



A.D. MDLXII

UNIVERSITÀ DEGLI STUDI DI SASSARI

SCUOLA DI DOTTORATO DI RICERCA IN SCIENZE BIOMEDICHE

Direttore della Scuola Prof. Franca Deriu

INDIRIZZO SCIENZE BIOMEDICHE INDIRIZZO DI GENETICA MEDICA,
MALATTIE METABOLICHE E NUTRIGENOMICA

Responsabile di Indirizzo Prof. Francesco Cucca

XXVI CICLO

**ANALISI GENETICA E CARATTERIZZAZIONE FUNZIONALE
DEL GENE *CBL-B* ASSOCIATO ALLA PREDISPOSIZIONE
NELLA SCLEROSI MULTIPLA**

Direttore *Prof.ssa Franca Deriu*

Tutor *Prof. Francesco Cucca*

Dottoranda *Dott.ssa Alessia Loi*

Anno Accademico 2012 – 2013

INDICE

1. INTRODUZIONE	Pag. 4
1.1 <i>Sclerosi Multipla; caratteristiche generali della patologia</i>	4
1.2 <i>Epidemiologia</i>	5
1.3 <i>Basi eziopatogenetiche</i>	7
1.4 <i>Modello patogenetico</i>	9
1.5 <i>Dissezione delle basi genetiche della Sclerosi Multipla</i>	11
1.6 <i>Geni identificati tramite studi di associazione su tutto il genoma (GWAS)</i>	13
1.7 <i>Studio GWAS nella popolazione sarda</i>	16
1.8 <i>Il gene CBL-B</i>	18
2. SCOPO DELLA RICERCA	Pag. 25
3. DISEGNO SPERIMENTALE DELLO STUDIO	Pag . 27
3.1 <i>Studio GWAS caso-controllo</i>	27
3.2 <i>Studi funzionali per chiarire il ruolo di varianti di CBL-B nella SM</i>	28
3.2.1 <i>Studio dell'espressione del gene CBL-B in linfociti T e B</i>	28
3.2.2 <i>Studio dei pathway di attivazione dei linfociti T e B</i>	29
4. MATERIALI E METODI	Pag. 30
4.1 <i>Descrizione della casistica</i>	30
4.2 <i>Estrazione DNA</i>	31
4.3 <i>Genotipizzazione mediante l'utilizzo di GeneChip arrays Affymetrix</i>	31
4.4 <i>Sequenziamento Next Generation di DNA genomico</i>	32
4.5 <i>Analisi delle sequenze NGS</i>	33
4.6 <i>Pannelli di referenza per l'Imputazione</i>	35
4.7 <i>Tipizzazione INDEL rs147983726</i>	37
4.8 <i>Test di associazione</i>	37

4.9	<i>Estrazione di cellule mononucleate del sangue periferico (PBMC) da sacca di buffy-coat</i>	38
4.10	<i>Separazione Linfociti T e B</i>	39
4.11	<i>Estrazione dell'RNA da linfociti T e B e RT-qPCR</i>	39
4.12	<i>Stimolazione dei Linfociti T e B</i>	42
4.13	<i>Esperimenti ELISA di quantificazione di IL-2 e dell'attivazione di NF-κB</i>	42
5.	RISULTATI	Pag. 44
5.1	<i>Studio caso-controllo GWAS</i>	44
5.2	<i>Studio di espressione del gene CBL-B</i>	48
5.2.1	<i>Esperimenti di RT-qPCR nelle cellule T</i>	49
5.2.2	<i>Esperimenti RT-qPCR su cellule B e linee Linfoblastoidi (LCL)</i>	52
5.3	<i>Studio dell'attivazione di cellule T mediante saggi ELISA</i>	55
5.4	<i>Studio dell'attivazione di cellule B mediante saggi ELISA</i>	56
6.	DISCUSSIONE	Pag. 58
7.	PROBLEMATICHE AFFRONTATE NELLO STUDIO	Pag. 61
8.	CONCLUSIONI E SVILUPPI FUTURI	Pag. 62
9.	BIBLIOGRAFIA	Pag. 64
10.	RINGRAZIAMENTI	Pag. 70

1 INTRODUZIONE

1.1 Sclerosi Multipla; caratteristiche generali della patologia

La Sclerosi Multipla (SM) è la più frequente malattia infiammatoria cronica demielinizzante del sistema nervoso centrale (SNC) e costituisce, se si escludono i traumi, la più frequente causa d'invalidità neurologica negli adolescenti ed adulti. Il nome della patologia deriva dalle molteplici aree "sclerotiche" visibili all'esame macroscopico del cervello. Le aree in cui la mielina viene danneggiata vengono chiamate "placche", da ciò deriva appunto la denominazione sclerosi a placche.

La SM è considerata una malattia multifattoriale, in quanto la sua insorgenza dipende dalla complessa interazione di fattori genetici di predisposizione in presenza di fattori ambientali scatenanti.

Le patologie idiopatiche croniche demielinizzanti rappresentano un gruppo eterogeneo e sono contraddistinte da una degenerazione della mielina e delle guaine assionali, con conseguente alterazione della velocità di trasmissione degli impulsi nervosi. La disabilità che ne consegue si manifesta a livello delle funzioni motorie, autonome e neuro-cognitive.

Le lesioni si distribuiscono preferenzialmente a livello periventricolare, a livello del corpo calloso e della sostanza bianca del cervelletto. Zone frequentemente colpite sono anche i nervi ottici e il midollo spinale (1). La diversa distribuzione e il numero delle placche danno luogo, nei pazienti affetti da SM, ad un'ampia variabilità di manifestazioni cliniche.

Resta ancora da comprendere la natura dei fattori che comportano tanta eterogeneità, sebbene l'ipotesi più plausibile sostenga che essa sia legata ad un determinato assetto genetico che conduce, in maniera diversa a seconda degli individui, ad alterazioni della risposta immune così come ad una particolare vulnerabilità del sistema nervoso centrale e alterata capacità di rigenerazione (2).

Dal punto di vista clinico sono state descritte quattro forme principali di SM, ciascuna delle quali può avere un decorso severo, moderato o leggero.

La più comune (circa l'85% dei casi) è la forma recidivante-remittente (RR), caratterizzata dal manifestarsi di sintomi acuti o recidive di deficit neurologici, seguiti da un recupero completo o incompleto chiamato remissione.

Dopo un numero variabile di anni, circa il 50% dei pazienti sviluppa una forma secondaria progressiva (SP), caratterizzata da una progressione della disabilità neurologica associata o meno ad episodi di recidive. Un'ipotesi in attesa di conferma sperimentale riconosce in queste due forme della malattia la manifestazione di un'unica patologia, che si sviluppa dapprima in seguito all'istaurarsi di aree di infiammazione, che comporta successivamente l'innescarsi di un processo strettamente neurodegenerativo (3).

Altri tipi meno comuni di SM includono una forma primaria progressiva (PP), che colpisce il 10% dei pazienti, in cui si manifesta una lenta ma costante progressione della disabilità e dei sintomi neurologici senza riacutizzazioni, generalmente associata ad una non responsività alle immunoterapie, tanto da essere ritenuta da molti una malattia a sé stante (4).

La forma progressiva-recidivante (PR), costituisce circa il 5% dei casi, in cui i pazienti sono colpiti da un'insidiosa progressione della disabilità con qualche raro episodio di recidiva, ma senza fasi di recupero.

Infine, esiste una quinta forma di SM detta benigna, che esordisce con uno o due episodi acuti, seguiti da un recupero completo che non lascia tracce di disabilità e tende a essere associata a sintomi sensitivi (parestesie) o visivi (neurite ottica).

I modelli animali di SM generalmente utilizzati, come l'encefalomielite autoimmune sperimentale (EAE), sono caratterizzati sia dalla forma progressiva che dalla recidivante-remittente della malattia, riflettendo così le due principali forme di SM (5).

1.2 Epidemiologia

L'Organizzazione Mondiale della Sanità (OSM), stima che nel mondo ci siano oltre 2,5 milioni di persone affette da SM. La patologia è molto frequente tra le popolazioni caucasiche (soprattutto tra quelle residenti nel Nord-Ovest europeo), nel Nord America, nel Sud-Est dell'Australia e in Nuova Zelanda, Sud-Africa e America Meridionale, mentre si riscontra una bassa incidenza in Asia e nelle regioni caraibiche (Fig.1). Dal punto di vista epidemiologico la SM colpisce con una frequenza doppia le donne rispetto agli uomini, ed ha un esordio perlopiù tra i 20 e 40 anni, aumentando progressivamente fino ai trentacinque anni, per poi diminuire gradualmente.

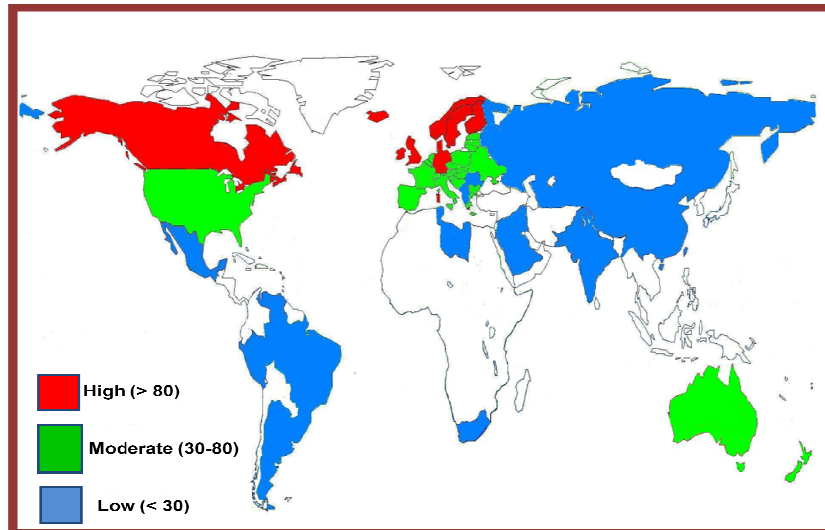


Fig. 1: Prevalenza della SM nel mondo (x 100.000)
(Atlas Multiple Sclerosis World Health Organization 2008)

In Europa si stima ci siano circa 600.000 persone affette, di cui circa 50.000 solo in Italia. La distribuzione sembra seguire un gradiente Nord-Sud, con una maggiore prevalenza nei Paesi del Nord, soprattutto in Scandinavia. La Sardegna rappresenta un'eccezione, con una prevalenza due volte superiore rispetto al resto della popolazione italiana e alla maggior parte delle popolazioni caucasiche (140 casi per 100,000 abitanti), equiparabile alla prevalenza dei Paesi Nord Europei, come indicato dalla figura 2 (6,7).

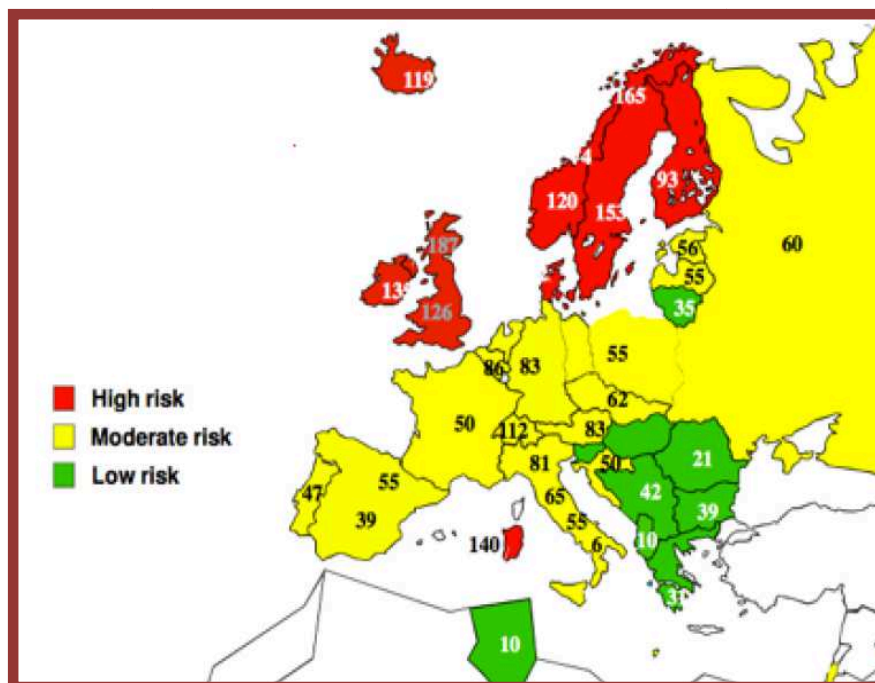


Fig. 2: Prevalenza della SM in Europa (x 100.000) (Pugliatti et al. 2006)

1.3 Basi eziopatogenetiche

Malgrado l'intensa attività di ricerca condotta dalla comunità scientifica negli ultimi decenni, i meccanismi che determinano la comparsa della SM e che influenzano il decorso di questa malattia sono ancora poco conosciuti, e a tutt'oggi, l'eziopatogenesi non è stata chiarita.

La maggior parte degli studi suggerisce che il danno della guaina mielinica, e la conseguente compromissione neurologica, avvengano in seguito ad una risposta anomala del sistema immunitario, che generalmente ha il compito di proteggere e difendere l'organismo da patogeni ed agenti estranei, ma che in alcuni casi attacca i propri tessuti. Attualmente non è noto cosa scateni questa reazione autoimmune, sebbene siano stati evidenziati diversi fattori che contribuiscono a determinare l'insorgenza di questa patologia, tra cui:

- **il genere:** le donne hanno una probabilità da 2 a 3 volte superiore di sviluppare la malattia rispetto agli uomini;
- **la storia clinica familiare:** una storia clinica di ricorrenza familiare suggerisce la presenza di un rischio maggiore di sviluppare SM. Il rischio è tanto maggiore, quanto più stretta è la relazione familiare; per esempio nel caso di fratelli gemelli monozigoti, il rischio di ammalarsi se uno dei fratelli è affetto è di circa il 30%;
- **l'etnia:** la SM è molto più frequente nelle popolazioni di origine caucasiche che in altri gruppi etnici;
- **fattori ambientali :** la SM è più frequente tra gli individui che vivono nei climi nordici, mentre le regioni tropicali e meridionali vengono considerate a basso rischio. Inoltre, diversi studi preliminari suggeriscono che il luogo in cui si vive nei primi 15 anni di vita possa influenzare il rischio di sviluppare SM. Se infatti dopo i 15 anni ci si trasferisce da una regione a basso rischio ad una ad alto rischio, la probabilità di sviluppare la SM rimane bassa. Questo aveva fatto ipotizzare un effetto protettivo dell'esposizione alla luce solare e di alti livelli di vitamina D sulla SM. Tuttavia gli studi condotti finora non sono conclusivi a riguardo e non sembrano confermare queste ipotesi;
- **infezioni:** in individui con una specifica predisposizione genetica alcuni tipi di infezioni batteriche e/o virali (tra cui l'Epstein-Barr) sono in grado di scatenare una reazione autoimmune. Infatti poiché le proteine (ed epitopi) di questi microrganismi sono simili a quelle prodotte dal nostro organismo, esse sono in grado di provocare un attacco ai nostri tessuti da parte di auto-anticorpi cross-reattivi anche dopo che l'infezione è risolta.

La SM viene definita una malattia complessa, o multifattoriale per via delle diverse componenti che concorrono a determinarne l'insorgenza, quali le complesse interazioni di fattori genetici di predisposizione con fattori ambientali scatenanti.

Le cause ambientali della malattia rimangono ancora ignote e difficili da dissezionare mentre la ricerca dei fattori genetici, per molti anni ha rappresentato, ed ancora oggi rappresenta, un'ardua sfida.

Solo negli ultimi, grazie ai progressi nella tecnologia ed in seguito all'avvento degli studi di associazione su tutto il genoma (GWAS) notevoli passi avanti sono stati compiuti sulla conoscenza delle componenti genetiche coinvolte nella patologia.

L'importanza del ruolo dei fattori genetici nella predisposizione alla malattia è stata messa in evidenza dai numerosi studi effettuati sull'ereditarietà della MS.

Questi hanno mostrato che il rischio di sviluppare la malattia aumenta con l'aumentare del grado di parentela con la persona affetta, con un rischio di ricorrenza familiare (sibling recurrence-risk ratio λ_s) di $\sim 6,3$ (8). In generale il rischio si riduce dal 3% nei parenti di primo grado (fratelli 5%, genitori 2% e figli 2%) all'1% nei parenti di secondo e terzo grado biologico. Questo conferisce ai parenti un rischio relativo rispettivamente di 9,2, 3,4 e 2,9, rispetto alla popolazione generale (9). Studi sui gemelli in Canada e Regno Unito hanno mostrato un tasso di concordanza maggiore nei gemelli omozigoti rispetto a quelli dizigoti (25% vs 5%) (10). Individui adottati subito dopo la nascita che hanno nella loro famiglia adottiva dei membri malati di SM hanno lo stesso rischio della popolazione generale e quindi inferiore rispetto ai parenti biologici del caso indice (11). Il rischio è inoltre elevato nei figli con entrambi i genitori affetti piuttosto che con un solo genitore affetto (12). Questi studi mostrano come, nei cluster familiari e nella suscettibilità individuale, i fattori genetici siano determinanti (13). Allo stesso tempo, queste osservazioni evidenziano una penetranza incompleta della malattia e sottolineano la necessità della presenza di fattori ambientali permissivi per la piena estrinsecazione del rischio (14).

Sebbene i geni giochino un ruolo importante nello sviluppo della SM, non possono spiegarne completamente l'eziologia. Secondo le stime più aggiornate, la genetica, infatti, spiega solo circa il 10-20% della suscettibilità alla malattia, attraverso l'azione congiunta di più varianti alleliche a diversi loci, dislocati lungo il genoma, e di cui circa l'80% sono ancora sconosciuti. Oltre all'aspetto genetico sta emergendo infatti l'importanza di meccanismi post-trascrizionali ed epigenetici che contribuiscono al rischio globale di sviluppare la SM.

Attualmente, non è chiaro quali siano i fattori ambientali coinvolti nell'eziologia della SM. Infatti la loro identificazione è complicata da numerose variabili. La difficoltà principale deriva dal fatto che la manifestazione clinica della malattia, si riscontra solo dopo molti anni dall'inizio del processo patogenetico. Sarebbero quindi necessari lunghi e dettagliati studi prospettici che seguano gli individui, reclutati nello studio, per diversi decenni precedenti all'esordio della malattia.

Tuttavia particolari infezioni nei primi anni di vita, l'esposizione ai virus Epstein Barr (EBV) e Herpes 6 (HHV6) umano, la carenza di vitamina D e l'esposizione al fumo, sono stati indicati più volte come possibili fattori di rischio (15,16).

Il campo della genetica e l'identificazione dei geni di malattia, potrebbero altresì, fornire un aiuto importante nell'indicare i fattori ambientali che svolgono un ruolo nel rischio di sviluppare la patologia. Infatti i geni di malattia, le proteine correlate ed i meccanismi d'azione, potrebbero fornire importanti indicazioni sui diversi fattori ambientali implicati nell'eziopatogenesi della SM, difficilmente evidenziabili con studi di tipo prospettico.

1.4 Modello patogenetico

Per molti anni le ricerche indirizzate a fornire una descrizione del modello patogenetico per la SM hanno sottolineato il ruolo dei linfociti T, cellule che partecipano alla regolazione e allo sviluppo della risposta immune cellulo-mediata, nei meccanismi che determinano i danni alle strutture del sistema nervoso tipici di questa malattia. Secondo questo modello si ha un'attivazione dei linfociti T pro-infiammatori (CD4+) nella periferia causata dal riconoscimento, da parte del recettore delle cellule T (T cell receptor, TCR), di antigeni presentati sul complesso maggiore di istocompatibilità di classe II (Major histocompatibility complex, MHC-II) dalle cellule presentanti l'antigene (Antigen Presenting Cell, APC). I linfociti T, migrano, aderiscono e penetrano la barriera ematoencefalica mediante meccanismi di adesione molecolare e con l'intervento di proteasi e citochine. All'interno del sistema nervoso centrale (SNC), i linfociti T vengono riattivati dalle molecole MHC-II sulle APC e iniziano a produrre citochine pro-infiammatorie, che promuovono lo stato infiammatorio nel SNC con conseguente attivazione di molecole effettrici come macrofagi, linfociti B e altri linfociti T. I macrofagi ed i linfociti T (CD8+) attaccano la guaina mielinica attraverso dei mediatori citotossici, tra cui soprattutto le specie radicaliche dell'ossigeno (O_2) e l'ossido nitrico (NO). I linfociti B si differenziano in plasmacellule che secernono anticorpi demielinizzanti.

