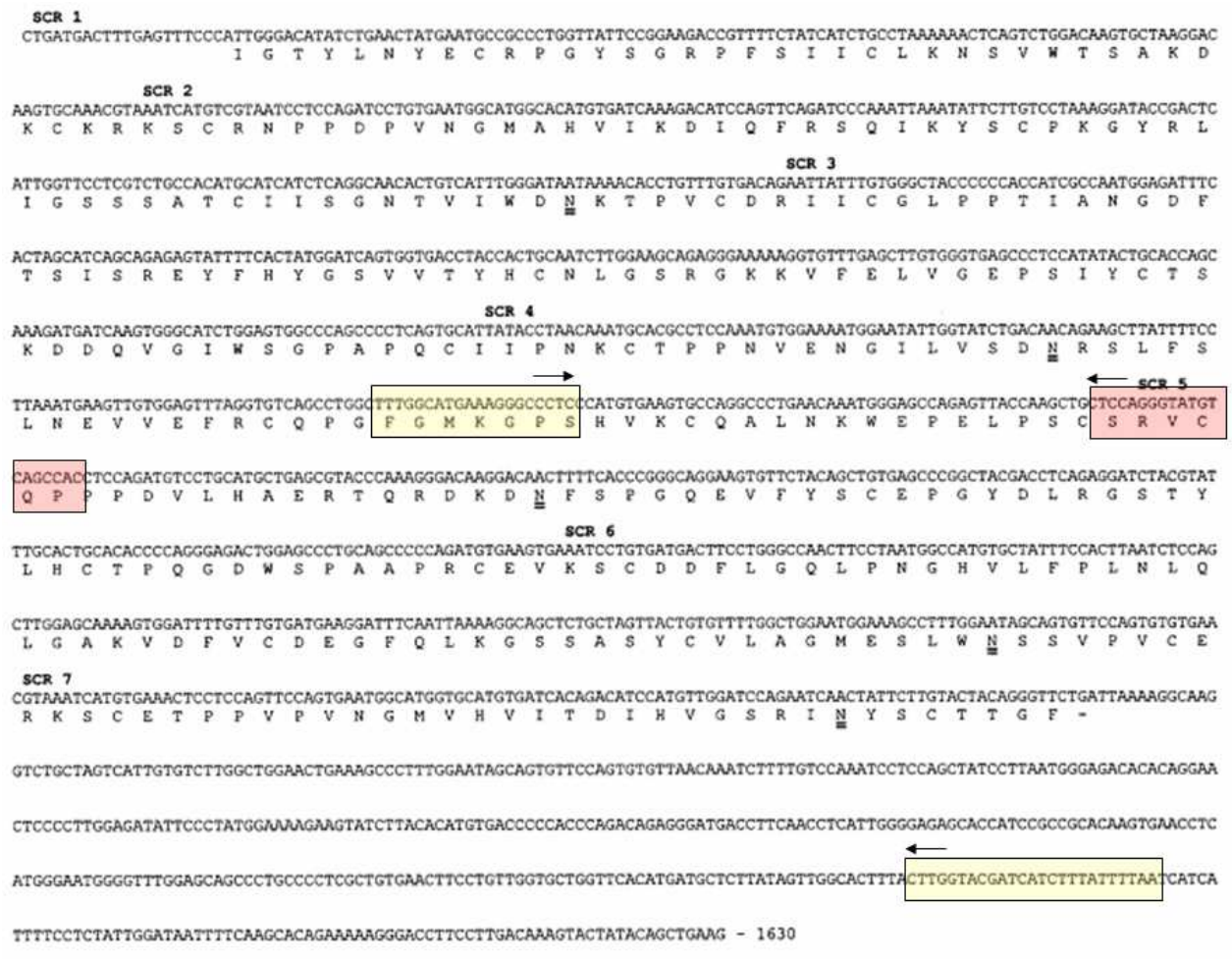


Fig R3 Le tre isoforme di *CR1L*



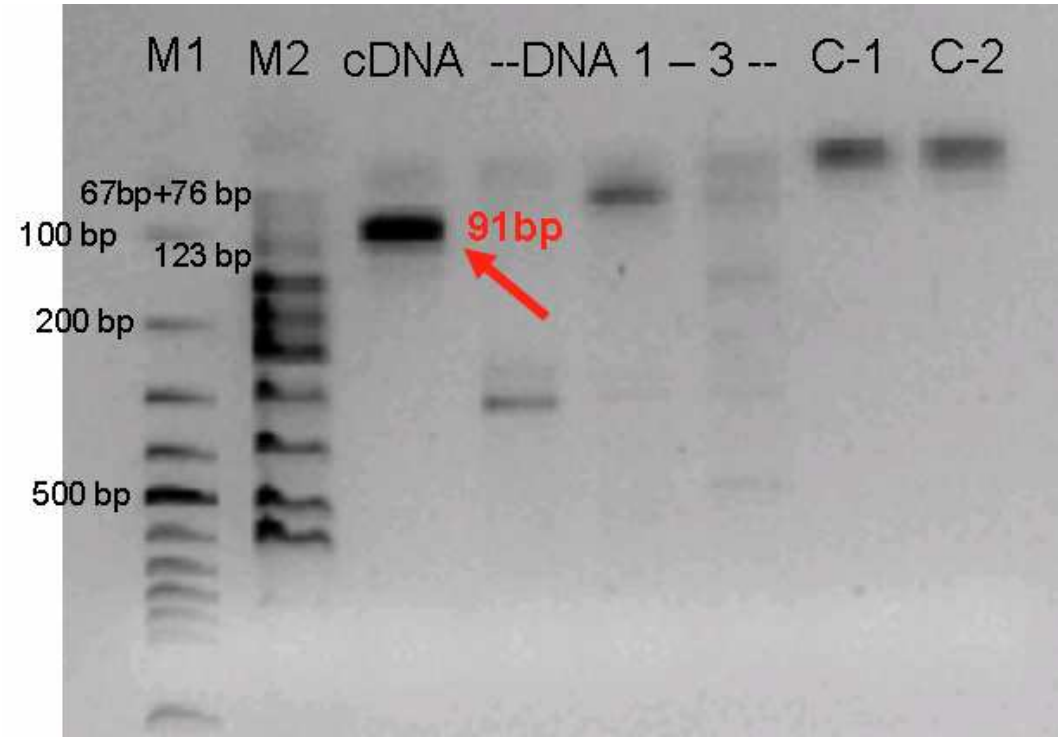
Le frecce indicano la posizione dei primers utilizzati per le RT-PCR.

Fig. R4 Sequenza di cDNA dell'isoforma 1 (Logar et al. 2004)



Sequenza nucleotidica e sequenza aminoacidica dedotta dell'isoforma 1 di *CR1L*, come è stata riportata da Logar et al. nel 2004. La sequenza nucleotidica è indicata in maiuscolo con il codice a singola lettera, e gli aminoacidi sono posizionati sotto ai rispettivi codoni. I primers per la RT-PCR long transcript sono stati evidenziati in giallo, mentre il primer antisenso utilizzato per l'emi-nested è evidenziato in rosso. Rispetto alla sequenza depositata nella banca dati ncbi.nlm.nih.gov quella riportata da Logar et al. è incompleta al 5,'manca cioè del primo esone e di parte del secondo. Questo fatto tuttavia non influenza la nostra analisi, in quanto i primers utilizzati sono stati scelti molto più a valle.

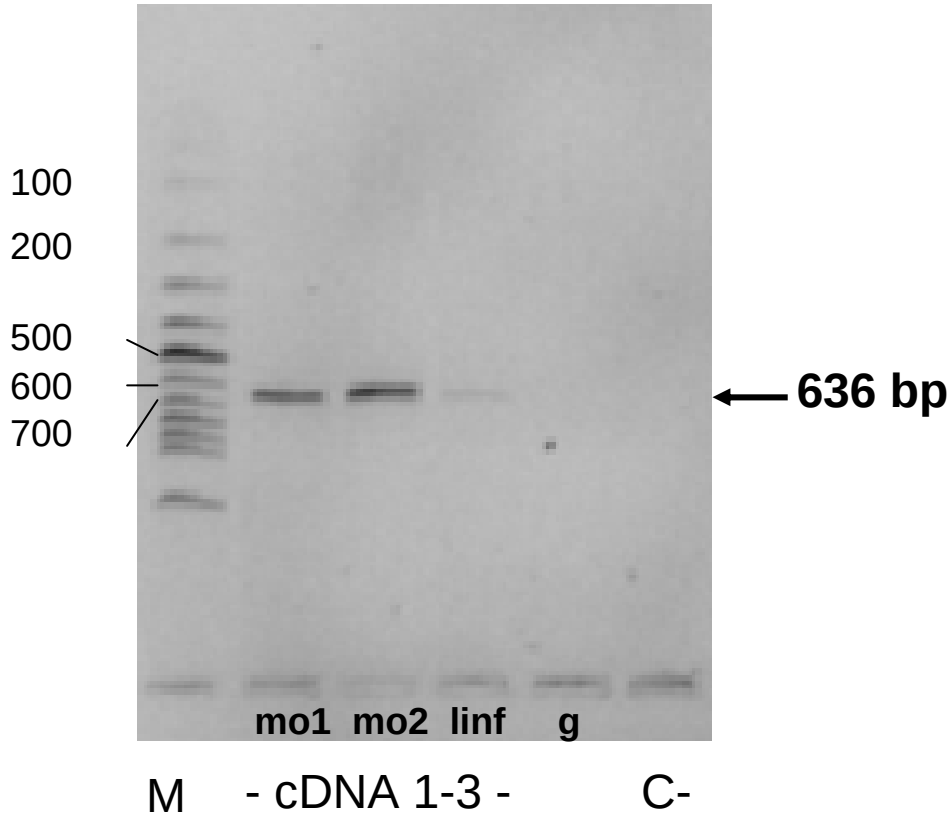
Fig. R5 Nested su RT-PCR di *CR1L* isoforma 1



Emi-nested dell'RT-PCR "long-transcript" di CR1L su cDNA di linfociti attivati.

Le linee 1 e2 sono state caricate con due differenti marcatori di peso molecolare (100bp DNA ladder e pBR322 DNA-Msp I digest). Sono stati indicati a lato i valori di peso molecolare utili per identificare la banda a 91bp (indicata dalla freccia rossa) nella terza linea e della banda più intensa.del marcatore 100bp DNA ladder (500bp). Le linee 4-6 contengono tre diversi campioni di DNA genomico amplificati co-amplificati con il cDNA in entrambe le fasi dell' amplificazione. Le bande che si possono osservare non hanno il peso molecolare corretto e sono prodotti aspecifici. Le ultime due linee (C-1 e C-2) contengono rispettivamente il controllo negativo della PCR "long transcript" riamplificato ed il controllo negativo della PCR nested.

Fig. R6 RT-PCR di *CR1L* isoforma 2



RT-PCR dell'isoforma 2 di *CR1L*. mo1 e mo2 = midollo osseo 1 e 2; linf = linfociti non attivati, g = DNA genomico. La prima linea è stata caricata col marcatore di peso molecolare 100bp DNA ladder. Nelle linee 2-4 si osserva una banda di peso molecolare compreso tra le 600 e le 700 bp, compatibile con le dimensioni del frammento che si stava cercando (636 bp). La banda non è visibile in corrispondenza della linea caricata con l'amplificato su DNA, per cui non è dovuta ad una contaminazione da DNA genomico.

Analisi mutazionale

DATI NON ANCORA PUBBLICATI

PRF1

Analisi mutazionale di PRF1

L'intera sequenza codificante del gene *PRF1* (cromosoma 10q22.1) è stata analizzata per la ricerca di nuove variazioni, tramite DHPLC e/o sequenziamento diretto, in 154 pazienti affetti da LES, 131 pazienti affetti da SSc e 559 controlli sani. Tutti i pazienti ed i controlli appartengono alla popolazione italiana.

Sono state riscontrate complessivamente 22 variazioni (11 non sinonime ed 11 sinonime), otto delle quali non sono state mai descritte precedentemente. Quattro di queste ultime sono varianti non sinonime (tabella R10).

Tra le sostituzioni non sinonime, c.403G>A, responsabile della sostituzione aminoacidica Val135Met è riportata sulle banche dati (rs12263572) con una MAF (minor allele frequency) di 0,0014 nella popolazione CEU (http://hapmap.ncbi.nlm.nih.gov/). Nella nostra popolazione è stata osservata in un solo paziente affetto da LES. La variazione c.11G>A – rs35418374 causa il cambiamento aminoacidico Arg4His nel peptide segnale e nella nostra casistica è stata riscontrata solo in un paziente affetto da SSc. Nei database pubblici è riportata con MAF = 0,013 nella popolazione AGI-ASP (Caucasoidi e Afro-Americani apparentemente sani). Tutte le altre variazioni non sinonime sono state osservate anche nella casistica di controlli. Due di queste (c.272C>T – rs35947132 e c.755A>G –rs28933375) sono polimorfismi del gene *PRF1* per i quali è già stato ipotizzato un ruolo funzionale (vedi discussione) e causano rispettivamente le sostituzioni p.Ala91Val e p.Asn252Ser.

Le variazioni nucleotidiche c.822C>T (rs995821) e c.900T>C (rs995822) sono state precedentemente riportate in letteratura come polimorfismi comuni di *PRF1* non associati a patologie. I due polimorfismi non predicono una sostituzione e non coinvolgono siti di splicing, le loro frequenze alleliche non differiscono in modo statisticamente significativo tra controlli ed i pazienti affetti da LES o da SSc e pertanto non sono stati ulteriormente analizzati. Sono inoltre state riscontrate altre due variazioni sinonime (c.435G>A e c.462A>G), già descritte in LD assoluto (D'=1, r^2=1) con c.755A>G (p.Asn252Ser), che infatti sono state osservate solo nei soggetti portatori di questa variazione. Le due varianti sinonime non sono state considerate nelle analisi successive. Ad eccezione di p.Ala91Val, tutte le altre variazioni sono sempre state osservate in eterozigosi.

Analisi in silico

Per predire la rilevanza funzionale delle sostituzioni non sinonime osservate sono stati utilizzati 4 diversi algoritmi: PolyPhen, SIFT SNAP e PMUT).

Le sostituzioni p.Arg232His, p.Val329Ile e p.Gly334Ser si trovano nel dominio MACPRF. In particolare p.Arg232His cade in una sequenza ad alfa elica anfipatica strutturalmente conservata, compresa tra i residui 212-241, omologa alle componenti C5b e C9 della cascata del complemento (Voskoboinik et al. 2006). Le varianti p.Phe421Cys, p.Ala437Val e p.His514Arg sono situate nella sequenza proteica di omologia con il dominio C2 della proteina chinasi C (PKC). Le altre sostituzioni aminoacidiche non ricadono in domini funzionali di PRF1.

Solo la sostituzione p.His514Arg è stata predetta come non tollerata o patologica da tre dei quattro programmi utilizzati (tabella R12). I risultati dell'allineamento con sequenze proteiche ortologhe e paraloghe ha dimostrato che l'istidina in posizione 514 è conservata tra le diverse specie (PolyPhen).

Per stabilire la rilevanza funzionale della variante p.Arg4His, localizzata nel peptide segnale, è stato utilizzato il programma Signal P 3.0. La variante non sembra influenzare né la posizione del sito di taglio (calcolata tra i nucleotidi 21 e 22, in accordo con il database SPdp - http://proline.bic.nus.edu.sg/spdb), né la probabilità predetta (signal prediction probability), calcolata sulla base della percentuale di variazione del valore di D score (0,21%), come suggerito da Jarjanazi et al (2007).

Per le variazioni sinonime è stata valutata la loro influenza sullo splicing tramite il programma SpliceView che predice l'eventuale predizione o rimozione di siti accettori o donatori di splicing e tramite l'algoritmo ESEfinder che predice l'eventuale introduzione o rimozione di sequenze di interazione con proteine coinvolte nel processo di splicing (ESE Exonic Splicing Enhancer). Inoltre è stato valutato se le sostituzioni nucleotidiche potessero determinare l'utilizzo di codoni più o meno frequenti rispetto a quelli normalmente usati per la codifica dell'amminoacido.

Tra quelle osservate solo nei pazienti, le variazioni sinonime 807C>T e c.999C>T potrebbero influenzare lo splicing attraverso interazioni con varie proteine regolatorie dello splicing: c.807C>T crea un nuovo sito per SRp35 e c.999C>T rimuove un sito di legame già esistente per SRp40. Inoltre c.273G>A crea un nuovo sito di legame per SRp40, c.1620A>G crea una nuova sequenza di interazione per SF2/ASF, mentre c.1356C>T introduce un nuovo sito di legame per due diverse proteine (SF2/ASF e SC35), ma ne rimuove uno già esistente per SRp40. c.1620A>G potrebbe anche introdurre un nuovo sito accettore di splicing (score 0,53). Il dato non è tuttavia significativo a causa del basso numero di campioni.

Analisi Funzionale. La rilevanza funzionale delle due varianti p.His514Arg e p.Val135Met è stata valutata tramite la determinazione dell'attività NK, dei livelli di espressione di perforina e della proporzione di cellule NK nel sangue periferico. Apparentemente le due variazionei non influiscono sull'attività NK, tuttavia abbiamo osservato una leggera diminuzione (al di sotto della soglia del quinto percentile) della proporzione di cellule esprimenti PRF1 nel paziente portatore della variante p.His514Arg. Inoltre il paziente con la sostituzione p.Val135Met ha rivelato bassi livelli di cellule NK $CD3^{-}CD16^{+}$ (tabella R13). L'analisi funzionale è stata condotta dal gruppo della professoressa Moretta del policlinico San Matteo di Pavia.

Parametri clinici

I parametri clinici dei pazienti affetti da LES e da SSc portatori di sostituzioni non sinonime del gene *PRF1* sono elencati nelle tabelle R10A e R10B. Tre pazienti affetti da SSc su quattro portatori di una variazione non sinonima di *PRF1* (75%) sono positivi per la presenza di anticorpi anti-topoisomerasi I. La frequenza di questo parametro nella nostra popolazione è del 39%. Non sono state rese disponibili informazioni sui dati clinici dei pazienti LES portatori della variante c.985G>A (p.Val329Ile).

Rilevanza per la suscettibilità alla malattia

Complessivamente la frequenza allelica comulativa delle variazioni di PRF1 è superiore sia nei pazienti affetti da LES (0.12) che in quelli con SSc (0.12) rispetto ai controlli sani (0.084), e questa differenza è statisticamente significativa per entrambe le patologie (LES: p=0.0038, OR=1.89 95%CI=1.22-2.94; SSc: p=0.0028, OR=1.99 95%CI=1.25-3.16). Se si considerano solo le varianti non sinonime le frequenze alleliche continuano a presentare una distribuzione significativamente diversa tra la casistica di controlli (0.062) e le due casistiche di pazienti (LES=0.097; SSc=0.11), nonostante la significatività statistica diminuisca (LES: p=0.043 OR=1.63 95%CI=1.01-2.61, SSc: p=0.0082 OR=1.89 95%I=1.17-3.05).

Parte di questa variabilità alle due varianti non sinonimie più comuni, p.Ala91Val e p.Asn252Ser. La stessa analisi è stata ripetuta dopo aver escluso queste variazioni dal computo totale, e le frequenze allliche (LES=0,029; SSc=0,027; controlli=0,0081) continuano a differire in modo statisticamente significativo tra pazienti e controlli (LES: p=0,00693 OR=3,71 95%CI=1,29-10,64; SSc: p=0,0198 OR=3,38 95%CI=1,06-10,30). Se si considerano solo le varianti non sinonimie ad eccezione di p.Ala91Val e p.Asn252Ser le differenze non sono più significative a causa della bassa numerosità campionaria, ma è ancora possibile osservare un leggero trend (LES=0,0097; SSc=0,015; controlli=0,0045).

Le frequenze alleliche della variante p.Asn252Ser misurate nei controlli differiscono in modo statisticamente significativo rispetto a quelle riscontrate sia nella casistica di LES (0.0027 vs. 0.019, p=0.0045, OR=7.34, 95%CI=1.64-37.41), sia nei pazienti con SSc (0.0027 vs. 0.015, p=0.027, OR=5.76, 95%CI=1.09-32.54).

Poiché p.Ala91Val e p.Asn252Ser sono le due variazioni di sequenza non sinonime più frequentemente riscontrate a carico del gene *PRF1*, e a causa delle differenze statisticamente significative osservate nella distribuzione delle frequenze alleliche di p.Asn252Ser per entrambe le malattie, abbiamo deciso di estendere l'analisi di questi due polimorfismi in altri 443 casi di LES, 73 pazienti con SSc e 1297 controlli sani. Nella casistica totale (597 LES, 1856 controlli) le frequenze alleliche del polimorfismo p.Asn252Ser nei LES (0.010) sono significativamente diverse (p=0.039 OR=2.35 95%CI=1.04-5.24) da quelle misurate nei controlli sani (0.0043) (tabella R12B). Inoltre abbiamo riscontrato una differenza statisticamente significativa (0.057 vs. 0.093, p=0.0054 OR=1.70 95%CI=1.16-2.47) tra controlli (n=1876) e pazienti affetti da SSc (n=204) nelle frequenze alleliche del polimorfismo p.Ala91Val (tabella R14A). La distribuzione dei genotipi nelle varie popolazioni non devia rispetto a quello atteso in base all'equilibrio di Hardy-Weinberg. Complessivamente la variante p.Ala91Val è stata osservata in 73 pazienti affetti da LES (69 eterozigoti e 4 omozigoti), 36 pazienti con SSc (34 eterozigoti e 2 omozigoti) e 204 controlli sani (196 eterozigoti e 8 omozigoti). Il polimorfismo p.Asn252Ser è portato complessivamente da 12 LES, 4 SSc e 16 controlli, sempre in eterozigosi. Uno dei pazienti affetto da LES era eterozigote sia per p.Ala91Val che per p.Asn252Ser. Tramite una reazione di PCR allele specifica appositamente allestita è stato stabilito che gli aminoacidi 91Val e 252Ser sono codificati su due alleli distinti (eterozigosi composta). Sul database delle mutazioni di PRF1 (http://bioinf.uta.fi/PRF1base/prf1pub) non è al momento riportato un paziente con questa combinazione genotipica.

Tab.R10 Riassunto delle variazioni di *PRF1* osservate nel primo pannello di pazienti (LES e SSc) e controlli.

Variazione nucleotidica	Cambiamento aminoacidico	esone	ID*	Segnalazioni precedenti§	SLE#	SSc#	controlli#
c.11G>A	**Arg4His**	2	rs35418374	AA, SoJIA, controlli	0/154	1/131	0/559
c.189A>C	Thr63Thr	2		nuova	1/154	0/131	0/559
c.272C>T	**Ala91Val**	2	rs35947132	FHL2, controlli	21/154	21/131	57/559
c.273G>A	Ala91Ala[1]	2		nuova	1/154	1/131	1/559
c.368G>A	**Arg123His**[2]	2		ALCL, controlli	0/154	1/131	1/559
c.403G>A	**Val135Met**	2	rs12263572	Database SNP	1/154	0/131	0/559
c.435G>A	Val145Val	2		Database SNP	6/154	4/131	3/559
c.462A>G	Ala154Ala	2		Database SNP	6/154	4/131	3/559
c.519G>A	Thr173Thr[3]	2		controlli	1/154	1/131	1/559
c.695G>A	**Arg232His**[4]	3		FHL2, ALCL, familiari sani	0/154	0/131	1/559
c.755A>G	**Asn252Ser**	3	rs28933375	FHL2, controlli	6/154	4/131	3/559
c.807C>T	His269His	3		nuova	1/154	0/131	0/559
c.822C>T	Ala274Ala	3	rs885821	Database SNP	54/163	48/136	258/772
c.900T>C	His300His	3	rs885822	Database SNP	110/162	92/136	562/771
c.985G>A	**Val329Ile**	3		nuova	1/154	0/131	0/559
c.999C>T	Pro333Pro	3		SM	0/154	1/131	0/559
c.1000G>A	**Gly334Ser**	3		nuova	0/154	1/131	1/559
c.1262T>C	**Phe421Cys**	3		FHL2, ALCL	0/154	1/131	1/559
c.1310C>T	**Ala437Val**	3		nuova	0/154	0/131	1/559
c.1356C>T	Thr452Thr	3		nuova	1/154	0/131	1/559
c.1541A>G	**His514Arg**	3		nuova	1/154	0/131	0/559
c.1620A>G	Gln540Gln	3		SM	1/154	0/131	1/559

La numerazione fa riferimento al clone cDNA di GeneBank M28393, ATG = +1.
Ad eccezione di p.Ala91Val, tutte le altre variazioni sono state osservate in eterozigosi.

*Numero di identificazione riportato sul database http://www.ncbi.nlm.nih.gov.
§FHL2 = Linfoistiocitosi Emofagocitica Familiare; ALCL = Linfoma Anaplastico a Grandi Cellule; SoJIA = Systemic onset Juvenile Idiopathic Arthritis; SM = Sclerosi Multipla; AA = Anemia Aplastica.
#numero di individui in cui è stata osservata la variazione sul totale dei soggetti analizzati per singole casistiche (LES, SSc e controlli)
[1]Il controllo sano è eterozigote anche per il polimorfismo p.Ala91Val.
[2]Sia il paziente SSc che il controllo sano portatori della variante p.Arg123His sono eterozigosi anche per la variazione sinonima c.519G>A.
[3] Sia il paziente SSc che il controllo sano sono eterozigoti anche per la sostituzione non sinonima p.Arg123His
[4]Il controllo è portatore anche di 91Val sullo stesso allele di 232His, come è stato stabilito tramite PCR allele-specifica.

Tab. R11 Caratteristiche cliniche ed immunologiche dei pazienti con LES (A) e SSc (B) che riportano varianti rare di *PRF1*.

A

Genotipo PRF1	sesso	MR	PS	S	Ar	GNF	AHA	LA	Lp/Lp	TP	NI	aPL	ANA	aDNA	SSA	LAC	RF
V135M/wt	F	-	-	-	-	+	+	-	-	+	-	-	+	+	-	-	-
H514R/wt	F	-	-	-	+	+	-	-	+	-	+	-	+	+	-	-	-

MR = malar rash; PS = fotosensibilità; LED = lesioni discoidi; S = sierositi; A = artrite; GNF = glomerulonefrite; LA= linfoadenopatia; AHA = anemia emolitica autoimmune; Lp/Lp = leucopenia e/o linfopenia; TP = trombocitopenia; NI = coinvolgimento neurologico; aPL = sindrome da antifosfolipidi; ANA= anticorpi antinucleo; aDNA= anticorpi anti-dsDNA; SSA = anticorpi antiRo (SSA); LAC = presenzadi attività lupus anticoagulante; RF = fattore rheumatoide.

Entrambi i pazienti in questa tabella sono negativi per la presenza di lesioni discoidi e di anticorpi anti-RNP e/o Sm. Le informazioni cliniche riguardanti la presenze di ulcere orali, fenomeno di Raynaud e anticorpi anti-La (SSB) non erano disponibili. Le informazioni cliniche relative ai pazienti che riportano la variantie c.985G>A (p.Val329Ile) non erano disponibili.

B

Genotipo PRF1	sesso	età alla diagnosi	diagnosi	fibrosi polmonare	PAH	ACA	anti-topo I	anti-RNApIII
R123H/T173T	F	50	lSSc	n.a	-	-	-	+
F421C/wt	M	39	dSSc	-	n.a	-	+	-
G334S/wt	F	12	SSc early	-	-	-	+	-
R4H/wt	F	56	lSSc	-	+	-	+	-

lSSc = forma limitata; dSSc = forma diffusa; SSc early = forma precoce della malattia, che non soddisfa tutti i criteri; SSc ss = sine scleroderma (è considerate una forma limitata, non si ha coinvolgimento cutaneo); PAH = ipertensione delle arterie polmonari; ACA= anticorpi anti-centromero; anti-topo I = anticorpi anti-topoisomerasi I; anti-RNApIII = anticorpi anti-RNA polimerasi III; n.a. = informazione clinica non disponibile.

Tab.R12 Predizione in silico dell'effetto delle variazioni missense di *PRF1* sulla struttura e sulla funzione della proteina effettuata utilizzando quattro programmi: Polyphen, SIFT, SNAP, e PMUT.

Cambiamento aminoacidico	Polyphen		SIFT		SNAP		PMUT	
	predizione	score[a]	predizione	score[b]	predizione	Precisione attesa[c]	predizione	score[d]
Arg4His	benigna	none	-	-	neutrale	60%	neutrale	0,2352
Arg123His	benigna	0,103	**non tollerata**	0,03	**non neutrale**	78%	neutrale	0,3133
Val135Met	benigna	1,380	**non tollerata**	0,02	**non neutrale**	82%	neutrale	0,2700
Arg232His	benigna	0,012	**non tollerata**	0,02	neutrale	69%	**patologica**	0,5239
Val329Ile	benigna	0,148	tollerata	0,44	neutrale	78%	neutrale	0,0468
Gly334Ser	benigna	1,378	tollerata	0,74	neutrale	92%	neutrale	0,3109
Phe421Cys	benigna	0,078	tollerata	0,07	**non neutrale**	63%	**patologica**	0,8781
Ala437Val	benigna	1,487	**non tollerata**	0,02	**non neutrale**	78%	neutrale	0,4791
His514Arg	**possibile danno**	1,518	**non tollerata**	0,00	neutrale	53%	**patologica**	0,7938

[a] = minore è il punteggio più benigna sarà la sostituzione (PolyPhen)
[b] = maggiore è il punteggio più il cambiamento sarà tollerato (SIFT)
[c] = maggiore è la percentuale maggiore sarà l'affidabilità della previsione (SNAP)
[d] = minore sarà il punteggio più la sostituzione sarà considerata neutrale (pMUT)

Tab. R13 Attività NK, espressione di perforina, e proporzione di cellule NK nel sangue periferico dei due pazienti LES portatori delle variazioni p.His514Arg e p.Val135Met.

Variazione	individuo	Attività NK* (rapporto effettore/target)			espressione di PRF1†		% NK in sangue periferico	
		100:1	30:1	10:1	%	MFI-R	$CD3^-$ $CD16^+$	$CD3^-$ $CD56^+$
H514R	**Pt.1**	54	34	19	16§	14	8	9
V135M	**Pt.2**	34	20	11	21	19	4§	12
-	**Ctr.1**	61	41	27	23	12	26	10
-	**Ctr.2**	52	39	24	32	16	31	11
	CONTROLLI¶	43 (15-62)	27 (9-43)	15 (4-28)	23 (17-29)	9 (5-20)	11 (6-31)	16 (7-27)

*: L'attività NK è espresso come % di lisi specifica.
†: L'espressione di perforina è espresso come percentuale di cellule positive e MFI-R (mean fluorescenza intensity ratio).
§: <5th percentile rispetto a 15 controlli.
¶: media (5-95th percentile) su 15 controlli.

Tab.R14 Frequenze alleliche e genotipiche delle varianti p.Ala91Val (A) e p.Asn252Ser (B) nei pazienti e nei controlli.

A	**CONTROLLI** n=1856	**LES** n=597			**SSc** n=204		
genotipo	**n(frequenza)**	**n(frequenza)**	**p**	**OR (95%CI)**	**n(frequenza)**	**p**	**OR (95%CI)**
CC	1652 (0.890)	524 (0.877)	0.45	0.89 (0.66-1.19)	168 (0.823)	**0.0070**	**0.58 (0.38-0.87)**
CT	196 (0.106)	69 (0.116)	0.52	1.11 (0.81-1.49)	34 (0.167)	**0.012**	**1.69 (1.12-2.56)**
TT	8 (0.004)	4 (0.007)	0.50	1.56 (0.39-5.71)	2 (0.010)	0.26	2.29 (inaccurato)
allele	**frequenza**	**frequenza**	**p**	**OR (95%CI)**	**frequenza**	**p**	**OR (95%CI)**
C	0.943	0.936			0.907		
T	0.057	0.064	0.38	1.14 (0.86-1.50)	0.093	**0.0054**	**1.70 (1.16-2.47)**

B	**CONTROLLI** n=1856	**LES** n=597			**SSc** n=204		
genotipo	**n(frequenza)**	**n(frequenza)**	**p**	**OR (95%CI)**	**n(frequenza)**	**p**	**OR (95%CI)**
AA	1840 (0.991)	585 (0.980)	**0.038**	**0.42 (0.19-0.96)**	200 (0.980)	0.13	0.43 (0.13-1.55)
AG	16 (0.009)	12 (0.020)	**0.038**	**2.36 (1.04-5.30)**	4 (0.020)	0.13	2.30 (0.64-7.41)
GG	0	0			0		
allele	**frequenza**	**frequenza**	**p**	**OR (95%CI)**	**frequenza**	**p**	**OR (95%CI)**
A	0.996	0.990			0.990		
G	0.004	0.010	**0.039**	**2.35 (1.04-5.24)**	0.010	0.13	2.29 (0.64-7.32)

La distribuzione complessiva delle frequenze genotipiche della variante p.Ala91Val è significativamente diversa tra controlli e casi di SSc (p=0.017), ma non tra controlli e LES (p=0.60).
Gli OR significativi sono evidenziati in grassetto.

TREX1

L'unico esone codificante del gene *TREX1* (lungo 945 bp) è stato analizzato per la ricerca di nuove variazioni su 210 pazienti affetti da LES, 58 affetti da Sindrome di Sjogren e 150 Sclerodermie. Sono state osservate 7 diverse sostituzioni di una singola base, tutte situate all'interno della sequenza codificante.

Due di queste varianti sono non sinonime (c.592G>A e c.694A>G) e causano rispettivamente le variazioni aminoacidiche p.Glu198Lys e p.Met232Val. La numerazione fa riferimento alla sequenza pubblicata sul database del sito del National Center for Biotechnology Information (NCBI) (NM_022517.17), ATG=1. Le due mutazioni sono state osservate ciascuna una sola volta, in eterozigosi, rispettivamente in un paziente con SS (p.Glu198Lys) e in un individuo affetto dalla forma limitata di SSc (p.Met232Val). Il paziente con SSc presenta ipertensione polmonare e la presenza di anticorpi anti-centromero. La variante p.Glu198Lys è già stata descritta in omozigosi nin un paziente con la Sindrome di Aicardi-Goutieres (AGS) (Ramantani et al. 2010), mentre p.Met232Val è nuova.

Entrambe le variazioni coinvolgono un residuo amminoacidico conservato tra le specie (analisi effettuata tramite il programma CLUSTAL W). La variazione p.Glu198Lys colpisce un amminoacido all'interno della sequenza conservata tra le diverse specie Exo3 che costituisce, assieme ad Exo 1 e Exo2, il sito catalitico del'enzima. La variazione p.Met232Val coinvolge invece un amminoacido che cade nella porzione proteica compresa tra Exo3 ed il dominio trans-membrana TMD/TMH. Un'analisi in silico effettuata con 4 diversi algoritmi (PolyPhen, SIFT, SNAP e pMUT) predice che entrambe le sostituzioni aminoacidiche di TREX1 (p.Glu198Lys e p.Met232Val) avranno un effetto dannoso sulla funzionalità della proteina, sia per le diverse caratteristiche chimiche degli amminoacidi della sequenza normale e mutata, sia per la conservazione del residuo amminoacidico normalmente tradotto (tabella R15).

Sono stati analizzati, al DHPLC, 200 controlli per verificare l'eventuale presenza delle 2 varianti non sinonime riscontrate nei pazienti. Non abbiamo riscontrato la presenza di tali variazioni in nessun controllo della casistica analizzata. Inoltre anche la presenza di queste 2 variazioni non risulta nei 1912 controlli sequenziati da Lee-Kirsch et al, 2007, suggerendo pertanto che non si tratta di polimorfismi rari.

Le altre 5 variazioni osservate (c.144C>G, c.198G>A, c.462T>C, c.531C>T e c.912G>A) sono sinonimie. Solo una di queste (c.144 C>: p.Pro48Pro) è stata osservata per la prima volta in questo studio, in un paziente affetto da LES. Le altre 4 sono invece già state descritte in letteratura in controlli sani (Lee-Kirsch, et al, 2007) e riportate su banche dati pubbliche. La variante c.531C>T -

rs11797 (p.Tyr177Tyr) è un polimorfismo comune del gene *TREX1*. La frequenza allelica osservata nelle tre casistiche (0,42-0,47) non è significativamente diversa da quella riportata in letteratura (0.42, Lee-Kirsh 2007) o sulle banche dati pubbliche (0.38, NCBI). Le altre quattro variazioni sono invece rare: nella nostra casistica la variante c.912G>A (p.Leu304Leu) è stata osservata in eterozigosi in due pazienti affetti rispettivamente da SS e da LES; mentre le variazioni c.198G>A (p.Lys66Lys) e c.462T>C (p.Asp154Asp) sono invece state osservate in eterozigosi nello stesso paziente, affetto da SS e c.144 C>G in paziente LES eterozigote.

La possibile rilevanza funzionale delle variazioni sinonime è stata valutata utilizzando i programmi SpliceView ed ESEfinder. Nessuna delle variazioni osservate sembra modificare i siti di splicing canonici o introdurne nuovi. Tuttavia la variante c.144C>G introduce nuove sequenze di interazione per le proteine SF2/ASF, SF2/ASF(IgM BRCA1) e SR35; la variazione c.592G>A rimuove siti di interazione già esistenti per SF2/ASF e SF2/ASF(IgM BRCA1) e la variante c.912G>A rimuove un sito di legame per SF2/ASF(IgM BRCA1). Inoltre è stata effettuata un'analisi per valutare se le sostituzioni nucleotidiche potessero determinare l'utilizzo di codoni più o meno frequenti rispetto a quelli normalmente usati per la codifica dell'amminoacido. Le differenze più evidenti riguardano la variazione c.144C>G la cui frequenza passa rispettivamente da 19,8‰ a 6,9‰ e la variazione c.912G>A la cui frequenza passa rispettivamente da 39.6‰ a 7.2‰, facendo ipotizzare un possibile effetto dannoso a carico della velocità di trascrizione della proteina.

Tabella R15: risultati dell'analisi in silico effettuata con l'utilizzo dei programmi PolyPhen, SIFT, SNAP e pMUT per le mutazioni non-sinonime Glu198Lys e Met232Val

	PolyPhen		**SIFT**		**SNAP**		**pMUT**	
Variazione	predizione	score[a]	predizione e	score[b]	predizione	Precisione attesa[c]	predizione	score[c]
Glu198Lys	Possibile danno	1.61	Non tollerato	0.00	Patologica	82%	non-neutrale	0.54
Met232Val	Probabile danno	2.27	Non tollerato	0.00	Patologica	78%	non-neutrale	0.58

[a] = minore è il punteggio più benigna sarà la sostituzione (PolyPhen)

[b] = maggiore è il punteggio più il cambiamento sarà tollerato (SIFT)

[c] = maggiore è la percentuale maggiore sarà l'affidabilità della previsione (SNAP)

[d] = minore sarà il punteggio più la sostituzione sarà considerata neutrale (pMUT)

Creazione di un algoritmo genetico di rischio

DATI NON ANCORA PUBBLICATI

Selezione dei marcatori per il calcolo del wGRS.

Allo scopo di definire un algoritmo che permetta di calcolare l'indice di rischio (weighted genetic risk score - wGRS) per il LES abbiamo selezionato in base ai dati presenti in letteratura 52 marcatori di tipo SNPs situati in 30 geni, distribuiti su tutto il genoma (vedi tabella R16). Basandoci sui recenti GWAS (dati riassunti da Graham et al. 2009), o su studi di tipo caso-controllo condotti su grosse casistiche, abbiamo scelto per questa analisi i loci che sono risultati più fortemente associati nella popolazione caucasoide, o la cui associazione è stata confermata in più studi indipendenti. Per ogni gene abbiamo selezionato il marcatore più fortemente associato con la suscettibilità al LES. Qualora per un gene siano stati riportati in letteratura due o più marcatori associati con la malattia in modo indipendente l'uno dall'altro (come nel caso del gene PTPN22) entrambi i marcatori sono stati inclusi.

In questa fase, oltre a selezionare i marcatori da includere nel progetto in maniera oculata, è stato importante assegnare ad ogni marcatore un valore di OR preciso, che verrà utilizzato per il calcolo del wGRS. La tabella R16 riporta i valori di OR e di p per ciascun polimorfismo, ed il numero totale (comprendente cioè tutti gli studi che hanno analizzato un dato SNP) di pazienti e controlli analizzati. Quando un marcatore è stato analizzato in più casistiche i valori di OR e di p sono stati ricalcolati sulla casistica complessiva, riferiti all'allele di rischio.

Sette dei geni selezionati (quelli marcati con un asterisco *) sono stati analizzati in studi di associazione condotti da gruppi di collaborazione internazionali a cui ha contribuito anche il nostro laboratorio, su ampie casistiche includenti campioni di pazienti e di controlli in nostro possesso. I dati derivanti da questi studi sono stati già pubblicati su riviste internazionali (Orru' 2009, Delgado-Vega 2009, Kozyrev 2008, Sanchez 2009, Liu 2009). Poiché questo progetto è in una fase iniziale, al momento per questi loci stiamo ottenendo da ogni centro i dati di genotipizzazione dei pazienti e dei controlli italiani. In futuro gli altri marcatori saranno tipizzati nel nostro laboratorio.

Studio di associazione su marcatori noti di suscettibilità al LES.

Al momento disponiamo di dati di genotipizzazione sulla popolazione italiana su 8 dei geni inclusi nella tabella R14: *TNFSF4* (per il quale sono stati tipizzati 6 SNP su 223 controlli e 259 pazienti affetti da LES), *IRF5* e *STAT4* (tipizzati per 4 marcatori ciascuno sulla stessa casistica formata da 264 controlli e 264 LES), *PXK, ATG5, BLK, SCUBE1* e *HLA-DRB1* (analizzati per 1268 casi e 1597 controlli, dei quali rispettivamente 278 e 333 di origine italiana). I dati di associazione

osservati nella nostra popolazione per i geni *IRF5*, *STAT4* e *TNFSF4* sono riportati nelle tabelle R17, R18 e R19. I dati di associazione relativi agli altri loci sono ancora confidenziali.
Osserviamo una forte associazione con i polimorfismi di *IRF5* (rs2070197: $p=7.69 \times 10^{-5}$ OR=2.33; indel: p=0.00063 OR=1.54) e *STAT4* (rs7574865: $p=4.049 \times 10^{-12}$ OR=2.77; rs3821236 $p=3.629 \times 10^{-7}$ OR=2.21; rs3024866: p=0.00036 OR=1.66). Osserviamo inoltre un'associazione più debole a carico del gene *TNFSF4* (SNP maggiormente associato: rs844644: p=0.00057 OR=0.64).

Creazione di un algoritmo di rischio

Abbiamo utilizzato i primi dati ottenuti per creare un algoritmo che permette di calcolare un indice di rischio (wGRS) che combina gli OR pesati di ciascuno dei loci di suscettibilità al LES. In questa prima analisi, abbiamo incluso nell'algoritmo 15 marcatori (tabella R20). Nove di questi sono SNP la cui associazione con la malattia era stata precedentemente osservata in letteratura, e che erano stati precedentemente selezionati per questo scopo (sono inclusi nella tabella R16). Questi marcatori (indicati con L in tabella R20) sono stati inseriti nell'algoritmo indipendentemente dal fatto che l'associazione riportata in precedenza sia significativa anche sulla nostra casistica. Il valore di OR riportato in tabella e che è stato utilizzato per il calcolo del wGRS è quello osservato negli studi originali da cui è emersa l'associazione, ed è stato ottenuto dai dati presenti in letteratura. Gli altri sei polimorfismi (indicati con N in tabella R20) non erano stati originariamente selezionati, ma sono stati inclusi nell'analisi perché hanno evidenziato nella nostra popolazione un'associazione fortemente significativa ($p<10^{-4}$) con la malattia. In questo caso i valori di OR utilizzati sono quelli osservati nella nostra casistica. Tutti i valori di OR riportati sono riferiti all'allele di rischio.
I pazienti ed i controlli al momento inclusi in questa elaborazione sono quelli per cui è disponibile un dato di genotipizzazione per tutti e 15 i marcatori utilizzati (171 casi e 164 controlli).
Per il calcolo dell'algoritmo abbiamo utilizzato la stessa formula messa a punto da De Jager et al (2009) sulla SM. Per ciascun campione (LES o controllo) abbiamo calcolato il numero di alleli di rischio per ogni marcatore (0, 1 oppure 2), e tale valore è stato moltiplicato per il "peso" del marcatore (logaritmo naturale dell'OR). Infine il wGRS è stato calcolato come sommatoria dei valori pesati di tutti e 15 gli SNP, secondo la seguente formula:

$$wGRS= \sum w_i X_i$$

dove i è il polimorfismo, w è il "peso" del polimorfismo (ln OR) e X_i è il numero di alleli di rischio.

Risk Score: risultati preliminari

Nella figura R7 i valori di wGRS così calcolati (in ascissa) sono plottati in rapporto alla percentuale di individui (in ordinata). Nei LES il wGRS medio è 4,62 ± 1,40, nei controlli wGRS = 3,77 ± 1,30.

Per ottenere una misura dell'accuratezza del test abbiamo misurato l'area sotto alla curva ROC (Reciever Operating Characteristic" (figura R8). La curva viene calcolata plottando per ogni valore di wGRS la sensibilità in ordinata e (100 – specificità) sull'asse delle ascisse. L'analisi è stata effettuata col software MedCalc. Come si può osservare l'area sottesa dalla curva (AUC) per per il wGRS calcolato sui soli dati genetici (AUC = 0,673) è significativamente maggiore ($p<0,0001$) rispetto a quella sottesa dalla diagonale (AUC = 0,5).

Per aumentare l'accuratezza della predizione abbiamo provato ad aggiungere all'algoritmo fattori demografici come il sesso, clinici come la presenza/assenza del fenomeno di Raynaud, delle lesioni discoidi o delle ulcere orali, e sierologici (presenza/assenza di anticorpi antiRNP/Sm o del fattore reumatoide). I valori di OR per i parametri clinici e immunologici sono stati calcolati provvisoriamente sulla nostra casistica sulla base dei dati clinici disponibili, assumendone l'assenza negli individui sani. Il valore di OR determinato dal sesso è stato calcolato sulla popolazione italiana (60.000.000) considerando la prevalenza del LES in Italia (0,071%) e il rapporto di incidenza femmine: maschi (9:1).

Effettivamente l'aggiunta del sesso e di alcuni parametri clinici (presenza del fenomeno di Raynaud e di anticorpi antiRNP/Sm) aumenta l'accuratezza dell'analisi, come abbiamo stabilito dal confronto delle curve ROC (figure R9 e R10). Il risultato migliore è stato ottenuto considerando un algoritmo di rischio che tenesse conto della suscettibilità genetica, del sesso e dalla presenza del fenomeno di Raynaud (AUC = 0,892).

Il sesso e il fenomeno di Raynaud, conferiscono un OR molto superiore rispetto a quello conferito da un singolo marcatore genetico (OR:1,19-2,33). Abbiamo quindi effettuato un confronto delle AUC per capire se la componente genetica cumulativa abbia un ruolo significativo nel determinare il rischio calcolato o se questo risultato sia influenzato prevalentemente dai due fattori clinico-demografici. Come si può osservare (figure R9 e R10) l'area sottesa dalla curva ROC comprendente sesso e fattori genetici è significativamente maggiore sia rispetto alla curva calcolata sulla base dei soli fattori genetici (differenza tra le AUC= 0,0815; $p<0,0001$) sia rispetto a quella determinata da un'ipotetica "curva" con soli due punti generata considerando un ipotetico algoritmo di rischio contenente una sola variabile: il sesso (differenza tra le AUC= 0,0788; $p=0,0009$). Abbiamo osservato risultati simili anche considerando la presenza/assenza di anticorpi antiRNP/Sm. Dal confronto delle curve si può evincere come sia la componente genetica cumulativa sia le variabili clinico-demografiche contribuiscano in misura significativa a determinare il risultato dell'algoritmo di rischio. Gli altri parametri clinico-sierologici considerati (lesioni discoidi, ulcere orali, fattore reumatoide) non aggiungevano molta accuratezza all'analisi, per cui non sono state ulteriormente considerate.

Tabella R16: Riassunto dei dati di associazione già riportati nei geni da analizzare

Gene	crom.	variante	posizione	OR	95% CI	p	Numero di individui analizzati		ref.
							LES	controlli	
***HLA-DR**	6p	**DR2**	**DRB1*1501**	1,87	1,51-2,32	$<10^{-7}$	356	975	Hartung 1992
Human Leukocyte Antigen		**DR3**	**DRB1*0301**	2,54	2,04-3,17	$<10^{-7}$	356	975	
***PTPN22**	1p	**rs2476601**	**R620W**	1,39	1.25-1.54	$<10^{-7}$	4234	9983	Lee 2007, Harley 2008
Protein Tyrosine Phosphatase Non-receptor 22		**rs33996649**	**R263Q**	1,57	1,16-2,12	0,0017	2093	2348	Orru' 2009
CRP									
Pentraxin C-reactive protein	1q	**rs3093061**	**promotore -70'7**	1,71	1,42-2,05	$<1X10^{-7}$	1118	1032	Edberg 2008
FCGR2A									
Fc fragment of IgG, low affinity IIa, receptor	1q	**rs1801274**	**H166R**	1,35	1,20-1,52	$6,78X10^{-7}$	720	2337	Harley 2008
SELP									
selectin P	1q	**rs3917815**	**N673S**			$5,74X10^{-6}$	787 famiglie		Jacob 2007
***TNFSF4 – OX40L**	1q	rs10489265	nella regione	1,3	1,19-1,43	$1,7X10^{-8}$			Graham 2009
Tumor necrosis factor (ligand) superfamily, member 4		rs1234317	nella regione	1,39		0,0009	1312	1801	Delgado-Vega 2009
		rs2205960	**nella regione**	1,22	1,15-1,30	$6,3X10^{-9}$	1963	4329	Gateva 2009
NMNAT2									
Nicotinamide nucleotide adenylyltransferase 2	1q	**rs2022013**	**introne 1**	1,18	1,10-1,27	$1,08X10^{-7}$	2566	4162	Harley2008
***IL10**	1q	**rs3024505**	**nella regione**	1,19	1,11-1,28	$4,0X10^{-8}$	1963	4329	Gateva 2009
Interleukin 10		IL10.G	promotore	1,78	1,19-2,66	0.042	172	164	D'Alfonso 2000
TLR5	1q		**R392X**			0,009	199	326genitori	Hawn 2005
Toll-like receptor 5								75 parenti	
***STAT4**									
Signal transducer and activator of transcription 4	2q	**rs7574865**	**introne 3**	1,55	1,47-1,65	$<10^{-7}$	4908	7101	Remmers 2007, Harley 2008
PDCD1									
Programmed cell death 1	2q	**PD1.3**	**introne 4**	RR=2,6	1,6-4,4	0,00001	880	356	Prokunina 2002
PXK									
PX domain containing serine/threonine kinase	3p	**rs6445975**	**introne 4**	1,27	1,15-1,39	$9,2X10^{-7}$	2566	4162	Harley2008
***BANK1**	4q	rs10516487	R61H	1,38	1,25-1,53	$3,74X10^{-10}$	2003	1968	Kozyrev 2008
B-cell scaffold protein with ankyrin repeats 1		**rs17266594**	**sito di spicing**	1,42	1,28-1,58	$4,74X10^{-11}$	1856	1774	
		rs3733197	A383T	1,23	1,11-1,36	$4,67X10^{-5}$	1819	1875	
PTTG1									
pituitary tumor-transforming 1	5q	**rs2431697**	**nella regione**	1,23	1,14-1,32	$1,00X10^{-10}$	2566	4162	Harley2008
ITPR3									
inositol 1,4,5-triphosphate receptor, type 3	6p	**rs3748079**	**promotore -1197**	1,88	1,51-2,35	$1,78X10^{-8}$			Oishi 2008
ATG5	6q	**rs573775**	**introne 6**	1,19	1,12-1,27	$1,36X10^{-7}$	2566	4162	Harley 2008
Autophagy protein 5		rs2245214	introne 2	1,15	1,09-1,21	$1,2X10^{-5}$	1963	4329	Gateva 2009
TNFAIP3	6q	**rs5029939**	**introne 2**	2,29		$2,89X10^{-12}$	431	2155	Graham 2008
Tumor necrosis factor, alpha-induced protein 3		rs13192841	nella regione	1,4	1,2-1,6	$5,4X10^{-8}$	1239	1629	Musone 2008
		rs2230926	F127C	2	1,4-3,0	$3,0X10^{-4}$	1239	1629	
		rs6922466	nella regione	1,3	1,1-1,4	$1,0X10^{-4}$	1239	1629	
ICA1									
Islet cell autoantigen 1	7p	**rs10156091**	**introne 8**	1,32	1,19-1,47	$1,9X10^{-7}$	2566	4162	Harley2008

IRF5 **Interferon regulatory factor 5**	7q	**rs2004640**	**sito di spicing**	1,5	1,39-1,61	$<10^{-7}$	2839	3262	Sigurdsson 2005, Graham
BLK	8p	rs13277113	promotore -2335	1,3	1,19-1,43	$1,7X10^{-8}$			
B lymphoid tyrosine kinase		**rs2248932**	**introne 1**	1,22	1,14-1,3	$7,00X10^{-10}$	2566	4162	Harley 2008
		rs10903340	nella regione	1,18	1,11-1,25	$1,46x10^{-7}$	2566	4162	
***LYN** **v-yes-1 Yamaguchi sarcoma viral related oncogene homolog**	8q	**rs7829816**	**introne 10**	1,22	1,11-1,33	$5,4X10^{-9}$	2566	4162	Harley 2008
		rs2667978	introne 1	1,23	1,13-1,35	$5,1X10^{-8}$	2566	4162	
MBL	10q	**rs1800450**	**G54D**	1,41	1,22-1,61	<0,001	1377	1504	Lee 2005
Mannose Binding Lectin		rs11003125	promotore -550	1,49	1,15-1,92	0,002	370	445	
		rs7096206	promotore -221	1,23	1,05-1,61	0,015	370	445	
PHRF1 **PHD and ring finger domains 1**	11p	**rs4963128**	**introne 4**	1,26	1,16-1,36	$3,00X10^{-10}$	2566	4162	Harley2008
***IL18** **Interleukin 18**	11q	**rs360719**	**promotore -1297**	1,37	1,21-1,54	$3,8X10^{-7}$	1358	1202	Sanchez 2009
ITGAM-ITGAX	16p	**rs1143679**	**R77H ITGAM**	1,74		$6,9X10^{-22}$			Nath 2008
integrin, alpha M - integrin, alpha X		rs11574637	introne 5 ITGAX	1,34	1,22-1,47	$<10^{-7}$	2104	4197	Hom 2008
		rs9888739	introne 14 ITGAM	1,7	1,51-1,92	$1,9X10^{-18}$	2566	4162	Harley 2008
TYK2 **Tyrosine Kinase 2**	19p	**rs2304256**	**V362F**	1,43	1,15-1,76	$5,60x10^{-5}$	589	377	Sigurdsson 2005
***callicreine**	19q	rs1054713	D135D KLK1	1,28	1,00-1,65	0,05	254	361	Liu 2009
		rs1897604	**introne 5 KLK5**	1,3	1,01-1,67	0,033	254	361	
UBE2L3 **Ubiquitin-conjugating enzyme E2L 3**	22q	**rs5754217**	**introne 1**	1,22	1,14-1,32	$7,53X10^{-8}$	2566	4162	Harley2008
SCUBE1 **signal peptide, CUB domain, EGF-like 1**	22q	**rs2071725**	**Val/Phe**	1,27	1,15-1,42	$1,21X10^{-7}$	2566	4162	Harley2008
IRAK1	Xq	rs2239673	introne 13						Jacob 2009
Interleukin-1 receptor associated kinase 1		**rs763737**	**introne 12**	1,19		$5,04X10^{-10}$	5337	5317	
		rs5945174	introne 10						
		rs70661789	introne 10						

Tab. R17 Associazione dei polimorfismi di *IRF5* (cromosoma 7) nella popolazione italiana

SNP	posizione (bp)	allele raro	freq_LES	freq_CT	CHISQ	P	OR	C.I. 95%	
rs10954213	128376587	G	0.3456	0.4065	3.726	0.05357	0.7708	0.5916	1.004
indel	1283651527		0.5324	0.4254	11.7	0.0006264	1.538	1.201	1.97
rs2004640	1283655307	G	0.4122	0.4791	4.144	0.04179	0.7627	0.5874	0.9902
rs2070197	1283761777	C	0.1522	0.07143	15.63	7.693e-005	2.333	1.519	3.584

Tab. R18 Associazione dei polimorfismi di *STAT4* (cromosoma 2) nella popolazione italiana

SNP	posizione (bp)	allele raro	freq_LES	freq_CT	CHISQ	P	OR	C.I. 95%	
rs1467199	191588747	G	0.2426	0.2313	0.1679	0.682	1.065	0.7893	1.436
rs3821236	191611003	A	0.3118	0.1704	25.88	3.629e-007	2.206	1.62	3.004
rs3024866	191631086	C	0.3596	0.2533	12.73	0.0003593	1.655	1.254	2.185
rs7574865	191672878	T	0.3926	0.1891	48.1	4.049e-012	2.771	2.067	3.715

Tab R19 Associazione dei polimorfismi di *TNFSF4* - *OX40L* (cromosoma 1) nella popolazione italiana

SNP	posizione (bp)	allele raro	freq_LES	freq_CT	CHISQ	P	OR	C.I. 95%	
rs1234314	171444015	G	0.4531	0.3879	4.156	0.04148	1.308	1.01	1.692
rs1234317	171454398	T	0.2961	0.2186	7.266	0.007025	1.503	1.117	2.024
rs844644	171476118	A	0.3934	0.5045	11.88	0.0005677	0.6369	0.4925	0.8236
rs12039904	171478896	T	0.2559	0.1816	7.625	0.005757	1.549	1.134	2.117
rs844648	171490486	A	0.4302	0.35	6.407	0.01137	1.402	1.079	1.823
rs844665	171515580	T	0.08696	0.05263	3.815	0.05078	1.714	0.9928	2.96

Tab. R20 SNPs utilizzati per l'algoritmo di rischio.

gene	SNP	N/L [c]	OR	weight (ln OR)	allele associato
STAT4	rs3821236	N	2,21	0,79	A
STAT4	rs7574865	L	1,55	0,44	T
IRF5	rs2004640	L	1,5	0,405	T
IRF5	rs2070197	N	2,33	0,845	C
OX40L	rs1234317	L	1,39	0,33	T
PXK	rs6445975	L	1,27	0,24	C
HLA-DRB1	HLA-DRB1_384plex_SNP1	N	1,94	0,66	T
HLA-DRB1	rs2187668 (DR3) [a]	L	1,76	0,565	T
ATG5	rs573775	L	1,19	0,17	T
BLK	rs13277113 [b]	L	1,3	0,26	A
BLK	rs2248932 [b]	L	1,22	0,20	T
SCUBE1	rs2071725	L	1,27	0,24	G
SCUBE1	SCUBE1_384plex_SNP10	N	1,68	0,52	C
SCUBE1	SCUBE1_384plex_SNP14	N	1,77	0,57	A
SCUBE1	SCUBE1_384plex_SNP21	N	1,82	0,60	G

[a] Il polimorfismo utilizzato è un tagger dell'aplotipo HLA DR3. L'associazione con questo SNP era già stata riportata da Graham et al. 2009

[b] I due marcatori non sono stati analizzati direttamente, ma è stato tipizzato un polimorfismo in LD elevato (D'=1, r2>0,9) con essi.

[c] In questa colonna è indicato se il polimorfismo era stato precedentemente selezionato da dati di letteratura (L) o se è stato incluso per via dell'associazione osservata nel presente studio (N).

Fig. R7 Grafico di distribuzione del wGRS nei pazienti e nei controlli

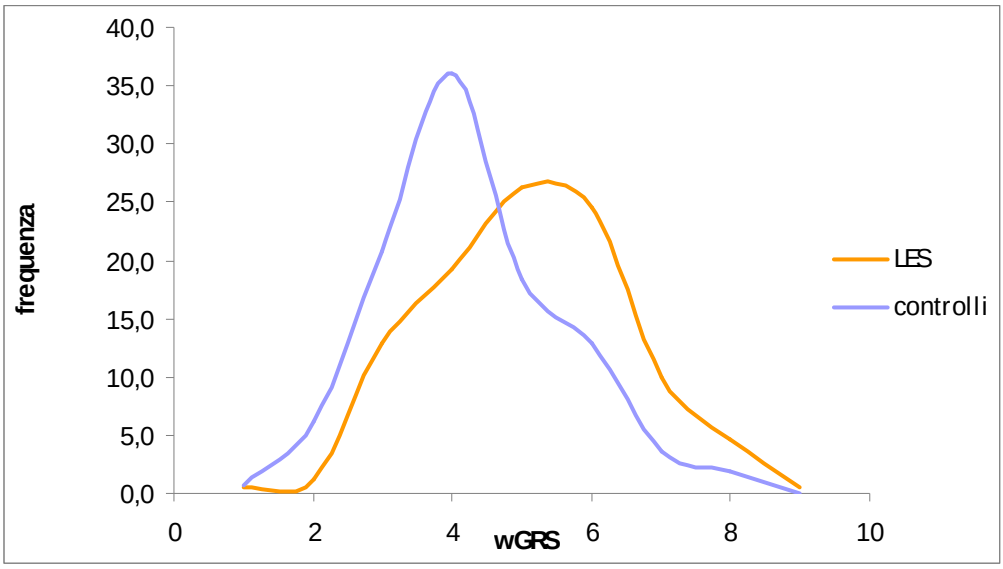


Fig. R8 Curva ROC per l'algoritmo wGRS calcolato con i soli dati genetici

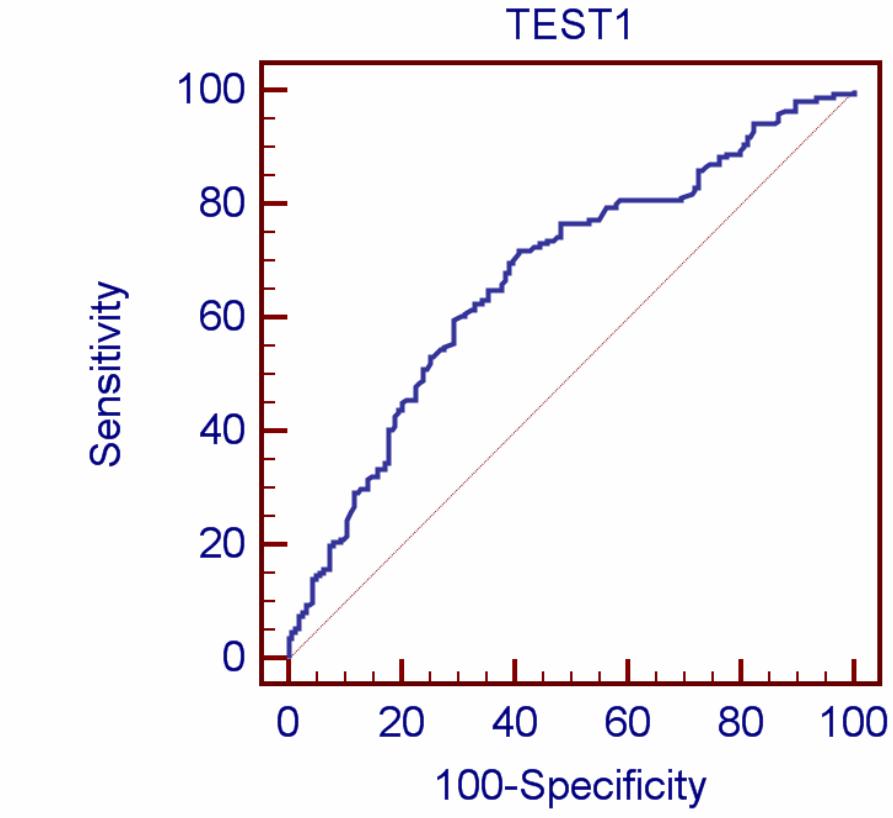


Fig R9 Confronto tra le curve ROC

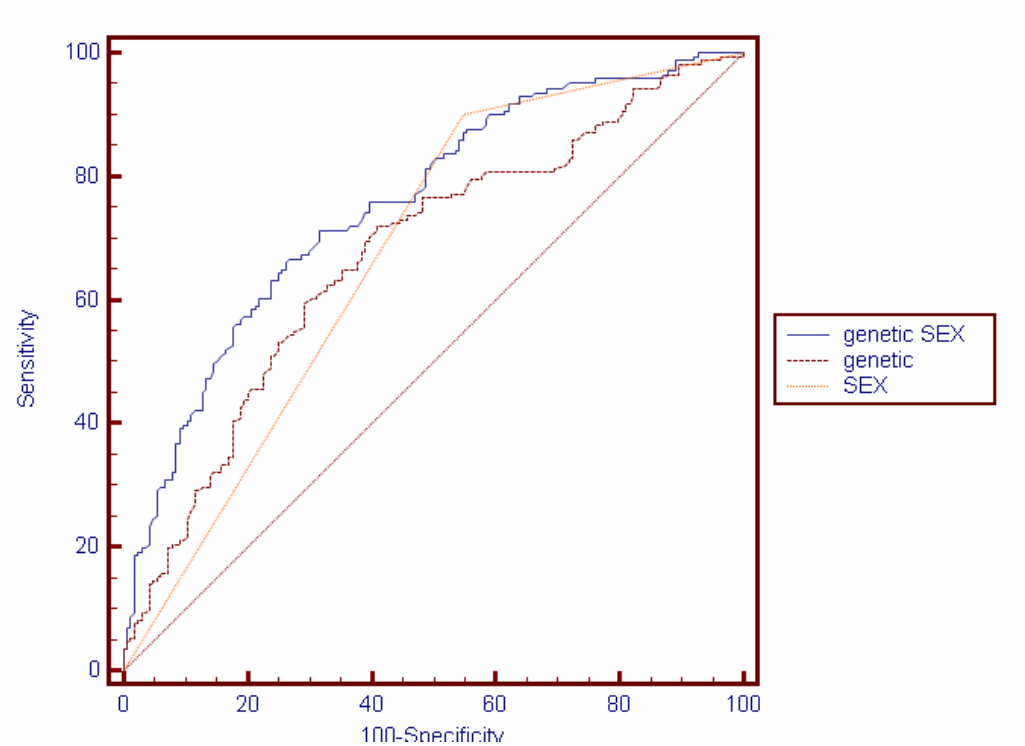
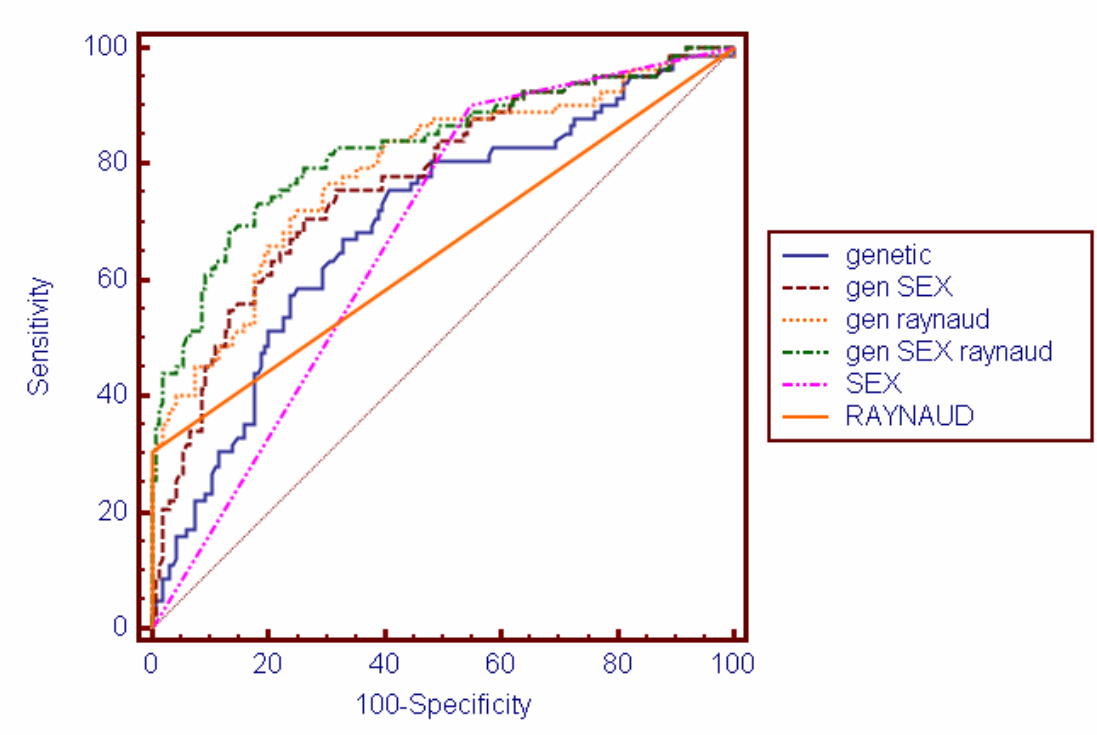


Fig R10 Confronto tra le curve ROC



Meta-analisi sui geni *TEC* e *BMPR1B*

DATI NON ANCORA PUBBLICATI

Nell'ambito di un progetto di collaborazione internazionale di follow up-fine mapping di un GWAS i cui dati non sono ancora stati pubblicati (da qui in poi denominato 100K GWAS) abbiamo analizzato l'associazione con il LES di due geni: *TEC* (tec protein tyrosine kinase) e *BMPR1B* (Bone morphogenetic protein receptor type IB precursor). Abbiamo analizzato i due geni prima su una casistica di 1268 casi e 1597 controlli analizzati nell'ambito dello studio di replicazione del 100K GWAS, poi sui dati derivanti dal 100K GWAS stesso. Successivamente abbiamo approfondito l'analisi effettuando l'imputazione dei polimorfismi non analizzati e una meta-analisi dei dati della replicazione con i dati derivanti due diversi studi GWAS: il 100K GWAS (condotto su 534 casi e 771 controlli) ed un secondo GWAS (da qui in poi denominato 300K GWAS) che è già risultato in una pubblicazione, nel 2008, a nome dell'International Consortium for Systemic Lupus Erythematosus (SLEGEN) e comprende 706 casi e 2313 controlli. Per questioni di riservatezza ai polimorfismi di questa analisi è stato assegnato un codice arbitrario, e non verrà riportata in questa sede né l'rsID né la posizione genomica. I tre studi sono stati condotti su pannelli diversi di polimorfismi, solo in minima parte sovrapponibili (vedi figura R11).
I risultati dell'analisi statistica sui marcatori analizzati nell'ambito dello studio di replicazione (15 in *TEC* e 28 in *BMPR1B*) sono riportati rispettivamente nelle tabelle R21 e R22.
Come si può osservare, le frequenze alleliche dei pazienti e dei controlli non si discostano in maniera fortemente significativa per nessuno dei polimorfismi esaminati. Tuttavia osserviamo, a carico del marcatore TEC_384plex_SNP11 situato nel primo introne del gene *TEC*, una debole associazione (p=0,0035) che, pur essendo lontana dalla soglia di significatività per la replicazione degli studi Genome-Wide, è comunque suggestiva di ulteriori analisi.

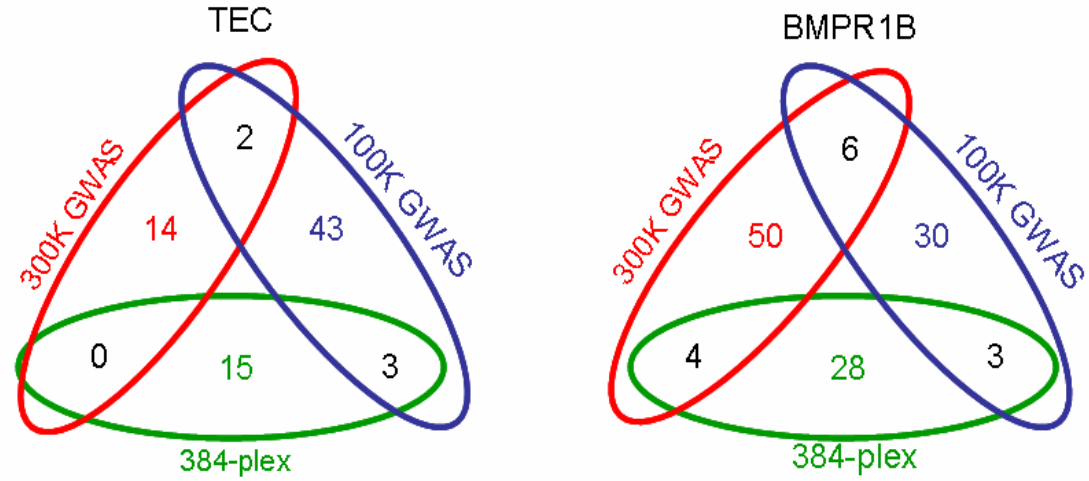
Imputazioni: impostazione del lavoro

Per ciascuno dei due geni sono state condotte 6 analisi di imputazione separate, una per ciascun pannello (300K GWAS controlli, 300K GWAS casi, 100K GWAS controlli, 100K GWAS casi, 384-plex controlli, 384-plex casi). E' stato utilizzato l'algoritmo IMPUTE.
Le imputazioni sono state condotte su una regione di 3Mb, compresa tra la base 46 000 000 e la base 49 000 000 del cromosoma 4 nel caso di *TEC* e su una regione di 2Mb, tra la base 95 000 000 e la base 97 000 000 del cromosoma 4 nel caso di *BMPR1B*. Inoltre sono state utilizzate due "buffer region", ciascuna di 250 Kb, a monte ed a valle della regione, i cui SNP sono stati utilizzati dall'algoritmo di imputazione per eseguire l'analisi, ma non compaiono nell'output finale. Come pannello di riferimento sono stati utilizzati i dati del progetto 1000 Genomi (112 aplotipi). Questo pannello è stato preferito a quello derivante dai dati di genotipizzazione di HapMap perché contiene un numero maggiore di polimorfismi.
Per quanto riguarda il gene *TEC*, sono stati utilizzati per l'analisi 11475 SNP, 9875 dei quali compaiono nel file dei risultati. Per il gene *BMPR1B* sono stati utilizzati 8134 SNP, e 6603 compaiono nel file di output.
I dati derivanti dalle imputazioni sono stati meta-analizzati con il software SNPTEST (vedi materiali e metodi), che tiene conto nell'analisi dell'incertezza del genotipo. Come soglia di probabilità per i genotipi (probability call) è stato utilizzato il valore comunemente usato in questo tipo di studi (0,9). Ciò significa che solo ai campioni che presentavano una probability call più alta

di 0,9 per uno dei tre possibili genotipi di un dato SNP è stato “assegnato” un genotipo per quello SNP.
Infine abbiamo condotto una selezione basata su criteri di qualità sull’output di SNPTEST.
Per i marcatori che sono stati analizzati in almeno uno dei tre pannelli caso-controllo (330K GWAS, 100K GWAS o 384-plex) abbiamo confrontato i dati “reali”, derivanti da genotipizzazione, con quelli imputati, e per ogni SNP abbiamo considerato i seguenti parametri:
- differenza tra le frequenze alleliche “reali” e quelle imputate <0,05
- differenza tra le frequenze genotipiche “reali” e quelle imputate <0,05
- differenza tra le frequenze alleliche “reali” e imputate / MAF (minor allele frequency) “reale” <0,1
- equilibrio di Hardy-Weinberg
- percentuale di campioni imputati (campioni con una probability call > 0,9 per un genotipo)

Per i marcatori che non sono stati analizzati in nessuna casistica abbiamo confrontato i risultati dell’imputazione ottenuti su ciascuna casistica con quelli ottenuti sulle altre due. Per ogni marcatore abbiamo considerato:
- differenza tra le frequenze alleliche <0,1
- differenza tra le frequenze genotipiche <0,1
- differenza tra le frequenze alleliche / MAF osservata su ciascuna casistica <0,25
- equilibrio di Hardy-Weinberg
- percentuale di campioni imputati

Fig. R11 SNP tipizzati nei tre studi



Numero di polimorfismi tipizzati e relazione tra i tre diversi studi. I numeri colorati all’interno degli insiemi rappresentano il numero complessivo di polimorfismi che sono stati tipizzati nello studio corrispondente per ciascuno dei due geni. I numeri neri all’incrocio di due insiemi rappresentano il numero di polimorfismi tipizzati in entrambi gli studi. Come si può osservare la sovrapposizione è minima, soprattutto per il gene *TEC*.

Risultati di imputazione e meta-analisi

In tal modo abbiamo selezionato 73 polimorfismi di *TEC* con un dato di imputazione che riteniamo sufficientemente affidabile, 56 dei quali non sono stati analizzati in nessuna delle tre casistiche, e 17 che sono stati tipizzati in almeno una di queste. Per il gene *BMPR1B* abbiamo selezionato 256 marcatori, 28 tipizzati in almeno una delle tre casistiche e 228 completamente imputati. I risultati sulla popolazione complessiva (2508 pazienti, 4681 controlli) dei marcatori che hanno superato la fase di controllo di qualità e che hanno evidenziato un p<0,01 sono contenuti nella tabella R23. Una rappresentazione grafica del risultato dell'analisi di imputazione su tutti gli SNP oggetto della meta-analisi (sia tipizzati che imputati) si può osservare nella figura R12.

Come si può vedere nessuno dei marcatori che hanno passato il controllo di qualità è fortemente associato al LES. Tuttavia uno dei polimorfismi di *TEC* derivanti dal 100K GWAS presenta un dato suggestivo (p= 0,00080; OR = 1,16). Inoltre un polimorfismo di *TEC* derivante dal progetto 384-plex (TEC_384-plex_SNP1) che nella casistica originaria aveva un valore di p=0,0277 ora presenta un dato leggermente più suggestivo(p=0,0079), anche se ancora lontano dalla soglia di replicazione per i GWAS. Purtroppo lo SNP che nella casistica 384-plex ha evidenziato il dato più interessante (TEC_384plex_SNP11) non ha passato i controlli di qualità per l'imputazione nelle altre due casistiche (300K GWAS e 100K GWAS).

Tab. R21 Associazione dei polimorfismi di *TEC* nello studio di replicazione

SNP	allele raro	freq_LES	freq_CT	CHISQ	P	OR	C.I. 95%	
TEC_384plex_SNP1	C	0.3224	0.3513	5.242	0.02205	0.8787	0.7866	0.9816
TEC_384plex_SNP2	A	0.3878	0.3655	2.999	0.08333	1.1	0.9875	1.225
TEC_384plex_SNP3	T	0.4578	0.4391	1.997	0.1576	1.079	0.9711	1.198
TEC_384plex_SNP4	A	0.3444	0.371	4.331	0.03743	0.8907	0.7986	0.9933
TEC_384plex_SNP5	G	0.4362	0.4245	0.7954	0.3725	1.049	0.9442	1.166
TEC_384plex_SNP6	A	0.4719	0.4752	0.06099	0.8049	0.9869	0.8888	1.096
TEC_384plex_SNP7	C	0.3871	0.3864	0.002799	0.9578	1.003	0.901	1.116
TEC_384plex_SNP8	T	0.3047	0.3001	0.1376	0.7107	1.022	0.9121	1.145
TEC_384plex_SNP9	C	0.4753	0.4924	1.627	0.2021	0.934	0.841	1.037
TEC_384plex_SNP10	A	0.3962	0.39	0.2245	0.6356	1.026	0.922	1.142
TEC_384plex_SNP11	T	0.1609	0.1905	8.5	0.003551	0.8146	0.7096	0.9351
TEC_384plex_SNP12	T	0.4244	0.4221	0.03149	0.8591	1.01	0.9084	1.122
TEC_384plex_SNP13	A	0.3931	0.3721	2.628	0.105	1.093	0.9816	1.217
TEC_384plex_SNP14	G	0.4728	0.4608	0.8144	0.3668	1.049	0.9452	1.165
TEC_384plex_SNP15	G	0.2099	0.2282	2.745	0.09754	0.8987	0.792	1.02

Tab. R22 Associazione dei polimorfismi di *BMPR1B* nello studio di replicazione

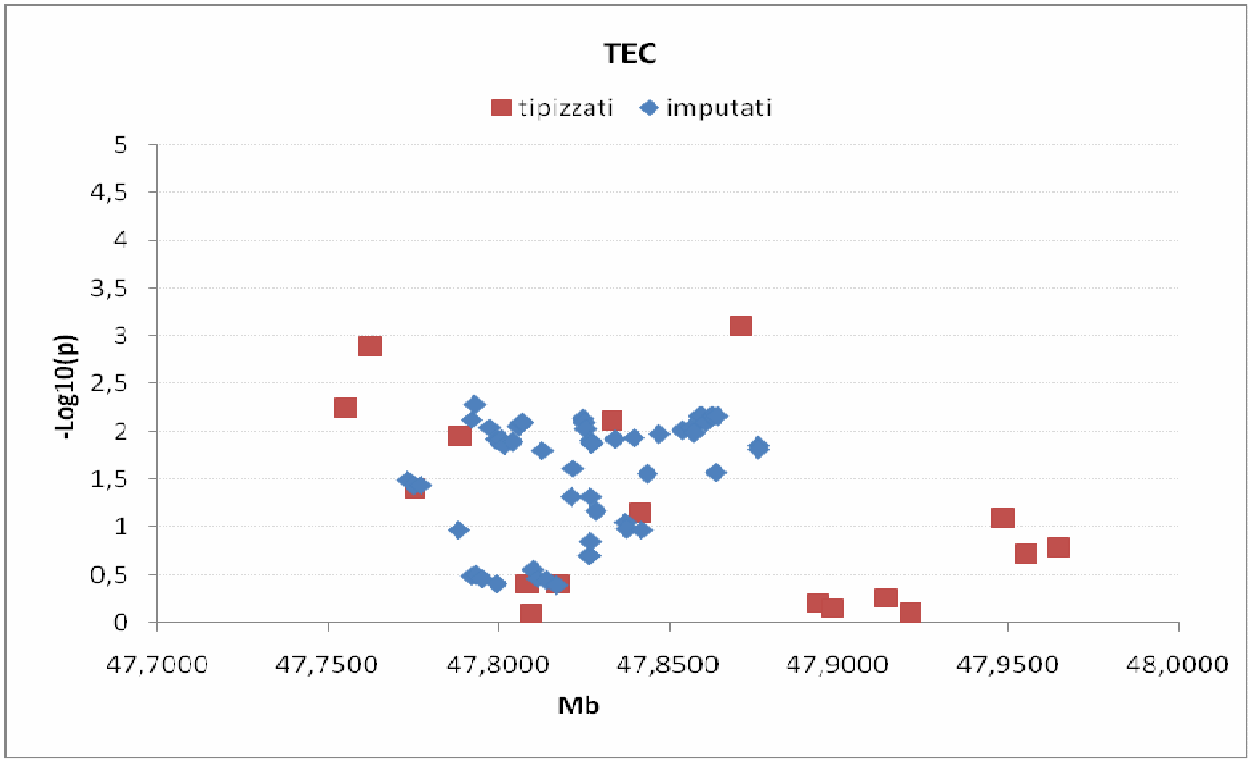
SNP	allele raro	freq_LES	freq_CT	CHISQ	P	OR	C.I. 95%	
BMPR1B_384plex_SNP1	A	0.4049	0.4027	0.02823	0.8666	1.009	0.9074	1.122
BMPR1B_384plex_SNP2	G	0.3941	0.402	0.3658	0.5453	0.9676	0.8695	1.077
BMPR1B_384plex_SNP3	C	0.3102	0.3055	0.144	0.7043	1.022	0.9129	1.144
BMPR1B_384plex_SNP4	T	0.1793	0.1901	1.087	0.2972	0.9309	0.8136	1.065
BMPR1B_384plex_SNP5	C	0.4431	0.443	2.406e-005	0.9961	1	0.9006	1.111
BMPR1B_384plex_SNP6	A	0.4787	0.4896	0.6779	0.4103	0.9571	0.8622	1.062
BMPR1B_384plex_SNP7	T	0.278	0.2639	1.414	0.2344	1.074	0.9549	1.207
BMPR1B_384plex_SNP8	A	0.1898	0.1813	0.6835	0.4084	1.058	0.9254	1.21
BMPR1B_384plex_SNP9	G	0.409	0.3975	0.7706	0.38	1.049	0.9429	1.167
BMPR1B_384plex_SNP10	T	0.3635	0.3451	2.082	0.1491	1.084	0.9716	1.209
BMPR1B_384plex_SNP11	A	0.1976	0.2137	2.259	0.1328	0.9056	0.7957	1.031
BMPR1B_384plex_SNP12	T	0.4326	0.4349	0.03242	0.8571	0.9904	0.8914	1.1
BMPR1B_384plex_SNP13	G	0.2725	0.2538	2.562	0.1095	1.102	0.9785	1.24
BMPR1B_384plex_SNP14	G	0.3424	0.3789	8.114	0.004392	0.8536	0.7655	0.9519
BMPR1B_384plex_SNP15	C	0.3828	0.3606	2.987	0.08396	1.1	0.9873	1.225
BMPR1B_384plex_SNP16	T	0.3316	0.3449	1.108	0.2926	0.9423	0.8436	1.053
BMPR1B_384plex_SNP17	G	0.1429	0.1328	1.213	0.2707	1.089	0.9359	1.266
BMPR1B_384plex_SNP18	G	0.2953	0.2718	3.815	0.0508	1.122	0.9996	1.26
BMPR1B_384plex_SNP19	T	0.4621	0.4872	3.551	0.05951	0.9044	0.8147	1.004
BMPR1B_384plex_SNP20	G	0.402	0.405	0.05415	0.816	0.9875	0.8878	1.098
BMPR1B_384plex_SNP21	G	0.4068	0.4059	0.004696	0.9454	1.004	0.9026	1.116
BMPR1B_384plex_SNP22	C	0.4057	0.408	0.03236	0.8572	0.9903	0.8905	1.101
BMPR1B_384plex_SNP23	A	0.4274	0.4375	0.5826	0.4453	0.9597	0.8634	1.067
BMPR1B_384plex_SNP24	G	0.267	0.2708	0.1073	0.7432	0.9805	0.8717	1.103
BMPR1B_384plex_SNP25	G	0.351	0.3505	0.00194	0.9649	1.002	0.8986	1.118
BMPR1B_384plex_SNP26	G	0.28	0.3038	3.859	0.04947	0.8912	0.7945	0.9998
BMPR1B_384plex_SNP27	A	0.4795	0.4972	1.769	0.1835	0.9317	0.8394	1.034
BMPR1B_384plex_SNP28	T	0.3951	0.3889	0.2322	0.6299	1.027	0.9226	1.142

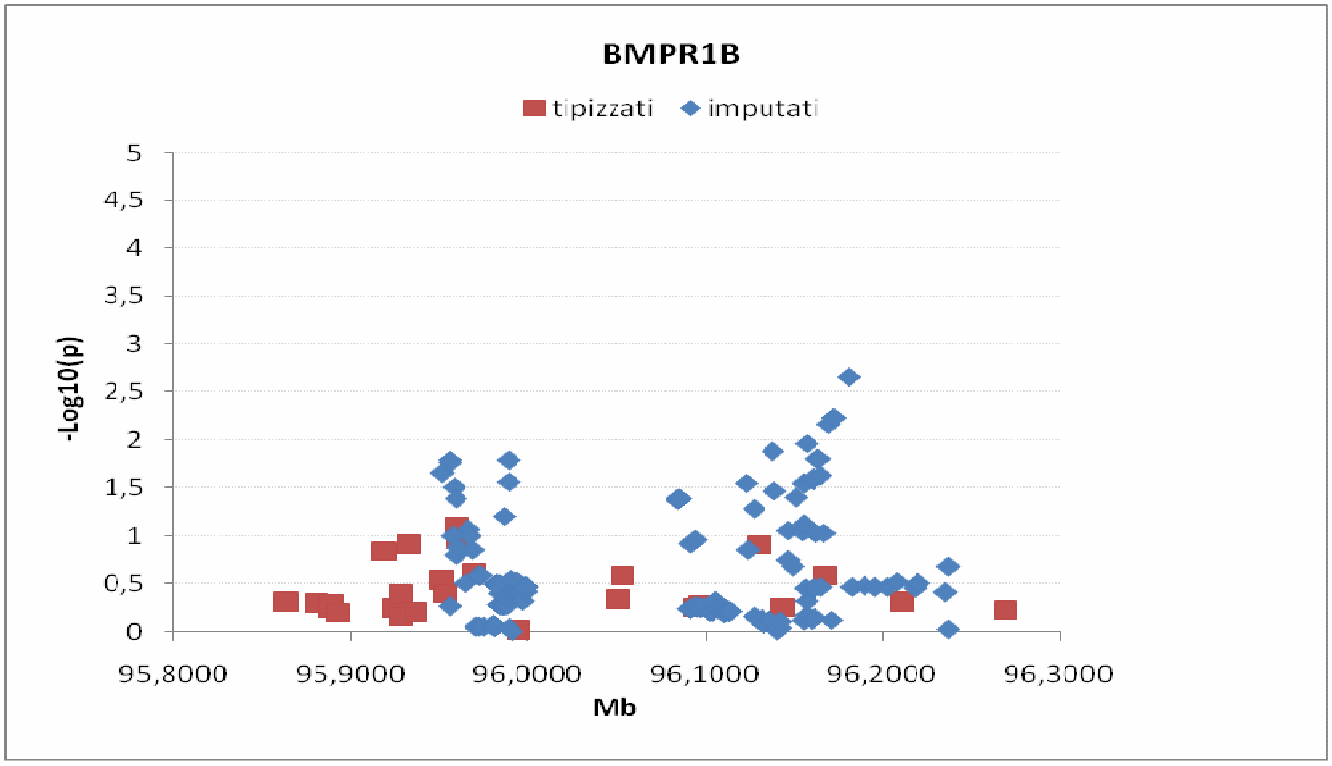
Tab.R23 Dati di associazione derivanti dall'analisi d'imputazione su polimorfismi con p<0,01

polimorfismo	allele_A	allele_B	freq_A controlli	freq_A casi	chi squared	P	OR	95% CI	
TEC_IMP_SNP1	T	A	0,6005	0,5707	7,13	0,007565	1,13	1,03	1,24
TEC_IMP_SNP2	T	C	0,6007	0,5695	7,78	0,005276	1,14	1,04	1,24
TEC_IMP_SNP3	C	T	0,6005	0,5712	6,82	0,009005	1,13	1,03	1,23
TEC_IMP_SNP4	C	T	0,6003	0,5710	6,84	0,008922	1,13	1,03	1,23
TEC_IMP_SNP5	A	T	0,2911	0,2680	6,99	0,008194	1,12	1,03	1,22
TEC_IMP_SNP6	A	T	0,2911	0,2680	6,99	0,008194	1,12	1,03	1,22
TEC_IMP_SNP7	T	C	0,6599	0,6858	7,18	0,007365	0,89	0,82	0,97
TEC_IMP_SNP8	A	T	0,6603	0,6858	7,00	0,008145	0,89	0,82	0,97
TEC_IMP_SNP9	T	C	0,6637	0,6889	6,74	0,009430	0,89	0,82	0,97
TEC_IMP_SNP10	C	T	0,6591	0,6845	6,68	0,009753	0,89	0,82	0,97
TEC_IMP_SNP11	C	G	0,6588	0,6844	6,82	0,009024	0,89	0,82	0,97
TEC_IMP_SNP12	A	G	0,6558	0,6823	7,28	0,006983	0,89	0,81	0,97
TEC_IMP_SNP13	C	T	0,6558	0,6823	7,28	0,006983	0,89	0,81	0,97
TEC_IMP_SNP14	C	T	0,6585	0,6844	7,00	0,008159	0,89	0,82	0,97
TEC_IMP_SNP15	C	G	0,6558	0,6823	7,28	0,006983	0,89	0,81	0,97
TEC_IMP_SNP16	C	T	0,6558	0,6823	7,28	0,006983	0,89	0,81	0,97
TEC_IMP_SNP17	C	T	0,6558	0,6823	7,28	0,006983	0,89	0,81	0,97
TEC GWAS SNP1	A	G	0,4681	0,4349	7,61	0,005815	0,87	0,79	0,96
TEC GWAS SNP2	T	C	0,4670	0,4283	10,32	0,001316	0,85	0,78	0,94
TEC_384plex_SNP1	T	C	0,6576	0,6830	7,06	0,007893	1,12	1,03	1,22
TEC GWAS SNP3	G	A	0,6089	0,6434	11,24	0,000800	1,16	1,06	1,26
BMPR1B_IMP_SNP1	A	G	0,6675	0,6437	7,32	0,006815	0,90	0,83	0,97
BMPR1B_IMP_SNP2	G	C	0,6679	0,6437	7,58	0,005891	0,90	0,83	0,97
BMPR1B_IMP_SNP3	A	G	0,6679	0,6437	7,58	0,005891	0,90	0,83	0,97
BMPR1B_IMP_SNP4	T	C	0,6679	0,6437	7,58	0,005891	0,90	0,83	0,97
BMPR1B_IMP_SNP5	C	T	0,6798	0,6529	9,39	0,002185	0,89	0,82	0,96

Le caselle relative a polimorfismi in elevato LD tra loro (r^2 >0,8) in base ai dati riportati su HapMap sono state colorate in modo uguale.

