

# Università degli Studi del Piemonte Orientale “Amedeo Avogadro”



**Dottorato di Ricerca**

**in**

**Medicina Molecolare**

*Ciclo XXIII*

Relazione 3° anno

TITOLO: RICERCA DI FATTORI GENETICI DI SUSCETTIBILITA' AL LUPUS  
ERITEMATOSO SISTEMICO

**Candidato: Barizzone Nadia**

***Tutor:* Prof.ssa Sandra D'Alfonso**

**SEZIONE 1**  
**RISULTATI SCIENTIFICI**

<b>SCOPO DEL LAVORO.....</b>	<b>pag. 3</b>
<b>INTRODUZIONE.....</b>	<b>pag. 4</b>
<b>MATERIALI E METODI.....</b>	<b>pag. 8</b>
<b>RISULTATI.....</b>	<b>pag. 11</b>
<b>DISCUSSIONE.....</b>	<b>pag. 16</b>
<b>BIBLIOGRAFIA.....</b>	<b>pag. 19</b>
<b>TABELLE.....</b>	<b>pag. 21</b>
<b>APPENDICE: CENNI SUI GENI COINVOLTI IN QUESTO STUDIO .....</b>	<b>pag. 30</b>
<b>ATTIVITA'FORMATIVA.....</b>	<b>pag. 37</b>

## SCOPO DEL LAVORO

Durante il III anno di dottorato ho disegnato e iniziato un nuovo progetto, nell'ambito della ricerca di fattori genetici di suscettibilità al Lupus Eritematoso Sistemico (LES). 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 prefiggiamo 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.

A tal scopo intendiamo:

- 1) Tipizzare su una casistica di almeno 300 casi e 300 controlli 30 polimorfismi in altrettanti geni, emersi come marcatori maggiormente associati da precedenti analisi di tipo GWAS (tabella 1). Per alcuni di questi sono già disponibili dati ottenuti tramite collaborazioni in studi multi centro internazionali che hanno incluso anche la nostra casistica (dati già pubblicati).
- 2) Unire a questi dati quelli ricavati da una collaborazione con un gruppo internazionale (ancora in corso d'opera) che ha analizzato 384 SNPs in 19 geni (tabella 2). Questi geni sono stati selezionati per la replicazione sulla base di un GWAS non ancora pubblicato.
- 3) Integrare le informazioni genetiche con informazioni di carattere clinico ed immunologico.

Per tre dei geni inclusi nella tabella 2 (CBLB, TEC e BMPR1B) intendiamo eseguire l'analisi di associazione sull'intera casistica (1268 casi e 1597 controlli), non solo sui campioni italiani. Inoltre i geni TEC e BMPR1B verranno analizzati tramite elaborazioni più approfondite, i cui obiettivi consistono in:

- 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.

# INTRODUZIONE

## LUPUS ERITEMATOSO SISTEMICO

Il *Lupus Eritematoso Sistemico* (LES) è una malattia autoimmune cronica a eziologia multifattoriale, con una prevalenza media nelle popolazioni Caucasoidi Europee e Nord-Americane di circa 15-50 casi ogni 100.000 abitanti. E' più diffuso tra le donne in età fertile, con un rapporto di incidenza donna/uomo di 9:1, e nella popolazione Afro-Americana (Simard et al. 2007). La sintomatologia del LES è molto complessa, e può variare notevolmente da un paziente ad un altro. I sintomi più comuni e caratteristici includono l'eritema malare "a farfalla" che insorge sul volto del paziente nelle pieghe naso-labiali, artrite, glomerulonefrite, pleuriti o pericarditi, altri disturbi cutanei come le lesioni discoidi, l'alopecia e il fenomeno di Raynaud (sbiancamento alle estremità indotto dall'esposizione a basse temperature) (Kotzin, 1996; Cervera et al. 1993). Il quadro immunologico è altrettanto complesso, ma è caratterizzato principalmente dalla presenza di autoanticorpi diretti contro antigeni nucleari (Cervera et al. 1993). La diagnosi di LES viene accertata qualora sussista la positività per almeno 4 dei criteri clinici ed immunologici definiti dall'*American Rheumatism Association* (ARA) (Wallace et al. 1993).

La patogenesi del LES è complessa, basata sull'interazione di fattori ambientali, come l'esposizione ai raggi U.V. e ad agenti infettivi (Kotzin, 1996; Hrycek et al. 2005; Simard et al. 2007), con altri fattori genetici ed immunologici. Nella maggior parte dei casi la causa principale del danno è dovuta alla formazione di immunocomplessi (IC) tra anticorpi IgG ed antigeni solubili ed alla precipitazione di IC e depositi di componenti del complemento a livello tissutale, soprattutto a livello renale (Brown et al. 2007; Sturfelt e Truedsson 2005).

Si hanno da tempo considerevoli evidenze di una base genetica nella suscettibilità al LES, derivanti da studi di aggregazione familiare e su gemelli. Il rischio per un parente di un malato di contrarre la malattia ( $\lambda$ ) è stimato tra le 20 e le 80 volte superiore rispetto a quello per la popolazione generale (Lindqvist et al. 1999). Il tasso di concordanza per la presenza della malattia nei gemelli monozigotici è del 24%, contro un 2% di concordanza nei gemelli dizigotici. Il tasso di concordanza per la presenza di anticorpi sierici può invece raggiungere il 92% nei gemelli monozigotici (Lindqvist et al. 1999; Deapen et al. 1992; Grennan et al. 1997). I fattori genetici coinvolti nella suscettibilità al LES sono moltissimi, ed includono:

- geni coinvolti nella presentazione dell'antigene (HLA-DR)

- geni coinvolti nella regolazione dell'attivazione dei linfociti T e B:

BANK1, PTPN22, PDCD1, BLK

- geni coinvolti nella via di segnalazione dell'IFN o stimolatori della produzione dell'IFN:

IL18, IRF5, IRAK1, TNFAIP3, STAT4, TLR5

- geni del complemento o che interagiscono con la cascata del complemento:

MBL, CR2 e altri geni del cluster, ITGAM (coinvolto nella fagocitosi di corpi opsonizzati dal complemento).

A oggi i diversi studi di associazione genetica, basati sull'approccio del gene candidato o di tipo GWA, hanno identificato almeno 30 geni la cui associazione è stata confermata in due o più studi indipendenti ed è ritenuta attendibile (tabella 1). Il nostro gruppo di ricerca ha contribuito all'identificazione di molti di questi geni sia nell'ambito di progetti italiani che grazie alla collaborazione con i principali gruppi di ricerca internazionali che si occupano della genetica del LES.

L'identificazione di questi geni è sicuramente assai importante per comprendere i meccanismi patogenetici alla base dello sviluppo del LES. Tuttavia queste osservazioni fino a questo momento hanno avuto uno scarso impatto nella pratica clinica. La presenza di ciascuno dei singoli alleli di suscettibilità identificati conferisce solamente un OR di 1.2-2.5 (tabella 1) rispetto alla sua assenza

per cui preso singolarmente nessuno dei loci di suscettibilità al LES ad oggi identificato può avere un valore predittivo-diagnostico.

Tuttavia è possibile che la valutazione dell'effetto combinato dei diversi geni, mediante l'analisi simultanea delle diverse varianti associate, possa aumentarne notevolmente il potere predittivo. Questo approccio è stato applicato con successo in altre patologie complesse quali malattie metaboliche cardiovascolari e sclerosi multipla. I risultati finora sono promettenti: considerare simultaneamente i diversi fattori genetici aumenta significativamente il potere predittivo della malattia sia rispetto all'utilizzo dei singoli geni sia rispetto all'analisi dei soli fattori di rischio non genetici (Yang 2003, Yang 2005, Kathiresan 2008, Lyssenko 2008, De Jager 2009). È stato dimostrato che, se molti fattori genetici distinti sono rilevanti nell'eziologia della malattia, allora, l'analisi simultanea e combinata di tutti questi fattori attraverso l'uso di un pannello di test genetici aumenta sostanzialmente il valore predittivo per la malattia. In particolare è stato stimato che 20-25 varianti genetiche di rischio con una frequenza maggiore del 10% e un OR di almeno 1.5 sono richieste per spiegare circa il 50% di tutti i casi affetti da una determinata patologia complessa (Yang 2005).

## **IMPUTAZIONE**

Per trovare varianti genetiche associate ad una patologia utilizzando un GWAS è tipicamente necessario tipizzare alcune migliaia di casi e di controlli per ottenere un potere statistico adeguato. Inoltre è necessario analizzare alcune centinaia di migliaia di SNPs per ottenere una copertura adeguata dell'intero genoma. Di conseguenza i GWAS necessitano di nuovi metodi efficienti di analisi computazionale.

Uno dei metodi per migliorare il potere statistico di un GWAS è di stimare gli aplotipi sulla base del Linkage Disequilibrium (LD) e di utilizzare questi per le analisi caso-controllo. Per via del LD gli aplotipi possono essere correlati con altre varianti nella stessa regione genomica, per cui l'analisi degli aplotipi può permettere di testare marcatori non genotipizzati.

Un secondo metodo per migliorare il potere dei GWAS è quello di utilizzare l'imputazione dei dati mancanti per inferire i genotipi di varianti note ma non genotipizzate, che possono così essere testate per l'associazione con la malattia. Attualmente i marcatori imputati sono di solito SNP genotipizzati dal progetto HapMap.

Un terzo approccio consiste nel combinare insieme risultati provenienti da studi diversi, imputando i genotipi dei marcatori che sono stati analizzati solo in alcuni studi, ma non in altri.

L'approccio seguito in questo studio è una combinazione dei metodi due e tre.

L'imputazione dei dati mancanti nell'elaborazione dei dati di un GWAS può essere usata per diversi scopi.

1) Per ottenere un dato per i genotipi che non hanno passato il controllo di qualità (imputazione dei genotipi mancanti). Le procedure di controllo di qualità che precedono l'analisi eliminano normalmente gli individui ed i polimorfismi con un alta percentuale di dati mancanti. Tuttavia ogni individuo ed ogni marcatore avrà una bassa percentuale di dati mancanti, che possono essere "riempiti" tramite le tecniche di imputazione. In questo caso non è necessario utilizzare un pannello di genotipi di riferimento.