Quest'ultimi attivano i macrofagi e la cascata del complemento che causano il danno mielinico (27).

Un nuovo filone di ricerca che ha assunto importanza crescente negli ultimi anni sta dimostrando che il ruolo dei linfociti B è cruciale almeno quanto quello dei linfociti T.

I linfociti B sono principalmente coinvolti nello sviluppo della risposta "umorale". Essi, infatti, una volta attivati si trasformano in plasmacellule che producono anticorpi, i quali liberati nel sangue e nei tessuti, inattivano e distruggono gli agenti potenzialmente dannosi.

Nel caso della SM, i linfociti B attivati rilasciano anticorpi che attaccano in maniera anomala i tessuti del proprio corpo (auto-anticorpi), ed in particolare la mielina. Questi anticorpi autoreattivi si possono dosare nel liquido cefalo-rachidiano dei malati di SM mediante una tecnica chiamata immunoelettroforesi che rileva la presenza di cosiddette bande oligoclonali di intensità direttamente proporzionale alla quantità di anticorpi prodotti. Questa metodica è utilizzata per definire il livello di attività e la prognosi della SM e secondo alcuni esperti, predice l'evoluzione da sindrome clinica isolata, (Clinically Isolated Syndrome CIS la sindrome è caratterizzata da un episodio neurologico che duri almeno 24 ore e che sia compatibile con una malattia demielinizzante del sistema nervoso centrale), a SM conclamata (42).

Si è rilevato inoltre che, sia i principali fattori di rischio genetico che alcuni fattori ambientali, sono in grado di influenzare la proliferazione e la funzione dei linfociti B. Per esempio, tra i geni non-MHC (vedi paragrafi successivi) che sono stati associati con alla suscettibilità alla SM, tra cui i geni *CD40*, *CXCR4*, e *CXCR5* appaiono particolarmente rilevanti per l'omeostasi e la funzione dei linfociti B.

Un'ulteriore dimostrazione del ruolo importante dei linfociti B nella patogenesi della SM, deriva dal fatto che l'unico anticorpo cellula-specifico altamente efficace contro la RRMS (relapsing remitting MS), rituximab, agisce sulle cellule B, eliminando specificamente quelle che esprimono il recettore CD20. Il CD20 è espresso in diversi stadi di sviluppo delle cellule B, da pre-B a naive e a cellule della memoria, ma è assente sia nelle cellule pro-B che nelle plasmacellule. Questo suggerisce che oltre alla produzione di auto-anticorpi, le cellule B potrebbero svolgere un ruolo centrale nella SM o attivando risposte T specifiche attraverso il loro ruolo di APC o promuovendo la produzione di citochine proinfiammatorie (28).

Definire il ruolo dei linfociti B, e delle loro sottopopolazioni, nel processo patogenetico della malattia assume quindi grande rilevanza in previsione della possibilità di individuare nuovi obiettivi farmacologici per il trattamento della SM.

1.5 Dissezione delle basi genetiche della Sclerosi Multipla

E' ormai accertato che fattori genetici giochino un ruolo importante nella predisposizione alla SM, contribuendo per circa il 10-20% alla probabilità di sviluppare la malattia.

In particolare, due osservazioni di carattere epidemiologico dimostrano il ruolo di fattori genetici nella SM:

- 1)** la constatazione dell'incremento dell'incidenza della malattia in ambito familiare, come è emerso dalle indagini sui gemelli e fratelli affetti;
- 2)** l'osservazione che nei vari gruppi etnici esistono delle differenze nella prevalenza della SM.

Due termini possono aiutare a inquadrare meglio il legame tra SM e genetica: suscettibilità e rischio relativo. Il termine suscettibilità appartiene alla genetica formale umana e viene usato in riferimento al comportamento additivo di diversi geni. La malattia si manifesterebbe quando la suscettibilità supera un determinato limite definito soglia.

Per rischio relativo si intende invece la probabilità di sviluppare la malattia in funzione della presenza di alcune varianti del DNA.

I primi studi genetici sono stati rivolti a verificare il coinvolgimento del sistema immunitario nella SM, ed in particolare l'associazione con il sistema di istocompatibilità HLA o con alcune molecole coinvolte nella risposta immune, quali, il recettore per le cellule T (TCR), il complemento e le immunoglobuline.

Mentre questi studi non hanno confermato in maniera conclusiva un ruolo nella predisposizione per il TCR, i geni del complemento, e quelli delle immunoglobuline, un ruolo per l'HLA è stato dimostrato inequivocabilmente.

Fin dai primi anni 70, infatti, ci si è concentrati sull'associazione tra SM e alleli del complesso maggiore di istocompatibilità (MHC)/antigeni leucocitari umani (HLA) localizzato nel braccio corto del cromosoma 6 (17).

L'MHC è costituito da diversi antigeni ovvero molecole appartenenti alla superfamiglia delle immunoglobuline, glicoproteine che si trovano sulla superficie di distinti tipi cellulari. Gli antigeni sono suddivisi in classe I e II. Quelli inclusi nella classe I sono definiti come HLA-A, HLA-B, HLA-C, mentre quelli che appartengono alla classe II sono designati con le sigle HLA-D, HLA-DR, HLA-DP, HLA-DQ. Gli antigeni di classe I sono distribuiti sulle cellule di ogni tessuto; viceversa quelli di classe II sono peculiari delle cellule immunocompetenti che presentano l'antigene, ovvero linfociti B, macrofagi e cellule dendritiche.

La funzione di tutti gli antigeni di classe I e II è quella di presentare l'antigene alle specifiche cellule immunocompetenti, rispettivamente i linfociti T CD8+ e CD4+.

Inizialmente le analisi genetiche hanno messo in evidenza l'associazione di alcuni aplotipi del locus HLA con la SM. In particolare, è stato identificato un aplotipo esteso predisponente di classe II, DRB1*1501-DQB1*0602 (anche definito sierologicamente, aplotipo DR2, DQw1), costantemente associato in diversi gruppi etnici (18). L'associazione con questo aplotipo è più forte nella popolazione del Nord Europa ma è stata confermata nella maggior parte delle popolazioni studiate, ad esclusione dei Sardi e di alcuni gruppi etnici nel Mediterraneo nei quali la SM si associa invece all'aplotipo DR4 (DRB1*0405-DQA1*0301-DQB1*0302) (19). Oltre a quest'ultimo, nella popolazione sarda, la SM è stata associata a diversi aplotipi dei loci DRB1 DQB1; quali il DRB1*0301-DQB1*0201 e DRB1*0405-DQB1*0301(20,21). In particolare sono stati descritti cinque aplotipi DRB1- DQB1 associati positivamente con la SM; DRB1*1303-DQB1*0301, DRB1*0405-DQB1*0301, DRB1*0301-DQB1*0201, DRB1*1501-DQB1*0602, DRB1*0405-DQB1*0302 mentre nessun aplotipo negativamente associato alla malattia in maniera significativa è stato osservato (22).

Per identificare ulteriori fattori genetici implicati nella SM, la ricerca si è focalizzata su geni non-HLA.

Gli approcci seguiti sono stati fondamentalmente due:

- 1) studi di linkage e di associazione su geni candidati scelti a priori sulla base di conoscenze sulla patogenesi (24);
- 2) studi di linkage sull'intero genoma alla ricerca di un legame con una regione cromosomica di interesse (25) e seguiti da studi di associazione per identificare il gene candidato di suscettibilità all'interno del picco di linkage (26).

I risultati ottenuti attraverso questi studi, che utilizzavano casistiche ridotte a poche centinaia di individui affetti, sono generalmente controversi e solo pochi sono stati riconfermati con i nuovi approcci di studio GWAS (vedi paragrafo successivo).

Tuttavia, essi sono stati utili ad evidenziare i limiti di questi disegni sperimentali per la ricerca di loci di predisposizione alla SM e l'esigenza di condurre studi su larga scala (sia in termini di numero di varianti che di individui studiati) con maggiore potere statistico.

Solo a partire dal 2006, con l'avvento degli studi di associazione su tutto il genoma (Genome-Wide Association Study, GWAS), si sono avuti notevoli avanzamenti nel campo della genetica delle malattie complesse in genere ed in particolare per la SM. Questo approccio è, infatti, risultato più idoneo all'identificazione di geni di suscettibilità in malattie genetiche

multifattoriali, secondo l'ipotesi del "common variant-common disease", per cui le malattie multifattoriali comuni sono causate da molteplici varianti comuni con effetto modesto.

1.6 Geni identificati tramite studi di associazione su tutto il genoma (GWAS)

Nel corso degli ultimi anni la conoscenza della variabilità genetica umana è migliorata notevolmente, grazie al completamento della sequenza del genoma umano, del progetto HapMap, e del progetto 1000 Genomi (1000 Genomes project, 1000G), con la descrizione di circa 38 milioni di polimorfismi a singolo nucleotide (SNP), 1,4 milioni di piccole inserzioni e delezioni (INDEL), ed oltre 14,000 varianti strutturali (Copy number variation, CNV).

Queste conoscenze sono scaturite dalla disponibilità di nuove tecnologie applicate agli studi genetici. La rapida evoluzione dei metodi di genotipizzazione implementate in piattaforme tecnologiche dalle maggiori compagnie del settore, ha permesso di condurre studi GWA (Genome Wide Association Study, GWAS), un approccio senza ipotesi a priori, che prevede l'analisi simultanea di centinaia di migliaia di varianti geniche in migliaia di individui a costi accessibili, attraverso l'utilizzo di gene chip-arrays.

L'applicabilità ed efficacia di questi gene chip-arrays è stata confermata a partire dal 2006 da un continuo susseguirsi di pubblicazioni su diverse riviste scientifiche internazionali, riguardo alla genetica di numerose malattie complesse e tratti quantitativi, fino a poco tempo fa scarsamente conosciuta.

Gli studi GWA si basano sul confronto delle frequenze alleliche o genotipiche in un campione di pazienti, definiti casi, rispetto ad un campione di individui sani, definiti controlli, non imparentati né tra loro né con i casi. Il test di associazione valuta quindi differenze statistiche nella frequenza di specifici alleli nei casi rispetto ai controlli.

Per convenzione per un singolo test statistico la soglia di significatività (nominale) è di $\leq 0,05$. Poiché gli studi GWA, prevedono una scansione dell'intero genoma ad alta densità, vengono condotti contemporaneamente molteplici test statistici, pertanto il livello di significatività deve tenere conto del numero di varianti testate. Considerando quindi una correzione per test multipli, generalmente vengono considerati significativi p-values sotto la soglia di 5×10^{-8} .

Di conseguenza, solo quando vengono utilizzate ampie casistiche (migliaia di casi e controlli) si raggiunge un potere statistico adeguato a mettere in evidenza eventuali differenze di frequenza tra casi e controlli. Tali casistiche dovranno essere tanto più ampie quanto più è bassa la frequenza dell'allele testato.

Il primo studio GWAS sulla SM nel 2007 ha permesso di mappare più precisamente il locus di rischio dell'MHC sull'HLA DRB*1501, e ha identificato i primi due loci di suscettibilità non-HLA in corrispondenza dei geni *IL2RA* and *IL7RA*, con p-values oltre la soglia di significatività genome-wide (29). Questi geni, la cui associazione è stata poi replicata in tutti gli studi successivi, sono di particolare interesse perché espressi sulle cellule T regolatorie, che sono state implicate nella regolazione della risposta immune nella SM (30).

Negli anni successivi diversi altri studi GWAS in diverse coorti di pazienti affetti da SM, hanno ulteriormente incrementato il numero di geni noti implicati nella suscettibilità e confermato loci precedentemente associati alla SM (31, 32, 33, 34, 35,36)

Tuttavia, è diventato presto evidente che questi studi GWAS iniziali avevano il limite di utilizzare un numero di campioni troppo basso, e di conseguenza il loro potere statistico non era sufficiente per l'identificazione di varianti con effetti modesti sul rischio malattia.

Così, nel 2011 è stato condotto il primo GWAS guidato dal consorzio internazionale per la SM (International Multiple Sclerosis Genetics Consortium, IMSGC), nel quale sono stati analizzati 9772 casi di origine Europea, collezionati da 23 gruppi di ricerca di 15 diversi Paesi.

Questo studio ha identificato 29 nuovi loci di predisposizione ed ha replicato la maggior parte di quelli precedentemente riportati (37), portando ad oltre 50 i geni di suscettibilità alla SM (Tab.1).

All'interno dell'MHC gli autori rifiniscono l'identità degli alleli di rischio del DRB1 (*DRB1*15:01*) e confermano che l'effetto protettivo indipendente attribuibile alla Classe I, si trova nel gene *HLA-A* (*HLA-A*02:01*). Da notare, che come atteso, tutti i loci identificati, ad eccezione dell'HLA, hanno un effetto genetico modesto (odds ratio OR 1,1-1,3) e la maggior parte delle varianti identificate sono comuni negli europei. Inoltre, quasi tutte si trovano all'interno o nelle vicinanze di geni che codificano per molecole del sistema immune, a supporto dell'ipotesi che la SM sia primariamente una malattia immuno-mediata.

In particolare, molti dei geni sono correlati alle funzioni dei linfociti, e giocano un ruolo nell'attivazione e proliferazione delle cellule T, in pathways di citochine, molecole co-stimolatorie e trasduzione del segnale.

Circa un terzo dei geni identificati è anche associato ad almeno un'altra malattia autoimmune, rafforzando l'idea che esistono dei meccanismi comuni alla base della maggior parte delle malattie autoimmuni.

Chr	Rs id	Position	Candidate Gene	Risk Allele	Odds Ratio	Risk Allele Freq Ctrls	Tags Autoimmune Disease SNP*
1	rs4648356	2699024	<i>MME11 (TNFRSF14)</i>	C	1.14	0.68	RA, CeD, UC
1	rs11810217	92920965	<i>EVIS</i>	A	1.15	0.25	
1	rs11581062	101180107	<i>VCAMI</i>	G	1.12	0.28	
1	rs1335532	116902480	<i>CD58</i>	A	1.22	0.87	
1	rs1323292	190807644	<i>RGS1</i>	A	1.12	0.82	CeD
1	rs7522462	199148218	<i>Clorf106 (KIF21B)</i>	G	1.11	0.73	UC, CeD, CrD, AS
2	rs12466022	43212565	<i>No gene</i>	C	1.11	0.73	
2	rs7595037	68500599	<i>PLEK</i>	A	1.11	0.56	CeD
2	rs17174870	112381672	<i>MERTK</i>	G	1.11	0.76	
2	rs10201872	230814968	<i>SP140</i>	A	1.14	0.16	CrD
3	rs11129295	27763784	<i>EOMES</i>	A	1.11	0.37	
3	rs669607	28046448	<i>No gene</i>	C	1.13	0.48	
3	rs2028597	107041527	<i>CBLB</i>	G	1.13	0.92	
3	rs2293370	120702624	<i>TMEM39A (CD80)</i>	G	1.13	0.82	
3	rs9282641	123279458	<i>CD86</i>	G	1.21	0.90	
3	rs2243123	161192345	<i>IL12A</i>	G	1.08	0.27	CeD, PBC
4	rs228614	103797685	<i>NFKB1 (MANBA)</i>	G	1.09	0.52	PBC
5	rs6897932	35910332	<i>IL7R</i>	G	1.11	0.73	T1D, PBC, UC
5	rs4613763	40428485	<i>PTGER4</i>	G	1.2	0.13	CrD, UC, AS
5	rs2546890	158692478	<i>IL12B</i>	A	1.11	0.52	Ps, CrD
6	rs12212193	91053490	<i>BACH2</i>	G	1.09	0.44	CeD, T1D, CrD, AITD, VIT
6	rs802734	128320491	<i>THEMIS</i>	A	1.1	0.70	CeD
6	rs11154801	135781048	<i>MYB (AHL1)</i>	A	1.13	0.36	
6	rs17066096	137494601	<i>IL22RA2</i>	G	1.14	0.24	
6	rs13192841	138008907	<i>No gene</i>	A	1.1	0.27	RA
6	rs1738074	159385965	<i>TAGAP</i>	G	1.13	0.58	CeD, CrD
7	rs354033	148920397	<i>ZNF746</i>	G	1.11	0.74	
8	rs1520333	79563593	<i>IL7</i>	G	1.1	0.26	
8	rs4410871	128884211	<i>MYC</i>	G	1.11	0.72	
8	rs2019960	129261453	<i>PVT1</i>	G	1.12	0.21	
10	rs3118470	6141719	<i>IL2RA</i>	G	1.12	0.33	RA, VIT
10	rs1250550	80730323	<i>ZMIZ1</i>	A	1.1	0.34	CeD, IBD
10	rs7923837	94471897	<i>HHEX</i>	G	1.1	0.62	
11	rs650258	60588858	<i>CD6</i>	G	1.12	0.64	
11	rs630923	118259563	<i>CXCR5</i>	C	1.12	0.84	PBC
12	rs1800693	6310270	<i>TNFRSF1A</i>	G	1.12	0.41	PBC
12	rs10466829	9767358	<i>CLECL1</i>	A	1.09	0.51	T1D
12	rs12368653	56419523	<i>CYP27B1</i>	A	1.1	0.48	RA
12	rs949143	122161116	<i>ARL6IP4</i>	G	1.08	0.30	
14	rs4902647	68323944	<i>ZFP36L1</i>	G	1.11	0.53	CeD, T1D, CrD
14	rs2300603	75075310	<i>BATF</i>	A	1.11	0.75	
14	rs2119704	87557442	<i>GALC (GPR65)</i>	C	1.22	0.92	
16	rs2744148	1013553	<i>SOX8</i>	G	1.12	0.17	
16	rs7200786	11085302	<i>CLECL1A (CHTA)</i>	A	1.15	0.46	T1D, PBC
16	rs13333054	84568534	<i>IRF8</i>	A	1.11	0.22	RA
17	rs9891119	37761506	<i>STAT3</i>	C	1.11	0.35	CrD
17	rs180515	55379057	<i>RPS6KB1</i>	G	1.09	0.34	
18	rs7238078	54535172	<i>MALTI</i>	A	1.12	0.76	
19	rs1077667	6619972	<i>TNFSF14</i>	G	1.16	0.79	
19	rs8112449	10381064	<i>TYK2 (ICAM3)</i>	G	1.08	0.67	T1D, Ps
19	rs874628	18165700	<i>MPV17L2 (IL12RB1)</i>	A	1.11	0.72	
19	rs2303759	54560863	<i>DKKL1 (CD37)</i>	C	1.11	0.26	
20	rs2425752	44135527	<i>CD40</i>	A	1.11	0.27	RA
20	rs2248359	52224925	<i>CYP24A1</i>	G	1.12	0.60	
20	rs6062314	61880157	<i>TNFRSF6B</i>	A	1.16	0.92	CrD, UC
22	rs2283792	20461125	<i>MAPK1</i>	C	1.1	0.51	
22	rs140522	49318132	<i>SCO2</i>	A	1.1	0.33	

Tab. 1: Varianti di rischio per la SM identificate dallo studio GWAS dell'IMSGC (Nature, 2011)

Infine, in uno studio successivo sempre guidato dall'IMSGC e pubblicato di recente (38), sono state identificate altre 48 varianti di suscettibilità alla SM, portando a 110 i loci noti a tutt'oggi. Di questi, 103 mappano fuori dall'HLA. In totale queste 110 varianti spiegano il 25% circa del rischio di ricorrenza familiare. Questo studio ha analizzato, in un disegno sperimentale a due stadi, oltre 80.000 individui di origine europea (~50.000 casi e ~30.000 controlli) per oltre 160.000 varianti presenti nell'ImmunoChip Illumina. Queste varianti sono state selezionate da un consorzio internazionale in base alla loro precedente associazione con malattie autoimmuni e infiammatorie per poter condurre il mappaggio fine di loci coinvolti in malattie e tratti immuno-relati.