Fig R12 Grafici rappresentanti il risultato dell'analisi di associazione effettuata sugli SNP di *TEC* e *BMPR1B* nella meta-analisi.





Altri articoli

Oltre alla linea di ricerca principale sulla suscettibilità genetica al LES, durante il periodo del mio dottorato ho anche contribuito ad altri progetti, interni al nostro laboratorio o in collaborazione con altri, riguardanti la genetica di malattie autoimmuni diverse dal LES (Sclerosi Multipla, Diabete di tipo I, Sclerosi Sistemica). Da questa collaborazione sono derivati i 6 articoli su riviste internazionali qui sotto elencati (uno dei quali è ancora in fase di revisione da parte della rivista a cui è stato inviato). Inoltre sono già pronti due articoli contenenti i dati dell'analisi mutazionale su *TREX1* e *PRF1*, il primo dei quali (**Barizzone N** et al. Rare variants in the TREX1 gene and susceptibility to autoimmune diseases) è pronto per la pubblicazione, il secondo (**Barizzone N** et al. Genetic analysis of perforin in two autoimmune diseases: Systemic Lupus Erithematosus and Systemic Sclerosis) è in fase di preparazione. Entrambi i lavori, anche se ancora provvisori, sono allegati in fondo alla trattazione.

Articolo 4:
D'Alfonso S, Bolognesi E, Guerini FR, **Barizzone N**, Bocca S, Ferrante D, Castelli L, Bergamaschi L, Agliardi C, Ferrante P, Naldi P, Leone M, Caputo D, Ballerini C, Salvetti M, Galimberti D, Massacesi L, Trojano M, Momigliano-Richiardi P.
A sequence variation in the MOG gene is involved in multiple sclerosis susceptibility in Italy.
Genes Immun. 2008 9(1):7-15

Articolo 5:
Cappellano G, Orilieri E, Comi C, Chiocchetti A, Bocca S, Boggio E, Bernardone IS, Cometa A, Clementi R, **Barizzone N**, D'Alfonso S, Corrado L, Galimberti D, Scarpini E, Guerini FR, Caputo D, Paolicelli D, Trojano M, Figà-Talamanca L, Salvetti M, Perla F, Leone M, Monaco F, Dianzani U.
Variations of the perforin gene in patients with multiple sclerosis.
Genes Immun. 2008 9(5):438-44.

Articolo 6:
Bergamaschi L, Leone M, Fasano ME, Guerini FR, Ferrante D Bolognesi E, **Barizzone N**, Corrado L, Naldi P, Agliardi C, Dametto E, Salvetti M, Visconti A, Galimberti D, Scarpini E, Cavalla P, Bergamaschi R, Monaco F, Caputo D, Momigliano-Richiardi P, D'Alfonso S.
HLA class I markers and Multiple Sclerosis susceptibility in the Italian population.
Genes Immun. 2010; 11(2):173-80.

Articolo 7:
Chiocchetti A, Orilieri E, Cappellano G, **Barizzone N**, D' Alfonso S, D' Annunzio G, Lorini R, Ravazzolo R, Cadario F, Martinetti M, Calcaterra V, Cerutti F, Bruno G, Larizza D, Dianzani U.
The osteopontin gene +1239A/C single nucleotide polymorphism is associated with type 1 diabetes mellitus in the Italian population.
Int J Immunopathol Pharmacol. 2010 23(1):263-9.

Articolo 8:
Corrado L, Bergamaschi L, **Barizzone N**, Fasano ME, Guerini FR, Salvetti M, Galimberti D, Benedetti MD, Leone M, D'Alfonso S.
Association of the CBLB gene with multiple sclerosis: new evidence from a replication study in an Italian population.
J Med Genet. 2010. [Epub ahead of print]

Articolo 9:

Barizzone N, Marchini M, Cappiello F, Chiocchetti A, Orilieri E, Ferrante D, Corrado L, Mellone S, Scorza R, Dianzani U, D'Alfonso S.

Association of Osteopontin (OPN) regulatory polymorphisms with Systemic Sclerosis (SSc).

Sottomesso a: *Human Immunology*.

DISCUSSIONE

Analisi di fattori genetici di suscettibilità localizzati nella regione cromosomica 1q32-43

La porzione cromosomica 1q32-43 era risultata associata al LES in più di uno studio di linkage su tutto il genoma (Whole Genome Linkage Studies – WGLS) precedentemente condotti da altri gruppi di ricerca, per questo in passato abbiamo intrapreso uno studio di associazione di tipo fine-mapping su questa regione. L'associazione più evidente era stata osservata con l'elemento genico *CR1L* (complement component receptor 1-like), che mappa all'interno del cluster RCA. Infatti ben tre polimorfismi, tutti intronici, sono risultati significativamente associati alla malattia. Di conseguenza il gene *CR1L* era stato analizzato più approfonditamente, ricercando variazioni nucleotidiche nelle sue sequenze codificanti e regolatorie che erano stati testati per l'associazione con la suscettibilità al LES in una prima casistica appartenente alla popolazione italiana.

Durante il mio dottorato abbiamo esteso la casistica italiana di pazienti LES e di controlli per 6 dei polimorfismi precedentemente analizzati, selezionati perché avevano evidenziato un'associazione con il LES nella casistica pilota. Inoltre abbiamo tipizzato gli stessi SNPs su una casistica di pazienti e controlli appartenenti alla popolazione spagnola, ed abbiamo analizzato altri SNPs nelle regioni centromeriche e telomeriche a *CR1L*.

E' stata osservata un'associazione statisticamente significativa, nella popolazione italiana, con cinque polimorfismi in stretto LD tra loro mappanti nella porzione 3'-terminale del gene.

Tra questi il marcatore che sembra primariamente associato è una variazione nucleotidica nella porzione 3'UTR dell'isoforma lunga di *CR1L*. In particolare l'allele più raro di questo SNP sembra essere protettivo contro il rischio di sviluppare il LES. L'associazione osservata nella popolazione italiana non si è però riconfermata in quella spagnola. Ciò può essere dovuto a differenze tra le popolazioni esaminate, ad errori di campionamento, o al fatto che le variazioni esaminate non hanno un effetto causale ma sono semplicemente in LD con la variazione determinante (e a differenze di LD tra la popolazione spagnola e quella italiana). In futuro intendiamo testare questi polimorfismi su altre popolazioni non correlate con quella italiana.

Il dato di associazione riscontrato appare quantomeno curioso poiché non è ancora stato completamente chiarito se *CR1L* sia uno pseudogene oppure un gene funzionalmente espresso.

CR1L presenta oltre il 90% di omologia con CR1 (complement component receptor 1), una proteina regolatrice del complemento espressa da vari tipi cellulari ma principalmente sulla membrana degli eritrociti. E' un recettore multifunzionale, in grado di legare C4b, C3b, C3bi, C1q e MBL. Il suo ruolo principale è la rimozione degli immunocomplessi. Poiché la deposizione degli

immunocomplessi è una delle principali cause del danno tessutale nel LES, e poiché il deficit congenito o acquisito di componenti della via classica di attivazione del complemento conferisce un rischio elevato di contrarre la patologia, CR1 è da tempo ritenuto un buon candidato per la suscettibilità al LES. Questa considerazione parrebbe confermata anche dal fatto che i livelli eritrocitari di CR1 (E-CR1) sono diminuiti nei pazienti LES. Numerosi studi volti ad individuare un'associazione genetica tra un polimorfismo responsabile della alta/bassa espressione di E-CR1 ed il LES sono però falliti (meta-analisi effettuata da Nath et al. 2005). Ridotti livelli di espressione nei LES sono stati riscontrati anche per CR2, un'altra proteina del cluster RCA, e si osservano oscillazioni del livello di espressione correlati con lo stato della malattia sia per CR1 che per MCP.

Per escludere l'eventualità che l'associazione osservata rifletta quella più forte con uno SNP in LD con quelli analizzati abbiamo tipizzato 14 tagSNPs situati nelle regioni centromeriche e telomeriche a *CR1L*, includenti i geni *CR1* ed *MCP*. Nessuno dei marcatori analizzati è risultato significativamente associato al LES, anche se due polimorfismi in posizione telomerica, situati rispettivamente nel gene *MCP* e nella porzione intergenica tra i geni *CR1L* ed *MCP*, contribuiscono a formare un aplotipo che amplifica l'effetto del polimorfismo del gene *CR1L* maggiormente associato al LES. Anche questa associazione non si riconferma però nella popolazione spagnola. Tra i marcatori telomerici rispetto a *CR1L* ed i polimorfismi situati nella porzione 3' terminale di *CR1L* si osserva un certo grado di LD, sia nella nostra popolazione che nelle famiglie CEPH del progetto HapMap. Non si osserva LD tra i polimorfismi statisticamente associati di *CR1L* ed i tagSNP del gene *CR1*, né nella nostra casistica, né nelle famiglie CEPH tipizzate dal progetto HapMap. In base ai nostri dati, quindi, l'associazione osservata sembrerebbe effettivamente imputabile prevalentemente a polimorfismi situati nella regione 3' terminale di *CR1L*.

CR1L è normalmente espresso e funzionale sulle membrane eritrocitarie dei primati non umani. Logar et al. hanno trovato prove che l'isoforma più lunga di *CR1L* sia espressa nel tessuto linfoide fetale umano ed in quello ematopoietico umano adulto, ma a livelli molto più bassi rispetto a *CR1*. Non sono però riusciti ad evidenziare anche la proteina. I nostri dati hanno rilevato tracce del cDNA dell'isoforma maggiore di *CR1L* in linfociti umani attivati e posti in coltura. Questo dato potrebbe però essere dovuto semplicemente a livelli basali di trascrizione. Inoltre abbiamo trovato in quantità più significativa il cDNA corrispondente ad una isoforma di *CR1L* costituita solo dai primi 4 esoni e dovuta a poliadenilazione alternativa, amplificandolo da cDNA di midollo osseo e di linfociti periferici non attivati. In base ai nostri dati sembra quindi che almeno l'isoforma più corta, e, anche se a livelli molto basi, l'isoforma a lunghezza completa, siano trascritte in *Homo Sapiens*. Per poter stabilire con sicurezza se la proteina CR1L viene effettivamente tradotta sarebbe necessario mettere

a punto un Western blot oppure un test di tipo ELISA. Al momento questo non è stato fatto poiché non si dispone di un anticorpo specifico per CR1L. Questo fatto ci impedisce anche di indagare la presenza della proteina umana CR1L sulla membrana degli eritrociti, dove potrebbe essere espresso in quantità nettamente superiore.

Una volta risolto il dubbio sull'effettiva espressione di CR1L resta comunque da chiarire la sua funzionalità. CR1 in *Homo Sapiens* è una proteina transmembranaria strutturata in ripetizioni in tandem di sequenze omologhe (SCR) organizzate nella sua isoforma più comune in 4 LHR (long homologous repeats) caratterizzate da elevata omologia. Ogni dominio LHR ha una sua attività di legame indipendente dagli altri (Tas et al. 1999, Krych et al. 1991, 1994 e 1998; Klickstein et al. 1988). Il gene *CR1L* è più corto, infatti è costituito da un solo dominio LHR e manca delle porzioni C-terminali contenenti il dominio trans-membrana. CR1L nei primati non umani è invece ancorato alla membrana tramite GPI, ed è perfettamente funzionale. La porzione C-terminale del CR1L umano il cui cDNA è stato amplificato da Logar et al. (isoforma 1) presenta differenze rispetto a quello degli altri primati, con l'introduzione di un codone di stop. La porzione C-terminale della proteina non esibisce una idrofobicità significativa o una sequenza consenso per il legame con GPI, per cui è plausibile che, se espresso, CR1L sia una proteina secretoria (Logar et al. 2004). Va ricordato che oltre alle proteine con domini trans-membrana CR1 e CR2, il cluster RCA comprende anche proteine ancorate alla membrana tramite GPI e proteine sieriche. Inoltre è stata identificata una forma ricombinante sierica di CR1, funzionalmente attiva. Anche se probabilmente non è in grado di restare ancorato alla membrana cellulare CR1L può dunque essere comunque funzionale. Logar et al. osservano una sua attività di binding con C4b, anche se non con C3b, di una proteina purificata da un esperimento di trascrizione e traduzione in vitro. Ricordiamo che C3 rappresentai il crocevia di entrambe le vie del complemento, sia quella classica che quella alternativa, mentre C4 è coinvolto esclusivamente nella via classica. E' interessante considerare anche l'importanza cruciale che la via classica di attivazione del complemento riveste nella patogenesi del LES: infatti i deficit genetici delle sue componenti (C1q, C2, C4) causano una forte predisposizione al LES. Questi fatti ben noti possono indurre a speculazioni sul ruolo di *CR1L* nel controllo dell'attivazione della via classica del complemento, e, di conseguenza nella suscettibilità al LES. Inoltre il gene murino *Crry*, costituito da 5 SCR omologhi a quelli di *CR1* e pertanto strutturalmente simile al corto CR1L è funzionale. Non si comporta però come recettore per C3b o C4b, ma presenta un'attività che riassume quelle delle proteine umane DAF e MCP.

Per quanto riguarda le due isoforme più corte è stato osservato che l'isoforma 2 sembra contenere al C-terminale un dominio di ancoraggio con il GPI. Inoltre questa isoforma sembrerebbe possedere

una specificità di legame per i frammenti C4c e C4d, derivati dal clivaggio di C4 e che normalmente svolgono il ruolo di opsonine (Irshaid, personal comunication).

La regione 3'UTR è fortemente implicata nel controllo della regolazione dell'espressione genica perché controlla specificamente la stabilità, l'attività trascrizionale, e l'esporto fuori dal nucleo del messaggero. Variazioni di sequenza a carico delle porzioni 3'UTR possono influenzare alcune di queste funzioni ed essere mutazioni causali di malattie genetiche. E' ipotizzabile che il gene *CR1L* sia in grado, anche se in stretta misura, di vicariare parzialmente *CR1*, e che una sua aumentata espressione possa conferire una protezione contro la malattia. Per cercare conferme a questa ipotesi vorremmo quantificare e confrontare i livelli di espressione di *CR1L* in individui portatori dell'aplotipo protettivo contro i livelli di individui portatori dell'aplotipo di rischio. E' stata effettuata un'analisi in silico con il programma UTR Scan, ma nessuno dei polimorfismi di interesse situati nella regione 3'UTR ricade in sequenze di legame per fattori trascrizionali.

Vista l'importanza del complemento nella patogenesi del LES, riteniamo che possa essere interessante approfondire l'analisi sia genetica che funzionale su questo gene. CR1 e CR2 sono stati proposti come bersagli terapeutici nella cura di alcune patologie. Se l'importanza di CR1L dovesse confermarsi non sarebbe impossibile ipotizzare un suo ruolo futuro per la terapia del LES.

Analisi Mutazionale

In questo studio ci siamo posti lo scopo di investigare il ruolo di varianti rare di due geni; *PRF1* e *TREX1* nel determinare suscettibilità a tre diverse malattie autoimmuni: LES , SSc e SS (per quest'ultima è stato per il momento analizzato solo il gene *TREX1*).

Le due varianti non sinonime più frequenti del gene *PRF1* sono p.Ala91Val e p.Asn252Ser. Entrambe queste variazioni sono state precedentemente osservate in associazione con altre malattie autoimmuni, in particolare con ALPS e T1D (p.Asn252Ser), DALD e MS (p.Ala91Val). In questo lavoro osserviamo un'associazione con le frequenze alleliche di *PRF1* anche nel LES (p.Asn252Ser) e con la SSc (p.Ala91Val), (anche se ammettiamo che quest'ultimo dato potrebbe essere inficiato dalla bassa numerosità campionaria, dovuta alla rarità della malattia). Studi basati su modelli sia umani (analisi dei linfociti citotossici su soggetti portatori dell'allele 91Val), che murini (cellule basofile leucemiche di ratto trasfettate con cDNA di *PRF1* contenenti la variazione) hanno evidenziato una rilevanza funzionale per la sostituzione p.Ala91Val, che causa un cambiamento conformazionale che a sua volta genera una diminuzione della stabilità della proteina e dell'attività NK (circa il 50% dell'attività della PRF1 wild-type) (Trambas et al. 2005, Voskoboinik et al. 2005,

Risma et al. 2006). Il significato funzionale di p.Asn252Ser è invece incerto. La variazione è localizzata nel membrane attack complex, un dominio critico per la funzione di PRF1 di formazione di pori nella membrana cellulare. Tuttavia più di un lavoro ha associato questa variante con livelli normali di espressione di PRF1 ed una normale attività NK (Stepp et al. 1999, Voskoboinik et al. 2005, Risma et al. 2006). Al contrario sono stati recentemente descritti due pazienti pediatrici, uno con ALPS e l'altro con T1D eterozigoti per la variante p.Asn252Ser che presentavano bassi livelli di attività NK nella prima infanzia (Clementi et al. 2006; Orilieri et al. 2008). Nel paziente con ALPs l'attività NK è revertita a livelli fisiologici intorno ai 12 anni di età (Clementi et al. 2006). Perciò è possibile che la variazione p.Asn252Ser, o un'altra variante associata ad essa, causi una diminuzione della funzione NK durante la prima infanzia, ma che successivamente questa venga normalizzata da un meccanismo compensatorio. Non sembra molto probabile che questo meccanismo ipotetico sia coinvolto nella suscettibilità a patologie come il LES o la SSc, che solitamente hanno esordio nell'età adulta, tuttavia è stato descritto anche un paziente adulto omozigote per l'allele 252Ser e con bassa attività NK (Orilieri et al. 2008). E' noto che p.Asn252Ser è in LD assoluto (D'=1, r^2=1) con due varianti sinonime (c.435G>A e c.462A>G) e la stessa osservazione è stata confermata anche nella nostra casistica. E' possibile che queste, oppure altre variazioni ignote, situate nelle regioni regolatorie di *PRF1* da noi non analizzate ed in LD assoluto con p.Asn252Ser possano avere un effetto sulla funzione proteica. Entrambe le variazioni c.435G>A e c.462A>G cambiano un codone comune con uno più raro, ed è stato osservato che questo fatto può portare ad anomalie nel folding della proteina (Komar 2007).

Oltre a p.Ala91Val e p.Asn252Ser sono state riscontrate altre 9 variazioni non sinonime. A parte p.Ala91Val e p.Asn252Ser, le altre varianti analizzate sono molto rare, per cui non è possibile effettuare una valutazione su base statistica della loro eventuale associazione individuale con le due malattie, ma, analogamente a quanto già osservato nella SM (Cappellano et al. 2008), l'incidenza complessiva di variazioni di sequenza non sinonime nel gene *PRF1* è più elevata nei pazienti rispetto ai controlli sia per la casistica di LES (p=0.043, OR = 1.63) che per quella di SSc (p=0.0082 OR = 1.89).

La variazione p.Arg4His è localizzata nel peptide segnale. Nella nostra popolazione è stata riscontrata in un paziente affetto da SSc limitata con ipertensione delle arterie polmonari. Precedentemente era già stata segnalata in eterozigosi in un paziente Afro-Americano di 21 anni affetto da anemia aplastica (Solomou et al. 2007), in un soggetto Europeo con SoJIA (Systemic onset Juvenile Idiopathic Arthritis) (Vastert et al. 2008), ma anche in 21/1156 controlli sani (frequenza allelica = 0,009) (Solomou et al. 2007). Inoltre è stato riportato con una MAF simile

anche nelle banche dati (popolazione CEU). La perforina è secreta in forma inattiva, ed attivata tramite taglio proteolitico. Sulla base dell'analisi in silico, questa variazione non sembra influenzare il taglio proteolitico del peptide segnale o il trasporto della proteina stessa. Considerando i dati in nostro possesso, questa variazione sembra un polimorfismo che probabilmente non influenza negativamente l'attività della proteina.

Altre tre variazioni sono state osservate ciascuna in un singolo paziente con LES. Una di queste (p.Val135Met) è riportata sulle banche dati come rs12263572, ed era stata identificata sulla popolazione AGI-ASP (Caucasoidi e Afro-Americani apparentemente sani), ma non è mai stata descritta in letteratura, le altre (p.His514Arg e p.Val329Ile) sono variazioni nuove. Le sostituzioni p.Val135Met e p.His514Arg compromettono residui conservati situati all'interno di regioni funzionalmente importanti (rispettivamente il dominio MACPRF ed il dominio C2), e p.His514Arg è stata predetta come non neutrale da tre diversi algoritmi. In base all'analisi funzionale sembrerebbe che nè p.Val135Met né p.His514Arg influenzino l'attività NK, almeno allo stato eterozigote. Tuttavia entrambi i pazienti in cui abbiamo osservato queste due variazioni presentavano una leggera alterazione in un parametro immunologico connesso con la perforina (rispettivamente: proporzione di cellule NK $CD3^{-}CD16^{+}$ e di linfociti esprimenti perforina). Tuttavia è difficile fare ipotesi sulla rilevanza clinica di queste osservazioni. La variazione p.Val329Ile è una sostituzione conservativa, che ricade nel dominio MACPRF ma non coinvolge un residuo conservato, ed è stata predetta come sostituzione neutrale da tutti i programmi utilizzati per l'analisi, pertanto la consideriamo un polimorfismo che probabilmente non influenza la funzione di PRF1.

Abbiamo osservato altre 5 variazioni non sinonime, ma tutte quante sono state riscontrate anche nella casistica di controlli. Tre di queste (p.Arg123His, p.Gly334Ser e p.Phe421Cys) sono state osservate ciascuna in un caso di SSc ed in un controllo sano, due (p.Arg232His and p.Ala437Val) ciascuna in un singolo soggetto di controllo. Tra le variazioni riscontrate nella casistica di SSc, p.Phe421Cys è localizzata nel dominio C2, responsabile del legame di PRF1 con la membrana plasmatica. Era stata precedentemente descritta in eterozigosi in un paziente pediatrico Italiano con linfoma anaplastico a grandi cellule (ALCL) (Cannella et al.2007) ed in eterozigosi composta con p.Ala91Val in un paziente caucasoide di FHL2 (Clementi et al. 2005). In entrambi i soggetti la variazione p.Phe421Cys è stata riscontrata a carico della linea germinale. Nella nostra casistica è stata trovata due volte, in un paziente con SSc diffusa ed in un controllo sano. Complessivamente, considerando le informazioni presenti in letteratura ed i dati ricavati dall'analisi in silico, il significato funzionale di questa sostituzione non è chiaro. Al contrario la variazione p.Arg123His non ricade in un dominio funzionale o in una posizione conservata. Era stata precedentemente

descritta in eterozigosi in un paziente di 14 anni affetto da ALCL (Cannella et al.2007), ed in 1/101 controlli sani, eterozigote anche per la variazione sinonima c.519G>A (Molleran Lee et al. 2004). E' considerata da molti autori un polimorfismo di *PRF1*. Nella nostra popolazione è stata osservata due volte, in un controllo sano ed in un paziente affetto da SSc limitata, entrambi portatori anche della variazione c.519G>A. Viene considerata neutrale dagli algoritmi Polyphen e PMUT. Complessivamente non pensiamo che questa variazione possa avere un effetto patologico sulla funzione proteica. La terza variazione, p.Gly334Ser, localizzata nel dominio MACPRF, non è mai stata descritta precedentemente. Coinvolge un residuo non conservato ed è considerata neutrale dai quattro algoritmi utilizzati per l'analisi, per cui sembra improbabile che possa avere un effetto sulla funzione della proteina.

La sostituzione p.Arg232His è stata riscontrata nello stesso allele con p.Ala 91Val in un controllo sano. E' stata precedentemente descritta e registrata sull' Human Gene Mutation Database – HGMD (http://www.hgmd.cf.ac.uk/ac/index) da Feldmann et al. (2002) in eterozigosi in due pazienti consanguinei con FHL2 ad esordio tardivo appartenenti alla popolazione italiana, e nella loro madre (sana). In questi pazienti gli autori hanno riscontrato circa il 27% dell'attività citotossica e livelli di espressione di PRF1 non rilevabili. In nessuno dei due è stata ritrovata la mutazione sul secondo allele. Uno dei pazienti era ancora in vita all'età di 14 anni senza aver avuto bisogno di ricevere un trapianto di midollo. Busiello et al. (2004) hanno trovato la variante p.Arg232His in due gemelli di 13 anni, uno dei quali affetto da FHL, mentre l'altro era sano. Entrambi i gemelli erano omozigoti per la variante p.Ala91Val ed eterozigoti per p.Arg232His (uno degli alleli era "doppiamente mutato". La variazione Arg232His era presente anche nel padre sano. Questa variazione è stata riscontrata anche in un altro paziente Italiano con FHL ad esordio tardivo, in eterozigosi composta con una mutazione missense (p.His222Gln) (Stadt et al. 2006). In uno studio condotto su un'ampia coorte Trizzino et al. (2008) hanno riscontrato questa variazione in quattro famiglie con membri affetti da FHL2, tre dei quali erano Italiani, e il quarto proveniva dalla Germania ma era di probabili origini italiane. La variazione è stata inoltre osservata in un paziente pediatrico italiano con ALCL, in eterozigosi composta con p.Ala91Val (Cannella et al. 2007). Voskoboinik et al (2005) hanno effettuato un'analisi funzionale della PRF1 Arg232His-mutata e Arg232His/Ala91Val-doppiamente mutata. La perforina Arg232His-mutata genera approssimativamente il 30% dell'attività NK rispetto alla proteina nativa, mentre la forma doppiamente mutata è completamente inattiva. Risma et al. (2006) hanno analizzato funzionalmente 21 mutazioni missense di *PRF1* tramite citofluorimetria, immunoistochimica ed immunoblotting, ed hanno dimostrato che la PRF1 con la sostituzione Arg232His è soggetta a maturazione proteolitica parziale. Complessivamente i dati disponibili in letteratura mostrano chiaramente che la sostituzione p.Arg232His compromette

parzialmente l'attività citotossica di PRF1. Tuttavia la penetranza di questa variazione sembra essere incompleta. Infatti individui con lo stesso pattern di variazioni presentano condizioni patologiche diverse, ed alcuni di essi sono sani. Queste osservazioni sono simili a quelle riportate per la variante p.Ala91Val, che molti autori considerano un polimorfismo di *PRF1*. Inoltre l'elevato numero di segnalazioni di p.Arg232His suggerisce che si tratti di una variante relativamente comune nella popolazione italiana, per cui non deve stupire il fatto di averla riscontrata in un controllo sano, in *cis* con la variazione comune p.Ala91Val. Sulla base dei dati di Voskoboinik et al (2005) potremmo speculare che questo soggetto abbia probabilmente il 50% dell'attività citotossica, sufficiente ad evitare lo sviluppo di FHL2. Purtroppo non è stato possibile effettuare un ulteriore follow up successivo al reclutamento di questo soggetto, per cui non possiamo sapere se esso abbia in seguito sviluppato un linfoma. L'ultima variazione (p.Ala 437Val) è stata osservata in un solo controllo, e non è mai stata segnalata prima. Nonostante codifichi per una sostituzione conservativa tra due aminoacidi non polari, è considerata non tollerata da due diversi algoritmi predittivi, ed è situata in una posizione conservata nel dominio C2, per cui sarebbe troppo semplicistico considerarla una variante neutrale. E' possibile che, così come p.Ala91Val e p.Arg232His, anche la variazione p.Ala437Val abbia un qualche effetto funzionale con bassa penetranza, ma sarebbero necessari ulteriori studi per chiarire questo punto.

Durante lo screening mutazionale di *PRF1* abbiamo identificato anche 11 variazioni sinonime, comprendenti due polimorfismi comuni (rs885821 e rs885822) e le due varianti in LD assoluto con p.Asn252Ser già menzionate. Tra le rimanenti 7 variazioni, solo 3 erano state precedentemente citate in letteratura, mentre le altre sono nuove. Tutte le variazioni sinonime sono state analizzate in silico per predire la loro influenza sui siti di splicing, sulle sequenze ESE e sulla frequenza relativa di utilizzo del codone che introducono. c.1620A>G e c.999C>T erano state precedentemente riportate in un paziente affetto da MS ciascuna (Cappellano et al, 2008), rispettivamente in omozigosi ed in eterozigosi. E' stato ipotizzato che c.1620A>G possa influenzare lo splicing, in quanto potrebbe creare un nuovo sito accettore di splicing. Inoltre, in base ai punteggi calcolati con la matrice ESEfinder, sia c.999C>T che c.1620A>G potrebbero influenzare la processazione del messaggero perché rimuovono (c.999C>T) o creano (c.1620A>G) una sequenza ESE. Abbiamo riscontrato la variazione c.519G>A tre volte, una in ciascun gruppo di campioni (LES, SSc, controlli). Il controllo sano ed il paziente con SSc erano anche eterozigoti per la sostituzione p.Arg123His. Come già menzionato precedentemente, c.519G>A è già stata descritta in letteratura unita a p.Arg123His nello stesso controllo sano, ed è stata considerata dagli autori come un polimorfismo. Pertanto riteniamo che si tratti di una variazione silente, relativamente comune nella popolazione italiana, probabilmente in LD con p.Arg123His.

E’ più difficile speculare sulla possibile rilevanza funzionale delle variazioni sinonime che vengono qui descritte per la prima volta. Nessuna di esse coinvolge un sito di splicing canonico, tuttavia, in base alle predizioni in silico effettuate con i programmi ESE finder e Splice View, alcune di esse (c.273G>A, c.807C>T, c.1356C>T) potrebbero influenzare lo splicing introducendo o rimuovendo una sequenza di legame per vari fattori di splicing. Le altre varianti osservate non sembrano influenzare lo splicing e non introducono codoni poco usati, per cui sono probabilmente neutre. E’ possibile tuttavia che le variazioni osservate non abbiano alcun effetto funzionale diretto, ma che siano in LD con altre variazioni ignote situate nella porzione 5’UTR o nelle sequenze regolatorie a monte ed a valle del gene.
PRF1 potrebbe influire sulla suscettibilità all’autoimmunità attraverso due meccanismi. Da un lato difetti nell’attività della perforina potrebbero causare una insufficiente clearance virale, che potrebbe favorire lo sviluppo di reazioni crociate tra antigeni virali ed antigeni self attraverso il mimetismo molecolare. Dall’altro PRF1 potrebbe essere coinvolta nella modulazione della risposta immune inducendo l’apoptosi in linfociti effettori e cellule presentanti l’antigene. Un suo deficit potrebbe causare un accumulo di linfociti e possibili reazioni autoimmuni.
La scelta di studiare il gene *TREX1* sulle malattie autoimmuni è stata motivata da uno studio del 2007 (Lee-Kirsh) che identificava mutazioni in *TREX1* in 9 su 417 pazienti affetti da LES e 1 su 169 pazienti affetti da Sindrome di Sjögren (SS). Inoltre 2 dei 9 pazienti affetti da LES descritti da Lee-Kirsh presentavano tra la sintomatologia una sindrome di Sjögren secondaria, considerazione che fornisce nuovi elementi a supporto di un eventuale coinvolgimento di *TREX1* nella suscettibilità alla SS. Mutazioni causali di *TREX1* sono state descritte in una rara forma monogenica di lupus eritematoso (FCL: familial chilblain lupus) e in una malattia autosomica recessiva che presenta alcune caratteristiche cliniche in comune con il LES (AGS: Aicardi-Goutières Syndrome)

Nel nostro studio sono stati analizzati 210 pazienti affetti da LES, 58 affetti dalla Sindrome di Sjogren e 150 affetti dalla Sclerosi sistemica, alla ricerca di nuove mutazioni all’interno della sequenza codificante e della regione 3’ UTR del gene *TREX1*. Il coinvolgimento di *TREX1* nella SSc è stato qui preso in considerazione per la prima volta.

In questa analisi abbiamo riscontrato la presenza di 7 sostituzioni di una singola base che ricadono tutte all’interno dell’unico esone codificante di *TREX1*. Due di queste (c.592G>A e c.694A>G) sono variazioni missense, che causano rispettivamente le variazioni aminoacidiche p.Glu198Lys (riscontrata in eterozigosi in 1 paziente affetto da Sindrome di Sjogren) e p.Met232Val (osservata in eterozigosi un paziente affetto dalla forma limitata di Sclerosi Sistemica). Le due varianti non sono state osservate né in 200 controlli italiani da noi analizzati né in 1712 controlli sani

precedentemente sequenziati da Lee-Kirsch. Coinvolgono entrambe un dominio altamente conservato tra le diverse specie, e in base ad una predizione in silico effettuata con 4 diversi programmi, sono entrambe ritenute possibilmente dannose da tutti e 4 gli algoritmi utilizzati.

La variazione Glu198Lys è localizzata all'interno di uno dei tre domini che costituiscono il sito catalitico dell'enzima, il dominio Exo3. E' stata precedentemente riportata in omozigosi in un paziente turco di 19 anni affetto da AGS. Il paziente presentava alcuni sintomi simili a quelli del LES (lesioni lupiche sui piedi, ulcere orali, ANA, anti-ENA e una diminuzione nei livelli di C3). L'importanza del dominio Exo3 nella funzionalità della proteina è provato dal fatto che nei pazienti AGS sono state descritte molte mutazioni a suo carico, sia in omozigosi che in eterozigosi. Per tre di esse (p.Asp200Asn, p.Asp201ins and p.Val201Asp) sono state effettuate analisi funzionali, che hanno dimostrato una drammatica diminuzione (fino a 35.000 volte inferiore per p.Asp201ins) nell'attività nucleasica Udesh de silva 2007, Lehtinen 2008). E' interessante considerare che entrambe le mutazioni che coinvolgono il residuo Asp-200 sembrano essere sufficienti, in eterozigosi, a causare l'AGS, che normalmente è una malattia recessiva. Lehtinen et al hanno dimostrato tramite esperimenti di cotrasfezione e saggio esonucleasico, che TREX1^{200Asn} agisce con un effetto dominante negativo sulla degradazione del DNA a doppia elica, inibendola anche in presenza di TREX1^{wt}.

La variazione Met232Val viene qui descritta per la prima volta. Cade nella porzione proteica compresa tra Exo3 ed il dominio trans-membrana TMD/TMH. La metionina in posizione 232 è un aminoacido altamente conservato tra le diverse specie. Anche questa mutazione è predetta come probabilmente dannosa dall'analisi in silico. La porzione proteica in cui ricade la mutazione Met232Val presenta un'elevata densità di mutazioni, sia missense che frameshift, precedentemente osservate in altre malattie come la RVLC (Richards et al, 2007) o il LES (Lee-Kirsch et al, 2007). Per alcune delle varianti frameshift è stata eseguita un'analisi funzionale tramite microscopia confocale, ed è stata osservata una non corretta distribuzione intracellulare della proteina, che perde lo sua localizzazione perinucleare e si trova diffusa nella cellula. Non ci sono però dati funzionali che ci aiutino ad ipotizzare quale può essere l'effetto di una variazione missense in questa regione.

Nel loro insieme questi dati suggeriscono fortemente che le variazioni descritte siano associate alla patogenesi delle due malattie (SS e SSc), confermando così il ruolo già precedentemente ipotizzato di *TREX1* nella SS, mentre la sua possibile associazione con la SSc è un dato nuovo. Al contrario non abbiamo riscontrato nessuna variazione missense nei 210 LES esaminati. Questo dato è significativamente diverso ($p = 0.034$) da quanto osservato da Lee-Kirsh et al. 2007 che identificava

9 mutazioni non-sinonime su 417 pazienti affetti da LES, però la significatività può essere dovuta al caso o a caratteristiche genetiche differenti tra nord-europei e popolazione italiana.

TREX1 potrebbe avere un ruolo nello sviluppo dell'autoimmunità in quanto la sua attività principale consiste nella risposta apoptotica mediata da Granzima A e da Perforina, e nella rimozione di ssDNA o ibridi di DNA:RNA derivati da retroelementi endogeni. Mutazioni di TREX1 che causino la perdita della funzione esonucleasica, potrebbero causare un accumulo di ssDNA e di ibridi DNA:RNA nel nucleo, scatenando di conseguenza una risposta autoimmune. A conferma di questa ipotesi risulta il fatto che sia nel LES che nella SSc si osserva la presenza di autoanticorpi diretti contro antigeni nucleari.

In conclusione questo studio avvalora l'ipotesi che varianti rare possano avere un'influenza sulla suscettibilità alle malattie autoimmuni. Per quanto riguarda il gene *PRF1*, osserviamo anche nel LES e nella SSc un'associazione con due polimorfismi con un probabile ruolo funzionale, p.Ala91Val (SSc) e p.Asn252Ser (LES), che era già stato osservato in altre malattie autoimmuni (ALPS, DALD, T1D e SM). Inoltre osserviamo un'associazione con la presenza di variazioni non sinonime rare. Una di queste (p.His514Arg), osservata in un paziente con LES, sembra evidenziare le prove più convincenti di un suo possibile effetto patologico. Per quanto riguarda il gene *TREX1* confermiamo il suo ruolo nella suscettibilità alla SS e identifichiamo un possibile coinvolgimento con la SSc fino ad ora mai analizzata. Nella nostra casistica non confermiamo il coinvolgimento di *TREX1* nella suscettibilità al LES. Infine è interessante considerare come i due geni *PRF1* e *TREX1* codifichino per proteine coinvolte in pathway connessi tra loro, entrambi coinvolti nel meccanismo apoptotico.

Creazione di un algoritmo genetico di suscettibilità

Negli ultimi anni i metodi d'indagine Genome-Wide hanno contribuito notevolmente ad ampliare la conoscenza delle basi genetiche del Lupus Eritematoso Sistemico, permettendo la scoperta di almeno 30 loci di suscettibilità che sono stati confermati in almeno due casistiche indipendenti. La maggior parte degli alleli di suscettibilità osservati sino ad oggi è relativamente comune (la frequenza degli alleli di rischio varia da 0.10 a 0.90 a seconda del polimorfismo nella popolazione sana). Inoltre la maggior parte di essi contribuisce al rischio di contrarre la malattia solo con un basso valore di OR (da 1,15 a 2,54). Questi dati comportano che singolarmente i singoli marcatori abbiano una scarsa rilevanza per la pratica clinica, sia a livello predittivo sia a livello diagnostico. Per un paziente la determinazione della positività o negatività a un singolo allele di rischio comune

nella popolazione e con un effetto ridotto non è informativa. Tuttavia aggregare tutti i fattori genetici di rischio validati mediante un unico algoritmo cumulativo (genetic risk score) permette di predire il rischio di malattia e quindi di identificare individui ad alto rischio di sviluppare il LES in futuro. Quest'operazione è stata già condotta con successo da De Jager et al. (2009) sulla Sclerosi Multipla (SM). Gli autori hanno messo a punto un algoritmo utilizzando 16 marcatori di suscettibilità alla SM, ciascuno con un OR compreso tra 1,1 e 2,75. All'inizio del nostro lavoro abbiamo stilato, sulla base dei dati presenti in letteratura, una lista di SNP da utilizzare per il calcolo dell'algoritmo, utilizzando criteri d'inclusione analoghi a quelli utilizzati da De Jager. (significatività statistica che superi la soglia dei GWAS ($p<10^{-8}$) in almeno uno studio, associazione confermata in almeno due studi indipendenti con $p<10^{-4}$ in almeno uno di questi, associazione confermata in più di uno studio a carico di varianti con rilevanza funzionale). Questi criteri ci hanno portato a selezionare 30 geni da includere nell'algoritmo.

Predire la probabilità che un individuo possa contrarre la malattia può essere uno strumento molto utile ai fini della sorveglianza sanitaria, per iniziare una terapia precoce e per la scelta della terapia più adatta.(Rhodes e Vyse 2010) In particolare l'uso del genetic risk score potrebbe trovare applicazione per la diagnosi precoce in individui con alcuni segni clinici di malattia che non soddisfano ancora completamente i criteri diagnostici per il LES oppure nei parenti di primo grado dei soggetti affetti. Gli studi di aggregazione famigliare dimostrano infatti che i parenti di primo grado una probabilità circa 30 volte superiore di sviluppare il LES rispetto alla popolazione generale, per cui potrebbe aver senso sottoporli ad una indagine di screening tramite l'algoritmo di rischio genetico. L' identificazione degli individui ad alto rischio di sviluppare la malattia potrebbe permettere un follow-up clinico più accurato e misure terapeutiche più precoci e di sottoporre gli individui a rischio a misure preventive (es. limitare l'esposizione ai raggi solari).

Il potere predittivo di tali algoritmi aumenterà con l'aumentare del numero di fattori di rischio genetici che verranno identificati. Inoltre in questi algoritmi è possibile includere anche fattori ambientali, clinici, demografici, familiari noti al momento o di futura identificazione, e pertanto il potere predittivo potrà essere ulteriormente migliorato nel tempo

Al momento disponiamo di dati di genotipizzazione sulla casistica raccolta nel nostro laboratorio di 9 dei marcatori che avevamo selezionato per calcolare il wGRS, situati in 8 geni (*STAT4*, *IRF5*, *TNFSF4*, *PXK*, *HLA-DRB1*, *ATG5*, *BLK* e *SCUBE1*). Per tutti e nove i marcatori osserviamo che l'allele di rischio (secondo i dati della letteratura) è più frequente nei malati rispetto ai controlli sani anche nella nostra popolazione. Tuttavia solo per due di questi (*STAT4* rs7574865 e *HLA-DRB1* rs2187668) questa differenza raggiunge il livello di significatività statistica comunemente accettato

come soglia per la replicazione di un GWAS ($p<10^{-4}$). Gli altri sette polimorfismi, con la sola eccezione di *PXK* rs6445975 e *ATG5* rs573775 presentano comunque un valore di p<0,5, soglia utilizzata normalmente per gli studi di associazione di tipo caso-controllo su gene candidato.

In aggiunta agli SNP preselezionati, osserviamo, nella nostra casistica, un'associazione con p<10-4 per altri sei marcatori, che durante la fase di progettazione dell'algoritmo non erano stati scelti. Sono tutti localizzati in loci per i quali è già stata riportata una forte associazione con il LES, e che sono stati inclusi nella lista di 30 geni da analizzare (*STAT4, IRF5, HLA-DRB1* e *SCUBE1*). Non abbiamo riscontrato nuovi loci di suscettibilità. Queste osservazioni non sorprendono, poiché, dato il numero abbastanza esiguo di campioni su cui è stata effettuata l'analisi, disponiamo solo del potere statistico per osservare associazioni relativamente forti, che inevitabilmente sono già emerse in uno o più GWAS.

Per la creazione dell'algoritmo di suscettibilità ci siamo attenuti alla formula messa a punto da De Jagger. I nove marcatori già selezionati sulla base di dati provenienti dalla letteratura sono stati tutti inclusi nel calcolo, indipendentemente dal fatto che l'associazione precedentemente descritta si sia confermata nei nostri campioni. Come valore di OR per il calcolo abbiamo utilizzato quello ottenuto unendo insieme le casistiche analizzate nei lavori ad oggi pubblicati in cui sia emersa un'associazione con quello specifico marcatore. Riteniamo questa scelta più corretta rispetto alla possibile alternativa di utilizzare l'OR calcolato sui nostri dati, poiché ci ha permesso di avere una stima sicuramente più corretta, in quanto effettuata su una popolazione molto più ampia. Tutti gli studi che abbiamo utilizzato sono stati condotti su popolazioni caucasoidi, per cui riteniamo ragionevole utilizzare, almeno nelle fasi preliminari del progetto, gli OR riportati in queste pubblicazioni su una casistica italiana.

In aggiunta ai marcatori noti, abbiamo deciso di includere nell'algoritmo per il calcolo del wGRS anche gli altri sei polimorfismi fortemente associati ($p<10^{-4}$) nella nostra casistica. Con il crescente uso dei metodi di indagine Genome-Wide, le nostre conoscenze sulla genetica del LES sono in continua e rapida espansione. Di conseguenza riteniamo sia scorretto considerare un algoritmo predittivo come un' unità statica; al contrario la lista di loci e di marcatori elencati nella tabella 2 dovrebbe essere periodicamente aggiornata, sulle basi sia di nuovi dati in letteratura che dei risultati da noi stessi ottenuti.

I nostri risultati riguardanti il wGRS sono al momento molto preliminari, in quanto solo 9 su 52 degli SNP selezionati vi sono stati al momento inclusi, e abbiamo potuto condurre l'elaborazione solo su una casistica estremamente ridotta. Tuttavia, dalla figura R7, si può osservare come le due

curve, che pongono il valore di wGRS in relazione con la sua frequenza rispettivamente nei pazienti e nei controlli, presentino un andamento leggermente diverso, con la mediana e la media spostate di oltre un'unità. Tuttavia gli intervalli (± st. dev.) sono sovrapponibili. L'accuratezza dell'analisi (valutata come area sottesa dalla curva ROC) migliora se nell'algoritmo vengono inseriti parametri demografici e clinici, come il sesso e la presenza/assenza del fenomeno di Raynaud. Il valore massimo di accuratezza ottenuto (AUC= 0,892), calcolato sull'algoritmo contenente, oltre ai dati di suscettibilità genetica, anche entrambi i parametri citati,è convenzionalmente ritenuto un buon valore per un test predittivo. Tuttavia questo risultato potrebbe essere inficiato dal bias introdotto nel considerare tutti i controlli sani come negativi per la presenza del fenomeno di Raynaud.

In conclusione il dato preliminare sembra promettente, ma è indispensabile ampliare la nostra popolazione, nonché il numero di loci da analizzare.

Meta-analisi

Abbiamo condotto questa analisi su due geni (*TEC* e *BMPR1B*) selezionati sulla base di un precedente genome-wide non ancora pubblicato, ma i cui marcatori, nella casistica utilizzata per la replicazione, non hanno evidenziato forti prove di associazione.

Per poter poter effettuare la meta-analisi sui dati derivanti da due genome-wide e da uno studio di replicazione che hanno sfruttato tre pannelli di SNP diversi abbiamo applicato la tecnica d'imputazione dei dati mancanti. Quest'ultima è una metodica bio-informatica che trova un utilizzo sempre più ampio nell'analisi degli studi Genome-Wide, ma soprattutto nelle meta-analisi che uniscono insieme più GWAS. Studi di questo tipo stanno fiorendo negli ultimi anni, e rappresentano la nuova frontiera nell'analisi genetica delle malattie complesse. Infatti ci si sta rendendo conto che spesso per queste patologie nemmeno un GWAS condotto su poche migliaia di pazienti e controlli e su centinaia di migliaia di marcatori è in grado di fornire risposte univoche. D'altra parte, per molte malattie autoimmuni, in letteratura esistono ormai numerosi GWAS a media e alta risoluzione, che uniti insieme con una meta-analisi consentirebbero di "raccogliere in silico" casistiche di alcune decine di migliaia di campioni. Questi studi sono stati condotti su piattaforme diverse, utilizzando set differenti di marcatori. La tecnica statistica che permette di meta-analizzare questi GWAS, e di ottenere dati anche per SNP che non sono stati tipizzati, è l'imputazione.

L’analisi d’imputazione e di meta-analisi ha aumentato il potere statistico dello studio, permettendoci di osservare associazioni più forti ad alcuni loci debolmente significativi nella casistica di partenza (*TEC*_384-plex_SNP1 e *TEC* GWAS SNP3). Inoltre osserviamo deboli associazioni a carico di loci che originariamente non erano stati tipizzati. La nostra esperienza sembra quindi confermare l’utilità delle imputazioni nelle analisi genomiche. Tuttavia va ricordato che i risultati derivanti da questo tipo di analisi vanno considerati con prudenza, in quanto soggetti ad errore soprattutto in regioni con basso LD, o per inference panels costituiti da pochi marcatori. Inoltre da questi errori potrebbe derivare l’osservazione di false associazioni. Crediamo che la nostra selezione basata su criteri di qualità abbia già escluso la maggioranza, se non la totalità, dei polimorfismi con un dato non attendibile. Tuttavia è indispensabile condurre altre prove di qualità per attestare l’affidabilità dell’analisi. Intendiamo inizialmente effettuare un controllo in silico ripetendo l’elaborazione utilizzando un diverso pannello di riferimento (i dati del progetto HapMap invece di quelli derivanti dal progetto 1000 genomi). Nell’ambito del progetto HapMap sono stati genotipizzati circa la metà degli SNP presenti nel progetto 1000 genomi, per cui molti dei marcatori della tabella R23 non potrebbero essere imputati con questo tipo di analisi. Tuttavia i dati sono stati ottenuti su 30 famiglie trios, contro i 112 aplotipi del progetto 1000 genomi. Questo implica che, per alcuni marcatori, i dati del progetto HapMap potrebbero rivelarsi un pannello di riferimento migliore, in quanto permetterebbero di ottenere un’imputazione più accurata. Successivamente occorrerebbe tipizzare in laboratorio alcuni degli SNP della tabella R23 su una sottocasistica di campioni imputati, per verificare che i dati ottenuti in silico e quelli ottenuti in vivo siano sovrapponibili.

Conclusioni generali

Lo scenario complessivo della genetica del LES e in generale delle malattie autoimmuni si è notevolmente modificato negli ultimi anni, grazie soprattutto ai progressi nella conoscenza apportati dai GWAS. Il numero di geni di suscettibilità confermati è sensibilmente aumentato, tuttavia questi non bastano ancora a spiegare il fenomeno in tutta la sua complessità. Indubbiamente quello della genetica del LES è un campo ancora aperto alla sperimentazione. L’apporto combinato dei GWAS e degli studi su geni candidati ha permesso di avere una visione più precisa della complessità della suscettibilità genetica al LES, che coinvolge più di un pathway molecolare e a cui concorrono sia varianti comuni a bassissima penetranza che mutazioni rare ad alta penetranza.

Il nostro gruppo sta contribuendo a dimostrare l'importanza delle varianti rare nella suscettibilità alle malattie autoimmuni grazie a due studi qui descriti, condotti sui geni *TREX1* e *PRF1*. Gli stessi studi costituiscono un'ulteriore conferma dell'opinione corrente secondo la quale malattie autoimmuni diverse spesso condividono un "background" comune di loci di suscettibilità, a cui si aggiungono fattori di rischio specifici per la malattia, sia genetici che ambientali.

Data la continua espansione degli studi Genome-Wide, e ultimamente delle meta-analisi su GWAS diversi, parrebbe che nel campo della genetica delle malattie complesse ci sia sempre meno spazio per studi su geni candidati. In realtà i risultati ottenuti in studi come quello sul gene CR1L, o lavori come quello su *IL18* dimostrano che anche questo tipo di analisi può fornire ancora contributi importanti alla comprensione della genetica del LES.

MATERIALI E METODI

Campioni utilizzati. Tutti i pazienti e i controlli utilizzati per questa analisi, quando non è diversamente specificato, sono di origine italiana, e concordi per sesso, età e provenienza geografica. Il materiale biologico di partenza (sangue periferico) proviene da diversi centri clinici di raccolta distribuiti sul territorio italiano e il DNA è stato purificato nel nostro laboratorio con metodiche standard. Tutti i pazienti affetti da LES (femmine: maschi = 9:1) soddisfano almeno 4 criteri ARA ed hanno firmato un consenso informato per la partecipazione allo studio. I controlli sani comprendono donatori di sangue, studenti e personale tecnico. I pazienti e i controlli di origine sarda sono stati esclusi dall'analisi. La numerosità della coorte di pazienti e controlli utilizzati varia da uno studio all'altro, ed è specificata nei risultati.

Inoltre sono state utilizzate una casistica di pazienti e di controlli spagnoli, provenienti dalla regione delle Asturie, da Madrid, da Granada e da Malaga. Sono altresì state utilizzate due casistiche di pazienti italiani affetti rispettivamente da SSc e da SS, provenienti dallo stesso centro clinico. I pazienti hanno firmato un consenso informato e non sono imparentati Tutti i casi di SSc (femmine: maschi = 12:1) rispondevano ai criteri preliminari dell'American College of Rheumatology per la classificazione dei pazienti con SSc (Subcommittee for Scleroderma Criteria, 1980) e sono stati classificati nelle categorie SSc limitata o SSc diffusa.

Tipizzazione degli SNP di CR1L: Tutti i polimorfismi sono stati analizzati con la metodica SNaPshot. Per ogni polimorfismo sono state messe a punto condizioni di amplificazione specifiche in singolo o in multiplex che escludono tutte le altre sequenze omologhe e condizioni di SNaPshot in multiplex. Data l'elevata omologia tra le sequenze da amplificare e le regioni vicine i primers di PCR specifici sono stati disegnati sfruttando i pochissimi punti in cui la sequenza da amplificare e quelle omologhe, allineate con il programma BLAST, presentano poche basi non appaiate. La sequenza dei primers utilizzati, le condizioni di amplificazione, e quelle utilizzate per lo SNaPshot sono disponibili su richiesta.

Studio dell'espressione di CR1L: L'RNA totale è stato estratto utilizzando il kit RNAwiz della ditta Ambion (Austin, TX) da cellule congelate di linfociti di sangue periferico, ottenuti tramite separazione su gradiente di densità, attivati e precedentemente posti in coltura, da linfociti non attivati o da cellule congelate di midollo osseo (materiale fornito dal laboratorio di ematologia). Il cDNA è stato retrotrascritto utilizzando il kit ThermoScript TM RT+PCR Systems (Invitrogen).

I primers specifici sono stati disegnati sfruttando i punti di non omologia tra le sequenze di cDNA di *CR1* e *CR1L* grazie al programma ncbi.nlm.nih.gov/BLAST. Sono stati seguiti tre differenti protocolli per i tre diversi frammenti amplificati.

-*CR1L* isoforma 1 L'amplificazione del messaggero è stata ottenuta attraverso due PCR successive: una prima amplificazione che genera un frammento di 918 basi seguita da una seconda PCR che utilizza un primer antisenso interno al frammento precedentemente amplificato e lo stesso primer forward utilizzato nel primo passaggio (SCR5 sense cDNAF).

SCR5 sense (cDNAF) = TTT GGC ATG AAA GGG CCC TC

Cr1L cDNA ex12-13 R = TTA AAA TAA AGA TGA TCG TAC CAA G

CR1L rna ex4 R=GTG GCT GAC ATA CCC TGG AG

La prima PCR è stata eseguita in un volume totale di 20 ml, utilizzando 2 ml di cDNA, 8.0 pmol di ciascun primer, 4.0 pmol di dNTPs, 2,5 mM MgCl2, 0,75 U di Taq Polimerasi (AmpliTaq Gold, Applied Biosystems) e 1X buffer II (10 mM Tris-HCl pH 8.3, 50 mM KCl alla concentrazione finale). L'amplificazione è avvenuta in un termociclatore di tipo Gene Amp PCR System 9700 (Applied Biosystems), seguendo il protocollo touch-down. Alla fase iniziale di denaturazione /attivazione della Taq sono seguiti 20 cicli in cui la temperatura di appaiamento dei primers è stata abbassata gradualmente da 65°C a 55°C, e 33 cicli ad una temperatura di 55°C. In tutti i cicli le fasi di appaiamento ed estensione avevano una durata di 30 secondi.

2 ml di amplificato sono stati utilizzati per la PCR emi-nested. Sono stati utilizzati 2 ml di cDNA, 8.0 pmol di ciascun primer, 4.0 pmol di dNTPs, 2,5 mM MgCl2, 0,75 U di Taq Polimerasi (AB Analitica) e 1X buffer (16 mM (NH4)2SO4, 67 mM TrisHCl pH 8.8, 0.01% Tween 20 alla concentrazione finale). Si è utilizzato un protocollo di amplificazione simile a quello adoperato per la prima PCR, ma sono stati eseguiti solo 25 cicli alla temperatura più bassa (55°C).

-*CR1L* isoforma 2 Il trascritto, di 636 basi, è stato amplificato con i primers specifici:

CR1L-2 CDNA F-SP = CAC TAA CCG GAC TCA GAA GGG ACT T

CR1L-2 CDNA R-branch = CCG GTT AAG ATT TTT ATT ATT CCA GAG

L'amplificazione è stata effettuata in un volume totale di 20 ml, utilizzando 2 ml di cDNA, 10.0 pmol di ciascun primer, 10.0 pmol di dNTPs, 2,5 mM MgCl2, 0,75 U di Taq Polimerasi (AmpliTaq Gold, Applied Biosystems) e 1X buffer II (10 mM Tris-HCl pH 8.3, 50 mM KCl alla

concentrazione finale). L'amplificazione è avvenuta in un termociclatore di tipo Gene Amp PCR System 9700 (Applied Biosystems). 10 minuti iniziali di denaturazione/attivazione dell'enzima sono stati seguiti da 35 cicli di denaturazione (96°C-30"), appaiamento dei primers (52°C-45") ed estensione (72°C-1'30") e dalla fase di estensione finale (72°C-10').

-CR1/CR1L I primers sono stati scelti in modo da coamplificare indifferentemente sia *CR1* che *CR1L*

CR1L rna ex4 F= CCT CCA AAT GTG GAA AAT GG

CR1L rna ex4 R= GTG GCT GAC ATA CCC TGG AG

Per la reazione (20 ml di volume finale) sono stati utilizzati 2.0 ml di cDNA, 8.0 pmol di ciascun primer, 4.0 pmol di dNTPs, 2,5 mM MgCl2, 0,75 U di Taq Polimerasi (AmpliTaq Gold, Applied Biosystems) e 1X buffer II (10 mM Tris-HCl pH 8.3, 50 mM KCl alla concentrazione finale). L'amplificazione è avvenuta in un termociclatore di tipo Gene Amp PCR System 9700 (Applied Biosystems), seguendo il protocollo touch-down.

Ricerca di nuove variazioni in TREX1 e PRF1. Le regioni codificanti dei geni *PRF1* e *TREX1* sono state analizzate per la ricerca di nuove variazioni di sequenza. Il DNA è stato isolato da sangue periferico con metodiche standard. L'intera sequenza codificante di PRF1 (esoni 2 e 3) è stata amplificata con condizioni standard di PCR. L'esone 2 è stato analizzato tramite sequenziamento diretto come descritto da Cappellano et al. L'esone 3 è stato amplificato in tre distinti frammenti parzialmente sovrapposti, e analizzato tramite Denaturing High Performance Liquid Chromatography (DHPLC) o sequenziamento diretto. L'unico esone codificante del gene TREX1 è stato amplificato in tre frammenti distinti parzialmente sovrapponibili di 375 bp, 428 bp e 466 bpp. Il primo frammento è stato analizzato tramite DHPLC, gli altri due tramite sequenziamento diretto. I frammenti analizzati tramite sequenziamento diretto sono quelli contenenti polimorfismi noti o la cui analisi tramite DHPLC avrebbe richiesto un numero eccessivo di temperature. Le temperature dell' analisi con DHPLC sono state calcolate tramite l'algoritmo "DHPLC melt program" (http://insertion.stanford.edu/melt). I prodotti di PCR sono stati esposti a rapida denaturazione termica ed a lenta rinaturazione su di un termociclatore, quindi analizzati sullo strumento TRANSGENOMIC WAVE (Transgenomic, Omaha, NE, USA). I risultati sono stati analizzati con il programma Navigator™ (Transgenomic, Omaha, NE, USA). Ai fini del sequenziamento diretto, i prodotti di PCR sono stati purificati con un sistema di filtrazione a vuoto (PCR96 Cleanup kit-

Millipore, Billerica, MA, USA). La reazione di sequenza è stata effettuata su un termociclatore utilizzando il kit ABI PRISM® BigDye Terminator kit v.1.1 (Applied Biosystems, Foster City, CA, USA), purificata tramite filtrazione a vuoto con il sistema Montage SEQ96 Sequencing Reaction Cleanup Kit (Millipore, Billerica, MA, USA) ed analizzata sul sequenziatore automatico 3100 Genetic Analyser (Applied Biosystems, Foster City, CA, USA).

Tipizzazione delle varianti p.Ala91Val and p.Asn252Ser. Le varianti di sequenza del gene *TREX1* p.Ala91Val e p.Asn252Ser sono state tipizzate utilizzando due saggi Taqman di discriminazione allelica pre-disegnati (Applied Biosystems, Foster City, CA, USA): C__25600964_20 (p.Ala91Val) e C__27529700_10 (p.Asn252Ser) e lo strumento per la Real Time PCR ABI PRISM 7000® Sequence Detection System (Applied Biosystems, Foster City, CA, USA). I genotipi sono stati attribuiti automaticamente dal programma per la discriminazione allelica SDS 1.3.

PCR allele-specifica. L'allele wild-type (91Ala) e quello "mutante" (91Val) sono stati amplificati separatamente utilizzando un primer forward specifico (primer F 91Ala-specifico: 5'-CTCCAGCGCCTGCCTCTGGC-3', primer F 91Val-specifico: 5'-CTCCAGCGCCTGCCTCT GGT-3') ed il primer reverse appropriato. I prodotti di PCR sono stati tipizzati per la seconda variazione (p.Asn252Ser o p.Arg232His) tramite sequenziamento diretto con il kit ABI PRISM® BigDye Terminator kit v.1.1 ed analizzati sul sequenziatore automatico 3100 Genetic Analyser.

Analisi in silico delle varianti di sequenza osservate. L'eventuale effetto delle variazioni sinonime sullo splicing è stato valutato utilizzando il programma SpliceView (http://www.itb.cnr.it/sun/webgene/) e l'algoritmo ESEfinder (http://www.rulai.cshl.edu/tools/ESE). Il significato funzionale delle variazioni non sinonime è stato predetto utilizzando 4 programmi diversi: PolyPhen (http://genetics.bwh.harvard.edu/pph), SIFT (http://blocks.fhcrc.org/sift/SIFT), SNAP (http://cubic.bioc.columbia.edu/services/SNAP/submit) e PMUT (http://mmb2.pcb.ub.es:8080/pMut/). L'effetto della variazione osservata nel peptide segnale sul taglio del peptide stesso è stata predetta utilizzando l'algoritmo 3.0 (www.cbs.dtu.dk/services/SignalP). L'analisi di allineamento delle sequenze proteiche umane di PRF1 e TREX1 con le sequenze ortologhe è stato effettuato tramite il programma ClustalW (http://saier-144-37.ucsd.edu/clustalw).

Analisi di associazione. La significatività statistica delle differenze riscontrate tra pazienti e controlli nelle frequenze fenotipiche, alleliche e genotipiche è stata valutata utilizzando il test del Chi quadro, oppure il test di Fisher a due code, se richiesto dalla numerosità del campione. L'associazione di ciascun polimorfismo con la suscettibilità alla malattia è stata misurata tramite Odd Ratio (OR) ed il suo intervallo di confidenza al 95% (95% CI). Per i polimorfismi del progetto 384-plex l'analisi è stata condotta utilizzando il software PLINK (pngu.mgh.harvard.edu/~purcell/plink) fornendo la seguente stringa:

"plink –ped <nome file.ped> --map <nome file.map> --allow-no-sex --assoc --ci 0.95 –out <nome file output>"

L'opzione <–allow-no-sex> fa si che i campioni per i quali non è stato indicato il sesso nel file di imput siano comunque considerati nell'analisi. L'opzione <ci 0.95> richiede al programma di calcolare l'intervallo di confidenza al 95% per il valore di OR.

Conversione dei dati nel formato "FILE FORMAT" Per le analisi di imputazione e la successiva meta-analisi è stato utilizzato un insieme di programmi disegnati appositamente per l'analisi dei GWAS (Genome-wide Association Study Sowtware), reperibili al sito web www.stats.ox.ac.uk/-marchini/software/gwas/gwas.html e funzionanti su piattaforma Linux. Gli stessi programmi sono stati utilizzati negli studi condotti dal Wellcome Trust Case-Control Consortium (WTCCC).

Tutti questi programmi utilizzano un unico formato di input dei dati (denominato "file format") che consiste in due parti:

1) Un "genotype file" che contiene i dati di genotipizzazione in formato "una linea per SNP". Le prime 5 informazioni per ciascuna linea consistono in: SNP ID, rs ID, posizione (in bp) allele A, allele B. I successivi tre numeri per ogni linea sono i valori di probabilità per i tre genotipi (AA, AB e BB) per quello SNP per il primo individuo della coorte, i successivi tre numeri sono riferiti al secondo individuo, e così via. Quando il dato genotipico è certo (perché derivato dalla tipizzazione dei campioni e non da analisi in silico) i valori di probabilità sono indicati come 1 o 0.

2) Un "sample file" contenente le informazioni per ciascun individuo, come sesso, status di malattia e proporzione di dati mancanti.

La conversione dei dati nel formato “file format” è stata effettuata con il software “GTOOL” del pacchetto “Genome-wide Association Study Sowtware”, utilizzando la modalità “PED to GEN Convertion Mode”. La stringa di comando utilizzata è:

“./gtool –P –ped <nome cartella>/<nome file.ped> --map <nome cartella>/<nome file.map> --discrete_phenotype 1”

Con l’opzione “discrete_phenotype” il programma crea file di output diversi per i due diversi fenotipi (pazienti o controlli). Se questa opzione è disattivata, il fenotipo è invece trattato come variabile continua.

Imputazione. L’analisi di imputazione è stata eseguita utilizzando il programma IMPUTE v2 del pacchetto “Genome-wide Association Study Sowtware”.Come pannello di riferimento sono stati utilizzati i dati di genotipizzazione del progetto 1000 genomi (famiglie CEPH).

La stringa di comando utilizzata è:

“./impute2 –m ./ <nome cartella>/genetic_map_chr4_combined_b36.txt –h ./ <nome cartella>/1kg_b36_aug09_ceu_chr4.hap –l ./ <nome cartella>/1kg_b36_aug09_ceu_chr4.legend –g ./ <nome cartella>/<nome file.gen> -int <lower> <upper> –Ne 1148 –iter 30 -o ./ <nome cartella>/<nome file output>”.

Segue la spiegazione delle diverse opzioni utilizzate:

<-m> (map file). Mappa di ricombinazione su scala fine del cromosoma da analizzare (cromosoma 4). Contiene tre colonne: posizione fisica (bp), tasso di ricombinazione tra la posizione corrente e la successiva (in cM/Mb) e posizione sulla mappa genetica (in cM). E’ stato scaricato dal sito del software IMPUTE.

<-h> (haplotype file) File di APLOTIPI noti (in questo caso ricavati dal progetto 1000 genomi), con una riga per SNP ed una colonna per aplotipo.

<-l> (legend file) Legenda contenente tutte le informazioni relative agli SNP dell’<-h> file. Contiene 4 colonne: rsID, posizione fisica (bp) allele 0 (A, C, G, T), allele 1 (A, C, G, T).

I file <-h> e <-l> costituiscono il pannello di riferimento (reference panel).

<-g> (inference panel) File contenente i dati di genotipizzazione noti

<-int> Intervallo genomico da usare per l'imputazione, specificato dai due limiti inferiore e superiore. Per esempio, nel caso di TEC, -int 46.0e6 49.0e6.

-<Ne> "Dimensioni effettive" della popolazione da cui è stata raccolta la nostra casistica. Comunemente si utilizza il valore di 1148 per i caucasoidi.

<-iter> Numero di iterazioni delle Markov chain Monte Carlo (MCMC) da effettuare. Il valore di default è 20, aumentandolo si ottengono imputazioni più accurate, ma si allunga di molto il tempo richiesto per ciascuna analisi (24-48 ore).

<-o> file di output.

Unione dei dati imputati con quelli tipizzati in unico file. Il file di output di IMPUTE contiene tutti gli SNP nell'intervallo considerato, tranne quelli genotipizzati, contenuti nel file di input originario, e il cui dato è certo e non probabilistico. Per poter effettuare la meta-analisi su casistiche tipizzate per marcatori diversi, è necessario unire casistica per casistica in un unico file i dati di imputazione con i dati di genotipizzazione originali.

Questo viene effettuato con il software "GTOOL" del pacchetto "Genome-wide Association Study Sowtware", utilizzando la modalità "Merge Mode", ch può essere utilizzata per riunire due o più dataset in GEN file format.

La stringa di comando utilizzata è:

"./gtool –M –g <nome cartella>/<nome file imputazione.impute2> -- g. <nome cartella>/<nome file genotipi.gen> --s <nome cartella>/<nome file.sample> -- s <nome cartella>/<nome file.sample> --threshold 0,9 –log <nome cartella>/<nome file output>".

Meta-analisi ed analisi di associazione sui dati imputati. Questa analisi è stata effettuata con il software SNPTEST del pacchetto "Genome-wide Association Study Sowtware", che nelle analisi tiene conto dell'incertezza del genotipo.

E' stata utilizzata la seguente stringa di comando:

"./snptest –cases ./ <nome cartella>/<file casi popolazione 1.gen> ./<nome cartella>/<file casi popolazione 1.sample> -controls ./ <nome cartella>/<file controlli popolazione 1.gen> ./<nome

cartella>/<file controlli popolazione 1.sample> –cases ./ <nome cartella>/<file casi popolazione 2.gen> ./<nome cartella>/<file casi popolazione 2.sample> -controls ./ <nome cartella>/<file controlli popolazione 2.gen> ./<nome cartella>/<file controlli popolazione 2.sample> –cases ./ <nome cartella>/<file casi popolazione 3.gen> ./<nome cartella>/<file casi popolazione 3.sample> -controls ./ <nome cartella>/<file controlli popolazione 3.gen> ./<nome cartella>/<file controlli popolazione 3.sample> -o ./ <nome cartella>/<file output> -hwe –frequentist –proper.

L’opzione <–hve> richiede di calcolare la probabilità relativa al’’equilibrio di Hardy-Weinberg, <-frequentist> richiede che nell’output compaiano i valori di p e <-proper> che questi ultimi siano corretti tenendo conto dell’incertezza del genotipo. Per l’assegnazione dei genotipi è stato utilizzato il valore soglia di default (0,9).

BIBLIOGRAFIA

[1].Adam C et al.. The effect of exogenous interferon: acceleration of autoimmune and renal diseases in (NZB/W) F1 mice. Clin Exp Immunol. 1980; 40:373-82
[2].Aguilar F et al. MCP-1 promoter polymorphism in Spanish patients with systemic lupus erythematosus. Tissue Antigens 2001; 58:335-338
[3].Aitman TJ et al. Copy number polymorphism in Fcgr3 predisposes to glomerulonephritis in rats and humans. Nature 2006; 439:851-855.
[4].Akahoshi M et al. Roles of genetic variations in signalling/ immunoregolatory molecules in susceptibility to systemic lupus erythematosus. Seminars in Immunology 2006; 18: 224-229
[5].Akira S et al. Role of adapters in Toll-like receptor signalling. Biochem. Soc. Trans. 2003; 31: 637–642
[6].Barcellos L et al. High density SNP screening of the Major Histocompatibility Complex in Systemic Lupus Erythematosus demonstrates strong evidence for independent susceptibility regions. PLoS Genetics 2009; 5:e1000696
[7].Bell JK et al. Leucine rich repeats and pathogen recognition in Toll-like receptors. TRENDS in Immunology 2003; 24(10):528-533
[8].Bennet L et al. Interferon and granulopoiesis signatures in systemic lupus erythematosus blood. J Exp Med 2003; 197:711-723
[9].Benucci M et al. Systemic lupus erythematosus (SLE) in Italy: an Italian prevalence study based on a two-step strategy in an area of Florence (Scandicci-Le Signe). Med Sci Monit. 2005; 11(9):CR420-5
[10].Boackle 2003. Complement and autoimmunity. Biomedicine & Pharmacology 2003; 57:269-273
[11].Borchers AT et al. The geoepidemiology of sistemi lupus erythematosus. Autoimmunity Reviews 2010; 9: A277-A287.
[12].Botto M et al. FcgRIIa polymorphism in systematic lupus erythematosus (SLE): no association with disease. Clinical and Experimental Immunology 1996; 104:264-268
[13].Botto M et al. Homozygous C1q deficiency causes glomerulonephritis associated with multiple apoptotic bodies (letter). Nature genetics 1998; 19: 56-59
[14].Brown et al. Mechanism of disease: the complement system in renal injury—new ways of looking at an old foe. Nat Clin Pract Nephrol. 2007 May;3(5):277-86
[15].Busiello R et al. Atypical features of familial hemophagocytic lymphohistiocytosis. Blood. 2004 Jun 15;103(12):4610-2. Epub 2004 Jan 22.
[16].Cambier JC, Johnson SA. Differential binding activity of ARTH1/TAM motifs. Immunol. Lett. 1995; 44(2-3):77-80
[17].Cannella S et al. Germline mutations of the perforin gene are a frequent occurrence in childhood anaplastic large cell lymphoma. Cancer. 2007 Jun 15;109(12):2566-71.
[18].Cappellano G et al. Variations of the perforin gene in patients with multiple sclerosis. Genes and Immunity 2008; 9: 438-444.
[19].Carrol MC. A protective role for innate immunity in systemic lupus erythematosus. Nature Reviews Immunology 2004; 4:825-831.
[20].Carrol MC. The lupus paradox. Nature Genetics 1998; 19: 3-4
[21].Cervera R et al. Systemic Lupus Erythematosus: Clinical and Immunologic Patterns of Disease expression in a cohort of 1,000 Patients. Medicine 1993; 72(2): 113-123
[22].Cervino AC et al. A genome-wide study of lupus: preliminary analysis and data release. Ann N Y Acad Sci 2007; 1110: 131-9
[23].Chaudary PM et al. Cloning and characterization of two Toll/interleukin-receptor-like genes TIL3 e TIL4: evidence for a multi-gene receptor family in humans. Blood 1998; 91:4020-4027
[24].Chen C et al. The sites and stage of anti-DNA B-cell deletion. Nature 1995; 373.
[25].Clark J et al. B lymphocyte hyperactivity in families of patients with systemic lupus erythematosus. J autoimmun 1996; 1: 59-65
[26].Clementi R et al. Variations of the perforin gene in patients with autoimmunity/lymphoproliferation and defective Fas function. Blood 2006; 108(9):3079-3084.

[27].Colten HR. Complement deficiencies. Annu. Rev. Immunol. 1992; 10: 809-834
[28].Crawley E et al. Polymorphic haplotypes of the interleukin-10 5' flanking region determine variable interleukin-10 transcription and are associated with particular phenotypes of juvenile rheumatoid arthritis. Arthritis & Rheumatism 1999; 6: 1101-1108
[29].Cunninghame DS et al. Association of IRF5 in UK SLE families identifies a variant involved in polyadenilation. Hum Mol Genet 2007; 16:579:91.
[30].D'Alfonso S et al. Association tests with Systemic lupus erythematosus (SLE) of IL10 markers indicate a direct involvement of a CA Repeat in the 5' regulatory region. Genes and Immunity 2002; 3(8):454-63
[31].D'Alfonso S et al. Systemic Lupus Erythematosus candidate genes in the Italian population. Evidence for a Significant Association with Interleukin-10. Arthritis & Rheumatism 2000; 43(1): 120-128
[32].De Jager PL et al. Integration of genetic risk factors into a clinical algorithm for multiple sclerosis susceptibility: a weighted genetic risk score. Lancet Neurol. 2009; 8(12):1111-9.
[33].Dean GS et al. Cytokines and systemic lupus erythematosus. Ann. Rheum. Dis. 2000; 59:243-251
[34].Deapen D et al. A revised estimate of twin concordance in Systemic Lupus Erythematosus.). Arthritis & Rheumatism 1992; 35(3): 311-318
[35].Delgado-Vega AM et al. Replication of the TNFSF4 (OX40L) promoter region association with systemic lupus erythematosus. Genes Immun. 2009; 10(3): 248-253
[36].Dragon-Durey MA et al. Lack of evidence of a specific role for C4A gene deficiency in determinating disease susceptibility among C4-deficient patients with systemic lupus erythematosus (SLE). Clinical and Experimental Immunology 2001; 123(1): 133-139
[37].Dragon-Durey MA et al. Molecular Basis of a Selective C1s Deficiency Associated with Early Onset Multiple Autoimmune Diseases. The Journal of Immunology 2001; 166: 7612-7616
[38].Duits AJ et al. Skewed distribution of IgG Fc receptor IIa (CD32) polimorphism is associated with renal disease in systemic lupus erythematosus. Arthritis & Rheumatism 1995; 38(12): 1832-1836
[39].Edberg JC et al. Genetic variation in the CRP promoter: association with systemic lupus erythematosus. Human Molecular Genetics 2008; 17(8): 1147-1155
[40].Edberg JC et al. Genetic variation in the CRP promoter: association with systemic lupus erythematosus. Hum Mol Genet. 2008; 17(8):1147-55.
[41].Fan Q et al. Complement component C4 deficiencies and gene alterations in patients with systemic lupus erythematosus. European Journal of Immunogenetics 1993; 20: 11-21
[42].Fanciulli M et al. FCGR3B copy number variation is associated with susceptibility to systemic, but not organ-specific, autoimmunity. Nat Genet 2007; 39:721-3.
[43].FanX et al. Upregulation of lymphoid and renal interferon-gamma mRNA in autoimmune MRL-Fas(lpr) mice with lupus nephritis. Inflammation 1997; 21:105-12
[44].Feldmann J et al. Functional consequencies of perforin gene mutations in 22 patients with familial haemophagocytic lynphohistiocytosis. British Journal of Haematology 2002; 117: 956-972
[45].Ferreiro-Neira I et al. Opposed independent effects and epistasis in the complex association of IRF5 to SLE. Genes Immun. 2007;8(5):429-38
[46].Fisher GH et al. Dominant interfering Fas gene mutations impair apoptosis in a human autoimmune lymphoproliferative syndrome. Cell 1995 81(6):935-46
[47].Fisher GH et al. Dominant interfering Fas gene mutetions impair apoptosis in a human autoimmune lynphoproliferative syndrome. Cell 1995; 81(6): 935-46
[48].Flierman R, Daha MR. The clearence of apoptotic cells by complement. Immunobiology 2007; 212:363-370
[49].Gabriel SB et al. The structure of haplotype blocks in the human genome. Science. 2002; 296(5576):2225-9.