2) Per imputare i genotipi dei marcatori che non sono stati imputati nello studio, ma che sono stati genotipizzati in un pannello di riferimento (di solito i dati del progetto Hap Map o del progetto

1000 genomi). Questo tipo di imputazione è strettamente connesso con il concetto di tagSNP. I tagSNP vengono selezionati da un pannello più ampio di marcatori (per esempio, tutti i polimorfismi di un gene) in modo che ogni polimorfismo del pannello più ampio sia in elevato LD con un tagSNP o con un aptotipo determinato dai tagSNP selezionati. In tal modo l'analisi di un numero ridotto di marcatori dovrebbe fornire, sia pure con una ridotta perdita di potere statistico, le stesse informazioni derivanti dall'analisi dell'intero pannello. I tagSNP possono essere usati anche per imputare marcatori in elevato LD che non sono stati genotipizzati. Prima dello sviluppo del progetto HapMap, per la selezione dei tag era necessario condurre uno studio pilota su tutti i polimorfismi noti in una regione. Ora è possibile selezionare i tagSNP dai dati del progetto HapMap, e combinare i propri dati con quelli di HapMap per imputare tutti gli altri SNP nella regione. Questa strategia permette di ridurre i costi, utilizzando arrays con un più basso numero di marcatori (quindi meno costosi).

3) Per combinare insieme dati di due o più studi. Sta diventando evidente che nemmeno i GWAS effettuati su diverse migliaia di casi e di controlli hanno un potere statistico sufficientemente alto per la ricerca di loci di suscettibilità in molte malattie comuni. L'approccio che viene utilizzato sempre più comunemente è quello di unire i dati provenienti da più studi per aumentare le dimensioni della casistica. Le tecniche di imputazione permettono di effettuare questa operazione con studi effettuati su pannelli diversi di marcatori. E' sempre indispensabile utilizzare un pannello di riferimento di individui che siano stati genotipizzati per la maggioranza dei marcatori di tutte le piattaforme che si vogliono combinare.

4) Negli studi basati su famiglie, quando non è disponibile il materiale biologico di uno dei membri della famiglia.

(Browning et al. 2008 )

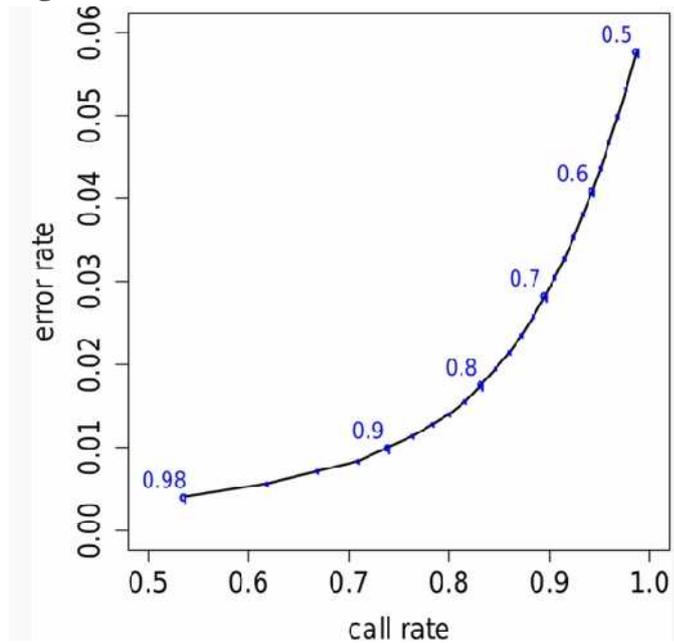
I programmi di imputazione utilizzano un algoritmo basato sulle catene di Markov. Teoricamente l'imputazione per un particolare SNP combina insieme le informazioni di tutti gli SNP del pannello di riferimento che si trovano sullo stesso cromosoma, ma l'influenza di ogni SNP diminuisce con l'aumentare della distanza genetica dal locus di interesse.

L'accuratezza (percentuale di genotipi assegnati correttamente) e l'efficacia (percentuale di genotipi assegnati) dell'imputazione dipendono dalla densità di polimorfismi tipizzati in una regione, dalla densità di polimorfismi nel pannello di riferimento, dalla somiglianza nel pattern di LD (struttura della popolazione) tra la popolazione genotipizzata e quella utilizzata come pannello di riferimento. Differenze nella struttura della popolazione potrebbero anche causare l'osservazione di false associazioni (Marchini et al. 2007). Inoltre l'accuratezza diminuisce con l'aumentare dell'eterozigosità del singolo marcatore e in regioni genomiche caratterizzate da un'alta variabilità tra gli individui (come per esempio il complesso HLA) (Nothnagel et al. 2009).

Al momento sono disponibili quattro diversi algoritmi per l'imputazione dei genotipi: BEAGLE (Browning and Browning 2007), IMPUTE (Marchini et al. 2007), MACH (Li and Abecasis 2006) e PLINK (Purcell et al. 2007). Nothnagel et al. 2009 hanno effettuato una validazione del metodo dell'imputazione, ed un confronto tra i 4 diversi algoritmi, imputando un gruppo di 449 individui genotipizzati da loro per i marcatori analizzati in tre diversi GWAS. Il software PLINK (utilizzato anche in questo studio) si è rivelato il migliore sia in termini di accuratezza dell'imputazione che in termini di efficacia.

I programmi di imputazione forniscono, per ciascuno dei tre possibili genotipi di ogni SNP di ogni individuo, un valore probabilistico. L'attribuzione dei genotipi viene effettuata fissando una soglia di probabilità, ed il dato viene assegnato solo ai genotipi che superano una data soglia (in gergo si dice che il genotipo viene "chiamato"). Valori di soglia più elevati generano un tasso di errore minore, ma anche una percentuale di genotipi imputati meno elevata. Per esempio nella figura 1, se si fissa la soglia di probabilità al 0,9 (valore utilizzato più comunemente) il tasso di errore (cioè di genotipi assegnati erroneamente) è dell'1%, e la percentuale di genotipi imputati che viene "chiamata" è approssimativamente del 74% (dati Guan et al. 2008, analizzati con IMPUTE).

**Fig. 1 Relazione tra tasso di errore (error rate) e percentuale di SNP imputati (call rate).**



Il grafico illustra la relazione tra tasso di errore e call rate al variare della soglia di probabilità per l'assegnazione di un genotipo. I numeri in blu indicano i valori di soglia che producono i tassi di errore e di call corrispondenti.

## 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 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.

Analisi di associazione. Tutte le analisi di associazione sono state condotte utilizzando il software PLINK ([pnu.mgh.harvard.edu/~purcell/plink](http://pnu.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 sì 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 Software), reperibili al sito web [www.stats.ox.ac.uk/~marchini/software/gwas/gwas.html](http://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 Software”, utilizzando la modalità “PED to GEN Conversion 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 Software". 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). È 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 Software", utilizzando la modalità "Merge Mode", che 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 SNPTTEST del pacchetto "Genome-wide Association Study Software", 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 <-hwe> richiede di calcolare la probabilità relativa all'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).

## RISULTATI

### **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 1). 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 caucasica, 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 state 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 1 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.

### **Progetto 384-plex**

In aggiunta ai marcatori di suscettibilità già noti, selezionati nella tabella 1, intendiamo inserire nell'algoritmo per il calcolo del wGRS anche nuovi SNPs, la cui associazione con il LES non sia ancora stata descritta. A tal scopo il nostro laboratorio sta contribuendo ad un progetto internazionale (384-plex project) di replicazione di un GWAS non ancora pubblicato. Sulla base dei dati del GWAS (ancora confidenziali) sono stati selezionati 18 geni (tabella 2), e per ciascuno di questi è stato eseguito un "fine mapping", ossia i marcatori da analizzare (tagSNP) sono stati selezionati (in base al LD) in modo da ottenere informazioni anche sugli altri polimorfismi noti della regione. Sono inoltre stati analizzati 4 marcatori situati nel gene HLA-DRB1.

Complessivamente il progetto ha comportato l'analisi di 384 SNP. Tra questi, 370 hanno passato i controlli di qualità, e sono stati inclusi nelle analisi successive. Oltre ad HLA-DRB1, altri 5 geni (PXK, ATG5, LYN BLK e SCUBE1) sono loci già noti di suscettibilità al LES, e sono già stati inclusi nella tabella 1. Per gli altri geni, invece, al momento non è ancora stata descritta un'associazione con il LES.

Lo studio di replicazione è stato condotto su 1268 casi e 1597 controlli di varia origine geografica. Il nostro laboratorio ha contribuito con 278 casi e 333 controlli italiani. Le associazioni più significative osservate nella popolazione italiana sono riportate nella tabella 3. Poiché i dati relativi a questo progetto sono ancora molto confidenziali, ad ogni SNP è stato attribuito un codice identificativo arbitrario, e non vengono riportati il numero ufficiale di identificazione del polimorfismo o la sua posizione genomica. Complessivamente, 15 marcatori presentano nella popolazione italiana, un  $p < 0,01$ , tutti localizzati in loci già noti di suscettibilità (vedi tabella 3). Solo 5 SNP (due situati nel gene HLA-DRB1 e tre situati nel gene SCUBE1) risultano associati con un  $p < 10^{-4}$  che rimane significativo dopo aver applicato la correzione di Bonferroni ( $p \times n$ . marcatori analizzati, cioè 384).

### **Studio di associazione su marcatori noti di suscettibilità al LES**

Al momento, oltre ai dati del progetto 384-plex, disponiamo dei dati di genotipizzazione sulla popolazione italiana di altri 3 geni inclusi nella tabella 1: TNFSF4 (per il quale sono stati tipizzati 223 controlli e 259 pazienti affetti da LES), IRF5 e STAT 4 (tipizzati per 4 marcatori ciascuno sulla stessa casistica formata da 264 controlli e 264 LES). I dati di associazione osservati nella nostra popolazione sono riportati nelle tabelle 4, 5 e 6.