Questo studio ha mostrato che circa il 22% dei segnali identificati coincide con almeno un segnale trovato associato con un'altra malattia autoimmune. Inoltre, ha anche rafforzato il coinvolgimento di NF- κ B nella patogenesi della SM. Questi dati espandono le conoscenze dei processi del sistema immunitario implicati nella SM, e possono in futuro promuovere lo sviluppo di terapie mirate e strategie di prevenzione.

1.7 Studio GWAS nella popolazione sarda

La popolazione sarda ha da sempre suscitato un notevole interesse in genetica umana sia per la peculiare distribuzione di diverse varianti genetiche, sia per le numerose malattie su base genetica particolarmente frequenti nell'isola. Vi sono numerose evidenze che la popolazione sarda, pur collocandosi nell'ambito della variabilità europea, manifesti tutta una serie di caratteristiche di unicità.

Queste caratteristiche peculiari si spiegano attraverso un isolamento plurimillenario rispetto ad altre popolazioni. Alcune varianti genetiche sono particolarmente frequenti in Sardegna e talvolta rare o assenti in altre popolazioni. Sono varianti antiche già presenti in quegli individui che diverse migliaia di anni fa hanno popolato l'isola (effetto fondatore). Per esempio, vi sono alleli e aplotipi dell'HLA piuttosto rari o assenti altrove, come DRB1*1501-DQB1*0602 e DRB1*1601-DQB1*0502 e altri molto comuni fuori dalla Sardegna ma molto rari nell'isola, come l'aplotipo esteso DRB1*0301-DQA1*0501-DQB1*0201-B18-A30 (40).

Tutto ciò rende questa popolazione ideale sia per estensivi studi di catalogazione della variabilità umana, che per la dissezione delle basi genetiche di malattie complesse come appunto la SM. Infatti, alcune delle varianti fondatrici sarde potrebbero spiegare una porzione del rischio di SM non ancora definito.

Pertanto con l'obiettivo di identificare nuovi geni implicati nella suscettibilità alla SM, il gruppo di ricerca presso cui ho svolto il dottorato di ricerca ha condotto uno studio GWA caso-controllo nella popolazione sarda, pubblicato nel 2010 su *Nature Genetics* che ha portato all'identificazione di varianti nel gene *CBL-B* associate alla SM (33).

Questo lavoro è stato articolato in due fasi principali. Nella fase preliminare sono stati analizzati per test di associazione circa 6.6 milioni di SNPs (direttamente genotipizzati con Genechip arrays Affymetrix 6.0 o imputati con pannello di riferimento HapMap II CEU, TSI e 1000 Genomes) in 882 casi di SM e 872 controlli sardi e di origine sarda.

L'inferenza statistica o imputazione permette di incrementare lo spettro delle varianti da testare con metodi che si avvalgono di algoritmi capaci di ricostruire i genotipi delle varianti non direttamente genotipizzate in tutti i campioni dello studio, basandosi sulla conoscenza e disponibilità di pannelli di riferimento.

Questo GWAS ha portato all'identificazione di diverse varianti significativamente associate nel locus *HLA*, ed in particolare ha confermato l'associazione nei sardi dell'aplotipo dell'*HLA* DRB1*0301-DQB1*0201.

Anche loci non *HLA*, precedentemente descritti (*CD58*, *IL2RA*, *EVI5*, *CD40*, *IRF8*, *CLEC16A*) sono stati confermati seppure con *p-values* sotto la soglia genome-wide ($p=5 \times 10^{-8}$).

Nella seconda fase, sono stati selezionati i 9 SNPs più significativamente associati da validati con un metodo indipendente (TaqMan®) in ulteriori 1.775 casi di SM and 2.005 controlli indipendenti provenienti sempre dalla popolazione sarda. Di questi, solo uno SNP l'rs9657904 (T>C) è stato confermato con un *p-value*= 9.34×10^{-6} .

Infine, l'analisi combinata dei due data-sets del GWAS e della replica indipendente ha poi ulteriormente rafforzato l'evidenza di associazione per lo SNP rs9657904, ottenendo un *p-value* oltre la soglia GWAS (*p-value*= 1.60×10^{-10}).

L'allele comune T dello SNP rs9657904, localizzato nell'introne 1 del gene *CBL-B*, è quello che conferisce suscettibilità alla SM (OR = 1.40, 95%CI = 1.27 - 1.57 per l'allele T, e OR = 2.5, 95% CI = 1.71 - 3.52, per il genotipo omozigote TT quando confrontati con il rischio conferito dal genotipo CC).

L'evidenza dell'associazione di questa variante sul gene *CBL-B* con la SM è stata successivamente replicata nella popolazione italiana geneticamente distante dalla popolazione sarda (39). In questo studio l'rs9657904 è stato analizzato in 1435 casi e 1466 controlli provenienti dalla penisola Italiana, confermando l'associazione dell'allele di predisposizione T alla SM con un *p-value*= 7.35×10^{-5} (OR=1.31, 95% CI 1.14 to 1.52). Recentemente l'associazione

è stata anche replicata dall'IMSGC (37,38), confermando il ruolo di questa variante nella predisposizione alla SM anche nelle altre popolazioni.

1.8 Il gene CBL-B

Il gene *CBL-B* (casitas b-lineage lymphoma), localizzato sul braccio lungo del cromosoma 3, è costituito da 21 esoni che danno origine a 11 diverse isoforme.

CBL-B codifica per una E3 ubiquitina ligasi di tipo RING, implicata in una varietà di funzioni. Come ubiquitina ligasi, *CBL-B* accetta l'ubiquitina da specifici enzimi coniuganti ubiquitina, E2, e la trasferisce a diversi substrati, promuovendo la loro degradazione mediata dal proteasoma. La funzione di maggiore importanza è il ruolo che svolge nel sistema immune. *CBL-B* rappresenta, infatti, un regolatore negativo dell'attivazione dei recettori delle cellule T (TCR) e B (BCR). La sua assenza nel topo knock-out risulta in un fenotipo di iperattivazione di cellule T e B e nello sviluppo di autoimmunità (43, 44). Inoltre, *CBL-B* agisce da regolatore negativo dell'attivazione dei linfociti T anche, promuovendo e mantenendo le cellule T in uno stato anergico, regolando così la tolleranza periferica. Altri studi hanno mostrato che topi knock-out per *Cbl-b* sono resistenti a vari tipi di cancro (probabilmente a causa del fatto che il sistema immunitario è iperattivato quindi ha una maggiore capacità di intercettare le cellule tumorali). Complessivamente si può quindi affermare che variazioni nell'attività di *CBL-B* hanno un ruolo critico nel mantenere l'equilibrio tra attivazione immunologica e tolleranza, con un impatto su autoimmunità e cancro.

In quale modo *CBL-B* contribuisca al rischio di SM non è noto, ma è plausibile ipotizzare che una sua ridotta attività possa portare ad un'iperattivazione della risposta immune adattiva mediata da cellule T e B dove *CBL-B* esercita la sua funzione che porta poi alla comparsa della patologia.

Chiarire il meccanismo attraverso cui *CBL-B* contribuisce alla predisposizione alla SM può sicuramente essere utile a comprendere l'eziologia di questa malattia e a sviluppare in futuro delle nuove strategie preventive e farmacologiche.

CBL-B appartiene alla famiglia di proteine CBL, altamente conservata nel corso dell'evoluzione dai nematodi fino all'uomo. Mentre negli organismi inferiori esiste un singolo gene *CBL*, in quelli superiori sono presenti, codificati nel loro genoma, almeno due geni *CBL*.

Nei mammiferi sono presenti tre differenti geni *CBL*: *C-CBL*, *CBL-B* e *CBL-3*. Questi geni sono codificati su cromosomi diversi e mostrano un'espressione tessuta specifica. Con l'eccezione di

Cbl-3, la cui funzione fisiologica negli epitelii è ancora sconosciuta, *C-CBL* ha un'espressione relativamente più elevata nel timo, mentre *CBL-B* è prevalentemente espressa negli organi linfoidi periferici, anche se la sua espressione è ubiquitaria ed è stata osservata nel polmone, rene, milza, testicoli, cervello, fegato e cellule ematopoietiche.

Le proteine della famiglia CBL hanno una struttura in domini altamente conservata. I vari domini sono implicati in interazioni proteina-proteina.

La presenza di diversi tipi di domini di legame permette alle proteine CBL di partecipare a molteplici vie di trasmissione del segnale.

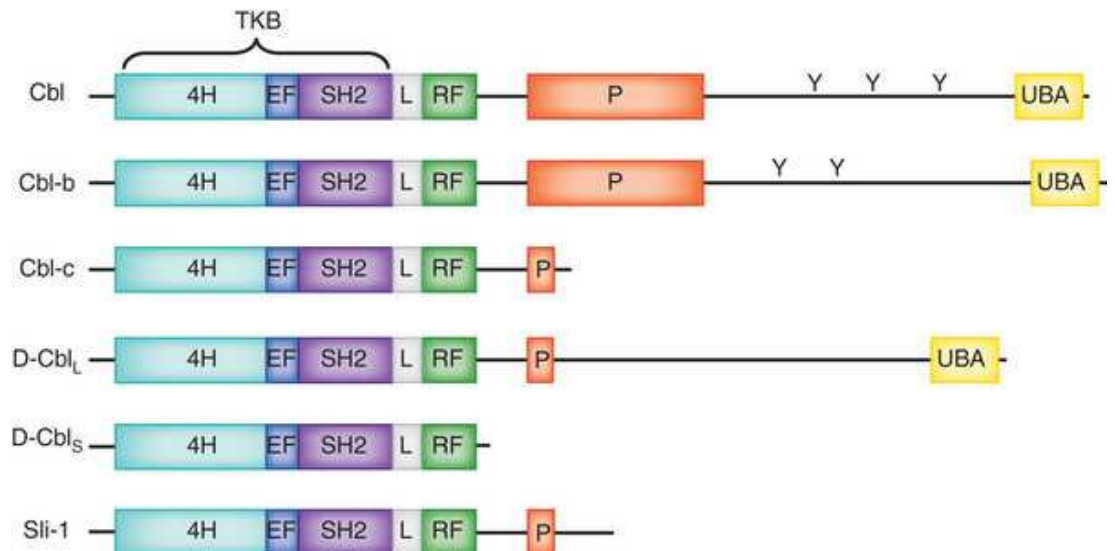
Tutte le proteine della famiglia CBL contengono un dominio catalitico, il dominio RING finger (RF), responsabile dell'attività E3-ligasi di queste proteine. Funzionalmente il dominio RF è implicato nel trasferimento dell'ubiquitina (Ub) tra l'enzima E2 e il substrato bersaglio.

La conservazione del dominio RF e dei diversi domini di interazione proteina-proteina suggerisce che questi siano cruciali affinché le proteine CBL possano esercitare la loro funzione di adattatori polivalenti e di E3-ligasi.

In particolare, la proteina CBL-B è caratterizzata, come tutte le proteine della famiglia *CBL*, da quattro principali domini (Fig.4), tra cui:

- 1) il dominio TKB (N-terminal tyrosine kinase binding domain) costituito da 3 differenti sottodomini: il four-helix bundle (4H), il calcium-binding EF hand e il dominio SH2;
- 2) il dominio RING- finger che come già ricordato possiede attività di E3-ubiquitin ligasi;
- 3) una regione ricca di prolina (P) coinvolta nell'interazione con proteine citosoliche ed il ruolo di adattatore della proteina;
- 4) il dominio UBA che interagisce con le proteine poli-ubiquitinate e domini ricchi di prolina.

L'isoforma principale codifica per una proteina di 982-aminoacidi. Le isoforme minori sono caratterizzate dall'assenza di uno o diversi motivi.



Nature Structural & Molecular Biology 19, 131–133 (2012)

Fig. 4: Famiglia delle proteine CBL codificate da differenti geni: Cbl (anche noto come c-Cbl, CBL2 or RNF55), Cbl-b (RNF56) and Cbl-c (Cbl-3, Cbl-SL or RNF57). D-Cbl_L and D-Cbl_S. Sli-1 è la proteina del *Caenorhabditis elegans*. Tutte le proteine Cbl contengono un dominio N terminale (TKB), una regione linker (L) e un C3HC4 RING finger domain (RF). La parte C terminale delle proteine CBLB presentano una tirosina (Y) che può andare incontro a fosforilazione. Il dominio C terminale contiene inoltre una regione ricca in prolina (P). Le forme più estese delle proteine Cbl presentano un dominio UBA (ubiquitin associated domain) all'estremità C terminale. *Nature Structural & Molecular Biology* 19, 131–133 (2012).

Diversi studi condotti su topi knock-out per il gene *Cbl-b* (*Cbl-b*^{-/-}) hanno consentito di caratterizzarne la funzione, e di chiarire il suo ruolo nella regolazione della risposta immune. Esistono due modelli indipendenti di topi knock-out per *Cbl-b*, che sono stati ampiamente studiati negli ultimi 10 anni. Entrambi i topi *Cbl-b*^{-/-} risultano avere un fenotipo apparentemente normale e sono fertili. Per quanto riguarda le cellule T e B, i topi *Cbl-b*^{-/-} non hanno difetti evidenti nello sviluppo, nella selezione timica e nel numero di recettori B e T di superficie. Tuttavia, la stimolazione *in vitro* di linfociti periferici isolati da topi *Cbl-b*^{-/-} ha rivelato diverse alterazioni nella attivazione di queste cellule.

Per quanto riguarda i linfociti T periferici *Cbl-b*^{-/-}, si è visto che non hanno bisogno di co-stimolazione del CD28 per essere pienamente attivati, e possono proliferare e secernere elevati livelli di interleuchina-2 (IL-2) anche in assenza di co-stimolazione. Inoltre, la mancata espressione di *Cbl-b* rende le cellule T resistenti alla stimolazione con ionomicina e TGF-β-mediata.

Dal punto di vista molecolare, l'attivazione dei linfociti T dipende da una serie di cascate di fosforilazione a valle della stimolazione del recettore delle cellule T (TCR) (Fig. 5). In particolare, in seguito allo stimolo si ha la fosforilazione dei residui tirosinici degli ITAMS (Immune-receptor-Tyrosine-based-Activation-Motif) da parte di tirosin-chinasi della famiglia Src, come Lck. Gli ITAMs fosforilati diventano ligandi ad alta affinità per proteine come la tirosin-chinasi della famiglia Syk, ZAP-70. L'attivazione di ZAP-70 la rende in grado di fosforilare LAT, proteina transmembrana adattatrice che media interazioni tra proteine e il loro reclutamento alla membrana. Questa, una volta fosforilata, forma un complesso con SLP-76 e VAV.

Infine, l'attivazione mediata da VAV permette la formazione di un complesso a valle, che determina cambiamenti nel citoscheletro, principalmente nella formazione di SMAC (supra-molecular activation cluster-cluster di molecole attivate), essenziali per la formazione della sinapsi immunologica, la produzione di interleuchina 2 (IL-2) e la proliferazione cellulare. Le chinasi Src attivano anche Itk chinasi, direttamente coinvolte nella fosforilazione e attivazione della fosfolipasi (PLC γ) che attivano i canali del calcio.

In assenza di *Cbl-b* è stato osservato che Vav1 è iperfosforilata anche in assenza di stimolazione del TCR e questo porta alla formazione del complesso SMAC (Supra Molecular Activation Cluster) ed alla attivazione costitutiva della sinapsi immunologica. Oltre a regolare negativamente VAV1, CBL-B mantiene i linfociti T in uno stato inattivo ubiquitinando numerose proteine a valle, come PLC γ e PKC θ e la subunità p85, con successiva degradazione lisosomiale. Questo porta all'interruzione del segnale con mancata attivazione dei fattori di trascrizione e produzione di citochine pro-infiammatorie a valle (Fig.5).

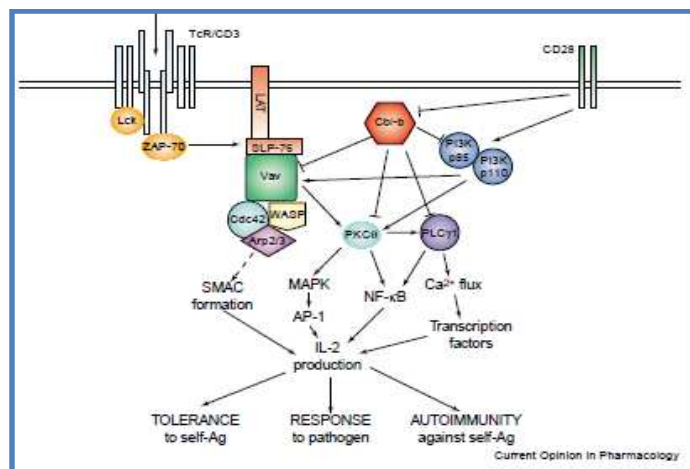
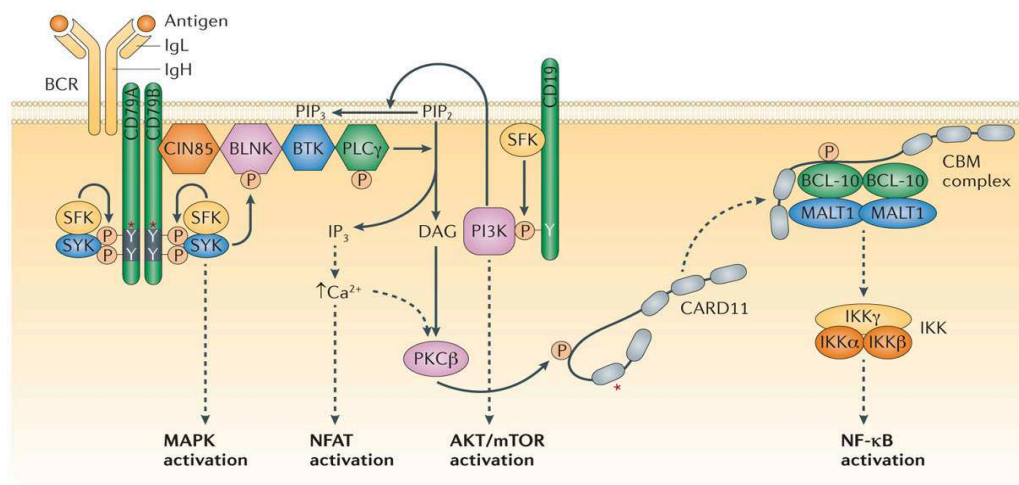


Fig 5: Pathway di attivazione del TCR (Curr Opin Pharmacol.2004 Aug;4(4):415-22.)

Per quanto riguarda i linfociti B, la funzione di CBL-B in queste cellule rimane in parte sconosciuta e non è chiaro come la sua assenza determini iperattivazione e proliferazione di queste cellule.

Le cellule B possono essere attivate secondo meccanismi T-indipendenti e T-dipendenti. L'attivazione T-indipendente avviene attraverso il legame dell'antigene ai recettori specifici delle cellule B, ovvero delle molecole IgM e IgD espresse sulla superficie dei linfociti B naive. Queste sono anche richieste per l'espressione superficiale delle molecole Ig e insieme ad esse formano il complesso recettoriale delle cellule B (complesso BCR). IgM e IgD sono dunque analoghe a CD3 e CD28 per i linfociti T. I domini citoplasmatici di queste due molecole contengono i motivi ITAM, e sono blandemente associati a tirosin chinasi della famiglia Src. Il cross-linking delle Ig di membrana porta le chinasi citoplasmatiche ad avvicinarsi e questo le attiva facendo loro fosforilare i domini ITAM. La fosforilazione di ITAM fornisce un sito di attacco per i domini SH2 della tirosin chinasi Syk, l'equivalente nelle cellule B di ZAP-70 dei linfociti T. Syk attivata va a fosforilare dei residui di tirosina su una proteina adattatrice detta SLP-65 facilitando il reclutamento su questa di altri domini SH2 di vari enzimi.