[50].Gaffney PM et al. A genome-wide search for susceptibility genes in human systemic lupus erythematosus sib-pair families. Proceedings of the National Academy of Science of the United States of America 1998; 95(25): 14875-14879
[51].Gaffney PM et al. Genome Screening in Human Systemic Lupus Erythematosus: results from a Second Minnesota Cohort and Combined Analyses of 187 Sib-Pair Families. American Journal of Human Genetics 2000; 66: 547-556
[52].Gateva V et al. A large-scale replication study identifies TNIP1, PRDM1, JAZF1, UHRF1BP1 and IL10 as risk loci for systemic lupus erythematosus. Nat Genet. 2009; 41(11):1228-33
[53].Ginzler EM, Antoniadis I. Clinical manifestations of systemic lupus erythematosus, measures of disease activity, and long-term complications. Current Opinion in Rheumatology 1992; 4: 672-680
[54].Graham R et al. Genetic linkage and transmission disequilibrium of marker haplotypes at chromosome 1q41 in human sistemi lupus erythematosus. Arthritis research 2001; 3: 299-305
[55].Graham RR 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(5): 550-5.
[56].Graham RR et al. Genetic variants near TNFAIP3 on 6q23 are associated with systemic Lupus Erythematosus (SLE). Nature Genetics 2008; 40(9): 1059-1061
[57].Graham RR et al. Review of recent genome-wide association scans in lupus. Journal of International Medicine 2009; 265:280-688
[58].Graham RR et al. Three functional variants of IFN regulatory factor 5 (IRF5) define risk and protective haplotypes for human lupus. Proc Natl Acad Sci USA. 2007; 104(16): 6758-63.
[59].Gramaglia I et al. The OX40 costimulatory receptor determines the development of CD4 memory by regulating primary clonal expansion J Immunol 2000; 165: 3043-50.
[60].Gray-McGuire C et al. Genome Scan of Human Systemic Lupus Erythematosus by regression Modeling: Evidence of linKage and Epistasis at 4p16-15.2. American Journal of Human Genetics 2000; 67: 1460-1469
[61].Grennan DM et al. Family and twin studies in systemic lupus erythematosus. Dis Markers 1997; 13(2): 93-98
[62].Haas C et al. IFN-gamma receptor deletion prevents autoantibody production and glomerulonephritis in lupus-prone (NZB X NZW)F1 mice. J Immunol 1998; 160:3713-8
[63].Han JW et al. Genome-wide association study in a Chinese Han population identifies nine susceptibility loci for systemic lupus erythematosus. Nat Genet 2009: 41:1234-1237.
[64].Harada H et al. Structure and regulation of the human interferon regulatory factor 1 (IRF-1) and IRF-2 genes: implications for a gene network in the interferon system. Mol Cell Biol. 1994; 14(2):1500-9.
[65].Harada H et al. The role of interferon regulatory factors in the interferon system and cell growth control. Biochimie. 1998; 80(8-9):641-50.
[66].Harley JB et al. Genetic susceptibility to lupus: new insights from fine mapping and genome-wide association studies. Nat Rev Genet 2009; 10(5):285-290.
[67].Harley JB et al. Genome-wide association scan in women with systemic lupus erythematosus identifies susceptibility variants in ITGAM, PXK, KIAA1542 and other loci. Nat genet 2008; 40:204-210.
[68].Harley JB et al. Unraveling the genetics of systemic lupus erythematosus. Springer Semin Immunopathol. 2006; 28(2):119-30.
[69].Hartung K et al. Major Histocompatibility Complex Haplotypes and Complement C4 Alleles in Systemic Lupus Erythematosus. J Clin Invest. 1992; 90(4): 1346–1351
[70].Hawn TR et al. A common dominant TLR5 stop codon polymorphism abolishes flagellin signalling and is associated with susceptibility to legionnaires' disease. J Exp Med. 2003; 198(10):1563-72

[71].Hawn TR et al. A stop codon polymorphism of Toll-like receptor 5 is associated with resistance to systemic lupus erythematosus. PNAS 2005; 102(30): 10593-10597
[72].Hibbs ML et al. Multiple defects in the immune system of Lyn-deficient mice, culminating in autoimmune disease. Cell 1995; 83:301-11.
[73].Hochberg MC. Updating the American College of Rheumatology revised criteria for the classification of systemic lupus erythematosus [letter]. Arthritis Rheum 1997; 40: 1725
[74].Hom G et al. Association of Systemic Lupus Erythematosus with C8orf13-BLK and ITGAM-ITGAX. N Engl J Med 2008; 358: 900-909
[75].Hrycek A et al. Human cytomegalovirus in patients with systemic lupus erythematosus. Autoimmunity 2005; 38(7):487-91
[76].Jacob CO et al. Identification of IRAK1 as a risk gene with critical role in the pathogenesis of systemic lupus erythematosus. PNAS 2009; 106(15):6256-6261
[77].Jacob CO et al. Identification of novel susceptibility genes in childhood-onset systemic lupus erythematosus using a uniquely designed candidate gene pathway platform. Arthritis Rheum. 2007; 56(12):4164-73
[78].Jarjanazi H et al. Biological implications of SNPs in signal peptide domains of human proteins. Proteins 2008; 70:394-403
[79].Johanneson B, Alarcón-Riquelme ME. An update on the genetics of systemic lupus erythematosus. Isr Med Assoc J. 2001; 3(2): 88-93.
[80].Johansson CM et al. Chromosome 17p12-q11 harbors susceptibility loci for systemic lupus erythematosus. Hum Genet. 2004; 115(3): 230-8.
[81].Karassa FB et al. Role of the Fcgamma receptor IIa polymorphism in susceptibility to systemic lupus erythematosus and lupus nephritis: a meta-analysis. Arthritis Rheum. 2002; 46(6): 1563-71.
[82].Karassa FB et al. The FC gamma RIIIA-F158 allele is a risk factor for the development of lupus nephritis: a meta-analysis. Kidney Int 2003; 63:1475-1482
[83].Kelly JA et al. Evidence for a susceptibility gene (SLEH1) on chromosome 11q14 for systemic lupus erythematosus (SLE) families with hemolytic anemia. Proc Natl Acad Sci U S A. 2002; 99(18):11766-71.
[84].Kelly JA et al. The genetics of systemic lupus erythematosus: putting the pieces together. Genes Immun. 2002; 3 Suppl 1:S71-85.
[85].Kirschfink M et al. Complete functional C1q deficiency associated with systemic lupus erythematosus (SLE). Clinical and Experimental Immunology. 1993; 94: 267-272
[86].Klickstein, LB et al. Identification of distinct C3b and C4b recognition sites in the human C3b/C4b receptor (CR1, CD35) by deletion mutagenesis. J. Exp. Med. 1988; 168:1699.
[87].Komar AA. Genetics. SNPs, silent but not invisibile. Science 2007; 315(5811): 466-7.
[88].Kotzin BL. Systemic Lupus Erythematosus Cell 1996; 85: 303-306
[89].Kozyrev SV et al. Functional variants in the B-cell gene BANK1 are associated with systemic lupus erythematosus. Nature Genetics 2008 February; 40(2):211-484
[90].Krishnan S et al. Autoimmunity in systemic lupus erythematosus: Integrating genes and biology. Seminars in Immunology 2006: 18:230-243
[91].Kroner A et al. A PD-1 polymorphism is associated with disease progression in multiple sclerosis. Ann. Neurol. 2005: 58:50-57
[92].Krych M et al. Sites within the complement C3b/C4b receptor important for the specificity of ligand binding. Proc. Natl. Acad. Sci. U.S.A. 1991; 88:4353.
[93].Lander ES, Kruglyak L. Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. Nature Genet 1995; 11:241-247
[94].Lehtinen DA et al. The TREX1 double-stranded DNA degradation activity is defective in dominant mutations associated with autoimmune disease J Biol Chem. 2008;283(46):31649-56
[95].Lee YH et al. CTLA-4 polymorphisms and systemic lupus erythematosus (SLE): a meta-analysis. Hum Genet 2005; 116:361-367

[96].Lee YH et al. The mannose-binding lectin gene polymorphisms and systemic Lupus Erythematosusl. Two case-control studies and a meta-analysis. Arthritis e Rheumatism 2005; 52(12): 3966-3974
[97].Lee YH et al. The PTPN22 C1858T functional polymorphism and autoimmune diseases--a meta-analysis. Rheumatology (Oxford). 2007; 46(1): 49-56.
[98].Lee-Kirsch MA et al. Familial chilblain lupus, a monogenic form of cutaneous lupus erythematosus, maps to chromosome 3p. The American Journal of Human Genetics 2006; 79: 731-737
[99].Lee-Kirsh MA et al. A mutation in TREX1 that impairs susceptibility to granzyme A-mediated cell death underlies familial chilblain lupus. Journal molecular medicine 2007; 85: 531-537
[100].Lee-Kirsh MA et al. Mutations in the gene encoding the 3'-5' DNA exonuclease TREX1 are associated with systemic lupus erythematosus. Nature Genetics 2007; 39: 1065-1067
[101].Lehtinen DA et al. The TREX1 double-stranded DNA degradation activity is defective in dominant mutations associated with autoimmune disease. J Biol Chem. 2008; 283(46): 31649-56
[102].Li X et al. A novel polymorphism in the Fc gamma receptor IIB (CD32B) transmembrane region alters receptor signalling. Arthritis Rheum 2003; 48: 3242-3252
[103].Lin T et al. The role of complement and Toll-like receptors in organ transplantation. Transplant International ISSN 2007; 20: (481-489)
[104].Lindqvist & Alarcón-Riquelme. The Genetics of Systemic Lupus Erythematosus (Review). Scandivian Journal of Immunology 1999; 50(6): 562-577
[105].Lindqvist AB et al. A Susceptibility Locus for Human Systemic Lupus Erythematosus (hSLE1) on Chromosome 2q. Journal of Autoimmunity 2000; 14: 169-178
[106].Liu K et al. Kallikrein genes are associated with lupus and glomerular basement membrane-specific antibody-induced nephritis in mice and humans. J Clin Invest. 2009; 119(4): 911-923
[107].Logar CM et al. A human CR1-like transcript containing sequence for a binding protein for iC4 is expressed in hematopoietic and fetal lymphoid tissue. Mol Immunol. 2004; 40(11):831-40.
[108].Machold KP, Smolen JS. Interferon-gamma induced exacerbation of systemic lupus erythematosus. J Rheumatol. 1990; 17(6):831-2
[109].Magistrelli G et al. A soluble form of CTLA-4 generated by alternative splicing is expressed by nonstimulated human T cells. Eur. J. Immunol. 1999: 29: 3596-3602
[110].Manger K et al. Fcg receptors IIa polymorphism in caucasian patients with systemic lupus erythematosus. Association with Clinical Syntoms. Arthritis & Rheumatism 1998; 41(7):1181-1189
[111].Mäurer M et al. A polymorphism in the human cytotoxic T-lymphocyte antigen 4 (CTLA4) gene (exon 1 +49) alters T-cell activation. Immunogenetics. 2002; 54(1):1-8.
[112].Miles S, Isenberg D. A review of serological abnormalities in relatives of SLE patients. Lupus 1993; 2(3): 145-150
[113].Mills JA. Systemic Lupus Erythematosus Medical Progress 1994; 330(26): 1871-1876
[114].Molleran Lee S et al. Characterization of diverse PRF1 mutations leading to decreased natural killer cell activity in North American families with haemophagocytic lymphohistiocytosis. J Med Genet 2004; 41:137-144.
[115].Morgan BP, Walport MJ. Complement deficiency and disease (overview). Immunology Today 1991; 12(9): 301-306
[116].Morita M et al. Gene-targeted mice lacking the Trex1 (DNase III) 3'→5' DNA exonuclease develop inflammatory myocarditis.Molecular and Cellular Biology 2004; 24: 6719-6727
[117].Moser KL et al. Genome scan of lupus erythematosus: Evidence for linkage on chromosome 1q in African-American pedigrees. Proceedings of the National Academy of Science of the United States of America 1998; 95(25): 14869-14874
[118].Mountz JD, Edwards CK 3rd. Murine models of autoimmune disease. Curr Opin Rheumatol. 1992; 4(5): 621-9.

[119].Musone S et al. Multiple polymorphisms in the TNFAIP3 region are independently associated with systemic lupus erythematosus. Nature Genetics 2008; 40(9): 1062-1064
[120].Nakashima H et al. The combination of polymorphisms within interferon-y receptor 1 and receptor 2 associated with the risk of systemic lupus erythematosus. FEBS 1999; 453(1-2):187-190
[121].Namijou B et al. Genome scan stratified by the presence of anti-double stranded DNA (dsDNA) autoantibody in pedigrees multiplex for systemic lupus erythematosus (SLE) establishes linkages at 19p13.2 (SLED1) and 18q21.1 (SLED2). Genes and Immunity 2002; 3(1):S35-S41
[122].Nath SK et al. A non synonimous functional variant in integrin-alpha(M) (encoded by ITGAM) is associated with systemic lupus erythematosus. Nat Genet 2008; 40:152-4
[123].Nath SK et al. Genetics of human systemic lupus erythematosus: the emerging picture. CurrOpin Immunol 2004; 16:794-800 236.
[124].Nath SK et al. Polymorphisms of complement receptor 1 and interleukin-10 genes and systemic lupus erythematosus: a meta-analysis. Hum Genet 2005; 118:225-234 234.
[125].Nishimura H et al. Autoimmune dilated cardiomyopathy in PD-1 receptor-deficient mice. Science 2001; 291:319-322
[126].Nishimura H et al. Development of lupus-like autoimmune diseases by disruption of the PD-1 gene encoding an ITIM motifcarrying immunoreceptor. Immunity 1999; 11:141-51
[127].Nishimura H, Honjo T. PD-1: an inibitory immunoreceptor involved in peripheral tolerance. Trends Immunol 2001; 22:256-8
[128].Oishi T et al. A functional SNP in the NKX2.5-binding site of ITPR3 promoter is associated with susceptibility to systemic lupus erythematosus in Japanese population.
[129].Ophoff RA et al. Hereditary vascular retinopathy, cerebroretinal vasculopathy, and hereditary endotheliopathy with retinopathy, nephropathy, and stroke map to a single locus on chromosome 3p21.1-p21.3. Am J Hum Genet. 2001; 69(2):447-53.
[130].Orilieri E et al. Variations of the perforin gene in patients with type 1 diabetes. Diabetes 2008; 57(4):1078-83
[131].Orozco G et al. Association of a functional single-nucleotide polymorphism of PTPN22, encoding lymphoid protein phosphatase, with rheumatoid arthritis and systemic lupus erythematosus. Arthritis Rheum. 2005; 52(1):219-24.
[132].Orrù V et al. A loss-of-function variant of PTPN22 is associated with reduced risk of systemic lupus erythematosus. Human Molecular Genetics 2009; 18(3): 569-579
[133].Pinkoski MJ & Green DR. Granzyme A: the road less travelled. Nature Immunology 2003; 4(2):107-108
[134].Prokunina L et al. A regulatory polymorphism in PDCD1 is associated with susceptibility to systemic lupus erythematosus in humans. Nature Genetics 2002; 32: 666-669
[135].Prokunina L et al. The systemic lupus erithematosus-associated PDCD1 polymorphism PD1.3A in lupus nephritis. Arthritis & Rheumatism 2004; 50(1):327-333
[136].Prokunina L, Alarcon-Riquelme M. The genetic basis of systemic lupus erythematosus – knowledge of today and thoughts for tomorrow. Human Molecular Genetics 2004; 13(1) R143-R148
[137].Ramantani G et al. Expanding the phenotypic spectrum of lupus erythematosus in Aicardi-Goutières syndrome. Arthritis Rheum. 2010; 62(5):1469-77.
[138].Remmers EF et al. STAT4 and the risk of rheumatoid arthritis and systemic lupus erythematosus. N Engl J Med 2007; 357:977-86
[139].Remmers RF et al. STAT4 and the risk of rheumatoid arthritis and systemic lupus erythematosus Am J Hum Genet 2005; 76:528-37
[140].Rhodes B, Vyse T. Using genetics to deliver personalized SLE therapy – a realistic prospect? Nature reviews rheumatology 2010 6: 373-377
[141].Rice G et al. Heterozygous mutations in TREX1 cause familial chilblain lupus and dominant Aicardi-Goutieres syndrome. The American Journal of Human Genetics 2007; 80: 811-815

[142].Rice G et al. Clinical and molecular phenotype of Aicardi-Goutieres syndrome The American Journal of Human Genetics 2007; 81: 713-725
[143].Richards A et al. C-terminal truncations in human 3'-5' DNA exonuclease TREX1 cause autosomal dominant retinal vasculopathy with cerebral leukodystrophy. Nature Genetics 2007; 39: 1068-1070
[144].Rieux-Laucat F et al. Mutations in Fas associated with human lynphoproliferative syndrome and autoimmunity. Science 1995; 268(5215): 1347-1349 269.
[145].Risma KA et al. Aberrant maturation of mutant perforin underlies the clinical diversity of hemophagocytic lymphohistiocytosis. The Journal of Clinical Investigation 2006; 116(1):182-192
[146].Rovin BH et al. A novel polymorphism in the MCP-1 gene regulatory region that influences MCP-1 expression. Biochem Biophys Res Commun 1999; 295:344-348
[147].Rupert KL et al. The molecular basis of complete complement C4A and C4 deficiencies in a systemic lupus erythematosus patient with homozygous C4A and C4B mutant genes. J Immunol 2002; 169(3): 1570-8
[148].Russell AI et al. Polymorphism at the Creactive protein locus influences gene expression and predisposes to systemic lupus erythematosus. Human Molecular Genetics 2004; 13(1):137-147
[149].Salmon JE et al. FcgRIIa Allels Are Hereditable Risk Factors for Lupus Nephritis in African Americans. The Journal of Clinical investigation 1996; 97(5):1348-1354
[150].Sánchez E et al. Identification of a new putative functional IL18 gene variant through an association study in systemic lupus erythematosus. Hum Mol Genet. 2009; 18(19):3739-48.
[151].Santulli-Marotto S et al. Autoreactive B Cell Regulation: Induction of Developmental Arrest by Lupus-Associated autoantigens. Immunity 1998; 8: 209-219
[152].Science. 1995 Nov 10; 270(5238): 985-8.
[153].Seguardo OG et al. Combined total deficiency of C7 e C4B with systemic lupus erythematosus (SLE). Clinical and Experimental Immunology 1992; 87: 410-414
[154].Shinohara T et al. Structure and chromosomal localization of the human PD-1 gene (PDCD1). Genomics 1994; 23: 704-706
[155].Shirai T et al. Genome Screening fot Susceptibility Loci in Systemic Lupus Erythematosus. American Journal of Pharmacogenomics 2002; 2(1): 1-12
[156].Sigurdsson S et al. Polymorphisms in the tyrosine kinase 2 and interferon regulatory factor 5 genes are associated with systemic lupus erythematosus. Am.J.Hum.Genet. 2005; 76:528-537
[157].Sigurdsson S et al. Polymorphisms in the tyrosinr kinase 2 and interferon regulatory factor 5 genes are associated with Systemic Lupus Erythematosus. A. J. Hum. Genet. 2005; 76:528-537
[158].Simard JF, Costenbader KH. What can epidemiology tell us about systemic lupus erythematosus. Int J Clin Pract 2007; 61:1170-1180
[159].Slingsby JH et al. Homozigous hereditary C1q deficiency and systemic lupus erythematosus. A new Family and Molecular Basis of C1q Deficiency in Three Families. Arthritis & Rheumatism 1996; 39(4): 663-670
[160].Smyth LJ et al. Fc gamma RIIa polymorphism in systemic lupus erythematosus. Ann. Rheum. Dis. 1997; 56: 744-746
[161].Solomou EE et al. Perforin gene mutations in patients with acquired aplastic anemia. Blood. 2007 Jun 15; 109(12):5234-7.
[162].Song YW et al. Abnormal distribution of Fcg receptors type IIa polymorphisms in korean patients with systemic lupus erythematosus. Arthritis & Rheumatism 1998; 41(3):421-426
[163].Stadt UZ et al. Mutation spectrum in children with primary hemophagocytic lymphohistiocytosis: molecular anf functional analyses of PRF1, UNC13D, STX11, and RAB27A. Human Mutation 2006; 27(1):62-68
[164].Stepp SE et al. Perforin gene defects in familial hemophagocytic lymphohistiocytosis. Science. 1999; 286(5446):1957-9.
[165].Stetson DB et al. Trex1 prevents cell-intrinsic initiation of autoimmunity. Cell 2008; 134: 587-598

[166].Sturfelt G, Truedsson L. Complement and its breakdown products in SLE. Rheumatology 2005; 44:1227-1232.
[167].Subcommittee for Scleroderma Criteria of the American Rheumatism Association Diagnostic and Therapeutic Criteria Committee. Preliminary criteria for the classification of systemic sclerosis (scleroderma). Arthritis Rheum. 1980; 23: 581–590.
[168].Takaoka A et al. Integral role of IRF-5 in the gene induction programme activated by Toll-like receptors. Nature. 2005; 434(7030): 243-9.
[169].Tanaka Y et al. Association of the interferon-gamma receptor variant (Val14Met) with systemic lupus erythematosus. Immunogenetics. 1999;49(4):266-71
[170].Taniguchi T et al. IRF family of transcription factors as regulators of host defense. Annu Rev Immunol. 2001;19:623-55.
[171].Tas SW et al. C1q and C4b bind simoultaneously to CR1 and additively support erythrocyte adhesion. The Journal of Immunology 1999; 163:5056-5063
[172].Tivol EA et al. Loss of CTLA-4 leads to massive lymphoproliferation and fatal multiorgan tissue destruction, revealing a critical negative regulatory role of CTLA-4. Immunity 1995: 3:541-7
[173].Trambas C et al. A single amino acid change, A91V, leads to conformational changes that can impari processing to the active form of perforin. Blood 2005; 106(3):932-937.
[174].Trendelenburg M et al. Lack of occurence of severe lupus nrphritis among anti-C1q autoantibody-negative patients. Arthritis Rheumatism 1999; 42(1):187-8
[175].Trizzino A et al. Genotype-phenotype study of familial haemophagocytic lymphohistiocytosis due to perforin mutations. J Med Genet. 2008; 45(1):15-21.
[176].Tsao BP et al. PARP alleles within the linked chromosomal region are associated with systemic lupus erythematosus. The Journal of Clinical Investigation 1999; 103(8): 1135-1140
[177].Tsao BP. The genetics of human systemic lupus erythematosus. Trends Immunol. 2003; 24(11):595-602.
[178].Turner MW. Deficiency of mannan binding protein-a new complement deficiency syndrome. Clin Exp Immun 1991; 86(s1):53-56
[179].Udesh de Silva et al. The crystal structure of TREX1 explains the 3' nucleotide specificity and reveals a polyproline II helix for protein partnering Journal of Biological Chemistry 2007; 282: 10537-10543
[180].Van der Pol WL, van der Winkel JGJ. IgG receptor polymorphisms: risk factors for disease. Immunogenetics 1998; 48:222-232
[181].Vastert B et al. Mutations in the perforin gene may be linked to macrophage activation syndrome in patients with systemic onset Juvenile Idiopathic Arthritis. The Turkish Journal of Pediatrics 2008; 50 (Supplement):2.
[182].Via CS, Handwerger BS. T-cell and B-cell function in lupus. Current Opinion in Rheumatology 1992; 4: 630-634
[183].Voskoboinik I et al. A functional analysis of the putative polymorphisms A91V and N252S and 22 missense perforin mutations associated with familial hemophagocytic lymphohistiocytosis. Blood 2005; 105(12): 4700-4706.
[184].Voskoboinik I et al. Perforin-mediated target-cell death and immune homeostasis. Nature Reviews Immunology 2006; 6:940-952
[185].Vyse TJ, Kotzin BL. Genetic susceptibility to Systemic Lupus Erythematosus. Annu. Rev. Immunol. 1998; 16:261-92
[186].Wallace JD, Hahn BH. Duboi's Lupus erythematosus. 1993; Lea & Febiger, Philadelphia and London; 4th edition
[187].Waterhouse P et al. Lymphoproliferative disorders with early lethality in mice deficient in Ctla-4.
[188].Wong M, Tsao BP. Current topics in human SLE genetics. Springer Seminars in Immunopatology 2006

[189].Wong M, Tsao BP. Current topics in human SLE genetics. Springer Semin Immunopathol. 2006; 28(2):97-107.
[190].Yamada M et al. Complement C1q regulates LPS-induced cytokine production in bone marrowderived dendritic cells. Eur. J. Immunol. 2004; 34:221-230
[191].Yamada R, Ymamoto K. Recent findings on genes associated with inflammatory disease. Mutation Research 2005; 573:136-151
[192].Yang W et al. Genome-wide association study in Asian populations identifies variants in ETS1 and WDFY4 associated with systemic lupus erythematosus. PLos Genet 2010; 6: e1000841
[193].Yang YG et al. Trex1 exonuclease degrades ssDNA to prevent chronic checkpoint activation and autoimmune disease. Cell. 2007; 131(5):873-86.

RINGRAZIAMENTI

Ringrazio le professoresse Sandra D'Alfonso e Mara Giordano per avermi dato l'opportunità di seguire questo dottorato di ricerca.

Ringrazio tutti i gruppi italiani e stranieri con cui ho avuto l'opportunità di collaborare durante questi tre anni e che hanno fornito un apporto importante ai risultati presentati in questa trattazione. In particolare vorrei ringraziare il gruppo svedese della professoressa Marta Alarcon-Riquelme e il gruppo di ricerca del professor Umberto Dianzani.

Infine vorrei ringraziare tutte le persone che grazie al loro supporto economico, o grazie al loro impegno scientifico, politico o sociale fanno si che anche in questi momenti così difficili per il mondo scientifico ed accademico la ricerca scientifica in Italia possa continuare ad esistere. A tutti loro va il mio più sentito "grazie". La ricerca scientifica non deve morire.

ARTICOLI ALLEGATI

nature genetics

Functional variants in the B-cell gene *BANK1* are associated with systemic lupus erythematosus

Sergey V Kozyrev[1,15], Anna-Karin Abelson[1,15], Jerome Wojcik[2], Ammar Zaghlool[1], MV Prasad Linga Reddy[1], Elena Sanchez[3], Iva Gunnarsson[4], Elisabet Svenungsson[4], Gunnar Sturfelt[5], Andreas Jönsen[5], Lennart Truedsson[6], Bernardo A Pons-Estel[7], Torsten Witte[8], Sandra D'Alfonso[9], Nadia Barizzone[9], Maria Giovanna Danieli[10], Carmen Gutierrez[11], Ana Suarez[11], Peter Junker[12], Helle Laustrup[12], Maria Francisca González-Escribano[13], Javier Martin[3,14], Hadi Abderrahim[2] & Marta E Alarcón-Riquelme[1]

Systemic lupus erythematosus (SLE) is a prototypical autoimmune disease characterized by production of autoantibodies and complex genetic inheritance[1–3]. In a genome-wide scan using 85,042 SNPs, we identified an association between SLE and a nonsynonymous substitution (rs10516487, R61H) in the B-cell scaffold protein with ankyrin repeats gene, *BANK1*. We replicated the association in four independent case-control sets (combined $P = 3.7 \times 10^{-10}$; OR = 1.38). We analyzed *BANK1* cDNA and found two isoforms, one full-length and the other alternatively spliced and lacking exon 2 (Δ2), encoding a protein without a putative IP3R-binding domain. The transcripts were differentially expressed depending on a branch point–site SNP, rs17266594, in strong linkage disequilibrium (LD) with rs10516487. A third associated variant was found in the ankyrin domain (rs3733197, A383T). Our findings implicate *BANK1* as a susceptibility gene for SLE, with variants affecting regulatory sites and key functional domains. The disease-associated variants could contribute to sustained B cell–receptor signaling and B-cell hyperactivity characteristic of this disease.

We genotyped 279 Swedish individuals with SLE and 515 control individuals using the 100K Affymetrix SNP array. As our purpose was to identify non-MHC genes and, if possible, important functional polymorphisms involved in SLE pathogenesis, we carried out an analysis of the genomic location of the associated SNPs, focusing on nonsynonymous substitutions. Among all associated SNPs, one (rs10516487) led to a substitution of arginine to histidine at amino-acid position 61 (R61H) of the BANK1 protein (allelic association, $P = 6.4 \times 10^{-3}$; genotypic association, $P = 2.01 \times 10^{-2}$). This SNP was ranked 679th in the allelic and 2,148th in the genotypic tests across the genome scan, with estimated false-discovery rates of 71.1% and 77.5%, respectively[4]. Four other SNPs in *BANK1* also showed association (**Supplementary Table 1** online). Because of the described B cell–specific expression of *BANK1* and its potential role in B cell receptor–mediated activation, we pursued this gene[5,6].

To provide better SNP coverage and refine the association signal, we genotyped 30 SNPs spanning the 284-kb *BANK1* genomic region (including the scan SNPs) in the Swedish SLE case and control samples. Two SNPs were not polymorphic, and nine SNPs were associated (**Table 1** bold). All associated SNPs were located between introns 1 and 7 (**Table 1**, **Supplementary Table 2** and **Supplementary Fig. 1** online).

Next, we carried out a detailed analysis of *BANK1* expression and structure. We observed that *BANK1* is indeed primarily expressed in $CD19^+$ B cells, with very low expression in other cell populations (**Fig. 1a**). To clone *BANK1* for functional analysis, we amplified full-length cDNA using distal primers. Of note, we detected two main bands following gel electrophoresis of the PCR products (**Fig. 1b**). *BANK1* is known to have two full-length alternative isoforms containing exon 1A or exon 1B[5]. Through subsequent cloning and sequencing, we identified a previously unknown isoform with an in-frame deletion of the entire exon 2 (Δ2 isoform). We analyzed cDNA from 83 healthy individuals and 30 individuals with SLE and found that this isoform was present in every sample, indicating that it is constitutively spliced. Moreover, we detected this isoform in cDNA from chimpanzee and mouse spleen (data not shown), suggesting that it is conserved across species.

[1]Department of Genetics and Pathology, Rudbeck Laboratory, Uppsala University, Uppsala 75185, Sweden. [2]Merck Serono, Chemin des Mines, Geneva 12202, Switzerland. [3]Instituto de Biomedicina López-Neyra, Granada 18100, Spain. [4]Department of Medicine, Unit of Rheumatology, Karolinska University Hospital, Solna 17176, Sweden. [5]Department of Rheumatology, University of Lund, Lund 22200, Sweden. [6]Department of Clinical Immunology and Microbiology, University of Lund, Lund 22200, Sweden. [7]Sanatorio Parque, Rosario 2000, Argentina. [8]Medical School Hannover, Hannover 30625, Germany. [9]Department of Medical Sciences and Interdisciplinary Research Center of Autoimmune Diseases (IRCAD), University of Eastern Piedmont, Novara 13100, Italy. [10]Dipartimento di Scienze Mediche e Chirurgiche, Università Politecnica delle Marche, Ancona 60121, Italy. [11]Department of Functional Biology, Hospital Universitario Central de Asturias, Universidad de Oviedo, Oviedo 33006, Spain. [12]Department of Internal Medicine C, Section of Rheumatology, Odense University Hospital, Odense 5000, Denmark. [13]Departamento de Inmunología, Hospital Virgen del Rocío, Sevilla 41013, Spain. [14]Consejo Superior de Investigaciones Científicas (CSIC), Grenada 18100, Spain. [15]These authors contributed equally to this work. Correspondence should be addressed to M.E.A.-R. (marta.alarcon@genpat.uu.se).

Received 4 September 2007; accepted 12 December 2007; published online 20 January 2008; corrected after print 27 March 2008; doi:10.1038/ng.79

Table 1 Association analysis of *BANK1* SNPs in Swedish SLE cases and controls

SNP	Associated allele	χ^2	*P*
rs7675129	T	0.155	0.6933
rs11726012	G	0.361	0.5479
rs11097755	C	1.433	0.2313
rs4522865	A	6.757	**0.0093**
rs4496585	A	2.618	0.1057
rs4572885	T	5.113	**0.0238**
rs10516487	G	8.091	**0.0044**
rs10516486	C	9.62	**0.0019**
rs17200824	A	4.265	**0.0389**
rs6849308	C	7.268	**0.007**
rs10516482	C	8.625	**0.0033**
rs10516483	C	11.437	**0.0007**
rs10516484	A	0.595	0.4404
rs4493533	C	3.184	0.0744
rs3733197	G	0.692	0.4054
rs2631271	G	6.356	**0.0117**
rs2850390	C	1.225	0.2684
rs2631265	C	0.016	0.8997
rs2631267	G	0.1	0.7524
rs2631268	T	1.446	0.2292
rs10516491	C	2.424	0.1195
rs1872701	G	1.842	0.1747
rs2850393	T	0.061	0.8053
rs2850396	C	0.682	0.4088
rs10516490	G	0.329	0.5665
rs10516489	T	0.338	0.5609
rs10516488	G	0.561	0.4538
rs1395306	T	1.97	0.1604

We carried out quantitative analysis of isoform expression in peripheral blood mononuclear cells. As exon 1B transcript was present at very low concentrations (data not shown), we continued the analysis, measuring common (exon 1A and exon 1B) full-length isoform concentrations. We noticed that the ratio of the full-length isoform to Δ2 (FL/Δ2) was not constant, which would be expected if Δ2 were equally expressed regardless of the genotypes of the samples. On the contrary, samples could be divided into groups according to the FL/Δ2 isoform ratio. After closely examining the genomic sequences surrounding exon 2, where putative signals affecting splicing could be located, we identified one SNP, rs17266594, located in the branch-point site. When we re-grouped the expression data, we observed a clear difference between genotypes (**Fig. 1c**). Individuals homozygous for the T allele and having the classical structure of the branch-point site[7] (YNY*T*G**A**YYN) showed higher expression of the full-length isoform; this expression was significantly suppressed (up to 40%) in homozygotes for the minor allele C, with concomitant upregulation of the Δ2 isoform expression. Total *BANK1* expression was not significantly affected by the SNP (**Fig. 1d**).

To determine whether other polymorphisms might contribute to the alternative splicing of exon 2, we sequenced the proximal promoter regions, exon 1A, exon 1B and exon 2 and 500 bp upstream and downstream of these exons, in 24 individuals with SLE and 8 controls. However, we found no previously unidentified SNPs in these regions that could be functional. Next, we identified five nonsynonymous substitutions in *BANK1* from the SNP databases. Although most were nonpolymorphic in our samples, we identified one polymorphic SNP, rs3733197, causing an alanine to threonine substitution at amino acid position 383 (A383T) in exon 7, which encodes the ankyrin repeat-like motif (**Supplementary Table 3** online).

To extend our association analyses, we genotyped rs10516487 (R61H), rs17266594 (branch-point variant) and rs3733197 (A383T) in four additional sets of SLE cases and controls (**Table 2**). We corroborated the genetic association with SLE for all three SNPs, although there were differences between individual populations. Using homogeneity and combinability tests according to the Breslow-Day method, we carried out a meta-analysis comprising 3,971 individuals. We then used the Mantel-Haenszel test to calculate pooled odds ratios of 1.38 ($P = 3.74 \times 10^{-10}$), 1.42 ($P = 4.74 \times 10^{-11}$) and 1.23

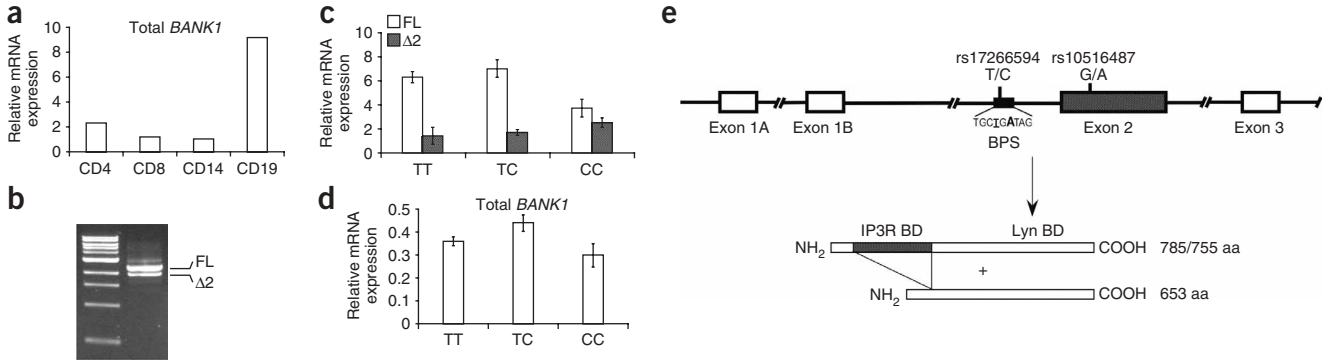


Figure 1 Correlation of rs17266594 genotypes with differences in FL/Δ2 isoform ratio of *BANK1*. (**a**) Expression of total *BANK1* mRNA in Clontech human blood fractions: $CD4^+$ and $CD8^+$ T cells, $CD14^+$ myeloid cells and $CD19^+$ B cells. (**b**) RT-PCR of the coding sequence of *BANK1* amplified from total human spleen cDNA shows two bands on a gel. Ladder (1 kb, New England Biolabs) is shown on the left. We confirmed the identity of both bands (2.3-kb upper band and 1.9-kb smaller band) by sequencing. (**c**) Relative mRNA expression of the full-length and Δ2 isoforms, as determined by quantitative real-time RT-PCR on total RNA purified from human PBMCs of healthy controls and cases. Data represent mean ± s.d. We analyzed 39 individuals with TT for the branch point–site SNP, 34 with TC and 10 with CC. Full-length transcript: TT versus CC, $P = 0.0004$ (Student's *t*-test); Δ2 transcript: TT versus CC, $P = 0.0088$. (**d**) Total *BANK1* expression did not correlate with genotypes of rs17266594. TT versus CC, $P = 0.229$. (**e**) Schematic structure of the 5′ end of *BANK1*. rs17266594, located in the branch point site of intron 1, alters splicing efficiency of the full-length and Δ2 transcripts. rs10516487 results in nonsynonymous substitution R61H. Alternative splicing gives rise to two isoforms, full-length and Δ2 with an in-frame deletion of entire exon 2. The short protein isoform lacks the putative domain for IP3R binding. IP3R BD, inositol 1,4,5-triphosphate receptor binding domain; Lyn BD, tyrosine kinase Lyn binding domain.

Table 2 Genotypic and allelic association of rs10516487 (R61H), rs17266594 and rs3733197 in five sets of SLE cases and controls and joint analysis with Mantel-Haenszel test

	Population	*n*	GG	GA	AA	χ^2	*P*	Odds ratio (CI)[a]	Allele G	Allele A	*P*	Odds ratio (CI)[b]
rs10516487	Scandinavia	Cases (536)	309 (57.6%)	200 (37.3%)	27 (5.0%)	11.7874	0.0028	GG: 2.12 (1.29–3.47)	818 (76.3%)	254 (23.7%)	7.3×10^{-4}	1.39 (1.14–1.68)
		Controls (565)	276 (48.8%)	238 (42.1%)	51 (9.0%)			GA: 1.59 (0.96–2.63)	790 (69.9%)	340 (30.1%)		
	Argentina	Cases (255)	164 (64.3%)	75 (29.4%)	16 (6.3%)	3.8013	0.1495	GG: 1.41 (0.73–2.72)	403 (79.0%)	107 (21.0%)	0.0564	1.31 (0.98–1.74)
		Controls (337)	190 (56.4%)	121 (35.9%)	26 (7.7%)			GA: 1.01 (0.51–2.00)	499 (74.3%)	173 (25.7%)		
	Germany	Cases (312)	181 (58.0%)	118 (37.8%)	13 (4.2%)	11.8503	0.0027	GG: 2.60 (1.32–5.14)	480 (76.9%)	144 (23.1%)	8.13×10^{-4}	1.52 (1.18–1.95)
		Controls (360)	166 (46.1%)	163 (45.3%)	31 (8.6%)			GA: 1.73 (0.87–3.44)	495 (68.8%)	225 (31.2%)		
	Italy	Cases (279)	166 (59.5%)	100 (35.8%)	13 (4.7%)	7.5139	0.0234	GG: 2.49 (1.22–5.09)	432 (77.4%)	126 (22.6%)	0.0078	1.46 (1.09–1.94)
		Controls (245)	123 (50.2%)	98 (40.0%)	24 (9.8)			GA: 1.88 (0.91–3.91)	344 (70.2%)	146 (29.8%)		
	Spain	Cases (702)	414 (59.0%)	243 (34.6%)	45 (6.4%)	11.3579	0.0034	GG: 1.26 (0.77–2.06)	1,071 (76.3%)	333 (23.7%)	0.0065	1.30 (1.07–1.58)
		Controls (446)	219 (49.1%)	197 (44.2%)	30 (6.7%)			GA: 0.82 (0.50–1.35)	635 (71.2%)	257 (28.8%)		
	Pooled	Cases (2,003)	1,187 (59.3%)	706 (35.2%)	110 (5.5%)				3,080 (76.9%)	926 (23.1%)	3.74×10^{-10}[c]	1.38 (1.25–1.53)
		Controls (1,968)	974 (49.9%)	817 (41.8%)	162 (8.3%)				2,763 (70.8%)	1,141 (29.2%)		
	Population		TT	CT	CC	χ^2	*P*	Odds ratio (CI)[a]	Allele T	Allele C	*P*	Odds ratio (CI)
rs17266594	Scandinavia	Cases (511)	296 (57.9%)	189 (37.0%)	26 (5.1%)	9.4399	0.0089	TT: 2.17 (1.28–3.66)	781 (76.4%)	241 (23.6%)	0.0036	1.36 (1.10–1.68)
		Controls (416)	210 (50.5%)	166 (39.9%)	40 (9.6%)			CT: 1.75 (1.03–2.99)	586 (70.4%)	246 (29.6%)		
	Argentina	Cases (274)	188 (68.6%)	77 (28.1%)	9 (3.3%)	14.1697	8.38×10^{-4}	TT: 3.26 (1.51–7.06)	453 (82.7%)	95 (17.3%)	1.06×10^{-4}	1.73 (1.30–2.31)
		Controls (346)	192 (55.5%)	124 (35.8%)	30 (8.7%)			CT: 2.07 (0.93–4.59)	508 (73.4%)	184 (26.6%)		
	Germany	Cases (241)	132 (54.8%)	98 (40.7%)	11 (4.6%)	7.7164	0.0211	TT: 2.46 (1.19–5.09)	362 (75.1%)	120 (24.9%)	0.0080	1.43 (1.09–1.87)
		Controls (335)	151 (45.1%)	153 (45.7%)	31 (9.3%)			CT: 1.81 (0.87–3.76)	455 (67.9%)	215 (32.1%)		
	Italy	Cases (231)	130 (56.3%)	87 (37.7%)	14 (6.1%)	10.1706	0.0062	TT: 2.42 (1.19–4.93)	347 (75.1%)	115 (24.9%)	0.0016	1.59 (1.18–2.14)
		Controls (219)	92 (42.0%)	103 (47.0%)	24 (11.0%)			CT: 1.45 (0.71–2.97)	287 (65.5%)	151 (34.5%)		
	Spain	Cases (678)	404 (59.6%)	231 (34.1%)	43 (6.3%)	14.8617	5.93×10^{-4}	TT: 1.04 (0.62–1.76)	1,039 (76.6%)	317 (23.4%)	0.010	1.29 (1.06–1.56)
		Controls (458)	225 (49.1%)	208 (45.4%)	25 (5.5%)			CT: 0.65 (0.38–1.09)	658 (71.8%)	258 (28.2%)		
	Pooled	Cases (1,856)	1,102 (59.4%)	655 (35.3%)	99 (5.3%)				2,859 (77.0%)	853 (23.0%)	4.74×10^{-11}	1.42 (1.28–1.58)
		Controls (1,774)	870 (49.0%)	754 (42.5%)	150 (8.5%)				2,494 (70.3%)	1,054 (29.7%)		
	Population		GG	GA	AA	χ^2	*P*	Odds ratio (CI)[a]	Allele G	Allele A	*P*	Odds ratio (CI)
rs3733197	Scandinavia	Cases (419)	167 (39.9%)	192 (45.8%)	60 (14.3%)	1.2365	0.5389	GG: 1.04 (0.69–1.58)	526 (62.8%)	312 (37.2%)	0.5832	1.06 (0.87–1.29)
		Controls (444)	163 (36.7%)	220 (49.6%)	61 (13.7%)			GA: 0.89 (0.59–1.33)	546 (61.5%)	342 (38.5%)		
	Argentina	Cases (287)	177 (61.7%)	97 (33.8%)	13 (4.5%)	9.6496	0.0080	GG: 2.36 (1.20–4.66)	451 (78.6%)	123 (21.4%)	0.0018	1.15 (0.95–1.40)
		Controls (363)	184 (50.7%)	147 (40.5%)	32 (8.8%)			GA: 1.62 (0.81–3.25)	515 (70.9%)	211 (29.1%)		
	Germany	Cases (272)	128 (47.1%)	112 (41.2%)	32 (11.8%)	4.1431	0.1260	GG: 1.65 (1.01–2.69)	368 (67.6%)	176 (32.4%)	0.0382	1.28 (1.00–1.63)
		Controls (362)	148 (40.9%)	153 (42.3%)	61 (16.9%)			GA: 1.40 (0.85–2.28)	449 (62.0%)	275 (38.0%)		
	Italy	Cases (253)	131 (51.8%)	102 (40.3%)	20 (7.9%)	8.2595	0.0161	GG: 1.74 (0.92–3.29)	364 (71.9%)	142 (28.1%)	0.0097	1.42 (1.08–1.87)
		Controls (251)	98 (39.0%)	127 (50.6%)	26 (10.4%)			GA: 1.04 (0.55–1.98)	323 (64.3%)	179 (35.7%)		
	Spain	Cases (588)	307 (52.2%)	234 (39.8%)	47 (8.0%)	3.4580	0.1775	GG: 1.14 (0.72–1.82)	977 (72.1%)	379 (27.9%)	0.1474	1.50 (1.15–1.96)
		Controls (455)	212 (46.6%)	206 (45.3%)	37 (8.1%)			GA: 0.89 (0.56–1.43)	630 (69.2%)	280 (30.8%)		
	Pooled	Cases (1,819)	910 (50.0%)	737 (40.5%)	172 (9.5%)				2,686 (70.4%)	1,132 (29.6%)	4.67×10^{-5}	1.23 (1.11–1.36)
		Controls (1,875)	805 (42.9%)	853 (45.5%)	217 (11.6%)				2,463 (65.7%)	1,287 (34.3%)		

[a]Genotypic odds ratio calculated using homozygosity for the protective allele as reference with OR = 1. [b]Calculated using the Robins, Breslow and Greenland method. [c]Calculated using the Mantel-Haenszel χ^2 with fixed effects.

Table 3 Effect sizes of individual SNPs and 2- and 3-SNP haplotypes of *BANK1*[a]

SNP or haplotype	Allele or haplotype	Frequency	OR	95% CI	Effect
rs17266594 (branch point)	T	0.738	1.18	1.09–1.23	Risk
	C	0.262	0.70	0.63–0.78	Protection
rs10516487 (R61H)	G	0.739	1.16	1.07–1.26	Risk
	A	0.261	0.72	0.66–0.80	Protection
rs3733197 (A383T)	G	0.680	1.13	1.04–1.23	Risk
	A	0.320	0.80	0.73–0.89	Protection
2-SNP haplotype[b]	TG	0.744	1.16	1.07–1.25	Risk
	CA	0.256	0.70	0.63–0.77	Protection
3-SNP haplotype	TGG	0.636	1.16	1.06–1.27	Risk
	CAA	0.211	0.69	0.62–0.77	Protection
	TGA	0.108	0.98	0.84–1.14	Neutral
	CAG	0.045	0.73	0.54–0.92	Protection

[a]Effect sizes were calculated using WHAP, and ORs were estimated using R language.
[b]2-SNP = rs17266594 + rs10516487; 3-SNPs are in the order rs17266594, rs10516487 and rs3733197.

($P = 4.67 \times 10^{-5}$) for rs10516485, rs17266594 and rs3733197, respectively, for the allelic association, supported by genotypic association (**Table 2**).

rs17266594 and rs10516487 are separated by 153 bp and are in strong LD ($D' = 0.95$; $r^2 = 0.90$ calculated for all sets jointly; **Supplementary Fig. 2** online). rs3733197 is 88 kb away from rs10516487 ($D' = 0.72$; $r^2 = 0.39$) and rs17266594 ($D' = 0.68$; $r^2 = 0.27$), and could segregate with a risk haplotype in some individuals (**Supplementary Figs. 2** and **3** online). To better define the relative contribution of each SNP, we carried out conditional logistic regression analyses using the three SNPs. We found that none of the SNPs is independent of the others, as a result of the LD between them (colineality in the multiple logistic regression analysis). Through haplotype-based logistic regression analysis using WHAP[8], we did not find any differences in the effect sizes (OR) of the individual SNP alleles or the 2- or 3-SNP haplotypes (**Table 3**). Thus, linkage disequilibrium, haplotype and conditional regression analyses suggested that all three SNPs, either individually or as haplotypes, confer susceptibility for SLE.

BANK1 is a B-cell adaptor protein[9,10]. The two full-length isoforms of 785 and 755 amino acids differ by 30 amino acids at the N terminus, encoded by the alternative exon 1A (**Fig. 1e**), and contain ankyrin repeat motifs and coiled-coil regions, structures very similar to other adaptor proteins[11]. B-cell activation through the B-cell receptor leads to tyrosine phosphorylation of BANK1, which in turn promotes its association with the tyrosine kinase Lyn and the calcium channel IP3R, facilitating phosphorylation and activation of IP3R by Lyn and release of Ca^{2+} from endoplasmic reticulum stores[5,12]. IP3R associates with the N-terminal domain of BANK1 encoded by exon 2, whereas Lyn interacts with the C-terminal portion[5]. Our own analysis predicts a pleckstrin homology domain in the N terminus, which could also participate in phosphatidylinositol-mediated signaling. rs10516487 lies within the region essential for binding of IP3R. We speculate that R61, being highly protonated under conditions of physiological pH, could potentially alter the affinity of BANK1 for IP3R, favoring stronger binding, although this has yet to be tested.

rs17266594 may affect the relative splicing efficiency, but not splicing *per se*, of the full-length and Δ2 isoforms of *BANK1*. Mutations affecting the thymidine of the branch point consensus sequence and altering splicing efficiency have been previously described[13]. Through more efficient splicing of a full-length transcript containing the arginine residue in the IP3R-binding domain, a more 'active' protein would be expected in individuals at risk. On the contrary, given that the Δ2 isoform lacks the entire exon 2 and thus the IP3R-binding and PH domains (**Fig. 1e**), it possibly functions as a dominant-negative or rather, a dose-dependent isoform attenuating BANK1-mediated signaling. This is supported by the observation of a strong protective genetic effect in individuals with the CC and CT genotypes of rs17266594 (CC: OR = 0.52, 95% CI = 0.0–0.67; CT: OR = 0.68, 95% CI = 0.59–0.78) that show increased concentrations of the Δ2 isoform relative to the full-length isoform (**Fig. 1c**). Experimental evidence for a dominant-negative effect of the Δ2 isoform is needed to validate this proposed mechanism.

The importance of mutations in ankyrin motifs for interactions with IP3R was recently highlighted by a discovery linking single amino acid substitutions in the adaptor protein ankyrin-B with cardiac arrhythmia and sudden cardiac death[14]. Although the A383 variant is associated with SLE, the minor allele 383T of rs3733197 might create a site for threonine kinases[15].

B cells are the primary cell type affected in SLE. Novel therapies are aimed at depleting hyperactivated B cells that may function as autoantibody-producing cells and as important regulators of innate and adaptive immune responses through antigen presentation and cytokine-mediated signaling[16]. Functional and expression abnormalities of signaling molecules in B cells have been described in individuals with lupus. Of note, Lyn, a binding partner of BANK1, is of key importance in both human and mouse lupus disease models[17–22].

Increased binding of BANK1 to downstream effector proteins may lead to a steady state marked by B-cell hyperresponsiveness or deregulated B-cell activation. The precise role of BANK1 in B cell receptor–mediated signaling remains unclear, as two reports published to date contain conflicting data regarding the stimulatory or inhibitory role of BANK1 on B-cell activation[5,6]. Given the previously unreported existence of the alternative splicing of exon 2, we can speculate that the negative role for BANK1 assigned from the knockout model was a result of, in part, the residual expression of the Δ2 isoform, as this exon was targeted by the knockout construct[6]. Further experiments are required to fully understand if, and how, *BANK1* polymorphisms lead to B-cell hyperactivity, breakage of B-cell tolerance and production of autoantibodies, which are the principal hallmarks of SLE.

METHODS

Clinical samples. In our initial scan, we genotyped 279 Swedish SLE cases and 515 controls using the Affymetrix 100K array. Of these, 279 cases and 352 controls were available for the additional genotyping of *BANK1* SNPs (**Table 1**). For the functional polymorphisms, we genotyped an additional 185 Swedish SLE cases; 465 of the controls were available for genotyping of rs17266594 and rs3733197. For the final Mantel-Haenszel analysis and OR estimation, we added 84 Danish SLE cases to the Swedish cases, comprising the Scandinavian set (**Table 2**). The replication sets included 384 North German SLE cases and 374 controls, 288 Argentine SLE cases and 372 controls, and 286 Italian SLE cases and 252 controls. The Spanish cohort included 799 SLE cases and 542 controls from several regions in Spain. Of these, we genotyped 707 SLE cases and 469 controls for rs10516487 and rs3733197, and 678 SLE cases and 457 controls for rs17266594, as DNA from a number of controls was not available. The German, Spanish and Argentine SLE cases have all been previously described[23]. The Italian cases are a multicenter collection of affected individuals and their matched controls from Rome, Siena, Milan and Naples (North- and Mid-Italy). All cases fulfill the 1982 American College of Rheumatology (ACR) criteria for

the classification of SLE[24]. All participating subjects provided informed consent for this study. The study was approved by the various institutional review boards and ethical committees at each of the participating locations.

Genotyping. We carried out genotyping using the 100K Affymetrix array according to the manufacturer's instructions. We carried out fine-mapping and replication for SNPs rs10516487, rs17266594 and rs3733197 using TaqMan SNP genotyping assays (Applied Biosystems). The Affymetrix genotyping and fine-mapping were done at Serono Genetics Institute (now MerckSerono SA). The functional polymorphism replications were done at Uppsala University. For verification, 106 samples were genotyped twice, showing 100% concordance. Genotyping success rate for all the samples was over 92%.

Statistical analysis. For the 100K Affymetrix whole-genome scan analysis, we applied the following pre-processing filters. Specifically, SNPs were discarded if: (i) the proportion of missing genotypes was higher than 5%, (ii) the relative minor allele frequency was lower than 1%, or (iii) the probability that the observed genotype distribution results from sampling a SNP which follows the Hardy-Weinberg equilibrium was lower than 0.02. After filtering, we used data from 85,042 SNPs. We retained only SNPs from autosomal chromosomes for the sake of homogeneity between male and female subjects. SNP sequences were mapped onto NCBI 36 human genome assembly, and SNPs with multiple localizations were discarded. For each remaining SNP, we calculated genotypic and allelic frequencies in cases and controls and computed the corresponding probability values using exact (non-asymptotic) and unbiased algorithms[25]. Detailed results from the scan will be published elsewhere. The false-discovery rate (FDR) was then estimated using a method previously described[4].

For fine-mapping analyses, we estimated genetic association, haplotype estimation, LD and r^2 using Haploview (v4.0RC2). The Breslow-Day test of combinability and the Mantel-Haenszel test were carried out using the StatsDirect software (v2.4.6). As the Breslow-Day test showed combinability of the strata, the Mantel-Haenszel test for fixed effects was used in the analysis. Haplotypes were constructed using the PHASE software (v2.1)[26,27]. Genotypic odds ratios were calculated using the Unphased software (v3.0.9)[28].

We carried out logistic regression analysis and conditional multiple logistic regression analysis using R language glm routines. Haplotype-based logistic regression analysis was done using WHAP[8]. Coefficients were estimated with WHAP, and ORs and confidence intervals (**Table 3**) were calculated using R language.

Sequencing. DNA fragments for sequencing were amplified with the corresponding primers (**Supplementary Table 3**), purified from agarose gel with QIAquick gel extraction kit (Qiagen) and sequenced using BigDye Terminator 3.1 (Applied Biosystems) at the Uppsala Genome Center.

RNA purification and *BANK1* expression analysis. Total RNA was purified with TRIZOL Reagent (Invitrogen) from peripheral blood mononuclear cells (PBMCs) obtained with agreed consent from healthy donors and SLE cases. 2 μg of RNA were reverse-transcribed with 2 U of MultiScribe transcriptase in PCR buffer II containing 5 mM $MgCl_2$, 1 mM dNTPs, 0.4 U of RNase inhibitor and 5 μM oligo-dT. All reagents were purchased from Applied Biosystems. cDNA synthesis was done at 42 °C for 80 min, and then the reaction was terminated at 95 °C for 5 min. All cDNA samples were diluted to 15 ng/μl.

BANK1 expression was determined by real-time PCR on an ABI PRISM 7700 Sequence Detector (Applied Biosystems) with SDS 1.9.1 software. Total *BANK1*, both alternative full-length isoforms and the Δ2 isoform were quantified with SYBR Green and relevant primers (**Supplementary Table 3**). We carried out initial denaturation at 95 °C for 5 min followed by 45 cycles of PCR (95 °C for 15 s, 62 °C for 15 s and 72 °C for 30 s). PCR buffer provided with enzyme was supplemented with 3 mM $MgCl_2$, 200 μM of each of dNTPs, primers, SYBR Green (Molecular Probes), 15 ng of cDNA and 0.5 U of Platinum Taq polymerase (Invitrogen). Expression levels were normalized to the levels of TBP in the same samples amplified with commercial reagents (Applied Biosystems). All experiments were run in triplicate. Independent cDNA synthesis was carried out twice.

Expression levels for total *BANK1*, both full-length isoforms and Δ2 isoform in separated blood cell populations ($CD4^+$, $CD8^+$, $CD14^+$ and $CD19^+$ cells) were determined using Human Blood Fractions MTC Panel (Clontech). Statistical calculations were performed with available online GraphPad Software using two-tailed *t*-test.

Cloning of human, mouse and chimpanzee Δ2 isoform. Purification of total RNA from mouse spleen and cDNA synthesis were conducted as described above for the human PBMCs. Total RNA from chimpanzee (*Pan troglodytes*) spleen was provided by T. Bergström and L. Cavelier (Uppsala University). The human gene was amplified from Human Spleen BD Marathon-Ready cDNA (Clontech). After initial denaturation at 95 °C for 5 min, 35 cycles (95 °C for 20 s, 60 °C for 15 s and 72 °C for 2 min 30 s) were performed in PCR buffer containing 2 mM $MgSO_4$, 200 μM of each of dNTPs, 0.4 μM of each of the corresponding primers (**Supplementary Table 3**) and 0.5 U of Platinum Taq-High Fidelity enzyme (Invitrogen). Chimpanzee cDNA was amplified with human-specific primers. PCR products were purified from agarose gel and cloned in pCR 4-TOPO vector (Invitrogen) according to the manufacturer's instructions. Plasmid DNA from positive clones was purified with QIAprep Spin Miniprep kit (Qiagen) and verified by sequencing.

Accession codes. GenBank: full-length isoforms of *BANK1* containing exon 1A or exon 1B, NM_017935 and AB063170, respectively. Δ2 transcript sequences have been deposited with the following accession codes: EU051376, human; EU051377, chimpanzee and EU051378, mouse.

URLs. Haploview, http://www.broad.mit.edu/mpg/haploview/; GraphPad Software, http://www.graphpad.com; protein analysis, http://www.ebi.ac.uk/saps/, http://smart.embl-heidelberg.de/, http://ca.expasy.org/prosite/, and http://www.cbs.dtu.dk/services/NetPhos/; WHAP, http://pngu.mgh.harvard.edu/~purcell/whap/.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

The authors would like to express their gratitude to S. Lewén for help with purification of PBMCs and total RNA, H. Yin for help with the preparation of DNA samples and to the members of the Uppsala Genome Center for help with sequencing. We thank T. Bergström and L. Cavelier (Uppsala University) for providing RNA from chimpanzee spleen. We also thank A.I. Scollo, A.M. Perichon and M.C.R. Tenaglia (CEDIM, Diagnóstico Molecular y Forense SRL, Rosario, Argentina) for their help in DNA preparation of the Argentine samples and A. Voss for clinical help with the Danish 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 all the patients for their contribution. The authors would also like to thank J.Osorio y Fortea (Pasteur Institut, Paris) for his contribution in the analyses that led to the identification of R61H early in this project.

This work has been supported in part by grants from the European CVDIMMUNE project from the European Commission LSHM-CT-2006-037227, Swedish Research Council for Medicine (12703), the Swedish Association against Rheumatism, the Magnus Bergwalls Foundation, the Gustaf V: 80th-year Jubilee Foundation, the Torsten and Ragnar Söderbergs Foundation and the Marcus Borsgtröms Foundation to M.E.A.-R. and by the Gurli and Edward Brunnbergs Foundation for rheumatologic research to S.V.K. M.E.A.-R. is supported by an award from the Knut and Alice Wallenberg Foundation through the Royal Swedish Academy of Sciences. This work was also partially supported by FISM, Regione Piemonte (CIPE), the Bundesministerium für Bildung und Forschung (BMBF) Kompetenznetz Rheuma C2.12, Germany and the Danish Rheumatism Association.

AUTHOR CONTRIBUTIONS

M.E.A.-R., S.V.K. and H.A. designed the experiments. S.V.K., A.-K.A., J.W., A.Z., M.V.P.L.R. and E.S. performed the experiments. M.E.A.-R., S.V.K., A.-K.A., J.W. and H.A. performed the analyses. I.G., E.S., G.S., L.T., A.J., T.W., S.D., N.B., M.G.D., C.G., A.S., P.J., H.L., B.A.P.-E., M.F.G.-E. and J.M. and their multicenter collaborators provided the samples. M.E.A.-R., S.V.K. and A.-K.A. wrote the manuscript.

The Argentine Collaborative Group participants are:
Hugo R. Scherbarth, Pilar C. Marino and Estela L. Motta (Servicio de Reumatología, Hospital Interzonal General de Agudos 'Dr. Oscar Alende', Mar del Plata, Argentina), Susana Gamron, Cristina Drenkard and Emilia Menso (Servicio de Reumatología de la UHMI 1, Hospital Nacional de Clínicas, Universidad Nacional de Córdoba, Córdoba, Argentina), Alberto Allievi and Guillermo A. Tate (Organización Médica de Investigación, Buenos Aires, Argentina), Jose L. Presas (Hospital General de Agudos Dr. Juán A. Fernandez, Buenos Aires, Argentina),

Simon A. Palatnik, Marcelo Abdala and Mariela Bearzotti (Facultad de Ciencias Medicas, Universidad Nacional de Rosario y Hospital Provincial del Centenario, Rosario, Argentina), Alejandro Alvarellos, Francisco Caeiro and Ana Bertoli (Servicio de Reumatología, Hospital Privado, Centro Medico de Córdoba, Córdoba, Argentina), Sergio Paira and Susana Roverano (Hospital José M. Cullen, Santa Fe, Argentina), Cesar E. Graf and Estela Bertero (Hospital San Martín, Paraná, Argentina), Cesar Caprarulo and Griselda Buchanan (Hospital Felipe Heras, Concordia, Entre Ríos, Argentina), Carolina Guillerón, Sebastian Grimaudo and Jorge Manni (Departamento de Inmunología, Instituto de Investigaciones Médicas 'Alfredo Lanari', Buenos Aires, Argentina), Luis J. Catoggio, Enrique R. Soriano and Carlos D. Santos (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, Fernando A. Ramos and Sandra M. Navarro (Servicio de Reumatología, Hospital Provincial de Rosario, Rosario, Argentina), Guillermo A. Berbotto, Marisa Jorfen and Elisa J. Romero (Servicio de Reumatología Hospital Escuela Eva Perón, Granadero Baigorria, Rosario, Argentina), Mercedes A. Garcia, Juan C. Marcos and Ana I. Marcos (Servicio de Reumatología, Hospital Interzonal General de Agudos General San Martín, La Plata, Buenos Aires, Argentina), Carlos E. Perandones and Alicia Eimon (Centro de Educación Médica e Investigaciones Clínicas (CEMIC), Buenos Aires, Argentina), and Cristina G. Battagliotti (Hospital de Niños Dr. Orlando Alassia, Santa Fe, Argentina). B.A.P.-E. is coordinator of the Argentine Collaborative Group.

The German Collaborative Group participants are:
K. Armadi-Simab and Wolfgang L. Gross (Abteilung Rheumatologie, University Hospital of Schleswig-Holstein, Campus Luebeck, Rheumaklinik Bad Bramstedt, Luebeck, Germany), Erika Gromnica-Ihle (Rheumaklinik Berlin-Buch, Berlin, Germany), Hans-Hartmut Peter (Medizinische Universitaetsklinik, Abteilung Rheumatologie und Klinische Immunologie, Freiburg, Germany), Karin Manger (Medizinische Klinik III derFAU Erlangen-Nuernberg, Erlangen, Germany), Sebastian Schnarr and Henning Zeidler (Abteilung Rheumatologie, Medizinische Hochschule Hannover, Hannover, Germany) and Reinhold E. Schmidt (Abteilung Klinische Immunologie, Medizinische Hochschule Hannover, Hannover, Germany).

The Spanish Collaborative Group participants are:
Norberto Ortego (Servicio Medicina Interna, Hospital Clínico San Cecilio, Granada, Spain), Enrique de Ramón (Servicio de Medicina Interna, Hospital Carlos Haya, Malaga, Spain), Juan Jiménez-Alonso (Servicio de Medicina Interna, Hospital Virgen de las Nieves, Granada, Spain), Julio Sánchez-Román (Servico de Medicina Interna, Hospital Virgen del Rocio, Sevilla, Spain) and Miguel Angel López-Nevot (Servicio de Inmunología, Hospital Virgen de las Nieves, Granada, Spain).

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 (Siena University, Siena, Italy), Sergio Migliaresi (Rheumatology Unit Second University of Naples, Naples, Italy). Also we would like to thank Armando Gabrielli, Clinica Medica di Scienze Mediche e Chirurgiche, Universitá Politecnica delle Marche.

COMPETING INTERESTS STATEMENT

The authors declare competing financial interests: details accompany the full-text HTML version of the paper at http://www.nature.com/naturegenetics/.

Published online at http://www.nature.com/naturegenetics
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions

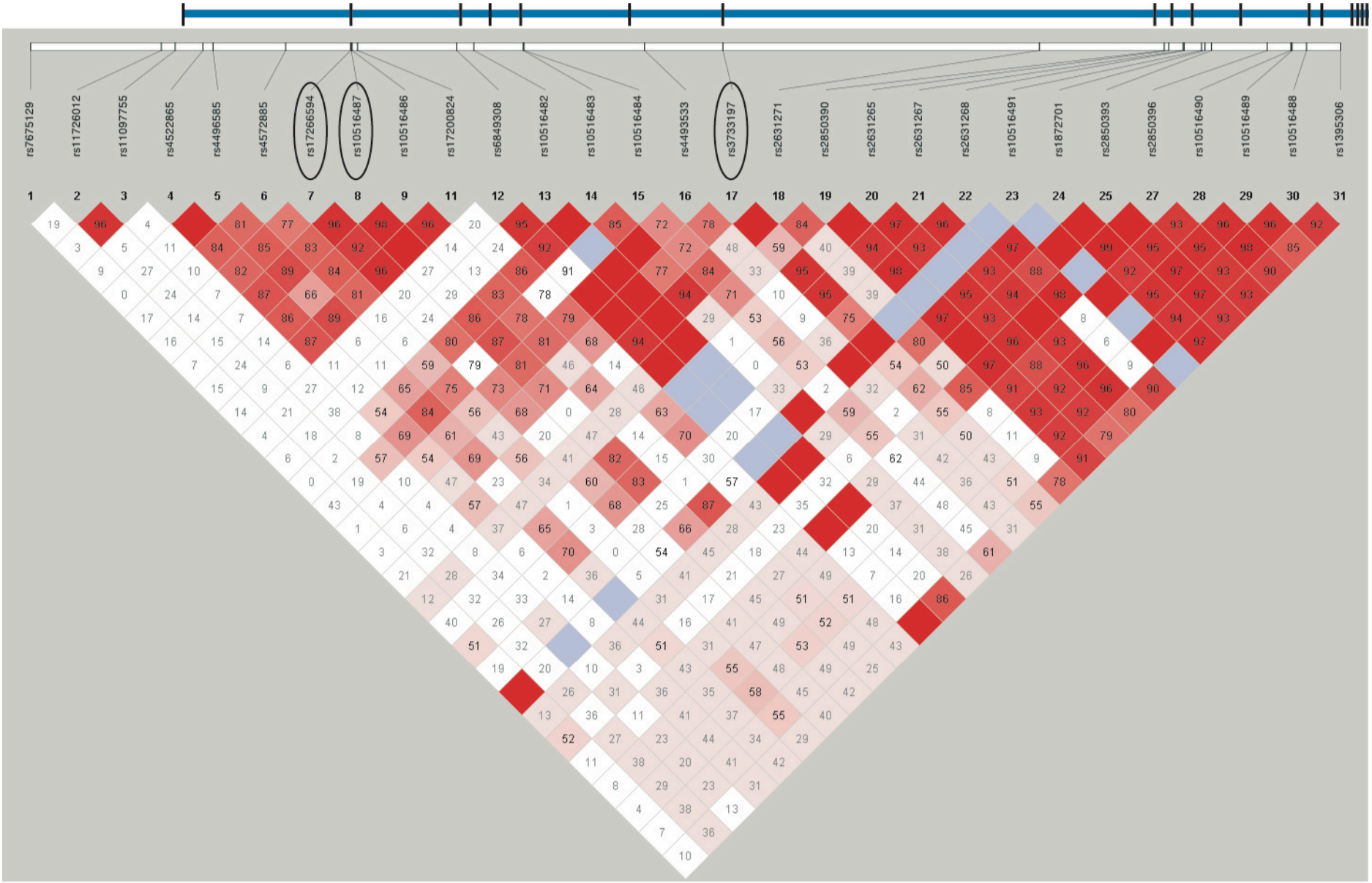
1. Sherer, Y., Gorstein, A., Fritzler, M.J. & Shoenfeld, Y. Autoantibody explosion in systemic lupus erythematosus: more than 100 different antibodies found in SLE patients. *Semin. Arthritis Rheum.* **34**, 501–537 (2004).
2. Sestak, A.L., Nath, S.K., Sawalha, A.H. & Harley, J.B. Current status of lupus genetics. *Arthritis Res. Ther.* **9**, 210 (2007).
3. Alarcon-Riquelme, M.E. The genetics of systemic lupus erythematosus: understanding how SNPs confer disease susceptibility. *Springer Semin. Immunopathol.* **28**, 109–117 (2006).
4. Forner, K., Guedj, M., Dauvillier, J. & Wojcik, J. Universal false discovery rate estimation methodology for genome-wide association studies. *Hum. Hered.* **65**, 183–194 (2008).
5. Yokoyama, K. *et al.* BANK regulates BCR-induced calcium mobilization by promoting tyrosine phosphorylation of IP(3) receptor. *EMBO J.* **21**, 83–92 (2002).
6. Aiba, Y. *et al.* BANK negatively regulates Akt activation and subsequent B cell responses. *Immunity* **24**, 259–268 (2006).
7. Burge, C.B., Tuschl, T. & Sharp, P.A. The RNA World II. (eds. Gesteland, R.F., Cech, T.R. & Atkins, J.F.) 525–560 (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1999).
8. Purcell, S., Daly, M.J. & Sham, P.C. WHAP: haplotype-based association analysis. *Bioinformatics* **23**, 255–256 (2007).
9. Jordan, M.S., Singer, A.L. & Koretzky, G.A. Adaptors as central mediators of signal transduction in immune cells. *Nat. Immunol.* **4**, 110–116 (2003).
10. Kurosaki, T. Regulation of B-cell signal transduction by adaptor proteins. *Nat. Rev. Immunol.* **2**, 354–363 (2002).
11. Okada, T., Maeda, A., Iwamatsu, A., Gotoh, K. & Kurosaki, T. BCAP: the tyrosine kinase substrate that connects B cell receptor to phosphoinositide 3-kinase activation. *Immunity* **13**, 817–827 (2000).
12. Patterson, R.L., Boehning, D. & Snyder, S.H. Inositol 1,4,5-trisphosphate receptors as signal integrators. *Annu. Rev. Biochem.* **73**, 437–465 (2004).
13. Li, M. & Pritchard, P.H. Characterization of the effects of mutations in the putative branchpoint sequence of intron 4 on the splicing within the human lecithin:cholesterol acyltransferase gene. *J. Biol. Chem.* **275**, 18079–18084 (2000).
14. Mohler, P.J. *et al.* Ankyrin-B mutation causes type 4 long-QT cardiac arrhythmia and sudden cardiac death. *Nature* **421**, 634–639 (2003).
15. Blom, N., Gammeltoft, S. & Brunak, S. Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. *J. Mol. Biol.* **294**, 1351–1362 (1999).
16. Anolik, J., Sanz, I. & Looney, R.J. B cell depletion therapy in systemic lupus erythematosus. *Curr. Rheumatol. Rep.* **5**, 350–356 (2003).
17. Liossis, S.N., Kovacs, B., Dennis, G., Kammer, G.M. & Tsokos, G.C. B cells from patients with systemic lupus erythematosus display abnormal antigen receptor-mediated early signal transduction events. *J. Clin. Invest.* **98**, 2549–2557 (1996).
18. Huck, S., Le Corre, R., Youinou, P. & Zouali, M. Expression of B cell receptor-associated signaling molecules in human lupus. *Autoimmunity* **33**, 213–224 (2001).
19. Liossis, S.N. *et al.* B-cell kinase lyn deficiency in patients with systemic lupus erythematosus. *J. Investig. Med.* **49**, 157–165 (2001).
20. Hibbs, M.L. *et al.* Sustained activation of Lyn tyrosine kinase *in vivo* leads to autoimmunity. *J. Exp. Med.* **196**, 1593–1604 (2002).
21. Flores-Borja, F., Kabouridis, P.S., Jury, E.C., Isenberg, D.A. & Mageed, R.A. Decreased Lyn expression and translocation to lipid raft signaling domains in B lymphocytes from patients with systemic lupus erythematosus. *Arthritis Rheum.* **52**, 3955–3965 (2005).
22. Cornall, R.J. *et al.* Polygenic autoimmune traits: Lyn, CD22, and SHP-1 are limiting elements of a biochemical pathway regulating BCR signaling and selection. *Immunity* **8**, 497–508 (1998).
23. Kozyrev, S.V. *et al.* Structural insertion/deletion variation in *IRF5* is associated with a risk haplotype and defines the precise isoforms expressed in SLE. *Arthritis Rheum.* **56**, 1234–1241 (2007).
24. Tan, E.M. *et al.* The 1982 revised criteria for the classification of systemic lupus erythematosus. *Arthritis Rheum.* **25**, 1271–1277 (1982).
25. Guedj, M., Wojcik, J., Della-Chiesa, E., Nuel, G. & Forner, K. A fast, unbiased and exact allelic test for case-control association studies. *Hum. Hered.* **61**, 210–221 (2006).
26. Stephens, M. & Donnelly, P. A comparison of bayesian methods for haplotype reconstruction from population genotype data. *Am. J. Hum. Genet.* **73**, 1162–1169 (2003).
27. Stephens, M., Smith, N.J. & Donnelly, P. A new statistical method for haplotype reconstruction from population data. *Am. J. Hum. Genet.* **68**, 978–989 (2001).
28. Dudbridge, F. Pedigree disequilibrium tests for multilocus haplotypes. *Genet. Epidemiol.* **25**, 115–121 (2003).

Corrigendum: Functional variants in the B-cell gene *BANK1* are associated with systemic lupus erythematosus

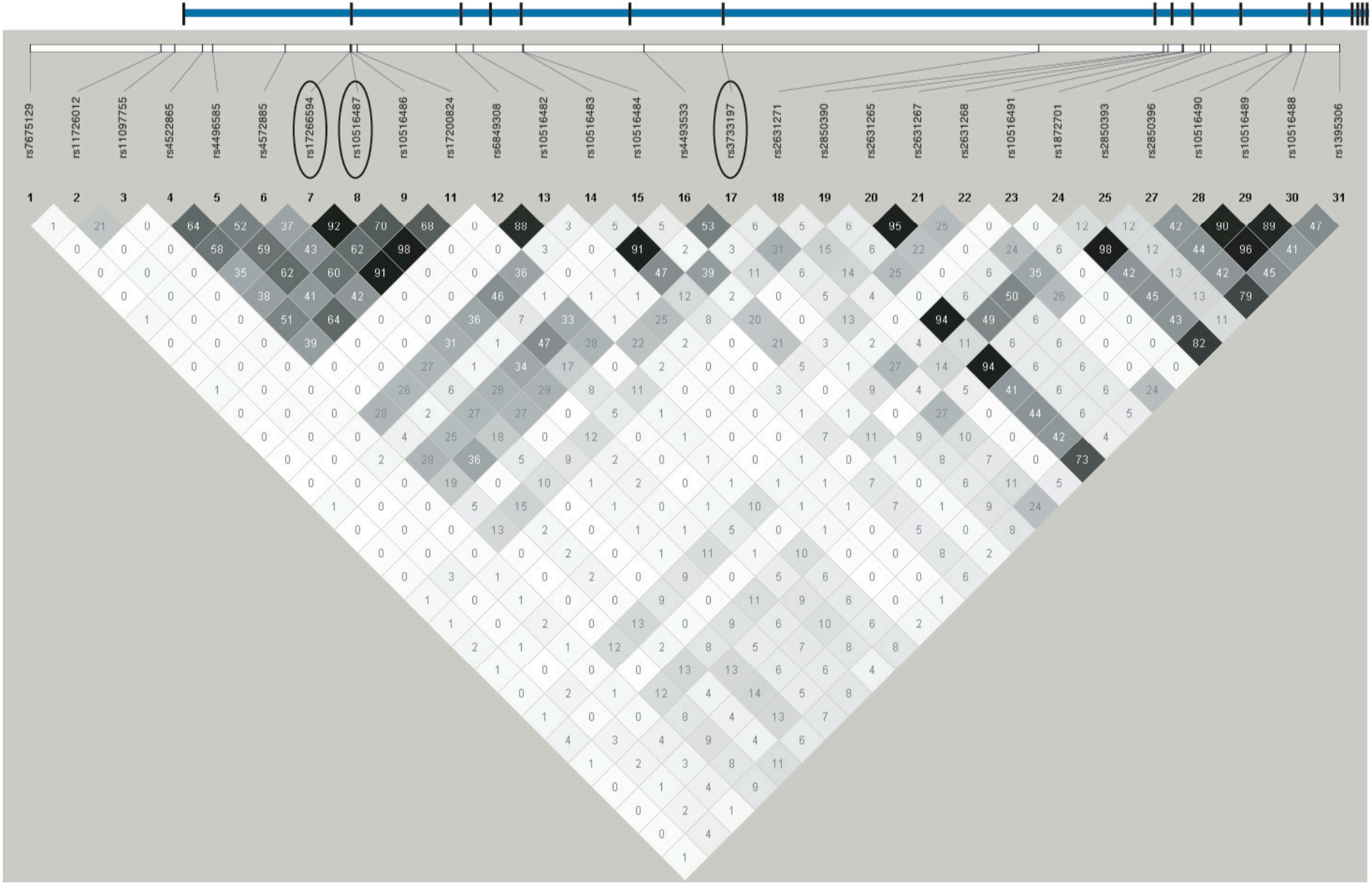
Sergey V Kozyrev, Anna-Karin Abelson, Jerome Wojcik, Ammar Zaghlool, MV Prasad Linga Reddy, Elena Sanchez, Iva Gunnarsson, Elisabet Svenungsson, Gunnar Sturfelt, Andreas Jönsen, Lennart Truedsson, Bernardo A Pons-Estel, Torsten Witte, Sandra D'Alfonso, Nadia Barrizzone, Maria Giovanna Danieli, Carmen Gutierrez, Ana Suarez, Peter Junker, Helle Laustrup, Maria Francisca González-Escribano, Javier Martin, Hadi Abderrahim & Marta E Alarcón-Riquelme

Nature Genetics **40, 211–216 (2008); published online 20 January 2008; corrected after print 27 March 2008**

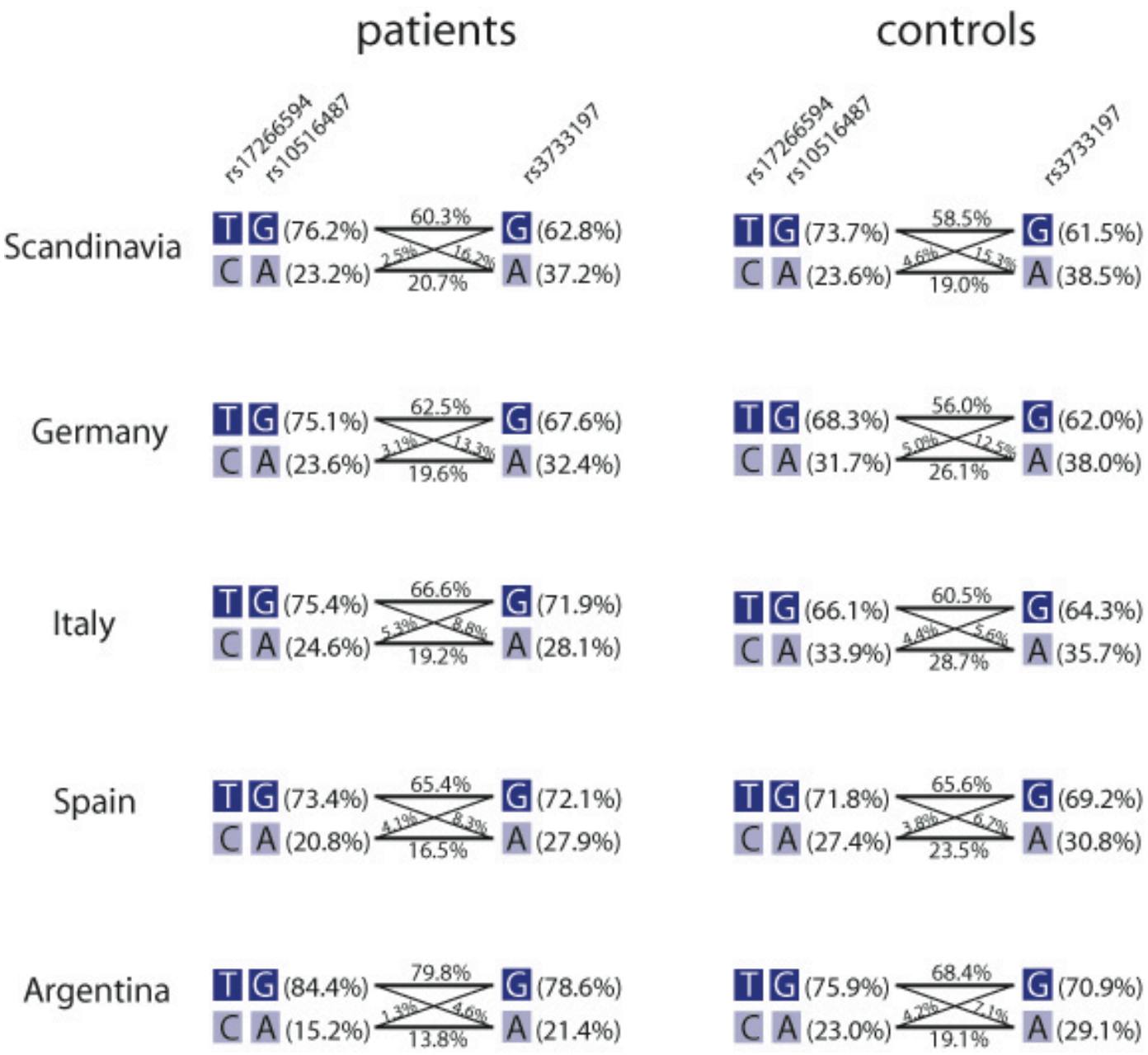
In the version of this article initially published, the name of the 15th author was misspelled. The correct author name is Nadia Barizzone. Also, the affiliation of Javier Martin was incomplete. Dr. Martin is affiliated with Instituto de Biomedicina López-Neyra, Grenada 18100, Spain and Consejo Superior de Investigaciones Científicas (CSIC), Grenada 18100, Spain. The errors have been corrected in the HTML and PDF versions of the article.



Supplementary Figure 1
Linkage disequilibrium structure (D') across *BANK1*. Data calculated with Haploview using the Swedish cases and controls run for 30 SNPs across the gene. In top of the figure, the genomic structure of *BANK1* is shown in blue, with exons shown as bars, being 5'-3' direction from left to right.



Supplementary Figure 2
Estimation of the R^2 between SNPs across *BANK1* using Haploview and data from the Swedish cases and controls. The genomic structure of *BANK1* is also shown in the top as in supplementary figure 1.



Supplementary Figure 3
Frequency and structure of the 2-SNP (rs17266594 and rs10516487) and 3-SNP haplotypes (including rs3733197) for each of the populations and cases and controls, separately.

Supplementary Table 1. BANK1 SNPs in the 100k Array			
SNP	**rs number**	**Position**	**(-log) P value**
SNP_A-1701374	rs10516487	103108254	2.27
SNP_A-1701494	rs10516486	103108454	2.79
SNP_A-1664926	rs6849308	103133261	2.22
SNP_A-1706628	rs10516482	103137348	2.52
SNP_A-1744756	rs10516483	103149083	3.25
SNP_A-1683131	rs2631271	103271574	n.s.
SNP_A-1697391	rs10516489	103331537	n.s.

Supplementary Table 2. Location of BANK1 SNPs

SNP rs number	Position**	Location in BANK1
rs7675129	102894046	intergenic, 36.8 kb upstream
rs11726012	102925041	distal promoter, 5.8 kb upstream
rs11097755	102928331	promoter, 2.5 kb upstream
rs4522865	102934911	intron 1
rs4496585	102937309	intron 1
rs4572885	102954536	intron 1
rs10516487	102970099	exon 2 coding (NS)*
rs10516486	102970299	exon 2 (synonymous)
rs17200824	102971612	intron 2
rs6849308	102995106	intron 2
rs10516482	102999193	intron 3
rs10516483	103010928	intron 5
rs10516484	103011108	intron 5
rs4493533	103039707	intron 6
rs3733197	103058310	exon 7 coding (NS)
rs2631271	103133419	intron 7
rs2850390	103163019	intron 8
rs2631265	103164099	intron 8
rs2631267	103167495	intron 9
rs2631268	103167753	intron 9
rs10516491	103171889	intron 10
rs1872701	103172704	intron 10
rs2850393	103174239	intron 10
rs2850396	103187471	intron 11
rs10516490	103193084	intron 11
rs10516489	103193382	intron 11
rs10516488	103196800	intron 11
rs1395306	103204873	intron 13

*NS: non-synonymous substitution

**Ensembl release 46

Supplementary Table 3. Primer sequences

Gene /gene fragment/isoform	Forward	Reverse
hBANK cDNA amplification	CACCTCAACCGCCACAATGCTGCCAGCA	ATAATAACCTTCTTTAATGATCTTTCTTGC
Total BANK1 qRT-PCR	AGAGGAAACTACACCTTACATAGCTC	GATGAGTTCTTCCTGACCATCAG
Total full-length isoforms	TCAAAGCAGATGGGAGATCTCAAC	
Δ2 isoform	CAGCGCCCCCAGATTCTGAAG	
Exon1A full-length isoform	CAGCGCCCCCAGGAAATACA	
Alternative exon1 full-length isoform	GCCTATTCTTTGTTTTGGAAATACA	
		Common reverse primer for all isoforms for qRT-PCR
		CACATGGAATTTCAGTGGGAAGCAC
		Common reverse primer for gel-analysis
		ATCACAGTAGACATTGACATGGAC
FOR GENOMIC SEQUENCING:		
promoter, exon 1A and 5'-part of intron1	TTGGAGAGGGTATTTAGAGCCATA	AAGCAGGGCTACCAATTCACCAG
Alternative exon1B	CTATGATACTGGAAATACTGTCAGT	AGCATATGACCAGCTGATCAG
Exon2	TTGATTTACTATGAAAATATCAAGC	TTACATAAGAAACCAGCTTCCAG
Mouse BANK1 cDNA	ACCTCCCGCAATGCTTCCTGT	ACATGGAATTTCCCCAGGAAGCAC

Kallikrein genes are associated with lupus and glomerular basement membrane–specific antibody–induced nephritis in mice and humans

Kui Liu,[1] Quan-Zhen Li,[1] Angelica M. Delgado-Vega,[2] Anna-Karin Abelson,[2] Elena Sánchez,[3] Jennifer A. Kelly,[4] Li Li,[1] Yang Liu,[1] Jinchun Zhou,[1] Mei Yan,[1] Qiu Ye,[1] Shenxi Liu,[1] Chun Xie,[1] Xin J. Zhou,[1] Sharon A. Chung,[5] Bernardo Pons-Estel,[6] Torsten Witte,[7] Enrique de Ramón,[8] Sang-Cheol Bae,[9] Nadia Barizzone,[10] Gian Domenico Sebastiani,[11] Joan T. Merrill,[4] Peter K. Gregersen,[12] Gary G. Gilkeson,[13] Robert P. Kimberly,[14] Timothy J. Vyse,[15] Il Kim,[9] Sandra D'Alfonso,[10] Javier Martin,[3] John B. Harley,[4] Lindsey A. Criswell,[5] The Profile Study Group,[16] The Italian Collaborative Group,[16] The German Collaborative Group,[16] The Spanish Collaborative Group,[16] The Argentinian Collaborative Group,[16] The SLEGEN Consortium,[16] Edward K. Wakeland,[1] Marta E. Alarcón-Riquelme,[2,4] and Chandra Mohan[1]

[1]Department of Immunology and Department of Medicine, University of Texas Southwestern Medical Center, Dallas, Texas, USA. [2]Department of Genetics and Pathology, Rudbeck Laboratory, Uppsala University, Uppsala, Sweden. [3]Instituto de Biomedicina y Parasitología López Neyra, CSIC, Granada, Spain. [4]Oklahoma Medical Research Foundation, Oklahoma City, Oklahoma, USA. [5]UCSF, San Francisco, California, USA. [6]Sanatorio Parque, Rosario, Argentina. [7]Hannover Medical School, Hannover, Germany. [8]Servicio de Medicina Interna, Hospital Carlos Haya, Malaga, Spain. [9]Department of Rheumatology, Hospital for Rheumatic Diseases, Hanyang University, Seoul, Republic of Korea. [10]Department of Medical Sciences and IRCAD, University of Eastern Piedmont, Novara, Italy. [11]U.O.C. di Reumatologia Azienda Ospedaliera San Camillo-Forlanini, Rome, Italy. [12]Feinstein Institute for Medical Research, Long Island Jewish Health System, Manhasset, New York, USA. [13]Medical University of South Carolina, Charleston, South Carolina, USA. [14]University of Alabama at Birmingham, Birmingham, Alabama, USA. [15]Section of Medical Genetics and Rheumatology, Hammersmith Hospital, Imperial College London, London, United Kingdom. [16]The collaborative groups are detailed at the end of this article.