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 7). 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 1). Questi marcatori (indicati con L in tabella 7) 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 7) 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 GRS è stato calcolato come sommatoria dei valori pesati di tutti e 15 gli SNP, secondo la seguente formula:

$$GRS = \sum w_i X_i$$

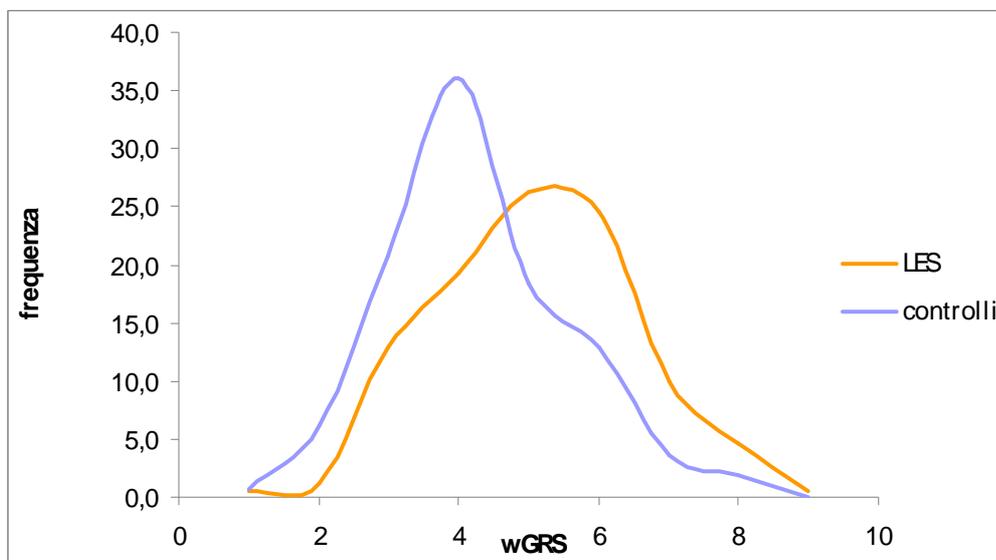
dove  $i$  è il polimorfismo,  $w$  è il “peso” del polimorfismo ( $\ln OR$ ) e  $X_i$  è il numero di alleli di rischio.

### Risk Score: risultati preliminari

Nella figura 2 i valori di wGRS così calcolati (in ascissa) sono plottati in rapporto alla percentuale di individui (in ordinata).

Nei LES il wGRS medio è  $4,62 \pm 1,40$ , nei controlli wGRS =  $3,77 \pm 1,30$ .

**Fig. 2 Grafico di distribuzione del wGRS nei pazienti e nei controlli**



### Associazione di CBLB, TEC e BMPR1B sulla casistica complessiva

Per tre dei geni inclusi nel progetto 384-plex (CBLB, TEC e BMPR1B) abbiamo avuto a disposizione i dati di genotipizzazione ottenuti sull’intera casistica (1268 casi e 1597 controlli), non solo sui campioni di origine italiana. I risultati dell’analisi di associazione su tutti i marcatori (8 in CBLB, 15 in TEC e 28 in BMPR1B) sono riportati rispettivamente nelle tabelle 8,9 e 10.

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 i geni TEC e BMPR1B abbiamo eseguito un’analisi più complessa, che consiste nell’imputazione dei polimorfismi non analizzati (vedi introduzione e materiali e metodi) e in una meta-analisi dei dati del progetto 384-plex (1268 casi e 1597 controlli) con i dati derivanti da due diversi studi GWAS. Il primo (da qui in poi denominato 300K GWAS) 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. Il secondo (100K GWAS), di proporzioni più

modeste, è stato condotto su 534 casi e 771 controlli, e i suoi dati non sono ancora stati pubblicati. Per questioni di riservatezza, anche 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 3).

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). È stato utilizzato l'algoritmo IMPUTE (vedi materiali e metodi). Le imputazioni sono state condotte su una regione di 3Mb, compresa tra la base 46 000 000 e la base 49 000 000 nel caso di TEC e su una regione di 2Mb, tra la base 95 000 000 e la base 97 000 000 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.

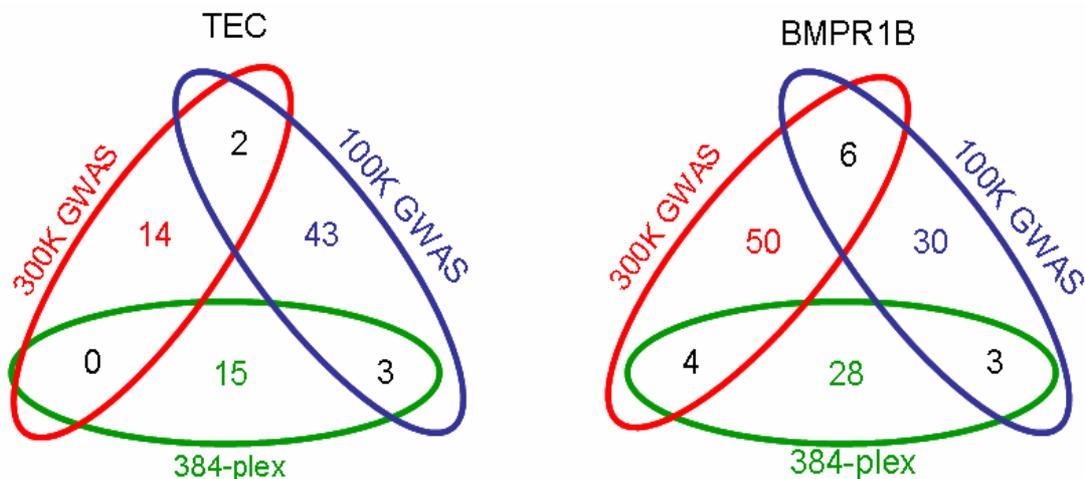
Per i marcatori che sono stati analizzati in almeno uno dei tre pannelli caso-controllo (330K GWAS, 100K GWAS o 384-plex) abbiamo confrontato i dati "reali", derivanti da genotipizzazione, con quelli imputati, e per ogni SNP abbiamo considerato i seguenti parametri:

- 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. 3 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 11.

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).

## DISCUSSIONE

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). Queste realtà 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 in un unico algoritmo cumulativo potrebbe permettere di stimare la probabilità che un singolo individuo contragga 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, contro i 16 selezionati da De Jager. Questa differenza potrebbe riflettere il fatto che il LES sia geneticamente più eterogeneo della SM, e pertanto la suscettibilità genetica al LES sia "dispersa" su un numero maggiore di loci, oppure potrebbe essere dovuto ad una conoscenza migliore della genetica del LES.

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).

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, PXX, 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 PXX 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. Riteniamo pertanto che questi dati non debbano essere considerati come una mancata replicazione dell'associazione precedentemente osservata, ma che siano da attribuire ad una mancanza di potere statistico dovuto essenzialmente alla bassa numerosità campionaria. E' pertanto necessario estendere l'analisi a un numero maggiore di pazienti e di controlli, che stiamo già raccogliendo nel nostro laboratorio.

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 2, 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 ( $\pm$  st. dev.) sono sovrapponibili, e la numerosità campionaria è troppo modesta per permetterci al momento di condurre elaborazioni statistiche più complesse. Anche in questo caso, quindi, il dato sembra promettente ma è indispensabile ampliare la nostra popolazione, nonché il numero di loci da analizzare.

Il secondo punto d'interesse nel lavoro svolto riguarda l'applicazione delle tecniche 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 "raccolgere 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.

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. 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 11 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 11 su una sottocasistica di campioni imputati, per verificare che i dati ottenuti in silico e quelli ottenuti in vivo siano sovrapponibili.

## BIBLIOGRAFIA

1. Brown KM, Sacks SH, Sheerin NS. 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
2. Browning S. Missing data imputation and haplotype phase inference for genome-wide association studies. *Hum Genet* 2008; 124(5): 439-450
3. Cervera R ... European Working Party on Systemic Lupus Erythematosus. Systemic Lupus Erythematosus: Clinical and Immunologic Patterns of Disease expression in a cohort of 1,000 Patients. *Medicine* 1993; 72(2): 113-123
4. D'Alfonso S ... Momigliano-Richardi P. Systemic lupus erythematosus candidate genes in the Italian population: evidence for a significant association with interleukin-10. *Arthritis Rheum.* 2000 Jan;43(1):120-8.
5. De Jager PL ... Karlson EW. Integration of genetic risk factors into a clinical algorithm for multiple sclerosis susceptibility: a weighted genetic risk score. *Lancet Neurol.* 2009 Dec;8(12):1111-9.
6. Deapen D ... Mack TM. A revised estimate of twin concordance in Systemic Lupus Erythematosus. *Arthritis & Rheumatism* 1992; 35(3): 311-318
7. Delgado-Vega AM ... Alarcón-Riquelme M: Replication of the TNFSF4 (OX40L) promoter region association with systemic lupus erythematosus. *Genes Immun.* 2009 April; 10(3): 248-253
8. Edberg JC ... Kimberly RP: Genetic variation in the CRP promoter: association with systemic lupus erythematosus. *Human Molecular Genetics* 2008; 17(8): 1147-1155
9. Gateva V ... Graham RR. A large-scale replication study identifies TNIP1, PRDM1, JAZF1, UHRF1BP1 and IL10 as risk loci for systemic lupus erythematosus. *Nat Genet.* 2009 Nov;41(11):1228-33
10. Graham RR ... Behrens TW: Review of recent genome-wide association scans in lupus. *Journal of International Medicine* 2009; 265:280-688
11. Graham RR ... Behrens TW: Review of recent genome-wide association scans in lupus. *Journal of International Medicine* 2009; 265:280-688
12. Graham RR ... Gaffney PM: Genetic variants near TNFAIP3 on 6q23 are associated with systemic Lupus Erythematosus (SLE). *Nature Genetics* 2008 September 40(9): 1059-1061
13. Grennan DM ... Badcock C. Family and twin studies in systemic lupus erythematosus. *Dis Markers* 1997; 13(2): 93-98
14. Guan Y, Stephens M. Practical issues in imputation-based association mapping. *PLoS Genetics* 2008; 4(12)
15. Hartung K ... Seuchter SA: Major Histocompatibility Complex Haplotypes and Complement C4 Alleles in Systemic Lupus Erythematosus. *J Clin Invest.* 1992 October; 90(4): 1346-1351
16. Hom G ... Association of Systemic Lupus Erythematosus with C8orf13-BLK and ITGAM-ITGAX. *N Engl J Med* 2008; 358: 900-909
17. Hrycek A ... Wilczok T. Human cytomegalovirus in patients with systemic lupus erythematosus. *Autoimmunity* 2005; 38(7):487-91
18. Jacob CO ... Mohan C: Identification of IRAK1 as a risk gene with critical role in the pathogenesis of systemic lupus erythematosus. *PNAS* 2009 April; 106(15):6256-6261
19. Kotzin BL. Systemic Lupus Erythematosus *Cell* 1996; 85: 303-306
20. Kotzin BL. Systemic Lupus Erythematosus *Cell* 1996; 85: 303-306
21. Kozyrev SV ... Alarcón-Riquelme M: Functional variants in the B-cell gene BANK1 are associated with systemic lupus erythematosus. *Nature Genetics* 2008 February; 40(2):211-484
22. Lee YH ... Sestak AL: The mannose-binding lectin gene polymorphisms and systemic Lupus Erythematosus. Two case-control studies and a meta-analysis. *Arthritis e Rheumatism* 2005 December; 52(12): 3966-3974
23. Lee YH... Harley JB The PTPN22 C1858T functional polymorphism and autoimmune diseases--a meta-analysis. *Rheumatology (Oxford).* 2007 Jan;46(1):49-56. Epub 2006 Jun 7.
24. Lindqvist & Alarcón-Riquelme. The Genetics of Systemic Lupus Erythematosus (Review). *Scandivian Journal of Immunology* 1999; 50(6): 562-577
25. Liu K ... D'Alfonso S...The Italian Collaborative Group...Mohan C: Kallikrein genes are associated with lupus and glomerular basement membrane-specific antibody-induced nephritis in mice and humans. *J Clin Invest.* 2009 April; 119(4): 911-923
26. Marchini J ... Donnelly P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nature Genetics* 2007; 39(7): 906-912
27. Musone S ... Criswell LA: Multiple polymorphisms in the TNFAIP3 region are independently associated with systemic lupus erythematosus. *Nature Genetics* 2008 September 40(9):1062-1064
28. Nath SK ... A nonsynonymous functional variant in the integrin-alpha(M) (encoded by ITGAM) is associated with systemic lupus erythematosus. *N Engl J Med* 2007; 357:977-986
29. Nothnagel M ... Franke A. A comprehensive evaluation of SNP genotype imputation. *Hum Genet* 2009; 125: 163-171
30. Oishi T... Matsuda K: A functional SNP in the NKX2.5-binding site of ITPR3 promoter is associated with susceptibility to systemic lupus erythematosus in Japanese population.
31. Orrù V... Nunzio Bottini: A loss-of-function variant of PTPN22 is associated with reduced risk of systemic lupus erythematosus. *Human Molecular Genetics* 2009; 18(3): 569-579