Nature Reviews | Drug Discovery

Fig. 6: Pathway di attivazione del BCR

L'attivazione T-dipendente dei linfociti B origina dall'interazione del recettore CD40 con il CD40 ligand (CD40L) espresso sulla superficie delle cellule T CD4+. L'interazione tra CD40L e il CD40 delle cellule B porta all'attivazione della proliferazione e differenziazione cellulare, nonché allo sviluppo di centri germinali e alla formazione di cellule B della memoria (46). Nei topi deleti per

Cbl-b è stata osservata una più alta proliferazione cellulare B in seguito a stimolazione con il CD40 (43). I partner Lyn e Syk sono indispensabili affinché *Cbl-b* possa fosforilare le tirosine in seguito a stimolazione del BCR (47). Nelle cellule B *Cbl-b* è descritto come regolatore negativo del segnale a partire dal BCR. Infatti, in cellule B di topi *Cbl-b*^{-/-} è stata osservata una sostenuta fosforilazione di Igα, Syk, e fosfolipasi C (plc-γ) e una prolungata mobilitazione del Ca⁺². Questo risulta in una iperproliferazione e nella produzione aumentata di anticorpi (Fig.6).

Per comprendere meglio il ruolo di CBL-B nella regolazione della risposta immune e nello sviluppo di autoimmunità, sono state studiate anche le cellule T regolatorie (T reg) di topi *Cbl-b*^{-/-}.

Le cellule T reg sono una sottopopolazione specializzata di linfociti T capaci di sopprimere l'attivazione del sistema immunitario, regolandone l'omeostasi e la tolleranza verso auto-antigeni. Modelli murini sperimentali dimostrano che l'azione immunodeprimente di queste cellule potrebbe essere utilizzata per trattare malattie autoimmuni, facilitare la tolleranza al trapianto o essere selettivamente eliminate per potenziare l'immunoterapia nei tumori. Esistono diverse sottoclassi di cellule T reg, che si differenziano per l'espressione di diversi recettori di membrana; alcune esprimono il CD8, altre CD4 CD25 e Foxp3 (fattore di trascrizione fondamentale per la differenziazione a Treg), e altre si differenziano in diversi tipi come ad esempio quelle con funzione soppressoria come TR1, Th3, CD8+CD28. In particolare, è stato visto che *Cbl-b* sembra avere un ruolo cruciale nello sviluppo delle cellule T regolatorie inducibili nella periferia e nell'espressione del fattore di trascrizione Foxp3, il quale è essenziale per lo sviluppo e la funzione delle cellule T reg (45).

La regolazione delle risposte immunitarie e l'autoimmunità dipendono da una serie di meccanismi di checkpoint che sono impiegati dall'organismo. Questi includono l'eliminazione di cellule T autoreattive nel timo (selezione negativa), il controllo dell'attivazione delle cellule T periferiche da parte delle cellule T regolatorie e l'induzione di anergia (48). Le cellule T autoreattive che sfuggono alla selezione negativa nel timo subiscono espansione clonale in seguito ad interazione con auto-antigeni nella periferia, ma questo normalmente li rende inattivi o non reattivi ad ulteriori stimoli, impedendo così autoimmunità. L'anergia è il meccanismo di immunotolleranza dove la cellula T è mantenuta in uno stato in cui non è in grado di rispondere in seguito allo stimolo iniziale antigene.

Visto il ruolo così importante di CBL-B nel regolare l'attivazione sia delle cellule T che B, è abbastanza evidente come una sua disregolazione possa portare allo sviluppo di autoimmunità. Questo è quanto si osserva, infatti, nei topi *Cbl-b*^{-/-}, che sono più suscettibili,

rispetto ai topi wild type (WT), allo sviluppo di encefalomielite autoimmune (EAE), forma sperimentale della SM, suggerendo che le alterazioni del segnale di trasduzione, modulate da *CBLB*, possono contribuire allo sviluppo di patologie autoimmune umane, come la SM.

Il possibile coinvolgimento di CBL-B nella SM è stato valutato misurandone i livelli di espressione dell'RNA messaggero (mRNA) e della proteina in linfociti T isolati dal sangue periferico di pazienti affetti (in fase progressiva e remittente) e da controlli (49). Questo studio ha messo in evidenza una diminuzione dell'espressione dell'mRNA e della proteina CBL-B nei pazienti. Inoltre, la riduzione di espressione della proteina, ma non dei livelli di mRNA, sono risultati più marcati nei campioni dei pazienti in fase progressiva rispetto a quelli in fase remittente. Ancora, minori livelli di mRNA correlano nei pazienti con una maggiore frequenza di ricadute cliniche. Complessivamente questi dati suggeriscono una possibile correlazione tra livelli di CBL-B in linfociti T periferici ed attacchi clinici nel corso della SM.

2 SCOPO DELLA RICERCA

Il progetto di ricerca, oggetto di questa tesi, è volto alla caratterizzazione genetica e funzionale delle varianti del gene *CBL-B* associate a SM, per chiarire il loro ruolo nella regolazione genica e quindi il meccanismo molecolare attraverso cui *CBL-B* contribuisce al rischio di sviluppare SM. Il primo passo in questa direzione è l'identificazione della variante causativa, ovvero la variante genetica che spiega l'associazione osservata in un determinato locus e che influenza il rischio di sviluppare la malattia.

Lo studio GWA iniziale aveva messo in evidenza che lo SNP rs9657904 nell'introne 1 di *CBL-B* era quello maggiormente associato e spiegava interamente l'associazione a questo locus (33). Tuttavia, altre varianti nella regione mostravano un grado di associazione simile ed erano in forte LD con il top SNP. Dal punto di vista statistico, quindi, queste varianti risultavano indistinguibili dall'rs9657904, ed era pertanto difficile definire su queste basi quale fosse quella causativa.

Inoltre, nello studio Sanna et. al., l'analisi di associazione era stata portata avanti utilizzando principalmente varianti comuni derivate dalla popolazione generale (vedi paragrafo 1.7), che non contengono varianti popolazione specifiche e/o varianti rare, e quindi con un'estrazione incompleta delle informazioni genetiche. In questi casi, spesso le varianti associate non sono quelle causative ma marcatori in LD con esse.

Per ovviare a questo limite ed avere una migliore estrazione della variabilità genetica, abbiamo condotto un nuovo GWAS utilizzando una mappa genetica ad alta densità che include varianti derivate dal sequenziamento di individui sardi e da diversi pannelli di referenza, e che ci ha consentito di ottenere un mappaggio fine del locus *CBL-B* (vedi disegno sperimentale).

Questa strategia ci ha permesso di rifinire l'associazione genetica, identificando poche varianti potenzialmente causative, e permettendoci quindi di condurre i successivi studi funzionali in maniera mirata. In particolare, gli studi funzionali descritti in questa tesi sono stati rivolti ad analizzare l'effetto dello SNP maggiormente associato, rs9657904, (e della variante INDEL ad esso correlata) sull'espressione di *CBL-B* e sull'attivazione dei linfociti T e B a valle (vedi disegno sperimentale). La disponibilità di una coorte di individui sani caratterizzati da un punto di vista genetico e fenotipico ci ha consentito di condurre gli esperimenti funzionali in cellule primarie.

L'identificazione e caratterizzazione delle varianti di rischio rappresenta un primo importante traguardo nella battaglia contro la SM. E' plausibile che una migliore comprensione dei fattori

e dei meccanismi che intervengono nell'attacco da parte del nostro sistema immune al sistema nervoso centrale, consentirà di sviluppare trattamenti più efficaci e anche per la prevenzione della malattia.

3 DISEGNO SPERIMENTALE DELLO STUDIO

Le attività di ricerca oggetto di questa tesi sono state finalizzate ad effettuare:

- 1) Studio caso-controllo GWA con una mappa genetica ad alta risoluzione, che ha consentito di effettuare un mappaggio fine del locus *CBL-B*;
- 2) Studi di espressione del gene *CBL-B* in cellule primarie T e B di individui selezionati in base al genotipo della variante genica maggiormente associata alla SM;
- 3) Studi per valutare il ruolo di *CBL-B* nell'attivazione di cellule primarie T e B di individui selezionati in base al genotipo della variante genica maggiormente associata alla SM.

3.1 Studio GWAS caso-controllo

Per condurre in un solo step un mappaggio fine del locus *CBL-B* ed identificare nuove varianti associate alla SM, abbiamo esteso gli studi genetici iniziali e portato avanti un nuovo studio GWA.

In particolare, per questo progetto abbiamo incrementato il numero dei campioni fino a 2000 casi e 2000 controlli tutti genotipizzati con la tecnologia Affymetrix (Genechip arrays 6.0). Per incrementare lo spettro delle varianti da testare per associazione, abbiamo successivamente inferito statisticamente (o imputato) i genotipi di varianti non direttamente genotipizzate (vedi metodi) utilizzando tre differenti pannelli di referenza:

- 1) un pannello costituito da 2120 sequenze di individui sardi che ha permesso di testare oltre 17 milioni di SNP;
- 2) uno costituito dalle sequenze di 382 individui Europei (1000 Genomes EUR) e che include oltre 30 milioni di SNP e 1,4 milioni di INDELS;
- 3) infine un pannello che si basa sulle sequenze di 1092 campioni di individui provenienti da tutto il mondo (1000 Genomes ALL) che comprende come il precedente oltre 30 milioni di varianti e 1,4 milioni di INDELS.

I risultati di tali analisi sono illustrati nei successivi paragrafi.

Il nostro gruppo ha in corso, infatti, un progetto che si propone di sequenziare l'intero genoma di 3.500 individui sardi a bassa copertura (in media 4x). Fino ad oggi, 2.120 sequenze sono state completate e utilizzate per costruire un pannello di riferimento "sardo" che parallelamente ai pannelli di referenza messi a disposizione dal progetto 1000 Genomi, viene usato nei test di associazione per l'imputazione dei genotipi in tutti i campioni analizzati con i

Genechip arrays. La strategia di analizzare i dati di genotipizzazione e di sequenza combinati permette di esaminare il genoma completo di un numero elevato di individui ottimizzando notevolmente i costi.

3.2 Studi funzionali per chiarire il ruolo di varianti di CBL-B nella SM

Gli studi funzionali condotti al fine di chiarire il contributo delle varianti associate nel gene *CBL-B* al rischio di SM, sono stati effettuati su cellule primarie T e B di individui selezionati per specifici genotipi della variante top (rs9657904; TT predisponente, CC protettivo). Come di seguito illustrato, gli esperimenti sono stati programmati per valutare variazioni genotipo-specifiche di 1) espressione di *CBL-B* e delle sue isoforme e 2) pathways di attivazione di linfociti T e B dove è coinvolto CBL-B.

In tutti gli esperimenti i linfociti T e B sono stati isolati da sacche di buffy coat di individui sani di genotipo noto arruolati nello studio in qualità di controlli provenienti dal centro trasfusionale Brotzu di Cagliari. L'utilizzo di cellule provenienti da soggetti sani, consente di valutare l'effetto della variante in esame in assenza di effetti confondenti, presenti invece nei soggetti affetti e difficili da controllare, correlati alla malattia stessa (assunzione di farmaci, eterogeneità del fenotipo clinico ecc).

3.2.1 Studio dell'espressione del gene CBL-B in linfociti T e B

L'espressione del gene *CBL-B* è stata valutata in linfociti T e B estratti da individui con genotipo noto per le varianti in esame, sia allo stato basale che stimolati con diversi induttori specifici.

Come già riportato, per il gene *CBL-B* sono state descritte, nei database pubblici, 11 isoforme derivanti da splicing-alternativi del trascritto principale (Fig. 7) di cui solo 10 sono predette codificare per una proteina.

Per valutare se tutte le 10 isoforme, potenzialmente codificanti, fossero espresse nelle cellule selezionate e se qualcuna di esse mostrasse un pattern di espressione differenziale genotipo specifico abbiamo disegnato coppie di oligonucleotidi (oligonucleotidi) da utilizzare in esperimenti di qRT-PCR (real time quantitative reverse transcription-PCR) che discriminavano 9 isoforme su 10. Purtroppo per una delle isoforme non è stato possibile disegnare degli oligonucleotidi specifici (vedi Fig. 7, tabella 2 e materiali e metodi). Questa metodica permette la precisa quantificazione dell'espressione genica dopo retrotrascrizione dell'RNA. Uno

screening iniziale condotto mediante chimica SyberGreen ha evidenziato l'espressione di 8 delle 10 isoforme selezionate. Abbiamo quindi valutato se ci fosse un'espressione differenziale degli 8 trascritti espressi in linfociti T e B di individui con genotipo omozigote per l'allele T (predisponente) e C (protettivo) del top SNP rs9657904. Le isoforme che hanno mostrato un'espressione differenziale significativa sono state quindi validate mediante saggi TaqMan® (vedi risultati).

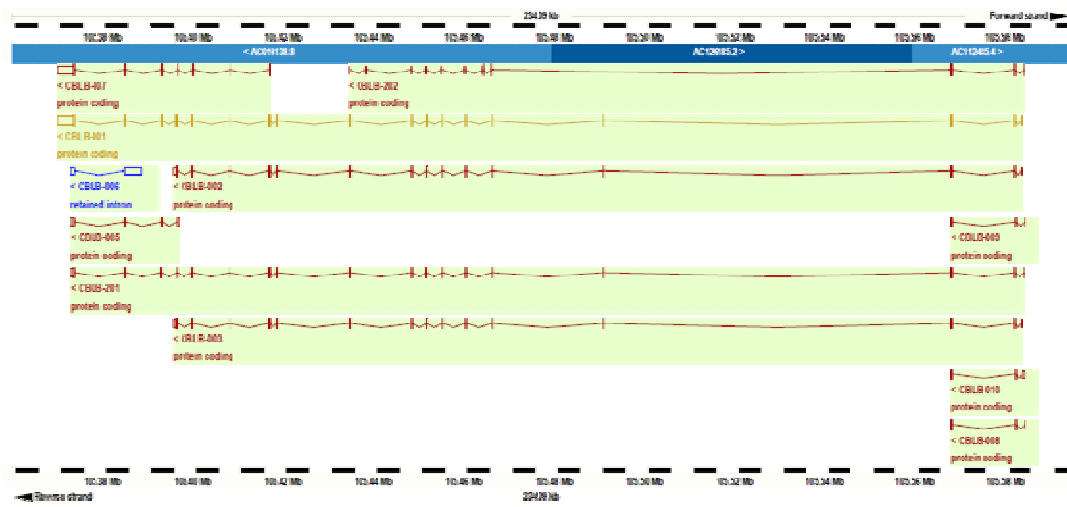


Fig. 7: Illustrazione grafica delle 11 isoforme descritte di CBLB nel database Ensemble

3.2.2 Studio dei pathway di attivazione dei linfociti T e B

Visto il ruolo di CBL-B nella regolazione dell'attivazione dei linfociti T e B, abbiamo valutato l'effetto genotipo-specifico a valle delle varianti trovate associate a SM, attraverso esperimenti ELISA. Nello specifico, per valutare lo stato di attivazione dei linfociti T, abbiamo misurato la produzione di IL-2, una citochina prodotta specificamente dai linfociti T in seguito a stimolazione del TCR.

Per quanto riguarda le cellule B, abbiamo misurato la traslocazione nucleare di NF- κ B e quindi l'attivazione di questo fattore di trascrizione, che avviene in seguito ad attivazione sia T-dipendente che T-indipendente di queste cellule.

4 MATERIALI E METODI

In questa sezione verranno descritte tutte le metodologie sperimentali utilizzate sia nella fase dell'analisi genetica, sia per le procedure utilizzate negli studi funzionali.

4.1 Descrizione della casistica

Tutti gli individui arruolati nello studio hanno compilato una scheda con informazioni anagrafiche, anamnestiche e biometriche e firmato il consenso informato, approvato nelle rispettive ASL di provenienza.

Tutti i volontari sono stati tutelati dal punto di vista della privacy mediante l'identificazione di ciascuno di loro con un codice numerico. Sono state raccolte ulteriori informazioni riguardo alle condizioni di salute del volontario e dei suoi familiari affinché si potessero avere a disposizione la maggior quantità di dati per l'analisi.

Sono stati arruolati solo individui sardi e di origine sarda, ovvero con almeno tre linee parentali native della Sardegna. Sono stati selezionati esclusivamente individui che non mostravano tra loro relazione, entro il primo grado, di parentela. Quest'analisi è stata primariamente effettuata attraverso una comparazione dei dati anagrafici ed informazioni personali ed a posteriori con una più accurata analisi dei dati genotipici, effettuata attraverso il software Relative Finder (http://www.isogg.org/wiki/Relative_Finder).

La raccolta dei campioni di pazienti SM è iniziata oltre 20 anni fa. I campioni provengono da diversi centri clinici dell'isola; Clinica Universitaria di Sassari, Centro Sclerosi Multipla ASL8 di Cagliari ed Azienda Ospedaliera "G. Brotzu". La diagnosi di tutti i pazienti è stata effettuata nel rispetto dei criteri McDonald ed il rapporto maschi-femmine osservato è pari a 1:2,2.

La raccolta degli individui sani (controlli) è stata effettuata presso diversi centri trasfusionali dell'isola; Centro Trasfusionale dell'Azienda Ospedaliera "G. Brotzu", Centro trasfusionale di Sassari e Centro Trasfusionale ASL6 di San Gavino Monreale. Il 75% dei controlli utilizzati nello studio sono maschi, che riflette la percentuale degli uomini rispetto alle donne, dei donatori sardi.

I controlli utilizzati per gli studi funzionali provengono dal centro trasfusionale dell'ospedale Brotzu di Cagliari.

4.2 Estrazione DNA

I campioni di DNA utilizzati sono stati ottenuti da sangue periferico, trattato con l'anticoagulante acido etilendiamminotetracetico (EDTA), attraverso la classica metodica dell'estrazione salina (*salting-out*).

Questa metodica si basa sul processo di solubilizzazione delle proteine, la quale è dipendente da caratteristiche fisico-chimiche, dalla temperatura, dal pH e dalla concentrazione salina della soluzione. Infatti, ad alte concentrazioni di sali, la solubilità delle proteine diminuisce, causandone la loro precipitazione.

Il metodo prevede l'isolamento delle cellule nucleate, presenti nel campione di sangue, dopo aver rimosso per lisi i globuli rossi. I leucociti isolati, vengono trattati con un tampone di lisi (Buffer A, SDS e proteinasi K) allo scopo di estrarre gli acidi nucleici e degradare le proteine presenti, che vengono allontanate mediante precipitazione con i sali (NaCl soprassaturo). Infine, mediante trattamento con isopropanolo si ottiene la precipitazione del DNA, il quale viene recuperato sotto forma di nubesola e risospeso in una soluzione contenente 1mM Tris-HCl, 0.1mM EDTA 1M (TE 1:0.1).

Il DNA genomico estratto è stato sottoposto a vari controlli di qualità che prevedono una valutazione visiva dopo corsa elettroforetica su gel d'agarosio al 1%, la stima della concentrazione e la valutazione della purezza (rapporto OD260/280), con analisi spettrofotometrica al Nanodrop 1000.

4.3 Genotipizzazione mediante l'utilizzo di GeneChip arrays Affymetrix

La genotipizzazione mediante il GeneChip arrays Affymetrix v 6.0, che include 1.800.000 varianti di cui 906.600 SNPs e 946.000 variazioni di numero di copie (Copy Number Variation, CNV), è stata eseguita secondo il protocollo della ditta produttrice, di seguito brevemente descritto.

Una quantità di 500 ng di DNA genomico è stata digerita utilizzando separatamente gli enzimi di restrizione NspI e Styl. Alle estremità dei frammenti prodotti, sono state aggiunte, tramite una ligasi, delle brevi sequenze nucleotidiche, a sequenza nota, dette adattatori, che riconoscono le 4 bp coesive dell'oligonucleotide. Un oligonucleotide generico complementare alla sequenza dell'adattatore viene, di seguito, utilizzato per l'amplificazione.

Come da protocollo è stato utilizzato il seguente programma di amplificazione: 94° C 3 min di denaturazione, 94° C 30 sec, 60° C 45 sec, 68° C 15 sec per 30 volte, 68° C per 7 min.