Immune-mediated nephritis contributes to disease in systemic lupus erythematosus, Goodpasture syndrome (caused by antibodies specific for glomerular basement membrane [anti-GBM antibodies]), and spontaneous lupus nephritis. Inbred mouse strains differ in susceptibility to anti-GBM antibody–induced and spontaneous lupus nephritis. This study sought to clarify the genetic and molecular factors that may be responsible for enhanced immune-mediated renal disease in these models. When the kidneys of 3 mouse strains sensitive to anti-GBM antibody–induced nephritis were compared with those of 2 control strains using microarray analysis, one-fifth of the underexpressed genes belonged to the kallikrein gene family, which encodes serine esterases. Mouse strains that upregulated renal and urinary kallikreins exhibited less evidence of disease. Antagonizing the kallikrein pathway augmented disease, while agonists dampened the severity of anti-GBM antibody–induced nephritis. In addition, nephritis-sensitive mouse strains had kallikrein haplotypes that were distinct from those of control strains, including several regulatory polymorphisms, some of which were associated with functional consequences. Indeed, increased susceptibility to anti-GBM antibody–induced nephritis and spontaneous lupus nephritis was achieved by breeding mice with a genetic interval harboring the kallikrein genes onto a disease-resistant background. Finally, both human SLE and spontaneous lupus nephritis were found to be associated with kallikrein genes, particularly *KLK1* and the *KLK3* promoter, when DNA SNPs from independent cohorts of SLE patients and controls were compared. Collectively, these studies suggest that kallikreins are protective disease-associated genes in anti-GBM antibody–induced nephritis and lupus.

Introduction

Immune-mediated nephritis is an important pathogenic determinant in SLE and Goodpasture syndrome (anti–glomerular basement membrane [anti-GBM] disease). In spontaneous lupus nephritis, both Ab-mediated and Ab-independent mechanisms lead to renal pathology (1–4). In particular, anti-DNA and antiglomerular Abs, as well a few other specificities, have been implicated in the pathogenesis of lupus nephritis, in both mice and humans (1–8). A useful experimental tool for dissecting out the molecular mechanisms leading to immune-mediated nephritis in lupus and Goodpasture disease is the experimental anti-GBM Ab–induced glomerulonephritis (AIGN) model, wherein the transfer of anti-GBM Abs elicits glomerulonephritis (GN) with reproducible kinetics. Although the specificities of the inciting Abs may

Authorship note: Kui Liu, Quan-Zhen Li, Angelica M. Delgado-Vega, Anna-Karin Abelson, Elena Sánchez, and Jennifer A. Kelly are co–first authors. Edward K. Wakeland, Marta E. Alarcón-Riquelme, and Chandra Mohan are co–senior authors.

Conflict of interest: The authors have declared that no conflict of interest exists.

Nonstandard abbreviations used: AIGN, anti-GBM Ab–induced glomerulonephritis; B6, C57BL/6; BK, bradykinin; GBM, glomerular basement membrane; GN, glomerulonephritis; OR, odds ratio.

Citation for this article: *J. Clin. Invest.* **119**:911–923 (2009). doi:10.1172/JCI36728.

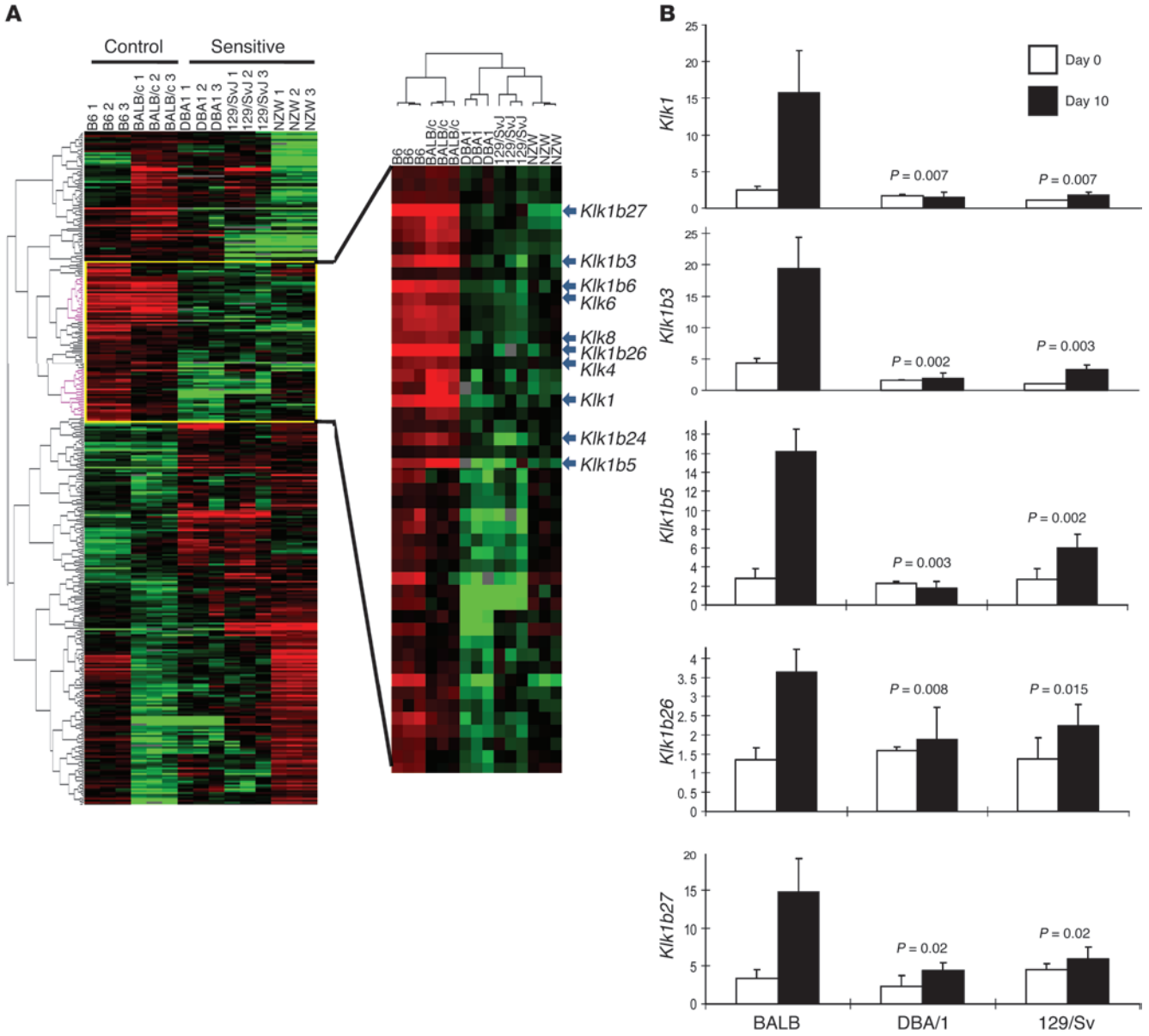


Figure 1
Strain-dependent gene expression differences in the renal cortex in AIGN. (**A**) Anti-GBM disease was induced in 3 disease-sensitive strains (DBA1, NZW, and 129/SvJ) and 2 control strains (BALB/c and B6), after which renal cortex RNA was analyzed using DNA microarrays on day 10 of disease (i.e., 5 days after injection of anti-GBM Abs). Three biological replicates were included for each strain. The left panel shows that a total of 360 gene transcripts were differentially expressed between the study strains (>2 fold, $P < 0.001$). The right panel (a higher-magnification view of the boxed region on the left) shows a cluster of gene transcripts that were increased in all control strains but not in the AIGN-sensitive strains (>2-fold difference, $P < 0.001$), including 10 *Klk* genes. (**B**) Renal cortex gene expression differences in *Klk1*, *Klk1b3*, *Klk1b5*, *Klk1b26,* and *Klk1b27* were confirmed by real-time PCR in the indicated strains, before (day 0) and after (day 10) anti-GBM Ab challenge. Each bar represents the mean of 6 samples. Similar changes were seen between B6 and NZW mice (data not shown). *P* values pertain to comparisons with BALB/c day 10 values. Error bars denote SD.

differ in experimental AIGN and spontaneous lupus nephritis, the downstream pathogenic cascades that lead to disease in the 2 scenarios appear to be shared, as reviewed recently (9).

Hence, the AIGN experimental model may be a useful tool for dissecting out the molecular and genetic basis of lupus nephritis. Notably, of more than 20 inbred mouse strains challenged with anti-GBM Abs, severe renal disease was noted in only 5 strains, including DBA/1, NZW, and 129/SvJ (10–12). Coincidentally, the latter 2 strains are known to develop spontaneous lupus nephritis (13–16). We had previously reported that the strain differences in AIGN susceptibility cannot be simply attributed to differences in systemic immune response (to the administered rabbit anti-mouse GBM Abs) or to differences in Th1 skewing (10, 11). On the other hand, the degree to which differences in renal-intrinsic

Table 1
Several kallikrein gene messages were significantly upregulated in the kidneys of B6 and BALB/c mice compared with NZW, DBA1, and 129/SvJ mice, upon anti-GBM Ab challenge

Gene	Accession no.	Average signal intensity					Fold change[A]	*P*[A]
		B6	BALB/c	NZW	129/SvJ	DBA1		
Klk1b3	NM_008693	36,397	26,179	16,453	9,715	7,056	2.6	<0.001
Klk1b27	NM_020268	26,608	22,255	10,360	7,185	5,164	2.9	<0.001
Klk1b26	NM_010644	17,694	17,596	7,907	7,932	5,602	2.3	<0.001
Klk1b5	NM_008456	14,546	15,007	3,608	3,847	4,286	3.7	<0.001
Klk1	NM_010639	12,779	13,115	5,079	4,682	3,300	2.8	<0.001
Klk1b24	NM_010643	269	129	77	43	58	3.0	<0.001
Klk1b8	NM_008457	246	350	47	62	35	6.2	<0.001
Klk4	NM_019928	7	103	1	1	3	30	<0.001
Klk6	NM_010639	19,269	28,119	9,375	13,804	17,533	2.0	<0.001

All values are normalized fluorescence intensity. [A]Fold change and *P* values were calculated by comparing B6 and BALB/c (pooled, $n = 6$) with the other 3 strains (pooled, $n = 9$).

processes may contribute to the observed strain differences in AIGN susceptibility remains to be elucidated.

Given the possibility that renal-intrinsic differences may be contributory, we undertook a microarray-based transcriptomic analysis of the renal cortex from 3 AIGN-sensitive strains and 2 control strains, after anti-GBM challenge. Surprisingly, we found that a significant fraction of the differentially expressed genes that distinguish the nephritis-sensitive strains from the control strains belong to the kallikrein (*Klk*) gene family. Importantly, this gene complex is encoded within an interval on chromosome 7 that had previously been associated with spontaneous lupus nephritis (14–21). Kallikreins constitute a multigene family of serine esterases with a wide spectrum of biological functions (22–39). These reported functions include the regulation of inflammation, apoptosis, redox balance, and fibrosis within the kidneys, as well as local blood pressure. In further genetic and functional studies, we demonstrate that *Klk* genes are renoprotective in immune-mediated renal disease and may constitute important disease susceptibility genes for experimental anti-glomerular Ab–induced nephritis as well as spontaneous lupus nephritis in mice and in humans.

Results

Displayed in Figure 1 are all genes that were significantly upregulated or downregulated (at least 2-fold difference, $P < 0.001$) in the strains that were highly sensitive to AIGN disease (i.e., NZW, DBA/1, and 129/SvJ) compared with either of the control strains (C57BL/6 [B6] or BALB/c), following challenge with anti-GBM serum. Though several strain-specific gene differences were also noted within this panel of genes, a subset of 50 genes within this panel were consistently downregulated in all 3 of the highly disease-sensitive strains compared with the control strains (shown enlarged on the right, Figure 1A). Intriguingly, 10 of these genes belonged to the kallikrein (*Klk*) family, with the highest expression levels and differences being noted in *Klk1*, *Klk1b3*, *Klk1b5*, *Klk1b26*, and *Klk1b27*, as summarized in Table 1. In addition to the *Klk* genes displayed in Table 1 (all of which were found to be different between the 2 sets of strains at $P < 0.001$), a few other *Klk* genes, notably *Klk1b9* and *Klk1b21*, exhibited similar expression differences between the disease-sensitive strains and the control strains, though these differences did not reach statistical *P* values of 0.001 (data not shown). Hence, in total, 12 *Klk* genes were underexpressed in the kidneys of AIGN-sensitive strains, following anti-GBM challenge, compared with the control strains.

Next, renal *Klk* gene expression was examined before and after induction of AIGN, using real-time PCR as an orthogonal approach. Real-time PCR analyses validated the above microarray results, indicating that whereas the B6 and BALB/c control strains successfully upregulated *Klk* following anti-GBM Ab challenge, the DBA/1, 129/SvJ, and NZW strains were ineffective at doing so (Figure 1B). In contrast, the basal, predisease levels of renal *Klk* were similar in all strains (Figure 1B). These differences were also confirmed at the protein level by Western blot analysis of renal cortex samples from these 5 strains, as illustrated for *Klk1* (Figure 2A and data not shown). Parallel differences in Klk enzymatic activity were also noted in urine samples from the same mice following anti-GBM Ab challenge (Figure 2B). Thus, whereas the BALB/c control strain exhibited a robust increase in urinary Klk activity 10–14 days after challenge, in the AIGN-sensitive strains DBA/1 and 129/SvJ, Klk was not significantly upregulated over the same period (Figure 2B).

Given that Klk upregulation in the kidneys and urine of anti-GBM Ab–challenged mice correlated well with the subdued nephritis noted in the B6 and BALB/c control strains, we next asked whether Klk might have a disease-protective role in immune nephritis. Kallikreins act through the generation of bradykinins (BKs), which in turn exert their biological effects by binding BK (BK B1 and B2) receptors on various cells (25, 26). Selective receptor blockade using pharmacological inhibitors further indicated that the biological effects of BK (and Klk) were mediated by the BK B2 receptor, since blocking this receptor aggravated proteinuria, azotemia, and GN following anti-GBM Ab challenge in BALB/c mice (Figure 3, A–D). Importantly, the mice in which BK B2 receptor was blocked exhibited significantly more severe GN compared with the other groups of mice (average GN score of 2.1 versus 0.2, $P < 0.001$, as partly illustrated in Figure 3, C and D). Similar differences were noted when the BK B2 receptor blockers were administered to anti-GBM–challenged B6 mice (data not shown). Conversely, the administration of BK dampened the severity of anti-GBM disease in 129/SvJ mice, which otherwise develop severe AIGN following the experimental insult (Figure 3E and data not shown).

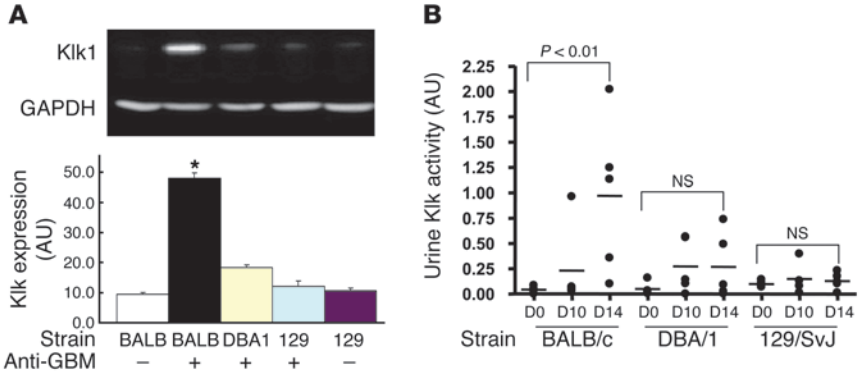


Figure 2

The differential renal expression of *Klk* in the AIGN-sensitive versus control strains was confirmed at the protein level. BALB/c, DBA1, and 129/SvJ mice were subjected to AIGN. (**A**) Fourteen days after anti-GBM challenge, kallikrein protein expression was assayed in the renal cortex by Western blotting, using a rabbit anti–mouse Klk1 Ab. The bar chart below ($n = 3$ mice per group) shows Klk expression normalized to GAPDH (AU). *$P < 0.001$, compared with all other study groups. (**B**) Twenty-four-hour urine samples collected from these mice on days 0, 10, and 14 after anti-GBM insult were also assayed for kallikrein enzymatic activity, using the synthetic chromogenic substrate HD-Val-Leu-Arg-pNA (S-2266), as detailed in Methods. Similar differences in renal and urinary kallikrein levels were noted between B6 and B6.*Sle3*^{z} mice (data not shown). Error bars in **A** denote SD. In **B**, each dot represents data from a single mouse, and the horizontal bars denote arithmetic group means.

With respect to the 3 AIGN-sensitive strains, DBA/1, 129/SvJ, and NZW, it is already known that several genomic intervals in the latter 2 strains (including loci on chromosomes 1 and 7) also contribute to spontaneous lupus nephritis (13–16). Indeed, the entire *Klk* gene complex is encoded within a lupus susceptibility interval on chromosome 7. The NZW-derived lupus susceptibility locus on chromosome 7, *Sle3*^{z} (which includes the *Klk* gene complex), has previously been introgressed onto the B6 genome as a congenic interval, and this locus had already been shown to facilitate development of spontaneous lupus nephritis (17–19). Through recursive backcrossing of B6.*Sle3*^{z} congenic mice to B6 parents and microsatellite-assisted selection, we generated B6.*Sle3*$^z_{157–158}$ recombinants harboring the NZW-derived *Klk* gene complex within a 4-Mb interval, with termini at *D7mit157* and *D7mit158*, as diagramed in Figure 4A.

Importantly, B6.*Sle3*^{z} mice and the newly generated B6.*Sle3*$^z_{157–158}$ congenic recombinants both displayed heightened sensitivity to AIGN, marked by elevated proteinuria and severe nephritis (Figure 4, B–D). Moreover, the renal cortex of both these congenics failed to efficiently upregulate *Klk* following anti-GBM Ab challenge, compared with the B6 controls (Figure 4E), as assessed by real-time PCR. Taken together with the functional data presented above, these findings suggest that the *z* allele of *Klk* positioned within the *Sle3*$^z_{157–158}$ subinterval may harbor important culprit genes for the heightened experimental anti-GBM disease (and spontaneous lupus nephritis) seen in NZW (and related strains of) mice.

Five of the most differentially expressed *Klk* genes, *Klk1*, *Klk1b3*, *Klk1b5*, *Klk1b26*, and *Klk1b27* were sequenced (GenBank accession numbers EU597301–EU597324). *Klk4* and *Klk1b8*, though differentially expressed, were not studied further because of their relatively low expression levels in all strains (Table 1). Several strain-specific differences were noted in the promoter regions of the 5 sequenced genes, as summarized in Table 2. The B6 and BALB/c *Klk* genes were almost identical, while the *Klk* genes from the AIGN-sensitive strains NZW and DBA/1 were closely related to each other, as illustrated by the phylogenetic trees in Figure 5A and the promoter region sequence exemplified in Figure 5B. The *Klk* genes from the 129/SvJ strain clustered more closely with the B6/BALB/c genes in some cases, and with the NZW/DBA/1 genes in others; despite repeated attempts, we could not amplify some of the critical regions of the 129/SvJ *Klk1b26* gene (Table 2). In contrast to the promoter regions, no sequence differences were noted in the coding regions or the 3′ untranslated regions of these *Klk* genes, when the 5 strains were compared (data not shown).

Some of the observed strain-specific promoter sequence differences in *Klk* fell within putative transcription factor binding sites. In the promoter region of *Klk1b3*, for example, SNPs were noted in potential transcription factor binding sites for PBF, TFIID, HoxD9, TCF-4E, NF-S, NF-E, and LBP-1, which distinguished the AIGN-sensitive strains from the controls. In particular, the GGCTT[C→G]AAAAT SNP in the promoter region of *Klk1b3* is predicted to abrogate TFIID binding. Importantly, promoter-luciferase studies indicated that sequence differences in the promoter region of *Klk1b3* may contribute in part to the reduced expression of this gene in the disease-sensitive strains (Figure 5C). Given that more than 200 sequence variations have been noted in the regulatory regions of the *Klk* genes (Table 2), each difference has to be systematically evaluated for its potential functional relevance.

As in murine lupus, the orthologous human interval encompassing the *KLK* genes on human chromosome *19q13* has also been implicated in human SLE susceptibility in previous genome scans (40–43). To ascertain whether *KLK* might also be a culprit gene in human SLE, we examined several SNPs in the renal-expressed *KLK* genes encoded within this interval, specifically human *KLK1*, *KLK5*, *KLK6*, and *KLK7*. We first genotyped a set of German SLE patients ($n = 340$) and a set of controls matched for ethnicity, age, and sex ($n = 400$). As detailed in Table 3, association to two *KLK1* SNPs, *rs1054713* (a synonymous coding SNP) and *rs2740502* (an intronic SNP), but not to the nonsynonymous substitution *rs5517*, was observed, with the strongest associations being noted in SLE patients with nephritis (*KLK1* SNP, *rs2740502*; $P = 0.007$ compared with non-nephritic SLE patients and $P = 0.01$ compared with healthy controls). A weak association was also observed for SNP *rs1897604* located in *KLK5*; however, this SNP was not in Hardy-Weinberg equilibrium in the controls (data not shown). No disease association was detected for any of the other *KLK* genes analyzed in this preliminary study.

To validate these preliminary associations, we genotyped 6 more cohorts of patients, including additional European, European-American, and Korean patients with SLE, as detailed in Tables 4 and 5. For the *KLK1* SNP *rs2740502*, we could replicate the genetic association to lupus nephritis compared with controls in the German and Italian samples but not in the remaining patient sets (Table 5). Likewise, the German SLE patients exhibited association

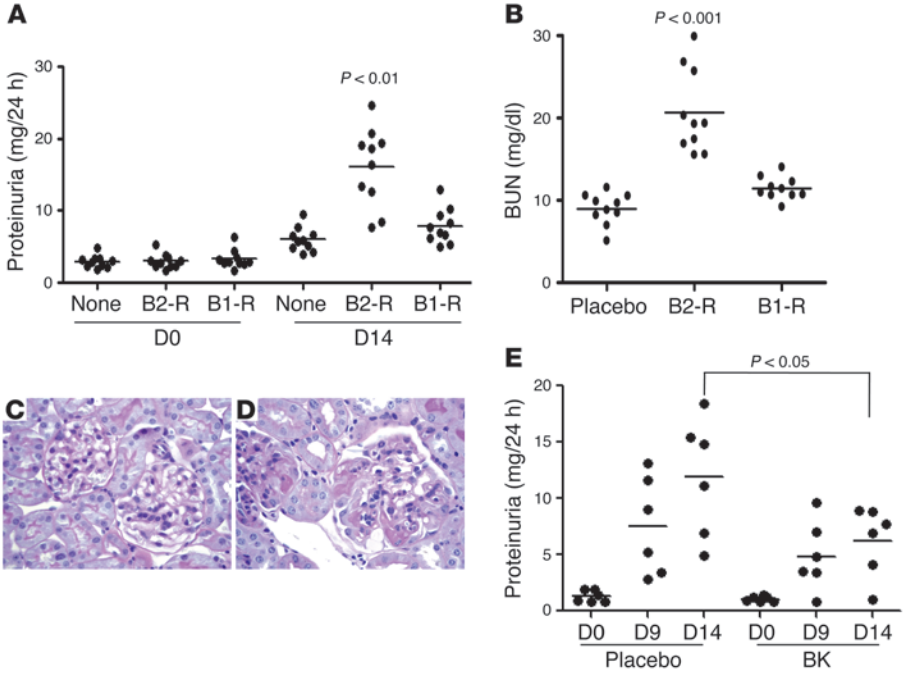


Figure 3
Impact of BK receptor blockade or activation on the severity of AIGN. BALB/c mice were treated with BK receptor antagonists (B1 receptor blocker H158 [B1-R] or B2 receptor blocker H157) or PBS vehicle alone as placebo (None), using osmotic pumps, for the 14-day duration of an anti-GBM challenge and phenotyped for proteinuria (**A**), azotemia (**B**), and GN (**C**, B1-R blocked; and **D**, B2-R blocked). Original magnification, ×400. (**E**) Conversely, the administration of BK into 129/SvJ mice using osmotic pumps ameliorated disease after anti-GBM challenge, compared with mice treated with vehicle placebo. In **A**, **B**, and **E**, each dot represents data from a single mouse, and the horizontal bars denote arithmetic group means.

to another KLK1 SNP, *rs5516* (Table 4). Due to likely heterogeneity in haplotype distribution and potential differences in linkage disequilibrium between the analyzed SNPs and functional SNPs in the different ethnic groups, we were unable to perform a Mantel-Haenszel metaanalysis with these data sets.

For further independent validation, we gained access to the genotype data for SNPs of the entire KLK region typed in 689 SLE patients and 3,718 healthy controls from the International Consortium for Systemic Lupus Erythematosus Genetics (SLEGEN) and 595 SLE patients from the UCSF Lupus Genetics Project cohort, both of which had been assembled for genome-wide association studies. Although the sets of SNPs spanning the entire KLK locus utilized in these studies were not completely overlapping, 56 SNPs were, and these were examined further for disease association. The strongest association with SLE was again noted to SNPs close to *KLK1*, in an intergenic region bordered at the centromeric end by *KLK1* and *KLK15* and at the telomeric end by *KLK3* and *KLK2* (Figure 6A and Supplemental Table 1; supplemental material available online with this article; doi:10.1172/JCI36728DS1). Additional associations to SNPs in *KLK4* promoter, *KLK5*, *KLK7*, *KLK11*, and *KLK12* were also uncovered when cases with nephritis were compared with cases without nephritis (Figure 6B and Supplemental Table 2). Next, we defined the haplotype blocks across the *KLK* locus (Figure 6A) and examined them for association to SLE. Again, the haplotype blocks harboring the promoter region of *KLK3* (block 1) displayed the strongest association, followed by the haplotype containing *KLK8* to *KLK11* genes (block 8) (Figure 6C and Supplemental Table 3). However, when only the patients with nephritis were considered as cases, the association to the *KLK3* promoter region was weaker; instead, the haplotypes harboring the *KLK4* promoter, *KLK5*, and the *KLK8–KLK11* block yielded stronger association. Of note, the 2 strongest associations observed corresponded to haplotypes with significantly lower frequencies in patients, reflecting a protective effect conferred by the KLK polymorphism.

Overall, these 3 sets of independent findings support a likely genetic association of the human *KLK1* gene and the *KLK3* promoter region with lupus and lupus nephritis.

Discussion

AIGN is an experimental tool that shares downstream molecular cascades with spontaneous lupus nephritis, as recently reviewed (9). Over the past decade, the roles of about 25 different molecules (including various complement proteins and TLR ligands, FcR, B7/CD28/CTLA4, LFA1/ICAM1, P-selectin, TNF-α, IL-1, IL-6, IL-12, IL-18, IFN-γ, M-CSF, PDGF, MCP-1, and NO) have been directly assessed (using gene knockouts or by deliberate experimental modulation of the molecules) in 2 disease settings — anti-GBM disease and lupus nephritis. Importantly, the effects of each of these molecules were consistent in both disease settings (9). In other words, molecules documented to influence the progression of experimental anti-GBM disease also impacted the development of lupus nephritis in the same direction. Hence, although experimental anti-GBM nephritis and spontaneous lupus nephritis differ in the nature of the inciting Abs and the localization of the immune deposits, pathology in both settings may be mediated by a shared network of downstream molecular pathways, including complement- and FcR-dependent activation of resident renal cells and infiltrating leukocytes, T cell help, proinflammatory mediators initially, and profibrotic molecules later in disease (9). In the present study, we have identified kallikrein as an additional molecule that appears to impact both diseases concordantly.

Of more than 20 inbred strains tested for AIGN susceptibility, DBA/1, 129/SvJ, and NZW are particularly sensitive to disease in this model. Though the genetic and molecular origins of increased disease susceptibility may vary among these strains, it is remarkable that all 3 strains shared approximately 50 genes that were underexpressed in their renal cortex during disease, compared with other strains. Even more remarkable is the finding that one-fifth of these 50 genes belonged to the same family

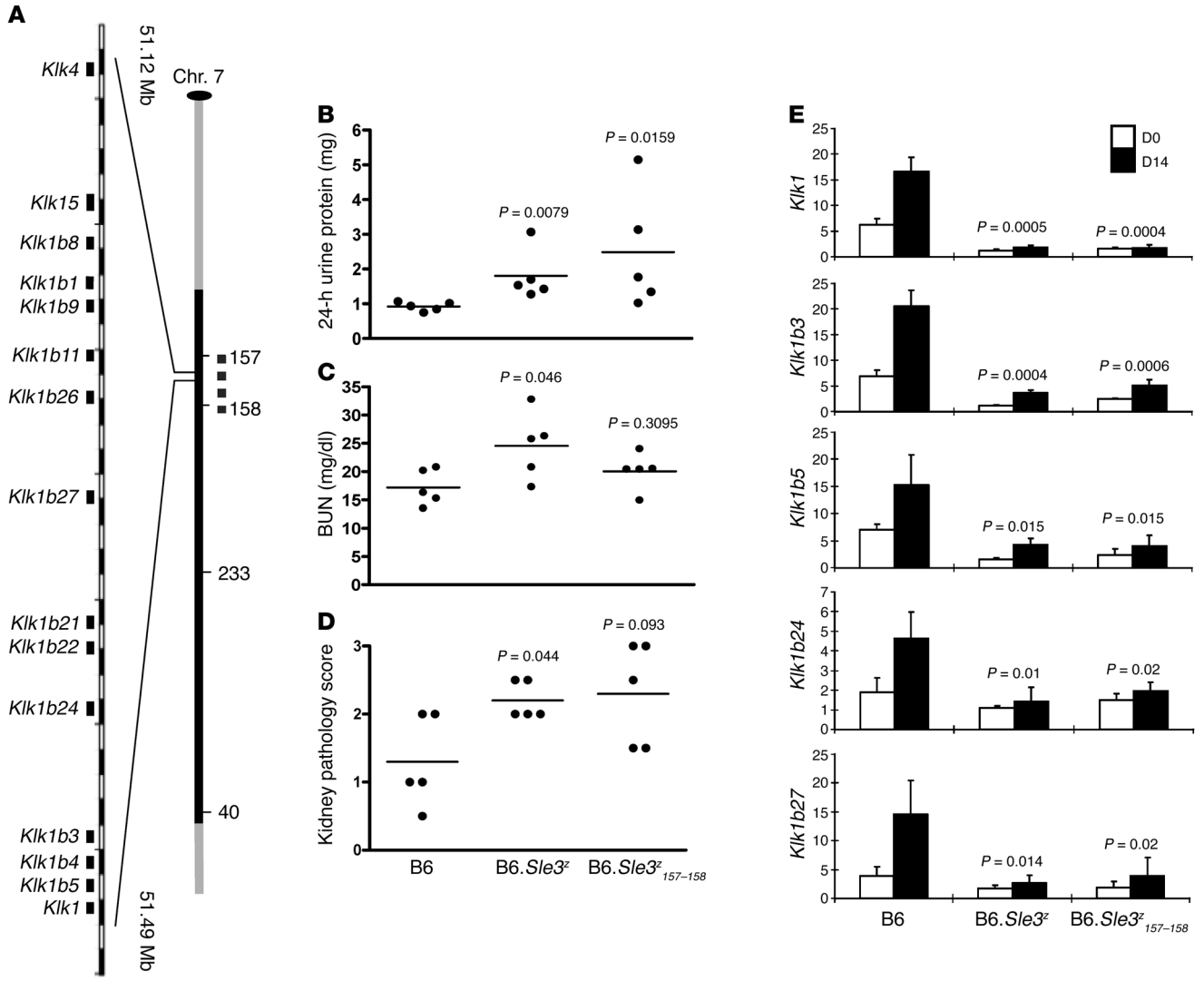


Figure 4

The *Sle3*^{z} locus, particularly the *Sle3*$^{z}_{157–158}$ subinterval on chromosome 7, may be responsible for the reduced *Klk* and enhanced nephritis susceptibility seen in NZW mice. (**A**) Shown are the *Sle3*^{z} lupus susceptibility interval on chromosome 7 (Chr. 7; black denotes the interval derived from NZM2410/NZW; gray denotes B6 origin); the 4-Mb subinterval spanning *D7mit157* to *D7mit158* (denoted by the dashed line on right); and the cluster of *Klk* genes harbored within the indicated subinterval. The numbers on the right refer to the positions of respective microsatellite markers (e.g., 157 represents *D7mit157*). Shown also are the 24-hour urine protein excretion profiles (**B**), blood urea nitrogen (BUN) (**C**), GN pathology score (**D**), and renal *Klk* message levels (**E**), 14 days after anti-GBM challenge of B6, B6.*Sle3*^{z}, and B6.*Sle3*$^{z}_{157–158}$ congenics (n = 5 each). The data shown in **B**–**D** were reproduced in at least 2 additional experiments. In the second study, for example, the B6.*Sle3*$^{z}_{157–158}$ congenics exhibited significantly higher 24-hour protein levels in urine ($P < 0.045$) and GN score ($P < 0.013$) and more severe tubulo-intersitial disease ($P < 0.001$), compared with the B6 control (data not plotted). All statistical comparisons were made with the respective B6 controls, using the Mann-Whitney U test. In **B**–**D**, each dot represents data from a single mouse, and the horizontal bars denote arithmetic group means. Error bars in **E** denote SD.

of genes, the kallikreins. This strain difference in *Klk* gene expression may originate from polymorphisms in the *Klk* genes themselves or they may be the consequence of yet other candidate genes in the AIGN-susceptible strains. The finding that a similar renal *Klk* expression profile was recapitulated in B6 mice bearing a 4-Mb congenic interval harboring the NZW allele of *Klk* (Figure 4) suggests that the observed strain differences are likely to originate from sequence differences in the *Klk* locus itself. The observation that the AIGN-sensitive strains have *Klk* sequences that are related to each other but diverge from the *Klk* sequences of the 2 control strains, B6 and BALB/c, offers further evidence that the strain differences in renal Klk upregulation must be intrinsically encoded within the *Klk* genetic locus.

Though many sequence polymorphisms in the promoter regions of the *Klk* genes were uncovered in this study, with several falling within transcription factor binding sites, the functional relevance of these regulatory sequence differences to the observed expression differences warrants systematic study. At the very least, our completed studies indicate that sequence polymorphisms within the promoter region of *Klk1b3* may be contrib-

Table 2
Sequence polymorphisms in the promoter region of the mouse *Klk* genes

Gene	position	B6 and BALB/c	129/SvJ	DBA1 and NZW
Klk1b3	–70 to –80	–	TGTCAGGGAGG	TGTCAGGGAGG
	–157	G	A	A
	–163	C	G	C
	–191	G	A	A
	–205	A	G	G
	–275	G	A	G
	–287	T	G	G
	–298	C	T	T
	–301	T	C	C
	–423	T	A	A
	–539	C	T	T
	–607	C	G	G
	–734	T	G	G
	–735 to –736	–	T	GG
	–822	G	A	G
	–885	G	T	G
	–912	C	C	T
	–973	T	C	C
	–1,004	A	A	–
	–1,095 to –1,096	GG	AA	AA
	–1,167	A	T	T
	–1,170	G	G	A
	–1,171	A	G	G
	–1,249	T	C	T
	–1,262	A	G	G
	–1,271	A	G	G
	–1,274	C	G	G
	–1,317	A	G	G
	–1,368	G	A	A
	–1,426	C	T	T
	–1,430 to –1,433	–	TCGA	TCGA
	–1,446	C	C	–
	–1,477	G	A	G
Klk1b5	–516	A	A	G
	–657	C	A	A
	–706	T	T	C
	–866	G	G	A
Klk1b26	–144	C	N/A	T
	–153	T	N/A	C
	–158	T	N/A	C
	–162	A	N/A	C
	–192	A	N/A	G
	–341	C	N/A	T
	–734 to –753	TATTTTATTTTATTTTATTT	N/A	–
	–1,034	A	N/A	G
	–1,282	C	N/A	A
Klk1	–170	G	G	A
	–202	G	T	G

uting to the observed expression differences in this gene, when B6.*Sle3*^{z}/NZW kidneys are compared with the B6 control (Figure 5B). Alternatively, the entire *Klk* gene complex may be differentially regulated in the 2 haplotypes by polymorphic regulatory regions located within or perhaps even outside this gene complex. Further sequencing of both haplotypes and *Klk* allele–specific knock-in studies are warranted to obtain definitive evidence that the observed *Klk* polymorphisms are indeed responsible for the strain-specific phenotypic differences.

Having established that the reduced Klk observed in the AIGN-sensitive strains is genetically encoded by the *Klk* locus, we next examined whether reduced Klk may be responsible for the increased severity of immune-mediated nephritis seen in these susceptible strains. This indeed appears to be the case, given that subduing Klk action (using B2 receptor blockers) rendered BALB/c control mice AIGN susceptible, while the administration of BK ameliorated nephritis in AIGN-susceptible 129/SvJ mice (Figure 3). These findings resonate well with previous genetic and pharmacological studies that also reveal a renoprotective role for Klk and BKs in nephritis following other insults, including hypertension, ischemic stroke, salt imbalance, and diabetes (26–35). Previous studies suggest that Klk may be playing a protective role in nephritis by modulating several different parameters, including local blood pressure, the inflammatory milieu, redox balance, and/or signaling within various cell types (25, 26, 31, 36–39).

The observation that the NZW-derived *Sle3*^{z} lupus-susceptibility locus confers susceptibility to spontaneous lupus nephritis raises the interesting possibility that polymorphisms in *Klk* may also confer susceptibility to spontaneous lupus nephritis. Though it is formally possible that additional genes within the *Sle3*^{z} interval may be contributing to the increased susceptibility to lupus nephritis associated with this locus (14, 17–19), the observation that a recombinant congenic harboring a 4-Mb interval spanning the *Klk* locus confers susceptibility to experimental immune nephritis (Figure 4) suggests that the *Klk* genes are likely to be major players in spontaneous lupus nephritis as well. Efforts are underway to evaluate whether long-term modulation of Klk levels in vivo can impact the severity of spontaneous lupus nephritis.

The corresponding interval on human chromosome *19q13* bearing the *KLK* gene complex has also been implicated in human SLE susceptibility in previous genome scans, particularly in patients of European descent (40–43). The observation that SNPs in *KLK1* show significant association to lupus nephritis in European-descent SLE patients suggests that *KLK* gene polymorphisms may also influence human lupus nephritis, at least in some ethnic groups. The observed heterogeneity among different populations in some of our earlier studies (as summarized in Tables 4 and 5) may be due to haplotype differences among the populations. However, the larger and more comprehensive SNP and haplotype analysis conducted in the SLEGEN and UCSF studies clearly validates the existence of a disease locus in *KLK1*. It is intriguing that most of the disease-associated SNPs are located in the regulatory region between *KLK1* and *KLK3*, with these genes being transcribed in opposite orientations to each other (Figure 6A). This region possesses many regulatory elements sensitive to steroid hormones extensively studied in prostate cancer, which have been shown to affect *KLK3* (prostate-specific antigen) transcription. Incidentally, the murine *Klk* genes that are

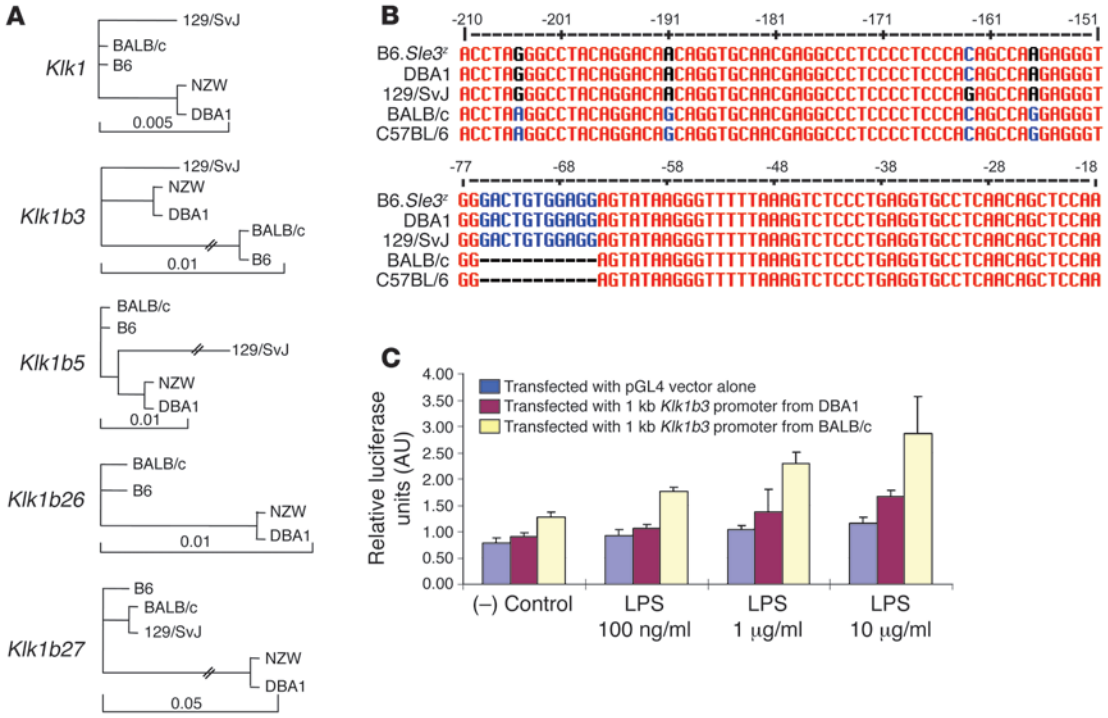


Figure 5

Sequence analysis of the 5′-regulatory regions of *Klk* genes reveals nucleotide polymorphisms that distinguish the AIGN-sensitive strains from the control strains. When 2 kb of the 5′-regulatory regions of *Klk1*, *Klk1b3*, *Klk1b5*, *Klk1b26*, and *Klk1b27* from the indicated strains were sequenced, several SNPs/deletions were identified, as detailed in Table 2 (GenBank accession numbers EU597301–EU597324). (**A**) Phylogenetic analysis revealed the sequence of the AIGN-sensitive strains to be closely related, compared with the sequences of the 2 control strains. Bars represent the fraction of sequence variation. (**B**) Part of the nucleotide sequence of the *Klk1b3* promoter (up to 200 bp upstream of transcription start site) from the different study strains indicated. Note that the B6.*Sle3*^{z} strain bears the NZW allele at *Klk*. (**C**) One kilobase of the promoter region of *Klk1b3* from both BALB/c and DBA/1 strains was inserted into the pGL4 luciferase vector and transfected into COS-7 cells, and luciferase activity was assayed 24 hours later, as detailed in Methods. Each bar represents the median of triplicate values, and the error bars denote SD. Cells transfected with vectors carrying the BALB/c-derived *Klk1b3* promoter showed significantly increased luciferase activity compared with cells with vectors bearing the DBA/1 promoter or vector alone ($P < 0.01$). Similar differences were noted when the B6 *Klk1b3* promoter was compared with the B6.*Sle3*^{z} *Klk1b3* promoter (data not shown).

most differentially expressed in disease-prone kidneys (as listed in Table 1) are most homologous to human *KLK1* and *KLK3*, raising the interesting possibility that regulatory polymorphisms shared by the two species could be at play in dictating reduced kallikrein production, which could potentially lead to enhanced nephritis in mice and patients with lupus.

In follow-up studies, it would be important to study SLE patients with clearly defined and graded nephritis (as determined using uniformly applied diagnostic criteria across different patient sets originating from different collaborative groups) to further validate the association to *KLK1/KLK3*, to elucidate any corresponding promoter polymorphisms, and to define the functional consequences of any such polymorphisms. Interestingly, familial clustering of reduced renal Klk expression as well as African American hypertensive patients with reduced renal Klk have been reported (44, 45). Given that African American ethnicity and coexisting hypertension are both risk factors for severe lupus nephritis (46, 47), it would be important to establish the potential role of *KLK* gene polymorphisms in driving disease severity in human SLE. Along these lines, there is some evidence that promoter polymorphisms in human *KLK1* may be associated with reduced renal and urinary Klk, as well as with hypertension and nephritis (48, 49).

Whereas some lupus susceptibility genes may impact the adaptive arm of the immune system, others appear to augment innate immunity (50–52). The present study uncovers yet another class of lupus susceptibility genes — those that may potentially modulate end-organ disease. Besides their central relevance to the genetics of lupus nephritis, these findings have additional clinical implications. Monitoring urinary kallikreins as a marker of renal disease in lupus and developing kallikrein-based therapeutics for modulating renal lupus may be future directions worthy of pursuit.

Methods

Mice and AIGN. BALB/c, DBA/1, 129/SvJ, B6, and NZW mice were purchased from The Jackson Laboratory. B6.*Sle3*^{z} congenic mice, described previously (17, 18), were bred in-house. Recursive backcrossing of B6.*Sle3*^{z} mice to B6 parents and microsatellite-assisted selection were used to generate B6.*Sle3*$^{z}_{157–158}$ recombinants harboring the NZW-derived *Klk* gene complex within a 4-Mb interval, with termini at *D7mit157* and *D7mit158*. All mice were maintained in a specific pathogen–free colony. Two- to 3-month-old females were used for all studies. To induce nephritis, we first sensitized mice on day 0 with rabbit IgG (250 μg/mouse, i.p.), in adjuvant, as described previously (10, 11). On day 5, the mice were challenged i.v. with rabbit anti-GBM Ig (200 μg per 25 g body weight, in a 300-μl volume).

Table 3
Association analysis of KLK SNPs in German SLE patients and controls

KLK gene	SNP	Allele	Allele frequencies[A] Cases	Reference group	χ^2	*P*[B]
SLE/total[C] vs. healthy controls						
KLK1	*rs5517*	C	0.325	0.321	0.025	0.874
KLK1	*rs1054713*	G	0.694	0.64	3.833	**0.050**
KLK1	*rs2740502*	C	0.385	0.346	1.94	0.164
KLK5	*rs1897604*[D]	C	0.33	0.275	4.441	**0.033**
KLK5	*rs268908*	A	0.873	0.854	0.981	0.322
KLK5	*rs2569522*	T	0.63	0.603	0.965	0.326
KLK6	*rs1654537*	G	0.345	0.34	0.033	0.855
KLK7	*rs1701924*[D]	C	0.346	0.345	0.001	0.977
SLE/nephritis vs. SLE/no nephritis						
KLK1	*rs5517*	T	0.716	0.654	1.006	0.316
KLK1	*rs1054713*	G	0.757	0.658	2.503	0.114
KLK1	*rs2740502*	C	0.5	0.326	7.271	**0.007**
KLK5	*rs1897604*[D]	T	0.706	0.672	0.292	0.589
KLK5	*rs268908*	A	0.914	0.871	0.974	0.324
KLK5	*rs2569522*	T	0.662	0.61	0.672	0.412
KLK6	*rs1654537*	A	0.712	0.65	0.909	0.341
KLK7	*rs1701924*[D]	C	0.348	0.346	0.002	0.968
SLE/nephritis vs. healthy controls						
KLK1	*rs5517*	T	0.716	0.679	0.425	0.514
KLK1	*rs1054713*	G	0.757	0.64	3.822	**0.051**
KLK1	*rs2740502*	C	0.5	0.346	6.563	**0.010**
KLK5	*rs1897604*	C	0.294	0.275	0.117	0.732
KLK5	*rs268908*	A	0.914	0.854	1.941	0.164
KLK5	*rs2569522*	T	0.662	0.603	0.993	0.319
KLK6	*rs1654537*	A	0.712	0.66	0.743	0.389
KLK7	*rs1701924*	C	0.348	0.345	0.003	0.957

[A]The allele frequencies were determined by SNP-typing an average of 722 normal control alleles, 508 SLE patient alleles, 260 non-nephritic SLE patient alleles, and 70 nephritic SLE patient alleles. [B]Significant associations are indicated in bold. [C]Total includes patients with and without nephritis. [D]These SNPs were not in Hardy-Weinberg equilibrium in controls.

The source and preparation of the anti-GBM serum/Ig have been detailed previously (10, 11). Twenty-four-hour urine samples were collected from experimental mice on days 0, 10, and 14, using metabolic cages, with free access to drinking water. Urinary protein concentration was determined using the Coomassie Plus protein assay kit (Pierce, Thermo Scientific). Serum was collected on days 0, 10, and 14 for measurement of blood urea nitrogen (BUN), using a urea nitrogen kit (Sigma-Aldrich). Animals were sacrificed on day 10 or 14, and the kidneys were processed for histopathological examination by light microscopy, as detailed previously (10, 11).

Microarray and real-time PCR analysis. Kidneys collected on day 10 of experimental nephritis were used for gene expression analysis by microarray. Total RNA was isolated from renal cortex using RNeasy (QIAGEN) and quality-checked using an Agilent Bioanalyzer (Agilent Technologies). Sentrix Mouse-6 Whole Genome Expression BeadChips (Sentrix Mouse-6 v1_1; Illumina) were used for the microarray analysis, according to the manufacturer's instructions. Microarray data were extracted using BeadStudio v3.1, background-subtracted, and normalized using a cubic spline algorithm. Genes differentially expressed between groups were identified using the Illumina custom error model implemented in BeadStudio. Genes were considered significantly differentially expressed when *P* values were less than 0.001 and the change was greater than 2-fold. Supervised hierarchical clustering of transcripts that were differentially expressed between the groups was performed. Data were median centered; in Figure 1A, green represents expression below median; red represents above-median expression; and gray represents missing data.

Renal cortex *Klk* gene expression differences were validated by quantitative RT-PCR using validated TaqMan assays (Applied Biosystems) and Mm00834006_g1 for *klk1*, Mm01203825_gH for *klk1b3*, Mm00833453_g1 for *klk1b5*, Mm01702809_m1 for *klk1b26*, and Mm00834759_gH for *klk1b27*. Transcription of eukaryotic 18S rRNA (assay ID Hs99999901_s1) was used as an internal control.

Generation of rabbit anti-mouse Klk1 Ab. Rabbit Abs against mouse Klk1 were developed by Abgent, using a peptide sequence from mouse Klk1, EKNSQPWQVAVYRFTKYQC, conjugated to keyhole limpet hemocyanin (KLH). The peptide was used to immunize rabbits (5 mg per rabbit, 3 injections administered 21 days apart). Sera obtained from the rabbits 2 months following the primary immunization reacted strongly with recombinant mouse Klk1 protein (titer, >1:4,000). Immune rabbit sera were harvested 2 weeks after the third injection, protein G purified, aliquoted, and stored at –80°C.

Assaying Klk protein. BALB/c, DBA1, and 129/SvJ mice were subjected to AIGN, as described above. Mice (3 per group) were sacrificed on day 0 or 14. Renal cortex was homogenized in lysis buffer (25 mM Tris-HCl, pH 7.4, 1% Triton X-100, 0.1% SDS, 2 mM EDTA) containing 1:100 protease inhibitor cocktail (Sigma-Aldrich) and centrifuged at 4°C for 30 minutes. The supernatants were stored at –80°C. Protein concentrations were determined by Lowry's method. These homogenates were subjected to Western blot analysis using the custom-generated rabbit anti–mouse Klk1 Ab (1:1,000; described above) and chemiluminescence detection (Pierce, Thermo Scientific). Urinary kallikrein enzymatic activity was assayed using the synthetic chromogenic substrate HD-Val-Leu-Arg-pNA (S-2266). Briefly, 50 μl of urine was added to 50 μl of assay buffer (0.2 M Tris-HCl, pH 8.2, containing 300 μg/l soybean trypsin inhibitor and 375 μg/l EDTA) and incubated at 37°C for 30 minutes. Then, 50 μl of S-2266 was added and incubated at 37°C for 3 hours, and the absorbance was read at 405 nm. A standard curve was constructed using purified kallikrein (human urinary kallikrein [HUK]; Calbiochem), and the optical densities were converted to activity units.

In vivo studies using BK agonists and antagonists. Two- to 3-month-old female BALB/c mice were divided into 3 groups (10 mice per group), and subjected to anti-GBM Ab challenge. The mice in the first 2 groups were infused with either kinin B2 receptor antagonist H157 (HOE140 or icatibant) or B1 receptor antagonist H158, using subcutaneous osmotic minipumps (DURECT; dosage, 0.74 μg/h) from day 3 to day 14. The third group was infused with PBS vehicle as control. Urine and sera were collected on days 0 and 14 for proteinuria and BUN assessment. All mice were sacrificed on day 14, and kidneys were scored for pathology. For the in vivo agonist studies, 129/SvJ mice were subjected to anti-GBM disease, as described above. One group received BK (Sigma-Aldrich; 0.5 μg/h released by osmotic minipumps, from day 3 to day 14), while the other group received PBS. All mice were phenotyped for disease as detailed above.

Amplification and sequencing of the 5′-flanking region of the Klk genes. Two-kilobase fragments of the 5′-flanking region of various *Klk* genes were sequenced after amplifying with the following primers: *Klk1*-F: 5′-AATCACCTC-TACCACCCAGTAC-3′, *Klk1*-R: 5′-TTGGAGCTGTAGAGGCTCCTG-3′,

Table 4
Genotypic and allelic association of *KLK1* rs5516 in 6 additional sets of SLE patients with nephritis (cases) and healthy controls

Population		CC	CG	GG	*P*	OR (CI)	Allele G	Allele C	*P*	OR (CI)
Germany	Cases (40)	7.50%	27.50%	65.00%	0.010	CG: 0.97 (0.26–3.61)	78.80%	21.20%	**0.008**	2.10 (1.18–3.92)
	Controls (352)	12.50%	47.40%	40.10%		GG: 2.70 (0.78–9.37)	63.80%	36.20%		
European-American	Cases (579)	11.74%	41.97%	46.29%	0.446	CG: 0.84 (0.63–1.13)	67.27%	32.73%	0.991	1.00 (0.87–1.14)
	Controls (3089)	10.49%	44.51%	45.00%		GG: 0.92 (0.69–1.23)	67.25%	32.75%		
Argentina	Cases (136)	4.40%	41.20%	54.40%	0.040	CG: 2.68 (1.08–6.63)	75.0%	25.0%	**0.028**	1.42 (1.03–1.98)
	Controls (367)	11.70%	40.90%	47.40%		GG: 3.05 (1.24–7.47)	67.8%	32.2%		
Spain	Cases (125)	16.80%	47.20%	36.00%	0.180	CG: 0.67 (0.38–1.19)	59.6%	40.4%	0.089	0.78 (0.59–1.05)
	Controls (569)	11.20%	46.90%	41.80%		GG: 0.58 (0.32–1.04)	65.3%	34.7%		
Italy	Cases (92)	10.90%	57.70%	30.40%	0.149	CG: 1.27 (0.57–2.85)	59.8%	40.2%	0.194	0.79 (0.55–1.15)
	Controls (216)	11.10%	47.20%	41.70%		GG: 0.75 (0.32–1.75)	65.3%	34.7%		
Korea[A]	Cases (296)	4.40%	35.10%	60.50%	0.140	CG: 1.79 (0.94–3.39)	78.00%	22.00%	0.122	1.19 (0.95–1.51)
	Controls (785)	7.8%	34.8%	57.5%		GG: 1.86 (1.00–3.47)	74.80%	25.20%		

[A]Hardy-Weinberg equilibrium, $P = 0.037$. Significant associations are indicated in bold.

Klk1b26-F: 5′-CAGAGAGAATCATAAAAGAGTGC-3′, *Klk1b26*-R: 5′-AATACCCTCATTCTCTCCATAGT-3′, *Klk1b3*-F: 5′-AGGTTTACCATGAAAGAGTTAGG-3′, *Klk1b3*-R: 5′-TTGGAGCTGTTGAGGCACCTC-3′, *Klk1b5*-F: 5′-GTCACCTCTAGCACCCAATGG-3′, *Klk1b5*-R: 5′-CTGTCCAGGAGCTGCAGGCT-3′, *Klk1b27*-F: 5′-AGAAAGACTCCTGGAAGAGTGG-3′, and *Klk1b27*-R: 5′-GTGAACTTGGAGCTGTTGAGGA-3′. The DNA sequences of these PCR fragments were analyzed on an ABI 3100 sequence analyzer (Applied Biosystems) and deposited into GenBank (accession numbers EU597301–EU597324). Based on the sequence information, phylogenetic trees were constructed using publicly available software (Protein Information Resource, http://pir.georgetown.edu/pirwww/search/multialn.shtml).

Promoter function studies. One kilobase of the promoter sequence upstream of *Klk1b3* was PCR amplified from B6, BALB/c, DBA/1, and NZW/B6.*Sle3*^{z} mice using the primers 5′-CCGGTACCGCCACCAAGCTTAACCCTGA-3′ and 5′-CCCTCGAGGCTTGGAGCTGTTGAGGCAC-3′, cloned separately into pGL4.10[luc2] luciferase reporter vector (Promega), and sequence verified. COS-7 cells (2×10^5 to 5×10^5 cells in 96-well plates) were transfected with 0.1 μg of pGL4.10[luc2]-promoter luciferase construct or vector using PolyFect (QIAGEN), together with 0.01 μg of an internal control plasmid, pGL4.73[hRluc/SV40] (Promega). Cells were lysed after 48 hours of LPS stimulation, using 1× passive lysis buffer (Promega) and assayed for luciferase activity, according to the manufacturer's instructions. Each construct was measured in at least 8 replicates. The results were normalized against *Renilla* luciferase control and presented as median values.

Human KLK genetics. Several independent sets of cases and controls were used in this study, originating from Germany, Italy, Spain, Argentina, the United States, Mexico, and Korea (53). The European sets and the Latin American sets have been previously described (54, 55). All patients fulfilled American College of Rheumatology criteria for the classification of SLE (56). All human studies were approved by the Central Ethical Review Board, Sweden, and the local ethical review boards of Instituto de Biomedicina y Parasitología López Neyra; Sanatorio Parque; Hospital Carlos Haya; Hospital for Rheumatic Diseases, South Korea; University of Hannover; University of Eastern Piedmont; U.O.C. di Reumatologia Azienda Ospedaliera San Camillo-Forlanini; Oklahoma Medical Research Foundation; University of Alabama at Birmingham; Feinstein Institute for Medical Research; Medical University of South Carolina; and Hammersmith Hospital. For the initial analysis, SNPs on human renal-expressed *KLK* genes *KLK1*, *KLK5*, *KLK6,* and *KLK7* were selected from the HapMap-CEU genotype data (http://www.hapmap.org) using Haploview version 3.2 (Broad Institute of MIT and Harvard), and were genotyped using protocols described earlier (55). Differences between patients and controls

Table 5
Genotypic and allelic association of *KLK1* rs2740502 in 6 additional sets of SLE patients with nephritis (cases) and healthy controls

		CC	CG	GG	*P*	OR (CI)	Allele C	Allele G	*P*	OR (CI)
Germany	Cases (35)	25.70%	48.60%	25.70%	0.036	CG: 1.831 (0.79–4.23)	50.0%	50.0%	**0.010**	1.89 (1.12–3.19)
	Controls (364)	12.40%	44.50%	43.10%		CC: 3.49 (1.31–9.31)	34.6%	65.4%		
European-American	Cases (579)	17.27%	48.53%	34.20%	0.277	CG: 1.09 (0.90–1.33)	41.50%	58.50%	0.113	1.11 (0.97–1.26)
	Controls (3089)	15.09%	47.94%	36.97%		CC: 1.24 (0.95–1.61)	39.10%	60.90%		
Argentina	Cases (134)	19.40%	41.00%	39.60%	0.057	CG: 0.85 (0.55–1.31)	39.9%	60.1%	0.242	1.19 (0.88–1.60)
	Controls (365)	11.50%	48.80%	39.70%		CC: 1.69 (0.95–3.03)	35.9%	64.1%		
Spain	Cases (119)	14.30%	48.70%	37.00%	0.915	CG: 0.96 (0.63–1.49)	38.7%	61.3%	0.693	0.94 (0.70–1.27)
	Controls (567)	15.70%	48.70%	35.60%		CC: 0.88 (0.48–1.62)	40.0%	60.0%		
Italy	Cases (93)	12.90%	47.30%	39.80%	0.023	CG: 0.59 (0.35–0.99)	36.6%	63.4%	**0.008**	0.63 (0.44–0.90)
	Controls (259)	21.60%	52.50%	25.90%		CC: 0.39 (0.18–0.81)	47.9%	52.1%		
Korea[A]	Cases (296)	5.10%	35.80%	59.10%	0.223	CG: 1.21 (0.91–1.61)	23.00%	77.00%	0.833	1.02 (0.81–1.29)
	Controls (785)	7.00%	31.10%	61.90%		CC: 0.76 (0.42–1.38)	22.50%	77.50%		

[A]Hardy-Weinberg equilibrium, $P = 0.003$. Significant associations are indicated in bold.

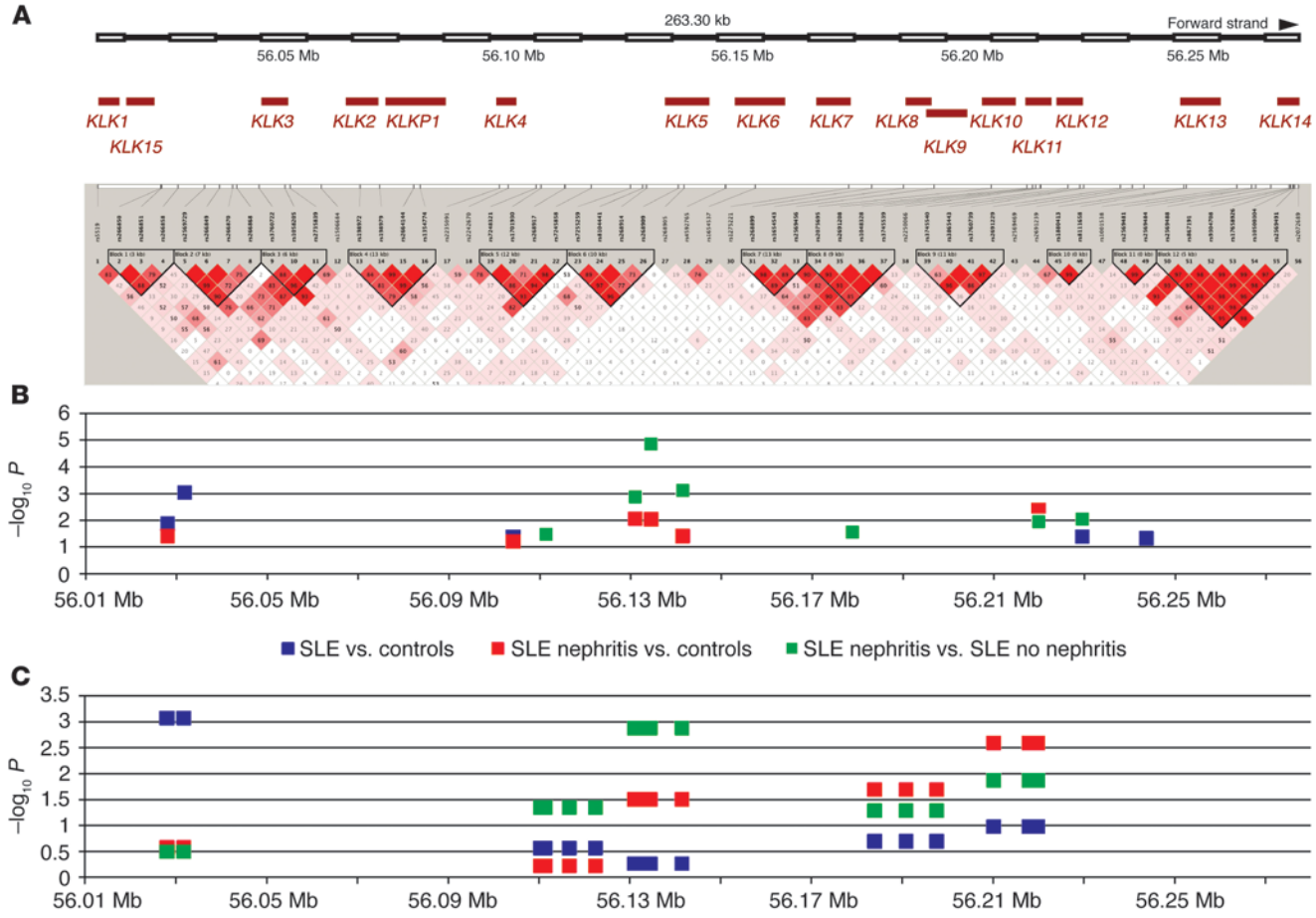


Figure 6
KLK association in SLE patients — second validation study. (**A**) The Haploview plot shows the genotyped markers in the KLK locus, from *KLK1* at the centromeric limit until *KLK14* at the telomeric end, as well as the linkage disequilibrium between them measured by the D prime coefficient. Blocks were defined by the solid spine method implemented in Haploview version 4.1. Dataset: UCSF (n = 595 SLE patients) plus SLEGEN (n = 689 SLE patients and n = 3,718 controls). (**B**) In the indicated numbers of SLE patients and healthy controls, 56 KLK SNPs were tested for disease association using logistic regression analysis, with the phenotype "SLE" as the outcome variable (shown in blue). To examine whether the risks conferred by the KLK polymorphisms were influenced by the presence of nephritis, we tested the KLK SNPs for association, considering the phenotype "nephritis" as the outcome variable. Red indicates significant differences compared with controls; green indicates significant differences between cases with nephritis and cases without nephritis calculated by a metaanalysis, in order to control for heterogeneity among the contributing clinical centers. (**C**) The observed linkage disequilibrium blocks across the *KLK* locus were tested for haplotype association, using both omnibus and haplotype-specific association statistics (T test) as implemented in PLINK. Shown are significant haplotype associations when SLE patients were compared with controls (blue), when patients with lupus nephritis were compared with controls (red), and the case-only analysis (green). Besides the SNPs indicated in **B** and the blocks indicated in **C**, none of the other SNPs/blocks shown in **A** showed significant disease associations. For a larger version of this figure, see Supplemental Figure 1.

were statistically analyzed using χ^2 test. Odds ratio (OR) with 95% CI was calculated using StatsDirect software for allelic differences and using UNPHASED software (http://www.mrc-bsu.cam.ac.uk/personal/frank/software/unphased/) for genotypic differences.

Validation of KLK associations using SLEGEN and UCSF datasets. We extracted the genotype data for the human KLK locus (Chr19:56014125-56277428) from 2 different genome-wide association studies (GWASs) conducted recently in SLE patients of European ancestry (57, 58). The patients and controls included in those studies, as well as their contributing centers and the genotyping platforms used, have been described in detail in the original publications. All patients fulfilled the American College of Rheumatology's revised criteria for classification of SLE (56) and were matched to healthy controls by sex, age, and ethnic origin (i.e., only individuals of European ancestry were included). From the GWAS performed by the SLEGEN, we obtained data for 63 *KLK* SNPs genotyped in 689 cases and 3,718 controls. From the UCSF Lupus Genetics Project, we added 595 patients who had genotype data for 113 *KLK* SNPs, of which 56 were also present in the SLEGEN dataset. Hence, in total, we had access to genotype data for 56 *KLK* SNPs pertaining to 1,284 (689 SLEGEN plus 595 UCSF) SLE patients and 3,718 controls, drawn from the 2 GWAS studies.

Statistics. For the murine studies, Student's t test (1-tailed) was used for statistical analysis, unless otherwise indicated. Error bars shown in figures represent standard deviations. For the human genetic studies, statistical analysis was performed using Haploview version 4.1 (59) and PLINK version 1.04 (60). We first verified that all the SNPs/alleles were concordant in terms of positive or negative strand. Then, quality control filters were applied to

remove individuals and SNPs with more than 5% of data missing, differential missing rate between cases and controls ($P < 0.05$), significant deviation from Hardy-Weinberg equilibrium in controls ($P < 0.001$), or a minor allele frequency of less than 1%. After this filtering step, we had 56 SNPs in 1,241 cases and 3,664 controls, with a total genotyping rate of 0.996054. Clinical data regarding the presence or absence of nephritis was available for 1,122 patients, of whom 319 had nephritis (28.4%).

In this filtered sample, we tested the allelic frequencies for significant association by logistic regression analysis using the phenotype "SLE" as the outcome variable. To examine whether the risk conferred by the KLK polymorphism on SLE was influenced by the presence of nephritis, we also performed the analysis considering the phenotype "nephritis" as the outcome variable. OR and 95% CI limits (L95, lower limit; U95, upper limit) were calculated controlling by the covariate "sex." We also performed a case-only χ^2 test to determine whether there were significant differences in allelic frequencies between the subsets of patients with and without nephritis. Since the patient samples had originated from different centers, combined ORs were estimated by Cochran-Mantel-Haenszel metaanalysis, including a Breslow-Day test for homogeneity of the OR between centers. Multiple testing was corrected by adjusting the *P* values using the false discovery rate (FDR) control.

Finally, we examined the linkage disequilibrium (LD) structure of the region in Haploview, and the observed blocks were tested for haplotype association. The data were analyzed using the omnibus test and haplotype-specific association statistics (T test) as implemented in PLINK. The case/control omnibus test is an H – 1 degree of freedom test, where H is the number of different haplotypes. We compared the haplotype frequencies in patients with SLE versus controls, patients with lupus nephritis versus controls, and patients with lupus nephritis versus patients without nephritis.

Note on authors. The members of the collaborative groups are as follows.

Italian Collaborative Group: Maria Giovanna Danieli, Clinica Medica, Dipartimento di Scienze Mediche e Chirurgiche, Università Politecnica delle Marche, Ancona; Mauro Galeazzi, Department of Clinical Medicine, Rheumatology Unit, Siena University; Patrizia Rovere Querini, IRCCS San Raffaele Scientific Institute, Milan; Sergio Migliaresi, Rheumatology Unit Second University of Naples.

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

German Collaborative Group: K. Armadi-Simab and Wolfgang L. Gross, Abteilung Rheumatologie, University Hospital of Schleswig-Holstein, Campus Luebeck, Rheumaklinik Bad Bramstedt; Erika Gromnica-Ihle, Rheumaklinik Berlin-Buch; Hans-Hartmut Peter, Medizinische Universitaetsklinik, Abteilung Rheumatologie und Klinische Immunologie, Freiburg; Karin Manger, Medizinische Klinik III der FAU Erlangen-Nuernberg; Sebastian Schnarr, Henning Zeidler, Abteilung Rheumatologie, Medizinische Hochschule Hannover; Reinhold E. Schmidt, Abteilung Klinische Immunologie, Medizinische Hochschule Hannover.

Spanish Collaborative Group: Norberto Ortego and José L Callejas, Servicio 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. Garcia-Hernandez, Servicio de Medicina Interna, Hospital Virgen del Rocio, Seville; Mayte Camps, Servicio Medicina Interna, Hospital Carlos Haya, Malaga; Miguel Angel López-Nevot, Servicio de Inmunología, Hospital Virgen de las Nieves, Granada; Maria F. González-Escribano, Servicio de Inmunología, Hospital Virgen de las Nieves, Seville.

SLEGEN consortium: John H. Harley, Oklahoma Medical Research Foundation, Oklahoma City, Oklahoma, USA; Marta Alarcon Riquelme, Uppsala University, Sweden, and Oklahoma Medical Research Foundation; Robert Kimberly, University of Alabama, Birmingham, USA; Lindsey Criswell, UCSF, San Francisco, California, USA; Carl Langefeld, Wake Forest University, Winston-Salem, North Carolina, USA; Betty Tsao, UCLA, Los Angeles, California, USA; and Chaim Jacob, University of Southern California, Los Angeles, California, USA.

Acknowledgments

The following funding sources are acknowledged: the Alliance for Lupus Research; the NIH (grant R01 AR50812); the Swedish Research Council (grant 12673); the Torsten och Ragnar Söderbergs stiftelser; the Gustaf V 80-Year Jubilee Foundation; the National Arthritis Foundation (Arthritis Investigator Award to K. Liu); the Swedish Association Against Rheumatism; the Knut and Alice Wallenberg Foundation through the Royal Swedish Academy of Sciences (to M.E. Alarcón-Riquelme); the Spanish Ministerio de Educacion y Ciencia (grant SAF2006-00398); and the Regione Piemonte (grants 2003 and 2004).

Received for publication July 9, 2008, and accepted in revised form February 4, 2009.

Address correspondence to: Chandra Mohan, Room Y8.204, MC 8884, University of Texas Southwestern Medical Center, 5323 Harry Hines Blvd., Dallas, Texas 75390-8884, USA. Phone: (214) 648-9675; Fax: (214) 648-7995; E-mail: chandra.mohan@

utsouthwestern.edu. Or to: Edward Wakeland, University of Texas Southwestern Medical Center, 5323 Harry Hines Blvd., Dallas, Texas 75390-8884, USA. Phone: (214) 648-7332; Fax: (214) 648-7995; E-mail: ward.wakeland@utsouthwestern.edu. Or to: Marta Alarcón-Riquelme, Department of Genetics and Pathology, Unit for Medical Genetics, Uppsala University, Dag Hammarsjkölds väg 20, 75185 Uppsala, Sweden. Phone: 46-18-4714805; Fax: 46-18-4714808; E-mail: marta.alarcon@genpat.uu.se.

1. Hahn, B.H. 1998. Antibodies to DNA. *N. Engl. J. Med.* **338**:1359–1368.
2. Pisetsky, D.S. 2000. Anti-DNA and autoantibodies. *Curr. Opin. Rheumatol.* **12**:364–368.
3. Lefkowith, J.B., and Gilkeson, G.S. 1996. Nephritogenic autoantibodies in lupus: current concepts and continuing controversies. *Arthritis Rheum.* **39**:894–903.
4. Foster, M.H., and Kelley, V.R. 1999. Lupus nephritis: update on pathogenesis and disease mechanisms. *Semin. Nephrol.* **19**:173–181.
5. Budhai, L., Oh, K., and Davidson, A. 1996. An in vitro assay for detection of glomerular binding IgG autoantibodies in patients with systemic lupus erythematosus. *J. Clin. Invest.* **98**:1585–1593.
6. Liang, Z., et al. 2004. Pathogenic profiles and molecular signatures of antinuclear autoantibodies rescued from NZM2410 lupus mice. *J. Exp. Med.* **199**:381–398.
7. Li, Q.Z., et al. 2005. Identification of autoantibody clusters that best predict lupus disease activity using glomerular proteome arrays. *J. Clin. Invest.* **115**:3428–3439.
8. Pankewycz, O.G., Migliorini, P., and Madaio, M.P. 1987. Polyreactive autoantibodies are nephritogenic in murine lupus nephritis. *J. Immunol.* **139**:3287–3294.
9. Fu, Y., Du, Y., and Mohan, C. 2007. Experimental anti-GBM disease as a tool for studying spontaneous lupus nephritis. *Clin. Immunol.* **124**:109–118.
10. Xie, C., Zhou, X.J., Liu, X., and Mohan, C. 2003. Enhanced susceptibility to end-organ disease in the lupus-facilitating NZW mouse strain. *Arthritis Rheum.* **48**:1080–1092.
11. Xie, C., Sharma, R., Wang, H., Zhou, X.J., and Mohan, C. 2004. Strain distribution pattern of susceptibility to immune-mediated nephritis. *J. Immunol.* **172**:5047–5055.
12. Xie, C., et al. 2008. Strain distribution pattern of immune nephritis — a follow-up study. *Int. Immunol.* **20**:719–728.
13. Kono, D.H., and Theofilopoulos, A.N. 2000. Genetics of systemic autoimmunity in mouse models of lupus. *Int. Rev. Immunol.* **19**:367–387.
14. Morel, L., Rudofsky, U.H., Longmate, J.A., Schiffenbauer, J., and Wakeland, E.K. 1994. Polygenic control of susceptibility to murine systemic lupus erythematosus. *Immunity.* **1**:219–229.
15. Wakeland, E.K., Wandstrat, A.E., Liu, K., and Morel, L. 1999. Genetic dissection of systemic lupus erythematosus. *Curr. Opin. Immunol.* **11**:701–707.
16. Bygrave, A.E., et al. 2004. Spontaneous autoimmunity in 129 and C57BL/6 mice — implications for autoimmunity described in gene-targeted mice. *PLoS Biol.* **2**:E243.
17. Morel, L., et al. 1997. Functional dissection of systemic lupus erythematosus using congenic mouse strains. *J. Immunol.* **158**:6019–6028.
18. Mohan, C., Yu, Y., Morel, L., Yang, P., and Wakeland, E.K. 1999. Genetic dissection of Sle pathogenesis: Sle3 on murine chromosome 7 impacts T cell activation, differentiation, and cell death. *J. Immunol.* **162**:6492–6502.
19. Mohan, C., et al. 1999. Genetic dissection of lupus pathogenesis: a recipe for nephrophilic autoantibodies. *J. Clin. Invest.* **103**:1685–1695.
20. Zhu, J., et al. 2005. T cell hyperactivity in lupus as a consequence of hyperstimulatory antigen-presenting cells. *J. Clin. Invest.* **115**:1869–1878.
21. Liu, K., et al. 2007. Sle3 and Sle5 can independently couple with Sle1 to mediate severe lupus nephritis. *Genes Immun.* **8**:634–645.
22. Yousef, G.M., Chang, A., Scorilas, A., and Diamandis, E.P. 2000. Genomic organization of the human kallikrein gene family on chromosome 19q13.3-q13.4. *Biochem. Biophys. Res. Commun.* **276**:125–133.
23. Evans, B.A., Drinkwater, C.C., and Richards, R.I. 1987. Mouse glandular kallikrein genes. Structure and partial sequence analysis of the kallikrein gene locus. *J. Biol. Chem.* **262**:8027–8034.
24. Mason, A.J., Evans, B.A., Cox, D.R., Shine, J., and Richards, R.I. 1983. Structure of mouse kallikrein gene family suggests a role in specific processing of biologically active peptides. *Nature.* **303**:300–307.
25. Bhoola, K.D., Figueroa, C.D., and Worthy, K. 1992. Bioregulation of kinins: kallikreins, kininogens, and kininases. *Pharmacol. Rev.* **44**:1–80.
26. Carretero, O.A., and Scicli, A.G. 1980. The renal kallikrein-kinin system. *Am. J. Physiol.* **238**:F247–F255.
27. Chao, J., Bledsoe, G., Yin, H., and Chao, L. 2006. The tissue kallikrein-kinin system protects against cardiovascular and renal diseases and ischemic stroke independently of blood pressure reduction. *Biol. Chem.* **387**:665–675.
28. Chao, J., and Chao, L. 2005. Kallikrein-kinin in stroke, cardiovascular and renal disease. *Exp. Physiol.* **90**:291–298.
29. Chao, J., Zhang, J.J., Lin, K.F., and Chao, L. 1998. Human kallikrein gene delivery attenuates hypertension, cardiac hypertrophy, and renal injury in Dahl salt-sensitive rats. *Hum. Gene Ther.* **9**:21–31.
30. Chao, J., Zhang, J.J., Lin, K.F., and Chao, L. 1998. Adenovirus-mediated kallikrein gene delivery reverses salt-induced renal injury in Dahl salt-sensitive rats. *Kidney Int.* **54**:1250–1260.
31. Bledsoe, G., et al. 2006. Kallikrein/kinin protects against gentamicin-induced nephrotoxicity by inhibition of inflammation and apoptosis. *Nephrol. Dial. Transplant.* **21**:624–633.
32. Bledsoe, G., et al. 2006. Reversal of renal fibrosis, inflammation, and glomerular hypertrophy by kallikrein gene delivery. *Hum. Gene Ther.* **17**:545–555.
33. Zhang, J.J., Bledsoe, G., Kato, K., Chao, L., and Chao, J. 2004. Tissue kallikrein attenuates salt-induced renal fibrosis by inhibition of oxidative stress. *Kidney Int.* **66**:722–732.
34. Schanstra, J.P., et al. 2002. In vivo bradykinin B2 receptor activation reduces renal fibrosis. *J. Clin. Invest.* **110**:371–379.
35. Kakoki, M., Takahashi, N., Jennette, J.C., and Smithies, O. 2004. Diabetic nephropathy is markedly enhanced in mice lacking the bradykinin B2 receptor. *Proc. Natl. Acad. Sci. U. S. A.* **101**:13302–13305.
36. Chao, J., et al. 2007. Kinin infusion prevents renal inflammation, apoptosis, and fibrosis via inhibition of oxidative stress and mitogen-activated protein kinase activity. *Hypertension.* **49**:490–497.
37. Marceau, F., Lussier, A., Regoli, D., and Giroud, J.P. 1983. Pharmacology of kinins: their relevance to tissue injury and inflammation. *Gen. Pharmacol.* **14**:209–229.
38. Zhang, X., Scicli, G.A., Xu, X., Nasjletti, A., and Hintze, T.H. 1997. Role of endothelial kinins in control of coronary nitric oxide production. *Hypertension.* **30**:1105–1111.
39. Yang, C.M., et al. 2005. Bradykinin-induced p42/p44 MAPK phosphorylation and cell proliferation via Src, EGF receptors, and PI3-K/Akt in vascular smooth muscle cells. *J. Cell. Physiol.* **203**:538–546.
40. Lindqvist, A.K., et al. 2000. A susceptibility locus for human systemic lupus erythematosus (hSLE1) on chromosome 2q. *J. Autoimmun.* **14**:169–178.
41. Moser, K.L., et al. 1998. Genome scan of human systemic lupus erythematosus: evidence for linkage on chromosome 1q in African-American pedigrees. *Proc. Natl. Acad. Sci. U. S. A.* **95**:14869–14874.
42. Forabosco, P., et al. 2006. Meta-analysis of genome-wide linkage studies of systemic lupus erythematosus. *Genes Immun.* **7**:609–614.
43. Xing, C., et al. 2007. Localization and replication of the systemic lupus erythematosus linkage signal at 4p16: interaction with 2p11, 12q24 and 19q13 in European Americans. *Hum. Genet.* **120**:623–631.
44. Margolius, H.S., Horwitz, D., Pisano, J.J., and Keiser, H.R. 1974. Urinary kallikrein excretion in hypertensive man. Relationships to sodium intake and sodium-retaining steroids. *Circ. Res.* **35**:820–825.
45. Zinner, S.H., Margolius, H.S., Rosner, B., and Kass, E.H. 1978. Stability of blood pressure rank and urinary kallikrein concentration in childhood: an eight-year follow-up. *Circulation.* **58**:908–915.
46. Freedman, B.I., et al. 1997. Genetic linkage analysis of growth factor loci and end-stage renal disease in African Americans. *Kidney Int.* **51**:819–825.
47. Austin, H.A., 3rd, Boumpas, D.T., Vaughan, E.M., and Balow, J.E. 1995. High-risk features of lupus nephritis: importance of race and clinical and histological factors in 166 patients. *Nephrol. Dial. Transplant.* **10**:1620–1628.
48. Slim, R., et al. 2002. Loss-of-function polymorphism of the human kallikrein gene with reduced urinary kallikrein activity. *J. Am. Soc. Nephrol.* **13**:968–976.
49. Yu, H., et al. 2002. Association of the tissue kallikrein gene promoter with ESRD and hypertension. *Kidney Int.* **61**:1030–1039.
50. Kumar, K.R., et al. 2006. Regulation of B cell tolerance by the lupus susceptibility gene Ly108. *Science.* **312**:1665–1669.
51. Pisitkun, P., et al. 2006. Autoreactive B cell responses to RNA-related antigens due to TLR7 gene duplication. *Science.* **312**:1669–1672.
52. Subramanian, S., et al. 2006. A Tlr7 translocation accelerates systemic autoimmunity in murine lupus. *Proc. Natl. Acad. Sci. U. S. A.* **103**:9970–9975.
53. Shin, H.D., et al. 2007. Replication of the genetic effects of IFN regulatory factor 5 (IRF5) on systemic lupus erythematosus in a Korean population. *Arthritis Res. Ther.* **9**:R32.
54. Kozyrev, S.V., et al. 2007. Structural insertion/deletion variation in IRF5 is associated with a risk haplotype and defines the precise IRF5 isoforms expressed in systemic lupus erythematosus. *Arthritis Rheum.* **56**:1234–1241.
55. Reddy, M.V., et al. 2007. Genetic association of IRF5 with SLE in Mexicans: higher frequency of the risk haplotype and its homozygozity than Europeans. *Hum. Genet.* **121**:721–727.
56. Tan, E.M., et al. 1982. The 1982 revised criteria for the classification of systemic lupus erythematosus. *Arthritis Rheum.* **25**:1271–1277.
57. Harley, J.B., et al. 2008. Genome-wide association scan in women with systemic lupus erythematosus identifies susceptibility variants in ITGAM, PXK, KIAA1542 and other loci. *Nat. Genet.* **40**:204–210.
58. Hom, G., et al. 2008. Association of systemic lupus erythematosus with C8orf13-BLK and ITGAM-ITGAX. *N. Engl. J. Med.* **358**:900–909.
59. 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.
60. Purcell, S., et al. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**:559–575.

Supplementary table S1 Association analysis of the human KLK region polymorphism with SLE and lupus nephritis.