32. Prokunina L...Alarcón-Riquelme M: A regulatory polymorphism in PDCD1 is associated with susceptibility to systemic lupus erythematosus in humans. *Nature Genetics* 2002 December; 32:666-669
33. Remmers RF et al.: STAT4 and the risk of rheumatoid arthritis and systemic lupus erythematosus *Am J Hum Genet* 2005; 76:528-37
34. Rhodes B, Vyse T. Using genetics to deliver personalized SLE therapy – a realistic prospect? *Nature reviews rheumatology* 2010 6: 373-377
35. Sigurdsson S... Syvänen AC: 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
36. Simard JF, Costenbader KH. What can epidemiology tell us about systemic lupus erythematosus. *Int J Clin Pract* 2007; 61:1170-1180
37. Sturfelt G, Truedsson L. Complement and its breakdown products in SLE. *Rheumatology* 2005; 44:1227-1232.
38. Wallace JD, Hahn BH. *Duboi's Lupus erythematosus*. 1993; Lea & Febiger, Philadelphia and London; 4th edition.
39. Yang Q ... Flanders WD. How many genes underlie the occurrence of common complex diseases in the population? *Int J Epidemiol.* 2005 Oct;34(5):1129-37. Epub 2005 Jul 25. PubMed PMID: 16043441.
40. Yang Q ... Flanders WD. Improving the prediction of complex diseases by testing for multiple disease-susceptibility genes. *Am J Hum Genet.* 2003 Mar;72(3):636-49. Epub 2003 Feb 14. Erratum in: *Am J Hum Genet.* 2004 Feb;74(2):372. PubMed PMID: 12592605; PubMed Central PMCID: PMC1180239

Tabella 1: 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	
<b>*HLA-DR</b> Human Leukocyte Antigen	6p	<b>DR2</b>	<b>DRB1*1501</b>	1,87	1,51-2,32	<10 <sup>-7</sup>	356	975	Hartung 1992
		<b>DR3</b>	<b>DRB1*0301</b>	2,54	2,04-3,17	<10 <sup>-7</sup>	356	975	
<b>*PTPN22</b> Protein Tyrosine Phosphatase Non-receptor 22	1p	<b>rs2476601</b>	<b>R620W</b>	1,39	1,25-1,54	<10 <sup>-7</sup>	4234	9983	Lee 2007, Harley 2008
		<b>rs33996649</b>	<b>R263Q</b>	1,57	1,16-2,12	0,0017	2093	2348	Orru' 2009
<b>CRP</b> Pentraxin C-reactive protein	1q	<b>rs3093061</b>	<b>promotore -70'7</b>	1,71	1,42-2,05	<1X10 <sup>-7</sup>	1118	1032	Edberg 2008
<b>FCGR2A</b> Fc fragment of IgG, low affinity IIa, receptor	1q	<b>rs1801274</b>	<b>H166R</b>	1,35	1,20-1,52	6,78X10 <sup>-7</sup>	720	2337	Harley 2008
<b>SELP</b> selectin P	1q	<b>rs3917815</b>	<b>N673S</b>			5,74X10 <sup>-6</sup>	787 famiglie		Jacob 2007
<b>*TNFSF4 – OX40L</b> Tumor necrosis factor (ligand) superfamily, member 4	1q	rs10489265	nella regione	1,3	1,19-1,43	1,7X10 <sup>-8</sup>			Graham 2009
		rs1234317	nella regione	1,39		0,0009	1312	1801	Delgado-Vega 2009
		<b>rs2205960</b>	<b>nella regione</b>	1,22	1,15-1,30	6,3X10 <sup>-9</sup>	1963	4329	Gateva 2009
<b>NMNAT2</b> Nicotinamide nucleotide adenyltransferase 2	1q	<b>rs2022013</b>	<b>introne 1</b>	1,18	1,10-1,27	1,08X10 <sup>-7</sup>	2566	4162	Harley2008
<b>*IL10</b> Interleukin 10	1q	<b>rs3024505</b>	<b>nella regione</b>	1,19	1,11-1,28	4,0X10 <sup>-8</sup>	1963	4329	Gateva 2009
		IL10.G	promotore	1,78	1,19-2,66	0,042	172	164	D'Alfonso 2000
<b>TLR5</b> Toll-like receptor 5	1q		<b>R392X</b>			0,009	199	326genitori 75 parenti	Hawn 2005
<b>*STAT4</b> Signal transducer and activator of transcription 4	2q	<b>rs7574865</b>	<b>introne 3</b>	1,55	1,47-1,65	<10 <sup>-7</sup>	4908	7101	Remmers 2007, Harley 2008
		<b>PDCD1</b> Programmed cell death 1	2q	<b>PD1.3</b>	<b>introne 4</b>	RR=2,6	1,6-4,4	0,00001	880
<b>PXK</b> PX domain containing serine/threonine kinase	3p	<b>rs6445975</b>	<b>introne 4</b>	1,27	1,15-1,39	9,2X10 <sup>-7</sup>	2566	4162	Harley2008
<b>*BANK1</b> B-cell scaffold protein with ankyrin repeats 1	4q	rs10516487	R61H	1,38	1,25-1,53	3,74X10 <sup>-10</sup>	2003	1968	Kozyrev 2008
		<b>rs17266594</b>	<b>sito di spicing</b>	1,42	1,28-1,58	4,74X10 <sup>-11</sup>	1856	1774	
		rs3733197	A383T	1,23	1,11-1,36	4,67X10 <sup>-5</sup>	1819	1875	
<b>PTTG1</b> pituitary tumor-transforming 1	5q	<b>rs2431697</b>	<b>nella regione</b>	1,23	1,14-1,32	1,00X10 <sup>-10</sup>	2566	4162	Harley2008
<b>ITPR3</b> inositol 1,4,5-triphosphate receptor, type 3	6p	<b>rs3748079</b>	<b>promotore -1197</b>	1,88	1,51-2,35	1,78X10 <sup>-8</sup>			Oishi 2008
<b>ATG5</b> Autophagy protein 5	6q	<b>rs573775</b>	<b>introne 6</b>	1,19	1,12-1,27	1,36X10 <sup>-7</sup>	2566	4162	Harley 2008
		rs2245214	introne 2	1,15	1,09-1,21	1,2X10 <sup>-5</sup>	1963	4329	Gateva 2009
<b>TNFAIP3</b> Tumor necrosis factor, alpha-induced protein 3	6q	<b>rs5029939</b>	<b>introne 2</b>	2,29		2,89X10 <sup>-12</sup>	431	2155	Graham 2008
		rs13192841	nella regione	1,4	1,2-1,6	5,4X10 <sup>-8</sup>	1239	1629	Musone 2008
		rs2230926	F127C	2	1,4-3,0	3,0X10 <sup>-4</sup>	1239	1629	
		rs6922466	nella regione	1,3	1,1-1,4	1,0X10 <sup>-4</sup>	1239	1629	
<b>ICA1</b> Islet cell autoantigen 1	7p	<b>rs10156091</b>	<b>introne 8</b>	1,32	1,19-1,47	1,9X10 <sup>-7</sup>	2566	4162	Harley2008

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

**Tab. 2 Geni del progetto 384-plex**

<b>Gene_</b>	<b>Descrizione</b>	<b>crom.</b>	<b>N. bp</b>	<b>N.SNP</b>	<b>Espressione</b>
PLA2G4A	Phospholipase A2 group IVA	1	160028	19	Fegato, piastrine
FN1	fibronectin 1	2	75632	20	Epatociti, macrofagi, monociti
PXK	PX domain containing serine/threonine kinase	3	92261	11	<b>Linfociti B?</b>
CBLB	E3 ubiquitin-protein ligase CBL-B	3	210778	9	<b>Linfociti B</b>
HLA-DRB1	HLA class II histocompatibility antigen, DRB1-	6	11079	4	
TEC	tec protein tyrosine kinase	4	134081	15	<b>Linfociti B</b>
BMPR1B	Bone morphogenetic protein receptor type IB precursor	4	396948	30	cartilagine, milza, timo
ITGA1	Integrin alpha-1 precursor	5	165711	37	condrociti, linf. T CD8+, midollo osseo, piastrine, endotelio vascolare
F13A1	Coagulation factor XIII A chain	6	176584	51	piastrine, endotelio vascolare, fegato
COL21A1	collagen, type XXI, alpha 1 [	6	337504	20	matrice extracellulare
BAI3	brain-specific angiogenesis inhibitor 3	6	754144	54	endoteli
ATG5	Autophagy protein 5	6	141344	9	leucociti, milza, timo
LYN	Tyrosine-protein kinase Lyn	8	131567	3	<b>B,T, monociti, macrofagi, piastrine, neutrofil</b>
BLK	B lymphocyte kinase	8	70603	25	<b>Linf. B e precursori dei linf. T</b>
CD163	macrophage-associated antigen	12	32623	5	Macrofagi
GZMB	T-cell serine protease 1-3E	14	3313	5	Linf. NK, milza, midollo osseo, linfonodi
CD226	T lineage-specific activation antigen 1	18	93968	9	Linf. T attivati, piastrine
SAMSN1	SAM domain, SH3 domain and nuclear localization signals, 1	21	400982	20	<b>Linf.B, midollo osseo, milza, timo</b>
SCUBE1	signal peptide-CUB domain-EGF-related 1	22	139327	38	Endoteli, piastrine