Le condizioni di PCR sono state ottimizzate per amplificare preferenzialmente frammenti da 200 a 1.100 bp. I prodotti di PCR corrispondenti a ciascuno dei due enzimi di restrizione sono stati uniti, purificati con biglie di polistirene (Invitrogen™) e quantificati con spettrofotometro (Spectramax, Molecular Devices). Il DNA amplificato è stato in seguito frammentato, marcato ed ibridato al GeneChip array. Dopo una serie di lavaggi, eseguiti in una stazione fluidica, è stata effettuata la lettura mediante uno scanner.

4.4 Sequenziamento Next Generation di DNA genomico

Le sequenze sono state prodotte attraverso l'utilizzo di sequenziatori di nuova generazione (Next Generation Sequence, NGS); Genome Analyzer IIx (GAIIx) e Hi-Seq 2000 (Illumina). Sono state dapprima preparate le librerie di DNA, in accordo con il protocollo Illumina(50).

Brevemente, il DNA genomico è stato frammentato in maniera casuale, mediante un processo di nebulizzazione o sonicazione (Covaris S, Applied Biosystems), in frammenti sotto 800 paia di basi. Successivamente le estremità dei frammenti, in 3' e 5', sono state riparate e fosforilate. I frammenti di DNA riparati sono stati adenilati in 3' con una DNA polimerasi *Klenow exo* (NEB) e poi sono stati aggiunti degli adattatori (IDT) con l'impiego di DNA ligase. I prodotti di ligazione, di dimensione compresa tra le 300 e le 400 paia di basi, sono stati selezionati su gel di agarosio 2%, purificati con il saggio Qiagen Gel Extraction e, successivamente, preamplificati mediante PCR, utilizzando degli oligonucleotidi (IDT) compatibili con gli adattatori. Gli ampliconi ottenuti sono stati purificati con saggio Qiagen Gel Extraction e successivamente è stata valutata la concentrazione e la distribuzione dei frammenti delle librerie mediante corsa su Genechip arrays DNA 1000 nel Bioanalyzer 2100 (Agilent Technologies).

Le librerie sono state ibridate e amplificate sulla superficie di uno specifico vetrino, detto flow-cell, mediante un processo di amplificazione a ponte "bridge amplification", formando dei clusters, che vengono sequenziati con il GAIIx, in corse paired-end da 240 basi (con Paired End Cluster Generation Kit versioni 3,4 e SBS Cycle Sequencing Kit versioni 2, 4, 5 di Illumina), o con l'Hi-Seq 2000 in corse da 202 basi, ottenendo una copertura media di 3-4X.

Il sequenziamento, nella piattaforma Illumina impiega un metodo ciclico di incorporazione, fluorescenza, l'acquisizione di immagine e lavaggio, sfruttando la chimica di particolari dinucleotidi modificati, sviluppati sulla base dei dideossinucleotidi utilizzati nella tecnologia

Sanger. Tali nucleotidi sono rappresentati dalla molecola 3'Ozidomethyl, che riproducono dei terminatori reversibili. Durante il sequenziamento è presente una miscela delle quattro basi dinucleotidiche modificate e fluorescinate con quattro diversi fluorocromi che competono tra di loro. A seguito dell'incorporazione della base complementare alla prima base del filamento, il laser eccita la molecola ed il segnale luminoso viene raccolto come immagine per identificare la base appena aggiunta. Nel passaggio successivo viene ripristinato il 3'OH del dinucleotide modificato, appena incorporato, che sarà quindi in grado di accogliere la seconda base. Questo processo si verifica contemporaneamente per tutti i filamenti all'interno della *flow-cell* e le immagini vengono raccolte indipendentemente per ciascuno degli otto canali della *flow-cell*. L'Hi-Seq 2000, in otto giorni di lavoro, è in grado di produrre circa 200 miliardi di basi. Le risorse computazionali, necessarie per ogni corsa, sono imponenti (30 terabasi) come del resto l'analisi bioinformatica che è prevista.

4.5 Analisi delle sequenze NGS

Le analisi dei dati di sequenza prodotte con sequenziatori NGS rappresentano un forte impegno per i ricercatori sia in termini di tempo che di risorse computazionali che devono essere dedicate. Infatti, i dati grezzi delle sequenze sono costituiti da brevi sequenze di 100-200 paia di basi, chiamate reads, che devono essere allineate con un genoma umano di referenza.

Nel nostro studio le corte sequenze sono state allineate attraverso il software Burrows-Wheeler Aligner (BWA), <http://bio-bwa.sourceforge.net>.

Dai dati totali sono state filtrate le reads duplicate, attraverso il programma Picard, <http://picard.sourceforge.net>. I duplicati si generano, verosimilmente, durante la fase d'amplificazione (Polymerase Chain Reaction, PCR) e possono alterare le frequenze alleliche delle varianti, e quindi, diminuire la sensibilità e specificità di rilevamento delle stesse (53).

La qualità delle basi, assegnata dal sequenziatore, è stata ricalibrata considerando diversi fattori confondenti, quali la presenza di regioni ricche in GC, la posizione della reads nella flowcell, e verificando la presenza nei database pubblici di un polimorfismo della stessa base.

La qualità delle basi è stata riassegnata in Phred Score: $-\log_{10}(p)$, dove p è la probabilità che la base sia errata. Sono state scartate tutte le basi con qualità < 20 . Questo processo è stato eseguito dapprima indipendentemente per ciascun campione e successivamente analizzando insieme tutti i campioni sequenziati. Attraverso il software SAMTOOL,

<http://samtools.sourceforge.net>, è stata creata una lista di basi nucleotidiche che rappresentano potenziali polimorfismi e quindi definita la probabilità dei possibili genotipi, salvati nel formato GLF (Genotype Likelihood File) (<http://genome.sph.umich.edu/wiki/GLF>) (54).

Sono state filtrate, poiché verosimilmente si trattava di regioni ripetute ed infine sono stati caratterizzati i genotipi. Quest'ultimo processo è stato eseguito con un programma scritto dai nostri collaboratori, che tiene conto delle reads lette in individui con aplotipi simili e delle relazioni familiari degli individui sequenziati. Poiché quest'algoritmo ricostruisce la trasmissione dei cromosomi all'interno delle famiglie, i genotipi risultanti sono sottoforma di aplotipi. La lista degli aplotipi del pannello di riferimento sardo è stato salvato in formato standard vcf (<http://www.1000genomes.org/wiki/Analysis/vcf4.0>).

Prima di procedere all'imputazione statistica, è stato effettuato un ulteriore controllo tra i dati di sequenza generati e i genotipi attribuiti dalla genotipizzazione Affymetrix. Ovvero per ciascun individuo sequenziato, sono stati confrontati tutti gli SNPs, caratterizzati con piattaforma Affymetrix, localizzati nel cromosoma 20 (scelto per la sua ridotta dimensione che consente tempi di analisi più rapidi) con i genotipi estrapolati dai dati di sequenza per i medesimi SNPs. Tale controllo ha permesso di escludere errori e artefatti nel processo di imputazione, dovuti a cross match di individui.

Il processo è stato eseguito utilizzando il software VerifyBamID con verifica delle reads mappate nel Bam file (Binary Alignment Map) versus un file di genotipi, inputGenotypes, definito da dati genotipici in strand positivo, formato binario PLINK. La linea di comando utilizzata è la seguente: `verifyBamID --reference [reference.fa] --in [inputReads.bam] --bfile [inputGenotypes] --out [outPrefix] --verbose`. Per ogni coppia di campioni è stato generato un score di probabilità chiamato P ibd, valore tra 0 e 1. Il valore 1 definisce la coincidenza tra il campione genotipizzato e quello sequenziato, mentre il valore 0 indica l'indipendenza dei campioni. Valori di P ibd uguali a 0.5 indicano un rapporto di parentela tra i campioni analizzati mentre un valore di 0.95 è indice di una probabile contaminazione del campione.

4.6 Pannelli di referenza per l'Imputazione

Imputazione statistica è il termine utilizzato per descrivere il processo di predizione dei genotipi di varianti non direttamente genotipizzate, in un campione definito di individui. Il termine si riferisce alla situazione generale in cui un pannello di riferimento, costituito da aplotipi di una densa mappa di SNPs, viene usato per ricostruire i genotipi in un campione di individui che sono stati caratterizzati solo per un sottoinsieme di essi.

Tali metodi, consentono di analizzare, in studi di associazione su tutto il genoma, varianti non direttamente genotipizzate, in quanto assenti nelle mappe dei GeneChip arrays commerciali. I metodi di imputazione, per questo motivo, aumentano il potere degli studi di associazione su tutto il genoma di svelare loci di suscettibilità nei tratti in esame.

Nel nostro studio, per incrementare lo spettro delle varianti testate per associazione, abbiamo utilizzato metodi d'inferenza statistica, usufruendo di pannelli aplotipici di referenza generati da dati di sequenza su tutto il genoma a bassa copertura (low pass).

L'imputazione genotipica combina due o più insiemi di dati e si basa sul confronto degli individui dello studio, genotipizzati per un numero relativamente alto di marcatori genetici, e del pannello aplotipico di referenza utilizzato, che include informazioni genotipiche dettagliate, di un numero molto maggiore di marcatori. L'algoritmo identifica i marcatori comuni tra i due gruppi e gli aplotipi condivisi affinché i genotipi mancanti possano essere compilati in ciascun campione dello studio, copiando gli alleli osservati nell'aplotipo di riferimento corrispondente (55).

Nel nostro studio abbiamo utilizzato il software MACH (<http://www.sph.umich.edu/csg/abecasis/MACH/download/>).

Per questa fase di analisi sono stati utilizzati diversi pannelli di referenza ad alta definizione.

Pannello sardo: sono stati sequenziali 2120 individui sardi, di cui 1123 facenti parte della coorte del progetto SardiNIA (51), uno studio longitudinale rivolto allo studio di tratti quantitativi, e 997 selezionati da uno studio caso-controllo su diabete di tipo 1 e SM condotto sulla popolazione sarda (Sanna et manoscritto in preparazione). Nello studio SardiNIA gli individui sono stati reclutati su base familiare dai paesi dell'Ogliastra Elini, Ilbono, Arzana e Lanusei, e sono stati ampiamente caratterizzati per oltre 300 tratti quantitativi (56). I campioni relativi allo studio caso-controllo sono stati invece collezionati da diversi centri clinici della Sardegna, in particolare localizzati nelle città di Cagliari e Sassari. Tutti i campioni sono stati selezionati in base alla loro origine sarda.

Questa strategia ha mirato alla selezione di individui appartenenti a famiglie trios (genitori e figlio) in quanto, ciò porta ad una migliore efficienza dell'algoritmo di imputazione nella chiamata genotipica delle varianti ed una migliore ricostruzione delle fasi aplotipiche, nonché un miglioramento delle prestazioni dell'inferenza statistica nel propagare probabilisticamente i dati della sequenza a individui supplementari non sequenziati. Gli individui selezionati sono stati sequenziati con un approccio low pass sequencing, con una copertura media del genoma di 4 volte (4x).

Pannello mondiale (1000G ALL): il progetto internazionale "1000 Genomes Project" (<http://www.1000genomes.org>) ha messo a disposizione un database pubblico dettagliato della variazione genetica umana, contenente una mappa con un livello di risoluzione senza precedenti. Per questo motivo attualmente, i pannelli più utilizzati per l'inferenza statistica sono quelli resi disponibili dal progetto 1000 Genomes (1000G), in quanto contengono il più alto numero di varianti rare o poco frequenti e permettono di ottenere un'imputazione di maggiore qualità. Il progetto 1000G ha descritto il genoma di 1092 individui (2184 aplotipi) provenienti da quattordici popolazioni originarie dell'Europa, dell'Asia Orientale, dell'Africa sub sahariana e delle Americhe. Dall'analisi di questi individui è stata generata una mappa aplotipica validata di 38 milioni di polimorfismi a singolo nucleotide (SNPs), 1,4 milioni di inserzioni e delezioni brevi (INDELs) e più di 14 mila grandi delezioni.

Pannello europeo (1000G EUR): questo pannello di referenza si basa sulle sequenze di 392 individui europei e, come il precedente, permette di analizzare oltre 30 milioni di SNP e 1,4 milioni di INDELs e più di 14 mila grandi delezioni derivanti dalla popolazione europea.

Dopo ciascuna fase d'imputazione è stata valutata la qualità dei genotipi inferiti. Il metodo prevede che tutti gli SNPs con bassa qualità d'imputazione vengono filtrati.

Ci sono diverse metriche con le quali viene valutata la qualità e l'accuratezza dell'imputazione; il parametro RSQR (Russell Square Quality Representatives), che rappresenta il rapporto della varianza media osservata rispetto all'atteso di ciascun SNP ed il parametro r^2 che indica il rapporto tra la varianza osservata del dosaggio allelico e la varianza attesa, in equilibrio di Hardy-Weinberg.

Entrambe le metriche sono dei valori che variano tra 0 e 1, nella quale il valore 1 indica la presunta certezza dei genotipi assegnati ed il valore 0, la totale incertezza circa i genotipi imputati.

E' consigliata l'applicazione della soglia di 0.3, la quale è statisticamente capace di rimuovere il 90% degli SNPs inferiti in maniera errata. In genere, si guardano con sospetto gli SNPs con

RSQR tra 0.3 e 0.5, con prudenza quelli con RSQR tra 0.5 e 0.8, e con maggiore confidenza quelli con valori al di sopra di 0.8.

In questo studio sono stati esclusi dall'analisi tutti gli SNPs con un RSQR minore di 0.3.

4.7 Tipizzazione INDEL rs147983726

La genotipizzazione dell'INDEL rs147983726 (-/TATT) che mappa in posizione chr3:105601077-105601080 è stata condotta mediante l'analisi di frammenti di PCR con il sequenziatore a 48 capillari "3730 DNA Analyzer", Applied Biosystem. Tale metodica prevede la separazione dei frammenti con differenti lunghezze mediante corsa elettroforetica all'interno del capillare ed il loro dimensionamento grazie ad un marcatore del DNA a lunghezze conosciute (GeneScan™ Size Standard 500). Lo standard è stato combinato con il campione di interesse e co-iniettato sul sistema di elettroforesi capillare.

La regione di interesse (contente il marcatore rs147983726) è stata amplificata usando un oligonucleotide marcato con il tag fluorescente FAM (molecola capace di assorbire l'energia fornita da un laser e riemetterla a differenti lunghezze d'onda specifiche). In particolare, al fine di ottimizzare tempi e risorse di lavoro sono stati utilizzati 4 sets di oligonucleotidi capaci di amplificare frammenti di diversa lunghezza, ovvero 230, 288, 343 e 394 nucleotidi, intorno alla sequenza di interesse. Questa parte del flusso di lavoro garantisce che frammenti di DNA possano essere etichettati per lunghezza ed consente di eseguire la corsa elettroferetica per 4 diversi individui contemporaneamente in ogni capillare.

In questo modo è stato possibile validare i genotipi imputati dell'INDEL mediante i pannelli di referenza selezionati ed estendere successivamente l'analisi a tutta la casistica a nostra disposizione.

4.8 Test di associazione

Nell'analisi d'associazione sono stati testati oltre 30 milioni di SNPs, direttamente genotipizzati con Affymetrix 6.0 ed imputati, in 2000 casi e 2000 controlli.

Abbiamo applicato il test d'associazione (basato sulle tipizzazioni di pazienti versus controlli) sia sui marcatori autosomici che localizzati sul cromosoma X e calcolato il chi quadro con correzione per sottostruttura utilizzando il software Eigenstrat [<http://genepath.med.harvard.edu/~reich/Software.htm>].

Abbiamo, quindi, valutato visivamente i plot della discriminazione genotipica di ciascun SNP che mostrava un *P-value* $<10^{-5}$, eliminando tutti i marcatori che presentavano un'alterata attribuzione dei genotipi.

4.9 Estrazione di cellule mononucleate del sangue periferico (PBMC) da sacca di buffy-coat

Per effettuare gli esperimenti funzionali descritti in questa tesi è stato necessario ottenere una grande quantità di cellule T e B per ciascun individuo analizzato. Infatti, gli esperimenti funzionali in linfociti T e B, sia di espressione che di attivazione, sono stati effettuati in parallelo per ciascun individuo, in modo da avere un quadro completo dell'effetto della variante associata in relazione al genotipo, senza possibili errori dovuti al momento in cui è stato effettuato il prelievo o introdotti dalle procedure di manipolazione delle cellule durante il processamento dei campioni. Grazie alla collaborazione con il centro trasfusionale dell'ospedale Brotzu di Cagliari, per gli individui con i genotipi di interesse è stato possibile ottenere sacche di Buffy-Coat da cui abbiamo isolato le cellule mononucleate del sangue periferico umano (PBMC). Queste sono state prodotte al centro trasfusionale mediante centrifugazione a 300 X e successivamente processate per la separazione di emazie, plasma e buffy coat. Successivamente, nell'arco di 1-2 ore dalla preparazione, presso i nostri laboratori, il buffy coat è stato diluito 1:3 con una soluzione salina (PBS) tamponata a pH 7 e stratificato per gradiente (Histopaque) nel rapporto buffy coat/soluzione 2:1. Dopo centrifugazione a 1800 rpm per 30 minuti, l'anello di PBMC che si forma all'interfaccia viene raccolto e lavato per 3 volte con PBS+2mM EDTA ed infine sospeso in MACS buffer. Il numero delle cellule ottenute viene determinato attraverso la conta in una camera di Burker e la percentuale di cellule vive determinata con il metodo dell'esclusione del colorante (trypan-blue). La resa media di un'estrazione di PBMC da Buffy Coat è di 5×10^8 cellule/ml, di cui il 70% circa linfociti T e il 5-10% linfociti B.

4.10 Separazione Linfociti T e B

Le cellule T e B sono state purificate dai PBMC attraverso una procedura di deplezione negativa con il sistema MACS (Magnetic cell Sortine System, Miltenyi Biotec) utilizzando saggi specifici. Il saggio utilizzato per separare le cellule T (Pan T cell Isolation Kit II, Miltenyi Biotec) permette di isolare le cellule T da PBMCs tramite l'utilizzo di un cocktail di anticorpi monoclonali coniugati alla biotina (biotin-antibody Cocktail). Questo cocktail consente di marcare cellule che espongono sulla propria superficie di membrana i markers CD14, CD16 e CD56, CD19, CD36, CF123 e glicoforina A che identificano rispettivamente monociti, cellule NK, linfociti B, cellule dendritiche ed eritrociti. Successivamente si procede con una marcatura con un anticorpo monoclonale secondario anti-biotina (anti-biotin Microbeads) coniugato a biglie magnetiche che consente l'eluizione in colonna delle sole cellule T, che in questo modo non vengono attivate rimanendo in uno stato quiescente. Sono state utilizzate le colonne MS (MS columns, Miltenyi Biotec). Questo tipo di colonna consente di purificare fino ad un massimo di $1 \cdot 10^7$ cellule marcate. La colonna viene montata sul separatore magnetico e idratata con 500ul di buffer MACS, dopo di che si aggiungono 500ul di sospensione cellulare e viene lavata con Buffer MACS per 3 volte. Le colonne vengono poi centrifugate per eliminare il MACS buffer e le cellule vengono eluite e risospese nel terreno di coltura (RPMI + L-glutamina 1% + antibiotici 1%+ HSA 10%). Stessa procedura è stata utilizzata per l'estrazione dei linfociti B attraverso l'utilizzo del saggio B Cell Isolation Kit (B-CELL), che contiene un cocktail di anticorpi diretti verso cellule non-B (CD2, CD4, CD11b, CD16, CD36, Anti-IgE, e Glycophorin A).

4.11 Estrazione dell'RNA da linfociti T e B e RT-qPCR

L'RNA totale è stato estratto da $1 \cdot 10^6$ linfociti T o B (RNeasy mini kit, Qiagen Crawley, UK) secondo il protocollo dell'azienda produttrice. Gli RNA sono stati retrotrascritti con SuperScript III (Invitrogen™). Campioni identici sono stati trattati in assenza di trascrittasi inversa e impiegati come controlli per eventuale contaminazione genomica del DNA.