Chr.	SNP	BP Position	Gene	Allele	Frequency Cases	Frequency Controls	SLE vs. controls[1]				SLE Nephritis vs. controls[1, 2]			
							OR	L95	U95	P-value	OR	L95	U95	P-value
19	rs5519	56014125	*KLK1*	C	0.27	0.26	1.04	0.93	1.16	0.51	1.01	0.84	1.22	0.92
19	rs266850	56028010		G	0.17	0.19	0.85	0.75	0.97	**0.01312**	0.79	0.63	0.99	**0.0388**
19	rs266851	56028152		T	0.20	0.18	1.11	0.99	1.26	0.08	1.02	0.83	1.26	0.83
19	rs266858	56031775		C	0.22	0.26	0.83	0.74	0.93	**0.0009081**	0.88	0.73	1.07	0.19
19	rs2569729	56037634		A	0.19	0.18	1.10	0.97	1.24	0.14	1.08	0.88	1.34	0.45
19	rs266849	56040903		G	0.19	0.20	0.92	0.82	1.04	0.17	0.96	0.79	1.18	0.73
19	rs266870	56043747		C	0.50	0.49	1.01	0.92	1.12	0.77	1.01	0.85	1.19	0.93
19	rs266868	56044750		A	0.30	0.30	0.97	0.87	1.08	0.57	1.01	0.84	1.21	0.93
19	rs3760722	56049629	*KLK3*	T	0.10	0.10	1.06	0.90	1.24	0.49	1.16	0.89	1.51	0.27
19	rs1058205	56055211	*KLK3*	C	0.17	0.19	0.92	0.81	1.04	0.16	0.93	0.75	1.15	0.52
19	rs2735839	56056436		A	0.14	0.15	0.95	0.83	1.08	0.41	0.97	0.77	1.22	0.78
19	rs1506684	56063232		A	0.46	0.46	1.03	0.94	1.13	0.52	1.07	0.91	1.26	0.39
19	rs198972	56071706	*KLK2*	T	0.32	0.31	1.03	0.93	1.14	0.63	1.10	0.93	1.31	0.27
19	rs198979	56075993	*KLK2*	C	0.10	0.10	1.11	0.95	1.30	0.20	0.98	0.73	1.31	0.89
19	rs2664144	56083469		C	0.10	0.10	1.07	0.92	1.25	0.37	0.98	0.75	1.29	0.89
19	rs1354774	56084931		G	0.37	0.37	1.00	0.91	1.11	0.99	1.06	0.89	1.25	0.53
19	rs2235091	56102284	*KLK4*	G	0.37	0.37	1.01	0.92	1.12	0.78	1.02	0.86	1.21	0.82
19	rs2242670	56104128	*KLK4*	A	0.46	0.44	1.11	1.00	1.22	**0.04103**	1.17	0.99	1.38	0.06
19	rs7248321	56110318		C	0.05	0.05	0.93	0.75	1.17	0.54	1.04	0.72	1.50	0.85
19	rs1701930	56111359		A	0.41	0.42	0.96	0.87	1.06	0.43	1.09	0.92	1.28	0.32
19	rs268917	56116667		A	0.44	0.45	0.95	0.87	1.05	0.34	1.06	0.89	1.24	0.53
19	rs7245858	56122409	*LOC390956*	G	0.03	0.03	0.80	0.60	1.06	0.12	0.92	0.58	1.47	0.74
19	rs7255259	56131022	*LOC390956*	G	0.08	0.08	1.09	0.92	1.30	0.33	1.44	1.10	1.88	**0.008607**
19	rs8104441	56133620		C	0.18	0.18	1.05	0.93	1.19	0.42	0.90	0.72	1.12	0.35
19	rs268914	56134512		G	0.32	0.33	0.96	0.87	1.06	0.44	1.25	1.06	1.49	**0.008991**
19	rs268909	56141477	*KLK5*	G	0.16	0.17	0.97	0.85	1.10	0.66	1.24	1.01	1.53	**0.03866**
19	rs268905	56142742	*KLK5*	T	0.28	0.28	1.00	0.90	1.11	0.94	0.91	0.75	1.09	0.30
19	rs4592765	56151844	*KLK6*	C	0.40	0.39	1.06	0.96	1.17	0.23	1.03	0.87	1.21	0.75
19	rs1654537	56158329	*KLK6*	G	0.34	0.33	1.01	0.91	1.12	0.85	0.98	0.82	1.17	0.82
19	rs1275221	56178880	*KLK7*	T	0.36	0.35	1.02	0.92	1.12	0.76	0.85	0.71	1.02	0.07
19	rs268899	56183902		C	0.37	0.36	1.04	0.94	1.15	0.43	1.01	0.86	1.20	0.87
19	rs1654543	56190802	*KLK8*	C	0.45	0.45	1.01	0.92	1.11	0.88	0.91	0.77	1.07	0.26
19	rs2569456	56197538	*KLK9*	A	0.34	0.33	1.07	0.96	1.18	0.21	1.08	0.91	1.28	0.37
19	rs2075695	56210110	*KLK10*	A	0.44	0.44	0.98	0.89	1.08	0.63	1.05	0.89	1.24	0.55
19	rs2691208	56217901	*KLK11*	A	0.16	0.17	0.90	0.79	1.02	0.10	0.86	0.68	1.08	0.19
19	rs1048328	56219177	*KLK11*	A	0.10	0.09	1.10	0.93	1.30	0.27	1.24	0.94	1.64	0.13
19	rs3745539	56219854	*KLK11*	T	0.07	0.08	0.86	0.72	1.04	0.12	0.57	0.39	0.84	**0.004004**
19	rs2250066	56220932	*KLK11*	T	0.15	0.13	1.13	0.99	1.30	0.08	1.00	0.79	1.27	1.00
19	rs3745540	56226943	*KLK12*	A	0.43	0.42	1.03	0.93	1.13	0.59	1.03	0.87	1.21	0.77
19	rs3865443	56229495	*KLK12*	A	0.23	0.25	0.89	0.80	0.99	**0.04048**	1.07	0.89	1.29	0.47
19	rs3760739	56230374	*KLK12*	T	0.35	0.33	1.06	0.96	1.17	0.26	1.00	0.84	1.19	0.99
19	rs2691229	56238455		C	0.49	0.49	1.03	0.94	1.13	0.54	0.95	0.80	1.12	0.51
19	rs2569469	56241011		A	0.31	0.30	1.08	0.98	1.20	0.12	1.08	0.90	1.28	0.42
19	rs2691239	56243625		A	0.36	0.38	0.91	0.82	1.00	**0.04669**	0.95	0.80	1.12	0.54
19	rs1880413	56252280	*KLK13*	C	0.33	0.34	0.94	0.85	1.04	0.21	0.95	0.80	1.13	0.56
19	rs8111658	56252836	*KLK13*	C	0.36	0.37	0.99	0.90	1.10	0.90	0.93	0.78	1.10	0.39
19	rs1001538	56263968		C	0.28	0.29	0.91	0.82	1.01	0.07	0.91	0.76	1.09	0.32
19	rs2569481	56267237		A	0.28	0.29	0.94	0.84	1.04	0.23	0.95	0.79	1.14	0.58
19	rs2569484	56268087		T	0.28	0.29	0.94	0.85	1.05	0.27	0.96	0.80	1.15	0.68
19	rs2569488	56271459	*KLK14*	C	0.35	0.37	0.92	0.84	1.02	0.11	0.95	0.81	1.13	0.59
19	rs867191	56275429	*KLK14*	C	0.06	0.06	1.03	0.85	1.25	0.74	0.98	0.70	1.38	0.91
19	rs9304708	56275749	*KLK14*	T	0.34	0.36	0.93	0.85	1.03	0.18	0.96	0.81	1.14	0.67
19	rs17658926	56276223	*KLK14*	A	0.20	0.19	0.99	0.88	1.11	0.85	1.05	0.85	1.28	0.67
19	rs10500304	56276411	*KLK14*	G	0.39	0.41	0.93	0.84	1.02	0.11	0.95	0.81	1.13	0.58
19	rs2569491	56276729	*KLK14*	A	0.33	0.34	0.94	0.85	1.04	0.22	0.99	0.84	1.18	0.92
19	rs2072689	56277428	*KLK14*	T	0.47	0.48	0.96	0.88	1.06	0.44	1.03	0.88	1.21	0.73

[1] The association analysis was performed by logistic regression and adjusted by gender introducing it as a covariable in the regression model. [2] Only the patients with nephritis were considered as cases. OR: odds ratio, L95: lower limit of the 95% confidence interval, U95: upper limit of the 95% confidence interval.

Supplementary table S2 Case-only meta-analysis of the human KLK region polymorphism comparing SLE cases with nephritis to SLE cases without nephritis.

CHR	SNP	Base pair position	Gene	A1	Frequency in cases with nephritis	Frequency in cases without nephritis	P_{CMH}	OR	L95	U95	P_{HOMOG}
19	rs5519	56014125	*KLK1*	C	0.26	0.27	0.82	0.98	0.79	1.21	0.20
19	rs266850	56028010		G	0.16	0.17	0.28	0.87	0.67	1.13	0.45
19	rs266851	56028152		T	0.19	0.20	0.38	0.89	0.70	1.15	0.83
19	rs266858	56031775		C	0.24	0.22	0.40	1.10	0.88	1.39	0.56
19	rs2569729	56037634		A	0.19	0.19	0.93	0.99	0.77	1.27	0.70
19	rs266849	56040903		G	0.20	0.20	0.66	1.06	0.83	1.34	0.60
19	rs266870	56043747		C	0.49	0.50	0.56	0.94	0.78	1.15	0.84
19	rs266868	56044750		A	0.31	0.29	0.56	1.06	0.86	1.31	0.78
19	rs3760722	56049629	*KLK3*	T	0.11	0.10	0.29	1.18	0.87	1.60	0.28
19	rs1058205	56055211	*KLK3*	C	0.18	0.18	0.91	1.02	0.79	1.31	0.44
19	rs2735839	56056436		A	0.15	0.14	0.74	1.05	0.80	1.37	0.26
19	rs1506684	56063232		A	0.47	0.46	0.25	1.12	0.92	1.37	0.99
19	rs198972	56071706	*KLK2*	T	0.33	0.32	0.36	1.10	0.90	1.35	0.20
19	rs198979	56075993	*KLK2*	C	0.09	0.11	0.28	0.83	0.59	1.16	0.33
19	rs2664144	56083469		C	0.10	0.11	0.38	0.87	0.63	1.20	0.21
19	rs1354774	56084931		G	0.38	0.36	0.38	1.09	0.90	1.34	0.40
19	rs2235091	56102284	*KLK4*	G	0.38	0.37	0.97	1.00	0.82	1.23	0.68
19	rs2242670	56104128	*KLK4*	A	0.48	0.45	0.23	1.13	0.93	1.37	0.49
19	rs7248321	56110318		C	0.05	0.04	0.25	1.30	0.83	2.04	0.30
19	rs1701930	56111359		A	0.44	0.40	**0.03345**	1.24	1.02	1.50	0.89
19	rs268917	56116667		A	0.47	0.42	0.07	1.20	0.99	1.45	0.73
19	rs7245858	56122409	*LOC390956*	G	0.03	0.02	0.25	1.40	0.79	2.50	0.41
19	rs7255259	56131022	*LOC390956*	G	0.11	0.08	**0.001313**	1.67	1.22	2.30	0.30
19	rs8104441	56133620		C	0.16	0.19	0.12	0.81	0.63	1.05	0.22
19	rs268914	56134512		G	0.39	0.30	**1.36E-05**	1.56	1.28	1.91	0.50
19	rs268909	56141477	*KLK5*	G	0.20	0.15	**0.0007579**	1.53	1.19	1.96	0.55
19	rs268905	56142742	*KLK5*	T	0.26	0.29	0.14	0.85	0.68	1.06	0.72
19	rs4592765	56151844	*KLK6*	C	0.39	0.40	0.84	0.98	0.80	1.19	0.58
19	rs1654537	56158329	*KLK6*	G	0.33	0.34	0.47	0.93	0.75	1.14	0.94
19	rs1275221	56178880	*KLK7*	T	0.32	0.37	**0.02725**	0.79	0.64	0.97	0.71
19	rs268899	56183902		C	0.37	0.37	0.31	0.90	0.74	1.10	0.99
19	rs1654543	56190802	*KLK8*	C	0.42	0.46	0.05	0.82	0.68	1.00	0.99
19	rs2569456	56197538	*KLK9*	A	0.35	0.34	0.55	0.94	0.76	1.16	0.94
19	rs2075695	56210110	*KLK10*	A	0.46	0.44	0.28	1.11	0.92	1.35	1.00
19	rs2691208	56217901	*KLK11*	A	0.15	0.16	0.19	0.83	0.63	1.09	0.68
19	rs1048328	56219177	*KLK11*	A	0.11	0.09	0.10	1.31	0.95	1.81	0.34
19	rs3745539	56219854	*KLK11*	T	0.05	0.08	**0.01117**	0.57	0.37	0.88	0.45
19	rs2250066	56220932	*KLK11*	T	0.13	0.16	0.23	0.84	0.63	1.12	0.12
19	rs3745540	56226943	*KLK12*	A	0.43	0.43	0.49	0.93	0.77	1.14	0.14
19	rs3865443	56229495	*KLK12*	A	0.27	0.22	**0.008919**	1.34	1.08	1.67	0.22
19	rs3760739	56230374	*KLK12*	T	0.33	0.35	0.35	0.91	0.74	1.11	0.98
19	rs2691229	56238455		C	0.47	0.50	0.40	0.92	0.76	1.12	0.60
19	rs2569469	56241011		A	0.31	0.30	0.74	1.04	0.84	1.28	0.79
19	rs2691239	56243625		A	0.37	0.36	0.60	1.06	0.86	1.29	0.48
19	rs1880413	56252280	*KLK13*	C	0.33	0.34	0.76	1.03	0.84	1.27	0.91

19	rs8111658	56252836	*KLK13*	C	0.35	0.36	0.43	0.92	0.75	1.13	0.34
19	rs1001538	56263968		C	0.27	0.28	0.67	0.95	0.77	1.19	0.25
19	rs2569481	56267237		A	0.28	0.27	0.73	1.04	0.84	1.29	0.25
19	rs2569484	56268087		T	0.28	0.27	0.65	1.05	0.85	1.31	0.23
19	rs2569488	56271459	*KLK14*	C	0.36	0.35	0.72	1.04	0.85	1.27	0.82
19	rs867191	56275429	*KLK14*	C	0.06	0.06	0.94	0.99	0.66	1.47	0.39
19	rs9304708	56275749	*KLK14*	T	0.35	0.34	0.94	1.01	0.82	1.24	0.77
19	rs17658926	56276223	*KLK14*	A	0.20	0.20	0.98	1.00	0.78	1.27	0.35
19	rs10500304	56276411	*KLK14*	G	0.40	0.39	0.95	0.99	0.82	1.21	0.77
19	rs2569491	56276729	*KLK14*	A	0.34	0.33	0.67	1.05	0.85	1.28	0.63
19	rs2072689	56277428	*KLK14*	T	0.48	0.47	0.23	1.13	0.93	1.37	0.75

P_{CMH}: P-value from the Cochran-Mantel-Haenzel meta-analysis performed within cases, which takes into account the variability across the different strata, in this, cases the patient's clinical center. P_{HOMOG}: P value for the homogeneity test between strata. The non-significant results indicate homogeneity of the association. OR: odds ratio, L95: lower limit of the 95% confidence interval, U95: upper limit of the 95% confidence interval.

Supplementary table S3 Haplotypic association analyses.

Haplotype block	Haplotype	Freq controls	Freq SLE	Freq SLE nephritis	Frequency SLE no nephritis	SLE vs. controls	SLE nephritis vs. controls	SLE nephritis vs. SLE no nephritis
						P-value	P-value	P-value
Block 1	OMNIBUS	--	--	--	--	**0.0008359**	0.2641	0.318
	GCC	***0.18***	***0.16***	***0.15***	0.16	**0.009036***	**0.06646**	0.6231
	ACC	0.08	0.06	0.08	0.06	**0.01237**	0.7238	0.1637
	ATT	0.18	0.20	0.18	0.20	**0.02412**	0.9759	0.2034
	ACT	0.56	0.58	0.60	0.58	0.115	0.1131	0.489
Block 2	OMNIBUS	--	--	--	--	0.2897	0.8511	0.8611
	GACA	0.28	0.27	0.28	0.27	0.6126	0.794	0.4326
	GACG	0.23	0.24	0.23	0.25	0.4375	0.7119	0.3049
	GGTG	0.19	0.19	0.19	0.19	0.3454	0.869	0.8013
	AATG	0.18	0.20	0.20	0.19	0.09253	0.3676	0.992
	GATG	0.11	0.11	0.10	0.11	0.2257	0.4436	0.967
Block 3	OMNIBUS	--	--	--	--	0.5282	0.4745	0.6379
	CCA	0.16	0.15	0.15	0.15	0.357	0.6816	0.9143
	TTG	0.10	0.11	0.12	0.10	0.4353	0.2312	0.3592
	CTG	0.74	0.75	0.73	0.75	0.8248	0.6258	0.4618
Block 4	OMNIBUS	--	--	--	--	0.4984	0.5133	0.3443
	TCCG	0.09	0.09	0.08	0.10	0.5916	0.5884	0.2564
	TTTG	0.20	0.20	0.22	0.19	0.9226	0.1862	0.1179
	CTTG	0.09	0.08	0.08	0.08	0.1426	0.4167	0.8718
	CTTA	0.62	0.63	0.61	0.63	0.5391	0.7646	0.4857
Block 5	OMNIBUS	--	--	--	--	0.2724	0.6005	**0.04465**
	TAAA	0.40	0.38	0.42	0.36	0.2038	0.3162	**0.01417**
	TAGA	0.07	0.08	0.07	0.08	0.2103	0.9174	0.3207
	TGGA	0.53	0.54	0.51	0.56	0.5518	0.3521	0.06413
Block 6	OMNIBUS	--	--	--	--	0.5431	**0.03104**	**0.001309**
	GTGG	0.08	0.08	0.11	0.08	0.5427	**0.00764**	**0.006601**
	ATGG	0.06	0.06	0.06	0.05	0.2113	0.9355	0.2056
	ATGA	0.20	0.19	0.22	0.18	0.4235	0.171	**0.02579**
	ACAA	0.18	0.18	0.16	0.19	0.3388	0.3489	0.1123
	ATAA	0.48	0.48	0.44	0.50	0.8665	0.05769	**0.008754**
Block 7	OMNIBUS	--	--	--	--	0.2005	**0.02015**	0.05108
	CCA	0.32	0.33	0.34	0.32	0.1611	0.2937	0.6253
	CCC	0.05	0.05	0.04	0.05	0.2258	0.05225	0.08427
	ACC	0.09	0.08	0.06	0.09	0.1487	**0.02109**	**0.04002**
	ATC	0.54	0.54	0.57	0.54	0.9702	0.2381	0.1626
Block 8	OMNIBUS	--	--	--	--	0.1044	**0.002529**	**0.01334**
	AGGT	***0.07***	***0.06***	***0.04***	0.07	**0.02648**	**0.001114***	**0.004292**
	GGAC	0.08	0.09	0.10	0.09	0.3057	0.1566	0.2941
	GAGC	0.17	0.16	0.15	0.16	0.2169	0.3436	0.5861
	AGGC	0.37	0.37	0.41	0.36	0.5268	**0.01757**	**0.0282**
	GGGC	0.31	0.32	0.29	0.32	0.3459	0.4491	0.3068
Block 9	OMNIBUS	--	--	--	--	0.4908	0.9606	0.3975
	GCTC	0.36	0.37	0.35	0.37	0.4104	0.8801	0.5035
	ACGC	0.13	0.13	0.13	0.13	0.9692	0.7942	0.7553
	GAGT	0.23	0.21	0.24	0.20	0.1277	0.6014	0.08535
	ACGT	0.29	0.29	0.28	0.30	0.615	0.8944	0.5415
Block 10	OMNIBUS	--	--	--	--	0.5014	0.3786	0.7334
	TC	0.37	0.36	0.35	0.36	0.8415	0.3604	0.5546
	CT	0.34	0.33	0.33	0.34	0.388	0.7012	0.8981
	TT	0.29	0.30	0.32	0.30	0.2685	0.1723	0.4562
Block 11	AT	0.29	0.28	0.28	0.27	0.2701	0.6023	0.7
	GG	0.71	0.72	0.72	0.73	0.2701	0.6023	0.7
Block 12	OMNIBUS	--	--	--	--	0.3867	0.7956	0.9749
	CTTAGA	0.21	0.21	0.22	0.21	0.897	0.6123	0.6668
	CTTGGA	0.15	0.14	0.14	0.14	0.0956	0.3428	0.9346
	TCGGTG	0.07	0.07	0.07	0.07	0.7415	0.9105	0.9249
	TTGGTG	0.57	0.58	0.58	0.58	0.2563	0.746	0.7134

The linkage disequilibrium structure of the region showed 12 haplotype blocks (see figure 6A). These blocks were tested for haplotype association by both omnibus and haplotype-specific association statistics (T test) as implemented in PLINK. The case/control omnibus test is an H-1 degree of freedom test, if there are H haplotypes. The two most significant associations (*) correspond to haplotypes with a lower frequency in patients, even less in the subgroup of patients with nephritis, reflecting a protective effect conferred by the KLK polymorphisms

Human Molecular Genetics, 2009, Vol. 18, No. 19 ***3739–3748***
doi:10.1093/hmg/ddp301
Advance Access published on July 7, 2009

Identification of a new putative functional *IL18* gene variant through an association study in systemic lupus erythematosus

Elena Sánchez[1,2,*,†], Rogelio J. Palomino-Morales[1,†], Norberto Ortego-Centeno[3], Juan Jiménez-Alonso[4], Miguel A. González-Gay[2], Miguel A. López-Nevot[5], Julio Sánchez-Román[6], Enrique de Ramón[7], M. Francisca González-Escribano[8], Bernardo A. Pons-Estel[9,‡], Sandra D'Alfonso[10], Gian Domenico Sebastiani[11], The Italian collaborative group, Marta E. Alarcón-Riquelme[12] and Javier Martín[1]

[1]Instituto de Parasitología y Biomedicina López-Neyra, CSIC, Granada 18100, Spain, [2]Hospital Xeral-Calde, Servicio de Reumatología, Lugo 27004, Spain, [3]Hospital Clínico San Cecilio, Unidad de enfermedades sistémicas autoinmunes, Granada 18012, Spain, [4]Hospital Virgen de las Nieves, Servicio de Medicina Interna, Granada 18014, Spain, [5]Hospital Virgen de las Nieves, Servicio de Inmunología, Granada 18014, Spain, [6]Hospital Virgen del Rocío, Servicio de Medicina Interna, Sevilla 41013, Spain, [7]Hospital Carlos Haya, Servicio de Medicina Interna, Málaga 29010, Spain, [8]Hospital Virgen del Rocio, Servicio de Inmunologia, Sevilla 41013, Spain, [9]Sanatorio Parque, Rosario 860, Argentina, [10]Department of Medical Sciences and IRCAD, University of Eastern Piedmont, Novara 28100, Italy, [11]U.O.C. di Reumatologia Ospedale San Camillo, Roma 00151, Italy and [12]Department of Genetics and Pathology, Rudbeck Laboratory, Uppsala University, Uppsala 75185, Sweden

Received April 14, 2009; Revised and Accepted June 26, 2009

Interleukin-18 (IL-18) is a proinflammatory cytokine that plays an important role in chronic inflammation and autoimmune disorders. In this study, we aimed to determine the potential role of the *IL18* gene in SLE. To define the genetic association of the *IL18* and SLE, we have genotyped nine SNPs in an independent set of Spanish cases and controls. The *IL18* polymorphisms were genotyped by PCR, using a predeveloped TaqMan allele discrimination assay. Two SNPs were still significant after fine mapping of the *IL18* gene. The SNP (rs360719) surviving correction for multiple tests was genotyped in two replication cohorts from Italy and Argentina. After the analysis, a significance with rs360719 C-allele remained across the sets and after the meta-analysis (Pooled OR = 1.37, 95% CI 1.21–1.54, *combined P* = 3.8E-07, *Pc* = 1.16E-06). Quantitative real-time PCR was performed to assess *IL18* mRNA expression in PBMC from subjects with different *IL18* rs360719 genotypes. We tested the effect of the *IL18* rs360719 polymorphism on the transcription of *IL18* by electrophoretic mobility shift assay and western blot. We found a significant increase in the relative expression of *IL18* mRNA in individuals carrying the rs360719 C-risk allele; in addition we show that the polymorphism creates a binding site for the transcriptional factor OCT-1. These findings suggest that the novel *IL18* rs360719 variant may play an important role in determining the susceptibility to SLE and it could be a key factor in the expression of the *IL18* gene.

*To whom correspondence should be addressed at: Instituto de Parasitología y Biomedicina 'López-Neyra', CSIC, Parque Tecnológico de Ciencias de la Salud, Avenida del Conocimiento s/n 18100-Armilla, Granada, Spain. Tel: +34 958181621; Fax: +34 958181632; Email: elena@ipb.csic.es
†Both authors have contributed equally to this work.
‡B.A.P.-E. is coordinator of the Argentine collaborative group. The list of participants is given in Appendix section along with Italian collaborative group.

INTRODUCTION

Systemic lupus erythematosus (SLE) is a prototypic autoimmune disease with a complex pathogenesis involving multiple genetic and environmental factors. The disease is characterized by enhanced autoantibody production, abnormalities of immune-inflammatory system function and inflammatory manifestation in several organs. Although the pathogenesis of SLE is unknown, a strong genetic component has been supported by studies on twins and families (1). Like most autoimmune diseases, the HLA genes make an important contribution, although other somewhat weaker but well established associations have been found with non-HLA genes (2).

Interleukin-18 (IL-18) is an important proinflammatory cytokine, member of the IL-1 cytokine family, which has been shown to exert innate and acquired immune responses (3–5). IL-18 is expressed by a wide range of immune cells (6) and has been found to have multiple biological functions. The IL-18 has recently been shown to be a pleiotropic cytokine that can mediate both Th1 and Th2 driven immune responses (7,8). In combination with IL-12, IL-18 induces IFN-γ production in Th1 cells, B cells and natural killer cells, promoting Th1-type immune responses (9,10), but it can also stimulate Th2 immune responses in the absence of IL-12 (11,12). Abnormalities in the production of Th1 and Th2 cytokines have been shown in SLE patients (13). In addition, IL-18 can accelerate spontaneous autoimmune disease in MRL/*lpr* mice, characterized by glomerulonephritis, vasculitis and symmetrical malar rash, suggesting that it is an important mediator of lupus-like disease (14). Likewise, elevated serum levels of IL-18 have been described in SLE patients compared with controls (15–17) and the elevation of IL-18 was positively correlated to SLE disease activity index (15). Because of its multiple functions in inflammation and immunological responses, a potential pathological role in the development of chronic inflammation has been suggested for IL-18, including autoimmune diseases such as SLE.

The *IL18* gene is located on chromosome 11q22.2–22.3 (18), a close linkage region with SLE in European populations (19,20). In addition, several polymorphisms within the *IL18* promoter gene have been associated with different inflammatory and autoimmune diseases (21–28). These findings suggest that *IL18* is a candidate proinflammatory cytokine gene involved in the susceptibility to autoimmune diseases, such as SLE. The aim of this study was to determine the relationship between the *IL18* gene and susceptibility to SLE.

RESULTS

Selection of markers for analysis

To investigate whether the *IL18* gene on Chromosome 11q22.2–22.3 associates with SLE, we selected a total of nine polymorphisms spanning this candidate gene for genotyping. Tagging SNPs as well as random SNPs with minor allele frequencies above 0.01 in Caucasian populations were included to ascertain maximum haplotype information for the gene and to ensure coverage of intergenic regions which may harbour regulatory polymorphisms.

IL18 is associated with susceptibility to SLE

In all populations, genotype frequencies were in Hardy–Weinberg equilibrium in patients and controls for all the polymorphisms analysed. The success rate of genotyping (that is the percentage of samples that could be analysed) was >95% for all polymorphisms in both SLE cases and controls.

Single SNP analysis performed in the first set of SLE patients ($n = 752$) and controls ($n = 595$) from Spain revealed significant association between rs360719 and rs1946518 SNPs in the *IL18* gene and SLE (allele-$P = 7.8E\text{-}07$ and allele-$P = 0.03$, respectively) (Table 1). However, only the rs360719 surviving after correction for multiple test ($P_{\text{Bonferroni correction}} = 0.01$, $P_{\text{FDR correction}} = 0.01$), and this variant also remains significantly associated after permutation analysis with 10 000 permutations ($P_{10000} < 0.00001$).

To delineate the haplotypic architecture of the whole gene, we estimated the underlying haplotype block structure of 19 SNPs with minor allele frequency >5% in the Caucasian population. We used imputation to fill in missing genotypes and to test ungenotyped variants within the gene of our existing Spanish genotyping. Figure 1 shows a plot of $-\log_{10}$ (*P*-value) for all the SNPs used in the imputed case–control association study against chromosomal position.

To seek replication of the Spanish association in *IL18*, we genotyped the rs360719 variant in two independent populations from Italy and Argentina (Table 2). Risk allele frequencies were similar in the three populations. Interestingly, in Argentinean and Italian populations, we found that the frequencies of the rs360719 CC genotype were slightly increased in SLE patients compared with healthy controls ($P = 0.02$, OR = 1.88 95% CI 1.07–3.31 in Italians and $P = 0.09$, OR = 1.70 95% CI 0.89–3.23 in Argentineans), although these differences did not reach statistical significance in the allele frequencies ($P = 0.1$, OR = 1.20 95% CI 0.95–1.51 in Italians and $P = 0.2$, OR = 1.18 95% CI 0.90–1.53 in Argentineans).

Using homogeneity and combinability test according to the Breslow–Day method, we carried out a meta-analysis comprising 2579 individuals to maximize the number of samples for the association analysis (Breslow–Day *P*-value=0.2). We then used the Mantel–Haenszel test to calculate pooled OR for *IL18* rs360719 polymorphism (Table 2), and corroborated the genetic association with SLE for this polymorphism (pooled OR = 1.37, 95% CI 1.21–1.54, $P = 3.8E\text{-}07$, $Pc = 1.16E\text{-}06$). In addition, the clinical and demographic features of patients with SLE were analysed for possible association with the different alleles or genotypes of all the *IL18* polymorphisms. No statistically significant differences were observed in the distribution of these variants (data not shown).

In view of these interesting results, we decided to perform functional experiments in order to confirm the potential role of the *IL18* rs360719 polymorphism in the pathogenesis of SLE.

Expression analysis

A relative quantification of mRNA was performed in total RNA from 23 healthy individuals carrying different genotypes for *IL18* rs360719 polymorphism (Fig. 2). A statistically

Table 1. Fine mapping of the *IL18* gene in the Spanish patients with SLE and matched controls

Chr.	Position	SNP	Cases (aa/Aa/AA)	Controls (aa/Aa/AA)	Frequency cases	Frequency controls	Genotypic test *P*-value	Allele test *P*-value	OR[a]	L95	U95	Multiple test correction Bonferroni	FDR
11	111541359	rs360719	97/314/325	34/235/324	0.34	0.25	7.9E−06	7.8E−07	1.53	1.29	1.81	0.01	0.01
11	111540668	rs1946518	146/375/215	91/304/197	0.45	0.41	0.03	0.03	1.19	1.02	1.38	0.2	0.08
11	111540198	rs187238	55/317/365	34/256/302	0.29	0.27	0.2	0.4	1.08	0.91	1.28	1	1
11	111536290	rs795467	56/301/380	41/229/321	0.28	0.26	0.6	0.3	1.09	0.91	1.29	1	1
11	111534931	rs4937113	146/346/244	97/283/211	0.43	0.40	0.1	0.1	1.13	0.96	1.32	1	1
11	111531913	rs360722	10/175/551	14/128/450	0.13	0.13	0.2	0.9	1.01	0.80	1.26	1	1
11	111527192	rs5744257	0/45/705	1/33/558	0.03	0.03	0.2	0.9	1.02	0.65	1.58	1	1
11	111522081	rs5744276	40/239/456	29/214/349	0.22	0.23	0.6	0.4	0.93	0.77	1.11	1	1
11	111521724	rs5744280	103/306/301	75/256/262	0.36	0.34	0.3	0.3	1.08	0.91	1.27	1	1

[a]Allele-OR of the minor allele.

significant deviation was observed when we compared the relative expression of the *IL18* in samples from healthy subjects stratified according to their *IL18* rs360719 genotypes, showing an increased expression in individual carriers of the C allele in each of the reference genes (CC + CT: $n = 11$, versus TT: $n = 12$; $P = 0.012$ for *β-actin*, $P = 0.03$ for *GADPH* and $P = 0.016$ for *ABL*).

Electrophoretic mobility shift assay

Our *in silico* analysis of the wild-type and variant sequences (http://www.cbrc.jp/research/db/TFSEARCH.html) indicates that transcription factor OCT-1 binds to the protective allele (T) but not to the risk allele (C) at position -1297 (rs360719). To investigate the effect of this polymorphism on transcription factor binding, we performed an electrophoretic mobility shift assay (EMSA) assay. We observed a higher level of binding of protein to the *IL18* -1297T allele than to the -1297C allele (Fig. 3), which support the sequence-based prediction of OCT-1 binding to the rs360719 alleles of *IL18* gene. Three protein complexes formed on the OCT-1 probe (complexes 1–3) were detected consistently (Fig. 3). To confirm the specificity of OCT-1 T or C-allele oligonucleotides, a competition assay was carried out using excess amounts of cold oligonucleotides as competitors (Fig. 3). This indicates that all three DNA–protein complexes represent a specific interaction between Jurkat nuclear protein and the OCT-1 sequence. To further confirm the results, we performed a supershift assay with anti-human OCT-1 antibody. As show in Figure 3, when anti-human OCT-1 antibody was added into the reaction mixture, the super-shifted band due to the antibody binding to complex appeared.

Immunoassays

To test the interaction between IL-18 and OCT-1, we analysed the expression patterns of both proteins in unstimulated and stimulated Jurkat cells. Western blot analysis showed an increased expression of the IL-18 after activation of Jurkat cells with a proinflammatory stimulus such as phorbol myrestate acetate (PMA) and ionomycin (Supplementary Material, Fig. S1). Conversely, we observed that the OCT-1 expression is decreased after stimulation. Similar results were obtained after analysis by flow cytometry (Supplementary Material, Fig. S2). We found an increased expression of IL-18 in Jurkat cells after activation with PMA-ionomycin (12.73 ± 1.58) with respect to the IL-18 expression in unstimulated Jurkat cells (7.03 ± 0.83) ($P = 0.01$). In addition, a decreased expression of OCT-1 was observed in stimulated Jurkat cell (5.42 ± 0.005) with respect to unstimulated cells (15.45 ± 1.45) ($P = 0.02$).

DISCUSSION

Although the role of IL-18 in inflammatory and autoimmune processes has been well established (14), the available genetic data are largely contradictory, reflecting the small samples used, the different diseases analysed and the ethnic groups investigated (29). The aim of this study was to investigate the role of *IL18* gene variations in SLE, for this purpose we performed the most powerful genetic study to date. Regarding the *IL18* rs187238 and rs1946518 polymorphisms, which have been found associated with susceptibility to SLE and different phenotypes of the disease in Asiatic populations (30–32), we could not confirm these previously reported associations. These contradictory data could be due to genetic or environmental ethnical heterogeneity, which is clearly present, since allele and genotype frequencies are significantly different between Chinese and Spanish populations. Similarly, these ethnic differences in allele frequency of autoimmune disease-associated polymorphisms have been found in other susceptibility genes to SLE, such as *PDCD1* and *PTPN22* (33).

We have identified a novel variation which seems to affect the expression of the *IL18* gene (*IL18* rs360719), which could be a key genetic variant in the role of the *IL18* gene in autoimmune diseases. Our results have shown that, in the combined analysis of three Caucasians populations, including 1356 SLE patients and 1223 healthy controls, the *IL18* rs360719 C-allele was associated with an increased risk of SLE (OR 1.37). Interestingly, we observed an increased expression of IL-18 levels in correlation with the *IL18* rs360719 polymorphism. Although these data support an evidence of association between *IL18* rs360719 polymorphism and SLE, the full inclusion of *IL18* gene as a candidate gene to SLE will require further independent studies in different

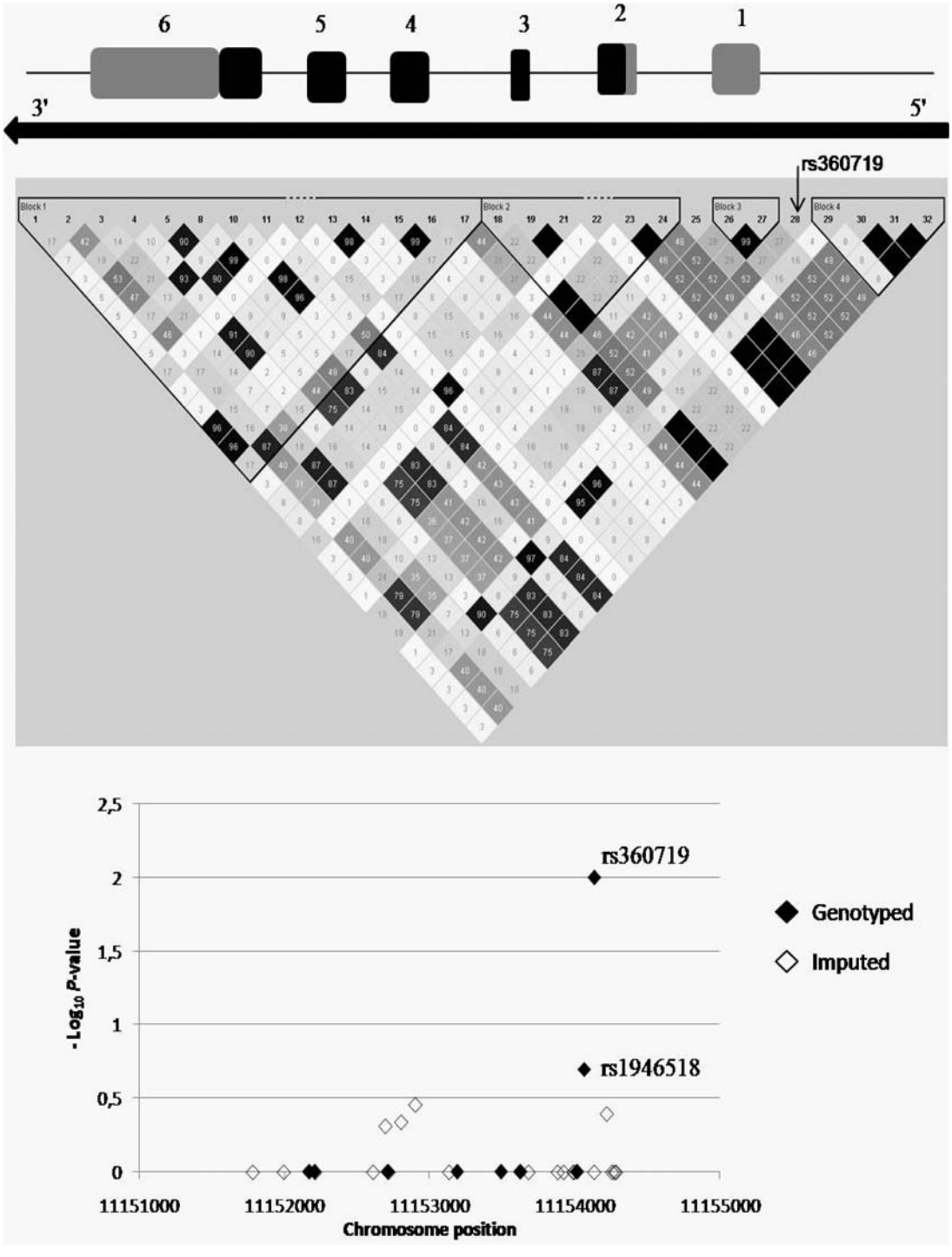


Figure 1. (**A**) Schematic representation of the *IL18* gene. Grey boxes indicate UTRs, black boxes coding exons. (**B**) LD r^2 chart from Haploview that summarized the pattern of LD in the Spanish population as a coloured plot. Black represents regions of high pairwise r^2, and white represents regions of low pairwise r^2. The numbers in the boxes are the pairwise r^2 values. In the diagram the SNPs used in the analysis have been recoded as numbers from 1 to 32: SNP1 (rs544354), SNP2 (rs3882891), SNP3 (rs5744280), SNP4 (rs5744276), SNP5 (rs360729), SNP8 (rs549908), SNP10 (rs5744258), SNP11 (rs360720), SNP12 (rs5744257), SNP13 (rs5744256), SNP14 (rs1834481), SNP15 (rs5744247), SNP16 (rs360722), SNP17 (rs4937113), SNP18 (rs795467), SNP19 (rs2043055), SNP21 (rs7106524), SNP22 (rs5744232), SNP23 (rs360717), SNP24 (rs360718), SNP25 (rs187238), SNP26 (rs1946518), SNP27 (rs1946519), SNP28 (rs360719), SNP29 (rs5744222), SNP30 (rs1293344), SNP31 (rs1290349) and SNP32 (rs11214105). (**C**) The graphic shows the significance of the association data (presented as $-\log_{10}$ *P*-value) for nine genotyped SNPs (solid diamonds) and 19 imputed SNPs (open diamonds) in 792 case patients and 595 controls.

populations. An important issue is how this *IL18* variant affects the expression of the gene. Several transcription factor binding sites that may be involved in the gene regulation of *IL18* were identified by Kalina *et al.* (34). The rs360719 polymorphism in *IL18* leads to loss of the OCT-1 transcription factor binding site. OCT-1 is known as a ubiquitously expressed factor and

Table 2. Individual and pooled[a] genetic association analysis of *IL18* rs360719 polymorphism in three sets of SLE cases and controls

	Population		TT, *n* (%)	TC, *n* (%)	CC, *n* (%)	*P*-value	Odds ratio (95% CI)	Allele T, *n* (%)	Allele C, *n* (%)	*P*-value	Odds ratio (95% CI)
IL18 rs360719											
	Spain SLE	Cases (735)	325 (44.2)	314 (42.7)	96 (13.1)	0.3	TC: 1.14 (0.91–1.41)	964 (65.6)	506 (34.4)		
		Controls (593)	324 (54.6)	235 (39.6)	34 (5.7)	7.9E−06	CC: 2.43 (1.61–3.63)	883 (74.5)	303 (25.5)	7.8E−07	1.53 (1.29–1.81)
	Italy SLE	Cases (348)	182 (52.3)	131 (37.6)	35 (10.1)	0.4	TC: 0.88 (0.65–1.19)	495 (71.1)	201 (28.9)		
		Controls (364)	200 (55)	144 (39.5)	20 (5.5)	0.02	CC: 1.88 (1.07–3.31)	544 (74.7)	184 (25.3)	0.1	1.20 (0.95–1.51)
	Argentina SLE	Cases (275)	128 (47.1)	119 (42.7)	28 (10.2)	0.4	TC: 0.86 (0.61–1.22)	375 (68.4)	173 (31.6)		
		Controls (245)	122 (49.8)	108 (44.1)	15 (6.1)	0.09	CC: 1.70 (0.89–3.23)	352 (71.8)	138 (28.2)	0.2	1.18 (0.90–1.53)
	Pooled	Cases (1358)	635 (46.8)	564 (41.5)	159 (11.7)	0.6	TC: 1.04 (0.89–1.22)	1834 (67.5)	882 (32.5)		
		Controls (1202)	646 (53.7)	487 (40.5)	69 (5.7)	1.2E−07	CC: 2.16 (1.61–2.89)	1779 (74)	625 (26)	3.8E−07*	1.37 (1.21–1.54)

*$Pc = 1.16E - 06$.

Figure 2. Relative quantification of *IL18* mRNA expression in PBMCs from 23 individuals (TT, $n = 12$; CC+CT, $n = 11$). The *Y*-axis shows the fold change in expression of the target gene (*IL18*) relative to the internal control genes (*β-actin, GADPH* and *ABL*) represented as Log_{10}.

is involved in the regulation of several genes. It can also repress the expression of certain genes, including some cytokines (35–39). We have shown that the presence of the T allele at position -1297 (rs360719) may play a key role in the transcription of the *IL18* gene and its role could be mediated through OCT-1 binding. This suppression would result in reduced IL-18 production and potential protection against IL-18 overexpression in disorders such as SLE, in which a persistent inflammatory response appears to be an underlying pathogenic process. Such a functional explanation would be consistent with the observation that the rs360719 C-allele is associated with SLE, in which IL-18 overexpression is generally observed (15–17). Thus, the functional experiments that we describe indicate differences in the

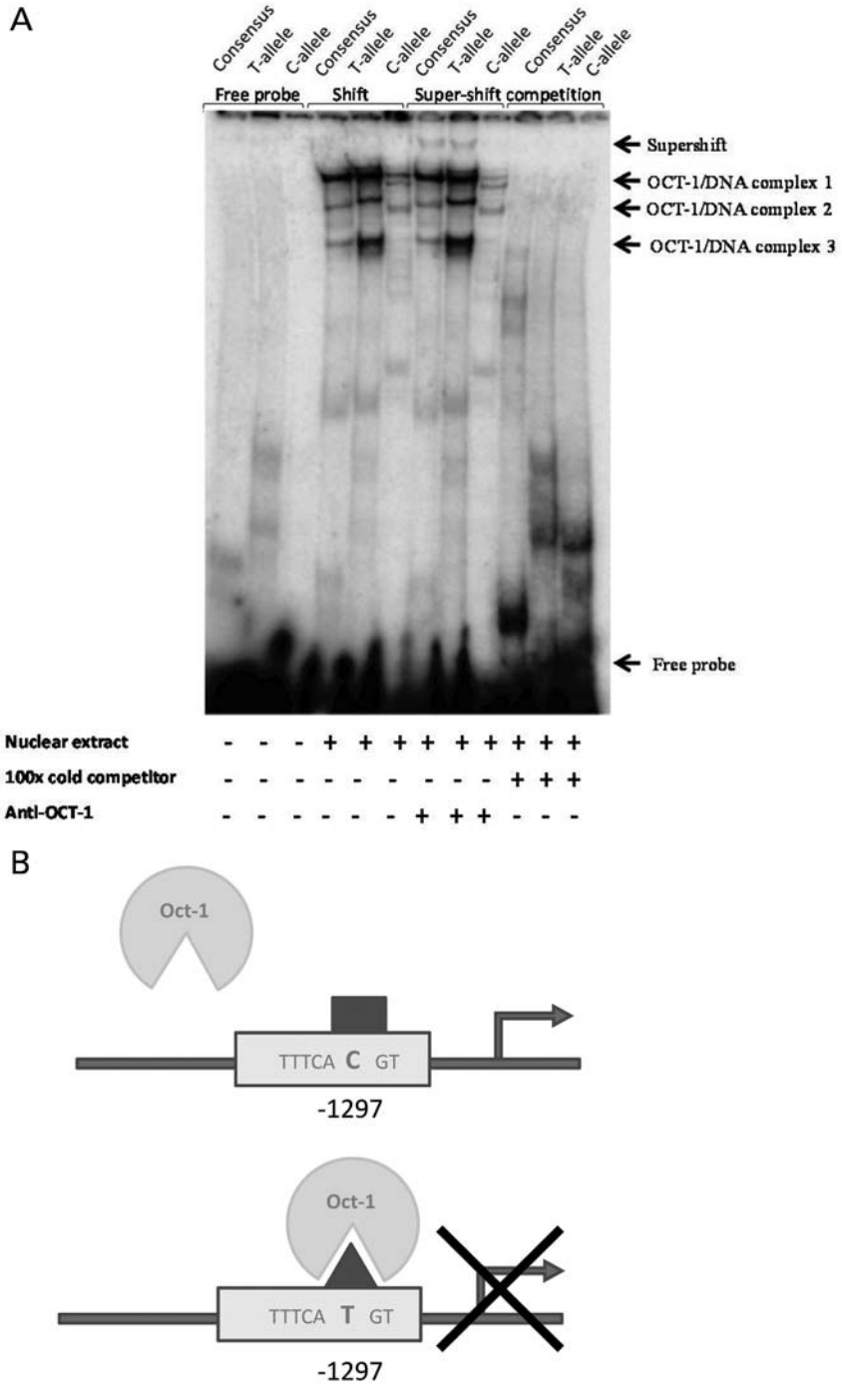


Figure 3. Action of OCT-1 in the preferential binding of the rs360719 T-allele and repression of transcription. (**A**) Electrophoretic mobility shift assay for the two alleles of the rs360719 polymorphism showing stronger protein binding to the T-allele compared with the C-allele. Lanes 1–3 contain labelled probe alone (consensus: lane 1; wild-type: lane 2 and mutant: lane 3). Lanes 4–6 contain labelled probe (consensus: lane 4; wild-type: lane 5 and mutant: lane 6) and nuclear extract. Lanes 7–9 shows supershifts experiments using an antibody specific to OCT-1 (consensus: lane 7; wild-type: lane 8 and mutant: lane 9). Competition experiments were performed using 100-fold molar excess of the cold probe consensus sequence (lanes 10–12). (**B**) Proposed model of interaction of the polymorphic site with the OCT-1 transcription factor in the *IL18* gene.

transcriptional activity of the rs360719 T and rs360719 C-alleles and also identify a possible major repressor site in the *IL18* promoter (the rs360719 T-allele). On this basis the results from the present study suggest that the *IL18* rs360719 gene polymorphism may play a major role in SLE; however, complete re-sequencing in the area to examine all possible functional variants and perhaps additional functional data in different cell types, such as antigen presenting cells or endothelial cells, would be needed to confirm this hypothesis.

Studies using animal models that develop spontaneous lupus-like autoimmune disease have provided further evidence that *IL18* is involved in the pathology (14,40). Interestingly, MRL/*lpr* mice have significantly elevated serum levels of IL-18 compared with MLR/++ controls, and MRL/*lpr* mice treated with IL-18 or IL-18 plus IL-12 resulted in accelerated proteinuria, glomerulonephritis, vasculitis and increased levels of proinflammatory cytokines. These data together with our findings suggest that IL-18 is a possible novel therapeutic target in the treatment of autoimmune SLE.

In conclusion, we have identified a putative functional variant within the *IL18* promoter region that seems to have an important role in IL-18 expression associated with susceptibility to SLE.

MATERIALS AND METHODS

Patients

Three independent case–control cohorts from Spain, Italy and Argentina were analysed. The study includes, 750 SLE patients and 595 controls from Spain, 330 SLE patients and 366 controls from Italy and 276 SLE patients and 262 controls from Argentina. The Spanish, Italian and Argentinean SLE cases have all been previously described (41–43). Both patient and control groups were matched for age and sex in each geographic region. All cases fulfil the American College of Rheumatology (ACR) criteria for the classification of SLE (44). The samples were collected according to the Helsinki Declaration. All subjects provided informed consent for this study. The study was approved by the various institutional review boards and Ethical Committees at each of the participating locations.

IL18 polymorphisms selection

SNPs spanning a 27 kb region from 2.9 kb upstream to 3.3 kb downstream of *IL18* transcribed sequence were surveyed in the NCBI-dbSNP (http://www.ncbi.nlm.nih.gov/SNP/index.html) and the International HapMap (http://www.hapmap.org) websites. We used the HapMap database of the CEU population (Utah residents with ancestry from northern and western Europe). Tag SNPs were selected using the pairwise method under a restriction of minor allele frequency >0.01 and r^2 threshold >0.8, aiming to identify a set of tag SNPs that efficiently captures all known common variants and is likely to tag most unknown variants. In all, seven tag SNPs were identified that capture all 26 alleles with a mean r^2 of 0.945. Six tag SNPs were located in an intron region and only one tag SNP was in the promoter region. No non-synomymous variations or polymorphisms that may interfere with mRNA splicing have been described in the *IL18* gene. In addition, variants in *IL18* with potential functional effects or locations in putative transcription factor binding sites were chosen for genotyping. They consisted of two promoter SNPs, the rs187238 polymorphism was selected because it was previously associated with SLE and have been suggested that this variant could alter the *IL18* promoter activity. The rs360719 was selected based on the minor allele frequency and its ability to bind the transcription factor OCT-1. The SNPs associated in the Spanish fine mapping, after quality control and correction for multiple testing, were typed in the Italian and Argentinean samples.

IL18 genotyping

DNA was obtained from peripheral blood mononuclear cells (PBMC), using standard methods. The genotyping of all *IL18* polymorphisms was performed using a pre-development TaqMan SNP Genotyping Assays (Applied Biosystems, Foster City, CA). The PCR reaction was carried out in a total reaction volume of 5 μl, containing 50 ng genomic DNA as template, 2 μl of TaqMan genotyping master mix, 0.1 μl of 20× assay mix and ddH_2O up to 5 μl of final volume. The amplification protocol used was initial denaturation at 95°C for 10 min followed by 40 cycles of denaturation at 92°C for 15 s and annealing/extension at 60°C for 1 min. After PCR, the genotype of each sample was automatically attributed by measuring the allele-specific fluorescence in the ABI Prism 7900 Sequence Detection System, using the SDS 2.2.2 software for allele discrimination (Applied Biosystems).

All samples were genotyped in the same centre to avoid genotyping inconsistencies and to verify the genotyping consistency; random samples were genotyped twice showing 99% identical genotypes.

Imputation analysis

Imputation was performed on the cases and controls using IMPUTE, using a method described by Marhcini *et al*. (45). We imputed all SNPs in the HAPMAP within the range of our data plus 3 kb either side, giving us 28 SNPs (19 imputed) for our case–control analysis.

The output from IMPUTE gives probabilities for each genotype, rather than point estimates. The use of probabilities allowed us to account for the uncertainty in imputation within the case–control analysis using SNPTEST.

Real-time quantitative PCR

To analyse constitutive *IL18* mRNA expression, PBMCs from 23 selected healthy individuals were isolated by Ficoll density gradient centrifugation. Total RNA was isolated with Trizol according to the manufacturer's protocol (Invitrogen, Carlsbad, CA). RNA integrity was verified both electrophoretically and by the 260/280 nm absorption ratio. Reverse-transcription was performed in a total volume of 20 μl with Superscript™ First-Strand Synthesis System for RT–PCR (Invitrogen) as recommended by the manufacturer. Real-time quantitative PCR was performed on an ABI PRISM 7500 Fast SDS (Applied Biosystems) using a TaqMan gene expression assay (Applied Biosystems) in a total volume of 20 μl using 10 μl of TaqMan Fast Universal PCR Master Mix, 1 μl of each probe and 200 ng of cDNA. Cycle conditions were 95°C for 20 s followed by 40 cycles at 95°C for 3 s and 60°C for 30 s. Each sample was tested in triplicate and a sample without template was included as a negative control. Relative expression levels of *IL18* mRNA were normalized according to β-actin, GAPDH and ABL expression using the $2^{-\Delta\Delta CT}$ method (46). We have used three different

endogenous controls in order to minimize the potential variability characteristic of each single housekeeping gene.

Cell culture

Jurkat cells were purchased from American Type Culture Collection (ATCC) and were cultured in RPMI (Rosweli Park Memorial Institute) 1640 medium (PAA Laboratories GmbH) supplemented with 10% fetal bovine serum (FBS, Gibco), 1 mM glutamine and 1% penicillin/streptomycin. They were kept at 37°C in a humidified 5% CO2/95% air incubator.

Electrophoretic mobility-shift assay

Nuclear extract from Jurkat cells was prepared by the mini-extraction procedure as described previously (47). The double-stranded oligonucleotides (50 ng) spanning the *IL18* rs360719 polymorphism were as follows: top-strand T allele oligonucleotide 5′-CACTTCGTGCTTTCATGTTATTGGCCCAA T-3′ and top-strand C allele oligonucleotide 5′-CACTTCGT GCTTTCACGTTATTGGCCCAAT-3′. A pair of oligonucleotides corresponding to the OCT-1 consensus binding sequence (48) (5′-TGTCGAATGCAAATCACTAGAA-3′ and 3′-TTCT AGTGATTTGCATTCGACA-5′) was end-labelled with (γ-^{32}P)adenosine 5′-triphosphate (ATP) by using T4 polynucleotide kinase (Promega corporation, Madison, WI). For EMSAs with nuclear extract, 20 000–50 000 cpm double-stranded oligonucleotides corresponding to ~0.5 ng were used for each reaction. The binding-reaction mixtures were set up containing 15 pmol DNA probe, 5 μg nuclear extract, 2 μg poly(dI-dC).poly(dI-dC) and binding buffer 2× (40 mM HEPES pH 7.5, 200 mM ClNa, 4 mM Cl_2Mg, 4 mM DTT, 10% glycerol, 200 μg/ml BSA) up to 20 μl. The mixtures were incubated on ice for 15 min before adding the probe, followed by another 20 min at room temperature. In the competition assay 100-fold excess amounts of cold oligonucleotides agains the probe used were incubated with the Jurkat nuclear extracts. In the super-shift assay a mouse monoclonal antibody that recognizes human OCT-1 (Anti-OCT-1) was used (Abnova Corporation, Taipei, Taiwan). The standard binding reaction mixtures were treated with anti-OCT-1 antibody for 15 min at room temperature. Samples were loaded onto 7% non-denaturing polyacrylamide gels and electrophoresed in 0.5×TBE (45 mM Tris base, 45 mM boric acid, 1 mM EDTA pH 8.0) buffer at 100 V, followed by transfer to Whatman paper and visualized by phosphorimager analysis.

Western blot

The protein levels of IL-18 and OCT-1 were determined by western blot analysis with specific polyclonal antibody for IL-18 (Santa Cruz Biotechnology Inc., Santa Cruz, CA) and monoclonal antibody for OCT-1 (Abnova). Fifty micrograms of nuclear proteins was separated on a 7.5% (for OCT-1) or 12% (for IL-18) SDS–PAGE and transferred to PVDF membrane (Immobilon P, Millipore). The membrane was blocked in washing solution (0.01 M Tris, 0.1 M NaCl, 0.1% Tween 20; pH 7.5) with 5% non-fat dried milk, for 1 h at room temperature. It was first incubated overnight with 1 μg/ml (for IL-18) or 10 μg/ml (for OCT-1) of primary antibody at 4°C and then with a peroxidise-conjugated secondary antibody for 1 h at room temperature. The bands were detected with a chemiluminescent system (ECL, Amersham, Arlington Heights, IL) and exposed to X-ray film.

Flow cytometry analysis

The expression of IL-18 and OCT-1 was determined by flow cytometry. Jurkat cells were cultured for 18 h with 5 ng/ml phorbol myristate acetate (PMA) and 1 μM ionomycin to induce IL-18 and OCT-1 expression, followed by 6 h incubation with monensin in order to trap cytokine production within the cells. The cultured cells were fixed and permeabilized according to manufacturer's recommendations with CtyoFix/CytoPerm Kit (BD Biosciences Inc, Franklin Lakes, NJ). Afterwards, cells were incubated with anti-human IL-18 (10 μg/ml) or with anti-human OCT-1 (10 μg/ml) (Santa Cruz and ABnova, respectively), followed by incubation with phycoerythrin (PE)-anti-rabbit IgG or fluorescein isothiocyanate (FITC)-anti-mouse IgG secondary antibodies, respectively. Samples were analysed in a FACSCalibur flow cytometer (BD Biosciences, Inc) using the CellQuest Software (BD Biosciences, Inc.).

Data analysis

Allele and genotype frequencies were obtained by direct counting. We used the χ^2 test for statistical analysis to compare allelic and genotypic distributions. We assessed the quality of the genotype data by testing for Hardy–Weinberg equilibrium for all samples using Fisher's exact test and found no differences. Odds ratios (OR) with 95% confidence intervals (95% CI) were calculated according to Woolf's method. All statistics described earlier were performed with PLINK (http://pngu.mgh.harvard.edu/~purcell/plink/). *P*-values below 0.05 were regarded as statistically significant. The Breslow–Day test of combinability and the Mantel–Haenszel test were carried out using the StatsDirect software v2.4.6. The pooled OR was calculated according to a fixed-effects model (Mantel–Haenszel meta-analysis). Genotypic ORs were calculated using the Unphased software with homozygosity for non-associated allele as reference with OR = 1.

Results relative to mRNA expression are shown as mean ± standard deviation. Because the variances were homogeneous (Bartlett's test $P > 0.05$), a statistical analysis of the mean of relative expression of the IL-18 and OCT-1 was performed using the ANOVA test included in GraphPad Prism 4 software (GraphPad Software; Inc, La Jolla, CA).

SUPPLEMENTARY MATERIAL

Supplementary Material is available at *HMG* online.

ACKNOWLEDGEMENTS

We thank Francisco Javier Oliver, PhD and Gema Robledo for their excellent technical assistance and all the patients and control donors for their essential collaboration.

Conflict of Interest statement. None declared.

FUNDING

This work was supported by Plan Nacional de I+D, Spain (grant: SAF06-00398) and Junta de Andalucía, grant: CTS-1180. This work was partially supported by RETICS Program, RD08/0075 (RIER) from Instituto de Salud Carlos III (ISCIII). The study was also funded by the Swedish Research Council, the Swedish Association Against Rheumatism and the Torsten and Ragnar Söderbergs Foundation to M.E.A.-R. University of Eastern Piedmont (ex 60%), Regione Piemonte (Ricerca Sanitaria Finalizzata Project and Ricerca Sanitaria Applicata-CIPE Project). B.A.P.-E. is coordinator of the Argentine collaborative group.

REFERENCES

1. Wakeland, E.K., Liu, K., Graham, R.R. and Behrens, T.W. (2001) Delineating the genetic basis of systemic lupus erythematosus. *Immunity*, **15**, 397–408.
2. Alarcon-Riquelme, M.E. (2005) The genetics of systemic lupus erythematosus. *J. Autoimmun.*, **25**(suppl.), 46–48.
3. Gracie, J.A., Robertson, S.E. and McInnes, I.B. (2003) Interleukin-18. *J. Leukoc. Biol.*, **73**, 213–224.
4. McInnes, I.B., Gracie, J.A., Leung, B.P., Wei, X.Q. and Liew, F.Y. (2000) Interleukin 18: a pleiotropic participant in chronic inflammation. *Immunol. Today*, **21**, 312–315.
5. Okamura, H., Tsutsui, H., Kashiwamura, S., Yoshimoto, T. and Nakanishi, K. (1998) Interleukin-18: a novel cytokine that augments both innate and acquired immunity. *Adv. Immunol.*, **70**, 281–312.
6. Dinarello, C.A. (1999) Interleukin-18. *Methods*, **19**, 121–132.
7. Nakanishi, K., Yoshimoto, T., Tsutsui, H. and Okamura, H. (2001) Interleukin-18 regulates both Th1 and Th2 responses. *Annu. Rev. Immunol.*, **19**, 423–474.
8. Nakanishi, K., Yoshimoto, T., Tsutsui, H. and Okamura, H. (2001) Interleukin-18 is a unique cytokine that stimulates both Th1 and Th2 responses depending on its cytokine milieu. *Cytokine Growth Factor Rev.*, **12**, 53–72.
9. Yoshimoto, T., Takeda, K., Tanaka, T., Ohkusu, K., Kashiwamura, S., Okamura, H., Akira, S. and Nakanishi, K. (1998) IL-12 up-regulates IL-18 receptor expression on T cells, Th1 cells, and B cells: synergism with IL-18 for IFN-gamma production. *J. Immunol.*, **161**, 3400–3407.
10. Tominaga, K., Yoshimoto, T., Torigoe, K., Kurimoto, M., Matsui, K., Hada, T., Okamura, H. and Nakanishi, K. (2000) IL-12 synergizes with IL-18 or IL-1beta for IFN-gamma production from human T cells. *Int. Immunol.*, **12**, 151–160.
11. Yoshimoto, T., Mizutani, H., Tsutsui, H., Noben-Trauth, N., Yamanaka, K., Tanaka, M., Izumi, S., Okamura, H., Paul, W.E. and Nakanishi, K. (2000) IL-18 induction of IgE: dependence on CD4+ T cells, IL-4 and STAT6. *Nat. Immunol.*, **1**, 132–137.
12. Yoshimoto, T., Tsutsui, H., Tominaga, K., Hoshino, K., Okamura, H., Akira, S., Paul, W.E. and Nakanishi, K. (1999) IL-18, although antiallergic when administered with IL-12, stimulates IL-4 and histamine release by basophils. *Proc. Natl Acad. Sci. USA*, **96**, 13962–13966.
13. Akahoshi, M., Nakashima, H., Tanaka, Y., Kohsaka, T., Nagano, S., Ohgami, E., Arinobu, Y., Yamaoka, K., Niiro, H., Shinozaki, M. *et al.* (1999) Th1/Th2 balance of peripheral T helper cells in systemic lupus erythematosus. *Arthritis Rheum.*, **42**, 1644–1648.
14. Esfandiari, E., McInnes, I.B., Lindop, G., Huang, F.P., Field, M., Komai-Koma, M., Wei, X. and Liew, F.Y. (2001) A proinflammatory role of IL-18 in the development of spontaneous autoimmune disease. *J. Immunol.*, **167**, 5338–5347.
15. Wong, C.K., Ho, C.Y., Li, E.K. and Lam, C.W. (2000) Elevation of proinflammatory cytokine (IL-18, IL-17, IL-12) and Th2 cytokine (IL-4) concentrations in patients with systemic lupus erythematosus. *Lupus*, **9**, 589–593.
16. Wong, C.K., Li, E.K., Ho, C.Y. and Lam, C.W. (2000) Elevation of plasma interleukin-18 concentration is correlated with disease activity in systemic lupus erythematosus. *Rheumatology (Oxford)*, **39**, 1078–1081.
17. Amerio, P., Frezzolini, A., Abeni, D., Teofoli, P., Girardelli, C.R., De Pita, O. and Puddu, P. (2002) Increased IL-18 in patients with systemic lupus erythematosus: relations with Th-1, Th-2, pro-inflammatory cytokines and disease activity. IL-18 is a marker of disease activity but does not correlate with pro-inflammatory cytokines. *Clin. Exp. Rheumatol.*, **20**, 535–538.
18. Nolan, K.F., Greaves, D.R. and Waldmann, H. (1998) The human interleukin 18 gene IL18 maps to 11q22.2-q22.3, closely linked to the DRD2 gene locus and distinct from mapped IDDM loci. *Genomics*, **51**, 161–163.
19. Johansson, C.M., Zunec, R., Garcia, M.A., Scherbarth, H.R., Tate, G.A., Paira, S., Navarro, S.M., Perandones, C.E., Gamron, S., Alvarellos, A. *et al.* (2004) Chromosome 17p12-q11 harbors susceptibility loci for systemic lupus erythematosus. *Hum. Genet.*, **115**, 230–238.
20. Lindqvist, A.K., Steinsson, K., Johanneson, B., Kristjansdottir, H., Arnasson, A., Grondal, G., Jonasson, I., Magnusson, V., Sturfelt, G., Truedsson, L. *et al.* (2000) A susceptibility locus for human systemic lupus erythematosus (hSLE1) on chromosome 2q. *J. Autoimmun.*, **14**, 169–178.
21. Higa, S., Hirano, T., Mayumi, M., Hiraoka, M., Ohshima, Y., Nambu, M., Yamaguchi, E., Hizawa, N., Kondo, N., Matsui, E. *et al.* (2003) Association between interleukin-18 gene polymorphism 105A/C and asthma. *Clin. Exp. Allergy*, **33**, 1097–1102.
22. Imboden, M., Nicod, L., Nieters, A., Glaus, E., Matyas, G., Bircher, A.J., Ackermann-Liebrich, U., Berger, W. and Probst-Hensch, N.M. (2006) The common G-allele of interleukin-18 single-nucleotide polymorphism is a genetic risk factor for atopic asthma. The SAPALDIA Cohort Study. *Clin. Exp. Allergy*, **36**, 211–218.
23. Tiret, L., Godefroy, T., Lubos, E., Nicaud, V., Tregouet, D.A., Barbaux, S., Schnabel, R., Bickel, C., Espinola-Klein, C., Poirier, O. *et al.* (2005) Genetic analysis of the interleukin-18 system highlights the role of the interleukin-18 gene in cardiovascular disease. *Circulation*, **112**, 643–650.
24. Kretowski, A., Mironczuk, K., Karpinska, A., Bojaryn, U., Kinalski, M., Puchalski, Z. and Kinalska, I. (2002) Interleukin-18 promoter polymorphisms in type 1 diabetes. *Diabetes*, **51**, 3347–3349.
25. Mojtahedi, Z., Naeimi, S., Farjadian, S., Omrani, G.R. and Ghaderi, A. (2006) Association of IL-18 promoter polymorphisms with predisposition to Type 1 diabetes. *Diabet. Med.*, **23**, 235–239.
26. Sugiura, T., Maeno, N., Kawaguchi, Y., Takei, S., Imanaka, H., Kawano, Y., Terajima-Ichida, H., Hara, M. and Kamatani, N. (2006) A promoter haplotype of the interleukin-18 gene is associated with juvenile idiopathic arthritis in the Japanese population. *Arthritis Res. Ther.*, **8**, R60.
27. Tamura, K., Fukuda, Y., Sashio, H., Takeda, N., Bamba, H., Kosaka, T., Fukui, S., Sawada, K., Satomi, M., Yamada, T. *et al.* (2002) IL18 polymorphism is associated with an increased risk of Crohn's disease. *J. Gastroenterol.*, **37**(Suppl. 14), 111–116.
28. Sivalingam, S.P., Yoon, K.H., Koh, D.R. and Fong, K.Y. (2003) Single-nucleotide polymorphisms of the interleukin-18 gene promoter region in rheumatoid arthritis patients: protective effect of AA genotype. *Tissue Antigens*, **62**, 498–504.
29. Thompson, S.R. and Humphries, S.E. (2007) Interleukin-18 genetics and inflammatory disease susceptibility. *Genes Immun.*, **8**, 91–99.
30. Xu, Q., Tin, S.K., Sivalingam, S.P., Thumboo, J., Koh, D.R. and Fong, K.Y. (2007) Interleukin-18 promoter gene polymorphisms in Chinese patients with systemic lupus erythematosus: association with CC genotype at position -607. *Ann. Acad. Med. Singapore*, **36**, 91–95.
31. Lin, Y.J., Wan, L., Lee, C.C., Huang, C.M., Tsai, Y., Tsai, C.H., Shin, T.L., Chao, K., Liu, C.M., Xiao, J.W. *et al.* (2007) Disease association of the interleukin-18 promoter polymorphisms in Taiwan Chinese systemic lupus erythematosus patients. *Genes Immun.*, **8**, 302–307.
32. Chen, D.Y., Hsieh, C.W., Chen, K.S., Chen, Y.M., Lin, F.J. and Lan, J.L. (2009) Association of interleukin-18 promoter polymorphisms with WHO pathological classes and serum IL-18 levels in Chinese patients with lupus nephritis. *Lupus*, **18**, 29–37.
33. Mori, M., Yamada, R., Kobayashi, K., Kawaida, R. and Yamamoto, K. (2005) Ethnic differences in allele frequency of autoimmune-disease-associated SNPs. *J. Hum. Genet.*, **50**, 264–266.
34. Kalina, U., Ballas, K., Koyama, N., Kauschat, D., Miething, C., Arnemann, J., Martin, H., Hoelzer, D. and Ottmann, O.G. (2000) Genomic

organization and regulation of the human interleukin-18 gene. *Scand. J. Immunol.*, **52**, 525–530.
35. Pankratova, E.V., Sytina, E.V., Luchina, N.N. and Krivega, I.V. (2003) The regulation of the Oct-1 gene transcription is mediated by two promoters. *Immunol. Lett.*, **88**, 15–20.
36. Ullman, K.S., Flanagan, W.M., Edwards, C.A. and Crabtree, G.R. (1991) Activation of early gene expression in T lymphocytes by Oct-1 and an inducible protein, OAP40. *Science*, **254**, 558–562.
37. Delhase, M., Castrillo, J.L., de la Hoya, M., Rajas, F. and Hooghe-Peters, E.L. (1996) AP-1 and Oct-1 transcription factors down-regulate the expression of the human PIT1/GHF1 gene. *J. Biol. Chem.*, **271**, 32349–32358.
38. Wu, G.D., Lai, E.J., Huang, N. and Wen, X. (1997) Oct-1 and CCAAT/enhancer-binding protein (C/EBP) bind to overlapping elements within the interleukin-8 promoter. The role of Oct-1 as a transcriptional repressor. *J. Biol. Chem.*, **272**, 2396–2403.
39. Kaushansky, K., Shoemaker, S.G., O'Rork, C.A. and McCarty, J.M. (1994) Coordinate regulation of multiple human lymphokine genes by Oct-1 and potentially novel 45 and 43 kDa polypeptides. *J. Immunol.*, **152**, 1812–1820.
40. Neumann, D., Tschernig, T., Popa, D., Schmiedl, A., Perez de Lema, G., Resch, K. and Martin, M.U. (2006) Injection of IL-12- and IL-18-encoding plasmids ameliorates the autoimmune pathology of MRL/Mp-Tnfrsf6lpr mice: synergistic effect on autoimmune symptoms. *Int. Immunol.*, **18**, 1779–1787.
41. Graham, R.R., Kozyrev, S.V., Baechler, E.C., Reddy, M.V., Plenge, R.M., Bauer, J.W., Ortmann, W.A., Koeuth, T., Gonzalez Escribano, M.F., Pons-Estel, B. *et al.* (2006) 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.*, **38**, 550–555.
42. Kozyrev, S.V., Abelson, A.K., Wojcik, J., Zaghlool, A., Linga Reddy, M.V., Sanchez, E., Gunnarsson, I., Svenungsson, E., Sturfelt, G., Jonsen, A. *et al.* (2008) Functional variants in the B-cell gene BANK1 are associated with systemic lupus erythematosus. *Nat. Genet.*, **40**, 211–216.
43. Sanchez, E., Gomez, L.M., Lopez-Nevot, M.A., Gonzalez-Gay, M.A., Sabio, J.M., Ortego-Centeno, N., de Ramon, E., Anaya, J.M., Gonzalez-Escribano, M.F., Koeleman, B.P. *et al.* (2006) Evidence of association of macrophage migration inhibitory factor gene polymorphisms with systemic lupus erythematosus. *Genes Immun.*, **7**, 433–436.
44. Hochberg, M.C. (1997) Updating the American College of Rheumatology revised criteria for the classification of systemic lupus erythematosus. *Arthritis Rheum.*, **40**, 1725.
45. 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.
46. Livak, K.J. and Schmittgen, T.D. (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(−Delta Delta C(T)) Method. *Methods*, **25**, 402–408.
47. Delgado, M., Munoz-Elias, E.J., Gomariz, R.P. and Ganea, D. (1999) Vasoactive intestinal peptide and pituitary adenylate cyclase-activating polypeptide prevent inducible nitric oxide synthase transcription in macrophages by inhibiting NF-kappa B and IFN regulatory factor 1 activation. *J. Immunol.*, **162**, 4685–4696.
48. Ishimaru, N., Kishimoto, H., Hayashi, Y. and Sprent, J. (2006) Regulation of naive T cell function by the NF-kappaB2 pathway. *Nat. Immunol.*, **7**, 763–772.

APPENDIX: THE LIST OF PARTICIPANTS

The Argentine collaborative participants: 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. Juán 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 Italian collaborative participants: Nadia Barizzone (University of Eastern Piedmont, Novara, Italy), Mauro Galeazzi (Siena University, Siena, Italy), Maria Giovanna Danieli (Clinica Medica di Scienze Mediche e Chirurgiche, Universitá Politecnica delle Marche), Sergio Migliaresi (Rheumatology Unit Second University of Naples, Naples, Italy), Enrica Bozzolo (IRCCS San Raffaele Hospital, Milan, Italy).

SUPPLEMENTARY MATERIAL

Figure legends:

Figure 1.Supplementary material. Protein expression of OCT-1 and IL18 in Jurkat cells using western blot analysis. Protein from cells incubated in medium alone (controls) or PMA + Ionomycin for 24 h were separated on a 7.5% (for OCT-1) or 12% (for IL-18) SDS-PAGE gel, transferred to a PVDF membrane and hybridized with an anti-OCT-1 or anti-IL-18 antibody. The molecular weights are 97 KDa for OCT-1 and 18 KDa for IL-18.

Figure 2. Supplementary material. IL-18 and OCT-1 expression were determined by flow cytometry. Values are the mean and SD. MCF= mean channel fluorescence. A) Overlaid histrogram showing IL-18 expression in unstimulated and PMA + Ionomycin stimulated Jurkat cells. B) Overlaid histogram showing OCT-1 expression in unstimulated and PMA + Ionomycin-stimulated Jurkat cells.

Figure 1.

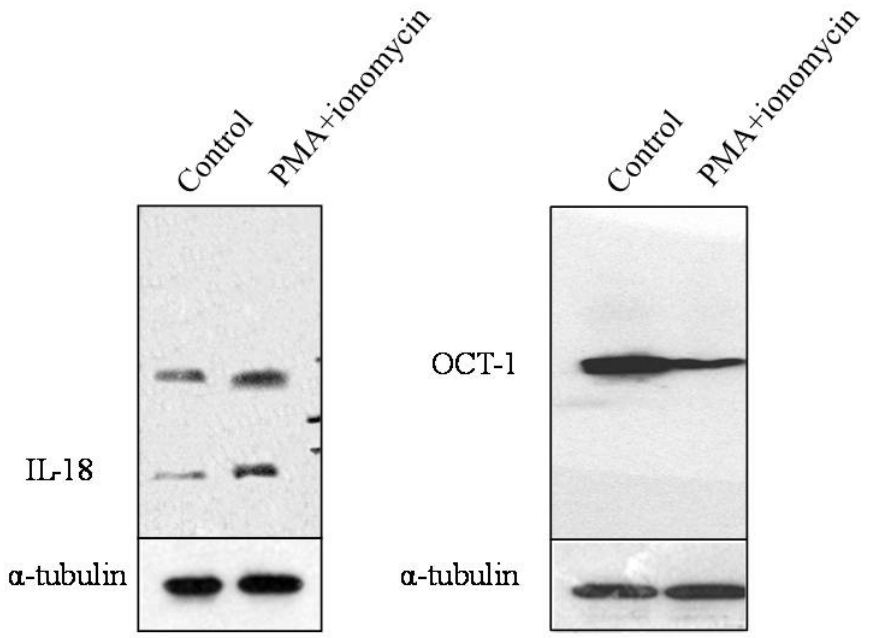
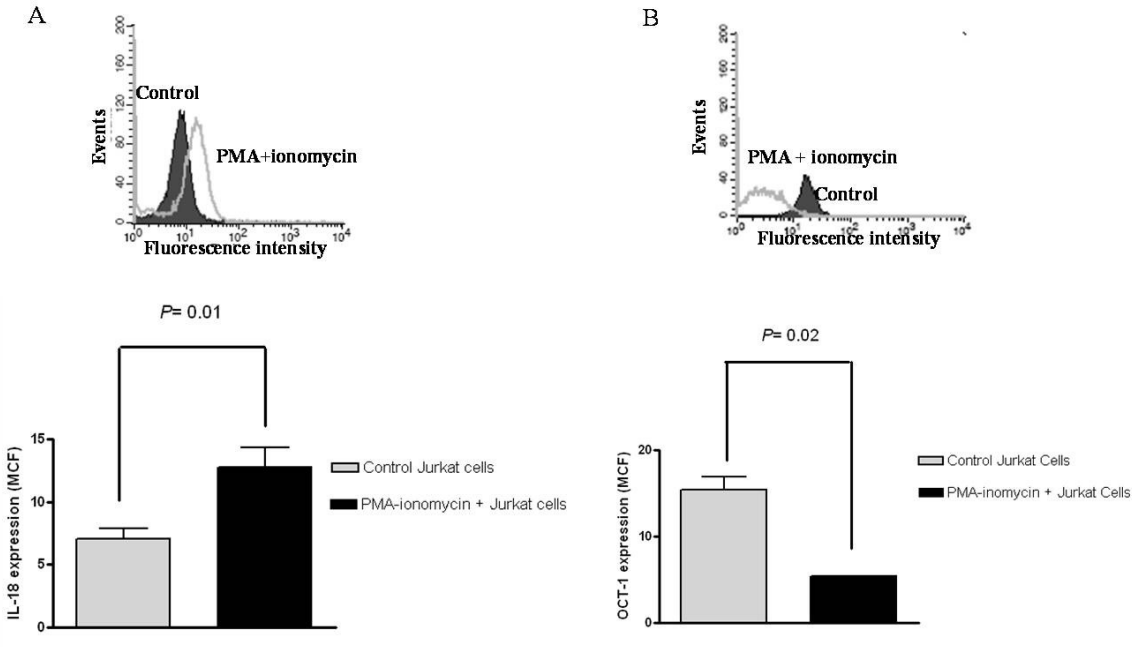

Control
PMA+ionomycin
IL-18
α-tubulin
Control
PMA+ionomycin
OCT-1
α-tubulin

Figure 2.


A
Control
PMA+ionomycin
Events
Fluorescence intensity
B
PMA + ionomycin
Control
Events
Fluorescence intensity
P= 0.01
IL-18 expression (MCF)
Control Jurkat cells
PMA-ionomycin + Jurkat cells
P= 0.02
OCT-1 expression (MCF)
Control Jurkat Cells
PMA-inomycin + Jurkat Cells

AM_7

Genes and Immunity (2008) 9, 7–15

www.nature.com/gene

ORIGINAL ARTICLE

A sequence variation in the MOG gene is involved in multiple sclerosis susceptibility in Italy

S D'Alfonso[1], E Bolognesi[1], FR Guerini[2], N Barizzone[1], S Bocca[1], D Ferrante[3], L Castelli[1], L Bergamaschi[1], C Agliardi[2], P Ferrante[2,4], P Naldi[5], M Leone[5], D Caputo[6], C Ballerini[7], M Salvetti[8], D Galimberti[9], L Massacesi[7], M Trojano[10] and P Momigliano-Richiardi[1]

[1]*Department of Medical Sciences, Interdisciplinary Research Center of Autoimmune Diseases (IRCAD), University of Eastern Piedmont, Novara, Italy;* [2]*Laboratory of Molecular Medicine and Biotechnology, Don C Gnocchi Foundation IRCCS, S Maria Nascente, Milan, Italy;* [3]*Unit of Medical Statistics and Cancer Epidemiology, Department of Medical Sciences, University of Eastern Piedmont and Ospedale Maggiore, Novara, Italy;* [4]*Department of Biomedical Sciences and Technology, University of Milan, Milan, Italy;* [5]*Department of Neurology, Ospedale Maggiore, Novara, Italy;* [6]*Multiple Sclerosis Unit, Don C Gnocchi Foundation ONLUS, IRCCS, Milan, Italy;* [7]*Department of Neurological and Psychiatric Sciences, University of Florence, Florence, Italy;* [8]*Neurology and Center for Experimental Neurological Therapy (CENTERS), Universitá La Sapienza, Roma, Italy;* [9]*Department of Neurological Sciences, Ospedale Maggiore Policlinico, University of Milan, Milan, Italy and* [10]*Department of Neurological and Psychiatric Sciences, University of Bari, Bari, Italy*

*Several studies suggest that the histocompatibility complex (HLA) class I region harbours genes modulating multiple sclerosis (MS) susceptibility independently from the effect of class II alleles. A candidate gene in this region is MOG, encoding the myelin oligodendrocyte glycoprotein. A significant association with the missense variation V142L (rs2857766) was previously reported in a small sample of 50 Italian MS patients. We confirmed this result in two independent Italian sample sets consisting of 878 MS patients and 890 matched controls ($P = 6.6 \times 10^{-4}$) and 246 trio families ($P = 1.5 \times 10^{-3}$). The comparison of genotype frequencies suggested a dominant-protective effect of L142. In the combined sample sets L142 conferred an odds ratio (OR) = 0.70 (95% confidence interval (CI): 0.60–0.82) that remained similar after accounting for HLA-DRB1*15 carrier status. The association with MOG V142L was still significant after conditioning for all DRB1 alleles ($P = 0.035$). Eleven additional single nucleotide polymorphisms in the MOG gene (namely −1077T/C, −910T/C, −875A/G, −93T/C, S5S, Indel L22, V145I, +814C/T, +900A/G, +1024A/T, +1059C/T), two microsatellites in the MOG 5′ flanking (MOGCA) and 3′ untranslated (MOGTAAA) regions and four microsatellites in the HLA-class I region, from HLA-B to HFE, (namely MIB, D6S265, D6S1683 and D6S2239) were tested by transmission disequilibrium test in 199 trio families. None of these polymorphisms or of their haplotypic combinations showed a significant transmission distortion, in the absence of V142L. In conclusion, MOG V142L, or an untested variant in tight-linkage disequilibrium with it, is an independent MS susceptibility-modulating factor in the HLA class I region.*

Genes and Immunity (2008) **9,** 7–15; doi:10.1038/sj.gene.6364437; published online 11 October 2007

Keywords: *MS; MOG; HLA; genetic association*

Introduction

Multiple sclerosis (MS) is a demyelinating autoimmune disease of the central nervous system[1] caused by an interplay of environmental and genetic factors.[2–4]

The only genetic factor that has been clearly demonstrated by linkage and association studies maps in the human major histocompatibility complex (HLA) region. In particular, the class II allele HLA-DRB1*1501 (DR15) is positively associated in all tested Caucasoid populations.[5–7] The only exception is represented by Sardinia where the frequency of DR15 is low both in patients and controls, while MS is principally associated with DR4 and DR3.[8] Although a weaker positive association with DR4[9–12] and DR3[13] has been reported also in other populations, this was not detected in a large case-control study in continental Italy.[6] Negative associations have also been reported.[6,7]

More recently, several studies suggested that genes located in the HLA class I region modulate susceptibility to MS independently of class II. Fogdell-Hahn *et al.*[14] and Harbo *et al.*[15] found a significantly increased frequency of the class I HLA-A*03 allele in northern Europe MS patients. The A*03 allele is part of the extended HLA haplotype (hp) characterized by DR15. However, the effect of this association was additive to the risk conferred by DR15 and not a consequence of linkage disequilibrium (LD) between the two alleles. In addition Fogdell-Hahn *et al.*[14] found a protective effect of HLA-A*02. This was confirmed in a larger Swedish

Correspondence: Dr S D'Alfonso, Department of Medical Sciences, Interdisciplinary Research Center of Autoimmune Diseases (IRCAD), University of Eastern Piedmont, Via Solaroli 17, Novara 28100, Italy.
E-mail: dalfonso@med.unipmn.it
Received 2 July 2007; revised 27 August 2007; accepted 31 August 2007; published online 11 October 2007

panel.[16] A protective effect independent of HLA class II was recently detected for HLA-Cw*05 in a large UK and US cohort of simplex families and sporadic cases.[17] Rubio *et al*.[18,19] reported in Tasmanian patients with European ancestry a significant association with D6S265 and MOGCA microsatellites mapping in the HLA class I region 100 kb centromeric and 500 kb telomeric to HLA-A, respectively. In addition they found that having class I and II susceptibility variants on the same hp provides an additive effect on MS risk. In the Sardinian population the D6S1683 microsatellite, located about 980 kb telomeric to HLA-A, was significantly associated[20] with MS independently of class II. Altogether these data point to the presence of at least one MS risk gene, either additive to or independent of class II, in a region of about 1 MB in the HLA class I region.