**Tab. 3 Associazioni più significative nella popolazione italiana tra i polimorfismi del progetto 384-plex**

crom.	gene e polimorfismo	allele raro	freq_LES	freq_CT	CHISQ	P	OR	C.I. 95%	
2	FN1_384plex_SNP9	C	0.3022	0.2364	6.686	0.009715	1.398	1.084	1.804
3	PXK_384plex_SNP2	G	0.3112	0.3897	8.161	0.004281	0.7073	0.5575	0.8973
3	PXK_384plex_SNP3	G	0.3651	0.4429	7.603	0.005826	0.7232	0.5743	0.9108
6	F13A1_384plex_SNP46	G	0.4532	0.3649	9.82	0.001726	1.443	1.147	1.816
6	HLA-DRB1_384plex_SNP1	T	0.1835	0.1036	16.05	6.158e-005	1.944	1.399	2.701
6	HLA-DRB1_384plex_SNP3	T	0.1295	0.06607	14.19	0.0001652	2.103	1.419	3.117
6	HLA-DRB1_384plex_SNP4	T	0.139	0.06607	18	2.208e-005	2.282	1.546	3.368
6	COL21A1_384plex_SNP17	C	0.05469	0.1306	18.59	1.618e-005	0.385	0.2462	0.6018
8	BLK_384plex_SNP9	G	0.3651	0.4383	6.719	0.009539	0.7371	0.5852	0.9286
8	BLK_384plex_SNP11	T	0.286	0.3619	7.926	0.004872	0.7063	0.5542	0.9001
8	BLK_384plex_SNP14	T	0.4353	0.3559	8.015	0.00464	1.395	1.108	1.757
22	SCUBE1_384plex_SNP5	C	0.3592	0.2598	14.11	0.0001723	1.597	1.25	2.041
22	SCUBE1_384plex_SNP6	A	0.214	0.1336	13.87	0.0001962	1.765	1.306	2.387
22	SCUBE1_384plex_SNP10	C	0.3141	0.2139	15.79	7.079e-005	1.683	1.3	2.179
22	SCUBE1_384plex_SNP14	A	0.2572	0.1637	16.19	5.717e-005	1.769	1.337	2.341
22	SCUBE1_384plex_SNP21	G	0.2816	0.1772	18.93	1.355e-005	1.82	1.387	2.389

**Tab. 5 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. 6 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 4 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. 7 SNPs utilizzati per l'algoritmo di rischio.**

gene	SNP	N/L <sup>c</sup>	OR	weight (ln OR)	allele associatio
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) <sup>a</sup>	L	1,76	0,565	T
ATG5	rs573775	L	1,19	0,17	T
BLK	rs13277113 <sup>b</sup>	L	1,3	0,26	A
BLK	rs2248932 <sup>b</sup>	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

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

<sup>b</sup> I due marcatori non sono stati analizzati direttamente, ma è stato tipizzato un polimorfismo in LD elevato ( $D'=1$ ,  $r^2>0,9$ ) con essi.

<sup>c</sup> 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).

**Tab. 8 Associazione dei polimorfismi di CBLB (cromosoma 3) nella popolazione complessiva**

SNP	allele raro	freq_LES	freq_CT	CHISQ	P	OR	C.I. 95%	
CBLB_384plex_SNP1	C	0.3564	0.3415	1.373	0.2412	1.068	0.957	1.191
CBLB_384plex_SNP2	G	0.2346	0.2577	4.035	0.04457	0.883	0.7821	0.997
CBLB_384plex_SNP3	A	0.2863	0.2923	0.2528	0.6151	0.9709	0.8654	1.089
CBLB_384plex_SNP4	G	0.389	0.4088	2.31	0.1286	0.9207	0.8276	1.024
CBLB_384plex_SNP5	A	0.1956	0.2152	3.327	0.06815	0.8865	0.7789	1.009
CBLB_384plex_SNP6	A	0.1606	0.1765	2.528	0.1118	0.8929	0.7764	1.027
CBLB_384plex_SNP7	T	0.2352	0.239	0.1147	0.7348	0.979	0.866	1.107
CBLB_384plex_SNP8	T	0.1893	0.1895	0.0003184	0.9858	0.9988	0.8743	1.141

**Tab. 9 associazione dei polimorfismi di TEC (cromosoma 4) nella popolazione complessiva**

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. 10 associazione dei polimorfismi di BMPR1B (cromosoma 4) nella popolazione complessiva

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.11 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.

## APPENDICE: CENNI SUI GENI COINVOLTI IN QUESTO STUDIO

**PTPN22 (Protein Tyrosine Phosphatase Non-receptor 22)** Il gene PTPN22 si trova sul cromosoma 1p13.2 e codifica per la proteina Lyp, una proteina tirosina fosfatasi intracellulare che inibisce la via di attivazione del recettore delle cellule T (TCR) attraverso la defosforilazione di vari substrati. La proteina Lyp agisce formando un complesso con la proteina Csk. È stata osservata un'associazione significativa tra una variazione aminoacidica di PTPN22 (R620W) con più di una malattia autoimmune: LES, artrite reumatoide, diabete di tipo 1, morbo di Graves, artrite giovanile idiopatica. L'SNP R620W influenza il legame di Lyp a Csk, causando un aumento dell'attività fosfatase con una conseguente ulteriore riduzione della via di segnalazione del TCR, che predispone alle malattie autoimmuni. È 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 un secondo polimorfismo missense di Lyp (R263Q), una variante con perdita di funzione che rende Lyp meno efficiente nell'inibire il segnale di TCR. 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.

**CRP (pentraxin C-reactive protein).** Questa proteina è un importante modulatore della risposta immunitaria innata, che opsonizza i detriti cellulari ed i corpi apoptotici facilitandone la rimozione. È una proteina di fase acuta, i cui livelli di espressione aumentano durante la risposta infiammatoria. Inoltre CRP è anche in grado di attivare la cascata del complemento e di legare alcuni recettori per il frammento costante (Fc) delle Ig. È stato osservato un suo ruolo protettivo in alcuni modelli murini di malattie autoimmuni. Recentemente è stata descritta un'associazione tra un polimorfismo nel promotore di questo gene ed il LES.

**FCGR2A (Fc fragment of IgG, low affinity IIa, receptor).** Questo gene codifica per uno dei membri della famiglia del frammento costante delle immunoglobuline. FCGR2A è un recettore membranario che si trova sulla superficie di cellule fagocitiche come i macrofagi ed i linfociti neutrofili, ed è coinvolto nel processo della fagocitosi e nella clearance degli immunocomplessi.

**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.

**TNFSF4 (tumor necrosis factor (ligand) superfamily, member 4).** La proteina codificata da questo gene è una citochina che appartiene alla famiglia di proteine che legano il fattore di crescita tumorale (TNF). TNFSF4 è coinvolta nell'interazione con le cellule presentanti l'antigene, e media direttamente l'adesione dei linfociti T attivati con le cellule vascolari endoteliali.

**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.

**Cluster interleuchine sul cromosoma 1** Questo cluster comprende le interleuchine: IL10, IL19, IL20 e IL24. Originariamente era stata descritta un'associazione con una variante nel promotore di IL10, 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.

**TLR5 (Toll-like receptor 5)** Il gene TLR5 mappa nella regione cromosomica di suscettibilità al LES 1q41-42. Le proteine TLR costituiscono una sottofamiglia di proteine transmembrinarie 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. Dati recenti indicano un ruolo dei TLR nel rigetto dei trapianti. 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. Il ligando di TLR5 è la flagellina, il principale componente del flagello batterico e noto fattore virulento riconosciuto dal sistema immune nelle piante, negli insetti e nei mammiferi. 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. Un polimorfismo che introduce un codone di stop a livello del codone 392 ha dimostrato un debole effetto protettivo nei confronti del LES. La variante p.392X non è funzionale ed è associata con una marcata diminuzione nella produzione di citochine proinfiammatorie, in particolare IL-6, TNF- $\alpha$  e IL-1 $\beta$ .

**STAT4 (Signal transducer and activator of transcription 4).** STAT 4 è un regolatore critico della risposta immunitaria Th1 mediata indotta da IL12. La sua attivazione porta alla secrezione acuta di IFN- $\gamma$  nei linfociti T CD4<sup>+</sup> 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.

**PDCD1 (programmed cell death 1).** Il gene PDC1 codifica per la proteina PD-1, un recettore di membrana che agisce come inibitore dei linfociti T attraverso l'interazione con i suoi ligandi PDL1 e PDL2. 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 fase G1 che una risposta proliferativa. Normalmente è espresso sulla membrana dei linfociti B e T attivati e regola il meccanismo della tolleranza immunitaria in entrambi i tipi cellulari. Topi BALB/c knockout per PDCD1 sviluppano una cardiomiopatia autoimmune fatale con deposito di IgG sulla superficie dei cardiomiociti. In un altro modello murino il deficit di PDCD1 causa un fenotipo simile al LES con glomerulonefrite ed artrite. Un polimorfismo nel quarto introne (chiamato PD1.3) è risultato significativamente

associato con il LES in più di una popolazione. Inoltre la frequenza dell'allele A di PD1.3 é significativamente più elevata nei pazienti con nefrite. Questo polimorfismo cade in un sito di legame per un fattore trascrizionale (RUNX1), e potrebbe pertanto influenzare l'espressione di PDCD1.

**PXK (PX domain containing serine/threonine kinase).** E' una proteina chinasi le cui funzioni sono state indagate solo recentemente. E' in grado di legarsi alla Na,K-ATPasi e di modularne l'attività di trasporto ionico e quella ATPasica. E' pertanto importante nel mantenimento dell'omeostasi cellulare.