La RT-qPCR è stata condotta utilizzando dapprima la chimica SYBR® Green, un fluorocromo che durante la fase dell'estensione della PCR è in grado di intercalarsi all'interno della doppia elica in formazione. In questa fase viene rilevata la fluorescenza che rispecchierà la quantità iniziale di cDNA (DNA complementare) nella reazione. Qualora durante le reazioni di PCR avvenga un'amplificazione di DNA non specifica, anche essa sarà rilevata. Questo problema può

comunque essere superato effettuando l'analisi della curva di dissociazione alla fine di ogni corsa.

I risultati di questo studio sono stati validati in un secondo step, utilizzando i saggi TaqMan® i quali , per generare il segnale fluorescente, prevedono un'ibridazione specifica tra il DNA-target e la sonda (probe); in questo caso le amplificazioni non specifiche, come ad esempio gli artefatti dovuti alla dimerizzazione degli oligonucleotidi, non genereranno alcun segnale.

Nella seguente tabella sono riportate le sequenze degli oligonucleotidi e delle sonde che sono state utilizzate per gli esperimenti svolti con chimica SYBR® Green e saggio TaqMan® per la discriminazione delle isoforme che hanno mostrato espressione.

Isoforma	Oligonucleotide Forward SYBR® Green	Oligonucleotide Reverse SYBR® Green
CBLB 001	TCGTTGCACTGTTGAGTGCAA	GCAGTGTGGAGCAATGAATGA
CBLB 002	GACTCACTGGGTTGTAGTTTGC	CCTGACTGATGCCCAACTACTACA
CBLB 003	GACTCACTGGGTTGTGAATTTGC	CCTGACTGATGCCCAACTACTACA
CBLB 004	GATGCAGTTGGACCCCTAA	AGGCAAATCATGGCGTCTGT
CBLB 005	AGGAATACTAGTTCCTCTACCTCTCAA	GACCATGTGTTCCATTGAGCAT
CBLB 006	TTCAGCCAATCAACAGAGATCATC	TTTGACATTTTCACCTGGTAACCTT
CBLB 008	ATGAATTTCTCCAGCGTTGGA	TTCTGCCATTGAGTTTGC
CBLB 007/201	CCTCCTGTTCCGGTCTTGATAA	CATCAAAGCATCTTACCCT
CBLB 009	CTGCCACCCCTTATTGAACTAAA	CAACTGCATCTGGATAGCATCA
CBLB 010	ACCTGCCACCCCTTATTGGT	CGCAGCACGTGAGAAAGGA

Tab. 2: Set di oligonucleotidi utilizzati per la discriminazione delle isoforme di CBLB con la chimica SYBR® Green

Isoforma	Oligonucleotide FORWARD	Probe TaqMan®	Oligonucleotide REVERSE
CBLB 001	GACGCCC TCTGATTATGATCTTCT	CCCTCCATTAG	ATGCTTTTATGATGCCCTCC
CBLB 003	AATCAAACATCCCTGACTTAAGCATA	ATTTAAAGGGAGATGTTTTTG	TCAGCCTCTGATCCCGTG
CBLB 002/003	GCCTGAAATTTGAATGGGTTCT	CAGGTGCCGGTGAC	TGGAGTAATGGTCTCGTGAACCTG

Tab. 3: Set di oligonucleotidi utilizzati per la discriminazione delle isoforme di CBLB con saggio TaqMan®

Per normalizzare i valori di espressione del gene è stato condotto uno studio per identificare i migliori geni di riferimento specifici per le cellule T e B, rispettivamente. Partendo da un set iniziale di 10 geni di riferimento (Fig. 8) sono stati selezionati, in seguito ad analisi con i software GeneNORM e NORMFINDER, la beta-2-microglobulina per i linfociti T e il Tyrosine3-monooxygenase/tryptophan 5-monooxygenase activation protein, zeta polipeptide per i linfociti B rispettivamente.

Internal control genes evaluated in this study								
Symbol	Accession number	Name	Function	Localization	Pseudo-gene*	Primers [†]	Alias	IMAGE [‡]
ACTB	NM_001101	Beta actin	Cytoskeletal structural protein	7p15-p12	+	S		510455
B2M	NM_004048	Beta-2-microglobulin	Beta-chain of major histocompatibility complex class I molecules	15q21-q22	-	S		51940
GAPD	NM_002046	Glyceraldehyde-3-phosphate dehydrogenase	Oxidoreductase in glycolysis and gluconeogenesis	12p13	+	D		510510
HMBS	NM_000190	Hydroxymethyl-bilane synthase	Heme synthesis, porphyrin metabolism	11q23	-	D	Porphobilinogen deaminase	245564
HPRT1	NM_000194	Hypoxanthine phosphoribosyl-transferase I	Purine synthesis in salvage pathway	Xq26	+	D		345845
RPL13A	NM_012423	Ribosomal protein L13a	Structural component of the large 60S ribosomal subunit	19q13	+	D	23 kDa highly basic protein	-
SDHA	NM_004168	Succinate dehydrogenase complex, subunit A	Electron transporter in the TCA cycle and respiratory chain	5p15	+	D		375812
TBP	NM_003194	TATA box binding protein	General RNA polymerase II transcription factor	6q27	-	D		280735
UBC	M26880	Ubiquitin C	Protein degradation	12q24	-	D		510582
YWHAZ	NM_003406	Tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, zeta polypeptide	Signal transduction by binding to phosphorylated serine residues on a variety of signaling molecules	2p25	+	S [§]	Phospholipase A2	416026

*Presence (+) or absence (-) of a retroseudogene in the genome determined by BLAST analysis of the mRNA sequence using the high-throughput genomic sequences database (htgs) or human genome as database. [†]Localization of forward and reverse primer in different exons (D) or the same exon (S). [‡]IMAGE cDNA clone number according to [14]. [§]A single-exon gene.

Fig. 8: Set dei 10 geni di riferimento analizzati e da cui sono stati selezionati quelli specifici per le cellule T e B

Gli esperimenti sono stati svolti in triplicato tecnico e ripetuti per tre volte. Per analizzare i dati della RT-qPCR è stato utilizzato il metodo del “delta-delta Ct”. L’analisi statistica di significatività è stata condotta tramite *T test*, utilizzando come valore soglia di significatività un $p \leq 0,05$.

4.12 Stimolazione dei Linfociti T e B

I linfociti T sono stati stimolati con apposite biglie CD3/CD28 (Dynabeads® Human T-Activator CD3/CD28 Invitrogen™), sulla cui superficie sono ancorate molecole di CD3 e CD28, utilizzati a differenti concentrazioni e per diversi tempi di induzione. In particolare, per gli esperimenti di espressione di *CBL-B* è stata utilizzata una concentrazione biglia/cellula 1:2, e le cellule sono state stimolate per 3, 6, 12, 24 ore. Per valutare la produzione di IL-2 tramite ELISA è stata utilizzata una concentrazione biglia/cellula 1:5 per 20 e 48 ore.

I linfociti B sono stati stimolati sia in modo T-indipendente, con anti-IgM (Fab)₂ (2ug/ml) + IL-4 (5ng/ml), che T- dipendente (CD40L + IL-4), per 12, 24, 48 e 72 ore. Per valutare la traslocazione nucleare di NF-κB e quindi l'attivazione delle cellule B tramite ELISA, i tempi di induzione analizzati sono stati di 15, 30 e 60 minuti.

4.13 Esperimenti ELISA di quantificazione di IL-2 e dell'attivazione di NF-κB

Gli esperimenti ELISA sono stati condotti utilizzando differenti saggi. Per i linfociti T è stata valutata la concentrazione di IL-2 utilizzando il saggio Human IL-2 Elisa Kit Immunological Sciences (IK4142) secondo le istruzioni del produttore. L'IL-2 è stata misurata nel sovranatante dei linfociti T stimolati, il quale è stato diluito secondo un fattore di diluizione 1:50. Gli esperimenti indipendenti sono stati portati avanti in duplicato tecnico e biologico. La lettura dell'assorbanza O.D.=450 è stata effettuata con luminometro Synergy2.

Per i linfociti B è stata valutata l'attivazione di NF-κB con il saggio Elisa TransAM® NFκB p65 Active Motif. Questo saggio è un DNA-binding ELISA e permette di valutare lo stato di attivazione della subunità p65 di del fattore di trascrizione NF-κB. In caso di attivazione, NF-κB cambia conformazione e trasloca nel nucleo dove si lega alla sua sequenza nucleotidica bersaglio (in questo caso rappresentata da un oligonucleotide a doppio filamento immobilizzato sulla piastra fornita dal saggio). La quantificazione di NF-κB legato avviene tramite l'uso di un anticorpo specifico (incluso nel saggio) per la forma attiva legata del fattore di trascrizione. Gli estratti nucleari delle cellule B utilizzati per l'analisi sono stati prodotti utilizzando il saggio TransAM® Nuclear Extration Active Motif. Anche in questo caso la lettura dell'assorbanza O.D.=550 è stata effettuata con luminometro Synergy2. A causa della difficoltà di ottenere quantità sufficienti di cellule B, gli esperimenti sono stati effettuati solo in duplicato tecnico ma non biologico.

Per calcolare la significatività degli esperimenti è stato utilizzato il T test e il valore soglia di significatività è $p \leq 0.05$.

5 RISULTATI

I risultati descritti in questa tesi sono il frutto di un grosso lavoro d'equipe, e di una serie di collaborazioni a livello sia nazionale che internazionale. Innanzitutto, questo studio è stato possibile grazie alla collaborazione da parte di diversi enti di ricerca (l'Istituto di Ricerca Genetica e Biomedica del CNR Monserrato Cagliari, l'Università di Sassari ed il Centro di Ricerca, Sviluppo e Studi Superiori in Sardegna, CRS4) e di diversi centri clinici dell'isola (i centri SM dell'Università ed ASL di Cagliari, di Ozieri, dell'Università di Sassari, dell'ospedale Brotzu di Cagliari), che hanno costituito un consorzio sardo della SM, al fine di generare un ampio dataset di pazienti affetti e controlli sani su cui condurre studi genetici per raggiungere l'obiettivo comune di aumentare le conoscenze sulla biologia della SM, trovare una spiegazione all'alta prevalenza di questa patologia in Sardegna e mettere in luce i meccanismi ed i pathways coinvolti nella patologia, al fine di trovare nuove ed efficaci cure nel trattamento e nella prevenzione della SM. Inoltre, sono da sottolineare le collaborazioni nazionali ed internazionali che hanno consentito di completare il sequenziamento "next generation" degli individui sardi e la generazione del pannello di riferimento utilizzato per condurre il GWAS descritto in questa tesi. Il sequenziamento dei campioni è stato condotto in parallelo utilizzando la tecnologia Illumina all'Università del Michigan e al CRS4 a Pula. Questi due centri hanno anche messo a disposizione le risorse computazionali necessarie per l'analisi statistica dei dati generati ed utilizzati in questo studio.

5.1 Studio caso-controllo GWAS

Per avere una migliore rappresentanza della variabilità genetica ed essere in grado di testare in associazione un numero maggiore di varianti che possono avere un ruolo nella patogenesi della SM, abbiamo condotto un nuovo studio GWAS su pazienti affetti e controlli sani di origine sarda, testando varianti sia genotipizzate che inferite da pannelli di riferimento generati dal sequenziamento di individui di diversa origine etnica. Per aumentare il potere statistico rispetto allo studio precedente, abbiamo inoltre ampliato la casistica includendo 2000 casi con SM e 2000 controlli. Dopo aver tipizzato i campioni inclusi nello studio con il GeneChip arrays Affymetrix 6.0, abbiamo condotto l'analisi utilizzando per l'imputazione tre diversi pannelli di riferimento (vedi disegno dello studio e metodi). In questa tesi verranno riportati nello specifico i risultati ottenuti nella regione *CBL-B*, mentre i risultati su altri loci non verranno trattati in

quanto non rappresentano l'oggetto del mio dottorato. Brevemente i segnali già noti sia su loci HLA che non-HLA sono stati confermati con p-values genome-wide. Altri nuovi segnali di associazione suggestivi, tra cui alcuni sardo-specifici, sono stati messi in evidenza e sono allo stato attuale in corso di approfondimento.

Il GWAS condotto utilizzando il pannello di riferimento sardo-specifico generato dal nostro gruppo (utilizzando le sequenze di 2120 individui sardi con copertura 4x), ha consentito di testare in associazione un totale di circa 17 milioni di varianti. Questa analisi ha identificato come variante maggiormente associata al locus *CBL-B* lo SNP rs11924742, con $p\text{-value}=1,97 \cdot 10^{-7}$ (Fig. 9). Questa variante è altamente correlata ($r^2=0.99$) con il top SNP rs9657904 precedentemente riportato dal nostro gruppo (33), e rappresenta quindi una conferma del segnale iniziale. Nessun segnale nuovo e/o sardo-specifico è stato messo in evidenza a questo locus utilizzando questo pannello di imputazione.

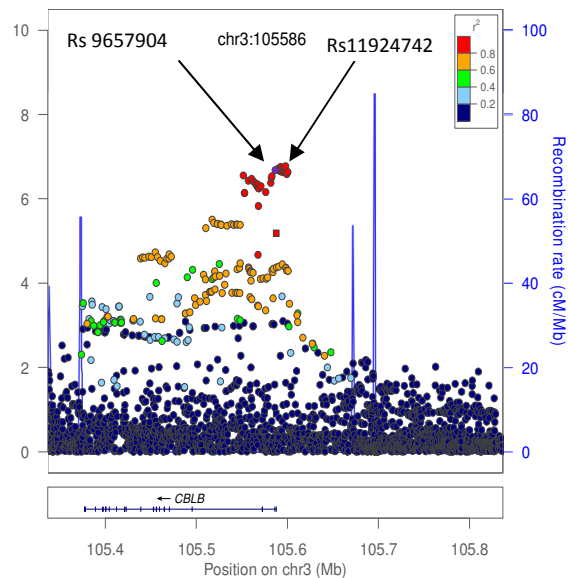


Fig.9: LocusZoom della regione che comprende il gene *CBLB*. Il pannello mostra i risultati di associazione in seguito all'imputazione con il pannello di riferimento delle sequenze sarde. Le frecce indicano le varianti rs9657904 e rs11924742.

MARKER	RS	CHR	POS	Frequenze casi e controlli	OR	P-VALUE
chr3:105598033	rs11924742	3	105598033	.8539(.8739,.8327)	1,369	$1,68 \times 10^{-7}$
chr3:105586714	rs9657904	3	105586714	.8528(.8728,.8318)	1,365	$2,08 \times 10^{-7}$

Tab. 4: Associazione con SM utilizzando il pannello di imputazione sardo

Il GWAS condotto utilizzando il pannello di riferimento 1000G europeo ha consentito di testare circa 30 milioni di SNPs e 1,4 milioni di INDELS per associazione con SM ed ha confermato come variante maggiormente associata nella regione *CBL-B* l'rs9657904, con un $p\text{-value}=1.15\times 10^{-7}$ (Fig. 10).

Infine, utilizzando il pannello 1000G mondiale, è stata identificata l'associazione con SM di una variante INDEL (delezione), rs14798726, strettamente correlata con lo SNP rs9657904 ($r^2=0.98$), con un $p\text{-value}=8.2\times 10^{-8}$ (Tab.5; Fig.10).

MARKER	RS	CHR	POS	Frequenze casi e controlli	OR	P-VALUE
chr3:105601076	rs147983726	3	105601076	.8428(.8626,.8218)	1.384	8.22×10^{-8}
chr3:105586714	rs9657904	3	105586714	.8528(.8728,.8317)	1.369	1.78×10^{-7}

Tab. 5: Associazione con SM utilizzando il pannello di imputazione 1000G mondiale

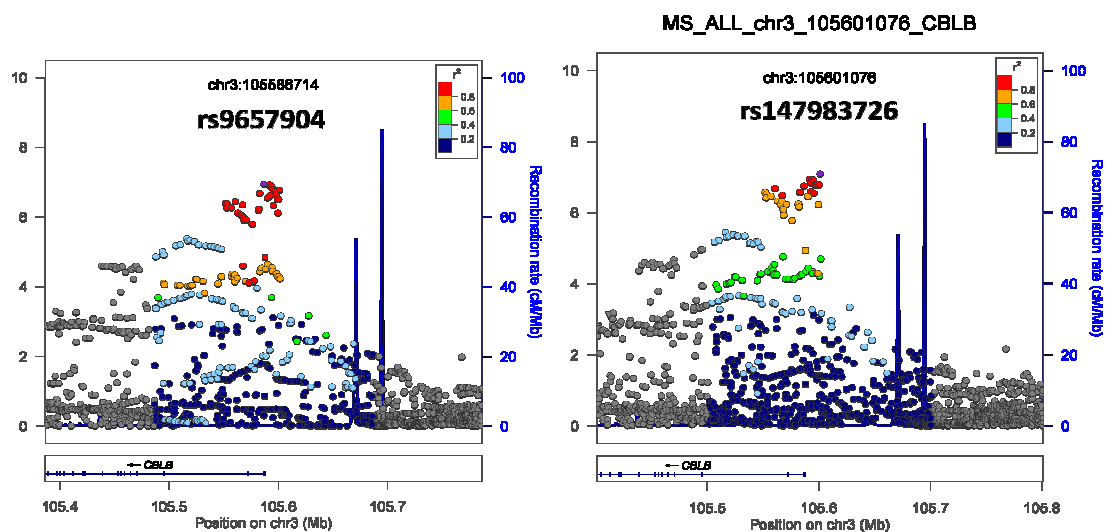


Fig.10: LocusZoom della regione che comprende il gene CBLB. Il pannello a sinistra mostra i risultati in seguito all'imputazione con il pannello di riferimento 1000G, mentre il pannello a destra illustra i risultati dopo l'imputazione con il pannello 1000G mondiale. In viola è indicato il top SNP per ciascun pannello.

Dal momento che i genotipi degli SNPs rs9657904 e rs147983726 sono stati inferiti (Tab. 6), l'associazione di queste due varianti è stata rianalizzata dopo aver confermato i genotipi, rispettivamente tramite TaqMan® e analisi di frammenti di PCR (vedi metodi), su tutta la casistica a disposizione di 4015 controlli e 3200 casi. Questa analisi ha convalidato l'associazione di entrambe le varianti, con $p\text{-value}=5.22 \times 10^{-9}$ per l'rs9657904, e $p\text{-value}=2,21 \times 10^{-8}$ per l'rs147983726.

L'rs9657904 si conferma quindi la variante più associata a SM nel locus *CBL-B*.

MARKER	RS	CHR	POS	Qualità di imputazione Pannello Sardo	Qualità di imputazione 1000G	Qualità di imputazione 1000G ALL
chr3:105586714	rs9657904	3	105586714	.9959	.9914	.9913
chr3:105601076	rs147983726	3	105601076	x	.9203	.9296

Tab 6: Qualità di imputazione per le varianti rs9657904 e rs11924742 nei diversi pannelli di inferenza utilizzati per l'analisi di associazione

Lo SNP rs9657904 è localizzato sul primo introne del gene mentre l'INDEL rs147983726, corrisponde ad una delezione di 4 nucleotidi (-/TATT), mappa a 15 kb a monte del 5' UTR del gene *CBL-B*. L'allele di rischio, T, per la variante rs9657904 e la delezione rs147983726 risultano mappare prevalentemente sullo stesso aptotipo: le due varianti infatti sono strettamente correlate tra di loro con un $r^2=0,98$. Le frequenze alleliche risultano essere quindi le stesse per le due varianti e sono pari a 0.16 (0.14 nei casi e 0.17 nei controlli) per l'allele protettivo C (corrispondente all'assenza della delezione), e 0.84 (0.86 nei casi e 0.83 nei controlli) per l'allele di rischio T.

Visto che i GWAS sono stati condotti utilizzando mappe dettagliate ad alta densità, le varianti identificate rappresentano quelle potenzialmente causative. Avendo un grado di correlazione così elevato e $p\text{-values}$ di associazione simili, le due varianti non possono essere distinte geneticamente e non si può quindi affermare quale sia quella primariamente responsabile dell'associazione e dell'effetto biologico sul rischio di SM. Questo potrà essere determinato solo attraverso esperimenti di dissezione molecolare e studi funzionali mirati.