A very interesting candidate gene mapping in this region is *MOG*, encoding the myelin oligodendrocyte glycoprotein, a quantitatively minor component of myelin which is specific of the central nervous system and is located exclusively on the surface of myelin sheaths and olygodendrocytes.[21] MOG is a potential target antigen of the central nervous system, known to induce an autoreactive T-cell response and pathogenic demyelinating anti-MOG antibodies in MS patients and in EAE (experimental autoimmune encephalomyelitis) mice, the animal model of MS.[22–24]

So far, association studies of MOG markers with MS susceptibility yielded conflicting results. No association was found in 169 French MS patients[25] with three MOG dinucleotide repeats, whereas a significant association was reported for one of these markers (MOGCA) in the above mentioned Tasmanian patients[18,19] and by Barcellos *et al*.[26] Single nucleotide polymorphisms (SNPs) located in the *MOG* 5′ flanking and coding region were not associated with MS susceptibility in 100 German paediatric patients,[27] 82 French adult patients[28] and in 397 Sardinian trio families.[29] Conversely, in the continental Italian population a significant association with the missense variation V142L in the MOG transmembrane region was reported in a small sample of 50 MS patients.[30] Possible explanations of the inconsistency among the different studies might be population differences and/or the small sample sizes.

The aim of the present study was to confirm the Italian finding in a larger independent sample. To this purpose an association study with the V142L SNP was performed by both a case-control and an intrafamilial association approach. We confirmed the result previously reported in the continental Italian population. Consequent steps were to examine whether this association was independent of known MS susceptibility factors in the HLA region and if other sequence variations in the *MOG* gene and in the class I region were more significantly associated with MS.

Results

Association with MOG—V142L polymorphism

The MS association with the V142L MOG missense variation was tested in two independent Italian sample sets consisting of 878 MS patients and 890 controls (sample set 1) and 246 simplex families including the MS patient and both parents (sample set 2). Significantly different case and control allele frequencies were observed in both sample sets ($P = 6.6 \times 10^{-4}$, $P = 1.5 \times 10^{-3}$) and in the combined sample ($P = 6.7 \times 10^{-6}$). In particular, the

Table 1 Allele frequencies of MOG V142L

Allele	*Sample*	*Frequencies*		P-*value*	*OR (95% CI)*
		MS	*Controls*		
	Set 1 (case/control)	$N^a = 1756$	$N^a = 1780$		
V		0.838	0.793		1.35 (1.13–1.61)
L		0.162	0.207		0.74 (0.62–0.88)
				6.6×10^{-4}	
	Set 2 (trio families)	$N^b = 492$	$N^b = 492$		
V		0.884	0.809		1.80 (1.24–2.61)
L		0.116	0.191		0.55 (0.38–0.80)
				1.5×10^{-3}	
	Set 1+2	$N^c = 2248$	$N^c = 2272$		
V		0.848	0.797		1.43 (1.22–1.67)
L		0.152	0.203		0.70 (0.60–0.82)
				6.7×10^{-6}	
	DRB1*15-	$N^d = 1236$	$N^d = 1556$		
V		0.837	0.787		1.39 (1.14–1.70)
L		0.163	0.213		0.72 (0.59–0.88)
				9.7×10^{-4}	

Abbreviations: CI, confidence interval; OR, odds ratio.

[a]Number of chromosomes.

[b]Number of chromosomes transmitted (cases) or non-transmitted (controls) to MS patients, calculated according to the AFBAC (affected family-based controls) method[31] in 246 simplex families.

[c]Sum of the chromosomes of sample sets 1 and 2. Patient chromosomes from set 1 and transmitted chromosomes from set 2 were counted as MS chromosomes and control chromosomes from set 1 and non-transmitted chromosomes from set 2 were counted as control chromosomes.

[d]Number of chromosomes in HLA-DRB1*15-negative individuals among the 890 controls from set 1 and a random subset of 887 MS patients (655 from set 1 and 232 from set 2) typed for HLA-DRB1*15.

Table 2 Genotype frequencies of MOG V142L

Genotype	*Frequencies*		*P-value*	*OR (95% CI)*
	MS (N = 1124)	*Controls (N = 890)*		
V/V	0.727	0.645	2.8×10^{-5}	1.50 (1.24–1.82)
V/L	0.250	0.314	5.9×10^{-4}	0.71 (0.58–0.86)
L/L	0.023	0.041	0.033	0.57 (0.34–0.96)
Overall			6.3×10^{-5}	

Abbreviations: CI, confidence interval; MS, multiple sclerosis; OR, odds ratio.
Genotypes were distributed according to Hardy–Weinberg equilibrium both in patients and controls.

minor allele (L142) was under-represented in the patients (Table 1). Transmission disequilibrium test (TDT) analysis in the 246 families showed a preferential non-transmission of L142 (transmitted:non-transmitted allele 45:82, $P = 0.001$). The comparison of genotype frequencies in all MS patients ($N = 1124$) and in the controls of sample set 1 ($N = 890$) suggested a dominant-protective effect of L142 since both L142 homozygotes and heterozygotes were significantly decreased among the patients (Table 2).

To eliminate the possible confounding effects of LD with HLA-*DRB1*15*, we conditioned the analysis of MOG 142 association on DRB1*15. L142 remained significantly less frequent in DRB1*15-negative patients than in DRB1*15-negative controls (0.163 vs 0.213, $P = 9.7 \times 10^{-4}$; Table 1). This result was in line with the very weak LD between L142 and DRB1*15 detected in this panel ($D' = -0.011$; $r^2 = 0.059$; $P = 0.02$). The effect of L142 and DRB1*15 alleles on MS risk was further evaluated by logistic regression modelling (Table 3). All individuals were categorized according to the DRB1*15 and L142 status (grouped as positive/negative for each of them). The OR conferred by L142 (0.70) remained similar after accounting for DRB1*15 carrier status (0.72). Moreover, the MS risk conferred by DRB1*15 (OR = 2.81) did not change when adjusted for the presence of L142 (OR = 2.77). The interaction between the two markers was also tested in the model: no evidence of interaction was found ($P = 0.40$). These data demonstrate that the involvement of MOG L142 in MS susceptibility was not secondary to the well-known major effect of DRB*15. Moreover, the effect of these two MS-associated alleles was not reciprocally modified.

Although the majority of the DRB1 association with MS is attributable to DRB1*15, other DRB1 alleles are also involved in MS susceptibility.[6,7,17] We therefore considered the relationship between the V142L association and all DRB1 alleles in a random subset of 540 MS cases and 558 controls fully typed for DRB1. A weak LD was detected between MOG V142L and DRB1 (global $D' = 0.19$, Cramer's $V = 0.17$ in the patients; $D' = 0.17$, Cramer's $V = 0.17$ in the controls). V142L was still significantly associated in this panel after conditioning on the DRB1 locus by the COCAPHASE program (unconditioned $P = 0.0031$; conditioned $P = 0.035$). Moreover, both in the total panel and in DRB1*15-negative individuals, the OR conferred by L142 remained similar after accounting for each DRB1 allele separately by logistic regression modelling (data not shown).

Table 3 Logistic regression analysis of MOG L142 and DRB1*15 alleles

Allele	*Crude values*			*Adjusted values*		
	OR	*95% CI*	P	*OR*	*95% CI*	P
L142	0.70	0.57–0.85	<0.0001	0.72	0.59–0.89	0.002
DRB1*15	2.81	2.21–3.57	<0.0001	2.77	2.18–3.52	<0.0001

Abbreviations: CI, confidence interval; OR, odds ratio.
Eight hundred eighty-seven MS patients and 890 controls (see note to Table 1) with complete data for MOG V142L and DRB1*15 were included in the analysis. All individuals were coded as positive or negative for the L142 and DRB1*15 alleles. All values were adjusted for sex.

Having proved that the V142L association is independent of HLA-DR, we then tested its relationship with other polymorphisms in the MOG gene and in the HLA class I region. This analysis was mainly performed in the trio families in order to be able to directly deduce haplotypic combinations from family segregation.

Test of association with other MOG *sequence variations*

The involvement of MOG in MS susceptibility was further tested by analysing the association with 11 SNPs, of which 4 in the 5′ flanking, 3 in the coding and 4 in the 3′ untranslated regions of the gene (Figure 1b). All the selected SNPs were detected directly in MS patients. SNPs in *MOG*-5′ flanking and coding regions were selected from the literature and were detected by extensive sequencing of these regions in Italian[30] and German[27] patients. Since no sequencing scan was available for the 3′-UTR in MS patients, we screened *MOG* 3′-UTR (1195 bp) in 25 patients by DHPLC and identified 4 variations. Three of these had been previously detected in normal individuals and are included in the SNP database (http://www.ncbi.nlm.nih.gov/SNP/). The association was first tested by TDT in the trio families already analysed for position 142. Complete genotyping results for all the tested SNPs were obtained for 199 families. In this random subset of families L142 was still preferentially non-transmitted (T:NT = 31:63; $P = 0.001$). Conversely, none of the other SNPs showed a significant transmission distortion. To increase the power of our analysis we further tested the association of MOG SNPs by a case-control analysis in 290 controls and 343 patients. Combining the two sets (trios and cases/controls) we were able to compare allele frequencies of 1084 case and 978 control chromosomes. Considering a range from 0.05 to 0.20 of the minor allele frequency of the tested SNPs, this sample size has an 80% power to detect a risk factor conferring an OR ranging from 1.35 to 1.6 at an α level of 0.05 (for comparison, the V142 allele in this panel conferred an OR = 1.5). Thus, this study had sufficient power to detect a risk factor with an OR ~ V142L for the majority of tested SNPs with the exception of the rare –1077T/C, –910T/C and Indel L22 variations. None of the tested MOG SNPs, besides MOG V142L, was significantly associated with MS (Table 4). The same result was obtained when including only

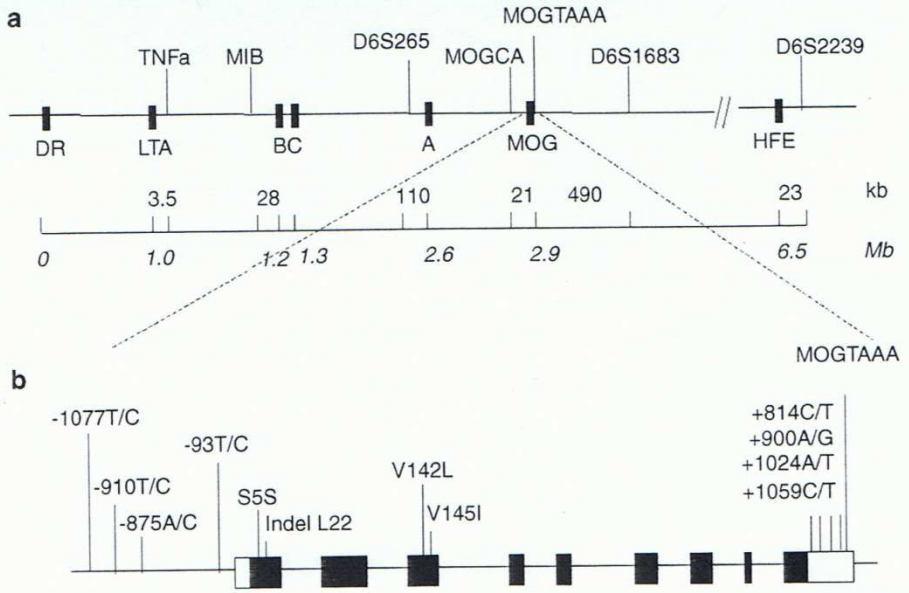


Figure 1 (a) Map of the HLA region. Microsatellites analysed in this study and reference HLA genes are shown. The distance of referenc genes from HLA-DR (Mb scale) and between each microsatellite and the closest reference gene (kb scale) is indicated.[32] (b) Detail of the *MO* gene. Coding and untranslated regions are indicated as black and empty boxes, respectively. The positions of *MOG* intragenic marke analysed in this study are shown.

Table 4 Association analysis of MOG SNPs

Variation	*Accession no.*[a]	*Allele*[b]	*Frequency*[c] *Patients* ($N^d = 1084$)	*Controls* ($N^e = 978$)	P
−1077T/C[f]	—	C	0.015	0.029	NS
−910T/C[f]	rs29235	C	0.041	0.047	NS
−875A/G[f]	rs29234	G	0.056	0.065	NS
−93T/C[f]	rs9468571	C	0.049	0.059	NS
G15A[f] (S5S)	rs3130250	A	0.135	0.134	NS
Indel L22[g]	—	Del	0.038	0.062	NS
V145I	rs3130253	I	0.063	0.057	NS
+814C/T[h]	rs9257936	T	0.151	0.144	NS
+900A/G[h]	rs2535243	G	0.281	0.274	NS
+1024A/T[h]	—	T	0.050	0.072	NS
+1059C/T[h]	rs3135049	T	0.130	0.153	NS
V142L[i]	rs2857766	L	0.131	0.191	$7.9 \times 10^{-}$

Abbreviation: NS, not significant ($P > 0.05$).
[a]Accession number in the National Center for Biotechnology Information SNP data bank (http://www.ncbi.nlm.nih.gov/SNP).
[b]The minor allele is indicated.
[c]Allele frequency.
[d]Sum of the patient chromosomes from the 343 cases of the case/control study and the transmitted chromosomes in 199 simplex familie SNPs−1077T/C and Indel L22 were analysed only in the families.
[e]Sum of the control chromosomes from the 290 controls of the case/control study and the non-transmitted chromosomes in 199 simple families. SNPs−1077T/C and Indel L22 were analysed only in the families.
[f]Numbering is relative to the A nucleotide (+1) of the first ATG corresponding to nucleotide no. 19403 of the GenBank sequence AL050328.2 (version GI: 7530094).
[g]Polymorphism consisting in the presence (+) or deletion (del) of Leucine at codon 22.
[h]SNPs in the 3′-UTR. Numbering is relative to the first (+1) nucleotide of the 3′-UTR corresponding to nucleotide no. 33428 of the GenBan sequence AL050328.24 (version GI: 7530094).
[i]Results obtained in the same subset of families, cases and controls analysed for the other MOG SNPs.

DRB1*15-negative individuals and families (data not shown).

Linkage disequilibria between V142L and the tested MOG SNPs are shown in Table 5. V142L showed a modestly significant LD only with S5S and V145I.

TDT results on the hps including all the analysed SNP are shown in Table 6. Ten hps, accounting for 89% of th total, had a frequency higher than 0.01. The remainin hps showed a frequency $\leqslant 0.01$ both in the patients an in controls. L142 was mainly included in one hp (hp no. 3

Table 5 Linkage disequilibria calculated in the simplex families

Marker	*DRB1 *15*	*D6S265-5*	*MOG CA-6*	*−910*	*−875*	*−93*	*S5S*	*Del22*	***V142L***	*V145I*	*+814*	*+900*	*+1024*
DRB1*15													
D6S265-5	NS												
MOG CA-6	NS	0.54*** (0.27)											
−910	NS	NS	NS										
−875	NS	NS	1* (0.01)	0.90*** (0.63)									
−93	NS	NS	1* (0.01)	0.90*** (0.64)	0.95** (0.88)								
S5S	NS	NS	NS	NS	NS	NS							
Del22	NS	NS	NS	0.63** (0.35)	0.79** (0.60)	0.74*** (0.51)	NS						
V142L	**NS**	**0.53*** (0.25)**	**0.76*** (0.55)**	**NS**	**NS**	**NS**	**0.86* (0.02)**	**NS**					
V145I	NS	1* (0.01)	NS	NS	NS	NS	1*** (0.45)	NS	**1* (0.01)**				
+814	NS	NS	NS	NS	NS	NS	NS	NS	**NS**	NS			
+900	NS	NS	0.47* (0.02)	NS	NS	NS	0.5* (0.02)	NS	**NS**	NS	0.88*** (0.32)		
+1024	NS	NS	1* (0.01)	NS	NS	NS	NS	NS	**NS**	NS	1* (0.01)	1* (0.03)	
+1059	NS	NS	0.91* (0.03)	NS	NS	NS	0.79*** (0.47)	NS	**NS**	0.81*** (0.21)	NS	NS	1* (0.01)

Abbreviation: NS, not significant.
For each pair of loci D' and r^2 (in brackets) values are indicated.
Only the microsatellite alleles with a significant LD with V142L are included.
*$P<0.05$.
**$P<0.01$.
***$P<0.001$; LDs involving V142L are indicated in bold.

which was preferentially non-transmitted ($P=0.018$). With the only exception of position 142, this hp shared the same markers with hp no. 1 which, as opposed to hp no. 3, was significantly over-transmitted ($P=0.002$). When V142L was removed from the analysis, the transmission of this hp combination was no longer distorted.

One further rare hp (no. 7) showed a transmission distortion ($P=0.03$) either including or excluding position 142 in the analysis. This association was no longer significant when considering only DRB1*15-negative families (Table 6). Conversely, hp no. 3 carrying L142 remained significantly under-transmitted also in DRB1*15-negative families ($P=0.007$).

Besides the above 11 SNPs, we tested in the same MS families the association of an *MOG* intragenic microsatellite, MOGTAAA, located in the 3′-UTR (Figure 1). None of the microsatellite alleles or of their hp combinations with the tested SNPs was significantly associated.

In conclusion, none of the tested variations in the MOG gene showed an association with MS⩾V142L. Thus the association with V142L was not secondary to any other tested MOG variation.

Test of association with other markers of the HLA class I region
To test whether the association observed for V142L extended outside the *MOG* gene, we analysed the MS association of six microsatellites (five in the class I region and tumour-necrosis factor α (TNFα); Figure 1a) in the 199 families already tested for all *MOG* SNPs. None of the microsatellite alleles showed a significant transmission distortion. Among those previously reported to be significantly associated,[15,18–20,26] a non-significant trend was observed only for the MOGCA allele 5 (T:NT 72:52, $P=0.073$).

L142 showed a significant LD with the microsatellite alleles MOGCA-6 and D6S265-5 (Table 5). The transmission of the three allele hps (L142, MOGCA-6, D6S265-5) was distorted (T:NT 12:20) but not significantly ($P=0.16$), while a significantly lower transmission (T:NT 13:28, $P=0.019$) was seen for the two allele hps (L142, D6S265-5). When considering only L142-negative chromosomes, no transmission distortion was observed either for D6S265-5 (T:NT 18:17) or for MOGCA-6 (T:NT 2:3).

Discussion

Our results strongly support the presence of an additional MS risk-modulating element in the HLA class I region, independent of DRB1 alleles, and point to MOG V142L as the possible involved factor in the Italian population.

The association of MOG V142L with MS susceptibility was detected by two independent, population- and family-based, association studies, thus confirming a previously reported result in a small sample of 50 Italian MS patients.[30] Notably V142L was the only SNP in the MOG gene significantly associated to MS, in accordance with previous negative results[28,33,34] obtained for MOG markers *not including* V142L. Besides the continental Italian population, the association of V142L was previously tested in one study performed in German paediatric MS patients and in a recent study in the

Table 6 TDT analysis of MOG haplotypes

hp no.	*5′ flanking*			*Coding*				*3′-UTR*				*Freq*[a]	*Total*[b]				*DR15 neg*[c]			
	−910	*−875*	*−93*	*5*	*22*[d]	*142*	*145*	*814*	*900*	*1024*	*1059*		*T*	*NT*	P	% T	*T*	*NT*	P	%
1	T	A	T	G	+	V	V	C	A	A	C	0.335	114	72	0.002	61	53	33	0.028	62
2	T	A	T	G	+	V	V	T	G	A	C	0.116	49	34	NS	59	22	16	NS	60
3	T	A	T	G	+	L	V	C	A	A	C	0.110	27	47	0.018	36	10	26	0.007	28
4	T	A	T	G	+	V	V	C	G	A	C	0.103	36	38	NS	49	14	16	NS	47
5	T	A	T	G	+	V	V	C	A	T	C	0.062	16	28	NS	36	12	12	NS	50
6	T	A	T	A	+	V	V	C	A	A	T	0.049	20	15	NS	57	12	8	NS	60
7	T	A	T	A	+	V	I	C	A	A	T	0.041	21	9	0.029	70	10	4	NS	71
8	T	A	T	G	+	V	V	C	A	A	T	0.036	10	14	NS	42	5	8	NS	38
9	C	G	C	G	Del	V	V	C	A	A	C	0.026	12	9	NS	57	8	4	NS	67
10	T	A	T	G	+	L	V	C	G	A	C	0.016	5	6	NS	45	3	4	NS	43

Abbreviation: NS, not significant ($P > 0.05$).
Only haplotypes with a frequency $\geqslant 0.01$ are displayed. They account for 89% of the total haplotype frequencies. They all carry the T allele a position −1077 (not shown). T = number of transmitted alleles. NT = number of non-transmitted alleles. % T = percentage of transmission c the indicated haplotype.
[a]Haplotype frequencies among non-transmitted haplotypes.
[b]TDT was performed in the same 199 trio families analysed in Table 4.
[c]TDT was performed on informative meioses of the 224 parents of the 112 families where both parents were DRB1*15 negative.
[d]Polymorphism consisting in the presence (+) or deletion (del) of Leucine at codon 22.

Sardinian population. In both studies it was not significantly associated with MS.[27,29] This discrepant result may be related to the different population, to the different age of onset of the disease or to the sample size.

The MOG142 association with MS was not modified when conditioning the analysis on HLA-DRB1 alleles. Thus the protective effect of L142 is independent of HLA-DRB1.

While it is possible that MOG142 is only a marker in LD with a causative sequence variation mapping in an unexplored sequence inside or outside the *MOG* gene, the possibility of a primary involvement of MOG142 is, however, indicated by the following circumstantial evidence:

(1) a V at position 142 is conserved among all tested mammalian species (mouse, rat, cow, macaca, orangutan), suggesting that it has a non-dispensable functional role;
(2) V142L is a conservative substitution located in the first MOG transmembrane domain. There are several instances of diseases in which V/L variations in transmembrane domains (for example in proteins encoded by *CFTR*, *ABC7*, *PSEN1*, *TSHR*, *ACHR*, *VKORC1*) were shown to be causative mutations (references[35–39] and www.genet.sickkids.on.ca/cftr).

What could be the role in MS susceptibility of a conservative V→L substitution in an MOG transmembrane domain?

One possible mechanism could be related to the role of MOG as an autoantigen.[22–24,40,41] Immunization of experimental animals with MOG peptides induces EAE.[22] Interestingly, there is a high similarity between MOG-induced animal models and biopsies from MS patients, both showing large demyelinating lesions and axonal loss.[42,43] MOG has an extracellular part including aa 1–122 with an immunnoglobulin-like domain, a transmembrane part from aa 123 to 152 and an intracellular and a second transmembrane part[21] comprising aa 123–218. Up to now, T-cell responses against the extracellular part of MOG have been mainly looked at. However mor recent data show that the intracellular part of MOG i much more immunogenic than the extracellular part.[44] I particular, there is a dominant MOG epitope recognize by CD4+ T cells within the intracellular part of MO comprising amino acids immediately flanking the 14 residue. The reported experiments were performed onl with the more frequent V142 sequence. As an attractiv hypothesis, the L residue could confer protection for M by decreasing the immunogenicity of this epitop Notably, analysis by computer programs (http:/ www.immuneepitope.org, http://www.syfpeithi.de/ showed that the presence of either V or L at positio 142 changes the affinity score for class I and class II HL molecules.

A second proposed mechanism refers to the possibl influence of transmembrane variations on signal trans mission. In fact, mutations in transmembrane sequence of different plasma membrane receptors alter the signa ling cascade by affecting the correct repartition into th glycosphingolipid-cholesterol membrane microdo mains[45,46] ('lipid rafts'). According to current model upon ligand or antibody cross-linking, different plasm membrane receptors undergo enhanced repartitionin into lipid rafts as an obligatory first step toward participation in early signal-transduction events. Lipi rafts provide a specialized environment for nov molecular interactions that can activate signal-transduc tion pathways. Recent data demonstrated that th signalling cascade activated by antibody cross-linkin of MOG is dependent on MOG repartition into lipi rafts.[47] It is tempting to speculate that the V142 substitution could influence MOG repartition into lipi rafts and hence signal transduction.

Finally, missense variations modifying exonic splicin enhancer (ESE) or silencer sites have been reported a causative mutations of genetic diseases.[48] However, th involvement of *MOG* exon 3 missense SNP in th mechanism seems unlikely since (1) the known alte natively spliced isoforms of human MOG do not involv

exon 3; (2) ESE prediction programs (ESEfinder, http://rulai.cshl.edu/tools/ESE/ and RESCUE ESE, http://genes.mit.edu/burgelab/rescue-ese/) did not predict a different score for candidate ESE motifs for the two MOG142 alleles and (3) no different splicing isoforms were observed in V142L heterozygous individuals when cDNA from T cells was amplified in nested PCR conditions (data not shown).

In conclusion, our data confirm the presence of an MS risk factor, independent of HLA-DR, in the HLA class I region and propose that it might be located in the MOG gene. The relationship between MOG L142 and the other recently identified HLA class I protective markers[14,16,17] (HLA-A2 and –Cw5) remains to be tested. In addition, its possible pathogenetic role should be clarified by functional analysis.

Materials and methods

Subjects

Two independent Italian sample sets were analysed, namely sample set 1 consisting of 878 MS patients (female:male ratio 2:1) and 890 controls (medical students, university and hospital staff, blood donors, female:male ratio 1:1), and sample set 2 consisting of 246 simplex families including the MS patients (female:male ratio 1.75:1) and their parents. Enrolment followed their informed consent. Patients were diagnosed according to McDonald *et al*.[49] All patients had a relapsing-remitting onset and included patients with relapsing-remitting and secondary progressive disease course defined according to Lublin and Reingold.[50] Patients with a primary progressive disease course were excluded as well as patients with Sardinian ancestors to avoid the introduction of confounding sources of heterogeneity.

Search for sequence variations in the MOG *3′-UTR region*

The *MOG* 3′-UTR region (1195 bp) was amplified from genomic DNA of 25 MS patients as 4 overlapping PCR fragments including the whole 3′-UTR with the exception of the most distal 113 bp containing the MOGTAAA microsatellite. PCR primers were designed on the genomic DNA Genbank sequence AL050328.24 (version GI: 7530094).

Search for sequence variations was performed by DHPLC (denaturing high-performance liquid chromatography) on an automated HPLC instrument (Wave, Transgenomic Santa Clara, CA, USA) as previously reported.[51] The temperature required for successful resolution of heteroduplex molecules was determined using a specific program (website http://insertion.stanford.edu/melt.html). Samples were analysed at the predicted temperatures (RTm) and at RTm + 2 °C as recommended by the software authors.[52] Primers used for amplification and DHPLC conditions are available upon request. The PCR products displaying a heteroduplex peak were sequenced and compared with the sequence of a homozygous sample.

Sequencing

Prior to sequencing, unincorporated deoxyribonucleotide triphosphates and primers were removed by 0.5 units of shrimp alkaline phosphatase) and 5 units of Exo (exonuclease) I (both from Amersham, Piscataway, NJ, USA) at 37 °C for 30 min, after which the enzymes were inactivated by incubation at 80 °C for 15 min. Samples were sequenced in both directions on an Applied Biosystem's (Foster City, CA, USA) (ABI) 3100 Genetic Analyzer using the Big-dye terminator cycle sequencing reaction kit (Applied Biosystems).

SNP typing

MOG V142L was typed by primer extension followed by DHPLC analysis.[53] Deletion of leucine at codon 22 (indel L22) was typed by electrophoresis on the ABI 3100 Genetic Analyzer of the PCR fragment obtained with a Fam-conjugated primer.

The remaining SNPs were typed by the SNaPshot approach (ABI). Two SNaPshot reactions were designed, one for the seven SNPs in the promoter and coding region amplified by a multiplex PCR in three amplicons and one for the four SNPs in the 3′-UTR region amplified in one fragment. Samples were electrophoresed on the ABI 3100 Genetic Analyzer. The resulting electropherograms were analysed using the ABI GeneScan 3.7 software.

All primers and conditions are available upon request.

Microsatellite typing

Microsatellite alleles were typed by electrophoresis on the ABI 3100 Genetic Analyzer of PCR fragments obtained with a fluorescence-conjugated primer. Resulting electropherograms were analysed using the ABI GeneScan 3.7 and Genotyper 3.6 NT softwares. PCR primers were as previously reported for TNFα,[54] MIB,[55] MOGCA,[25] D6S1683[56] and D6S2239.[56] Newly designed primers were utilized for D6S265 (forward: 5′-ACGTTCGTACCCATTAACCT-3′; reverse: 5′-TAACTGG AGGTTCTCATATTA-3′) and MOGTAAA (forward: 5′-GTTGCAGTGAGCTGAGATCG-3′; reverse: 5′-TGACCT CTGGGGTAATGAGG-3′).

Numbering of microsatellite alleles was assigned according to the size of the repeat (the smallest allele = 1). On this basis, the microsatellite allele typing of the International Histocompatibility Workshop reference cell line no. 9031 was TNFα-2, MIB-1, D6S265-6, MOGCA-1, MOGTAAA-10, D6S1683-8, D6S2239-3.

HLA typing

MS cases (540) and 558 controls of the case/control set were typed for HLA-DRB1 by the DR low-resolution PCR-SSP (Dynal or BAG, Interlabo diagnostici, Trezzano Rosa, Italy). HLA-DR-typed patients and controls in part overlap with those included in a previous paper.[6]

Additional 332 controls and 115 patients of the case/control set and the members of 232 simplex families (patients and parents) were specifically typed for DRB1*1501. DRB1*1501 typing was performed by allele-specific PCR (with DR15-specific primers: DRB1-15F CCTGTGGCAGCCTAAGAGG and DRB1-15R CCG CGCCTGCTCCAGGAT and an internal control: primers CF TGTTCTGTATTTGTGTTGTCTGATG and CR GTGC TCAGAGAGGCAAGGTT).

Statistical analysis

The association of each polymorphism with the disease was measured by the odds ratio (OR) and its 95% confidence interval (CI). The statistical significance of the difference of gene, genotype and hp frequencies among

cases and controls was evaluated by the χ^2 test with Yates's correction. When required by the small number of expected cases, the two-tailed Fisher's exact test was used. Reported *P*-values were not corrected for the number of comparisons.

The Haploview program (version 3.11)[57] was used to estimate hp structures, compare gene and hp frequencies among cases and controls and for TDT analysis.

TDT for microsatellite markers and for their haplotypic combinations was performed by the Genehunter program (version 2.1).[58]

AFBAC (affected family-based controls) was performed according to Thomson.[31]

Pairwise linkage disequilibria (D' and r^2) were evaluated by the Haploview program (version 3.11).[57] Estimates for Global D' and Cramer's V (measures of LD between multiallelic loci) were calculated using the COCAPHASE program, part of the UNPHASED suite.[59]

The effect of both MOG L142 and DRB1*15 alleles on MS risk was evaluated by binary logistic regression modelling. OR and 95% CI were computed. The goodness-of-fit between two models was tested by the likelihood ratio test. All analyses were adjusted for sex. Multiplicative interaction between the two alleles was also tested adding an interaction term in the model. Analyses were carried out using SAS v.8.

The main-effects test of the COCAPHASE program[59] was used for conditional analysis on DRB1.

Acknowledgements

This work was supported by the Italian Foundation for Multiple Sclerosis (FISM grants 2001/R/44, 2002/R/40 and 2005/R/10); CARIPLO Foundation, Regione Piemonte (grants 2003 and 2004), Ministry of Health (ricerca finalizzata grant 2004.80), Eastern Piedmont University (grants to SD and PMR). NB was supported by a fellowship from FISM (2003/B/2). We thank Dr R Tosi and Dr S Sawcer for their critical suggestions. We are grateful to the patients and their parents.

References

1 Compston A, Coles A. Multiple sclerosis. *Lancet* 2002; **359**: 1221–1231.
2 Compston A, Sawcer S. Genetic analysis of multiple sclerosis. *Curr Neurol Neurosci Rep* 2002; **2**: 259–266.
3 Giordano M, D'Alfonso S, Momigliano-Richiardi P. Genetics of multiple sclerosis: linkage and association studies. *Am J Pharmacogenomics* 2002; **2**: 37–58.
4 Dyment DA, Ebers GC, Sadovnick AD. Genetics of multiple sclerosis. *Lancet Neurol* 2004; **3**: 104–110.
5 Olerup O, Hillert J. HLA class II-associated genetic susceptibility in multiple sclerosis: a critical evaluation. *Tissue Antigens* 1991; **38**: 1–15.
6 Ballerini C, Guerini FR, Rombola G, Rosati E, Massacesi L, Ferrante P *et al*. HLA-multiple sclerosis association in continental Italy and correlation with disease prevalence in Europe. *J Neuroimmunol* 2004; **150**: 178–185.
7 Barcellos LF, Sawcer S, Ramsay PP, Baranzini SE, Thomson G, Briggs F *et al*. Heterogeneity at the HLA-DRB1 locus and risk for multiple sclerosis. *Hum Mol Genet* 2006; **15**: 2813–2824.
8 Marrosu MG, Murru MR, Costa G, Murru R, Muntoni F, Cucca F. DRB1-DQA1-DQB1 loci and multiple sclerosis predisposition in the Sardinian population. *Hum Mol Gen* 1998; **7**: 1235–1237.
9 Saruhan-Direskeneli G, Esin S, Baykan-Kurt B, Ornek Vaughan R, Eraksoy M. HLA-DR and -DQ associations wi multiple sclerosis in Turkey. *Hum Immunol* 1997; **55**: 59–65.
10 Coraddu F, Reyes-Yanez MP, Parra A, Gray J, Smith SI, Tayl CJ *et al*. HLA associations with multiple sclerosis in t Canary Islands. *J Neuroimmunol* 1998; **87**: 130–135.
11 Allcock RJ, de la Concha EG, Fernandez-Arquero M, Vigil Conejero L, Arroyo R *et al*. Susceptibility to multiple scleros mediated by HLA-DRB1 is influenced by a second ge telomeric of the TNF cluster. *Hum Immunol* 1999; **6** 1266–1273.
12 Brassat D, Salemi G, Barcellos LF, McNeill G, Proia Hauser SL *et al*. The HLA locus and multiple sclerosis Sicily. *Neurology* 2005; **64**: 361–363.
13 Masterman T, Ligers A, Olsson T, Andersson M, Olerup Hillert J. HLA-DR15 is associated with lower age at onset multiple sclerosis. *Ann Neurol* 2000; **48**: 211–219.
14 Fogdell-Hahn A, Ligers A, Gronning M, Hillert J, Olerup Multiple sclerosis: a modifying influence of HLA class I gen in an HLA class II associated autoimmune disease. *Tiss Antigens* 2000; **55**: 140–148.
15 Harbo HF, Lie BA, Sawcer S, Celius EG, Dai KZ, Oturai Genes in the HLA class I region may contribute to the HI class II-associated genetic susceptibility to multiple scleros *Tissue Antigens* 2004; **63**: 237–247.
16 Brynedal B, Duvefelt K, Jonasdottir G, Roos IM, Akesson Palmgren J *et al*. HLA-A confers an HLA-DRB1 independe influence on the risk of multiple sclerosis. *PLoS ONE* 2007; e664.
17 Yeo TW, De Jager PL, Gregory SG, Barcellos LF, Walton Goris A *et al*. A second major histocompatibility compl susceptibility locus for multiple sclerosis. *Ann Neurol* 2007; 228–236.
18 Rubio JP, Bahlo M, Butzkueven H, van Der Mei IA, Sale M Dickinson JL *et al*. Genetic dissection of the human leukoc antigen region by use of haplotypes of Tasmanians w multiple sclerosis. *Am J Hum Genet* 2002; **70**: 1125–1137.
19 Rubio JP, Bahlo M, Stankovich J, Burfoot RK, Johnson Huxtable S *et al*. Analysis of extended HLA haplotypes multiple sclerosis and narcolepsy families confirms a pred posing effect for the class I region in Tasmanian MS patier *Immunogenetics* 2007; **59**: 177–186.
20 Marrosu MG, Murru R, Murru MR, Costa G, Zavattari Whalen M *et al*. Dissection of the HLA association w multiple sclerosis in the founder isolated population Sardinia. *Hum Mol Genet* 2001; **10**: 2907–2916.
21 Johns TG, Bernard CC. The structure and function of mye oligodendrocyte glycoprotein. *J Neurochem* 1999; **72**: 1–9.
22 Amor S, Groome N, Linington C, Morris MM, Dornmair Gardinier MV *et al*. Identification of epitopes of mye oligodendrocyte glycoprotein for the induction of experim tal allergic encephalomyelitis in SJL and Biozzi AB/H mi *J Immunol* 1994; **153**: 4349–4356.
23 Bernard CC, Johns TG, Slavin A, Ichikawa M, Ewing Liu J *et al*. Myelin oligodendrocyte glycoprotein: a no candidate autoantigen in multiple sclerosis. *J Mol Med* 19 **75**: 77–88.
24 Iglesias A, Bauer J, Litzenburger T, Schubart A, Linington T- and B-cell responses to myelin oligodendrocyte glycop tein in experimental autoimmune encephalomyelitis a multiple sclerosis. *Glia* 2001; **36**: 220–234.
25 Roth MP, Dolbois L, Borot N, Amadou C, Clanet M, Pontar P *et al*. Three highly polymorphic microsatellites at the hun myelin oligodendrocyte glycoprotein locus, 100 kb telom to HLA-F. Characterization and relation to HLA haplotyp *Hum Immunol* 1995; **43**: 276–282.
26 Barcellos LF, Thomson G, Carrington M, Schafer J, Begov AB, Lin P *et al*. Chromosome 19 single-locus and multilo haplotype associations with multiple sclerosis. Evidence

new susceptibility locus in Caucasian and Chinese patients. *JAMA* 1997; **278**: 1256–1261.
27 Ohlenbusch A, Pohl D, Hanefeld F. Myelin oligodendrocyte gene polymorphisms and childhood multiple sclerosis. *Pediatr Res* 2002; **52**: 175–179.
28 Roth MP, Dolbois L, Borot N, Pontarotti P, Clanet M, Coppin H. Myelin oligodendrocyte glycoprotein (MOG) gene polymorphisms and multiple sclerosis: no evidence of disease association with MOG. *J Neuroimmunol* 1995; **61**: 117–122.
29 Marrosu MG, Murru R, Costa G, Melis MC, Rolesu M, Schirru L *et al*. Variation of the myelin oligodendrocyte glycoprotein gene is not primarily associated with multiple sclerosis in the Sardinian population. *BMC Genet* 2007; **178**: 25.
30 Gomez-Lira M, Moretto G, Bonamini D, Benedetti MD, Pignatti PF, Rizzuto N *et al*. Myelin oligodendrocyte glycoprotein polymorphisms and multiple sclerosis. *J Neuroimmunol* 2002; **133**: 241–243.
31 Thomson G. Mapping disease genes: family-based association studies. *Am J Hum Genet* 1995; **57**: 487–498.
32 Stewart CA, Horton R, Allcock RJ, Ashurst JL, Atrazhev AM, Coggill P *et al*. Complete MHC haplotype sequencing for common disease gene mapping. *Genome Res* 2004; **14**: 1176–1187.
33 Rodriguez D, Della Gaspera B, Zalc B, Hauw JJ, Fontaine B, Edan G *et al*. Identification of a Val 145 Ile substitution in the human myelin oligodendrocyte glycoprotein: lack of association with multiple sclerosis. *Mult Scler* 1997; **3**: 377–381.
34 Lincoln MR, Montpetit A, Cader MZ, Saarela J, Dyment DA, Tiislar M *et al*. A predominant role for the HLA class II region in the association of the MHC region with multiple sclerosis. *Nat Genet* 2005; **37**: 1108–1112.
35 Maguire A, Hellier K, Hammans S, May A. X-linked cerebellar ataxia and sideroblastic anaemia associated with a missense mutation in the ABC7 gene predicting V411L. *Br J Haematol* 2001; **115**: 910–917.
36 Janssen JC, Lantos PL, Fox NC, Harvey RJ, Beck J, Dickinson A. Autopsy-confirmed familial early-onset Alzheimer disease caused by the L153V presenilin 1 mutation. *Arch Neurol* 2001; **58**: 953–958.
37 Esapa CT, Duprez L, Ludgate M, Mustafa MS, Kendall-Taylor P, Vassart G *et al*. A novel thyrotropin receptor mutation in an infant with severe thyrotoxicosis. *Thyroid* 1999; **9**: 1005–1010.
38 Shen XM, Ohno K, Tsujino A, Brengman JM, Gingold M, Sine SM *et al*. Mutation causing severe myasthenia reveals functional asymmetry of AChR signature cystine loops in agonist binding and gating. *J Clin Invest* 2003; **111**: 497–505.
39 Rost S, Fregin A, Ivaskevicius V, Conzelmann E, Hortnagel K, Pelz HJ *et al*. Mutations in VKORC1 cause warfarin resistance and multiple coagulation factor deficiency type 2. *Nature* 2004; **427**: 537–541.
40 Kerlero de Rosbo N, Mendel I, Ben-Nun A. Chronic relapsing experimental autoimmune encephalomyelitis with a delayed onset and an atypical clinical course, induced in PL/J mice by myelin oligodendrocyte glycoprotein (MOG)-derived peptide: preliminary analysis of MOG T cell epitopes. *Eur J Immunol* 1995; **25**: 985–993.
41 Linington C, Bradl M, Lassmann H, Brunner C, Vass K. Augmentation of demyelination in rat acute allergic encephalomyelitis by circulating mouse monoclonal antibodies directed against a myelin/oligodendrocyte glycoprotein. *Am J Pathol* 1988; **130**: 443–454.
42 Storch MK, Stefferl A, Brehm U, Weissert R, Wallstrom E, Kerschensteiner M *et al*. Autoimmunity to myelin oligodendrocyte glycoprotein in rats mimics the spectrum of multiple sclerosis pathology. *Brain Pathol* 1998; **8**: 681–694.
43 Kornek B, Storch MK, Weissert R, Wallstroem E, Stefferl A, Olsson T *et al*. Multiple sclerosis and chronic autoimmune encephalomyelitis: a comparative quantitative study of axonal injury in active, inactive, and remyelinated lesions. *Am J Pathol* 2000; **157**: 267–276.
44 Weissert R, Kuhle J, de Graaf KL, Wienhold W, Herrmann MM, Muller C *et al*. High immunogenicity of intracellular myelin oligodendrocyte glycoprotein epitopes. *J Immunol* 2002; **169**: 548–556.
45 Simons K, Gruenberg J. Jamming the endosomal system: lipid rafts and lysosomal storage diseases. *Trends Cell Biol* 2000; **10**: 459–462.
46 Simons K, Ehehalt R. Cholesterol, lipid rafts, and disease. *J Clin Invest* 2002; **110**: 597–603.
47 Marta CB, Montano MB, Taylor CM, Taylor AL, Bansal R, Pfeiffer SE. Signaling cascades activated upon antibody cross-linking of myelin oligodendrocyte glycoprotein: potential implications for multiple sclerosis. *J Biol Chem* 2005; **280**: 8985–8993.
48 Cartegni L, Chew SL, Krainer AR. Listening to silence and understanding nonsense: exonic mutations that affect splicing. *Nat Rev Genet* 2002; **3**: 285–298.
49 McDonald WI, Compston A, Edan G, Goodkin D, Hartung HP, Lublin FD *et al*. Recommended diagnostic criteria for multiple sclerosis: guidelines from the International Panel on the diagnosis of multiple sclerosis. *Ann Neurol* 2001; **50**: 121–127.
50 Lublin FD, Reingold SC. Defining the clinical course of multiple sclerosis: results of an international survey. *Neurology* 1996; **48**: 907–911.
51 D'Alfonso S, Giordano M, Mellai M, Lanceni M, Barizzone N, Marchini M *et al*. Association tests with systemic lupus erythematosus of IL10 markers indicate a direct involvement of a CA repeat in the 5′ regulatory region. *Genes Immun* 2002; **3**: 454–463.
52 Jones AC, Austin J, Hansen N, Hoogendoorn B, Oefner PJ, Cheadle JP *et al*. Optimal temperature selection for mutation detection by denaturing HPLC and comparison to single-stranded conformation polymorphism and heteroduplex analysis. *Clin Chem* 1999; **45**: 1133–1140.
53 Hoogendoorn B, Owen MJ, Oefner PJ, Williams N, Austin J, O'Donovan MC. Genotyping single nucleotide polymorphisms by primer extension and high performance liquid chromatography. *Hum Genet* 1999; **104**: 89–93.
54 Udalova IA, Nedospasov SA, Webb GC, Chaplin DD, Turetskaya RL. Highly informative typing of the human TNF locus using six adjacent polymorphic markers. *Genomics* 1993; **16**: 180–186.
55 Grimaldi MC, Clayton J, Pontarotti P, Cambon-Thomsen A, Crouau-Roy B *et al*. New highly polymorphic microsatellite marker in linkage disequilibrium with HLA-B. *Hum Immunol* 1996; **51**: 89–94.
56 Foissac A, Salhi M, Cambon-Thomsen A. Microsatellites in the HLA region: 1999 update. *Tissue Antigens* 2000; **55**: 477–509.
57 Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 2005; **21**: 263–265.
58 Kruglyak L, Daly MJ, Reeve-Daly MP, Lander ES. Parametric and nonparametric linkage analysis: a unified multipoint approach. *Am J Hum Genet* 1996; **58**: 1347–1363.
59 Dudbridge F. Pedigree disequilibrium tests for multilocus haplotypes. *Genet Epidemiol* 2003; **25**: 115–121.

Genes and Immunity (2008) 9, 438–444

www.nature.com/gene

ORIGINAL ARTICLE

Variations of the perforin gene in patients with multiple sclerosis

G Cappellano[1,11], E Orilieri[1,11], C Comi[2,3], A Chiocchetti[1], S Bocca[1], E Boggio[1], IS Bernardone[1], A Cometa[4], R Clementi[1], N Barizzone[1], S D'Alfonso[1], L Corrado[1], D Galimberti[5], E Scarpini[5], FR Guerini[6], D Caputo[7], D Paolicelli[8], M Trojano[8], L Figà-Talamanca[9], M Salvetti[9], F Perla[10], M Leone[1,2], F Monaco[1,2] and U Dianzani[1]

[1]*Department of Medical Sciences, Interdisciplinary Research Center of Autoimmune Diseases, A Avogadro University of Eastern Piedmont, Novara, Italy;* [2]*Department of Neurology, A Avogadro University of Eastern Piedmont, Novara, Italy;* [3]*Neurorehabilitation Center, ML Novarese, Moncrivello, Italy;* [4]*Laboratory of Transplant Immunology, Division of Pediatric Hematology/Oncology, IRCCS Policlinico San Matteo, Pavia, Italy;* [5]*Department of Neurological Sciences, University of Milan, IRCCS Fondazione Ospedale Maggiore Policlinico, Milan, Italy;* [6]*Laboratory of Molecular Medicine and Biotechnology, Don C Gnocchi Foundation ONLUS, IRCCS, Milan, Italy;* [7]*Multiple Sclerosis Unit, Don C Gnocchi Foundation ONLUS, IRCCS, Milan, Italy;* [8]*Department of Neurological and Psychiatric Sciences, University of Bari, Bari, Italy;* [9]*Neurological Center of Experimental Therapy, University of Rome, La Sapienza, S Andrea Hospital, Rome, Italy and* [10]*Neurology Division, S Croce e Carle Hospital, Cuneo, Italy*

Perforin is involved in cell-mediated cytotoxicity and mutations of its gene (PRF1) cause familial hemophagocytic lymphohistiocytosis (FLH2). PRF1 sequencing in 190 patients with multiple sclerosis and 268 controls detected two FLH2-associated variations (A91V, N252S) in both groups and six novel mutations (C999T, G1065A, G1428A, A1620G, G719A, C1069T) in patients. All together, carriers of these variations were more frequent in patients than in controls (phenotype frequency: 17 vs 9%, P = 0.0166; odds ratio (OR) = 2.06, 95% confidence interval (CI): 1.13–3.77). Although A91V was the most frequent variation and displayed a trend of association with multiple sclerosis (MS) in the first population of patients and controls (frequency of the 91V allele: 0.076 vs 0.043, P = 0.044), we used it as a marker to confirm PRF1 involvement in MS and assessed its frequency in a second population of 966 patients and 1520 controls. Frequency of the 91V allele was significantly higher in patients than in controls also in the second population (0.075 vs 0.058%, P = 0.019). In the combined cohorts of 1156 patients and 1788 controls, presence of the 91V allele in single or double dose conferred an OR = 1.38 (95% CI = 1.10–1.74). These data suggest that A91V and possibly other perforin variations indicate susceptibility to MS.
Genes and Immunity (2008) **9**, 438–444; doi:10.1038/gene.2008.35; published online 22 May 2008

Keywords: *MS; perforin; autoimmune diseases*

Introduction

Multiple sclerosis (MS) is a chronic inflammatory demyelinating disease of the central nervous system.[1] Its clinical course varies; at onset, approximately 15% of patients display a primary progressive (PP) form, whereas the remainder start out with a relapsing remitting (RR) form and most of them switch to a secondary progressive (SP) form within 10–30 years.[2] Both environmental and genetic factors are involved in the development/progression of MS and several studies point to a complex inheritance involving interactions between combinations of loci that may influence the immune response.[3,4]

Demyelination is obviously a pathological hallmark of MS, but recent evidence has suggested that the clinically relevant cause of functional disability is injury to the axon.[5] This neurodegenerative model posits that demyelination is a permissive factor that creates an environment in which the axon becomes susceptible to injury mediated either by loss of axo–glial trophic interactions or immune-mediated attack of the denuded axon. The cellular effectors responsible for injuring demyelinated axons are currently unidentified. The fact that $CD8^+$ T cells are the most abundant lymphocytes within MS lesions[6] and correlate with axon injury[7] suggests that class I-restricted cytotoxic T cells (CTL) may be the culprit.

Cytolytic granules of $CD8^+$ CTL and natural killer (NK) cells contain perforin and granzymes, and are released on the target cell upon its recognition by the cytotoxic cell. Perforin polymerizes on the target cell membrane and forms pores allowing entry of granzymes

Correspondence: Professor U Dianzani, Department of Medical Sciences, Interdisciplinary Research Center of Autoimmune Diseases, A Avogadro University of Eastern Piedmont, Via Solaroli 17, Novara I-28100, Italy.
E-mail: dianzani@med.unipmn.it
[11]These authors contributed equally to the work.
Received 19 February 2008; revised 3 April 2008; accepted 11 April 2008; published online 22 May 2008

that trigger apoptosis of the target cell by cleaving caspases.[8]

Biallelic loss-of-function mutations of the perforin gene (*PRF1*) have been classically associated with about 30% of cases of familial hemophagocytic lymphohistiocytosis (FLH2), a rare life-threatening immune deficiency that occurs in infants and young adults.[9,10] FLH2 has been classically ascribed to decreased capacity of CTL and NK cells to clear viral infections; viral persistence is thought to cause the lymphoproliferative pattern. FLH2 is a recessive disease and subjects carrying heterozygous *PRF1* mutations are generally healthy. However, some heterozygous variations may favor development of autoimmune diseases. This has been initially suggested for the autoimmune lymphoproliferative syndrome (ALPS), a rare pediatric autoimmune disease due to defective function of the Fas death receptor involved in both downmodulation of the immune response and cell-mediated cytotoxicity.[11] ALPS is primarily due to mutations of the Fas gene or other genes involved in Fas function, but other genetic factors may concur. We have detected two FLH2-associated amino-acid substitutions of *PRF1* that are associated with ALPS, that is, N252S and A91V.[11] A subsequent work on patients with type 1 diabetes mellitus (T1DM) detected association with N252S, but not A91V, and a patient displayed a novel mutation causing a P477A amino-acid change decreasing NK function.[12] This work was aimed to evaluate whether *PRF1* also contributes to MS development in the light of recent studies showing that a region of chromosome 10q22.1, located near *PRF1*, may be a susceptibility locus for MS.[13–15]

Results

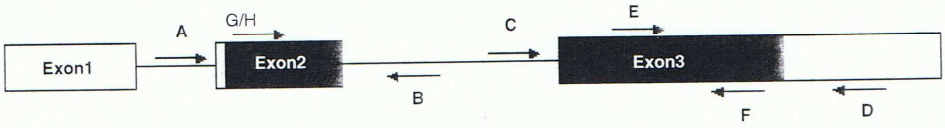
Analysis of the whole coding region of PRF1

The entire coding region of *PRF1* was sequenced in 190 MS and 268 controls to look for variations associated with FLH2 or novel variations (Figure 1). Four missense variations were detected, that is, C272T (rs35947132), A755G (rs28933375), G719A and C1069T (numerations are referred to the GenBank cDNA clone M28393, ATG = +1) causing A91V, N252S, R240H and R357W amino-acid substitutions, respectively. A91V and N252S are variations previously associated with FLH2, whereas R240H and R357W are new.

Four other novel variations, C999T, G1065A, G1428A and A1620G, were detected, but they were synonymous variations (P333P, P355P, G476G and Q540Q, respectively); analysis of their putative effect on splice sites using the Spliceview software and ESEfinder scoring matrix showed that only A1620G (Q540Q) may have an effect by creating a novel acceptor splice site (Spliceview) and a novel binding site for Ser/Arg-rich proteins (ESEfinder), a family of conserved splicing factors.

Finally, we detected the two nucleotide variations, C822T (rs885821) and T900C (rs885822), previously reported as common polymorphisms not associated with FLH2; they did not change the amino acid, nor influence the splicing sites. Their frequency was similar in the patients and the controls. Two other synonymous variations (G435A and A462G) are known to be in perfect linkage disequilibrium with N252S and were in fact only detected in the two subjects (one patient and one control) carrying this variation.[11]

A91V was detected 29 times in 26 patients (23 heterozygotes, 3 homozygotes) and 23 controls (heterozygotes); N252S in 1 patient and 1 control (heterozygotes); R240H in 2 patients (heterozygotes); R357W, P333P, P355P, and G476G in 1 patient each (heterozygotes); and Q540Q in 1 patient (homozygote). The R357W and P355P carriers were also heterozygous for A91V, and the two variations were found to be on different alleles by allele-specific PCR. All together, frequency of the FLH2-associated and novel variations was higher in patients than in controls (allele frequencies: 0.100 vs 0.045, $P = 0.0016$; phenotype frequency: 17 vs 9%, $P = 0.0166$; odds ratio (OR) = 2.06, 95% confidence interval (CI): 1.13–3.77; Table 1).



Nucleotide variation	Predicted amino acid change	Exon	Previous reports
C272T	A91V	2	FLH2-associated
A755G	N252S	3	FLH2-associated
G719A	R240H	3	Novel
C999T	P333P	3	Novel
G1065A	P355P	3	Novel
C1069T	R357W	3	Novel
G1428A	G476G	3	Novel
A1620G	Q540Q	3	Novel

A	5' -cccttccatgtgccctgataatc-3'
B	5' -ccaatcaaacttggaggctgct-3'
C	5' -ccagtcctagttctgcccacttac-3'
D	5' -gtatgcttggactgaaggggttc-3'
E	5' -caggtcaacataggcatccacg-3'
F	5' -gaacagcaggtcgttaatggagg-3'
G	5' -ctccagcgcctgcctccggC-3'
H	5' -ctccagcgcctgcctccggT-3'

Figure 1 Graphical representation (not in scale) of the *PRF1* gene, primers used for typing and the variations found in MS patients. The upper panel shows a scheme of the gene and the relative position of the primers; boxes represent the exons (the coding region is shown in black), lines the introns. Letters and arrows indicate the primers used to amplify and sequence the gene (see Materials and methods) and their sequence is shown in the lower right table. The lower left panel shows a summary of the FLH2-associatcd (*) and novel *PRF1* variations detected in 190 MS patients and 268 controls.

Table 1 Summary of the genotypes of 190 MS patients and 268 controls carrying *PRF1* variations

Allele 1	*Allele 2*	*MS (n = 190)*[a]	*Controls (n = 268)*[a]
A91V	A91V	3	0
A91V	R357W	1	0
A91V	P355P	1	0
Q540Q	Q540Q	1	0
A91V	wt	21	23
N252S	wt	1	1
R240H	wt	2	0
P333P	wt	1	0
G476G	wt	1	0
Total		32 (17%)	24 (9%)
		[b]OR = 2.06, 95% CI: 1.13–3.77; $P = 0.0166$	

Abbreviations: CI, confidence intervals; MS, multiple sclerosis; OR, odds ratio; wt, wild type.
[a]Number of subjects (frequency in the brackets).
[b]OR and 95% CI limits; *P*-values are two-tailed.

The PolyPhen algorithm was used to predict the functional effect of the two novel R240H and R357W missense variations and showed that both may damage the function and structure of the protein (R240H: score = 2.335; R357W: score = 2.690). Therefore, we directly assessed whether R240H affects perforin function by evaluating NK activity in the 2 patients carrying the variation and 15 controls. Results showed that, at low effector/target ratios, NK activity was defective in one patient and in the low level (that is, within the first quartile) of the normal range in the other (Figure 2). This analysis was not performed in the R357W carrier because his cells were not available. Intriguingly, both patients carrying R240H displayed an early switch from the RR to the SP course (5 and 6 years from onset, respectively) and a multiple sclerosis severity score (MSSS) of 7.65 and 7.38, respectively. By contrast, this aggressive clinical evolution was not displayed by the R357W carrier.

Search for the A91V variation in a second population of patients and controls

Although A91V was the most frequent variation and displayed a trend of association with MS in the first population of patients and controls (frequency of the 91V allele: 0.076 vs 0.043, $P = 0.044$), we used it as a marker to confirm *PRF1* involvement in MS, and assessed its frequency in a second independent population of 966 patients and 1520 ethnically and geographically matched controls. The 91V allele was carried by 138 patients (131 heterozygotes and 7 homozygotes) and 168 controls (160 heterozygotes and 8 homozygotes) and its frequency was significantly higher in patients than in controls (0.075 vs 0.058%, $P = 0.019$). In the combined cohorts of 1156 patients and 1788 controls, presence of the 91V allele in single or double dose conferred an OR = 1.38 (95% CI = 1.10–1.74; Table 2).

No differences were found between subjects carrying or not carrying A91V in terms of gender distribution, MS clinical form (RR, PP and SP) and MSSS (data not shown). Moreover, frequency of the MS susceptibility allele HLA-DR15 was not different in patients carrying A91V or not, as HLA-DR15 was carried by 32% of patients with the 91V allele, 29% of patients without it and 12% of the controls.

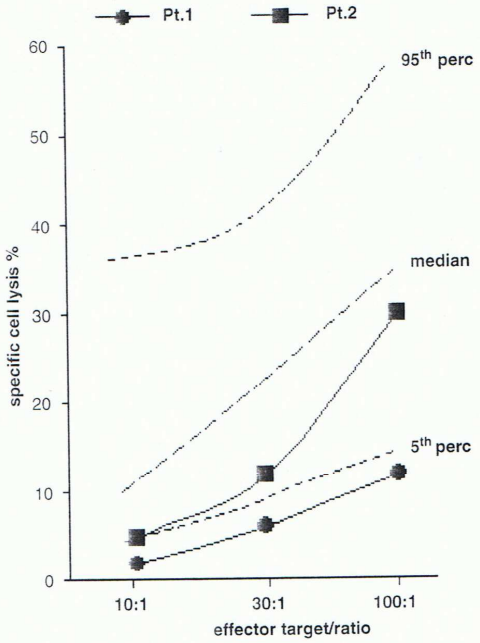


Figure 2 NK activity in PBMC of MS patients carrying the R240H perforin variations and controls. NK activity was assessed at the 100:1, 30:1, and 10:1 effector/target (E:T) ratios; continuous lines indicate patients; stripped and dotted lines indicate the median values and interquartile ranges of 15 controls.

Table 2 Genotype frequencies of A91V in the combined cohorts of MS patients ($n = 1156$) and healthy controls ($n = 1788$)

Genotypes	*MS (n = 1156)*[a]	*Controls (n = 1788)*[a]
AA	992 (0.858)	1597 (0.893)
AV	154 (0.133)	183 (0.102)
VV	10 (0.009)	8 (0.005)
AV+VV vs AA	[b]OR = 1.38, 95% CI = 1.10–1.74; $P = 0.005$	

Abbreviations: CI, confidence intervals; MS, multiple sclerosis; OR, odds ratio.
[a]Number of subjects; frequencies are shown in the brackets. Genotypic distribution did not deviate significantly from the Hardy–Weinberg equilibrium in any group (data not shown).
[b]OR and 95% CI limits; *P*-values are two-tailed.

Discussion

Multiple sclerosis is a complex disease that is probably the result of multiple genetic and environmental factors. Several genes have been involved in its development,[4] and some of them are important in the immune response. This work shows that *PRF1* may also be involved, as MS patients displayed higher frequency of *PRF1* variations

than the controls. This confirms data obtained by other authors showing that the chromosome region 10q22.1, where *PRF1* is located, contains susceptibility genes for MS development.[13–15]

The most frequent variation was A91V, as frequency of the 91V allele was increased in two independent populations of MS patients than in the respective controls, and increased the risk of MS by about 1.4-fold in the combined cohorts. By contrast, A91V did not seem to influence the disease course as MSSS was not different between patients with or without A91V. Studies based on analysis of cytotoxic lymphocytes from A91V carriers or rat basophil leukemia cells transfected with variants of the perforin cDNA have shown that A91V decreases perforin function by altering its conformation, decreasing its cleavage to the active form and increasing its degradation.[16,17] Risma *et al*.[18] classified A91V as a class 1 missense mutation with limited functional impact that allows partial maturation of the protein. Voskoboinik *et al*.[19] have recently used a complementation assay with perforin-knockout primary CTL to show that A91V reduces both the steady-state level of perforin expression in effector cells ('presynaptic' dysfunction) and its intrinsic lytic capacity on target cells, and also displays some dominant-negative effect on the wild-type protein ('postsynaptic' dysfunction).

Our previous work showed that frequency of the 91V allele was also increased in an incomplete variant of ALPS, whereas N252S was associated with the typical form of ALPS and T1DM.[11,12] The functional significance of N252S is debated, but we showed that it may be associated with decreased NK activity in the early childhood.[11,12] Although frequency of N252S was apparently not different in MS and controls, it is possible that *PRF1* variations favor development of several autoimmune diseases, with differences reflecting their effects on perforin function. A91V has also been associated with other immune diseases, such as lymphomas and acute childhood lymphocytic leukemia, and atypical (late-onset) FLH2.[20,21]

Besides A91V and N252S, we detected two new missense *PRF1* mutations in MS patients that cause R240H and R357W amino-acid substitutions. R240H occurs nearby N252S within the membrane-attack complex, a region critically involved in the pore-forming activity of perforin.[22] However, analysis of NK activity in the two R240H carriers showed that it was near the low limit of the normal range. Although both carriers were heterozygous, we suggest that R240H causes a mild decrease of perforin function without exerting a dominant-negative effect on the wild-type form. R357W is located in the same domain, but we could not evaluate its functional effect because fresh cells from the carrier were not available. However, both R357W and R240H were predicted to damage perforin function and structure by *in silico* analysis with the PoliPhen program.

Four other novel mutations were detected in MS patients, but they were synonymous (P333P, P355P, G476G, Q540Q). Q540Q may have an effect on RNA splicing, as it seems to create a new acceptor splice site. The others did not influence canonical splicing sites, but they might theoretically influence perforin expression by disturbing exonic splicing enhancers, mRNA processing and transport, efficiency of codon usage by tRNA stability of mRNA secondary structure, protein folding or interaction with microRNA.[23–28] An alternative possibility is that they do not have a direct effect, but they are linkage disequilibrium with other unknown *PRF1* mutations in the 5′ UTR. However, we could not assess perforin expression because fresh cells from the carriers were not available.

Besides A91V, the other mutations are too rare to draw conclusions about their individual association with MS, but they raise the possibility that the overall effect of *PRF1* variations on MS development may be substantially higher than that detected by A91V alone. In line with this possibility, these variations conferred a global OR = 2.06 for MS development in the first population of patients and controls whose entire *PRF1*-coding region was sequenced. It is intriguing that two patients were compound heterozygous for A91V, and R357W or P355P, respectively, and another patient was homozygous for Q540Q, which raises the possibility that the biallelic variations may have contributed to their MS. The MS association with several rare *PRF1* variations is in line with reports on systemic lupus erythematosus and inflammatory bowel disease indicating that private/rare variations as well as common polymorphisms of other genes may be important in common complex diseases.[29,30]

Perforin-mediated cytotoxicity has been classically associated with clearance of virus-infected cells. Therefore, it is possible that defects of perforin activity favor MS development by delaying virus clearance, which may favor development of crossreactions between viral and self-antigens by molecular mimicry. In this context, it is noteworthy that EBV infections are crucial in FLH2 pathogenesis, and have also been suggested to be important as triggering factors in MS.[31,32]

On the other hand, an increasing bulk of data suggests that perforin and cell-mediated cytotoxicity may also be involved in downmodulation of the immune response. This regulatory activity may involve several mechanisms including perforin-mediated killing of effector lymphocytes and antigen-presenting cells. Defective immune response switching off may favor both lymphocyte accumulation and autoimmunity.[8,33–36] It is noteworthy that involvement of inherited defects of the immune response switching off in MS development may not be limited to *PRF1*, but may also involve defective apoptosis of activated lymphocytes induced through the Fas or the activation-induced cells death (AICD) mechanisms. This possibility is suggested by our previous work showing that substantial proportions of MS patients carry inherited defects of Fas function similar to those displayed by ALPS patients.[37] Moreover, several reports detected high serum levels of osteopontin, a cytokine capable to inhibit AICD, in MS patients and we found that this is partly associated with variants of the osteopontin gene.[38–41]

In conclusion, this work suggests that *PRF1* variations may be a predisposing factor for MS by affecting either the antiviral response or the immune response switching off. Defects of both of these functions may favor development of autoimmunity by prolonging the immune response and increasing the risk of crossreactions between viral and self-antigens. Similar defects may be caused also by alterations of other genes and may be a general predisposing factor for autoimmunity.

Materials and methods

Patients

We analyzed two independent cohorts of Italian patients (391 men, 765 women; M/F: 1/1.96) with MS, diagnosed according to McDonald *et al.*'s criteria[42] and randomly selected ethnically matched healthy controls. The first population was composed of 190 patients and 268 controls, the second by 966 patients and 1520 controls.

Patients were consecutive patients enrolled from the Multiple Sclerosis Centers of the 'Amedeo Avogadro' University of Eastern Piedmont (Novara), the University of Milan, IRCCS Maggiore Policlinico Hospital (Milan), the Don Gnocchi Institute (Milan), the Santa Croce Hospital (Cuneo), the University of Rome 'La Sapienza', S Andrea Hospital (Rome) and the University of Bari (Bari). Their clinical and demographic features were similar to those of other series.[43,44] Controls were consecutive Italian donors obtained from the transfusion services of the respective hospitals. Patients and controls were unrelated, Caucasian and Italian, matched for age and gender, and analyzed as follows:[45]

1. RR: Occurrence of exacerbations, each lasting at least 24 h and separated by at least 1 month of inactivity, with full recovery or sequelae ($n = 852$).
2. PP: Steady worsening of symptoms and signs from onset for at least 6 months, whether superimposed with relapses or not, with occasional plateau and temporary minor improvements ($n = 92$).
3. SP: Initial RR course followed by steady worsening of symptoms and signs for at least 6 months, whether superimposed with relapses or not, with minor remissions and plateaux ($n = 212$).

Progression of disability was assessed with the MSSS.[46] In RR patients, MSSS score was assessed in remission phase.

All patients gave their informed consent according to the Declaration of Helsinki.[47] The research was approved by the Novara ethical committee.

Amplification of PRF1 and mutation detection

Genomic DNA was isolated from peripheral blood mononuclear cells (PBMCs) using standard methods and exons 2 and 3 of the *PRF1*-coding region were amplified in standard PCR conditions. PCR products were purified with the EXO/SAP kit (GE Healthcare, Piscataway, NJ, USA). In the first population, the entire coding region was sequenced with the ABI PRISMR BigDye Terminator kit (Applied Biosystems, Foster City, CA, USA) on an automatic sequencer (Applied Biosystems 3100 Genetic Analyser) according to the manufacturer's instructions. Figure 1 shows primers used for amplification, sequencing and typing. Briefly, exon 2 was amplified with primers A + B (755 bp) and sequenced with the same oligonucleotides. Exon 3 was amplified with C + D (1289 bp) and sequenced with these and with two additional internal primers (E and F). In the second population, the +272 C/T (A91V) variation was typed by sequencing (233 patients and 548 controls) or by the TaqMan 5′-allelic discrimination assay (733 patients and 972 controls; Applied Biosystems). Allelic-specific primers and probes used for discrimination have been previously described.[20] Genotyping of each sample was automatically attributed by the SDS 1.3 software for allelic discrimination. Similar results were obtained in patients typed by the two methods. All variations were confirmed twice by sequencing independent DNA samples. The genotypic distribution of the variation did not deviate significantly from the Hardy–Weinberg equilibrium in any group.

Allele-specific PCR

The wild-type (91A) and mutant (91V) alleles were separately amplified using specific PCR amplification of genomic DNA (forward primer: G or H; reverse primer: D). PCR products were typed for P355P and R357W by sequencing with the ABI PRISMR BigDye Terminator kit on the 3100 Genetic Analyser using the internal primer E.

HLADRB1 typing

Patients and controls were specifically typed for DRB1*1501 allele as previously described.[48]

Cytotoxicity assays

Natural killer activity of PBMC was assessed by a standard 4 h ^{51}Cr-release assay with K562 cells as the target. Results are expressed as specific lysis % calculated as follows: (sample ^{51}Cr release−spontaneous release)/(maximal release−spontaneous release) × 100.

Statistical analysis

Phenotype frequencies were calculated as the number of individuals carrying an allele (either homozygotes or heterozygotes) divided by the total number of individuals.

Allelic and phenotype frequencies were compared with the χ^2-test with the Yates correction. All *P*-values are two-tailed and the significance cutoff was $P < 0.05$. Putative effect of the variation on splicing sites was evaluated using the SpliceView program on the WebGene website (http://www.itb.cnr.it/sun/webgene/) and the ESEfinder scoring matrix (http://www.rulai.cshl.edu/tools/ESE). Putative functional significance of the missense variations was evaluated with the PolyPhen program (http://genetics.bwh.harvard.edu/pph).

Acknowledgements

This work was partially supported by FISM grant 2005/R/10 (Genoa), Telethon grant E1170 (Rome), AIRC (Milan), PRIN Project (MIUR, Rome), Compagnia di San Paolo (Turin), Fondazione Cassa di Risparmio di Cuneo (Cuneo), Regione Piemonte (Ricerca Sanitaria Finalizzata Project and Ricerca Sanitaria Applicata-CIPE Project), Associazione 'Amici di Jean' (Turin), Fondazione Lagrange (Turin) and Centro Dino Ferrari (Milan).

Disclosure

The authors report no conflict of interest.

References

1 Compston A, Coles A. Multiple sclerosis. *Lancet* 2002; **359**: 1221–1231.
2 Weinshenker BG. The natural history of multiple sclerosis. *Neurol Clin* 1995; **13**: 119–146.

3 Sospedra M, Martin R. Immunology of multiple sclerosis. *Annu Rev Immunol* 2005; **23**: 683–747.
4 Steinman L, Martin R, Bernard C, Conlon P, Oksenberg JR. Multiple sclerosis: deeper understanding of its pathogenesis reveals new targets for therapy. *Annu Rev Neurosci* 2002; **25**: 491–505.
5 Howe CL, Rodriguez M. Remyelination as neuroprotection. In: Waxman SG (ed). *Multiple Sclerosis as a Neuronal Disease*. Elsevier Academic Press: San Diego, 2005, pp 389–419.
6 Babbe H, Roers A, Waisman A, Goebels N, Hohlfeld R, Friese M *et al*. Clonal expansions of CD8(+) T cells dominate the T cell infiltrate in active multiple sclerosis lesions as shown by micromanipulation and single cell polymerase chain reaction. *J Exp Med* 2000; **192**: 393–404.
7 Bitsch A, Schuchardt J, Bunkowski S, Kulnann T, Brück W. Acute axonal injury in multiple sclerosis. Correlation with demyelination and inflammation. *Brain* 2000; **123**: 1174–1183.
8 Trapani JA, Smyth MJ. Functional significance of the perforin/granzyme cell death pathway. *Nat Rev Immunol* 2002; **2**: 735–747.
9 Stepp SE, Dufourcq-Lagelouse R, Le Deist F, Bhawan S, Certain S, Mathew PA *et al*. Perforin gene defects in familial hemophagocytic lymphohistiocytosis. *Science* 1999; **286**: 957–959.
10 Clementi R, zur Stadt U, Savoldi G, Varoitto S, Conter V, De Fusco C *et al*. Six novel mutations in the *PRF1* gene in children with haemophagocytic lymphohistiocytosis. *J Med Genet* 2001; **38**: 643–646.
11 Clementi R, Chiocchetti A, Cappellano G, Cerutti E, Ferretti M, Orilieri E *et al*. Variations of the perforin gene in patients with autoimmunity/lymphoproliferation and defective Fas function. *Blood* 2006; **108**: 3079–3084.
12 Orilieri E, Cappellano G, Clementi R, Cometa A, Ferretti M, Cerutti E *et al*. Variations of the perforin gene in patients with type 1 diabetes. *Diabetes* 2008; **57**: 1078–1083.
13 Goertsches R, Villoslada P, Comabella M, Montalban X, Navarro A, de la Concha EG. A genomic screen of Spanish multiple sclerosis patients reveals multiple loci associated with the disease. *J Neuroimmunol* 2003; **143**: 124–128.
14 Goertsches R, Comabella M, Navarro A, Perkal H, Montalban X. Genetic association between polymorphisms in the ADAMTS14 gene and multiple sclerosis. *J Neuroimmunol* 2005; **164**: 140–147.
15 Goertsches R, Baranzini SE, Morcillo C, Nos C, Camiña M, Oksenberg JR *et al*. Evidence for association of chromosome 10 open reading frame (C10orf27) gene polymorphisms and multiple sclerosis. *Mult Scler* 2008; **14**: 412–414.
16 Trambas C, Gallo F, Pende D, Marcenaro S, Moretta L, De Fusco C *et al*. A single amino acid change, A91V, leads to conformational changes that can impair processing to the active form of perforin. *Blood* 2005; **106**: 932–937.
17 Voskoboinik I, Thia MC, Trapani JA. A functional analysis of the putative polymorphisms A91V and N252S and 22 missense perforin mutations associated with familial hemophagocytic lymphohistiocytosis. *Blood* 2005; **105**: 4700–4706.
18 Risma KA, Frayer RW, Filipovich AH, Sumegi J. Aberrant maturation of mutant perforin underlies the clinical diversity of hemophagocytic lymphohistiocytosis. *J Clin Invest* 2006; **116**: 182–192.
19 Voskoboinik I, Sutton VR, Ciccone A, House CM, Chia J, Darcy PK *et al*. Perforin activity and immune homeostasis: the common A91V polymorphism in perforin results in both pre- and post-synaptic defects in function. *Blood* 2007; **110**: 1184–1190.
20 Mehta PA, Davies SM, Kumar A, Devidas M, Lee S, Zamzow T *et al*. Perforin polymorphism A91V and susceptibility to B-precursor childhood acute lymphoblastic leukaemia: a report from the Children's Oncology group. *Leukemia* 2006; **20**: 1539–1541.
21 Clementi R, Locatelli F, Dupré L, Garaventa A, Emmi L, Bregni M *et al*. A proportion of patients with lymphoma may harbor mutations of the perforin gene. *Blood* 2005; **105**: 4424–4428.
22 Voskoboinik I, Smyth MJ, Trapani JA. Perforin-mediated target-cell death and immune homeostasis. *Nat Rev Immunol* 2006; **6**: 940–952.
23 Cartegni L, Chew SL, Krainer AR. Listening to silence and understanding nonsense: exonic mutations that affect splicing. *Nat Rev Genet* 2002; **3**: 285–298.
24 Lavner Y. Kotlar codon bias as a factor in regulating expression via translation rate in the human genome. *Gene* 2005; **345**: 127–138.
25 Kotlar D, Lavner Y. The action of selection on codon bias in the human genome is related to frequency, complexity, and chronology of amino acids. *BMC Genomics* 2006; **7**: 67.
26 Shabalina SA, Ogurtsov AY, Spiridonov NA. A periodic pattern of mRNA secondary structure created by the genetic code. *Nucleic Acids Res* 2006; **34**: 2428–2437.
27 Komar AA. Genetics. SNPs, silent but not invisible. *Science* 2007; **315**: 466–467.
28 Martin MM, Buckenberger JA, Jiang J, Malana GE, Nuovo GJ, Chotani M *et al*. The human angiotensin II type 1 receptor +1166 A/C polymorphism attenuates microRNA-155 binding. *J Biol Chem* 2007; **282**: 24262–24269.
29 Lee-Kirsch MA, Gong M, Chowdhury D, Senenko L, Engel K, Lee YA *et al*. Mutations in the gene encoding the 3′-5′ DNA exonuclease TREX1 are associated with systemic lupus erythematosus. *Nat Genet* 2007; **39**: 1065–1067.
30 Lesage S, Zouali H, Cézard JP, Colombel JF, Belaiche J, Almer S *et al*. CARD15/NOD2 mutational analysis and genotype–phenotype correlation in 612 patients with inflammatory bowel disease. *Am J Hum Genet* 2002; **70**: 845–857.
31 Lunemann JD, Munz C. Epstein–Barr virus and multiple sclerosis. *Curr Neurol* 2007; **7**: 253–258.
32 Chuang HC, Lay JD, Hsieh WC, Su IJ. Pathogenesis and mechanism of disease progression from hemophagocytic lymphohistiocytosis to Epstein–Barr virus-associated T-cell lymphoma: nuclear factor-kappaB pathway as a potential therapeutic target. *Cancer Sci* 2007; **98**: 1281–1287.
33 Su MW, Pyarajan S, Chang JH, Yu CL, Jin YJ, Stierhof YD *et al*. Fratricide of CD8+ cytotoxic T lymphocytes is dependent on cellular activation and perforin-mediated killing. *Eur J Immunol* 2004; **34**: 2459–2470.
34 Badovinac VP, Hamilton SE, Harty JT. Viral infection results in massive CD8+ T cell expansion and mortality in vaccinated perforin-deficient mice. *Immunity* 2003; **18**: 463–474.
35 Zhou S, Ou R, Huang L, Moskophidis D. Critical role for perforin-, Fas/FasL-, and TNFR1-mediated cytotoxic pathways in down-regulation of antigen-specific T cells during persistent viral infection. *J Virol* 2002; **76**: 829–840.
36 de Saint Basile G, Fischer A. Defective cytotoxic granule-mediated cell death pathway impairs T lymphocyte homeostasis. *Curr Opin Rheumatol* 2003; **15**: 436–445.
37 Comi C, Leone M, Bonissoni S, DeFranco S, Bottarel F, Mezzatesta C *et al*. Defective T cell Fas function in patients with multiple sclerosis. *Neurology* 2000; **55**: 921–927.
38 Chiocchetti A, Comi C, Indelicato M, Castelli L, Mesturini R, Bensi T *et al*. Osteopontin gene haplotypes correlate with multiple sclerosis development and progression. *J Neuroimmunol* 2005; **63**: 172–178.
39 Hur EM, Youssef S, Haws ME, Zhang SY, Sobel RA, Steinman L. Osteopontin-induced relapse and progression of autoimmune brain disease through enhanced survival of activated T cells. *Nat Immunol* 2007; **8**: 74–83.
40 Comabella M, Pericot I, Goertsches R, Nos C, Castello M, Blas Navarro J *et al*. Plasma osteopontin levels in multiple sclerosis. *J Neuroimmunol* 2005; **158**: 231–239.
41 Vogt MH, Lopatinskaya L, Smits M, Polman CH, Nagelkerken L. Elevated osteopontin levels in active relapsing-remitting multiple sclerosis. *Ann Neurol* 2003; **53**: 819–822.
42 McDonald WI, Compston A, Edan G, Goodkin D, Hartung HP, Lublin FD *et al*. Recommended diagnostic criteria for multiple sclerosis: guidelines from the International Panel on the diagnosis of multiple sclerosis. *Ann Neurol* 2001; **50**: 121–127.

43 Weinshenker BG, Bass B, Rice GP, Noseworthy J, Carriere W, Baskerville J *et al*. The natural history of multiple sclerosis: a geographically based study: I. Clinical course and disability. *Brain* 1989; **112**: 133–146.
44 Trojano M, Avorio C, Mannari C, Calò A, De Robertis F, Serio G *et al*. Multivariate analysis of predictive factors of multiple sclerosis course with a validated method to assess clinical events. *J Neurol Neurosurg Psychiatry* 1995; **58**: 300–306.
45 Lublin FD, Reingold SC. Defining the clinical course of multiple sclerosis: results of an international survey. *Neurology* 1996; **48**: 907–911.
46 Roxburgh RH, Seaman SR, Masterman T, Hensiek AE, Sawcer SJ, Vukusic S *et al*. Multiple sclerosis severity score: using disability and disease duration to rate disease severity. *Neurology* 2005; **64**: 1144–1151.
47 International Committee of Medical Journal Editors. Protection of patients' rights to privacy. *BMJ* 1995; **311**: 1272.
48 D'Alfonso S, Bolognesi E, Guerini FR, Barizzone N, Bocca S, Ferrante D *et al*. A sequence variation in the MOG gene is involved in multiple sclerosis susceptibility in Italy. *Genes Immun* 2008; **9**: 7–15.

GENE Open

Genes and Immunity (2010) 11, 173–180

www.nature.com/gene

ORIGINAL ARTICLE

HLA-class I markers and multiple sclerosis susceptibility in the Italian population

L Bergamaschi[1], MA Leone[2], ME Fasano[3], FR Guerini[4], D Ferrante[5], E Bolognesi[1], N Barizzone[1], L Corrado[1], P Naldi[2], C Agliardi[4], E Dametto[3], M Salvetti[6], A Visconti[6], D Galimberti[7], E Scarpini[7], M Vercellino[8], R Bergamaschi[9], F Monaco[10,11], D Caputo[12], P Momigliano-Richiardi[1] and S D'Alfonso[1]

[1]*Department of Medical Sciences and Interdisciplinary Research Center of Autoimmune Diseases (IRCAD), University of Eastern Piedmont, Novara, Italy;* [2]*Department of Neurology, Ospedale Maggiore and IRCAD, Novara, Italy;* [3]*Transplantation Immunology, San Giovanni Battista Hospital, Turin, Italy;* [4]*Laboratory of Molecular Medicine and Biotechnology, Don C Gnocchi Foundation IRCCS, S Maria Nascente, Milan, Italy;* [5]*Department of Medical Sciences, Unit of Medical Statistics and Cancer Epidemiology, University of Eastern Piedmont and CPO Piemonte, Novara, Italy;* [6]*Neurology and Center for Experimental Neurological Therapy (CENTERS), Università La Sapienza, Roma, Italy;* [7]*Department of Neurological Sciences, Dino Ferrari Center, University of Milan, Fondazione Ospedale Maggiore Policlinico, Milan, Italy;* [8]*Department of Neurology, Ospedale San Giovanni Battista, Torino, Italy;* [9]*Neurological Institute C Mondino, IRCCS, Pavia, Italy;* [10]*Department of Neurology, Ospedale Maggiore, Novara, Italy;* [11]*Department of Clinical and Experimental Medicine, University of Eastern Piedmont, Novara, Italy and* [12]*Multiple Sclerosis Unit, Don C Gnocchi Foundation, IRCCS, S Maria Nascente, Milan, Italy*

*Previous studies reported an association with multiple sclerosis (MS) of distinct HLA-class I markers, namely HLA-A*02, HLA-Cw*05 and MOG-142L. In this work, we tested the association with MS of A*02 and Cw*05 in 1273 Italian MS patients and 1075 matched controls, which were previously analyzed for MOG-142, and explored the relationship among these three markers in modulating MS risk. HLA-A*02 conferred a statistically robust MS protection (odds ratio, OR = 0.61; 95% confidence intervals, CI = 0.51–0.72, $P < 10^{-9}$), which was independent of DRB1*15 and of any other DRB1* allele and remained similar after accounting for the other two analyzed class I markers. Conversely, the protective effect we previously observed for MOG-142L was secondary to its linkage disequilibrium with A*02. Cw*05 was not associated considering the whole sample, but its presence significantly enhanced the protection in the HLA-A*02-positive group, independently of DRB1: the OR conferred by A*02 in Cw*05-positive individuals (0.22, 95% CI = 0.13–0.38) was significantly lower than in Cw*05-negative individuals (0.69, 95% CI = 0.58–0.83) with a significant ($P = 4.94 \times 10^{-5}$) multiplicative interaction between the two markers. In the absence of A*02, Cw*05 behaved as a risk factor, particularly in combination with DRB1*03 (OR = 3.89, P = 0.0006), indicating that Cw*05 might be a marker of protective or risk haplotypes, respectively.*
Genes and Immunity (2010) **11,** 173–180; doi:10.1038/gene.2009.101; published online 12 November 2009

Keywords: *multiple sclerosis; HLA-class I markers; extended HLA haplotypes; Myelin Oligodendrocyte Glycoprotein; genetic association*

Introduction

Multiple sclerosis (MS) is an autoimmune disease of the central nervous system characterized by disseminated and focal damage of myelin and axons, resulting in a disabling condition.[1] The pathogenesis of MS is not understood yet, but several studies suggest an interaction of environmental and genetic factors.[2] The genetic factor showing the strongest association with MS is localized in the human Major Histocompatibility complex (HLA) region on chromosome 6p21.3. The risk haplotype that has been identified for the Caucasian populations is *HLA DRB1*1501-DQB1*0602* (also known as *DR15* haplotype) in the HLA-class II region. In the Italian[3] as well as in other European populations,[4–6] *DR15* confers an odds ratio (OR) of about 3.