**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.

**PTTG1 (pituitary tumor-transforming 1)** PTTG1 è omologo alla proteina di lievito "securina", che impedisce alla proteina separina di stimolare la separazione dei cromatidi. PTTG1 è un substrato del complesso proteico che promuove l'anafasi (APC). Ha un'attività tumorigenica in vivo, ed è espressa a livelli elevati in molti tumori.

**ITPR3 (inositol 1,4,5-triphosphate receptor, type 3).** Le proteine della famiglia degli ITPR mediano la trasmissione del segnale al secondo messaggero attraverso il rilascio di Ca<sup>++</sup> dai depositi intracellulari, in risposta all'inosotolo trifosfato. Ci sono 3 sottotipi di ITPR, ampiamente espressi in tutti i tessuti.

**ATG5 (Autophagy protein 5).** E' un elemento indispensabile per l'autofagia.

**TNFAIP3 (tumor necrosis factor, alpha-induced protein 3).** E' un elemento chiave nella regolazione della via di segnale di NF-κB. Agisce a valle di TNF-α e dei Toll-like receptors attraverso l'ubiquitinazione delle proteine adattatrici RIP e TRAF6.

**ICA1 (islet cell autoantigen 1).** La proteina ICA1 è stata osservata sia nel citosol che legata alla membrana dell'apparato di Golgi e dei granuli segretori immaturi. Si ritiene che costituisca un autoantigene nel diabete mellito insulino-dipendente e nella sindrome di Sjogren primaria.

**IRF5 (interferon regulatory factor 5).** L'IFN-γ (interferone gamma) é un importante mediatore dell'immunità virale. I pazienti affetti da LES presentano livelli aumentati di IFN-α e IFN-γ ed è stato osservato che i livelli ematici di espressione di geni IFN-α e IFN-γ inducibili correlano con l'attività di malattia del LES. I topi lpr hanno livelli elevati di IFN-γ e topi knockout IFN-γ -/- presentano una regressione della malattia. Inoltre la somministrazione terapeutica di IFN-γ può

causare l'instaurarsi di sintomi simili a quelli del LES sia in modelli umani che murini. Non sono state riportate al momento associazioni convincenti con il gene dell'IFN. Associazioni ben più significative sono state osservate con i geni che codificano per proteine della via dell'interferone. I fattori regolatori dell'interferone (IRFs) costituiscono una famiglia di regolatori trascrizionali che possiedono un dominio legante il DNA con una struttura elica-giroelica. IRF5 è responsabile della risposta immunitaria innata durante le infezioni virali. E' espresso nei linfociti B e nelle cellule dendritiche. IRF5 attiva i geni delle citochine infiammatorie, tra cui TNF- $\alpha$ , IL-12, IL-6, IFN e i geni responsivi a IFN. E' stata recentemente osservata una forte associazione tra polimorfismi di IRF5 ed il LES. Il dato, inizialmente riscontrato in popolazioni scandinave, è stato successivamente confermato in altre popolazioni.

**BLK (B lymphoid tyrosine kinase) e LYN (v-yes-1 Yamaguchi sarcoma viral related oncogene homolog).** Sono entrambe tirosine chinasi non recettoriali. Questa famiglia di proteine ha un ruolo cruciale nel controllo del ciclo cellulare, dell'adesione, della migrazione, della proliferazione cellulari, del differenziamento e della sopravvivenza cellulari negli organismi complessi.

**MBL (Mannose Binding Lectin).** MBL è una proteina sierica che inizia la via di attivazione della cascata del complemento mediata dalla lectina legandosi al mannosio che si trova sulla superficie microbica. L'eccessiva attivazione del sistema del complemento, e la sua deposizione a livello tissutale è un fattore importante nella patogenesi della malattia. Il sistema del complemento è coinvolto nell'opsonizzazione e nella rimozione degli immunocomplessi e delle cellule apoptotiche. E' accertato che scompensi in queste funzioni possono causare infiammazioni ed autoimmunità, e costituiscono il principale meccanismo patogenetico nel LES. Il deficit genetico di alcune componenti della via classica di attivazione del complemento (C1, C2, C4) è un fortissimo fattore di predisposizione alla patologia. Sono stati osservati polimorfismi sia nella regione del promotore che codificanti che causano un'alterata espressione di MBL (ridotta o assente). Per alcuni di questi è emersa un'associazione con la suscettibilità al LES, confermata in una meta analisi su popolazioni di origine europea, africana e asiatica. Oltre all'effetto predisponente queste varianti sembrano avere anche un effetto modificatore sul fenotipo clinico, ma sono state osservate associazioni con parametri clinici diversi in popolazioni diverse, per cui potrebbe essere utile delucidare questo effetto nella popolazione italiana. La prevalenza di questo deficit nella popolazione generale è stimata tra il 5% e il 10%.

**PHRF1 (PHD and ring finger domains 1).** E' un fattore trascrizionale, i cui bersagli specifici non sono ancora noti.

**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- $\gamma$  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.

**ITGAM -ITGAX (integrin, alpha M - integrin, alpha X).** I geni ITGAM e ITGAX codificano per 2 proteine della classe delle integrine. L'integrina- $\alpha$ M (ITGAM), in combinazione con l'integrina- $\beta$ 2 forma un'integrina specifica per i leucociti. L'integrina risultante è importante nell'aderenza dei neutrofili e dei monociti all'epitelio stimolato e nella fagocitosi dei corpi opsonizzati dal complemento. Il legame della componente complementare attivata C3b con l'integrina- $\alpha$ M sulle cellule presentanti l'antigene determina la produzione di TGF $\beta$ 2 e IL10, un processo essenziale per l'induzione della tolleranza immunitaria.

**TYK2 (Tyrosine Kinase 2).** Questo gene codifica per una proteina della famiglia delle Janus chinasi (JAK). La proteina si lega ai domini citoplasmatici dei recettori per le citochine e li fosforila attuando così la trasmissione del segnale. Inoltre TYK2 è anche un componente delle vie di segnalazione degli interferoni di tipo I e III, e pertanto potrebbe avere un ruolo nell'immunità antivirale. La sostituzione aminoacidica (p.Val362Phe) associata al LES ricade in un dominio importante per le interazioni proteina-proteina.

**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- $\kappa$ B.

**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-1 $\beta$  e TNF $\alpha$ . Di conseguenza SCUBE1 potrebbe giocare un ruolo importante nella risposta infiammatoria e in quella trombotica.

**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- $\alpha$  e IFN- $\gamma$ . 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.

**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.

**TEC (Tec protein tyrosine kinase).** Le chinasi della famiglia a cui appartiene TEC sono coinvolte nelle vie di segnalazione intracellulare che fanno capo ai recettori delle citochine, agli antigeni di superficie dei linfociti, ai recettori accoppiati con proteine G eterotrimeriche ed alle integrine. Hanno un ruolo chiave nella regolazione della funzione immunitaria.

**BMPRI1B (Bone morphogenetic protein receptor, type IB)** E' una serina/treonina chinasi trans-membrana. I ligandi di questo recettore (BMP, Bone Morphogenetic Proteins) appartengono alla superfamiglia TGF-beta e sono coinvolti nella formazione dell'osso endocondrale e dell'embriogenesi.

**ITGA1 (Integrin, alpha 1)** Codifica per la sub unità alpha 1 del recettore per le integrine. Forma un eterodimero con la sub unità beta 1 creando un recettore per il collagene e la laminina. Il recettore eterodimerico è coinvolto nell'adesione cellulare e potrebbe giocare un ruolo nell'infiammazione e nella formazione di fibrosi.

**F13A1 (Coagulation factor XIII, A1 polypeptide)** Codifica per la subunità A del fattore XIII della coagulazione. Il fattore XIII è l'ultimo ad essere attivato nella cascata coagulativa. La sub unità A è quella che possiede l'attività catalitica, favorendo la formazione di legami crociati tra le molecole di fibrina.

**COL2A1 (Collagen, type II, alpha 1)** Codifica per la catena alpha-1 del collagene di tipo II, che si trova nella cartilagine e nell'umor vitreo dell'occhio. Mutazioni a carico di questo gene sono state associate con l'acondrogenesi, la condrodiplosia e con l'osteoartrite familiare ad esordio precoce.

**BAI3 (Brain-specific angiogenesis inhibitor 3)** E' un bersaglio della proteina p53. Codifica per un inibitore dell'angiogenesi.

**GZMB (granzyme B)** E' contenuto nei granuli secretori delle cellule NK e dei linfociti T citotossici. Entra nelle cellule bersaglio attraverso pori formati dalla perforina ed innesca la risposta apoptotica.

**SAMSN1 (SAM domain, SH3 domain and nuclear localization signals 1)** Appartiene ad una nuova famiglia di adattatori putativi e di proteine "scaffold" che contengono domini SH3 e SAM (steril alpha motif).

**PLA2G4A (phospholipase A2, group IVA (cytosolic, calcium-dependent))** Appartiene alla famiglia delle fosfolipasi citosoliche. Catalizza l'idrolisi dei fosfolipidi di membrana per rilasciare acido arachidonico, che viene successivamente metabolizzato in eicosanoidi. Gli eicosanoidi, che includono le prostaglandine ed i leucotrieni regolano l'emodinamica e la risposta infiammatoria. PLA2G4A è attivato dall'aumento della concentrazione intracellulare di  $Ca^{2+}$ , e tramite fosforilazione.

**FN1 (Fibronectin 1)** La fibronectina è una glicoproteina presente in forma dimerica solubile nel plasma, e come dimero o multimero sulla superficie cellulare e nella matrice extracellulare. È coinvolta nell'adesione e nel processo di migrazione cellulare, durante l'embriogenesi, la cicatrizzazione, la coagulazione, e la formazione di metastasi.

**CBLB (Cas-Br-M (murine) ecotropic retroviral transforming sequence b)** Codifica per una proteina multifunzionale, che funge da ubiquitina ligasi nella regolazione negativa dell'attivazione dei recettori dei linfociti B e T. Inoltre questa proteina è critica per l'induzione dell'anergia nelle cellule NKT. In un recente GWAS è emerso un suo probabile ruolo nella suscettibilità alla Sclerosi Multipla (SM). Topi *Cblb* <sup>-/-</sup> sono altamente suscettibili all'encefalomielite sperimentale autoimmune, il modello murino di SM.

**SEZIONE 2:**  
**ATTIVITA'FORMATIVA**

CORSI FREQUENTATI (anno 2007-2008)

-Corso di Inglese – Prof. Colin Irving-Bell.