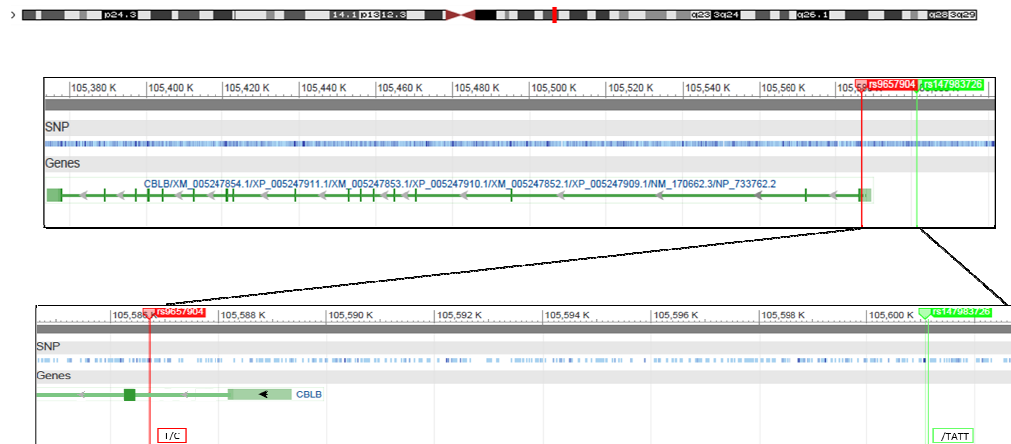


Fig. 11: Localizzazione sul cromosoma 3 delle varianti associate a SM. In rosso l' rs9657904 nell'introne 1 del CBL-B e in verde l'INDEL rs147983726 che mappa a 15 Kb a monte del 5' UTR del gene

5.2 Studio di espressione del gene CBL-B

Per studiare l'effetto delle varianti rs9657904 e rs147983726 associate a SM, sull'espressione del gene *CBL-B* abbiamo condotto esperimenti di RT-qPCR su linee primarie T e B di donatori sani selezionati per essere omozigoti per il genotipo protettivo, CC, o predisponente, TT, dello SNP rs9657904 ($r^2=0.98$ per rs147983726). Il genotipo di tutti gli individui utilizzati è stato confermato tramite sequenziamento; come atteso, in tutti all'allele T rs9657904 corrisponde la delezione rs147983726, e all'allele C corrisponde l'assenza di delezione. Gli esperimenti sono stati condotti in più stadi e con differenti metodiche (vedi metodi). In totale abbiamo analizzato le cellule T derivate da 20 individui (10 individui TT e 10 CC) e le cellule B provenienti da 10 individui (5 individui TT e 5 CC).

In una fase iniziale, è stata verificata l'espressione degli 10 trascritti alternativi riportati dalle banche dati pubbliche (*Ensembl*) per il gene *CBL-B*.

Nella tabella 6 vengono indicati per ciascuna isoforma l'estensione del trascritto e tradotti della putativa proteina da essi codificata. L'isoforma principale, che codifica per la proteina più estesa, è l'isoforma 001 lunga 6780 pb, mentre l'isoforma 006 che presenta un'estensione del trascritto di 4722 pb è predetta non risultare in alcuna proteina tradotta.

<i>Nome Isoforma</i>	<i>Estensione trascritto (pb)</i>	<i>Estensione tradotta (aa)</i>
<i>CBLB-001</i>	6780	982
<i>CBLB-007</i>	4475	321
<i>CBLB-201</i>	3430	960
<i>CBLB-002</i>	3350	770
<i>CBLB-003</i>	3237	810
<i>CBLB-202</i>	1632	227
<i>CBLB-005</i>	1493	197
<i>CBLB-010</i>	944	139
<i>CBLB-008</i>	561	139
<i>CBLB-009</i>	519	161
<i>CBLB-006</i>	4722	NON TRADOTTA IN PROTEINA

Tab 6: Estensione trascritto (pb) e della proteina (amminoacidi aa) per le 11 isoforme note di CBLB e descritte nei database (Ensemble)

5.2.1 Esperimenti di RT-qPCR nelle cellule T

Per stabilire i livelli di espressione delle principali isoforme del gene, sono stati condotti esperimenti di RT-qPCR preliminari. In questa prima fase i linfociti T sono stati estratti da 3 campioni con genotipo TT e 3 con genotipo CC. L'espressione delle varie isoforme è stata valutata utilizzando la chimica del SYBR® Green.

Data la forte omologia di sequenza, per alcune di esse non è stato possibile una discriminazione univoca del trascritto. Per questo motivo sono stati disegnati dei oligonucleotidi comuni che riconoscevano le isoforme 001/007, 002/003 e 007/201.

Come illustrato nella figura 12, solo le isoforme 001, 002, 003, 005 e 007/201 hanno mostrato di essere espresse nei linfociti T di individui TT (in rosso) e CC (in verde) allo stato basale.

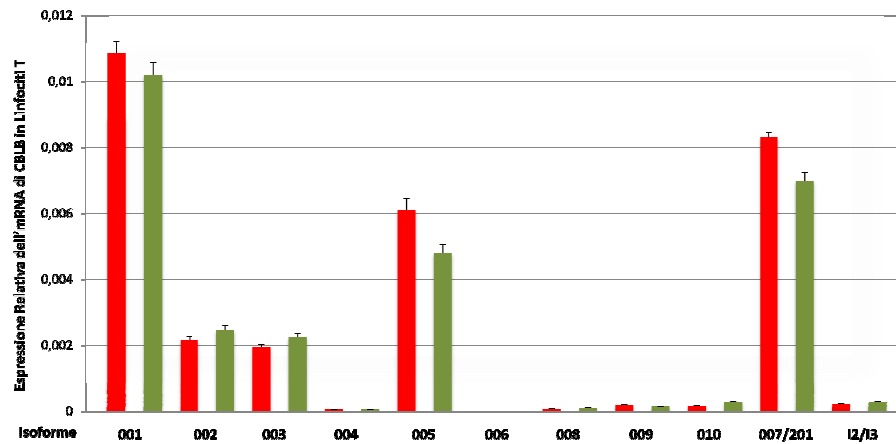


Fig. 12: Espressione relativa dell' mRNA delle isoforme di *CBL-B* delle medie di 3 individui con genotipo di predisposizione TT (in rosso) e 3 individui con genotipo protettivo CC (in verde).

In seguito ai test preliminari i successivi esperimenti sono stati eseguiti utilizzando i linfociti T estratti da 5 individui con genotipo di rischio TT e 5 individui con genotipo protettivo CC. L'espressione differenziale genotipo specifica è stata valutata in condizioni basali e in seguito a stimolazione con CD3/CD28 per 3, 6, 12 e 24 ore.

I risultati ottenuti non hanno evidenziato differenze genotipo specifiche significative per le isoforme analizzate.

Alla luce di questi risultati il campione è stato ampliato fino ad un totale costituito da 10 individui con genotipo di rischio TT e 10 campioni con genotipo protettivo CC. In questa terza fase gli esperimenti di RT-qPCR sono stati eseguiti utilizzando sonde TaqMan®, una tecnica più specifica del SYBR® Green.

Per analizzare l'espressione differenziale delle isoforme 001 e 003, non è stato possibile generare un saggio TaqMan® con oligonucleotidi che permettessero di discriminare le due isoforme separatamente. Quindi, abbiamo amplificato una regione compresa tra l'esone 15 e 16, che le comprende entrambe. Il saggio TaqMan® ha evidenziato un'espressione differenziale significativa per queste isoforme, con una minore espressione del gene nei campioni con genotipo predisponente TT, rispetto ai campioni con genotipo protettivo CC, sia in condizioni basali ($p\text{-value}=0.005$), sia in seguito a stimolazione con CD3/CD28 per 12 ore ($p\text{-value}=0.007$) e 24 ore ($p\text{-value}=0.04$) (Fig.13).

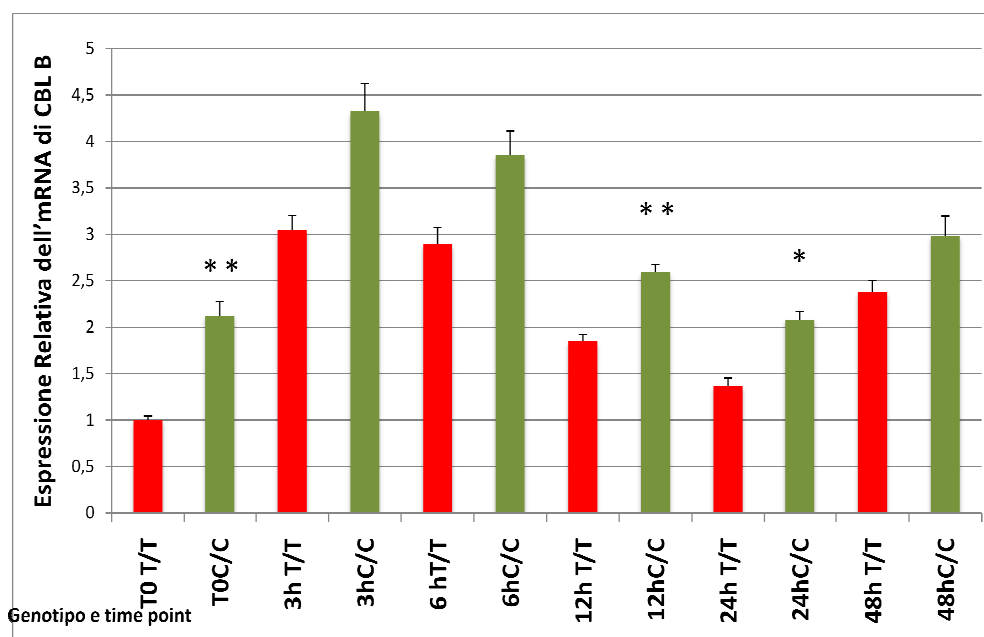


Fig. 13: Espressione relativa dell'mRNA delle isoforme di CBLB in linfociti T di 10 individui con genotipo di predisposizione TT (in rosso) e 10 individui con genotipo protettivo CC (in verde) a differenti tempi di induzione. Il singolo asterisco(*) o doppio asterisco(**) indica rispettivamente $p < .05$ o $p < .01$.

Per stabilire quale delle due isoforme, tra la 001 e la 003, fosse differenzialmente espressa, sono stati condotti ulteriori saggi TaqMan® in linfociti T di 10 individui omozigoti rispettivamente per i genotipi TT e CC e 5 individui eterozigoti C/T in condizioni basali.

In un set di esperimenti è stata utilizzata una coppia di oligonucleotidi localizzati nel 3'UTR delle isoforme 002 e 003, che le amplifica entrambe e non analizza quindi l'isoforma 001. Come riportato nella figura 14, non sono state evidenziate differenze significative genotipo-specifiche in linfociti T di individui con genotipo TT, CT e CC. La mancanza di differenze significative probabilmente esclude il coinvolgimento dell'isoforma 003 nell'espressione genotipo specifica individuata nell'esperimento precedente condotto su 20 campioni totali.

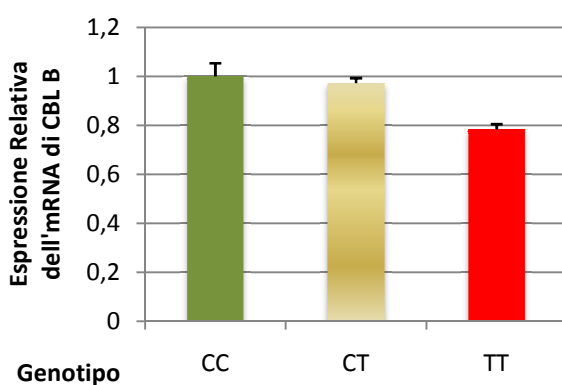


Fig. 14: Espressione Relativa dell'mRNA di CBL-B in 10 campioni con genotipo TT (rosso) e 10 campioni con genotipo CC (verde) e 5 campioni con genotipo eterozigote CT (in giallo) in condizioni basali utilizzando un saggio TaqMan® che analizza il 3' UTR delle isoforme 002 e 003.

Per confermare che l'espressione differenziale genotipo-specifica fosse dovuta esclusivamente all'isoforma 001, è stato condotto un saggio TaqMan® utilizzando questa volta una coppia di oligonucleotidi che mappano tra gli esoni 16 e 17, che amplificano esclusivamente questa isoforma. Quest'ultimo saggio ha confermato l'espressione significativamente più bassa dell'isoforma 001 del gene *CBL-B* in individui con genotipo predisponente TT ($p\text{-value}=0,005$) rispetto a quelli omozigoti CC (Fig. 15).

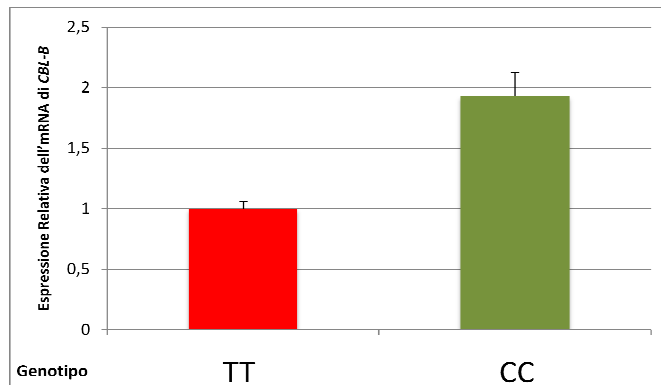


Fig. 15: Espressione Relativa dell'mRNA di *CBL-B* in 10 campioni con genotipo TT (in rosso) e 10 campioni con genotipo CC (in verde) in condizioni basali utilizzando un saggio TaqMan® posizionato tra gli esoni 16 e 17.

Complessivamente, gli esperimenti di RT-qPCR hanno quindi permesso di identificare un'espressione differenziale genotipo-specifica dell'isoforma principale 001 del gene *CBL-B* nei linfociti T, dovuto all'effetto delle varianti prese in esame, rs9657904 e rs147983726. In particolare, la riduzione dell'espressione del trascritto principale di *CBL-B* negli individui con genotipo predisponente TT, sia in condizioni basali che dopo induzione, dovrebbe risultare in una minore espressione della proteina. Questo a sua volta dovrebbe portare ad un fenotipo di iperattivazione dei linfociti T e quindi ad una predisposizione allo sviluppo di autoimmunità.

5.2.2 Esperimenti RT-qPCR su cellule B e Linee Linfoblastoidi (LCL)

Lo stesso disegno sperimentale è stato seguito per le cellule B. Per valutare quali trascritti del *CBL-B* fossero espresse in questa linea cellulare primaria l'esperimento di RT-qPCR è stato condotto su cellule B non stimolate estratte da un singolo individuo il cui genotipo relativo alla variante rs9657904 era quello di rischio TT. In queste cellule sono risultate espresse solo le isoforme 001, 005, 006 e 007/201 (Fig. 16).

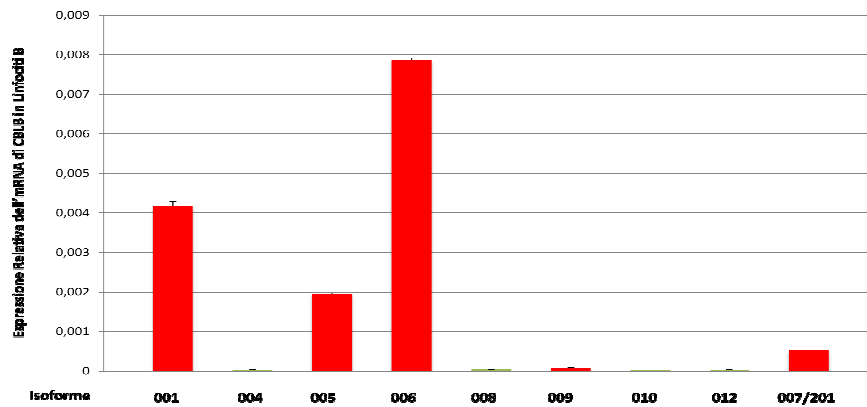


Fig. 16: Espressione relativa dell' mRNA delle isoforme di *CBL-B* in linfociti B di un individuo con genotipo di predisposizione TT all' rs9657904.

Nelle cellule B di 10 individui di cui 5 con genotipo TT e 5 CC esperimenti di RT-qPCR non hanno riscontrato differenze significative nell'espressione di *CBL-B* in nessuna delle isoforme analizzate sia in condizioni basali che in seguito a stimolazione T-indipendente con α -IgM (Fab)₂ e IL-4 a diversi tempi di induzione. La figura 17 mostra i risultati ottenuti per l'isoforma 001 rappresentativa di tutte le restanti isoforme studiate(Fig.17). La stimolazione T-indipendente utilizzando il CD40L e IL-4 non è risultata efficace per osservare differenze di espressione di *CBL-B* in queste cellule, e non è stata pertanto utilizzata negli esperimenti successivi (dati non mostrati).

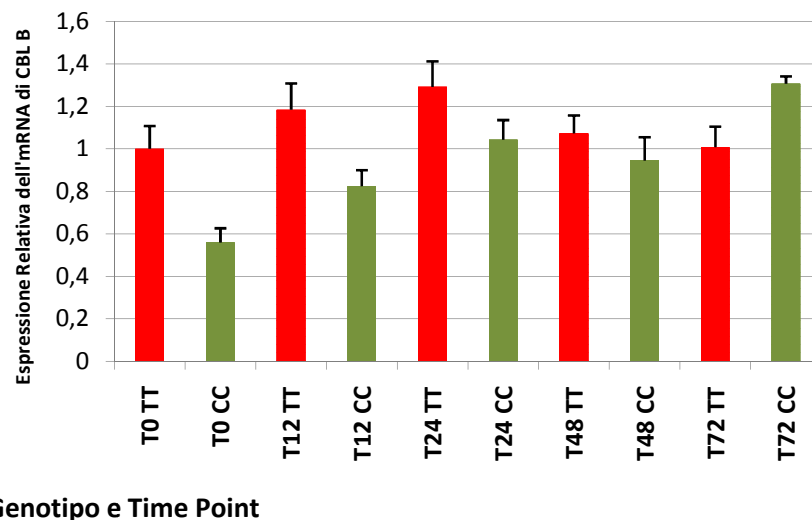


Fig. 17: Espressione relativa dell' mRNA dell' isoforma 001 di *CBL-B* in linfociti B di 5 individui con genotipo di predisposizione TT (in rosso) e 5 individui con genotipo protettivo CC (in verde) a differenti tempi di induzione.

Per estendere lo studio sulle cellule B, abbiamo utilizzato linee linfoblastoidi (LCL), ovvero delle linee generate da cellule B immortalizzate tramite infezione con il virus Epstein-Barr (EBV) estratte da volontari sardi arruolati nel progetto SardiNIA (51) e tipizzati per le varianti in esame. Infatti, si è dimostrato molto complicato reclutare un numero sufficiente di volontari con il genotipo d'interesse da cui isolare cellule B primarie; l'allele C è raro nella popolazione sarda, con una frequenza del genotipo CC di circa 2%. Inoltre, le cellule B sono poco rappresentate nel sangue periferico (5-10% dei PBMC totali), e l'RNA estratto da un semplice prelievo di sangue è insufficiente ai fini dell'esecuzione degli esperimenti di RT-qPCR. Sulle LCL di 10 individui con genotipo TT e 10 con genotipo CC abbiamo quindi valutato l'espressione dell'isoforma 001 di *CBL-B* con un saggio TaqMan®. Le cellule sono state analizzate solo in condizioni basali, in quanto in queste linee cellulari diversi pathways immuno-correlati risultano già attivati in seguito alla sola infezione con il EBV. Anche in questo caso non è stata osservata alcuna differenza genotipo-specifica di espressione per l'isoforma 001 in esame (Fig. 18).