In recent years, several studies have searched for the presence of MS susceptibility factors in the HLA region with an effect independent of *HLA-DRB1*. No evidence of an HLA-A or -B association was found by Chao *et al*.[7] in 294 Canadian families. Moreover, none of the 1068 single-nucleotide-polymorphisms (SNPs) from a high-density panel spanning the entire HLA genomic region showed additional association in 1185 Canadian and Finnish families after conditioning for *HLA-DRB1*.[8] Conversely, other studies have detected the effect at

Correspondence: Dr S D'Alfonso, Department of Medical Sciences and Interdisciplinary Research Center of Autoimmune Disease (IRCAD), University of Eastern Piedmont, Via Solaroli 17, Novara 28100, Italy.
E-mail: dalfonso@med.unipmn.it
Received 13 April 2009; revised 22 September 2009; accepted 23 September 2009; published online 12 November 2009

least of one class I genetic factor independently of *HLA-DRB1*. Brynedal *et al*.[9] confirmed in a Nordic cohort of 1084 MS patients and 1347 controls the results of a previous study performed in a smaller Swedish panel,[10] showing that HLA-A*02 is negatively associated with MS (OR = 0.63, $P = 7 \times 10^{-12}$). Yeo *et al*.[11] analyzed over 1600 UK MS patients and 3600 controls and found that *HLA-Cw*05* exerts an MS-protective effect (OR = 0.49, 95% confidence intervals, CI = 0.34–0.69, $P = 3.3 \times 10^{-5}$) after excluding all individuals carrying *DRB1* alleles associated with MS. In a panel of 1124 Italian MS patients and 1136 controls, we found a significant association with the missense variant *V142L* (rs2857766) in the gene encoding the Myelin Oligodendrocyte Glycoprotein (MOG) mapping in the HLA-class I region, telomeric to *HLA-A* (0.3 Mb) and *HLA-C* (1.6 Mb).[12] The *142L* allele (*MOG-142L)* conferred an OR = 0.70 (95% C.I = 0.60–0.82) that remained similar after accounting for *HLA-DRB1*15* carrier status.[12] Burfoot *et al*.[13] showed similar data in a small Tasmanian population. Moreover, they reported a negative association with *HLA-A*02*, but they did not analyze if this variation was primarily associated or was dependent on *MOG-142L*.[13]

The relationship among the three reported HLA-class I markers, *A*02*, *Cw*05, MOG-142L* and their association with MS has not been examined so far.

In this work, we tested the association of *A*02* and *Cw*05* in a large cohort of Italian MS patients and controls, most of whom had been previously typed for *MOG-142* polymorphism, and we explored the relationship among *MOG-142L*, *HLA-A*02* and *HLA-Cw*05* in modulating MS risk.

Results

*Association with MS of MOG-142L, HLA-A*02 and HLA-Cw*05 in the Italian population*

The MS association of *MOG-142L*, *HLA-A*02* and *HLA-Cw*05* was tested in a case–control study (Table 1). The frequency of *MOG-142L* and *HLA-A*02*-positive individuals was significantly lower in cases than in controls, confirming the previously reported protective effect,[9,10,12,13] whereas the association with *Cw*05* was not significant. The comparison of genotype frequencies suggested a dominant-protective effect of *MOG-142L* and of *HLA-A*02*, as their frequency was significantly decreased both among homozygous (OR = 0.55, 95% CI = 0.34–0.89 for *142L* and OR = 0.55, 95% CI = 0.37–0.82 for *A*02*) and heterozygous (OR = 0.75, 95% CI = 0.62–0.90 for *142L* and OR = 0.68, 95% CI = 0.57–0.81 for *A*02*) MS patients. The results did not change after adjustment for sex and were not different among male and female cases in a stratified analysis (data not shown). The results were not significantly different in relapsing remitting and primary progressive patients (data not shown).

To eliminate the possible confounding effect of linkage disequilibrium (LD) with *HLA-DRB1*15*, we performed the same analysis in *DRB1*15*-negative individuals. The OR values for the three tested markers remained substantially the same (Table 1).

The relative effect of the three tested HLA-class I markers on MS risk was evaluated by logistic regression analysis (Table 2). *DRB1*15* was also included in the model. All individuals were categorized according to the presence or absence of each of the four considered alleles. The OR conferred by *A*02* remained similar after accounting for each of the other markers included in the model and for all the markers together, showing that the protective effect was independent of any other analyzed HLA marker. Conversely, the protective effect of *MOG-142L* was abolished by adjustment for *A*02*. The lack of a significant association with *Cw*05* and the significant risk effect of *DRB1*15* were similar with and without the adjustment for the remaining markers. This analysis clearly showed that the association with *MOG-142L* was not independent of *HLA-A*02*. This was also indicated by an analysis stratified for *HLA-A*02*: *MOG-142L* was not significantly associated either in the *A*02* positive (OR = 0.96 95% CI = 0.74–1.23, $P = 0.76$) or in the *A*02* negative (OR = 0.80, 95% CI = 0.60–1.07, $P = 0.13$). These results were explained by the strong LD ($D' = 0.59$; $r^2 = 0.5$) between the two alleles both in the control and in the MS patient population (Table 3).

Considerable LD was also detected between *A*02* and *Cw*05* in the control population (Table 3). Notably, it was completely absent in the MS patients. This difference was explained by the significantly decreased frequency of the

Table 1 Frequency of *MOG-142L*, *HLA-A*02* and *HLA-Cw*05*-positive individuals among MS patients and controls

Marker	*Sample*	*Frequencies*		P-*value*	*OR (95% CI)*
		Cases	*Controls*		
	Total	N = *1273*	N = *1075*		
MOG-142L		0.285	0.364	4.93×10^{-5}	0.70 (0.59–0.83)
*HLA-A*02*		0.331	0.448	5.28×10^{-9}	0.61 (0.51–0.72)
*HLA-Cw*05*		0.101	0.112	NS	0.85 (0.65–1.10)
	DR15 negative	N = *883*	N = *935*		
MOG-142L		0.293	0.377	1.42×10^{-4}	0.68 (0.56–0.84)
*HLA-A*02*		0.334	0.464	1.44×10^{-8}	0.58 (0.48–0.70)
*HLA-Cw*05*		0.108	0.119	NS	0.89 (0.66–1.21)

Abbreviations: CI, confidence intervals; MS, multiple sclerosis; OR, odds ratio; NS, not significant.
Genotype frequencies (+/+, +/−, −/−) in the total MS patient sample vs total controls were 0.02, 0.26, 0.72 vs 0.04, 0.32, 0.64 for *MOG-142L* and 0.04, 0.29, 0.67 vs 0.07, 0.38, 0.55 for *HLA-A*02*. These frequencies did not deviate from Hardy Weinberg equilibrium either in MS patients or controls.

Table 2 Logistic regression analysis of *MOG-142L*, *HLA-A*02*, *HLA-Cw* 05* and *DRB1*15* alleles

Markers	*OR (95% CI) adjusted values for:*				
	MOG-142L	*HLA-A*02*	*HLA-Cw*05*	*DRB1*15*	*All markers in the model*
MOG-142L	**0.70 (0.59–0.83)**	0.87 (0.71–1.06) $P = 0.1761$	0.70 (0.59–0.83) $P = 6.2 \times 10^{-5}$	0.73 (0.61–0.87) $P = 0.0004$	0.90 (0.73–1.10)
*HLA-A*02*	0.65 (0.53–0.79) $P = 1.0 \times 10^{-5}$	**0.61 (0.51–0.72)**	0.61 (0.52–0.72) $P = 8.2 \times 10^{-9}$	0.63 (0.53–0.75) $P = 1.0 \times 10^{-7}$	0.66 (0.54–0.81)
*HLA-Cw*05*	0.87 (0.67–1.13) $P = 0.2970$	0.90 (0.69–1.17) $P = 0.4167$	**0.85 (0.65–1.10)**	0.88 (0.67–1.14) $P = 0.3348$	0.92 (0.71–1.21)
*DRB1*15*	2.90 (2.34–3.59) $P = 2.8 \times 10^{-24}$	2.89 (2.33–3.58) $P = 6.3 \times 10^{-24}$	2.94 (2.37–3.64) $P = 4.6 \times 10^{-25}$	**2.95 (2.38–3.65)**	2.87 (2.32–3.56)

Abbreviations: CI, confidence intervals; OR, odds ratio.
Each marker in row is adjusted for markers in columns. In the last column the model includes all the markers.
The bolded diagonal values contain the crude values for each marker.
P-values are obtained from likelihood ratio test comparing the likelihood of the two-gene additive vs the single marker model considering as single marker the marker used for adjustment.

Table 3 Linkage disequilibria among the three considered HLA-class I markers and *DRB*15* in the control and in the MS patient population

	Controls				*MS patients*			
	MOG-142L	*A*02*	*Cw*05*	*DRB1*15*	*MOG-142L*	*A*02*	*Cw*05*	*DRB1*15*
MOG-142L	–				–			
*A*02*	0.587*** (0.469)	–			0.588*** (0.510)	–		
*Cw*05*	0.194* (0.091)	0.403*** (0.160)	–		0.002 (0.001)	−0.003 (0.015)	–	
*DRB1*15*	−0.011** (0.074)	−0.014** (0.082)	−0.001 (0.012)	–	−0.006 (0.028)	−0.003 (0.011)	−0.003 (0.031)	–

Abbreviation: MS, multiple sclerosis.
Numbers represent D′ (Lewontin's delta) and r^2 (in brackets) values.
*$P < 0.05$, **$P < 0.005$ and ***$P < 0.000001$.

Table 4 Phenotypic combinations of *HLA-A*02* and *HLA-Cw*05* in MS patients and controls

*A*02*	*Cw*05*	*Sample*	*Frequency*		P-*value*	*OR (95% CI)*
			Cases	*Controls*		
		Total	N = *1273*	N = *1075*		
−	−		0.599	0.513	2.4×10^{-5}	1.42 (1.21–1.68)
+	+		0.031	0.078	4.1×10^{-7}	0.38 (0.26–0.56)
+	−		0.299	0.370	2.8×10^{-4}	0.73 (0.61–0.86)
−	+		0.070	0.039	0.001	1.85 (1.27–2.69)
		DR15 negative	N = *883*	N = *935*		
−	−		0.589	0.498	1.1×10^{-4}	1.44 (1.20–1.73)
+	+		0.031	0.081	1.8×10^{-6}	0.36 (0.23–0.56)
+	−		0.303	0.383	3.6×10^{-4}	0.70 (0.58–0.85)
−	+		0.077	0.038	2.4×10^{-4}	2.15 (1.41–3.26)

Abbreviations: CI, confidence intervals; MS, multiple sclerosis; OR, odds ratio.

*A*02+*, *Cw*05+* phenotypic combination in the MS patient as opposed to the control population (Table 4). When analyzing in more detail the effect of the different phenotypic combinations of the *A*02* and *Cw*05* alleles on MS risk (Table 4), it appeared that the *A*02+*, *Cw*05+* combination conferred an OR (0.38) significantly lower than that (OR = 0.73) consequent to the presence of *A*02* in the absence of *Cw*05*. Conversely, in the absence of *HLA-A*02*, *Cw*05* showed a significant positive association with MS (OR = 1.85). The same results were seen when considering only *DRB1*15*-negative individuals (Table 4).

Stratification for the *HLA-Cw*05* revealed a substantial difference in risk conferred by the *A*02* among

*Cw*05*-positive individuals (OR = 0.22, 95% CI = 0.13–0.38) and *Cw*05*-negative individuals (OR = 0.69, 95% CI = 0.58–0.83). When stratifying for *A*02*, *Cw*05* behaved as a protective factor among *A*02*-positive individuals (OR = 0.50, 95% CI = 0.33–0.74) and as a risk factor among *A*02*-negative individuals (OR = 1.53, 95% CI = 1.04–2.24). The likelihood ratio test used to compare the full model (including the two markers and the interaction term) to the additive model (including only the two markers) was statistically significant ($P = 4.9 \times 10^{-5}$). Thus, although *HLA-Cw*05* and *A*02* are not reciprocally confounders, as there was no difference between *crude* and *adjusted* ORs (Table 2), the stratification analysis showed a significant multiplicative interaction between the two markers.

From the analysis in the case–control population, it was not possible to distinguish whether the *A*02+*, *Cw*05+* individuals carried the two alleles in *cis* or in *trans*. We, therefore, analyzed the transmission of these two alleles in 201 family trios. No Mendelian errors were observed. Notably, out of 10 haplotypes that carried both *A*02* and *Cw*05*, none were transmitted to the MS patients (T:NT = 0:10; $P = 0.002$). The only *A*02+*, *Cw*05+* MS patient present in this panel inherited the two alleles in trans. *A*02*-positive individuals of the family trios were sequenced to define the *A*02* alleles at a higher resolution. In agreement with the literature (http://www.allelefrequencies.net/), in these samples, *A*0201* was the most frequent *A*02* allele (90%) followed by *A*0205* (8%) and *A*0217* (2%). *A*0201* was the most represented *A*02* allele both in the *A*02+*, *Cw*05+* and in the *A*02+*, *Cw*05−* haplotypes. The same proportion of *A*02* subtypes was seen among transmitted and non-transmitted haplotypes (data not shown).

The joint effect of class I (*A*02*, *Cw*05*) and class II (*DRB1*15*) markers on MS risk is reported in Figure 1. All the combinations of these three alleles showed a significantly increased risk relative to the *A*02+*, *Cw*05+*, *DRB1*15−* phenotype, which is characterized both by the presence of the class I markers conferring the highest protection and by the absence of the class II allele conferring the highest risk. The three combinations carrying *DRB1*15* in the absence of the highly protective *A*02+*, *Cw*05+* phenotype (that is carrying only *A*02* or only *Cw*05* or none of them) showed a similar eightfold increased risk. In the presence of both *A*02* and *Cw*05*, the *DRB1*15* risk approximately halved to 4.57, although this difference was not statistically significant

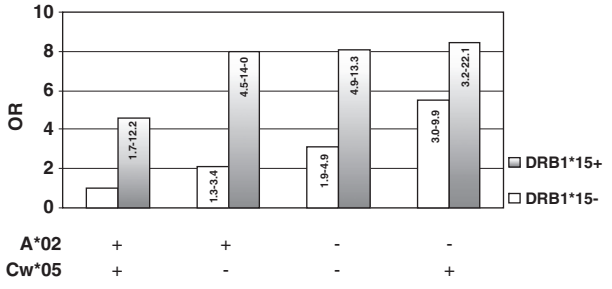


Figure 1 ORs for the different combinations of *A*02*, *Cw*05* and *DRB1*15* alleles. The combination carrying the HLA-class I alleles conferring the highest protection (*A*02+*, *Cw*05+*) and without the class II risk allele (*DRB1*15−*) was used as a reference (OR = 1) for the calculation of ORs. Figures in each bar correspond to the 95% CI.

owing to the low frequency of this phenotype ($N = 13$ MS patients and 8 controls). In the absence of *DRB1*15*, a differently increased risk was observed according to the presence of only *A*02* (OR = 2.11) or only *Cw*05* (OR = 5.47) or none of them (OR = 3.14).

Stratification for DRB1 alleles

To test whether the detected associations were consequent on LD with *DRB1* alleles different from *DRB1*15*, we evaluated the relationship of *HLA-A*02*, *Cw*05* and their combinations with the *DRB1* locus in a random subset of 562 MS cases and 888 controls fully typed for *DRB1* at low resolution. In this subgroup, *DRB1*15* was positively MS associated with OR = 2.86 (95% CI = 2.14–3.83; $P = 8.4 \times 10^{-14}$). When considering *DRB1*15*-negative individuals, there was an additional significantly positive association with *DRB1*04* (OR = 1.88, 95% CI = 1.34–2.63, $P = 1.6 \times 10^{-4}$) and a negative association with *DRB1*07* of borderline significance (OR = 0.69, 95% CI = 0.49–0.98, $P = 0.036$). At variance with other studies, *DRB1*03* was not associated (OR = 1.0, 95% CI = 0.72–1.39). A modest global LD was detected between *HLA-A*02* and *DRB1* (Global $D' = 0.149$, Cramer's $V = 0.176$) and between *HLA-Cw*05* and *DRB1* (Global $D' = 0.280$, Cramer's $V = 0.194$). The association with *A*02* remained significant after conditioning on the *DRB1* locus by the COCAPHASE program (unconditioned $P = 1.9 \times 10^{-5}$; conditioned $P = 2.2 \times 10^{-4}$), thus showing that it is independent of *DRB1*. *Cw*05* was not significantly associated (unconditioned $P = 0.09$; conditioned $P = 0.19$). The OR conferred by *HLA-A*02* did not substantially change after conditioning for each *DRB1* allele separately (data not shown). In addition, *Cw*05* showed very similar results after conditioning for each *DRB1* allele with the notable exception of *DRB1*03*: a protective effect with borderline significance was evidenced for *HLA-Cw*05* when considering only *DRB1*03*-negative individuals (OR = 0.58, 95% CI = 0.36–0.92; $P = 0.02$). When the same procedure was applied to the different *A*02*, *Cw*05* combinations, there was no substantial difference after conditioning for each *DRB1* allele separately, again with the exception of *DRB1*03* (Table 5). Interestingly, the *A*02* negative, *Cw*05* positive combination was significantly increased among *DRB1*03*-positive individuals (OR = 3.89), whereas it was not significantly associated with MS among *DRB1*03*-negative individuals (OR = 0.96). The *A*02* positive, *Cw*05* positive combination remained significantly protective both in *DRB1*03* positive (OR = 0.28) and negative (OR = 0.39) individuals. Thus, the presence of *Cw*05* was significantly protective in combination with *A*02*, independently of *DRB1*, whereas in the absence of *A*02*, it behaved as a risk marker in combination with *DRB1*03* (Table 5).

Discussion

This work stems from previous studies undertaken by our group and others, each reporting an MS association with distinct HLA-class I markers, namely *HLA-A*02*,[9,13] *HLA-Cw*05*[11] and *MOG-142L*.[12] We here analyze the relationship among these three markers for MS susceptibility.

Table 5 Distribution of the phenotypic combinations of *HLA-A*02* and *HLA-Cw*05* in MS patients and controls among *DRB1*03*-positive and *DRB1*03*-negative individuals

	*A*02*	*Cw*05*	*Cases* N *(%)*	*Controls N (%)*	*OR*	*95% CI*	P-*value*
*DRB1*03+*							
	−	+	20 (0.19)	10 (0.06)	3.89	1.74–8.66	0.0006
	−	−	57 (0.54)	82 (0.46)	1.34	0.83–2.18	NS
	+	+	5 (0.05)	26 (0.15)	0.28	0.11–0.77	0.006
	+	−	24 (0.23)	59 (0.33)	0.59	0.33–1.01	NS
*DRB1*03−*							
	−	+	16 (0.04)	26 (0.04)	0.96	0.51–1.81	NS
	−	−	274 (0.60)	360 (0.51)	1.46	1.15–1.86	0.002
	+	+	13 (0.03)	49 (0.07)	0.39	0.21–0.74	0.002
	+	−	153 (0.34)	276 (0.39)	0.79	0.62–1.01	NS

Abbreviations: CI, confidence intervals; MS, multiple sclerosis; OR, odds ratio; NS, not significant.

We confirmed a strong protective effect of *HLA-A*02*, which was independent of *DRB1*15* as well as of any other *DRB1* allele. This association has now been detected, with similar ORs, in three populations with a different genetic background, respectively, from Italy (this paper), Sweden[9] and Tasmania[13] based on a total of over 2600 MS patients and 2700 controls. These data concur to demonstrate a role in MS of this allele (or of a variation in LD with it). Conversely, Chao *et al*.[7] failed to observe a transmission distortion of the *HLA-A*02* allele in a large family sample from Canada and no SNPs in the *HLA-A* region showed association with MS independently of *HLA*-class II loci in Canadian and Finnish families.[8] As both these studies were based on an intrafamilial association method, an obvious explanation of the discrepancy could be a stratification problem related to the case–control approach used in the present as well as the other two studies.[9,13] However, as a significant *HLA-A*02* association was seen in independent studies and in three different populations, this simple explanation seems unlikely and further studies are needed to address this point.

The *HLA-A*02* protection remained similar after accounting for the other two analyzed class I markers. Conversely, the protective effect of *MOG-142L* was secondary to its high linkage disequilibrium with *HLA-A*02*. This excluded a direct role of the missense *V142L* variation in the *MOG* gene, a previously suggested strong MS susceptibility candidate (reviewed in D'Alfonso *et al*.[12]).

In this study, *Cw*05* was not associated with MS either in the whole sample or in *DRB1*15*-negative individuals. However, when stratifying the results separately for all *DRB1* alleles, *Cw*05* conferred a significant protection for MS in the subgroup of individuals negative for *DRB1*03*. This is partially in line with the data reported by Yeo *et al*.[11] who observed a protective effect of *Cw*05* in the whole sample as well as in individuals negative for *DRB1*03* and other MS-associated *DRB1* alleles (*DRB1*15*, *DRB1*0103*). Conversely, in *DRB1*03*-positive individuals, and in the absence of *A*02*, *Cw*05* was a risk marker and conferred a risk similar to *DRB1*15* (OR = 3.89, $P = 0.0006$). Moreover, *Cw*05* significantly enhanced the protection effect of *HLA-A*02*: the OR conferred by *A*02* among *Cw*05*-positive individuals (OR = 0.22, 95% CI = 0.13–0.38) was about 1/3 smaller than among *Cw*05*-negative individuals (OR = 0.69, 95% CI = 0.58–0.83). Thus, although *Cw*05* itself was not significantly associated to MS, it behaved as a modifier of the *HLA-A*02*-mediated protection. The protective effect of the *Cw*05-A*02* combination was independent of the presence of *DRB1*03* (Table 5).

These data suggest that the association of *Cw*05* with MS varies according to its haplotypic context. *Cw*05* is a marker of three ancestral or conserved extended HLA haplotypes,[14–16] one of which carries *DRB1*03* and is negative for *A*02*, namely 18.2 [*A*30*, *Cw*05*, *B*18*, *DRB1*03*], and the other two carry *A*02*, but not *DRB1*03*, namely 44.1 [*A*0201*, *Cw*0501*, *B*4402*, *DRB1*0401*] and 18.3 [*A*0201*, *Cw*0501*, *B*1801*, *DRB1*1102*].

The 18.2 HLA-extended haplotype is typical of the Sardinian population and, to a lesser extent, of other Mediterranean populations. Conversely, in the populations of northern European origin, this DR3 haplotype is very rare and the majority of *DRB1*03*-positive individuals carry the 8.1 [*A1, Cw*07, B*08, DRB1*03*] extended haplotype. The positive association of the *A*02−*, *Cw*05+*, *DRB1*03+* combination observed in this study may reflect the effect of the 18.2 haplotype. It is tempting to speculate that the strong positive association of MS with *DRB1*03* observed in Sardinia is related to the 18.2 haplotype and not only to *DRB1*03* itself.

As stated above, the *A*02-Cw*05* combination is carried by two HLA ancestral haplotypes (44.1 and 18.3). However, as the individuals included in this study were not typed for HLA-B and other HLA markers, from our data it is not possible to conclude whether the enhanced MS-protective effect of the *A*02-Cw*05* combination is due to an haplotype effect (that is the presence of a primarily associated protective factor carried by the extended haplotype marked by *A*02* and *Cw*05*) or to a direct interactive role of the two markers.

The evidence in favor of an haplotype effect is as follows: (i) *Cw*05* was not associated with MS in the absence of *A*02*. Actually, *Cw*05* behaved as a protective factor among *A*02*-positive individuals (OR = 0.50, 95% CI 0.33–0.74) and as a risk factor among *A*02*-negative individuals (OR = 1.53, 95% CI 1.04–2.24). (ii) In family trios, out of 10 haplotypes that carried both *A*02* and *Cw*05*, none were transmitted to the MS patients (T:NT = 0:10; $P = 0.002$). The only *A*02+*, *Cw*05+* MS patient present in this panel inherited the two alleles in trans. (iii) Each of the *A*02*, *B*44* and *Cw*05* alleles

(characterizing the 44.1 ancestral haplotype) was significantly decreased in the study of Yeo *et al*.[11] However, the association with the *A*02-B*44-Cw*05* haplotypic combination was not investigated by the authors.

On the other hand, evidence reported in this study and from work undertaken using an experimental animal model[17] points to a direct role of *A*02*. In our study, A*02 was also significantly protective in the absence of Cw*05 ($OR = 0.73$, $P < 10^{-4}$). Thus, an effect of A*02 alone cannot be excluded. Moreover, Friese *et al*.[17] recently reported that the MS-like disease developed by double transgenic mice expressing both a human HLA-class I allele (A*03) and a human myelin-specific autoreactive T-cell receptor is completely prevented by further adding an *HLA-A*0201* transgene. Thus, A*02, which protects against MS in human populations, also prevents an MS-like disease in transgenic humanized mice. This protection resulted from thymic deletion of autoreactive T cells, which greatly reduced their number in the periphery.[17]

Both *HLA-A* and *-C* genes are interesting candidates for a direct role in MS pathogenesis, as their encoded molecules present antigenic peptides and interact with NK receptors.[18,19] However, several other genes in this region might be primarily associated with MS. The 1.3 Mb interval within HLA-A and -C genes contains 46 genes (24% of which encoding molecules involved in immune functions) and about 4000 SNPs, according to the recently reported sequence of eight HLA ancestral haplotypes[20] and HapMap data (http://www.hapmap.org/). A genome-wide association study[21] in about 1000 US MS patients and 1000 matched controls identified several MS-associated SNPs in the HLA-class I region after conditioning for HLA-DRB1*15. Among these, the top signals were localized around members of the tripartite motif (TRIM) gene family, mapping about 200 kb centromeric to HLA-A. Although their function is unknown, the presence of a RING domain suggests DNA-binding activity.[21]

In conclusion, this study provides additional supportive evidence indicating that the HLA-class I region does indeed exert an additional influence on the risk of MS, analogous to that reported for other autoimmune diseases.[22–24] Moreover, it identifies haplospecific markers conferring a high MS protection. Notably, although in general DRB1*15-negative individuals have an about threefold lower MS risk relative to DRB1*15-positive individuals, their risk was significantly decreased (about eightfold) if they also carried A*02 and Cw*05 (Figure 1). The highly protective A*02-Cw*05 combination is rare (0.078 phenotypic frequency in our controls, 0.025 haplotype frequency in our family trios) and, therefore, unless it tags a primary factor with higher frequency, it confers a modest modification of the HLA attributable MS risk in the population. In any case, the identification of the mechanism mediating the protective effect might throw new light on MS pathogenesis.

Materials and methods

Subjects

A total of 1273 Italian MS patients (female:male ratio 2:1) diagnosed according to McDonald *et al*.,[25] were genotyped for the three HLA-class I markers considered in this study and for DRB1*15. For 201 of these, a DNA sample of the parents was also available (family trios). The mean age of the MS patients at disease onset was 31 ± 10.16 years, the mean age at time of analysis was 40 ± 15.52 years and the mean disease duration was 12 ± 9.31 years. Eighty-four percent of the patients were affected by the relapsing remitting, 7% by the secondary progressive and 9% by the primary progressive form of the disease, defined according to Lublin and Reingold.[26] MS patients with Sardinian ancestors were excluded to avoid the introduction of confounding sources of heterogeneity. Enrolment of the MS patients followed their informed consent. The study was approved by the Ethical Committees of the collaborating clinical centers.

Controls included 1075 Italian individuals (medical students, university and hospital staff, blood donors; female:male ratio 1:1.1) matched for age and regional origin with the MS patients and also typed for all considered markers.

MS patients (82%) and controls (30%) in part overlap with those included in a previous paper[12] and were selected for inclusion on the basis of the availability of DNA and of HLA genotypes.

MOG-V142L typing

The newly included samples were typed using a pre-designed TaqMan SNP Genotyping Assay (probe code: C__25474376_10). Reactions were performed according to the manufacturer's protocol using 25 ng of DNA. The PCR reaction was set up on a 7000 Applied Biosystems instrument. Genotypes were detected using the 7000 System Software (Applied Biosystems, Foster City, CA, USA).

A sub-sample of 60 individuals were typed both with this method and with the method used in the previous paper.[12] The results were consistent for all tested samples.

*HLA-A*02 typing*

HLA-A exon 2 was amplified using a specific couple of primers (Forward: 5′-CGACGCCGCGAGCCAGA**R**GAT-3′, Reverse: 5′-GGCCCGTCCGTGGGGGATGA-3′). The PCR product (213 bp) was digested using the restriction enzyme Kpn2 I. This enzyme recognizes and cuts the rs3173427-T sequence that is specific of *HLA-A*02*. By this approach, it was possible to distinguish *HLA-A*02* homozygotes (displaying two fragments of 159 and 54 bp) from heterozygotes (displaying three fragments of 159, 54 and 213 bp). The digestion was performed at 55 °C for 4 h, and then the enzyme was inactivated at 80 °C for 20 min.

*HLA-A*02*-positive individuals of family trios were also analyzed by sequencing exon 2 and exon 3 to define *HLA-A*02* alleles at a higher resolution. Exon 2 and 3 were amplified together in the same fragment (Forward: 5′-CGACGCCGCGAGCCAGA**R**GAT-3′, Reverse: 5′-AACGGGAAGGAGACGCTGC-3′). The reaction mix was performed using 0.02 U μl^{-1} TaqAB, 0.2 pmol μl^{-1} of every primers, 1.75 mM $MgCl_2$ and glycerol to 7.4%. PCR was made at 60 °C annealing temperature for 35 cycles. PCR products were sequenced using nested primers: 5′-GGCCCGTCCGTGGGGGATGA-3′ for exon 2 and 5′-TCAGTTTAGGCCAAAAATCC-3′ for exon 3. Sequences were analyzed with the automatic sequencer Applied Biosystems (ABI) 3100.

*HLA-Cw*05 allele-specific PCR*
All the samples were typed for *HLA-Cw*05* by an allele-specific PCR after the conditions of the 12th International Histocompatibility Workshop.[27] In detail, specific primers pairs were used to amplify in the same tube HLA-Cw*05 (Forward: 5′-CCGAGTGAACCTGCGGAAA-3′, Reverse: 5′-CGCGCGCTGCAGCGTCTT-3′) and a 796 bp internal control fragment (Forward: 5′-TGCCAAGTGGA GCACCCAA-3′, Reverse: 5′-GCATCTTGCTCTGTGCAG AT-3′). The reaction mix contained 0.02 U μl^{-1} TaqAB (AB Analitica), 1 pmol μl^{-1} of *HLA-Cw*05*-specific primers, 0.33 pmol μl^{-1} of internal control primers and 2 mM of $MgCl_2$. By this approach, it was possible to specifically identify all the samples positive for *HLA-Cw*05*, but not to distinguish between Cw*05 homozygotes and heterozygotes.

DRB1 locus analysis
For 562 MS cases and 888 controls, a complete low-resolution *DRB1* typing was already available. *DRB1* alleles were typed by the DR low-resolution PCR-SSP (Sequence Specific Primer amplification) kit (Dynal or BAG, Formedic, Milan, Italy).

The remaining MS patients and controls were typed only for *DRB1*15* by an allele-specific PCR (Forward primer: 5′-CCTGTGGCAGCCTAAGAGG-3′, Reverse primer: 5′-CCGCGCCTGCTCCAGGAT-3′) with an internal control fragment (Forward: 5′-TGTTCTGTATTTGTGTTG TCTGATG-3′, Reverse: 5′-GTGCTCAGAGAGGCAAGG TT-3′). The reaction mix contained 0.02 U μl^{-1} TaqAB (Applied Biosystems), 0.5 pmol μl^{-1} DRB1*15-specific primers, 0.25 pmol μl^{-1} of internal control primers and 1.5 mM of $MgCl_2$. By this approach, it was possible to specifically identify all the samples positive for DRB1*15, but not to distinguish between DRB1*15 homozygotes and heterozygotes.

Quality control of allele-specific HLA typing
The genotype methods used to type A*02, Cw*05 and DRB1*15 alleles were validated by typing 51 HLA homozygous typing cell lines from the reference panel of the 12th International Histocompatibility Workshop[27] and 55 individuals previously typed with a commercial kit.

Statistical analysis
Unconditional logistical regression was carried out to determine the effect of the considered markers on MS susceptibility. The association of each polymorphism with the disease was measured by the OR and its 95% CI. Reported *P*-values were not corrected for the number of comparisons.

The potential confounding variables were assessed individually by comparing the log-likelihood ratios derived from a model with and without the variable. This analysis was set up using multivariate models using the four considered markers. All analyses were adjusted for sex. The different models were compared by the likelihood ratio test.

The interaction (modification) effect was assessed by comparing ORs across levels of potential modifying variables. Inclusion of appropriate interaction terms in the logistic regression model was used to assess the statistical significance of the interactions. For each marker, the potential effect modification by sex variable was also tested.

The main-effects test of the COCAPHASE program, part of the UNPHASED suite,[28] was used for conditional analysis on DRB1. This program provides association tests conditioning on additional loci, which may already be associated and in linkage disequilibrium with the test loci. The EM algorithm is used to obtain maximium-likelihood estimates of haplotypes.

LD were calculated from phenotypes according to Mattiuz *et al*.[29] Estimates for Global D′ and Cramer's V (measures of LD between multiallelic loci) were calculated using the COCAPHASE program.[28]

Conflict of interest

The authors declare no conflict of interest.

Acknowledgements

This work was supported by the Italian Foundation for Multiple Sclerosis (FISM grants 2001/R/44, 2002/R/40 and 2005/R/10, 2008/R/11); CARIPLO Foundation, Regione Piemonte Ricerca Sanitaria Finalizzata (grants 2003, 2004, 2007, 2008), Regione Piemonte CIPE 2004, Ministry of Health (ricerca finalizzata grant 2004.80), Eastern Piedmont University (grants to SD and PMR). Compagnia di San Paolo (Turin), Fondazione CRT (Turin). DG and ES were supported by Fondazione Monzino, Ospedale Maggiore Policlinico and Ing. Cesare Cusan. NB was supported by a fellowship from FISM (2003/B/2). LB is a 'Biotechnology for Health' PhD student supported by a PhD Lagrange Fellowship. We are grateful to the patients and their parents and also to Professor Corrado Magnani for helpful discussion and to Dr Roberto Tosi for critical suggestions.

References

1 Compston A, Coles A. Multiple sclerosis. *Lancet* 2008; **372**: 1502–1517.
2 Ebers GC. Environmental factors and multiple sclerosis. *Lancet Neurol* 2008; **7**: 268–277.
3 Ballerini C, Guerini FR, Rombolà G, Rosati E, Massacesi L, Ferrante P *et al*. HLA-multiple sclerosis association in continental Italy and correlation with disease prevalence in Europe. *J Neuroimmunol* 2004; **150**: 178–185.
4 Fogdell A, Hillert J, Sachs C, Olerup O. The multiple sclerosis- and narcolepsy-associated HLA class II Haplotype incluse DRB5*0101 allele. *Tissue Antigens* 1995; **46**: 333–336.
5 Olerup O, Hillert J. HLA class II-associated genetic susceptibility in multiple sclerosis: a critical evaluation. *Tissue Antigens* 1991; **38**: 1–15.
6 Barcellos LF, Sawcer S, Ramsay PP, Baranzini SE, Thomson G, Briggs F *et al*. Heterogeneity at the HLA-DRB1 locus and risk for multiple sclerosis. *Hum Mol Genet* 2006; **15**: 2813–2824.
7 Chao MJ, Barnardo MC, Lui GZ, Lincoln MR, Ramagopalan SV, Herrera BM *et al*. Transmission of class I/II multi-locus MHC haplotypes and multiple sclerosis susceptibility: accounting for linkage disequilibrium. *Hum Mol Genet* 2007; **16**: 1951–1958.
8 Lincoln MR, Montpetit A, Cader MZ, Saarela J, Dyment DA, Tiislar M *et al*. A predominant role for the HLA class II region

in the association of the MHC region with multiple sclerosis. *Nat Genet* 2005; **37**: 1108–1112.
9 Brynedal B, Duvefelt K, Jonasdottir G, Roos IM, Akesson E, Palmgren J *et al.* HLA-A confers an HLA-DRB1 independent influence on the risk of multiple sclerosis. *PLoS One* 2007; **7**: e664.
10 Fogdell-Hahn A, Ligers A, Gronning M, Hillert J, Olerup O. Multiple sclerosis: a modifying influence of HLA class I genes in an HLA class II associated autoimmune disease. *Tissue Antigens* 2000; **55**: 140–148.
11 Yeo TW, De Jager PL, Gregory SG, Barcellos LF, Walton A, Goris A *et al.* A second major histocompatibility complex susceptibility locus for multiple sclerosis. *Ann Neurol* 2007; **61**: 228–236.
12 D'Alfonso S, Bolognesi E, Guerini FR, Barizzone N, Bocca S, Ferrante D *et al.* A sequence variation in the MOG gene is involved in multiple sclerosis susceptibility in Italy. *Genes Immun* 2008; **9**: 7–15.
13 Burfoot RK, Jensen CJ, Field J, Stankovich J, Varney MD, Johnson LJ *et al.* SNP mapping and candidate gene sequencing in the class I region of the HLA complex: searching for multiple sclerosis susceptibility genes in Tasmanians. *Tissue Antigens* 2008; **71**: 42–50.
14 Cattley SK, Williamson JF, Tay GK, Martinez OP, Gaudieri S, Dawkins RL. Further characterization of MHC haplotypes demonstrates conservation telomeric of HLA-A: update of the 4AOH and 10IHW cell panels. *Eur J Immunogenet* 2000; **27**: 397–426.
15 Alper CA, Larsen CE, Dubey DP, Awdeh ZL, Fici DA, Yunis EJ. The haplotype structure of the human major histocompatibility complex. *Hum Immunol* 2006; **67**: 73–84.
16 Dorak MT, Shao W, Machulla HK, Lobashevsky ES, Tang J, Park MH *et al.* Conserved extended haplotypes of the major histocompatibility complex: further characterization. *Genes Immun* 2006; **7**: 450–467.
17 Friese MA, Jakobsen KB, Friis L, Etzensperger R, Craner MJ, McMahon RM *et al.* Opposing effects of HLA class I molecules in tuning autoreactive CD8+ T cells in multiple sclerosis. *Nat Med* 2008; **14**: 1227–1235.
18 Moretta L, Moretta A. Killer immunoglobulin-like receptors. *Curr Opin Immunol* 2004; **16**: 626–633.
19 Thananchai H, Gillespie G, Martin MP, Bashirova A, Yawata N, Yawata M *et al.* Cutting edge: allele-specific and peptide-dependent interactions between KIR3DL1 and HLA-A and HLA-B. *J Immunol* 2007; **178**: 33–37.
20 Horton R, Gibson R, Coggill P, Miretti M, Allcock RJ, Almeida J *et al.* Variation analysis and gene annotation of eight MHC haplotypes: the MHC Haplotype Project. *Immunogenetics* 2008; **60**: 1–18.
21 Baranzini SE, Wang J, Gibson RA, Galwey N, Naegelin Y, Barkhof F *et al* Genome-wide association analysis of susceptibility and clinical phenotype in multiple sclerosis. *Hum Mol Genet* 2009; **18**: 767–778.
22 Howson JM, Walker NM, Clayton D, Todd JA, Diabetes Genetics Consortium. Confirmation of HLA class II independent type 1 diabetes associations in the major histocompatibility complex including HLA-B and HLA-A. *Diabetes Obes Metab* 2009; **11**(Suppl 1): 31–45.
23 Eike MC, Becker T, Humphreys K, Olsson M, Lie BA. Conditional analyses on the T1DGC MHC dataset: novel associations with type 1 diabetes around HLA-G and confirmation of HLA-B. *Genes Immun* 2009; **10**: 56–67.
24 Bolognesi E, Karell K, Percopo S, Coto I, Greco L, Mantovani V *et al.* Additional factor in some HLA DR3/DQ2 haplotypes confers a fourfold increased genetic risk of celiac disease. *Tissue Antigens* 2003; **61**: 308–316.
25 McDonald WI, Compston A, Edan G, Goodkin D, Hartung HP, Lublin FD *et al.* Recommended diagnostic criteria for multiple sclerosis: guidelines from the International Panel on the diagnosis of multiple sclerosis. *Ann Neurol* 2001; **50**: 121–127.
26 Lublin FD, Reingold SC. Defining the clinical course of multiple sclerosis: results of an international survey. *Neurology* 1996; **48**: 907–911.
27 Tonks S, Marsh SGE, Bunce M, Moses JH, Krausa P, Sadler AM *et al.* HLA class I DNA typing study. In: D Charron (ed). *Genetic Diversity of HLA: Functional and Medical Implications*. EDK: Sevres, 1997, pp 199–215.
28 Dudbridge F. Pedigree disequilibrium tests for multilocus haplotypes. *Genet Epidemiol* 2003; **25**: 115–121.
29 Mattiuz PL, Ihde D, Piazza A, Ceppellini R, Bodmer WD. New approaches to the population genetics and segregation analysis of the HLA system. In: P Terasaki (ed). *Histocompatibility Testing*. Munksgaard: Copenhagen, 1970, pp 193–205.

INTERNATIONAL JOURNAL OF IMMUNOPATHOLOGY AND PHARMACOLOGY Vol. 23, no. 1, 263-269 (2010)

THE OSTEOPONTIN GENE +1239A/C SINGLE NUCLEOTIDE POLYMORPHISM IS ASSOCIATED WITH TYPE 1 DIABETES MELLITUS IN THE ITALIAN POPULATION

A. CHIOCCHETTI, E. ORILIERI, G. CAPPELLANO, N. BARIZZONE, S. D'ALFONSO, G. D'ANNUNZIO[1], R. LORINI[1], R. RAVAZZOLO[1], F. CADARIO, M. MARTINETTI[2], V. CALCATERRA[3], F. CERUTTI[4], G. BRUNO[5], D. LARIZZA[3] and U. DIANZANI

Interdisciplinary Research Center of Autoimmune Diseases and Department of Medical Sciences, "A. Avogadro" University of Eastern Piedmont, Novara; [1]Department of Pediatric Sciences and Laboratory of Molecular Genetics, Gaslini Institute-IRCCS, University of Genoa, Genoa; [2]Immunohaematology and Transfusion Center Fondazione IRCCS Policlinico San Matteo, Pavia; [3]Department of Pediatric Sciences University of Pavia, Fondazione IRCCS Policlinico San Matteo, Pavia; [4]Department of Pediatrics and [5]Department of Internal Medicine, University of Turin, Turin, Italy

Received June 20, 2009- Accepted November 27, 2009

The first two authors contributed equally to this work

Secreted phosphoprotein 1, also known as Osteopontin (Opn), is a proinflammatory cytokine involved in the TH1 response and is highly expressed in the islets and pancreatic lymph nodes of non-obese diabetic mice before the onset of diabetes. In humans, typing of the +1239A/C single nucleotide polymorphism (SNP) in the 3'UTR of the Opn gene (SPP1) showed that +1239C carriers displayed higher Opn serum levels than +1239A homozygotes and a higher risk of developing autoimmune/lymphoproliferative syndrome, multiple sclerosis, and systemic lupus erythematosus. The aim of this work is to evaluate whether +1239A/C is also associated with type 1 diabetes mellitus (T1DM). We typed +1239A/C in an initial cohort of 184 T1DM patients and 361 controls, and confirmed our data in a second cohort of 513 patients and 857 controls. In both cohorts, +1239C carriers displayed a significantly higher risk of T1DM than +1239A homozygotes (combined cohorts: OR=1.63, 95%CI: 1.34-1.97). Clinical analysis did not detect any differences between patients carrying or not +1239C in terms of gender distribution and age at T1DM diagnosis. These data suggest that SPP1 variants marked by +1239C are associated with T1DM development in the Italian population. The predisposing effect may depend on its effect on Opn levels.

Type 1 diabetes (T1DM) is the result of selective destruction of the insulin-producing beta-cells by autoimmune aggression, mainly driven by pro-inflammatory type 1 T helper cells (TH1) and cytotoxic T lymphocytes (CTL) (1-2). Its development seems to involve environmental factors acting on a predisposing genetic background (3-4). The most important gene involved in T1DM

Key words: type 1 diabetes mellitus, Osteopontin, single nucleotide polymorphism, genetic susceptibility

Mailing address: Prof. Umberto Dianzani,
IRCAD
Department of Medical Sciences,
Via Solaroli, 17,
I-28100, Novara, Italy
Tel: ++39 0321660644 Fax: ++39 0321620421
e-mail: umberto.dianzani@med.unipmn.it

0394-6320 (2010)

susceptibility maps in the human MHC locus has a strong association with the HLA-DR3 and/or -DR4 antigens (5). However, non-MHC loci are also involved and may account for more than 50% of the genetic susceptibility (6-8).

Secreted phosphoprotein 1, also known as Osteopontin (Opn), is a cytokine expressed by many immune cells, including macrophages, T cells and NK cells. It is involved in the inflammatory and immune responses by activating macrophages and enhancing the TH1 cell response (9-13). We previously showed high Opn serum levels in patients with autoimmune lymphoproliferative syndrome (ALPS), multiple sclerosis (MS), and systemic lupus erythematosus (SLE) (14-16). Sequencing of the Opn gene (SPP1), mapping at 4q21-q25, detected two single nucleotide polymorphisms (SNPs) displaying associations with these diseases, namely +1239A/C (rs9138) in the 3'UTR and –156G/GG (rs7687316) in the 5' flanking region. Carriers of +1239C displayed higher risk of developing ALPS (8-fold higher), MS (1.5-fold), and SLE (1.5-fold) than +1239A homozygotes (14-16). Moreover, +1239C mRNA was more stable than +1239A mRNA, and healthy +1239C carriers displayed about 2-fold higher serum levels of Opn than +1239A homozygotes. The -156G allele was associated with SLE only (16).

An Opn role in T1DM has been suggested by detection of a striking up-regulation of Opn in the early phases preceding the diabetes onset in the islets and pancreatic lymph nodes of non-obese diabetic (NOD) mice. In humans, some of us recently typed -156G/GG and a second SNP in the SPP1 promoter (-66T/G) (rs28357094) in Italian T1DM patients and detected a significant association with the -66G allele (17), whereas no association was found with -156G/GG. Since that study did not analyze the 3'UTR region, the aim of this work is to extend the analysis of Italian T1DM patients to this region by typing +1239A/C.

MATERIALS AND METHODS

Patients and randomly selected, ethnically matched, healthy controls were enrolled from the Diabetes Centers of five main hospitals in north western Italy: the Maggiore Hospital of Novara (Novara, Italy), the San Giovanni Battista Hospital of Turin and the Regina Margherita Children Hospital of Turin (Turin, Italy), IRCCS Policlinico S. Matteo (Pavia, Italy), and Giannina Gaslini Children's Hospital (Genoa, Italy). Patients were consecutive Caucasian Italian patients followed by the Diabetes Centers of Novara (n=97), Turin (n=358), Pavia (n=69), and Genoa (n=173); controls were consecutive Italian donors obtained from the transfusion services of the respective hospitals. Patients and controls were unrelated and Italian. Overlaps between different sites were ruled out.

We typed the +1239A/C SNP in 2 independent cohorts of consecutive patients and controls. The first cohort consisted of 184 patients and 361 controls, the second of 513 patients and 857 controls. All subjects gave their informed consent according to the Declaration of Helsinki (International Committee of Medical Journal Editors, 1995). The research was approved by the local ethics committee.

DNA analysis

Genomic DNA was isolated from PBMC using standard methods. Primers used to sequence the 3'UTR were Opn-for (5'-GCCGTGAATTCCACAGCC-3') and Opn-rev (5'-GCTCTACACCACCAAATTCTT ATTACATTCAAG-3'). For amplification, 200 ng of genomic DNA were used as template and PCR products were purified with the EXO/SAP kit (GE Healthcare, Piscataway, NJ, USA). Sequencing was performed with the ABI PRISMR BigDyeTM Terminator kit (Applied Biosystems, Foster City, CA, USA) on an automatic sequencer (Applied Biosystems 3100 Genetic Analyser) according to the manufacturer's instructions with the amplification primers. In the second cohort, genotyping of +1239A/C SNP was performed with the TaqMan 5' allelic discrimination assay (Applied Biosystems). Allelic specific primers and fluorogenic probes were used for discrimination (ASSAY ID C_8826997, Cod.4351376, Applied Biosystems). Genotyping of each sample was automatically attributed by the SDS 1.3 software for allelic discrimination.

HLA-DQA1 and DQB1 genomic typing

HLA-DQA1 and DQB1 genes were typed at high resolution level using the reverse PCR-SSO technique, as previously described (18), in 92 patients (39 AA subjects and 53 non-AA subjects) and 677 healthy children recruited from the cord blood bank of the Pavia center. The controls were ethnically matched with the patients and checked for absence of diabetes in their families.

Statistical analysis

Genotype distributions were analyzed with the chi square (χ^2) test. All p values are 2-tailed and the significance cut-off was $p<0.05$.

RESULTS

We typed +1239A/C in an initial cohort of 184 T1DM patients and 361 ethnically- and geographically-matched controls by sequencing genomic DNA (Fig. 1A). Results showed that frequency of +1239C was significantly higher in T1DM patients than in controls (39% *vs* 29%, p=0.0005); analysis of the genotypic distribution showed that +1239C carriers (heterozygotes and homozygotes), displayed a significantly higher risk of T1DM than +1239A homozygotes (OR=1.90, CI 1.30-2.78, p=0.00074) (Table I).

To confirm these results, +1239A/C was typed in a second cohort of 513 patients and 857 controls by TaqMan 5' allelic discrimination assay (Fig. 1B). Here, too, frequency of +1239C was significantly higher in patients than in controls (34% *vs* 28%, p=0.00035), and +1239C carriers displayed higher risk of T1DM than +1239A homozygotes (OR=1.54, 1.23-1.94, p=0.00015) (Table I). Genotypic distribution did not deviate significantly from the Hardy-Weinberg equilibrium in any group.

Clinical analysis did not detect any differences between T1DM patients carrying or not +1239C in terms of gender distribution (+1239A homozygotes: Males/Females=163/116; +1239C carriers: 233/185) and age at T1DM diagnosis (+1239A homozygotes: median age=10 years, interquartile range: 6-14; +1239 carriers: median age=12, interquartile range: 7-16).

Lastly, we evaluated the frequencies of DQαβ diabetogenic heterodimers in 92 patients carrying or not +1239C, and 677 healthy controls. In this

Table I. *Frequency distribution of SPP1 +1239 A/C alleles and genotypes in 2 independent cohorts of type 1 diabetes patients and healthy controls.*

Cohort	Subject group	Allele*		p†	Genotype‡			p§
		A	C		AA	AC	CC	
Cohort 1	Controls N=361‡	514 (71%)	208 (29%)	0.00053	186 (52%)	142 (39%)	33 (9%)	0.0017
	Patients N=184‡	223 (61%)	145 (39%)		66 (36%)	91 (49%)	27 (15%)	
					OR‖: 1.90 (1.30-2.78) p=0.00074			
Cohort 2	Controls N=857‡	1240 (72%)	474 (28%)	0.00035	448 (52%)	344 (40%)	65 (8%)	0.00055
	Patients N=513‡	675 (66%)	351 (34%)		213 (41%)	249 (49%)	51 (10%)	
					OR‖: 1.54(1.23-1.94) p=0.00015			
Total Cohort	Controls N=1218‡	1754 (72%)	682 (28%)	0.0000012	634 (52%)	486 (40%)	98 (8%)	0.0000019
	Patients N=697‡	898 (64%)	496 (36%)		279 (40%)	340 (49%)	78 (11%)	
					OR‖: 1.63 (1.34-1.97) p=0.0000005			

**: number of chromosomes; proportions are shown in the brackets; †: p value for allele frequencies; ‡: number of subjects; proportions are shown in the brackets. Genotypic distribution did not deviate significantly from the Hardy-Weinberg equilibrium in any group (data not shown)*

§: overall p value for genotype frequencies; ‖: Odds ratio (OR) and 95% confidence limits (CI) calculated for carriers of the C allele (i.e. AC and CC genotypes against AA); P values are calculated with the χ^2 test and are 2-tailed

Table II. *Distribution of subjects with different numbers of type 1 diabetes predisposing HLA-DQαβ heterodimers in 677 controls and 92 type 1 diabetes patients carrying or not carrying the +1239C allele.*

HLA DQαβ susceptibility heterodimer*	Patients	Controls†		Genotype of patients	
	(N=92)	(N=677)		AA (N=39)	AC+CC (N=53)
0	11 (12)‡	356 (52.6)‡		6 (15.4)‡	5 (9.4)‡
1	6 (6.5)	185 (27.3)		1 (2.6)	5 (9.4)
2	18 19.5)	125 (18.5)		8 (20.5)	10 (18.9)
4	57 (62)	11 (1.6)		24 (61.5)	33 (62.3)
Overall p§	$<10^{-8}$ ‖			0.51¶	
4 carriers	$<10^{-8}$ ‖			0.88¶	

**: Number of susceptible heterodimers (HLA-DQα52Arg and DQβ57nonAsp). †: healthy babies recruited from the cord blood bank of the transfusion center; ‡: number of subjects; proportions are shown in the brackets; §: χ^2-square test comparing the frequency of patients and controls with the same number of predisposing HLA-DQαβ heterodimers; p values are 2-tailed; ‖: Controls vs patients ; ¶:: AA patients vs AC+CC patients*

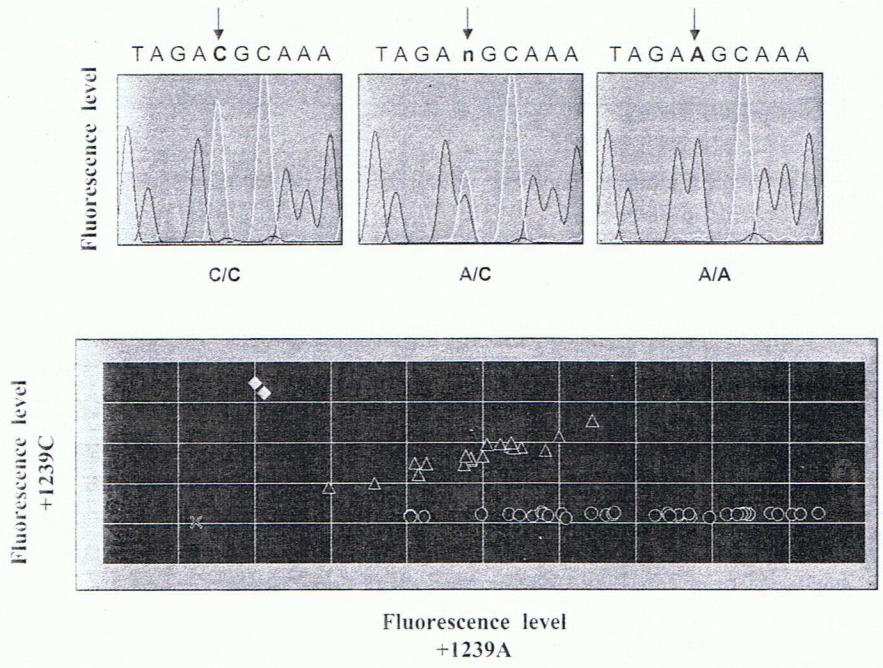


Fig. 1. *Genomic DNA typing of the +1239 A/C SNP. Upper Panel: Electropherograms from genomic DNA sequencing performed in three representative subjects carrying the +1239 CC, AC, and AA genotypes; the +1239 SNP is marked with an arrow. Lower Panel: Allelic discrimination plot of 65 representative samples: +1239A/A genotypes are shown with circles (n=33), +1239A/C with triangles (n=18), and +1239C/C with diamonds (n=1). Samples were amplified by the Taqman allelic discrimination method and positive controls were used for each genotype. Results were plotted showing the absolute fluorescence of each probe.*

analysis, susceptible heterodimers were those comprising a DQα chain with an arginine at position 52 and a DQβ chain with a non-aspartic acid at position 57; a subject can have 1, 2 or 4 susceptible heterodimers, the higher the number, the higher the risk of T1DM (19). Results showed that, as expected, the overall distribution of subjects carrying 0, 1, 2, or 4 susceptible heterodimers was significantly different in the patients and the controls, but it was similar in the patients carrying or not carrying +1239C (Table II). Moreover, the proportion of subjects carrying 4 predisposing heterodimers was significantly higher in the patients than in the controls (62% *vs* 1.6%, $p<10^{-5}$), but it was similar in the patients carrying or not carrying +1239C (63.3% *vs* 61.5%, p=0.88). These data suggests that the predisposing role of these HLA alleles is similar in these patient groups.

DISCUSSION

This work stems from our previous observation that the +1239C SNP of SPP1 is associated with production of increased amounts of Opn and favors development of ALPS, MS and SLE, and shows that it also predisposes to T1DM. This is suggested by the finding that frequency of +1239C carriers was higher in two independent cohorts of T1DM patients than in the respective controls; in the combined cohorts of 697 T1DM patients and 1218 controls, this allele conferred an OR=1.63 for T1DM development (CI: 1.34-1.97 p=0.0000005). Moreover, HLA typing of a small group of patients suggests that this variation did not influence the role played by HLA predisposing alleles, whose frequency was similarly increased in patients carrying or not +1239C.

The effect of +1239C on T1DM development may be related to its effect on Opn production, since it is associated with a high "baseline" production that may be due to the higher stability of its mRNA than the +1239A mRNA, as we previously showed in transfected cells (14-15). An alternative possibility is that +1239C does not have a direct effect, but is in LD with other variations. Indeed, SPP1 lies in a long LD block (126kb) and, according to HapMap (http://www.hapmap.org/) and literature, 18 validated SNPs (with minor allele frequency >0.05) are located in the SPP1 genomic region. The +1239 SNP shows a perfect LD with two SPP1 synonymous variations in the coding region (+282T/C, rs4754; +750C/T, rs11226616), and a high LD ($r^2 \geq 0.7$) with another 3'UTR SNP (+1083A/G, rs1126772), and 10 other markers located in the SPP1 3' flanking region or in intronic sequences of the other gene contained in the SPP1 LD block (PKD2, Polycystic Kidney Disease 2). Moreover, it is in significant LD (D'=0.9, r^2=0.15, $p<0.05$) with the -66T/G SNP (16) recently found to be associated with T1DM in the Italian population (17). Therefore, further studies are needed to define the primarily associated variation.

Recent genome-wide association studies (GWAS) reported no association with T1DM for markers in the SPP1 region in large cohorts from Canada, US, and UK (20-21). Although the +1239 SNP was not included in these studies, SNPs in LD ($r^2 \geq 0.6$) with +1239 were included in both studies and they did not display association. In particular, the synonymous +282T/C (rs4754) variation, that is in high LD (r^2=0.94 in HapMap CEU data, r^2=1.0 in our samples) with the +1239 SNP, showed no association ($p>0.05$) in about 1,000 patients from Canada (C Polychronakos, personal communication). This inconsistency might be explained by a population-specific association of SPP1 in T1DM. Several causes might explain a population-specific association of SNPs in complex diseases (22). For instance, the association detection may be favored by the peculiar allelic frequencies of the SNPs or by their LD with untyped causal alleles in that population. A substantial influence may also be exerted by interactions with population-specific environmental exposures or unlinked genetic factors. Classical population genetic studies highlighted the genetic distances of the Italian and other Mediterranean populations from northern European populations (23). Since the GWAS so far reported for T1DM have been performed in populations mainly of northern European origin, it is likely that other associations will emerge from studies of different populations.

In conclusion, this work detects an SPP1 variation that is clearly associated with T1DM development in the Italian population. The predisposing effect may depend on Opn-mediated support of inflammation and TH1 cell differentiation, which may build up an autoimmune-prone background favoring development of the autoimmune aggression of beta cells.

ACKNOWLEDGEMENTS

We thank Dr. Constantin Polychronakos for sharing data from his genome-wide association studies. This work was partially supported by PRIN Project (MIUR, Rome), Compagnia di San Paolo (Turin), Regione Piemonte "Ricerca Finalizzata" project, "Ricerca Applicata CIPE" project (Turin), AIRC (Milan) and "Alfieri project" (CRT, Turin).

REFERENCES

1. Yoon JW, Jun HS. Autoimmune destruction of pancreatic beta cells. Am J Ther 2005; 12:580-91.
2. Santamaria P. Effector lymphocytes in autoimmunity. Curr Opin Immunol 2001; 13:663-9.
3. Hirschhorn JN. Genetic epidemiology of type 1 diabetes. Pediatr Diabetes 2003; 4:87-100.
4. Jahromi MM, Eisenbarth GS. Genetic determinants of type 1 diabetes across populations. Ann NY Acad Sci 2006; 1079:289-299.
5. Buzzetti R, Galgani A, Petrone A, et al. Genetic prediction of type 1 diabetes in a population with low frequency of HLA risk genotypes and low incidence of the disease (the DIABFIN study). Diabetes Metab Res Rev 2004; 20:137-43.
6. Becker KG. Comparative genetics of type 1 diabetes and autoimmune disease: common loci, common pathways? Diabetes 1999; 48:1353-8.
7. Marre M. Genetics and the prediction of complications in type 1 diabetes. Diabetes Care 1999; 2:B53-58.
8. Todd JA. Genetics of type 1 diabetes. Pathol Biol 1997; 45:219-27.
9. Gravallese EM. Osteopontin: a bridge between bone and the immune system. J Clin Invest 2003; 112:147-9.
10. O'Regan A, Berman JS. Osteopontin: a key cytokine in cell-mediated and granulomatous inflammation. Int J Exp Pathol 2000; 81:373-90.
11. O'Regan AW, Nau GJ, Chupp GL, Berman JS. Osteopontin (Eta-1) in cell-mediated immunity: teaching an old dog new tricks. Immunol Today 2000; 21:475-8.
12. Denhardt DT, Noda M, O'Regan AW, Pavlin D, Berman JS. Osteopontin as a means to cope with environmental insults: regulation of inflammation, tissue remodeling, and cell survival. J Clin Invest 2001; 107:1055-61.
13. Ashkar S, Weber GF, Panoutsakopoulou V, et al. Eta-1 (osteopontin): an early component of type-1 (cell-mediated) immunity. Science 2000; 287:860-4.
14. Chiocchetti A, Indelicato M, Bensi T, et al. High levels of osteopontin associated with polymorphisms in its gene are a risk factor for development of autoimmunity/lymphoproliferation. Blood 2004; 103:1376-82.
15. Chiocchetti A, Comi C, Indelicato M, et al. Osteopontin gene haplotypes correlate with multiple sclerosis development and progression. J Neuroimmunol 2005; 163:172-8.
16. D'Alfonso S, Barizzone N, Giordano M, et al. Two single-nucleotide polymorphisms in the 5' and 3' ends of the osteopontin gene contribute to susceptibility to systemic lupus erythematosus. Arthritis Rheum 2005; 52:539-47.
17. Marciano R, D'Annunzio G, Minuto N, Pasquali L, Santamaria A, Di Duca M, Ravazzolo R, Lorini R. Association of alleles at polymorphic sites in the osteopontin encoding gene in young type 1 diabetic patients. Clin Immunol 2009; 131:84-91.
18. Iafusco D, Stazi MA, Cotichini R, et al. Early Onset Diabetes Study Group of the Italian Society of Paediatric Endocrinology and Diabetology. Permanent diabetes mellitus in the first year of life. Diabetologia 2002; 45:798-804.
19. Lipton RB, Kocova M, LaPorte RE, Dorman JS, Orchard TJ, Riley WJ, Drash AL, Becker DJ, Trucco M. Autoimmunity and genetics contribute to the risk of insulin-dependent diabetes mellitus in families: islet cell antibodies and HLA DQ heterodimers. Am J Epidemiol 1992; 136:503-12.
20. Cooper JD, Smyth DJ, Smiles AM, et al. Meta-analysis of genome-wide association study data identifies additional type 1 diabetes risk loci. Nat Genet 2008; 40:1399-401.
21. Hakonarson H, Grant SF, Bradfield JP, et al. A genome-wide association study identifies KIAA0350 as a type 1 diabetes gene. Nature 2007; 448:591-4.
22. McCarthy MI, Abecasis GR, Cardon LR, Goldstein DB, Little J, Ioannidis JP, Hirschhorn JN. Genome-

wide association studies for complex traits: consensus, uncertainty and challenges. Nat Rev Genet 2008; 9:356-69.

23. Cavalli-Sforza L, Menozzi P, Piazza A. The history and geography of human genes. Princeton University press (Princeton, New Jersey) 1994; p. 255-99.

Association of the *CBLB* gene with multiple sclerosis: new evidence from a replication study in an Italian population

Lucia Corrado, Laura Bergamaschi, Nadia Barizzone, et al.

J Med Genet published online October 30, 2010
doi: 10.1136/jmg.2010.081380

JMG Online First, published on October 30, 2010 as 10.1136/jmg.2010.081380

Association of the *CBLB* gene with multiple sclerosis: new evidence from a replication study in an Italian population

Lucia Corrado,[1] Laura Bergamaschi,[1] Nadia Barizzone,[1] Maria Edvige Fasano,[2] Franca R Guerini,[3] Marco Salvetti,[4] Daniela Galimberti,[5] Maria Donata Benedetti,[6] Maurizio Leone,[7] Sandra D'Alfonso[1]

[1]Department of Medical Sciences and Interdisciplinary Research Center of Autoimmune Diseases (IRCAD), University of Eastern Piedmont, Novara, Italy
[2]SCDU Transplantation Immunology, A.O.U. S. Giovanni Battista di Torino, Turin, Italy
[3]Laboratory of Molecular Medicine and Biotechnologies, Don C. Gnocchi Foundation IRCCS, S Maria Nascente, Milan, Italy
[4]Neurology and Center for Experimental Neurological Therapy (CENTERS), Università La Sapienza, Rome, Italy
[5]Università di Milano, Fondazione Cà Granda, IRCCS Ospedale Maggiore Policlinico, Milan, Italy
[6]Section of Neurology, Department of Neurological and Vision Sciences, University of Verona, Policlinico G Rossi, Verona, Italy
[7]Department of Neurology, A.O. U. Maggiore della Carità, and IRCAD, Novara, Italy

Correspondence to
Professor Sandra D'Alfonso, Department of Medical Sciences, Eastern Piedmont University, Via Solaroli 17, 28100 Novara, Italy; dalfonso@med.unipmn.it

Received 4 June 2010
Revised 9 September 2010
Accepted 13 September 2010

ABSTRACT

Background The T allele of rs9657904 within the *CBLB* gene was recently found to be significantly associated with multiple sclerosis (MS) in a genome-wide association study in Sardinia.

Objective To replicate this association in an independent population with a different genetic background.

Methods The rs9657904 variant was typed in a sample of 1435 cases and 1466 controls from the Italian mainland.

Results It was found that in this sample also, the common allele T of rs9657904 is significantly positively associated (one-tailed $p=7.35\times10^{-5}$) and with a comparable effect size with MS (OR=1.31, 95% CI 1.14 to 1.52).

Conclusion These data provide further evidence of the association of MS disease with variation within *CBLB*.

INTRODUCTION

Multiple sclerosis (MS) is a multifactorial neuroinflammatory and autoimmune disorder characterised by a progressive demyelination of axons of the central nervous system and neuronal cell degeneration, resulting in a severe disabling condition. Interactions between unknown environmental factors and alleles of many susceptibility loci across the genome contribute together to the development of the disease.[1] Until recently, the only genes consistently associated with MS mapped to the human major histocompatibility complex or HLA (human leucocyte antigen) region.

Study of the genetics of MS, after years of difficulty, is undergoing a period of rapid development. This is owing to the use of large datasets as well as improved genotyping techniques that have allowed the first genome-wide association scans (GWAS).[2–6] Once those significant associations are replicated in different populations, they acquire a profound importance for understanding a disease and for focusing lines of investigation for functional and bioinformatics analysis.[7 8]

Thus far, the major GWAS findings have come from analyses of populations with northern European origin in which MS is particularly common. Recently, a novel association with risk for MS of some markers within the *CBLB* gene (Cas-Br-M (murine) ecotropic retroviral transforming sequence b, 3q13.11) was observed in the island population from Sardinia.[6] To assess the associated variant in an independent sample set from the Italian mainland is cogent for several reasons. First, while sharing a very similar environment with much of continental Italy, Sardinians are, by most measures, genetically different from continental Italians, although to a lesser extent than populations from northern Europe.[9] Furthermore, Sardinia has an MS prevalence at least three times higher than that of the Italian mainland.[10] Also, the main genetic risk factor for MS at the HLA class II *DRB1-DQB1* loci shows a very different allelic distribution in Italy than in Sardinia. In the Sardinian population, the HLA association is mainly accounted for by the *HLA-DRB1*03:01* allele,[11] included within an extended or ancestral HLA haplotype—namely, *HLA-A*30, B*18, Cw*5, DRB1*03:01*, which is very rare elsewhere. Conversely, in continental Italy the MS association, as in most other populations, is marked by the *HLA-DRB1*15:01* allele.[12]

Hence we tested the positive association observed with the T allele of the top associated variant (rs9657904) observed in the Sardinian study in a large cohort of continental patients with MS and controls.

PATIENTS AND METHODS

The SNP rs9657904 C→T in intron 1 of the *CBLB* gene was genotyped in 1435 patients with MS and 1466 regionally matched controls from continental northern–central Italy.

Patients with MS had a female:male ratio of 2:1, a mean age of onset of 31.6±10.3 years, a mean expanded disability status scale 3.10±2.23 and a mean multiple sclerosis severity score 3.91±2.72.[13] Ninety per cent of the patients presented a relapsing remitting while 10% presented a primary progressive disease course. The controls (female:male 1.3:1) were blood donors who shared the same ethnicity background with cases. Individuals with Sardinian origin were selectively excluded. All the samples were collected after informed consent and appropriate ethical approval.

Genotyping was performed with a Taqman genotyping assay (assay ID C_1499397_10, Applied Biosystems, Foster City, CA). The genotype success rate of this Taqman assay was 97%.

The statistical significance of the difference of allele and genotype frequencies between patients with MS and controls was evaluated using the χ^2 test with Yates's correction. The strength of association was evaluated by the odds ratio (OR) and its 95% CIs.

Table 1 Association results of *CBLB* rs9657904

	Cases N (%)	Controls N (%)	OR (95% CI)
Genotype			
TT	1054 (73.4)	974 (66.4)	1.40 (1.19 to 1.64)
TC	347 (24.2)	451 (30.8)	0.72 (0.61 to 0.85)
CC	34 (2.4)	41 (2.8)	0.84 (0.52 to 1.37)
Allele			
T	2455 (85.5)	2399 (81.8)	1.31 (1.14 to 1.52)
C	415 (14.5)	533 (18.2)	

The frequency of the T allele was significantly increased in the patients (one-tailed $p = 7.35\times10^{-5}$).
N indicates the number of individuals or alleles.

RESULTS AND DISCUSSION

The same allele (T) of rs9657904 single nucleotide polymorphism (SNP), positively associated with MS in Sardinian population, was significantly associated with MS risk (one-tailed $p=7.35\times10^{-5}$), and even showed a comparable effect size (OR=1.31) also in the Italian mainland (table 1).

The genotype frequencies did not deviate from Hardy–Weinberg equilibrium either in cases (p=0.39) or controls (p=0.19). The association of the T allele seems to be consistent with a recessive model, since it showed a significant increase in patients with MS only in homozygosis (table 1). Moreover, no significant interaction was detected in a case-only analysis with *HLA-DRB1*15:01* allele, with no difference of allele frequencies between *HLA-DRB1*15:01* positive (n=361) and negative (n=825) patients (minor allele frequency: 0.15 vs 0.14 p = 0.67). This is consistent with the reported lack of interaction with *HLA-DRB1*03:01* in the Sardinian study[5]and indicates that the same *CBLB* variant is associated with MS in two populations with distinct HLA associations. Moreover, the associated variant exhibits nearly overlapping frequencies in Sardinia[5] and in the Italian mainland; it is thus unlikely that variation at *CBLB* explains the higher disease prevalence seen in Sardinia.

Conversely, the *CBLB* gene does not appear in the list of MS-associated loci that satisfy the genome-wide significance threshold from previous GWAS, despite this SNP is tagged ($r^2\geq0.9$) by at least one SNP in the different platforms used so far.[1 2–4] This could suggest that in these populations, mainly of northern European origin, rs9657904 SNP shows a weaker association with MS. This observation supports the hypothesis that the tested SNP in the *CBLB* gene is not the primarily associated variant and may indicate that the linkage disequilibrium structure of populations from southern Europe might favour the detection of this association.

Since the disease association might be affected by many variables, in particular by the linkage disequilibrium between marker allele and causal allele, further cross-population comparisons using samples from more distantly related populations along with additional resequencing/fine mapping work appear necessary to reduce the MS association to its essential, potentially causal elements.

Acknowledgements We are grateful to the patients and their parents. We are grateful to Professor Francesco Cucca for helpful discussions and to Dr Elizabeth Grass for editing the manuscript.

Funding This work was supported by the Italian Foundation for Multiple Sclerosis (FISM grants 2001/R/44, 2002/R/40 and 2005/R/10, 2008/R/11); Regione Piemonte Ricerca Sanitaria Finalizzata (grants 2003 2004, 2007, 2008, 2009), Italian MIUR Ministry (PRIN 2008), Eastern Piedmont University, Compagnia di San Paolo (Turin), Fondazione CRT (Turin). DG was supported by Italian Ministry of Health "Progetto Giovani Ricercatori 2008". NB and LC were supported by a fellowship from FISM (2003/B/2, 2009/B/1). LB was supported by a PhD Lagrange Fellowship.

Competing interests None.

Ethics approval This study was conducted with the approval of the local ethic committees.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

1. **Compston A,** Sadovnick AD. Epidemiology and genetics of multiple sclerosis. *Curr Opin Neurol Neurosurg* 1992;**5**:175–81.
2. Australia and New Zealand Multiple Sclerosis Genetics Consortium (ANZgene). *Nat Genet* 2009;**41**:824–8.
3. **De Jager PL,** Jia X, Wang J, de Bakker PI, Ottoboni L, Aggarwal NT, Piccio L, Raychaudhuri S, Tran D, Aubin C, Briskin R, Romano S; International MS Genetics Consortium,Baranzini SE, McCauley JL, Pericak-Vance MA, Haines JL, Gibson RA, Naeglin Y, Uitdehaag B, Matthews PM, Kappos L, Polman C, McArdle WL, Strachan DP, Evans D, Cross AH, Daly MJ, Compston A, Sawcer SJ, Weiner HL, Hauser SL, Hafler DA, Oksenberg JR. Meta-analysis of genome scans and replication identify CD6, IRF8 and TNFRSF1A as new multiple sclerosis susceptibility loci. *Nat Genet* 2009;**41**:776–82.
4. International Multiple Sclerosis Genetics Consortium (IMSGC). *N Engl J Med* 2007;**357**:851–62.
5. **Baranzini SE,** Wang J, Gibson RA, Galwey N, Naegelin Y, Barkhof F, Radue EW, Lindberg RL, Uitdehaag BM, Johnson MR, Angelakopoulou A, Hall L, Richardson JC, Prinjha RK, Gass A, Geurts JJ, Kragt J, Sombekke M, Vrenken H, Qualley P, Lincoln RR, Gomez R, Caillier SJ, George MF, Mousavi H, Guerrero R, Okuda DT, Cree BA, Green AJ, Waubant E, Goodin DS, Pelletier D, Matthews PM, Hauser SL, Kappos L, Polman CH, Oksenberg JR. Genome-wide association analysis of susceptibility and clinical phenotype in multiple sclerosis. *Hum Mol Genet* 2009;**18**:767–78.
6. **Sanna S,** Pitzalis M, Zoledziewska M, Zara I, Sidore C, Murru R, Whalen MB, Busonero F, Maschio A, Costa G, Melis MC, Deidda F, Poddie F, Morelli L, Farina G, Li Y, Dei M, Lai S, Mulas A, Cuccuru G, Porcu E, Liang L, Zavattari P, Moi L, Deriu E, Urru MF, Bajorek M, Satta MA, Cocco E, Ferrigno P, Sotgiu S, Pugliatti M, Traccis S, Angius A, Melis M, Rosati G, Abecasis GR, Uda M, Marrosu MG, Schlessinger D, Cucca F. Variants within the immunoregulatory CBLB gene are associated with multiple sclerosis. *Nat Genet* 2010;**42**:495–7.
7. **Liu DV,** Maier LM, Hafler DA, Wittrup KD. Engineered interleukin-2 antagonists for the inhibition of regulatory T cells. *J Immunother* 2009;**32**:887–94.
8. **Baranzini SE,** Galwey NW, Wang J, Khankhanian P, Lindberg R, Pelletier D, Wu W, Uitdehaag BM, Kappos L; GeneMSA Consortium,Polman CH, Matthews PM, Hauser SL, Gibson RA, Oksenberg JR, Barnes MR. Pathway and network-based analysis of genome-wide association studies in multiple sclerosis. *Hum Mol Genet* 2009;**18**:2078–90.
9. **Lampis R,** Morelli L, Congia M, Macis MD, Mulargia A, Loddo M, De Virgiliis S, Marrosu MG, Todd JA, Cucca F. The inter-regional distribution of HLA class II haplotypes indicates the suitability of the Sardinian population for case-control association studies in complex diseases. *Hum Mol Genet* 2000;**9**:2959–65.
10. **Pugliatti M,** Rosati G, Carton H, Riise T, Drulovic J, Vécsei L, Milanov I. The epidemiology of multiple sclerosis in Europe. *Eur J Neurol* 2006;**13**:700–22.
11. **Marrosu MG,** Murru R, Murru MR, Costa G, Zavattari P, Whalen M, Cocco E, Mancosu C, Schirru L, Solla E, Fadda E, Melis C, Porru I, Rolesu M, Cucca F. Dissection of the HLA association with multiple sclerosis in the founder isolated population of Sardinia. *Hum Mol Genet* 2001;**10**:2907–16.
12. **Ballerini C,** Guerini FR, Rombolà G, Rosati E, Massacesi L, Ferrante P, Caputo D, Talamanca LF, Naldi P, Liguori M, Alizadeh M, Momigliano-Richiardi P, D'Alfonso S. HLA-multiple sclerosis association in continental Italy and correlation with disease prevalence in Europe. *Neuroimmunol* 2004;**150**:178–85.
13. **Roxburgh RH,** Seaman SR, Masterman T, Hensiek AE, Sawcer SJ, Vukusic S, Achiti I, Confavreux C, Coustans M, le Page E, Edan G, McDonnell GV, Hawkins S, Trojano M, Liguori M, Cocco E, Marrosu MG, Tesser F, Leone MA, Weber A, Zipp F, Miterski B, Epplen JT, Oturai A, Sørensen PS, Celius EG, Lara NT, Montalban X, Villoslada P, Silva AM, Marta M, Leite I, Dubois B, Rubio J, Butzkueven H, Kilpatrick T, Mycko MP, Selmaj KW, Rio ME, Sá M, Salemi G, Savettieri G, Hillert J, Compston DA. Multiple Sclerosis Severity Score: using disability and disease duration to rate disease severity. *Neurology* 2005;**64**:1144–51.

Elsevier Editorial System(tm) for Human Immunology
Manuscript Draft

Manuscript Number:

Title: Association of Osteopontin (OPN) regulatory polymorphisms with Systemic Sclerosis (SSc).

Article Type: Research Article

Keywords: Systemic Sclerosis; Osteopontin; genetic association; polymorphisms

Corresponding Author: Dr. Nadia Barizzone,

Corresponding Author's Institution: University of Eastern Piedmont "Amedeo Avogadro"

First Author: Nadia Barizzone

Order of Authors: Nadia Barizzone; Maurizio Marchini; Francesca Cappiello; Annalisa Chiocchetti; Elisabetta Orilieri; Daniela Ferrante; Lucia Corrado; Simona Mellone; Raffaella Scorza; Umberto Dianzani; Sandra D'Alfonso

Abstract: To test the involvement of osteopontin (OPN) in Systemic Sclerosis susceptibility, two OPN SNPs previously reported to be associated with SLE, namely 156G/GG (5'flanking) and +1239A/C (3'UTR), were tested in 357 Italian patients and 864 matched controls. OPN serum levels were determined by ELISA in 32 patients and 116 controls.
Compared with the controls, in SSc patients there was a significantly increased frequency of the alleles -156G (p=0.0086), and +1239C (p=0.00064), paralleling the association reported for SLE.
According to logistic regression analysis this association is primarily due to the effect of +1239 SNP.
OPN serum levels were significantly higher in SSc patients than in controls (p=0.00025).
These data suggest that OPN genetic variations have a role in SSc susceptibility, reporting for the first time an involvement of this molecule in SSc pathogenesis and emphasizing that SSc shares pathogenetic mechanisms with other autoimmune diseases.

Suggested Reviewers: Koen Vandenbroeck
Neurogenomiks Group, Universidad del País Vasco (UPV/EHU), Leioa, Spain
k.vandenbroeck@ikerbasque.org

Marta Alarcón-Riquelme
Department of Genetics and Pathology, Uppsala University, Uppsala, Sweden
marta.alarcon@genpat.uu.se

An Goris
Katholieke Universiteit Leuven, Belgium
An.Goris@med.kuleuven.be

Editor of **Human Immunology**

Novara 29th December 2010

Dear Editor,

Enclosed please find our manuscript: "Association of Osteopontin (OPN) regulatory polymorphisms with Systemic Sclerosis (SSc)" (Barizzone et al) that we would like to submit to Human Immunology as a "Research Article". This paper describes the genetic association between two functional polymorphisms of Osteopontin (*OPN*) gene with Systemic Sclerosis. The study arise from a previous work by our own group, where the same polymorphisms where described in association with Systemic Lupus Erythematosus. The result was obtained on a quite large sample set for a rare disease (357 patients and 864 matched controls) and correlates with serum levels of OPN protein.

On behalf of all coauthors, I declare that:

- all coauthors have seen and agree with the contents of the manuscript;
- none of the authors has any financial interest related to this study;
- this study has not been published and is not under consideration for publication elsewhere.
- this work does not involve any experiment on humans or animals.

We hope that our manuscript may fulfill the interests of the readers of **Human Immunology**

Best regards

Nadia Barizzone

Nadia Barizzone
Department of Medical Sciences
Eastern Piedmont University
Via Solaroli 17
28100 Novara, Italy
Telephone: 39 321 660606, Fax : 39 321 620421
Email: barizzo@med.unipmn.it

Association of Osteopontin (OPN) regulatory polymorphisms with Systemic Sclerosis (SSc).

Nadia Barizzone[1], Maurizio Marchini[2], Francesca Cappiello[2], Annalisa Chiocchetti[1], Elisabetta Orilieri[1], Daniela Ferrante[3], Lucia Corrado[1], Simona Mellone[1], Raffaella Scorza[2], Umberto Dianzani[1], Sandra D'Alfonso[1]

[1]Department of Medical Sciences and Interdisciplinary Research Center of Autoimmune Diseases, "A. Avogadro" University of Eastern Piedmont, Novara, Italy

[2]Unit of Clinical Immunology and allergology IRCCS Fondazione Policlinico Mangiagalli e Regina Elena and University of Milano

[3]Unit of Medical Statistics and Cancer Epidemiology, CPO Piemonte and University of Eastern Piedmont, Novara, Italy

Corresponding author
Barizzone Nadia
Department of Medical Sciences and IRCAD (Interdisciplinary Research Center of Autoimmune Diseases)
A. Avogadro University
Via Solaroli 17
28100 Novara
Italy
Telephone: 39 321 660606
Fax : 39 321 620421
Email: barizzo@med.unipmn.it

ABSTRACT

To test the involvement of osteopontin *(OPN)* in Systemic Sclerosis susceptibility, two *OPN* SNPs previously reported to be associated with SLE, namely -156G/GG (5'flanking) and +1239A/C (3'UTR), were tested in 357 Italian patients and 864 matched controls. OPN serum levels were determined by ELISA in 32 patients and 116 controls.

Compared with the controls, in SSc patients there was a significantly increased frequency of the alleles -156G (p=0.0086), and +1239C (p=0.00064), paralleling the association reported for SLE. According to logistic regression analysis this association is primarily due to the effect of +1239 SNP.

OPN serum levels were significantly higher in SSc patients than in controls (p=0.00025).

These data suggest that *OPN* genetic variations have a role in SSc susceptibility, reporting for the first time an involvement of this molecule in SSc pathogenesis and emphasizing that SSc shares pathogenetic mechanisms with other autoimmune diseases.

Key words: Systemic Sclerosis; Osteopontin; genetic association; polymorphisms

1 INTRODUCTION

Systemic Sclerosis (scleroderma, SSc) is a rare (prevalence in Caucasians: 2.6/10,000 [1]) multisystemic disease characterized by widespread fibrosis involving skin as well as internal organs such as lung and heart, activation of immune cells, production of anti-nuclear autoantibodies (ANA), and injury to vascular and microvascular structures. Its pathogenesis is unclear and both genetic and environmental factors are believed to contribute to its susceptibility and clinical expression. Relative risk for siblings, calculated in three US cohorts, is 10-27 fold higher than that of the general population [2]. Although the clinical concordance rates for MZ and DZ twins are similar (<5%), there is a significantly higher concordance for ANA positivity in MZ twins (90%) than in DZ twins (40%) [2], suggesting that inherited factors have an important role for disease susceptibility, but a triggering environmental factor is needed for its development. This model is also supported by expression profile data [3]. SSc fibroblasts from both lesional and nonlesional skin sites exhibit an activated phenotype characterized by increased expression of extracellular matrix components, cytokines and growth factors, including collagen and osteonectin. This pattern has been found to be concordant in 40-50% clinically discordant MZ twins, but no DZ twins [3].

Many candidate gene association studies are present in the literature, but so far, besides HLA, only few genes have been consistently associated with the disease [4], and they include pathways involved in autoimmunity, endothelial function, and extracellular matrix deposition. Many of the most convincing associations with SSc were also detected in other autoimmune diseases. Among these, *BANK1* [5] *STAT4* [6,7], *IRF5* [6,8] and *PTPN22* [9] had previously been associated with Systemic Lupus Erythematosus (SLE). The associations with *STAT4* and *IRF5* were confirmed also in a GWAS on samples of European ancestry [10]. Another GWAS was performed in an

Asiatic population [11] but it only led to the identification of susceptibility loci in the HLA region.

Accordingly, there is an increasing evidence that several autoimmune diseases share a common pool of susceptibility genes despite their otherwise unique clinical features and varying pathophysiologies [12]. In past years, single nucleotide polymorphisms (SNPs) in the Osteopontin (*OPN*) gene (4q21-q25) have been associated with susceptibility to several autoimmune diseases [13-15]. Osteopontin is a 60-kd secreted phospho-protein functioning as a free cytokine in the body fluids or as an immobilized extracellular matrix molecule in mineralized tissues [16]. As a cytokine, it plays a role in cell-to-cell and cell-to-extracellular matrix interactions and is involved in inflammation, tissue repair, and lymphocyte differentiation [16,17].

Its involvement in SSc has never been tested, but several features of OPN function suggest that it might play a role. First, OPN may be implicated through its immunoregulatory effects, enhancing the proinflammatory Th1 cell response, believed to be crucial in SSc pathogenesis [17,18]. Second, transgenic mice hyper-expressing OPN display hypergammaglobulinemia and anti-DNA autoantibodies [19]. Third, OPN might also be involved in the control of SSc fibrosis because of its key role in the synthesis and/or turnover of matrix components in both human [20] and mice [21] models. Finally, polymorphisms in the *OPN* gene were reported to be associated to SLE [14,22,23]

The aim of the present study was to analyse the involvement of OPN in SSc pathogenesis. To this purpose we tested the association of two *OPN* SNPs, namely rs3841116 (-156G/GG; 5'flanking) and rs9138 (+1239A/C; 3'UTR) in a panel of Italian SSc patients and controls. These two markers were selected on the basis of both genetic and functional data (see discussion), arising from our previous association study with SLE (5). In addition to the genetic analysis, we also compared OPN serum levels in SSc patients and healthy individuals, and correlated them with *OPN* genotypes.