CORSI FREQUENTATI (anno 2008-2009)

-Journal Club in lingua inglese – Prof.ssa Mara Giordano.

CORSI FREQUENTATI (anno 2009-20010)

-Journal Club in lingua inglese – Prof.ssa Mara Giordano.

SEMINARI FREQUENTATI (anno 2008-2009)

14 ottobre 2009, Università del Piemonte Orientale, Dipartimento di Scienze Mediche, via Solaroli 17, Novara: **"Regolazione del metabolismo del ferro: nuove acquisizioni"** Prof.ssa C. Camaschella Università Vita-Salute e IRCCS San Raffaele, Milano

4 Novembre 2009, Università del Piemonte Orientale, Dipartimento di Scienze Mediche, via Solaroli 17, Novara: **"Naproxinod, un nuovo anti-infiammatorio con elevata tollerabilita' cardiovascolare"** Dott.Ennio Ongini Direttore NicOx Research Institute Bresso, Milano

17 novembre 2009 Università del Piemonte Orientale, Dipartimento di Scienze Mediche, via Solaroli 17, Novara: **"ALK e tumorigenesi"** Prof. Giorgio Inghirami Università di Torino

30 novembre 2009, Università del Piemonte Orientale, Dipartimento di Scienze Mediche, via Solaroli 17, Novara: **"Use of MAR epigenetic regulators for therapeutic expression"** Prof. Nicolas Mermod Institute of Biotechnology University of Lausanne, Switzerland

14 dicembre 2009, Università del Piemonte Orientale, Dipartimento di Scienze Mediche, via Solaroli 17, Novara: **"Controllo funzionale dei linfociti T da parte di adenosina trifosfato (ATP) extracellulare"** Fabio Grassi Università di Milano

14 gennaio 2010, Università del Piemonte Orientale, Dipartimento di Scienze Mediche, via Solaroli 17, Novara: **"Clinical significance of Antinuclear Antibodies in Autoimmune liver disease"** Prof. Pietro Invernizzi Dep. of Internal Medicine, IRCCS, Istituto Clinico Humanitas (Milano)

28 gennaio 2010, Università del Piemonte Orientale, Dipartimento di Scienze Mediche, via Solaroli 17, Novara: **"Dal difetto genetico all'infiammazione"** Prof. Luigi Maiuri Università di Foggia

9 febbraio 2010, Università del Piemonte Orientale, Dipartimento di Scienze Mediche, via Solaroli 17, Novara: **"R & D of magnetic-nanoparticles and carriers for drugs"** Dott. Giovanni BALDI Colorobbia Italia Ce.Ri.Col Centro di Ricerche Colorobbia Vinci (Fi)

17 febbraio 2010, Università del Piemonte Orientale, Dipartimento di Scienze Mediche, via Solaroli 17, Novara: **"Computational analysis of protein-protein interactions"** Prof. Stefano Pieraccini Dipartimento di Chimica Fisica ed Elettrochimica Università di Milano

6 maggio 2010, Università del Piemonte Orientale, Dipartimento di Scienze Mediche, via Solaroli 17, Novara: "**The Molecular Biology of Human Papillomavirus Infection and Disease**" Dr. John Doorbar Division of Virology, MRC, National Institute for Medical Research, (London-UK)

10 maggio alle ore 2010 Università del Piemonte Orientale, Dipartimento di Scienze Mediche, via Solaroli 17, Novara: "**4D Imaging of Tumor Proteolysis: Impact of microenvironment**" Prof.ssa Bonnie F. Sloane Department of Pharmacology, Wayne State University, School of Medicine Distinguished Professor, Karmanos Cancer Institute, Adjunct Professor, Department of Biological Sciences, University of Windsor, Ontario, Canada

26 maggio 2010 Università del Piemonte Orientale, Dipartimento di Scienze Mediche, via Solaroli 17, Novara: "**Prediction of RNA-RNA interaction**" Dr. Hakim Tafer Bioinformatics Group Department of Computer Science and Interdisciplinary Center for Bioinformatics University of Leipzig - Germany

4 giugno Università del Piemonte Orientale, Dipartimento di Scienze Mediche, via Solaroli 17, Novara: "**Functional nanocomposite for pharmaceutical application from physicochemical viewpoints**" tenuto dal Dr. Mamoru Senna, RIKEN Innovation Center, Wako, Japan

10 giugno 2010 Università del Piemonte Orientale, Dipartimento di Scienze Mediche, via Solaroli 17, Novara: "**A novel control of interferon induction by HCV at the early hours of infection**" Prof.ssa **Eliane Meurs** Hepacivirus and Innate Immunity Unit, Pasteur Institut (Paris, France)

14 giugno 2010 Università del Piemonte Orientale, Dipartimento di Scienze Mediche, via Solaroli 17, Novara: "**fenotipo e cellule staminali tumorali nei gliomi maligni**" tenuto da Davide Schiffer Professore Emerito di Neurologia, Università di Torino, Direttore Centro di Neuro-Bio-oncologia della Fondazione Policlinico di Monza-Vercelli

15 Giugno 2010 Università del Piemonte Orientale, Dipartimento di Scienze Mediche, via Solaroli 17, Novara: "**Nuove strategie anticancro: i microRNA come agenti differenziativi per i tumori solidi**" Carola Ponzetto, Università di Torino.

30 giugno Università del Piemonte Orientale, Dipartimento di Scienze Mediche, via Solaroli 17, Novara: "**Identificazione di antigeni per la diagnosi e la terapia del carcinoma pancreatico mediante l'analisi sierologica del proteoma**" Prof. Franco Novelli, Università di Torino

5 luglio 2010 Università del Piemonte Orientale, Dipartimento di Scienze Mediche, via Solaroli 17, Novara: "**Role of membrane lipids in neuronal synapses organization**" Paola Camoletto, Università di Torino

17 giugno 2010 Università del Piemonte Orientale, Dipartimento di Scienze Mediche, via Solaroli 17, Novara: "**Human CD38: a long journey from prognostic marker to key element in the pathogenesis, maintenance and progression of chronic lymphocytic leukemia**" Dott.ssa Silvia Deaglio, Dipartimento di Genetica, Biologia e Biochimica, Università di Torino

17 giugno 2010 Università del Piemonte Orientale, Dipartimento di Scienze Mediche, via Solaroli 17, Novara: "**Combating Bioterrorism: New Antidotes for an Old Poison**" Dr. Steven R. Ellis University of Louisville Kentucky, USA

28 giugno 2010 Università del Piemonte Orientale, Dipartimento di Scienze Mediche, via Solaroli 17, Novara: "**Insufficienza Venosa Cerebrospinale Cronica: una nuova prospettiva per la Sclerosi Multipla ?**" Prof. Paolo Zamboni, Direttore Centro Malattie Vascolari, Università di Ferrara

28 giugno 2010 Università del Piemonte Orientale, Dipartimento di Scienze Mediche, via Solaroli 17, Novara: "**Le cellule staminali mesenchimali: potenziale terapeutico nelle malattie del motoneurone e della mielina**" Prof.ssa Franca Fagioli, Direttore OncoEmatologia Pediatrica AOU Regina Margherita – S. Anna – Torino

5 luglio 2010 28 giugno 2010 Università del Piemonte Orientale, Dipartimento di Scienze Mediche, via Solaroli 17, Novara: "**Role of membrane lipids in neuronal synapses organization**" Paola Camoletto, Università di Torino

6 luglio 2010 28 giugno 2010 Università del Piemonte Orientale, Dipartimento di Scienze Mediche, via Solaroli 17, Novara: "**Tissue Transglutaminase, a polyhedric enzyme: friend or foe?**" Dario Fortunati, Istituto Superiore di Sanità -Roma

6 settembre 2010 Università del Piemonte Orientale, Dipartimento di Scienze Mediche, via Solaroli 17, Novara: "**Ruolo delle microvescicole nella comunicazione tra cellule**", Giovanni Camussi, Università di Torino

17 settembre 2010 Università del Piemonte Orientale, Dipartimento di Scienze Mediche, via Solaroli 17, Novara: "**Prevention and cure of autochthonous mammary carcinomas in her2 transgenic mice?**" Federica Cavallo, Università di Torino

22 settembre 2010 Università del Piemonte Orientale, Dipartimento di Scienze Mediche, via Solaroli 17, Novara "**Le cellule staminali tumorali: dalla biologia alla clinica**" Dr. Ruggero De Maria, Istituto Superiore di Sanità

23 settembre Università del Piemonte Orientale, Dipartimento di Scienze Mediche, via Solaroli 17, Novara "**Preventing Persisters: Targeting Epigenetic Changes in Cancer Chemotherapy**" Dr. Steven R. Ellis, University of Louisville, Kentucky, USA

Venerdì 24 settembre 2010 Dipartimento di Genetica, Biologia e Biochimica sezione di Genetica -- via Santena 19, Torino: "**The 1000 genomes project**" Gilean Mc VEAN (Oxford, UK)

#### CONGRESSI FREQUENTATI (elenco completo: denominazione congresso, sede, data)

-20/12/2007 - Prof. Ciro Isidoro: **Symposium in: NANOTECHNOLOGY AND BIOIMAGING novel approaches in diagnosis and drug delivery**

-XI Congresso Nazionale SIGU Genova, 23-25 novembre 2008

-28/01/2009 – ARESS: **iniziativa formativa: L'appropriatezza dei test genetici nel campo delle malattie complesse (Celiachia, Diabete tipo I, Trombofilia) e della farmaco genetica**

-XII Congresso Nazionale SIGU Torino, 8-10 novembre 2009

-Autoimmunity and Immune Mediated Disorders, Milan, 23-24 aprile 2010

-24th European Immunogenetics and Histocompatibility conference (EFI) Florence 15-18 maggio 2010

#### COMUNICAZIONI A CONGRESSI (elenco completo: autori, titolo, denominazione congresso, sede, data)

A) *Comunicazioni presentate personalmente (orali o poster)*

**Barizzone N**, Monti S, Mellone S, Marchini M, Scorza R, Danieli MG, D'Alfonso S.

Variazioni rare del gene TREX1 nella suscettibilità alle malattie autoimmuni. (**poster**)

XII Congresso Nazionale SIGU – Torino, Centro Congressi Lingotto, 8-10 novembre 2009. Agli atti: poster 424

**Barizzone N**, Cappellano G, Orilieri E, Chiocchetti A, Carlomagno Y, Marchini M, Scorza R, Danieli MG, Sabbadini MG, Sebastiani GD, Migliaresi S, Galeazzi M, Dianzani U, D'Alfonso S.