2 MATERIAL AND METHODS

2.1 Subjects.

A total of 357 Italian patients with SSc (13.2:1 female:male ratio, mean ± SD age at onset 46.7 ± 14.8) were included in this study. All patients fulfilled the American College of Rheumatology preliminary criteria for the classification of SSc [24] and were categorised as having the limited cutaneous (lcSSc) or the diffuse cutaneous (dcSSc) subset according to LeRoy et al [25].

All patients provided their informed consent. The clinical and immunological features of the SSc patients are shown in Table 1.

Disease onset was determined by the patient's recall of the first non-Raynaud symptom clearly ascribable to scleroderma.

Pulmonary hypertension, estimated by echocardiography, was defined as a right-ventricular systolic pressure ≥ 40 mmHg. Pulmonary fibrosis was evaluated by high resolution TC (HR-TC).

A total of 864 Italian individuals (comprising medical students, university and hospital staff, and blood donors) were used as control subjects (2.6:1 female:male ratio). The regional origin of the patients and controls was similar. Individuals with Sardinian ancestors were excluded. About 50% of the controls (N = 400) are the same included in a previous work [14]. The remaining are newly genotyped individuals.

2.2 Enzyme-linked immunosorbent assay (ELISA).

Serum OPN concentrations were evaluated by a capture ELISA in accordance with the protocol provided by the manufacturer (Calbiochem, La Jolla, CA). The optical density was measured at 450 nm with a microplate reader (Bio-Rad, Hercules, CA). The I-smart program was used to create a regression curve. Serum OPN levels for all controls were already reported in a previous paper [14].

2.3 Genotyping of *OPN* SNPs.

The two selected SNPs located at the 5'flanking (-156G/GG, rs3841116) and 3'UTR (+1239A/C, rs9138) of the *OPN* gene were co-amplified in a multiplex PCR reaction and genotyped in a unique SNaPshot reaction. Samples were electrophoresed on the ABI 3100 Genetic Analyzer. Resulting electropherograms were analyzed using the ABI GeneScan, version 3.7, software. PCR and SNaPshot conditions are available upon request.

2.4 Statistical analysis. The statistical significance of the difference in gene and genotype frequencies between patients and controls was evaluated for each polymorphism using the chi-square test with Yates' correction. The association of each polymorphism with the disease was measured by the odds ratio (OR) and its 95% confidence interval (95% CI), The effect of the two tested polymorphisms relative to each other was tested by conditional analysis using the main-effects test of the COCAPHASE program' part of the UNPHASED suite [26].
Haplotypes and LD values were calculated with the Haploview software (Cambridge, MA 02141, USA).

The OPN serum levels in SSc patients and controls were compared by the Mann-Whitney U test.

3 RESULTS

3.1 Genetic association analysis

The two selected *OPN* SNPs, -156G/GG and +1239A/C, were analysed in 357 SSc patients and 864 controls.

Both SNPs showed a significantly different allele frequency ($p = 0.0086$ and 0.00064 respectively for -156 and for +1239) (Table 2). For both SNPs the overall distribution of the genotype frequencies was significantly different in SSc patients and in controls.

For simplicity, throughout this paper alleles significantly increased in the patients (-156G and +1239C) are named 1 (i.e. susceptibility alleles), whereas alleles significantly decreased in the patients (-156GG and +1239A) are named 0 (i.e., protective alleles).

A significant linkage disequilibrium (LD) between the two SNPs was observed in SSc patients (D'=0.87, r^2=0.17, $p<10^{-6}$) as previously seen in Italian controls and SLE patients [14]. The two SNPs form three equally frequent haplotypes plus a rare one (frequency < 2% both in SSc cases and in controls; Table 3). Two of the frequent haplotypes showed a significantly different distribution in patients and in controls, with the haplotype containing both susceptibility alleles significantly increased (0.345 vs. 0.273, p=0.00036) and the haplotype characterized by two protective alleles significantly decreased in the patients (0.268 vs. 0.322, p=0.010). The third common haplotype (-156 1 +1239 0), had the same frequency in the patients and in controls, suggesting that -156 SNP alone has no effect on disease susceptibility. The same was observed for the rare haplotype (-156 0 +1239 1), but this latter category was highly underpowered due to its low frequency. Therefore this analysis does not allow to distinguish if the observed association is due to a haplotypic effect or to the main effect of +1239C allele.

The relative effect of the two associated alleles on SSc risk was further evaluated by a regression analysis. The association with +1239 remained statistically significant also after adjustment for -156 (p=0.0096), indicating that its effect is independent of LD with the other variant. Conversely, the association with -156 was no longer significant after adjustment (p=0.14), indicating that most of its effect is consequent to LD with +1239.

When we compared patients with diffuse (dcSSc) disease (n=73) with those affected by limited (lcSSc) disease (n=250) we saw an increased frequency, with a borderline significance, of both the –156GG allele (p=0.026) and -156GG/GG genotype (p=0.031) in the dcSSc sample set, indicating that this variation might be a phenotype modifying factor. Instead no significant difference, both in gene or in genotype frequencies, was detected for the +1239 SNP (data not shown).

3.2 Correlation with OPN serum levels.

In order to investigate a possible functional basis for the observed genetic association, we analyzed OPN protein production. OPN serum levels measured by ELISA were significantly (Mann-Whitney U test p=0.00025) higher in 32 SSc patients (median 225.5 ng/ml, interquartile range 145.8-394.4) than in 116 controls (median 148.8 ng/ml, interquartile range 95.5-201.3 [13-14].

At difference from what previously reported for controls [13], patient OPN serum levels were not associated with +1239 or -156 genotypes (data not shown).

4 DISCUSSION

This work follows the finding of an association with SLE of two *OPN* SNPs (+1239 A/C in the 3'UTR of the gene and 156 G/GG in the 5' flanking) [14]. These were the only two primarily associated OPN polymorphisms out of the exonic and 5' flanking tested SNPs. The association with both promoter and 3'end OPN polymorphisms has been reported also in other populations both for SLE [22,23] and other autoimmune diseases (T1D [27,], ALPS [13], MS [28,15].

We here found that the same *OPN* alleles that confer susceptibility to SLE are significantly increased also in SSc patients suggesting that OPN is involved in SSc pathogenesis. Among the two tested SNPs, the primarily associated one seems to be +1239, although functional studies would be necessary to test the possible modifying effect of the 5' polymorphism.

This marker confers higher susceptibility also to other autoimmune diseases. In particular, the +1239 C allele increased by 8-fold the risk of ALPS (Autoimmune Lymphoproliferative Syndrome [13]), by 1.6 the risk of SLE [14] and type 1 Diabetes [29] and by 1.5-fold the risk of Multiple Sclerosis in the Italian population [15]. Recent findings suggest that there is a common polygenic background for autoimmune diseases, on which disease-specific factors act with a

pleiotropic effect [30]. This hypothesis has been confirmed more recently by data obtained from GWAS, from which we are beginning to understand that some immune pathways are shared between autoimmune diseases [31]. This observation can explain why more than one autoimmune disease can coexist in the same patient (polyautoimmunity) or family (familial autoimmunity), as it has been observed also for SSc with SLE, Rheumatoid Arthritis, Sjögren Sindrome, and T1D [32]. Thus, if these data will be confirmed in further diseases and populations, *OPN* can be added to the elongating list of genes conferring risk for several autoimmune diseases, besides *HLA*, *PTPN22*, *CTLA4*, *STAT4*.

The *OPN* gene does not appear in the list of SSc-associated loci that satisfy the genome-wide significance threshold from the GWAS by Radstake et al. [10], despite in the platform used the +1239 variant is tagged by one SNP (rs4754, r2=0.941). This might indicate that +1239 shows a weaker association with SSc in populations of northern European ancestry, to whom the samples used for the GWAS mostly belonged.

The SSc associated variant (+1239) is located in a regulative region of the *OPN* gene and previously (5) reported data indicate a possible direct functional role for this SNP. Particularly, it seems to influence *OPN* mRNA stability since mRNA carrying the +1239C allele was 4.4-fold more expressed than mRNA carrying +1239A [14]. Accordingly, the +1239C allele was significantly associated with a higher OPN serum level in healthy controls [13,14]. Based on these observations, we can speculate that this sequence variant (or others in perfect linkage disequilibrium with it) induces production of higher basal levels of OPN. High OPN levels might, in turn, confer susceptibility to autoimmune diseases. Differently from what observed in controls, *OPN* genotypes did not correlate with the protein level in the patients. This result is analogous to what seen in other autoimmune diseases (ALPS, SLE and SM [13-15]) where we found significantly increased OPN serum levels in patients compared with controls, and a significant association between *OPN* alleles and disease, but no correlation between OPN serum levels and genotypes. This may be explained by an overwhelming effect of the immune

activation in the patients following the initial trigger. This finding gives support to the observation of a shared common gene expression profiles in autoimmune diseases [33].

An alternative possibility is that the associated SNPs do not have a direct effect, but are in linkage disequilibrium (LD) with other known or unknown variations. The OPN gene lies in a long LD block (126kb). The *SSc* primarily associated variation (+1239C/T) shows a high LD (r2>0.69) with 3 intragenic *OPN* SNPs (two synonymous variations and one in the 3'UTR) and 9 markers located in the *OPN* 3'flanking region and in the intronic sequences of a second gene (*PKD2*, Polycystic Kidney Disease 2) contained in the same LD block. *OPN* -156 SNP shows, in turn, a high LD with one intronic polymorphism in the *OPN* gene and with other 11 SNPs located in the same LD block, one of which maps in the PKD2 intronic region. According to HapMap (http://www.hapmap.org/) and literature data, 18 validated SNPs (minor allele frequency >0.05) are located in the genomic region containing the *OPN* gene and its proximal regulative sequences, including 5 tag SNPs. The two SNPs analysed in this paper have the power to capture ($r^2 \geq 0.6$) three of these tag SNPs. The two uncaptured tag markers are respectively an intronic and a promoter variant. The latter (-616) was previously analysed in the SLE population [14], but no association was detected.

Genetically determined high levels of OPN may increase susceptibility to SSc through its capacity of stimulating proliferation and inhibiting lymphocyte death [13] or modulating the immune response by inducing Th1 responses and potentiating polyclonal activation of B cells [17,18]. Moreover, a genetically determined higher production of OPN might create a predisposition to SSc fibrosis. According to recent reports, OPN is strongly expressed in alveolar macrophages accumulating in fibrotic areas in bleomicin-treated mice [21] and in alveolar epithelial cells from human idiopathic pulmonary fibrosis patients [20], and an increased expression of OPN in the lung is associated with the development of the fibrotic process. Considering all these data, OPN might be involved in the pathogenesis of SSc by influencing both fibrosis and autoimmune processes, two of the components characterising SSc disease.

In conclusion, this study indicates an involvement of OPN in SSc pathogenesis and emphasizes that SSc shares pathogenetic mechanisms with other autoimmune diseases.

ACKNOWLEDGEMENTS

This work was supported by FISM grant 2008/R/711 (Genoa), PRIN Project (MIUR, Rome), Fondazione Cariplo Ricerca (Milan), Associazione Italiana Ricerca sul Cancro grant IG10237 (Milan), Fondazione Cassa di Risparmio di Cuneo (Cuneo), Regione Piemonte (Turin), and Associazione "Amici di Jean" (Turin) .

POLICY AND ETHICS

This work does not involve any experiment on humans or animals. The authors have no actual or potential conflict of interest to declare. The final article has been approved by all the authors.

REFERENCES

[1] Arnett FC, Cho M, Chatterjee S, Aguilar MB, Reveille JD, Mayes MD. Familial occurrence frequencies and relative risk for Systemic Sclerosis (Scleroderma) in three United States Cohorts. Arthritis Rheum. 2001;44:1359-62.

[2] Assassi S, Tan FK. Genetics of scleroderma: update on single nucleotide polymorphism analysis and microarrays. Curr Opin Rheumatol 2005;17:761-7.

[3] Zhou X, Tan FK, Xiong M, Arnett FC, Feghali-Bostwick C. Monozygotic twins clinically discordant for scleroderma show concordance for fibroblast gene expression profiles. Arthritis Rheum. 2005; 52: 3305-13.

[4] Agarwal SK. The genetics of systemic sclerosis. Discov Med. 2010;10:134-43.

[5] Dieudé P, Guedj M, Wipff J, Avouac J, Fajardy I, Diot E, et al. Association between the IRF5 rs2004640 functional polymorphism and Systemic Sclerosis. Arthritis Rheum. 2009;60**:**225-33.

[6] Dieudé P, Wipff J, Guedj M, Ruiz B, Melchers I, Hachulla E, et al. BANK1 is a genetic risk factor for diffuse cutaneous systemic sclerosis and has additive effects with IRF5 and STAT4. Arthritis Rheum. 2009;60:3447-54.

[7] Rueda B, Broen J, Simeon C, Hesselstrand R, Diaz B, Suárez H, et al. The STAT4 gene influences the genetic predisposition to systemic sclerosis phenotype. Hum Mol Genet. 2009;18:2071-7.

[8] Ito I, Kawaguchi Y, Kawasaki A, Hasegawa M, Ohashi J, Hikami K, et al. Association of a functional polymorphism in the IRF5 region with systemic sclerosis in a Japanese population. Arthritis Rheum. 2009;60:1845-50.

[9] Dieudé P, Guedj M, Wipff J, Avouac J, Hachulla E, Diot E, et al. The PTPN22 620W allele confers susceptibility to Systemic Sclerosis. Arthritis Rheum. 2008;58:2183-8

[10] Radstake TR, Gorlova O, Rueda B, Martin JE, Alizadeh BZ, Palomino-Morales R, et al. Genome-wide association study of systemic sclerosis identifies CD247 as a new susceptibility locus. Nat Genet. 2010;42:426-9.

[11] Zhou X, Lee JE, Arnett FC, Xiong M, Park MY, Yoo YK, et al. HLA-DPB1 and DPB2 are genetic loci for systemic sclerosis: a genome-wide association study in Koreans with replication in North Americans. Arthritis Rheum. 2009;60:3807-14.

[12]. Becker KG. The common genetic hypothesis of autoimmune/inflammatory disease. Curr Opin Allergy Clin Immunol. 2001;1:399-405.

[13] Chiocchetti A, Indelicato M, Bensi T, Mesturini R, Giordano M, Sametti S, et al. High levels of osteopontin associated with polymorphisms in its gene are a risk factor for development of autoimmunity/lymphoproliferation. Blood. 2004;103:1376-82.

[14] D'Alfonso S, Barizzone N, Giordano M, Chiocchetti A, Magnani C, Castelli L, et al. Two single-nucleotide polymorphisms in the 5' and 3' ends of the osteopontin gene contribute to susceptibility to systemic lupus erythematosus. Arthritis Rheum. 2005;52:539-47.

[15] Chiocchetti A, Comi C, Indelicato M, Castelli L, Mesturini R, Bensi T, et al. Osteopontin gene haplotypes correlate with multiple sclerosis development and progression. J Neuroimmunol. 2005;163:172-8.

[16] Denhardt DT, Noda M, O'Regan AW, Pavlin D, Berman JS. Osteopontin as a means to cope with environmental insults: regulation of inflammation, tissue remodeling, and cell survival. J Clin Invest. 2001;107:1055–61.

[17] O'Regan AW, Nau GJ, Chupp GL, Berman JS. Osteopontin (Eta-1) in cell-mediated immunity: teaching an old dog new tricks. Immunol Today 2000;21:475–78.

[18] Ashkar S, Weber GF, Panoutsakopoulou V, Sanchirico ME,Jansson M, Zawaideh S, et al. Eta-1 (osteopontin): an early component of type-1 (cell-mediated) immunity. Science. 2000;287:860-64.

[19] Iizuka J, Katagiri Y, Tada N, Murakami M, Ikeda T, Sato M, et al. Introduction of an osteopontin gene confers the increase in B1 cell population and the production of anti-DNA autoantibodies. Lab Invest. 1998;78:1523–33.

[20] Pardo A, Gibson K, Cisneros J, Richards TJ, Yang Y, Becerril C, et al. Up-regulation and profibrotic role of osteopontin in human idiopathic pulmonary fibrosis. PLoS Med. 2005;2:891-903.

[21] Takahashi F, Takahashi K, Okazaki T, Maeda K, Ienaga H, Maeda M, et al. Role of osteopontin in the pathogenesis of bleomycin-induced pulmonary fibrosis. Am J Respir Cell Mol Biol. 2001;24:264-71

[22] Han S, Guthridge JM, Harley IT, Sestak AL, Kim-Howard X, Kaufman KM, et al. Osteopontin and Lupus Erythematosus Association: a probable gene-gender interaction. PLoS ONE. 2008;3:1-8.

[23] Xu AP, Bai J, Lü J, Liang YY, Li JG, Lai DY, et al. Osteopontin gene polymorphism in association with systemic lupus erythematosus in Chinese patients. Chin Med J 2007;120:2124-8).

[24] Subcommittee for Scleroderma Criteria of the American Rheumatism Association Diagnostic and Therapeutic Criteria Committee. Preliminary criteria for the classification of systemic sclerosis (scleroderma). Arthritis Rheum. 1980;23:581–90.

[25] LeRoy EC, Black C, Fleischmajer R, Jablonska S, Krieg T, Medsger TA Jr, et al. Scleroderma (systemic sclerosis): classification, subset and pathogenesis. J Rheumatol. 1988;15:202-5.

[26] Dudbridge F. Pedigree disequilibrium tests for multilocus haplotypes. *Genet Epidemiol* 2003; **25**: 115-121

[27] Marciano R, D'Annunzio G, Minuto N, Pasquali L, Santamaria A, Di Duca M, et al. Association of alleles at polymorphic sites in the Osteopontin encoding gene in young type 1 diabetic patients. Clin Immunol. 2009;131:84-91.

[28] Niino M, Kikuchi S, Fukazawa T, Yabe I, Tashiro K. Genetic polymorphisms of osteopontin in association with multiple sclerosis in Japanese patients. J Neuroimmunol. 2003;136:125-9.

[29] Chiocchetti A, Orilieri E, Cappellano G, Barizzone N, D' Alfonso S, D' Annunzio G, et al. The osteopontin gene +1239A/C single nucleotide polymorphism is associated with type 1 diabetes mellitus in the Italian population. Int J Immunopathol Pharmacol. 2010;23:263-9.

[30] Anaya JM, Gòmez L, Castilblanco J. Is there a common genetic basis for autoimmune diseases? Clin Dev Immunol 2006;13:185-95.

[31] Zhernakova A, van Diemen CC, Wijmenga C. Detecting shared pathogenesis from the shared genetics of immune-related diseases. Nat Rev Genet 2009;10:43-55.

[32] Hudson M, Rojas-Villarraga A, Coral-Alvarado P, Lòpez-Guzmàn S, Mantilla RD, Chalem P et al. Polyautoimmunity and familial autoimmunity in systemic sclerosis. J Autoimmun 2008;31:156-9.

[33] Aune TM, Maas K, Parker J, Moore JH, Olsen NJ. Profiles of gene expression in human autoimmune disease. Cell Biochem Biophys 2004;40:81-96

Table 1. Clinical and immunologic features of the patients with SSc

Feature		% of patients
SSc subset:	Diffuse cutaneous	23
	Limited cutaneous	77
Clinical manifestations:	Fibrosis	42
	Pulmonary artery hypertension (PAH)[a]	13
Autoantibodies:	ACA	40
	anti-topo I	38

[a]defined as a right-ventricular systolic pressure ≥ 40 mmHg;

ACA=anticentromere antibodies; anti-topo I = anti-topoisomerase I antibodies

Table 2 Gene and genotype frequencies of -156 and +1239 SNPs in patients and in controls

SNP	Frequency		p	OR (95% CI)
	Patients n=357	Controls n=864		
-156[a]	genotype			
11	0.52	0.45	0.032	1.32 (1.02-1.70)
01	0.40	0.42	0.42	0.89 (0.69-1.16)
00	0.08	0.13	0.045	0.64 (0.41-0.99)
	allele			
1	0.72	0.66	0.0086	1.30 (1.07-1.58)
0	0.28	0.34		
+1239[a]	genotype			
11	0.12	0.09	0.13	1.38 (0.91-2.08)
01	0.48	0.40	0.010	1.40 (1.08-1.80)
00	0.40	0.51	0.00046	0.64 (0.49-0.82)
	allele			
1	0.36	0.29	0.00064	1.39 (1.15-1.67)
0	0.64	0.71		

[a]Susceptibility alleles (increased in the patients: -156G and +1239C) are named 1, while protective alleles (decreased in the patients: -156GG and +1239A) are named 0.

Overall p values for genotype distributions in patients vs. controls were 0.030 and 0.0015 for –156 and +1239 SNPs respectively.

Table 3 Haplotype frequencies

Haplotype[a]		Patients (N=357)	Controls (N=864)	p	OR (95% CI)
-156	+1239				
0	0	0.268	0.322	0.010	0.77 (0.63-0.94)
1	0	0.372	0.389	0.48	0.93 (0.78-1.12)
0	1	0.015	0.016	0.79	0.85 (0.38-1.83)
1	1	0.345	0.273	0.00036	1.41 (1.17-1.71)

[a]Susceptibility alleles (increased in the patients: -156G and +1239C) are named 1, while protective alleles (decreased in the patients: -156GG and +1239A) are named 0.

Rare variants in the TREX1 gene and susceptibility to autoimmune diseases

Introduction

Trex1 (DNase III) is a 3'-5' exonuclease involved in response to oxidative stress and apoptosis. It is active as a homodimer and normally associates to the endoplasmatic reticulum, as part of a protein complex (SET complex) that is traslocated to nucleus in response to granzyme A-mediated cell death. Moreover Christmann et al. has recently demonstrated that Trex1 protein is up-regulated and translocated to nucleus in cells treated with UV light and that murine trex1 -/- cells are more susceptible to DNA damage caused by UV. Trex1 knockout mice develop an inflammatory myocarditis which results in progressive cardiomyopathy, circulatory failure and reduced survival (Morita et al. 2004).

Heterozygous mutations in *TREX1* gene have previously been observed in 9/417 patients with Systemic Lupus Erithematosus (SLE) and in 1/169 subjects with Sjögren's Syndrome (SS) (Lee-Kirsh, 2007). Subsequently de Vries et al. found an additional heterozygous missense variant in one out of 50 patients affected by neuropsychiatric-SLE (NPSLE). Those variations had not been found in 1712 controls (Lee-Kirsh, 2007).

Heterozygous mutations of *TREX1* have also been described in Familial Chilblain Lupus (FCL), a rare autosomal-dominant form of cutaneous lupus erythematosus manifesting in early childhood (Hedrich et al. 2008) and in Retinal Vasculopathy with Cerebral Leukodistrophy (RVCL), an adult-onset disease characterized by central nervous system degeneration, retinal vasculopaty and nephropathy (Richards et al. 2007). Furthermore mutations of *TREX1* cause Aicardi-Goutières Syndrome (AGS), a genetically heterogeneous, autosomal-recessive disorder presenting with early-onset progressive encephalopathy (Ramantani, …). It has been reported that some AGS patients shows a partial clinical overlap with lupus erythematosus. Particularly, AGS patients may present lupus-like rash, oral ulcers, arthritis, thrombocytopenia, leukocytopenia and presence of ANAs (Ramantani).

Altogether these data highlight a relevant role of TREX1 in autoimmune rheumatollogical diseases. On these bases we decided to investigate the role of *TREX1* gene in susceptibility to three different autoimmune diseases in the Italian population: SLE, SS and Systemic Sclerosis (SSc). In particular, this work is aimed at replicating in the Italian population the involvement of TREX1 in SLE and SS and to extend the investigation to a further autoimmune disease. The latter was included in the study because it shares some features in common with SLE, namely the fact that they are both connectivitis and the presence of anti-nuclear autoantibodies.

Patients and Methods

Subjects. In this study we performed a mutational screening on 210 SLE, 58 SS and 150 SSc patients. Furthermore we analyzed 200 healthy control subjects, comprising blood donors. The geographic origin of controls and patients of the three sample sets was similar. All subjects (patients and controls) belong to the Italian population and are unrelated. Individuals with Sardinian ancestors were excluded. All patients provided their informed consent.

All SLE patients (female: male ratio = 7.5:1) fulfilled at least 4 of the American College of Rheumatology 1997 revised criteria for the classification of SLE (Hochberg). Enrolment was completed when the patients provided their informed consent.

All SSc patients (female: male ratio = 13:1) fulfilled the American College of Rheumatology preliminary criteria for the classification of SSc (Subcommittee for Scleroderma Criteria) The 16% of patients was affected by the diffuse cutaneous (dcSSc) form of the disease, and 84% by the limited cutaneous (lcSSc) form. Pulmonary hypertension, estimated by echocardiography, was defined as a right-ventricular systolic pressure ≥ 40 mmHg and was observed in 29% of patients. Pulmonary fibrosis was evaluated by high resolution TC (HR- TC) and was present in 36% of SSc subjects. Disease onset was determined by the patient's recall of the first non-Raynaud symptom clearly ascribable to scleroderma.

SS patients (female:male ratio = 12:1) fulfilled the revised criteria proposed by the American-European Consensus Group (Vitali et al. 2002)

Search for new variations in TREX1 gene. Genomic DNA was isolated from peripheral blood nuclear cells with standard methods. The entire coding sequence of *TREX1* gene (945 bp) is contained in one single exon. It was amplified in three overlapping fragments, and analysed by Denaturing High Performance Liquid Chromatography (DHPLC) or direct sequencing. DHPLC temperatures of analysis have been calculated with the algorithm "DHPLC melt program" (http://insertion.stanford.edu/melt). PCR product have been exposed to rapid thermal denaturation and slow renaturation on a thermal cycler, then analyzed on TRANSGENOMIC WAVE (Transgenomic, Omaha, NE, USA). Results were analysed with NavigatorTM software (Transgenomic, Omaha, NE, USA). Primer sequences and analysis conditions are available on demand. PCR products were purified for sequencing application with a vacuum aspiration system (PCR96 Cleanup kit-Millipore, Billerica, MA, USA). Sequencing reactions were performed using the ABI PRISM® BigDye Terminator kit v.1.1 (Applied Biosystems, Foster City, CA, USA), purified with the Montage SEQ96 Sequencing Reaction Cleanup Kit (Millipore, Billerica, MA, USA) and analysed with the 3100 Genetic Analyser automatic sequencer (Applied Biosystems, Foster City, CA, USA).

In silico analysis. Putative functional relevance of the missense variations was evaluated with the programs PolyPhen (http://genetics.bwh.harvard.edu/pph), SIFT (http://blocks.fhcrc.org/sift/SIFT), SNAP (http://cubic.bioc.columbia.edu/services/SNAP/submit) and pMUT (http://mmb2.pcb.ub.es:8080/pMut/). Putative effect of the variations on splicing sites was evaluated using the SpliceView program (http://www.itb.cnr.it/sun/webgene/) and the ESEfinder scoring matrix (http://www.rulai.cshl.edu/tools/ESE).

Results

We searched for new variants of *TREX1* gene in 210 SLE, 58 SS and 150 SSc patients. We detected 7 single nucleotide variants, located in the coding sequence. Two of them were non synonymous: c.592G>A (p.Glu198Lys) and c.694A>G (p.Met232Val). Numeration is referred to the sequence NM_022517.17, (ATG=1) from NCBI public database. Both the substitutions have been observed only once, in one SS (p.Glu198Lys) and one SSc (p.Met232Val) heterozygous patient, respectively. The SSc patient was affected by the limited form of the disease, and she shows pulmonary hypertension and anti-centromer auto-antibodies, age of onset = 49. The p.Glu198Lys variant has recently been reported in literature in homozygosity in one AGS patient (Ramantani et al. 2010), while p.Met232Val is novel.

An in silico analysis has been performed to predict the possible functional role of the two missense variations. We used four different prediction algorithms (PolyPhen, SIFT, SNAP and pMUT). Both the substitutions seem to have a possibly damaging effect according to all the four software. The p.Glu198Lys variant is located in the highly conserved third exonuclease (Exo3) domain, one of the three regions that, together with Exo1 and Exo2, form the catalytic core of the protein. Instead p.Met232Val falls in a region of uncertain function, between the domain Exo3 and the membrane binding region TMH.

The presence of the two observed missense variants was searched in 200 geographically-matched healthy controls. None of them carried these variations, nor had the two substitutions been observed in further 1712 controls from literature, sequenced by Lee-Kirsch et al. (2007).

Moreover, we observed other 5 variants , all of them synonymous (c.144 C>G, c.198 G>A, c.462 T>C, c.531 C>T and c.912 G>A). Only c.144 C>G (p.Pro48Pro) is novel, while the others have already been described in healthy controls (Lee-Kirsch, et al, 2007) and on public databases. The c.531C>T - rs11797 (p.Tyr177Tyr) variant is a common polymorphism of *TREX1* gene and it was observed with an allelic frequency of 0,42-047 in the three sample sets, not significantly different from the frequencies reported in literature (0.42, Lee-Kirsh 2007) and on public databases (0.38,

NCBI). The other four are instead rare variants: in this study c.198G>A (p.Lys66Lys) and c.462 T>C (p.Asp154Asp) have been observed in the same SS patient at heterozygous form, c.912 G>A (p.Leu304Leu) in two heterozygous patients (affected by SS and SLE respectively) and c.144 C>G in one heterozygous SLE patient, not dissimilarly from what reported by Lee-Kirsh or in public databases.

According to Splice View algorithm none of the variants affects canonical splice sites. Conversely, the analysis performed with ESEfinder showed thatc.144 C>G, c.592 G>A and c.912 G>A might have some functional effect by introducing or removing a site of interaction with splicing proteins. Particularly c.144 C>G adds new interacting sites for the proteins SF2/ASF, SF2/ASF(IgM BRCA1) and SR35; c.592 G>A removes two existing interacting sites for SF2/ASF and SF2/ASF(IgM BRCA1) and c.912 G>A removes an existing site for SF2/ASF(IgM BRCA1). Furthermore, the analysis of the codon usage frequencies showed that the two variants 144 C>G and c.912 G>A both cause the introduction of a rarer codon (144 C>G: 19.8‰ vs. 6.9‰ c.912 G>A: 39.6‰ vs. 7.2‰).

DISCUSSION

This work stems from recent papers which strongly suggest a role of TREX1 in autoimmune diseases. In this study we analysed patients affected by three different autoimmune diseases, namely SLE, SS and SSc belonging to the Italian population. Interestingly, the role of TREX1 in SSc was analysed for the first time.

The present study identified two *TREX1* mutations, each in one patient, which were never observed in the 200 Italian controls analysed in this study nor in 1712 controls from literature, fully sequenced for TREX1 coding region by Lee-Kirsch et al, 2007.

One of them is a novel *TREX1* mutation (p.Met232Val) identified in an SSc patient affected by the limited form of the disease. This substitution affects a conserved nucleotide, and in silico analysis performed with four prediction algorithms predicts for it a non-neutral role. This position is not

located in a functional region, but falls in the protein portion comprised between the exonucleasic domain and the membrane binding domain TMH. Several mutations, both frameshift and non synonymous, located in the same protein portion have been reported in SLE and in RVCL patients (figure 1). Although most of them were frameshift variants, missense mutations in this region (p.Gly227Ser, p.Arg240Ser and p.Ala247Pro) have been described in two Afro-Caribbean patients with SLE and secondary SS (p.Gly227Ser and p.Ala247Pro were found on the same allele) (Lee-Kirsch). These variants were not functionally characterized, however we can hypothesize a possible effect of p.Met232Val on Trex1 subcellular targeting, or on protein folding. However this hypothesis needs confirmation with functional analysis, which is our next goal.

The second mutation (p.Glu198Lys) was observed in a SS patient. This variant has previously been reported in homozygosity in one 19 year old Turkish AGS patient with lupus-like features (chilblain lesions on feet, oral ulcers, ANA, anti-ENA and decreased C3 levels (Ramantani et al.). Clinical informations about the parents were not available. The substitution affects the functional domain Exo3, and it is predicted as damaging by the in silico analysis. The importance of this sequence in the functionality of the enzyme is confirmed by the fact that several other mutations affecting Exo3 domain have been reported in AGS patients, both in homozygous and in heterozygous form (fig. 1). In vitro functional analysis have been performed for some of these mutations (p.Asp200Asn, p.Asp201ins and p.Val201Asp), demonstrating a dramatic decrease (up to 35,000 fold for p.Asp201ins) in the exonuclease activity (Udesh de silva 2007, Lehtinen 2008). Intriguingly, both the mutations involving the Asp-200 seem to be sufficient, at heterozygous form, to cause AGS, usually a dominant disorder. A dominant negative effect can be hypothesized since Lehtinen et al. demonstrated with experiments of co-expression in *E. Coli* cells, protein purification and with an exonuclease assay that $TREX1^{200Asn}$ inhibits the DNA degradation even in the presence of $TREX1^{wt}$. In fact the $TREX1^{wt/200Asn}$ heterodimer dsDNA activity showed a 200-fold decrease if compared to $TREX1^{wt}$ dsDNA activity. Nuclease activity on ssDNA was not affected, as it showed

only a slight decrease (1,5 fold). The authors suggested that TREX1^{200Asn} acts with a dominant negative effect, inhibiting TREX1^{wt} DNA degradation.

Considering the physiological role of TREX1 protein it is tempting to speculate that defective TREX1 may result in the failure to degrade ssDNA or dsDNA leading to immune activation and development of autoantibodies against these macromolecules.

This study confirm the presence of TREX1 mutation in patients affected by SS with a frequency (1/58) not dissimilar to that reported in the literature (1/169). Moreover the finding of a mutation, previously reported as deleterious, in a SS patient strengthens the hypothesis of a *TREX1* involvement in this disorder.

Moreover, this study identified for the first time a TREX1 mutation in one SSc patient. TREX1 mutations seem rare in this disease (1/158), nevertheless this finding suggests an involvement of this gene in SSc which deserves a confirmation in further studies. Conversely, we did not observe any TREX1 mutations in 210 SLE patients thus not replicating in the Italian population the results observed in 9/417 patients from Northern Europe (UK, Germany and Netherlands) where a frequency of 0,021 have been reported. A different genetic background characterizing the Italian population or a clinical heterogeneity among the Italian and Northern Europe cohorts might be likely explanations of this inconsistency

In conclusion, this study contributes to the demonstration that TREX1 is involved in autoimmune diseases and proposes that the spectrum of involved autoimmune diseases can be broader since we detected a previously unreported possible association between *TREX1* mutations and SSc. Moreover, our study confirms that rare private variants play a role in the susceptibility of multifactorial diseases.

Figure legends:

Figure 1 Schematic of disease-associated *TREX1* mutations. The two mutations found in the present study are in squares. Normal type: AGS, *italic: RVCL*, underlined: FCL, **boldface: SLE**, ***boldface and italic: SSc***, **boldface and starred(*): SS.** Exo 1, 2, 3 domains; PII: polyproline II domain; TMH: trans-membrane domain.

REFERENCES

Ramantani G, Kohlhase J, Hertzberg C, Innes AM, Engel K, Hunger S, Borozdin W, Mah JK, Ungerath K, Walkenhorst H, Richardt H, Buckard J, Bevot A, Siegel C, von Stülpnagel C, Ikonomidou C, Thomas K, Proud V, Niemann F, Wieczorek D, Häusler M, Niggemann P, Baltaci V, Conrad K, Lebon P, Lee-Kirsch MA. Expanding the phenotypic Spectrum of Lupus Erythematosus in Aicardi-Goutières Syndrome. *Arthritis e Rheumatism* 2010; 62(5):1469-1477

Christmann M, Tomicic MT, Aasland D, Berdelle N, Kaina B. Three prime exonuclease I (TREX1) is Fos/AP-1 regulated by genotoxic stress and protects against ultraviolet light and benzo(a)pyrene-induced DNA damage. *Nucleic Acids Research* 2010; advanced access 1-15

Lee-Kirsch M.A., Gong M., Chowdhury D., Senenko L., Engel K.,. Lee Y.-A, Udesh de Silva,. Bailey S.L, Witte T., Vyse T.J., Kere J., Pfeiffer C., Harvey S., Wong A., Koskenmies S., Hummel O., Rohde K., Schmidt R.E., Dominiczak A.F., Gahr M., Hollis T., Perrino F.W., Lieberman J. and Hübner N. Mutations in the gene encoding the 3'-5' DNA exonuclease TREX1 are associated with systemic lupus erythematosus. *Nature Genetics* 2007; 39, 1065-1067

De Vries B, Steup-Beekman G, Haan J, Bollen EL, Luyendijk J, Frants RR, Terwindt GM, van Buchen MA, Huizinga TWJ, van den Maagdenberg AMJM, Ferrari MD. TREX1 gene variant in Neuropsychiatric Systemic Lupus Erythematosus. *Ann Rheum Dis*

Morita M., Stamp G., Robins P., Dulie A., Rosewell I., Hrivnak G., Daly G., Lindahl T. and Barnes D.E. Gene-targeted mice lacking the Trex1 (DNase III) 3'→5' DNA exonuclease develop inflammatory myocarditis. *Molecular and Cellular Biology* 2004; 24, 6719-6727

Vitali C, Bombardieri S, Jonsson R, Moutsopoulos HM, Alexander EL, Carsons SE, Daniels TE, Fox PC, Fox RI, Kassan SS, Pillmer SR, Talal N, Weisman MH, European Study Group on Classification Criteria for Sjögren's Syndrome. Classification criteria for Sjögren's Syndrome: a revised version of the European criteria proposed by the American-European Consensus Group. *Ann Rheum Dis* 2002; 61, 554-558

Subcommittee for Scleroderma Criteria of the American Rheumatism Association Diagnostic and Therapeutic Criteria Committee. Preliminary criteria for the classification of systemic sclerosis (scleroderma). *Arthritis Rheum*. 1980; 23: 581–590.

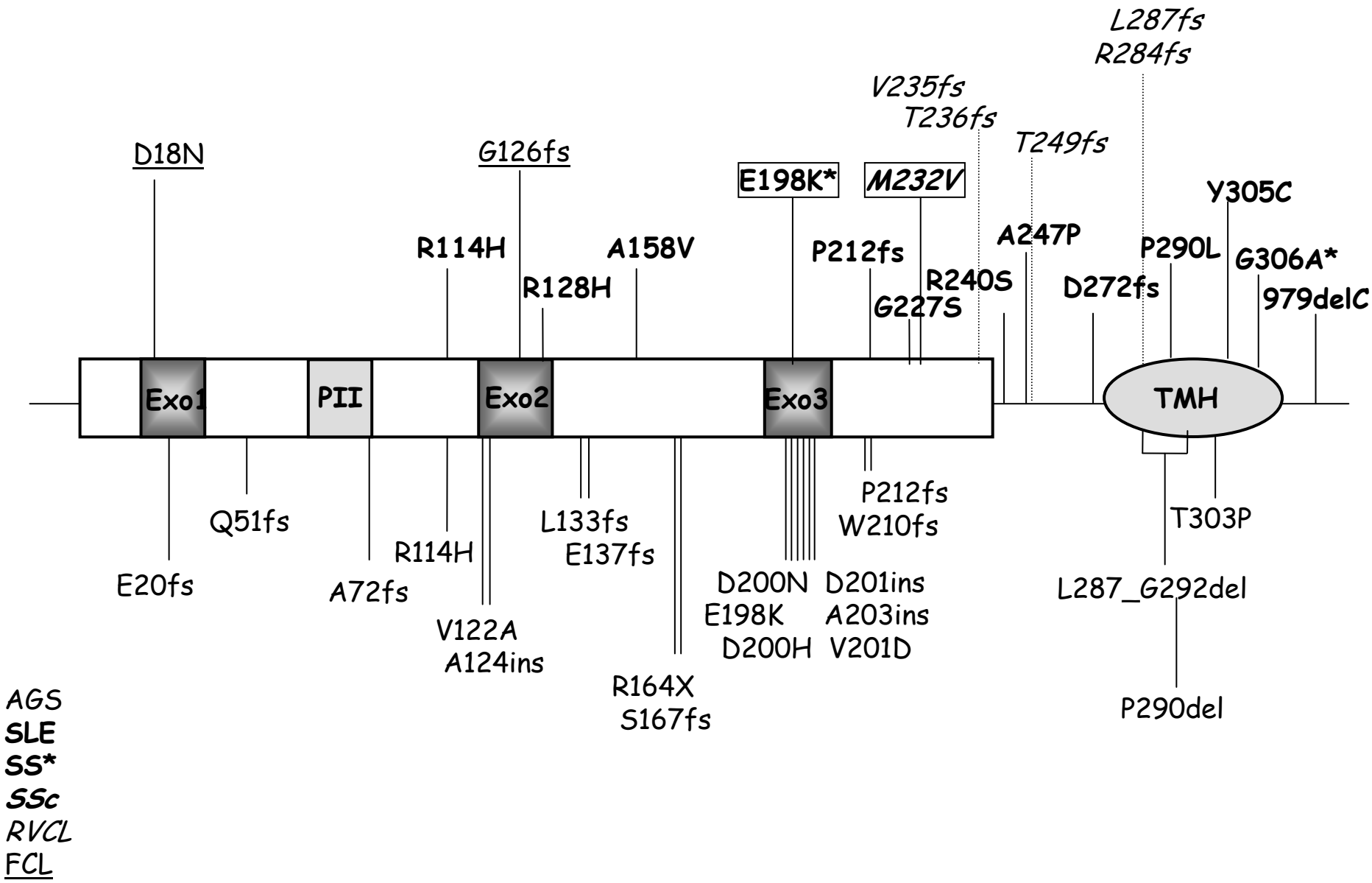
Hochberg MC. Updating the American College of Rheumatology revised criteria for the classification of systemic lupus erythematosus [letter]. *Arthritis Rheum* 1997; 40: 1725.

Hedrich CM, Fiebig B, Hauk FH, Sallman S, Hahn G, Pfeiffer C, Heubner G, Gahr M. Chilblain lupus erythematosus – a review of literature. *Clin Rheumatol* 2008

Richards A., Van den Maagdenberg A.M.J.M., Jen J.C., Kavanagh D., Bertram P., Spitzer D., Liezewski M.K., Barilla-LaBarca M.-L., Terwindt G.M., Kasai Y., Mclellan M.,. Grand M.G, Vanmolkot K.R.J., de Vries B., Wan J., Kane M.J., Mamsa H., Schäfer R., Stam A.H., Haan J., de Jong P.T.V.M., Storimans C.W., Van Schooneveld M.J., Oosterhuis J.A., Gschwendter A., Dichgans M., Kotschet K.E., Hodgkinson S., Hardy T.A., Delatycki M.B., Hajj-Ali R.A., Kothari P.H., Nelson S.F., Frants R.R., Baloh R.W., Ferrari M.D. and Atkinson J.P. C-terminal truncations in human 3'-5' DNA exonuclease TREX1 cause autosomal dominant retinal vasculopathy with cerebral leukodystrophy. *Nature Genetics* 2007: 39, 1068-1070

Lehtinen DA, Harvey S, Mulcahy MJ, Hollis T, Perrino FW. The TREX1 double-stranded DNA degradation activity is defective in dominant mutations associated with autoimmune disease. J Biol Chem. 2008;283(46):31649-56

Figure 1 *TREX 1* mutations



Genetic analysis of perforin in two autoimmune diseases: Systemic Lupus Erithematosus and Systemic Sclerosis

Introduction

Perforin (PRF1) is a 67 kDa protein composed by 555 amino acids, stored in secretory vesicles of cytotoxic T lymphocytes (CTLs) and of natural killer (NK) cells. It is involved in immune surveillance, as CTLs and NK cells recognize tumoral or virus infected cells, and disrupt them through granules release, by the activation of caspase mediated apoptosis. PRF1 polimeryzes on target cell membrane, and forms pores which allow the entry of granzymes into the cells. Granzymes are a class of serine proteases that trigger apoptosis through cleavage of caspases. This cytotoxicity is crucial to kill virus-infected cells, but may also be involved in down-modulation of the immune response by fratricide of activated immune cells.

Biallelic loss of function mutations of the *PRF1* gene have been associated with familial haemophagocitic lymphohistiocytosis (FHL2), a rare, life-threatening immune deficiency that occurs in pediatric age, characterized by fever, hepatosplenomegaly and thrombocytopenia. FHL2 has been classically ascribed to decreased capacity of CTL and NK cells to clear viral infections; viral persistence is thought to cause the lymphoproliferative pattern. FHL2 is a recessive disease and subjects carrying heterozygous *PRF1* mutations are generally healthy. However, some heterozygous variations may favour development of autoimmune diseases.

This has been initially suggested for the autoimmune lymphoproliferative syndrome (ALPS), and Dianzani Autoimmune Lymphoproliferative Disease (DALD) where an increased frequency of two functional variants (p.Asn252Ser and p.Ala91Val respectively) was observed (Clementi,….. Dianzani U. et al. 2006). ALPS is a rare pediatric autoimmune disease due to defective function of the Fas death receptor involved in both down-modulation of the immune response and cell-mediated cytotoxicity. [11] It is primarily due to mutations of the *FAS* gene (*TNFRSF6*) or other genes involved in Fas function, but other genetic factors may concur. Subsequently, findings similar to those already reported for ALPS and DALD have been observed also for Type 1 Diabetes (Orilieri et al. 2008) and Multiple Sclerosis (Cappellano et al. 2008). In these two diseases new PRF1 variants have also been reported, and the authors proposed for them a functional role (Orilieri et al. 2008, Cappellano et al. 2008).

In this paper we investigated the role of *PRF1* variations in the susceptibility to two autoimmune diseases: *Systemic Lupus Erithematosus* (SLE) and *Systemic Sclerosis* (SSc). The two pathologies share some features, such as the presence of anti-nuclear antibodies and connective tissue damage. SLE is a systemic autoimmune disease with multifactorial etiology, characterized by a wide

variability in the spectrum of clinical features, including malar rash, joint inflammation, nephritis, serositis and hematologic disorders. SSc is a rare multisystemic disease characterized by widespread fibrosis involving both skin and internal organs such as lung and hearth, activation of immune cells, production of anti-nuclear autoantibodies (ANA), and injury to vascular and microvascular structures.

Results

Analysis of the whole coding region of PRF1.

The entire coding region (1665 nucleotides) of the *PRF1* gene (chr.10q22.1) was analyzed by DHPLC and/or direct sequencing in 154 SLE patients, 131 SSc and 559 controls to look for novel, disease-associated, variations in the *PRF1* gene. We found a total of 22 variations (11 missense and 11 synonymous; table 2). Eight were novel variants, 4 of which predicted to be non-synonymous substitutions. All of them were detected in heterozygous individuals.

In addition to these novel variations we observed 14 previously described variants: 7 non synonymous and 7 synonymous.

Among the non synonymous substitutions, c.403G>A, responsible for the amino acid substitution Val135Met has been reported on public databases (rs12263572) with a minor allele frequency (MAF) of 0.014 in the CEU population (http://hapmap.ncbi.nlm.nih.gov/). In our sample set it was observed only in one SLE patient. The c.11G>A - rs35418374 variation causes the amino acid change Arg4His in the signal peptide and it was detected, in our sample set, only in one SSc patient. In public databases it is reported with a MAF of 0.013 in a population of Caucasoid and Afro-American apparently healthy subjects (AGI-ASP, http://www.ncbi.nlm.nih.gov). The substitutions p.Val329Ile and p.His514Arg are instead novel variants, observed in one SLE patient each. All the other non synonymous variations were found also in our control sample set (table 2). Two of these (c.272C>T - rs35947132 and c.755A>G - rs28933375) are polymorphisms of the *PRF1* gene for whom a functional role has already been described (see discussion), and they cause the substitutions Ala91Val and Asn252Ser respectively.

The two synonymous variations (p.Ala274Ala, p.His300His; not reported in Table 2) have been previously reported as common, not disease-associated polymorphisms (Cappellano et al.). They do not change the amino acid, nor influence the splice sites. Allelic frequencies of the two polymorphisms did not significantly differ between controls and SLE or SSc patients and therefore they have not been further analysed in this work. Two other synonymous variations (c.435G>A and c.462A>G) are known to be in perfect linkage disequilibrium (r^2=1) with c.755A>G (p.Asn252Ser),

and were in fact detected only in subjects carrying this variant. They have not been considered in further analyses.

In silico analysis.

Four different algorithms (PolyPhen, SIFT, SNAP and PMUT) were used to predict the functional relevance of the detected non synonymous variations (Table 3). The p.Arg232His, p.Val329Ile and p.Gly334Ser substitutions lie in the membrane attack complex domain (MACPRF, Risma et al. 2006), in particular p.Arg232His falls in a structurally conserved amphipathic α-helix region which spans between amino acids 212-241, homologous to the complement components C5b to C9 (Voskoboinik et al.2006); the variations p.Phe421Cys, p.Ala437Val and p.His514Arg are located in a protein region homologous to the C2 domain of PKC. This is a calcium-binding domain, and is responsible for perforin membrane binding. The other variations do not fall in a functional domain of the PRF1 protein. Only one substitution (p.His514Arg) was predicted as non tolerated or pathological by three of the four utilised programs. In fact, alignment with ortholog and paralog proteins showed that the histidine at position 514 is conserved among species.

To assess the possible functional relevance of the p.Arg4His substitution located in the signal peptide we also used the program Signal P 3.0. The variation does not seem to modify the predicted cleavage site (calculated between amino acid positions 21 and 22, in accordance with SPdb database - http://proline.bic.nus.edu.sg/spdb) or the signal prediction probability, assessed considering the percentage of variation of the D score value (0.12%), as suggested by Jarjanazi et al (2007).

Synonymous variations were evaluated for their influence on splicing through the software SpliceView, which predicts the introduction or the removal of donor or acceptor splice sites and through the algorithm ESEfinder, which predicts the introduction or the removal of sites of interaction with various arginine/serine rich splicing factors (ESE: Exonic Splicing Enhancer). Additionally we considered whether or not these variants coded for more or less common codons in respect to those normally used. Among the synonymous variants observed only in patients c.807C>T and c.999C>T may have an effect on splicing through interactions with splicing factors: C807T creates a novel binding site for SRp35 and c.999C>T removes an existing binding site for SRp40. Furthermore c.273G>A creates a novel binding site for SRp40, c.1620A>G creates a novel binding site for SF2/ASF, while c.1356C>T creates a novel binding site for two different proteins (SF2/ASF and SC35), but removes an existing binding sites for SRp40. Besides c.1620A>G may introduce a novel putative acceptor splice site (score 0,53).

Functional Analysis. We checked the functional relevance of the p.His514Arg and p.Val135Met variants by evaluation of the NK activity, perforin expression and percentage of NK cells in peripheral blood. Apparently the two variants do not affect NK activity, but we detected a slight decrease (below 5^{th} percentile range) in the proportion of PRF1 positive cells for the patient carrying the p.His514Arg variant. Also, the patient with the p.Val135Met variation showed low levels of $CD3^{-}CD16^{+}$ NK cells (table 4).

Clinical parameters analysis.

Clinical parameters of SLE and SSc patients which carry rare non synonymous *PRF1* variants are listed in tables 5A and 5B. Three out of four (75%) SSc patients carrying a non synonymous variation were positive for anti-topoisomerase I antibodies. The frequency of this immunologic parameter in our population was 39% (Table 1B). This difference was not statistically significant due to the small number of samples. Clinical data were not available for the SLE patient carrying the p.Val329Ile variation.

Relevance for disease susceptibility.

The overall allelic frequency of all considered *PRF1* variations was higher both in SLE (0.12) and in SSc patients (0.12) than in healthy controls (0.084), and this difference was statistically significant for both diseases (SLE: p=0.0038, OR=1.89 95%CI=1.22-2.94; SSc: p=0.0028, OR=1.99 95%CI=1.25-3.16). When considering only non synonymous variants, allelic frequencies were still different between healthy controls (0.062) and the two disease sample sets (SLE=0.097; SSc=0.11), although the statistical significance decreased (SLE: p=0.043 OR=1.63 95%CI=1.01-2.61, SSc: p=0.0082 OR=1.89 95%CI=1.17-3.05). Part of this variability was due to the two most frequent non synonymous variants, p.Ala91Val and p.Asn252Ser. The same analysis was repeated after excluding them from the analysis and the allelic frequencies (SLE=0,029; SSc=0,027; controls=0,0081) were still statistically different (SLE: p=0,00693 OR=3,71 95%CI=1,29-10,64; SSc: p=0,0198 OR=3,38 95%CI=1,06-10,30). When considering only non synonymous variants minus p.Ala91Val and p.Asn252Ser the differences were not significant due to small numbers, but a suggestive trend was still visible (SLE=0,0097; SSc=0,015; controls=0,0045).

The 91Val allele was detected in 21 SLE patients (all heterozygous), 21 SSc patients (all heterozygous) and 57 controls (53 heterozygous and 4 homozygous); 252Ser was detected in 6 SLE patients, 4 SSc patients and 3 controls, all heterozygous. The allelic frequency of the 252Ser variation in healthy controls (0.0027) was significantly lower from that observed both in SLE

(0.019, p=0.0045, OR=7.34, 95%CI=1.64-37.41) and in SSc (0.015, p=0.027, OR=5.76, 95%CI = 1.09-32.54) patients.

As p.Ala91Val and p.Asn252Ser are the most common non synonymous variants detected in the *PRF1* genomic sequence, and because of the statistically significant difference observed for both diseases with p.Asn252Ser allelic frequency distribution, we decided to extend the analysis of these two polymorphisms in further patients and geographically-matched healthy controls for a total of 597 SLE patients, 204 SSc patients and 1856 controls (Table 6). The 252Ser allele had a significantly increased frequency among SLE patients versus controls (p=0.039 OR=2.35 95%CI=1.04-5.24). Conversely, the 91Val allele was significantly increased among SSc patients (p=0.0054 OR=1.70 95%CI = 1.16-2.47). Genotype distributions did not deviate from Hardy-Weinberg equilibrium. One of the SLE patients was heterozygous for both p.Ala91Val and p.Asn252Ser variants: allele-specific PCR established that 91Val and 252Ser were carried by two distinct alleles (compound heterozygosity). This genotype combination is not as yet reported in the *PRF1* mutations database (http://bioinf.uta.fi/PRF1base/prf1pub).

We also tested the correlation of allelic frequencies of p.Ala91Val and p.Asn252Ser variations with the clinical and serological parameters listed in table 1A and 1B. The only statistically significant result was a slight decrease in 91Val allele in SLE patients positive for serositis (showing pleuritis and/or pericarditis) (0.038 vs. 0.082, p=0.022, OR=0.45, 95%CI=0.22-0.90). However this result doesn't withstand Bonferroni's correction for multiple testing.

Discussion

This study was aimed to investigate the role of *PRF1* variants in autoimmune susceptibility. Our group has already contributed to define a possible association with type 1 diabetes (T1D) (Orilieri et al. 2008) and Multiple Sclerosis (MS) (Cappellano et al. 2008). Here we extend the analysis to two autoimmune systemic connectivitis, namely Systemic Lupus Erithematosus and Systemic Sclerosis.

The two most frequent non synonymous variations in *PRF1* gene are p.Ala91Val and p.Asn252Ser. Both this variants have previously been observed in association with autoimmune diseases, in particular ALPS and T1D (p.Asn252Ser), DALD and MS (p.Ala91Val). In this work we observed an association of *PRF1* allelic frequencies also with SLE (p.Asn252Ser) and SSc (p.Ala91Val), though we admit that the latter may be biased by the small number of the samples, due to the rarity of the disease (prevalence in Caucasians: 2,6 cases/10,000 Arnett et al.2001). Studies based either on human (analysis of cytotoxic limphocytes from 91Val carriers) or mice (rat basophile leukaemia cells transfected with variants of the *PRF1* cDNA) models have shown a functional relevance for the p.Ala91Val substitution, which causes a conformational change that results in a decrease in

protein stability and NK activity (about 50% of the wild-type perforin) (Trambas et al. 2005, Voskoboinik et al. 2005, Risma et al. 2006). The functional meaning of p.Asn252Ser variation is instead debated. It is located in the membrane attack complex domain, a critical region for the pore-forming function of the protein. However, several works have associated this variant with normal perforin expression and NK function (Stepp et al. 1999, Voskoboinik et al. 2005, Risma et al. 2006). By contrast our group has recently described both ALPS and T1D pediatric patients heterozygous for p.Asn252Ser substitution with low NK activity in their early childhood (Clementi et al. 2006; Orilieri et al. 2008). In the ALPS patient NK function has reverted to normal levels by the age of 12 years (Clementi et al. 2006). Therefore it is possible that p.Asn252Ser variation, or another associated variant, may decrease NK function in early childhood, but that later on this might be normalized by a compensatory mechanism. It does not seem very likely that this putative mechanism may be involved in susceptibility to adult onset diseases such as SLE and SSc, however an adult patient homozygous for 252Ser allele with low NK activity has also been described (Orilieri et al. 2008). p.Asn252Ser is known to be in perfect LD (r^2=1) with two synonymous variations (c.435G>A and c.462A>G), and the same finding was confirmed also in our population. It is also possible that these or other unknown variants in LD with p.Asn252Ser located in the regulative regions of the gene may have an effect on protein function. Both c.435G>A and c.462A>G nucleotide substitutions change a common codon with a rare one, and it has been demonstrated that this may lead to anomalies in protein folding (Komar 2007).

Besides p.Ala91Val and p.Asn252Ser, other 9 non synonymous substitutions were found. Apart from p.Ala91Val and p.Asn252Ser, the other variations are to rare to allow a statistical analysis of their individual association with the two diseases but, paralleling what previously observed with MS, the overall incidence of *PRF1* missense variations is higher in patients than in controls, both for SLE (p=0.043, OR = 1.63) and for SSc (p=0.0082 OR = 1.89).

p.Arg4His variation is located in the signal peptide. In our cohort it was found in a single SSc patient with limited disease and pulmonary artery hypertension. It has previously been detected at heterozygous state in a 21-year-old African American patient with aplastic anemia by Solomou et al. 2007, who observed it also in 21/1156 healthy controls (allelic frequency=0,009), and in a European patient with systemic onset juvenile idiopathic arthritis (SoJIA) (Vastert et al. 2008). It has also been reported with similar allelic frequency in public databases (AGI-ASP population). Perforin is secreted as inactive protein and the NK activity is activated by proteolytic cleavage. According to in silico analysis, this variant does not seem to affect peptide cleavage or protein transportation. Therefore p.Arg4His is probably a rare *PRF1* polymorphism that doesn't affect protein function.

Other 3 substitutions were found in one SLE patient each and they have never been previously described in literature, although p.Val135Met is reported on pubblic databases with an allelic frequency of 0,014 (rs12263572). The p.Val135Met and the p.His514Arg substitutions affect conserved positions situated in functionally important domains (the MACPRF and the C2 domain respectively), and p.His514Arg is considered as non neutral by three different prediction algorithms (table 3). According to functional analysis they do not seem to affect NK activity, at least at heterozygous form. However, both the patients carrying these two variants showed a slight impairment in one perforin-related feature (proportion of $CD3^{-}CD16^{+}$ NK cells and of PRF1 positive cells respectively). Anyway the clinical relevance of this data is difficult to state. The third variation (p.Val329Ile) is a conservative substitution, which falls in the MACPRF domain but does not affect a conserved position, and it has been considered as a neutral substitution by all the programs used for the in silico prediction, therefore we consider it as a neutral polymorphism which probably does not affect protein function.

Five other non synonymous variations were found, but they have been detected also in the control sample set, three of them (p.Arg123His, p.Gly334Ser and p.Phe421Cys) in one SSc case and one control each, and two (p.Arg232His and p.Ala437Val) only in one healthy control each. Among the variants detected in the SSc population, p.Phe421Cys is located the C2 domain, responsible for perforin membrane binding. It has previously been described in heterozygosity in an Italian pediatric patient with anaplastic large cell lymphoma involving lymph nodes and spleen (Cannella et al.2007) and in compound heterozygosity with p.Ala91Val variation in a Caucasoid FHL2 patient (Clementi et al. 2005). In both patients the variation was found in the germ-line. In our sample set we found this variation twice, in a SSc patient affected by the diffuse form of the disease, and in a healthy control. Thus, considering both literature data and in silico analysis, the functional meaning of this substitution is unclear. Instead p.Arg123His does not fall in a conserved position or in a functional domain of perforin protein. It has previously been found in heterozygosity in a 14 years-old boy with anaplastic large cell lymphoma involving mediastinum (Cannella et al.2007), and in one out of 101 healthy controls, who was also heterozygous for the synonymous c.519G>A variant (Molleran Lee et al. 2004). By many authors it is considered a polymorphism of PRF1 gene. In our population it was found twice, in a SSc patient with limited disease and in a healthy control, both times together with the c.519G>A variation. It is considered as a neutral substitution both by Polyphen and by PMUT scoring matrixes. On the whole we do not think that this variant may have a pathogenic effect on protein function. The third variation, p.Gly334Ser, located in the MACPRF domain, has never been described before. It falls in an unconserved position and is considered a

neutral substitution by the four prediction algorithms (table 3), so it seems unlikely that it may affect the function of the protein.

The p.Arg232His substitution was detected in the same allele with p.Ala91Val in a healthy control. It has previously been described and registered on the Human Gene Mutation Database – HGMD (http://www.hgmd.cf.ac.uk/ac/index) by Feldmann et al. (2002) at heterozygous state in two Italian consanguineous patients of FHL with late-onset of the disease and in their mother (healthy). In these patients the authors detected about 27% of cytotoxic activity and undetectable levels of perforin expression. The mutation on the second allele was not detected in neither of the two patients. One of them was still alive at 14 without receiving bone marrow transplantation. Busiello et al. (2004) found this variation in two 13-year-old fraternal twins (one of which was affected by FHL, while the other one was healthy), who both were homozygous for p.Ala91Val and heterozygous for p.Arg232His (one allele was double-mutated). The Arg232His substitution was present also in the father (healthy). This variation was found also in another Italian FHL patient with late-onset of the disease, in compound heterozygosity with another missense mutation (p.His222Gln) (Stadt et al. 2006). In a large study Trizzino et al. (2008) found this variation in four families with FHL2 affected members, three of which from Italy, and the fourth from Germany but with probable Italian descent. This variation has been found also in an Italian pediatric patient with anaplastic large cell lymphoma (ALCL) involving lymph nodes and bone marrow, in compound heterozygosity with Ala91Val substitution (Cannella et al. 2007). Voskoboinik et al (2005) made a functional analysis of the Arg232His mutated and Arg232His/Ala91Val double-mutated perforin. The Arg232His perforin generates approximately the 30% activity of the native protein, while the double-mutated protein is completely inactive. Risma et al. (2006) functionally analyzed 21 missense mutations by flow cytometry, immunohistochemistry, and immunoblotting and demonstrated that perforin with Arg232His substitution undergoes to partial proteolytic maturation. On the whole literature data clearly show that p.Arg232His substitution partially affects cytotoxic activity. However the penetrance of this variation seems incomplete, as individuals with the same variation pattern show different disease affection status and some of them are healthy. These observations are similar to those made for p.Ala91Val variant, which is considered by many authors as a polymorphism of *PRF1* gene. The significant number of reports concerning this variant suggests that p.Arg232His is also quite common in the Italian population, so it not astonishing to find it in a healthy control, in *cis* with p.Ala91Val common variation. From Voskoboinik et al (2005) data we can speculate that this subject would probably have 50% of cytotoxic activity, sufficient to avoid development of FHL2. Unfortunately it was not possible to make a further follow up of this subject after recruitment, so we do not know if he has subsequently developed a

lymphoma. The latter variation (p.Ala 437Val), found in one healthy control, is novel. Although it codes for a conservative substitution among two non-polar amino acids, it is considered as non tolerated by two different programs, and it falls in a conserved position in the C2 domain, so it is difficult to simply consider it as a neutral variation. It is likely that, similarly to p.Ala91Val and p.Arg232His, also p.Ala437Val variation has some functional effect with incomplete penetrance, but further studies are needed to clarify this topic.

During the mutational screening of the *PRF1* gene we identified also 11 synonymous variations, comprising two common polymorphisms (rs885821 and rs885822) and the two over reported variants in LD with p.Asn252Ser. Among the remaining 7 variants, only three have previously been described in literature, while the other are new. All synonymous variants have been tested in silico to predict their influence on splicing sites, ESE sequences and codon usage. c.1620A>G and c.999C>T have been previously observed in one MS patient each (Cappellano et al, 2008), respectively in homozigosity and in heterozygosity. It has been hypothesised that c.1620A>G may affect splicing, as it seems to create a new acceptor splice site. Furthermore, according to ESEfinder scoring matrix, both c.999C>T and c.1620A>G may influence RNA processing as they remove/create an ESE sequence respectively. The c.519G>A variation was observed three times, once in all of the three sample sets. The healthy control and the SSc patient are also heterozygous for the p.Arg123His substitution. As reported above, c.519G>A has already been described in literature together with p.Arg123His in the same healthy control, and was considered by the authors as a polymorphism. Therefore we think that this is a silent variation, relatively common in the Italian population, and probably in LD with p.Arg123His. It is more difficult to speculate about the theoretical functional meaning of the newly described synonymous variants. None of them falls in a canonical splice site, anyway, according to in silico prediction, some of them (c.273G>A, c.807C>T, c.1356C>T) may influence splicing by adding or removing binding sites for various arginine/serine rich splicing factors. The other variations do not seem to influence splicing, nor they introduce seldom used codons, so they are probably neutral. It is possible, anyway, that the observed variants have not a direct functional effect, but are in LD with other unknown *PRF1* variations in the 5'UTR or in the regulative flanking sequences of the gene.

PRF1 might affect autoimmune susceptibility by two different pathways. First, a deficit in PRF1 activity may cause defective viral clearance, which in turn may favour the development of cross-reactions between viral and self antigens through molecular mimicry. Second, PRF1 might be involved in immune response modulation by perforin-mediated killing of effector lymphocytes and antigen-presenting cells. Defective immune response switching off may favour both lymphocyte accumulation and autoimmunity.

Materials and methods

Subjects. In this study we performed a mutational screening on 154 SLE cases, 131 SSc patients and 559 healthy controls, and a subsequent scanning analysis on two SNPs (enlarged population: 597 SLE, 204 SSc, 1856 controls). All subjects (patients and controls) belongs to the Italian population and are unrelated. Healthy control subjects comprise medical students, university and hospital staff and blood donors. The geographic origin of patients and controls of both panels was similar. Individuals with Sardinian ancestors were excluded. All patients (SSc and SLE) provided their informed consent.

All SLE patients (female: male ratio = 7.5:1) fulfilled at least 4 of the American College of Rheumatology 1997 revised criteria for the classification of SLE (18). Enrolment was completed when the patients provided their informed consent. The clinical and immunologic features of the SLE patients are shown in Table 1A.

All SSc patients (female: male ratio = 12:1) fulfilled the American College of Rheumatology preliminary criteria for the classification of SSc[13] and were categorised as having the limited cutaneous (lcSSc) or the diffuse cutaneous (dcSSc) subset according to LeRoy et al[14]. The clinical and immunological features of the SSc patients are shown in Table 1B.

Disease onset was determined by the patient's recall of the first non-Raynaud symptom clearly ascribable to scleroderma. Pulmonary hypertension, estimated by echocardiography, was defined as a right-ventricular systolic pressure ≥ 40 mmHg. Pulmonary fibrosis was evaluated by high resolution TC (HR- TC).

Search for new variations in PRF1 gene. Genomic DNA was isolated from peripheral blood nuclear cells with standard methods. The entire coding sequence of *PRF1* gene (exons 2 and 3) was amplified in standard PCR conditions. Exon 2 was analysed as previously described by Cappellano et al. Exon 3 was amplified in three overlapping fragments, and analysed by Denaturing High Performance Liquid Chromatography (DHPLC) or direct sequencing. DHPLC temperatures of analysis have been calculated with the algorithm "DHPLC melt program" (http://insertion.stanford.edu/melt). PCR product have been exposed to rapid thermal denaturation and slow renaturation on a thermal cycler, then analyzed on TRANSGENOMIC WAVE (Transgenomic, Omaha, NE, USA). Results were analysed with Navigator™ software (Transgenomic, Omaha, NE, USA). Primer sequences and analysis conditions are available on demand. PCR products were purified for sequencing application with a vacuum aspiration system

(PCR96 Cleanup kit-Millipore, Billerica, MA, USA). Sequencing reactions were performed using the ABI PRISM® BigDye Terminator kit v.1.1 (Applied Biosystems, Foster City, CA, USA), purified with the Montage SEQ96 Sequencing Reaction Cleanup Kit (Millipore, Billerica, MA, USA) and analysed with the 3100 Genetic Analyser automatic sequencer (Applied Biosystems, Foster City, CA, USA).

SNPs genotyping. In the second population p.Ala91Val and p.Asn252Ser polymorphisms were typed using two pre-designed TaqMan 5'-allelic discrimination assays (Applied Biosystems, Foster City, CA, USA): C__25600964_20 (p.Ala91Val) and C__27529700_10 (p.Asn252Ser) and the ABI PRISM 7000® Sequence Detection System (Applied Biosystems, Foster City, CA, USA). Genotyping of each sample was automatically attributed by the SDS 1.3 software for allelic discrimination.

Allele-specific PCR. The wild-type (91Ala) and the mutant (91Val) alleles were separately amplified using an allele-specific forward primer (91Ala-specific forward primer: 5'-CTCCAGCGCCTGCCTCTGGC-3', 91Val-specific forward primer: 5'-CTCCAGCGCCTGCCTCTGGT-3') and the proper reverse primer. PCR products were typed for the second variation (p.Asn252Ser or p.Arg232His) by sequencing with the ABI PRISM® BigDye Terminator kit v.1.1 on the 3100 Genetic Analyser.

Statistical and in silico analysis. The statistical significance of the differences in phenotype, allele and genotype frequencies between patients and controls was evaluated using the chi-square test with Yates' correction or the Fisher's two-tailed test, as appropriate. The association of each polymorphism with the disease was measured by the odds ratio (OR) and its 95% confidence interval (95% CI). The significance cut-off was $P<0.05$.

Putative effect of the variations on splicing sites was evaluated using the SpliceView program on the Web-Gene website (http://www.itb.cnr.it/sun/webgene/) and the ESEfinder scoring matrix (http://www.rulai.cshl.edu/tools/ESE). Putative functional meaning of the missense variations was evaluated with the programs PolyPhen (http://genetics.bwh.harvard.edu/pph), SIFT (http://blocks.fhcrc.org/sift/SIFT), SNAP (http://cubic.bioc.columbia.edu/services/SNAP/submit) and pMUT (http://mmb2.pcb.ub.es:8080/pMut/). The effect of the variant in signal peptide on peptide cleavage was determined with the server SignalP 3.0 (www.cbs.dtu.dk/services/SignalP). Alignment analysis was performed with the program ClustalW (http://saier-144-37.ucsd.edu/clustalw).

References

Arnett FC, Cho M, Chatterjee S, Aguilar MB, Reveille JD, Mayes MD. Familial occurrence frequencies and relative risk for Systemic Sclerosis (Scleroderma) in three United States Cohorts. *Arthritis Rheum*. 2001; **44(6)**: 1359-1362.

Cappellano G, Orilieri E, Comi C, Chiocchetti A, Bocca S, Boggio E, Bernardone IS, Cometa A, Clementi R, Barizzone N, D'Alfonso S, Corrado L, Galimberti D, Scarpini E, Guerini FR, Caputo D, Paolicelli D, Trojano M, Figà-Talamanca L, Salvetti M, Perla F, Leone M, Monaco F and Dianzani U. Variations of the perforin gene in patients with multiple sclerosis. *Genes and Immunity* 2008; 9: 438-444.

Clementi R, Chiocchetti A, Cappellano G, Cerutti E, Ferretti M, Orilieri E, Dianzani I, Ferraini M, Bregni M, Danesino C, Bozzi V, Putti MC, Cerutti F, Cometa A, Locatelli F, Maccario R, Ramenghi U, Dianzani U. Variations of the perforin gene in patients with autoimmunity/lymphoproliferation and defective Fas function. *Blood* 2006; 108(9):3079-3084.

Jarjanazi H, Savas S, Pabalan N, Tennis JW, Ozcelik H. Biological implications of SNPs in signal peptide domains of human proteins. *Proteins* 2008; 70:394-403.

Komar AA. Genetics. SNPs, silent but not invisibile. *Science* 2007; 315(5811):466-7.

Molleran Lee S, Villanueva J, Sumegi J, Zhang K, Kogawa K, Davis J, Filipovich AH. Characterization of diverse PRF1 mutations leading to decreased natural killer cell activity in North American families with haemophagocytic lymphohistiocytosis. *J Med Genet* 2004;41:137-144.

Orilieri E, Cappellano G, Clementi R, Cometa A, Ferretti M, Cerutti E, Cadario F, Martinetti M, Larizza D, Calcaterra V, D'Annunzio G, Lorini R, Cerutti F, Bruno G, Chiocchetti A, Dianzani U. Variations of the perforin gene in patients with type 1 diabetes. *Diabetes* 2008; 57(4):1078-83.

Risma KA, Frayer RW, Filipovich AH, Sumegi J. Aberrant maturation of mutant perforin underlies the clinical diversity of hemophagocytic lymphohistiocytosis. *The Journal of Clinical Investigation* 2006; 116(1):182-192

Stepp SE, Dufourcq-Lagelouse R, Le Deist F, Bhawan S, Certain S, Mathew PA, Henter JI, Bennet M, Fischer A, de Saint Basile G, Kumar V. Perforin gene defects in familial hemophagocytic lymphohistiocytosis. *Science* 1999;286:1957-1959.

Trambas C, Gallo F, Pende D, Marcenaro S, Moretta L, De Fusco C, Santoro A, Notarangelo L, Arico M, Griffiths GM. A single amino acid ch'ange, A91V, leads to conformational changes that can impari processing to the active form of perforin. *Blood* 2005;106(3):932-937.

Vastert B, Van Wijk R, D'Urbano L, De Vooght K, Ravelli A, Cortis E, Hoppenreijs E, Van Solinge W, Prakken B, Wulffraat N, De Benedetti F, Kuis W. Mutations in the perforin gene may be linked to macrophage activation syndrome in patients with systemic onset Juvenile Idiopathic Arthritis. *The Turkish Journal of Pediatrics* 2008; 50 (Supplement):2.

Voskoboinik I, Smyth MJ, Trapani JA. Perforin-mediated target-cell death and immune homeostasis. *Nature Reviews Immunology* 2006;6:940-952

Voskoboinik I, Thia MC, Trapani JA. A functional analysis of the putative polymorphisms A91V and N252S and 22 missense perforin mutations associated with familial hemophagocytic lymphohistiocytosis. *Blood* 2005;105(12):4700-4706.

Tab.1A Clinical and immunological features of SLE patients.

Feature		**% of patients**
Clinical manifestations:	Malar rash	43
	Photosensitivity	43
	Discoid lesions	11
	Serositis*	31
	Oral ulcers	10
	Arthritis	71
	Glomerulonephritis	37
	Autoimmune haemolytic anemia	9
	Raynaud's phenomenon	33
	Lymphadenopathy	13
	Leukopenia and/or lymphopenia[§]	53
	Thrombocytopenia[#]	23
	Neurologic involvement[@]	17
	Antiphospholipid syndrome	26
Autoantibodies:	ANA	94
	anti-dsDNA	78
	anti-La (SSB)	6
	anti-Ro (SSA)	38
	anti-RNP and/or Sm	28
	Antiphospholipids[┼]	46
	Rheumatoid factor	20

ANA = antinuclear antibodies; anti-dsDNA = anti-double-stranded DNA.
*Includes pericarditis and/or pleuritis.
[§]Leukocyte count <4,000/mm^3; lymphocyte count <1,500/mm^3.
[#]Platelet count <100,000/mm^3.
[@]Includes epilepsy and/or psycosis and/or transient ischemic attack.
[┼]Presence of anticardiolipidin antibodies and/or lupus anticoagulant activity.

Tab.1B Clinical and immunologic features of SSc patients.

Feature		**% of patients**
SSc subset:	Diffuse cutaneous	19
	Limited cutaneous	81
Clinical manifestations:	Fibrosis	42
	Pulmonary artery hypertension (PAH)*	13
Autoantibodies:	ACA	41
	anti-topo I	39
	anti-RNApIII	5

*defined as a right-ventricular systolic pressure ≥ 40 mmHg;
ACA=anti-centromere antibodies; anti-topo I = anti-topoisomerase I antibodies; anti-RNApIII = anti-RNA polymerase III antibodies

Tab.2 *PRF1* variations detected in a panel of SLE and SSc patients and controls

nucleotide variation	predicted amino acid change	**exon**	**ID***	**previous reports**[§]	SLE N = 154	SSc N = 131	controls N = 559
c.11G>A	Arg4His	2	rs35418374	AA, SoJIA, controls	0	1	0
c.189A>C	Thr63Thr	2		novel	1	0	0
c.272C>T	Ala91Val	2	rs35947132	FHL2, controls	21	21	57
c.273G>A	Ala91Ala[1]	2		novel	1	1	1
c.368G>A	Arg123His[2]	2		ALCL	0	1	1
c.403G>A	Val135Met	2	rs12263572	NCBI SNPs database	1	0	0
c.435G>A	Val145Val	2		NCBI SNPs database	6	4	3
c.462A>G	Ala154Ala	2		NCBI SNPs database	6	4	3
c.519G>A	Thr173Thr[3]	2		controls	1	1	1
c.695G>A	Arg232His[4]	3		FHL2, ALCL, healthy relatives	0	0	1
c.755A>G	Asn252Ser	3	rs28933375	FHL2, controls	6	4	3
c.807C>T	His269His	3		novel	1	0	0
c.985G>A	Val329Ile	3		novel	1	0	0
c.999C>T	Pro333Pro	3		MS	0	1	0
c.1000G>A	Gly334Ser	3		novel	0	1	1
c.1262T>C	Phe421Cys	3		FHL2, ALCL	0	1	1
c.1310C>T	Ala437Val	3		novel	0	0	1
c.1356C>T	Thr452Thr	3		novel	1	0	1
c.1541A>G	His514Arg	3		novel	1	0	0
c.1620A>G	Gln540Gln	3		MS	1	0	1

Numerations are referred to the GeneBank cDNA clone M28393, ATG = +1.
Besides p.Ala91Val all the other variations have been found in heterozigosity.
In addition two common synonymous polymorphisms have been observed: c.822C>T Ala274Ala – rs885821 (reported in 54/163 SLE, 48/136 SSc, 258/772 controls) and c.900T>C His300His – rs885822 (reported in 110/162 SLE, 92/136 SSc, 562/771 controls)

[*]Identification number according to http://www.ncbi.nlm.nih.gov database.
[§]FHL2 = Familial Haemophagocytic Lymphohistiocytosis; ALCL = Anaplastic Large Cell Lymphoma; SoJIA = Systemic onset Juvenile Idiopathic Arthritis; MS = Multiple Sclerosis; AA = Aplastic Anemia.
[1]The healthy control is heterozygous also for p.Ala91Val.
[2]Both SSc patient and the healthy control are heterozygous for the c.519G>A synonymous variant.
[3]Both SSc patient and the healthy control are heterozygous for the non synonymous substitution p.Arg123His.
[4]The healthy control carries also 91Val on the same allele with 232His, as assessed by allele-specific PCR.

Tab.3 In silico prediction of the effect of the newly identified *PRF1* missense variations on protein structure and function using the Polyphen, SIFT, SNAP, and pMUT programs

Aminoacid change	Polyphen		SIFT		SNAP		pMUT	
	prediction	score[a]	prediction	score[b]	prediction	Expected accuracy[c]	prediction	score[d]
Arg4His	benign	none	-	-	neutral	60%	neutral	0.2352
Arg123His	benign	0.103	**non tolerated**	0.03	**non neutral**	78%	neutral	0.3133
Val135Met	benign	1.380	**non tolerated**	0.02	**non neutral**	82%	neutral	0.2700
Arg232His	benign	0.012	**non tolerated**	0.02	neutral	69%	**pathological**	0.5239
Val329Ile	benign	0.148	tolerated	0.44	neutral	78%	neutral	0.0468
Gly334Ser	benign	1.378	tolerated	0.74	neutral	92%	neutral	0.3109
Phe421Cys	benign	0.078	tolerated	0.07	**non neutral**	63%	**pathological**	0.8781
Ala437Val	benign	1.487	**non tolerated**	0.02	**non neutral**	78%	neutral	0.4791
His514Arg	**possibly damaging**	1.518	**non tolerated**	0.00	neutral	53%	**pathological**	0.7938

[a] The lower the score, the more benign the substitution. Scores above 1.5 are considered non-benign (PolyPhen)

[b] The higher the score, the more the substitution will be tolerated. Amino acids with score <0.05 are considered deleterious (SIFT).

[c] The higher the percentage, the greater the confidence of the prediction (SNAP).

[d] A pathogenicity index ranging from 0 to 1. The lower the score, the more the substitution will be considered as neutral. Mutations with an index above 0.5 are taken as pathological (pMUT).

Table 4: NK activity, perforin expression, and proportion of NK cells in PBMC of SLE patients carrying the p.His514Arg and p.Val135Met perforin variations.

PRF1 Variation	SUBJECT	NK activity* (effector/target ratio)			Perforin expression†		peripheral blood NK cells %	
		100:1	30:1	10:1	%	MFI-R	$CD3^-$ $CD16^+$	$CD3^-$ $CD56^+$
H514R	**Pt.1**	54	34	19	16§	14	8	9
V135M	**Pt.2**	34	20	11	21	19	4§	12
-	**Ctr.1**	61	41	27	23	12	26	10
-	**Ctr.2**	52	39	24	32	16	31	11
	CONTROLS¶	43 (15-62)	27 (9-43)	15 (4-28)	23 (17-29)	9 (5-20)	11 (6-31)	16 (7-27)

*: NK activity is expressed as specific cell lysis % and it is the mean of triplicate assays, whose standard deviation was always <10% of the mean. Spontaneous cell lysis was always <10% of maximal cell lysis.

†: perforin expression is shown as proportion of positive cells (%) and mean fluorescence intensity ratio (MFI-R).

§: $<5^{th}$ percentile of controls. .

¶: median ($5\text{-}95^{th}$ percentile range) from 15 controls.

Tab. 5 Clinical and immunologic features of SLE (A) and SSc (B) patients carrying PRF1 variations.

A

Genotype	sex	MR	PS	S	Ar	GNF	AHA	LA	Lp/Lp	TP	NI	aPL	ANA	aDNA	SSA	LAC	RF
V135M/wt	F	-	-	-	-	+	+	-	-	+	-	-	+	+	-	-	-
H514R/wt	F	-	-	-	+	+	-	-	+	-	+	-	+	+	-	-	-

MR = malar rash; PS = photosensitivity; LED = discoid lesions; S = serositis; A = arthritis; GNF = glomerulonephritis; LA= limphadenopathy; AHA = autoimmune haemolytic anemia; Lp/Lp = leukopenia and/or lymphopenia; TP = thrombocytopenia; NI = neurologic involvement; aPL = antiphospholipid syndrome; ANA= antinuclear antibodies; aDNA= anti-dsDNA autoantibodies; SSA = antiRo (SSA) antibodies; LAC = presence of lupus anticoagulant activity; RF = rheumatoid factor; n.a. = clinical information is not available.

Both patients were negative for discoid lesions and anti-RNP and/or Sm antibodies. Clinical informations about presence of oral ulcers. Raynaud's phenomenon and anti-La (SSB) antibodies are not available for these patients. Clinical informations for the patient carrying the c.985G>A (p.Val329Ile) variation were not available.

B

Genotype	sex	age at diagnosis	diagnosis	lung fibrosis	PAH	ACA	anti-topo I	anti-RNApIII
R123H/T173T	F	50	lSSc	n.a	-	-	-	+
F421C/wt	M	39	dSSc	-	n.a	-	+	-
G334S/wt	F	12	SSc early	-	-	-	+	-
R4H/wt	F	56	lSSc	-	+	-	+	-

lSSc = limited form; dSSc = diffuse form; SSc early = early form of disease. without satisfaction of all ACR criteria; PAH = pulmonary artery hypertension; ACA=anti-centromere antibodies; anti-topo I = anti-topoisomerase I antibodies; anti-RNApIII = anti-RNA polymerase III antibodies; n.a. = clinical information is not available.

Tab.6 Allelic and genotype frequencies of p.Ala91Val (A) and p.Asn252Ser (B) PRF1 variations in patients and controls (overall population).

A	**CONTROLS** n=1856	**SLE** n=597			**SSc** n=204		
genotypes	**n(frequency)**	**n(frequency)**	**p**	**OR (95%CI)**	**n(frequency)**	**p**	**OR (95%CI)**
CC	1652 (0.890)	524 (0.877)	0.45	0.89 (0.66-1.20)	168 (0.823)	**0.0070**	**0.58 (0.39-0.88)**
CT	196 (0.106)	69 (0.116)	0.54	1.11 (0.81-1.49)	34 (0.167)	**0.012**	**1.69 (1.10-2.54)**
TT	8 (0.004)	4 (0.007)	0.50	1.56 (0.34-5.84)	2 (0.010)	0.26	2.29 (0.23-11.56)
alleles	**frequency**	**frequency**	**p**	**OR (95%CI)**	**frequency**	**p**	**OR (95%CI)**
C	0.943	0.936			0.907		
T	0.057	0.064	0.38	1.14 (0.86-1.50)	0.093	**0.0054**	**1.70 (1.16-2.47)**

B	**CONTROLS** n=1856	**SLE** n=597			**SSc** n=204		
genotypes	**n(frequency)**	**n(frequency)**	**p**	**OR (95%CI)**	**n(frequency)**	**p**	**OR (95%CI)**
AA	1840 (0.991)	585 (0.980)	**0.038**	**0.42 (0.19-0.99)**	200 (0.980)	0.13	0.43 (0.14-1.81)
AG	16 (0.009)	12 (0.020)	**0.038**	**2.36 (1.01-5.35)**	4 (0.020)	0.13	2.30 (0.55-7.22)
GG	0	0			0		
alleles	**frequency**	**frequency**	**p**	**OR (95%CI)**	**frequency**	**p**	**OR (95%CI)**
A	0.996	0.990			0.990		
G	0.004	0.010	**0.039**	**2.35 (1.04-5.24)**	0.010	0.13	2.29 (0.64-7.32)

Overall p values for genotype distributions relative to p.Ala91Val polymorphism in patients vs. controls were 0.60 for SLE and **0.017** for SSc.
Significant OR and p values are in boldface.