Analisi del gene PRF1 in 2 malattie autoimmuni: Lupus Eritematoso Sistemico e Sclerodermia. **(poster)**  
XII Congresso Nazionale SIGU – Torino, Centro Congressi Lingotto, 8-10 novembre 2009. Agli atti: poster 425

**Barizzone N**, Cappellano G, Orilieri E, Chiocchetti A, Carlomagno Y, Marchini M, Scorza R, Danieli MG, Sabbadini MG, Sebastiani GD, Migliaresi S, Galeazzi M, Dianzani U, D'Alfonso S.

Genetic analysis of perforin in two autoimmune diseases: Systemic Lupus Erythematosus and Systemic Sclerosis. **(presentazione orale)**

24th European Immunogenetics and Histocompatibility Conference (EFI) Firenze, palazzo dei congressi, 15-18 maggio 2010

**Barizzone N**, Monti S, Mellone S, Marchini M, Scorza R, Danieli MG, D'Alfonso S.

Rare variants in the TREX1 gene and susceptibility to autoimmune diseases.

**(presentazione orale)**

24th European Immunogenetics and Histocompatibility Conference (EFI) Firenze, palazzo dei congressi, 15-18 maggio 2010

#### *B) Altre comunicazioni*

-D'Alfonso S, Bolognesi E, Guerini FR, Bergamaschi L, **Barizzone N**, Ferrante D, Castelli L, Bocca S, Agliardi C, Ferrante P, Naldi P, Maurizio Leone M, Caputo D, Ballerini C, Salvetti M, Galimberti D, Massacesi L, Trojano M, Momigliano-Richiardi P.

Associazione di una variazione di sequenza nel gene MOG (Myelin Oligodendrocyte Glycoprotein) con la suscettibilità alla Sclerosi Multipla (SM)

X congresso nazionale SIGU, Salsomaggiore Terme, novembre 2007

**Barizzone N**, Marchini M, Cappiello F, Chiocchetti A, Nicola S, Ferrante D, Corrado L, Mellone S, Scorza R, Dianzani U, Momigliano-Richiardi P, D'Alfonso S.

Associazione di un polimorfismo del gene dell'Osteopontina con la Sclerosi Sistemica (Sclerodermia)

XVI Convegno Nazionale AIBT, Verbania, Scuola di Formazione di Polizia Penitenziaria, 1-3 ottobre 2009, pag. 16-17

Bergamaschi L, Leone M, Fasano ME, Guerini FR, Ferrante, Bolognesi E, **Barizzone N**, Corrado L, Naldi P, Agliardi C, Dametto E, Salvetti M, Visconti A, Galimberti D, Scarpini E, Cavalla P, Bergamaschi R, Caputo D, Momigliano-Richiardi P, D'Alfonso S.

Marcatori HLA di Classe I modulano la suscettibilità a Sclerosi Multipla.

XVI Convegno Nazionale AIBT, Verbania, Scuola di Formazione di Polizia Penitenziaria, 1-3 ottobre 2009, pag. 19-20.

Mellone S, Mellai M, **Barizzone N**, Bolognesi E, Momigliano-Richiardi P, Greco L, Giordano M.

Ricerca di fattori genetici di suscettibilità alla malattia celiaca nella regione 5q31-33.

XVI Convegno Nazionale AIBT, Verbania, Scuola di Formazione di Polizia Penitenziaria, 1-3 ottobre 2009, pag. 49-50.

**Barizzone N**, Monti S, Mellone S, Marchini M, Scorza R, Danieli MG, D'Alfonso S.

Variazioni rare del gene TREX1 nella suscettibilità alle malattie autoimmuni.

XII Congresso Nazionale SIGU – Torino, Centro Congressi Lingotto, 8-10 novembre 2009. Abstract n. 424

**Barizzone N**, Cappellano G, Orilieri E, Chiocchetti A, Carlomagno Y, Marchini M, Scorza R, Danieli MG, Sabbadini MG, Sebastiani GD, Migliaresi S, Galeazzi M, Dianzani U, D'Alfonso S.

Analisi del gene PRF1 in 2 malattie autoimmuni: Lupus Eritematoso Sistemico e Sclerodermia.

XII Congresso Nazionale SIGU – Torino, Centro Congressi Lingotto, 8-10 novembre 2009. Abstract n. 425

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, Vercellino M, Bergamaschi R, Collimedaglia L, Vecchio D, Caputo D, Momigliano-Richiardi P, D'Alfonso S.

Marcatori HLA di classe I modulano la suscettibilità alla Sclerosi Multipla.

XII Congresso Nazionale SIGU – Torino, Centro Congressi Lingotto, 8-10 novembre 2009. Abstract n. 453

Castillejo-Lopez C, Delgado-Vega A M., Wojcik J, Kozyrev S, Egido Juan R. Lopez, Sanchez E, Pollmann D, Fineschi S, Dominguez N, Lu R, James JA., Merrill J T., Kelly J A., Kaufman KM., Moser K L., Gilkeson GS, Pons-Estel B, **Barizzone N**, Witte T, Callejas JL, Harley Jb, Gaffney P M., Martin J, Guthridge J M., Alarcon-Riquelme M E.

Genetic Interactions Reveal a Novel B-Cell Pathway in Systemic Lupus Erythematosus

Arthritis & Rheumatism, Volume 60, October 2009 Abstract Supplement

The 2009 ACR/ARHP Annual Scientific Meeting Philadelphia October 16-21, 2009.

Gonzalez A, Suarez-Gestal M, Calaza MI, Liz M, Ordi-Ros J., Balada E., Bijl M, Kallenberg C.G.M., Papasteriades C, Kappou-Rigatou I, Carreira PE., Skopouli Fotini N., Mavromati M, Schmidt R E., Witte T, Endreffy E, Kovacs A, Marchini M, Scorza R, Migliaresi S, Sebastiani G D, Santos M J., Vinagre F, Suarez A, Gutierrez C, Rego I, Blanco F.J., **Barizzone N**, D'Alfonso S, Pullmann Jr R, Pullmann R, Ruzickova S, Dostal Cr, Gomez-Reino J. J.

Weak Association of Systemic Lupus Erythematosus Clinical Features with Susceptibility Alleles

Arthritis & Rheumatism, Volume 60, October 2009 Abstract Supplement

The 2009 ACR/ARHP Annual Scientific Meeting Philadelphia October 16-21, 2009

**Barizzone N**, Cappellano G, Orilieri E, Chiocchetti A, Carlomagno Y, Marchini M, Scorza R, Danieli MG, Sabbadini MG, Sebastiani GD, Migliaresi S, Galeazzi M, Dianzani U, D'Alfonso S.

Genetic analysis of perforin in two autoimmune diseases: Systemic Lupus Erythematosus and Systemic Sclerosis. (presentazione orale)

24th European Immunogenetics and Histocompatibility Conference (EFI) Firenze, palazzo dei congressi, 15-18 maggio 2010.

Tissue Antigen 2010, 75(5): Abstract O19, pag. 474.

**Barizzone N**, Monti S, Mellone S, Marchini M, Scorza R, Danieli MG, D'Alfonso S.

Rare variants in the TREX1 gene and susceptibility to autoimmune diseases.

(presentazione orale)

24th European Immunogenetics and Histocompatibility Conference (EFI) Firenze, palazzo dei congressi, 15-18 maggio 2010

Tissue Antigen 2010, 75(5): Abstract O18, pag. 473

Bergamaschi\* L., **Barizzone\*** N, Ban M, Fasano ME, Guerini FR, Corrado L, Vecchio D, Ruggerone S, Calzoni S, Agliardi C, Dametto E, Salvetti M, Mechelli R, Galimberti D, Scarpini E, Cavalla P, Bergamaschi R, Caputo D, Cordera S, Momigliano-Richiardi P, D'Alfonso S.

Associazione di marcatori HLA di classe II con la Sclerosi Multipla nelle popolazioni italiana e inglese: evidenziamo due effetti protettivi indipendenti.

Verrà presentato XIII Congresso Nazionale SIGU 2010, Firenze

Bergamaschi L, Ban M, Barizzone N, Leone M, Fasano ME, Guerini FR, Corrado L, Naldi P, Agliardi C, Dametto E, Salvetti M, Visconti A, Galimberti D, Scarpini E, Cavalla P, Bergamaschi R, Caputo D, Cordera S, Monaco F, Momigliano-Richiardi P, D'Alfonso S.

Association of HLA class I markers with Multiple Sclerosis in the Italian and UK population: evidence of two independent protective effects

Verrà presentato a. 60<sup>th</sup> congress of American Society of Human Genetics 2010, Washington DC.

Barizzone N, Monti S, Mellone S, Cappellano G, Orilieri E, Chiocchetti A, Marchini M, Scorza R, Danieli MG, Carlomagno Y, Sabbadini MG, Sebastiani GD, Migliaresi S, Galeazzi M, Dianzani U, Momigliano-Richiardi P, D'Alfonso S.

Rare variants in two functionally-related genes (TREX1 and PRF1) and susceptibility to autoimmune diseases.

Verrà presentato a. 60<sup>th</sup> congress of American Society of Human Genetics 2010, Washington DC.

ARTICOLI SCIENTIFICI PUBBLICATI NEL CORSO DEL DOTTORATO (elenco completo)

-Ferreiro-Neira I, Calaza M, Alonso-Perez E, Marchini M, Scorza R, Sebastiani GD, Blanco FJ, Rego I, Pullmann R Jr, Pullmann R, Kallenberg CG, Bijl M, Skopouli FN, Mavromati M, Migliaresi S, **Barizzone N**, Ruzickova S, Dostal C, Schmidt RE, Witte T, Papasteriades C, Kappou-Rigatou I, Endreffy E, Kovacs A, Ordi-Ros J, Balada E, Carreira P, Gomez-Reino JJ, Gonzalez A. Opposed independent effects and epistasis in the complex association of IRF5 to SLE.

*Genes Immun.* 2007 8(5):429-38.

-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.

-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, Lastrup H, González-Escribano MF, Martin J, Abderrahim H, Alarcón-Riquelme ME.

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

*Nat Genet.* 2008 40(2):211-6.

Erratum in: *Nat Genet.* 2008 Apr;40(4):484. Barizzone, Nadia [corrected to Barizzone, Nadia].

-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.

-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.

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

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

-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. Collaborators: **Barizzone N**, Galeazzi M, Danieli MG, Migliaresi S, Bozzolo E.

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.

-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 Mar;11(2):173-80.

-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.