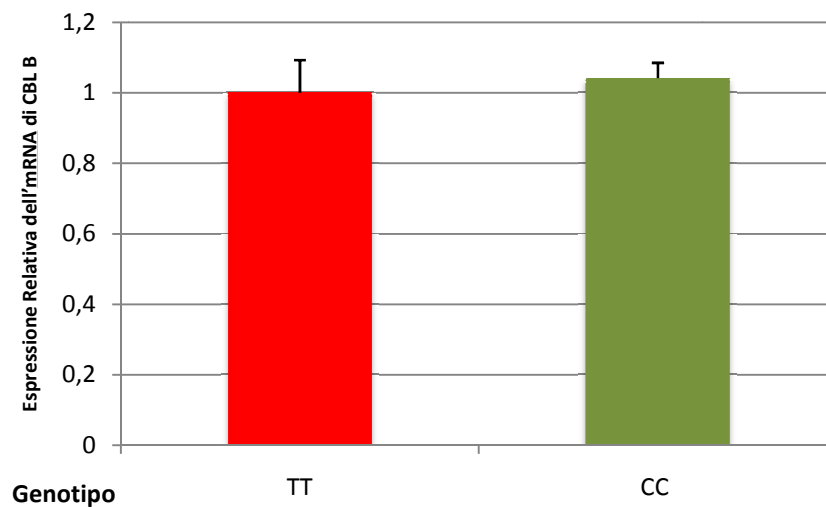


Fig. 18: Espressione relativa dell' mRNA dell' isoforma 001 di *CBL-B* in linee linfoblastoidi (LCL) di 10 individui con genotipo di predisposizione TT (in rosso) e 10 individui con genotipo protettivo CC (in verde) in condizioni basali.

5.3 Studio dell'attivazione di cellule T mediante saggi ELISA

Per valutare differenze genotipo-specifiche nell'attivazione dei linfociti T, abbiamo analizzato la produzione di IL-2 tramite saggi ELISA in 10 individui con genotipo di predisposizione TT e 10 individui con genotipo protettivo CC, sia in condizioni basali che in seguito a stimolazione del TCR con anti-CD3 e CD28 a 20 e 48 ore (vedi metodi). La stimolazione del TCR (CD3) e del suo co-recettore CD28, risulta infatti nell'attivazione dei linfociti T, con conseguente produzione di IL-2, che raggiunge un picco nell'arco di 20-48 ore dallo stimolo.

Come illustrato in Fig.19, questi esperimenti hanno messo in evidenza una consistente maggiore produzione di IL-2 nei linfociti di individui con genotipo predisponente TT per i tempi di induzione analizzati. Questa differenza, tuttavia, non è risultata significativa al T-test. Nei nostri esperimenti è stata riscontrata una marcata variabilità inter-individuale nella produzione di IL-2 in seguito a stimolazione, e questa, in un campione ridotto come quello studiato, potrebbe essere alla base della mancata significatività statistica osservata.

Questa ipotesi è supportata dai risultati ottenuti con le cellule allo stato basale (T0), invece, dove non riscontrando alcuna variabilità inter individuale nei livelli di IL-2 prodotti, la differenza tra linfociti TT e CC è risultata statisticamente significativa, con una produzione maggiore nei TT rispetto ai CC (p -value=0.02).

Questa maggiore espressione di IL-2 allo stato basale in linfociti T con genotipo TT è coerente con una minore espressione del trascritto 001 del *CBL-B* osservata in queste cellule, e quindi con l'esistenza di una iperattivazione del pathway già allo stato basale. L'ampliamento del campione e di conseguenza del potere statistico del saggio potrà aiutare a confermare o meno questa ipotesi.

Il pathway che porta alla produzione di IL-2 a valle dell'attivazione del TCR è complesso ed influenzato da numerosi fattori, genetici e non, soggetti a variabilità. E' quindi plausibile che per mettere in evidenza differenze significative nella produzione di IL-2 correlate a varianti che modificano solo finemente l'espressione genica, sia necessario analizzare un ampio numero di campioni, possibilmente in parallelo, in modo da diminuire i fattori confondenti dovuti alla manipolazione sperimentale delle cellule.

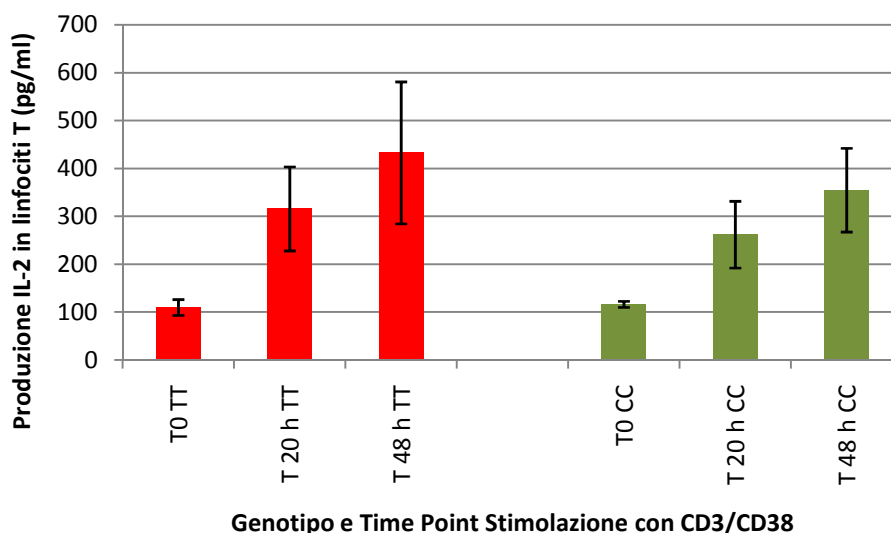


Fig 19: Produzione IL-2 (pg/ml) in 10 individui con genotipo TT (in rosso) e 10 individui con genotipo CC (in verde) in condizioni basali e in seguito a stimolazione con CD3/CD28 per 20 e 48 ore.

5.4 Studio dell'attivazione di cellule B mediante saggi ELISA

Gli esperimenti ELISA effettuati sulle cellule B sono stati indirizzati ad identificare differenze genotipo specifiche nell'attivazione di NF- κ B in individui con genotipo di rischio TT e protettivo CC per la variante rs 9657904.

Nella fase iniziale di studio si è valutato se in queste cellule fosse più efficace, ai fini di attivare di NF- κ B, una stimolazione T-dipendente (CD40L + IL-4) o T-indipendente (anti-IgM (Fab)₂ + IL-4). Le stimolazioni sono state effettuate per diversi tempi di induzione (0, 15', 30', 60') e hanno mostrato che il CD40L era in grado di stimolare le cellule in maniera più efficace rispetto all'altro (dati non mostrati). Inoltre, questo induttore ha una cinetica di attivazione più veloce dell'anti-IgM, come atteso dal fatto che la traslocazione nucleare di NF- κ B avviene molto rapidamente in seguito alla dalla stimolazione e raggiunge un plateau intorno ai 30 min, per poi tornare a livelli basali.

Gli esperimenti di attivazione sono stati condotti su 5 individui rispettivamente con genotipo TT e CC. Vista la difficoltà di reperire un numero sufficiente di cellule B per portare avanti in parallelo sia gli esperimenti di espressione che di attivazione, abbiamo potuto valutare per tutti gli individui studiati solo i livelli basali e l'induzione a 30'. Inoltre, visto il numero scarso di cellule disponibili, la concentrazione delle proteine negli estratti nucleari ottenuti dalle cellule in coltura non sono stati sufficienti per condurre gli esperimenti ELISA in tutti i campioni né tantomeno ci hanno permesso di effettuare duplicati biologici. Come mostrato nella figura 20,

abbiamo potuto analizzare solo due individui TT e CC rispettivamente, senza mettere in evidenza differenze significative nell'attivazione di NF- κ B.

Questi risultati non sono conclusivi considerato il numero di campioni studiato.

Attualmente è in corso sia un ampliamento del numero di campioni su cui ripetere l'esperimento in linee primarie che un'analisi su linee LCL più facilmente reperibili.

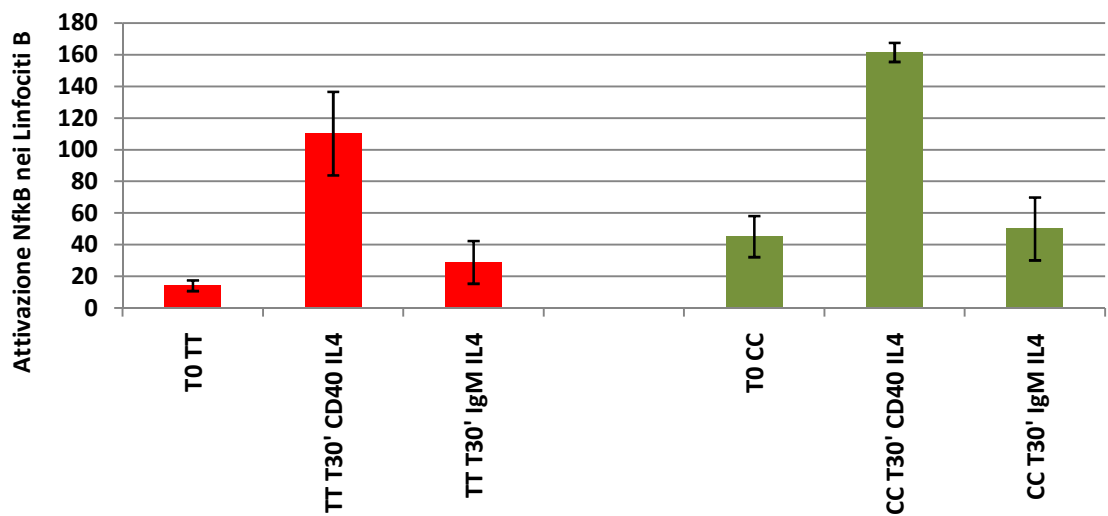


Fig 20: Attivazione di NF κ B 2 individui con genotipo TT e 2 individui con genotipo CC in condizioni basali e in seguito a stimolazione con CD40L+ IL4 e IgM+IL4.

6 DISCUSSIONE

La sclerosi multipla è una patologia invalidante ed incurabile, di cui ancora non si conoscono i meccanismi scatenanti. Pertanto, riuscire a comprendere le basi eziopatogenetiche è uno dei più importanti traguardi per la comunità scientifica mondiale, e per la Sardegna in particolare, in quanto nell'isola si osserva una delle più alte incidenze di SM del mondo (7).

Nel corso degli ultimi 40 anni sono stati condotti numerosi studi per comprendere le dinamiche molecolari e ambientali alla base di questa patologia. I maggiori avanzamenti nel campo sono venuti recentemente dalla genetica, che in seguito al grande progresso nella conoscenza della variabilità umana e negli studi di associazione su tutto il genoma, ha permesso di identificare numerosi geni di suscettibilità alla malattia. Questo passaggio è fondamentale per svelare i meccanismi ed i pathways coinvolti nell'insorgenza dei danni al SNC nelle persone affette da SM, al fine di comprendere anche i fondamenti per la protezione dello stesso e di identificare nuovi bersagli per lo sviluppo di nuove ed efficaci terapie. Tuttavia, ad oggi il quadro non è ancora completo, e solo circa il 28% del rischio di ricorrenza familiare è spiegato dalle oltre 100 varianti di suscettibilità identificate finora.

In questo contesto il nostro progetto si propone di contribuire a chiarire le cause della malattia, studiando le basi genetiche della SM nella popolazione sarda ad alto rischio. Il genoma della popolazione fondatrice sarda potrebbe contenere, infatti, varianti di predisposizione specifiche (come già noto per alcuni aplotipi dell'HLA) o presentare varianti rare di rischio con una frequenza più alta che in altre popolazioni. Queste varianti non sono rappresentate nei GeneChip arrays di genotipizzazione e nei pannelli disponibili generalmente usati per l'inferenza e pertanto non possono essere testate per associazione.

Recentemente il nostro gruppo, attraverso uno studio caso-controllo GWAS nella popolazione sarda, ha identificato varianti associate alla SM nel gene *CBL-B*. Questo gene al momento della pubblicazione non era stato ancora riportato da altri gruppi, mentre studi successivi di replica (39) e studi GWAS su ampie coorti di casi e controlli europei (37,38), ne hanno poi confermato l'associazione e quindi il ruolo cruciale in questa patologia nella popolazione generale.

Per chiarire il ruolo di varianti di *CBL-B* nella SM, in questo progetto di tesi abbiamo condotto un nuovo GWAS su un numero maggiore di campioni genotipizzati, utilizzando per l'associazione una mappa genetica ad ancora più alta risoluzione (Sanna et al, dati non pubblicati), basata su varianti derivanti dall'analisi delle sequenze di 2120 individui di origine sarda e delle sequenze di oltre 1000 individui provenienti da tutto il mondo (pannello 1000G).

Oltre ai risultati generali del GWAS, che non sono oggetto di questa tesi, queste analisi hanno consentito di effettuare un mappaggio fine e di rifinire ulteriormente l'associazione nella regione del *CBL-B*, dimostrando che lo SNP precedentemente riportato, rs9657904, insieme ad una nuova variante di tipo INDEL (rs147983726) con un elevatissimo grado di correlazione genetica ($r^2=0.98$) derivante dal pannello 1000G mondiale, sono le varianti più associate con la SM nel locus.

Queste due varianti sono geneticamente indistinguibili e, visto la mappa dettagliata utilizzata, rappresentano entrambe varianti potenzialmente causative. Nessuna delle due presenta a livello di annotazione funzionale (sui data base pubblici come UCSC, Ensemble, Encode) delle caratteristiche che possano suggerire quale sia primariamente responsabile dell'effetto biologico alla base dell'associazione. La variante maggiormente associata si trova nell'introne 1, mentre l'INDEL si trova a 15 kb a monte del gene, in regioni che sono attive trascrizionalmente e che potrebbero quindi modulare l'espressione di *CBL-B* in maniera genotipo-specifica.

Questa ipotesi è stata parzialmente confermata dagli studi funzionali mirati ad identificare differenze genotipo-specifiche nell'espressione delle varie isoforme di *CBL-B*. Questi esperimenti hanno, infatti, messo in evidenza livelli di espressione moderatamente, ma significativamente, ridotti per l'isoforma principale (001) di *CBL-B* in cellule T di individui sani con genotipo predisponente (TT) della variante rs9657904 e presenza della delezione di 4 basi della variante correlata INDEL. Queste differenze sono state osservate sia in condizioni basali che dopo stimolazione specifica del TCR. E' quindi presumibile che una diminuzione, anche lieve, nell'espressione del regolatore negativo centrale *CBL-B* e della sua proteina possa generare nel linfocita T uno stato di iperattivazione e proliferazione anche in assenza di uno stimolo antigenico specifico. Questo a sua volta, come sappiamo avvenire nel topo *Cbl-b*^{-/-}, potrebbe portare allo sviluppo di autoimmunità. Inoltre, ridotti livelli di espressione di *CBL-B* nei linfociti T periferici sono stati osservati anche in pazienti affetti da SM e correlati con il numero di episodi clinici e la fase della malattia.

Per accertare l'effetto delle varianti sull'attivazione dei linfociti T, abbiamo valutato la produzione di IL-2 in cellule con genotipo predisponente e protettivo delle stesse. Questi esperimenti, se da una parte hanno mostrato una maggiore produzione di IL-2 negli individui con genotipo di rischio in seguito a stimolazione del TCR, i risultati non sono risultati statisticamente significativi (se non nei linfociti allo stato basale), mettendo in evidenza alcune

problematiche di variabilità inter-individuale e l'esigenza di ampliare il campione per poter ottenere un dato conclusivo dal punto di vista statistico.

Sebbene il fenotipo del topo knock-out per *Cbl-b* abbia mostrato un'iperattivazione anche delle cellule B di questi animali, le nostre analisi non hanno messo in evidenza espressione differenziale genotipo-dipendente di alcuno dei trascritti studiati, sia in cellule B primarie che in un campione più numeroso di linee linfoblastoidi. Inoltre, non siamo riusciti a rilevare differenze nell'attivazione dei linfociti B a valle delle varianti studiate, attraverso la misurazione della translocazione nucleare di NF- κ B; questi ultimi esperimenti sono risultati particolarmente complicati da eseguire e frammentari, e devono essere ripetuti in un campione più ampio per poter essere conclusivi.

Complessivamente possiamo dire che questo progetto ha due punti principali di forza:

- 1) da una parte l'analisi genetica dettagliata che ha consentito di definire nel locus *CBL-B* le due varianti responsabili dell'associazione e potenzialmente causative da studiare funzionalmente;
- 2) dall'altra la possibilità di condurre studi funzionali in cellule primarie di individui richiamati per genotipo delle varianti d'interesse.

Queste caratteristiche potranno consentire di completare la caratterizzazione funzionale di queste varianti con esperimenti molecolari mirati a separarne gli effetti, ed a capire più precisamente in quale tipo cellulare *CBL-B* svolga la sua funzione predisponente alla SM.

7 PROBLEMATICHE AFFRONTATE NELLO STUDIO

In questo progetto sono state affrontate numerose problematiche relative sia allo studio genetico che a quello funzionale, alla maggior parte delle quali si è data una soluzione rapida ed efficace.

Tra quelle che hanno in modo particolare reso complicato svolgere gli esperimenti funzionali, una importante problematica è stata quella di reperire un numero sufficiente di cellule primarie T e B su cui effettuare parallelamente sia gli esperimenti di espressione che quelli di attivazione delle cellule d'interesse. Grazie alla collaborazione con il centro trasfusionale dell'ospedale Brotzu di Cagliari, che ci ha permesso di prelevare sacche di buffy coat di volontari che avevano dato il loro consenso alla partecipazione al progetto, abbiamo potuto ottenere un numero di linfociti T sufficiente per poter svolgere gli esperimenti pianificati, mentre per le cellule B, che sono solo il 5-10% dei leucociti circolanti totali, non tutti gli esperimenti si sono potuti effettuare. Un'ulteriore difficoltà è stata quella del reclutamento di un numero adeguato di individui con genotipo protettivo CC; questi si presentano, infatti, con una frequenza del solo 2% nella popolazione generale ed anche tra i donatori non siamo riusciti ad identificarne più di 10 disponibili al prelievo. Inoltre, nonostante la disponibilità della maggior parte dei donatori, ripetere gli esperimenti sullo stesso individuo non è stato semplice, in quanto questi devono attendere almeno tre mesi prima di sottoporsi ad un successivo prelievo (donazione di sangue).

Gli esperimenti di espressione con RT-qPCR hanno anche essi presentato numerose difficoltà, in quanto la discriminazione delle diverse isoforme del gene *CBL-B* è stata tutt'altro che semplice. Le varie isoforme presentano numerose regioni omologhe, rendendo difficile l'amplificazione univoca. Un attento studio e un'analisi critica degli oligonucleotidi utilizzati ci ha permesso però seppure con molteplici passaggi di discriminare coerentemente le isoforme analizzate.

8 CONCLUSIONI E SVILUPPI FUTURI

Gli studi di associazione sull'intero genoma hanno, in pochi anni, consentito l'identificazione di numerose varianti geniche legate a numerosi tratti e fenotipi complessi.

Oltre 120 varianti sono state descritte associate alla SM, con soglie di significatività genome wide ($p > 5 \times 10^{-8}$), tra cui l'associazione del gene *CBL-B* da parte del nostro gruppo di ricerca, successivamente replicata in altre popolazioni. Nel loro insieme i risultati ottenuti dal nostro e da altri gruppi definiscono in maniera inequivocabile la natura immunologica, autoimmune della SM, in quanto una larga parte dei geni associati con la malattia sono coinvolti nella regolazione delle risposte immuni e coincidono con loci associati con altre patologie autoimmuni. Questi studi definiscono anche il modello genetico della componente ereditabile della malattia, con una regione genica, la regione HLA, con effetti più rilevanti sul rischio di sviluppare la malattia e altre varianti coinvolte, al di fuori della regione HLA, con effetti genetici più modesti e con una larga fetta di ereditabilità, molto verosimilmente legata a varianti rare e molto rare, ancora da spiegare.

Per identificare le varianti rare potenzialmente associate alla SM, o ad altre malattie e tratti complessi, vengono utilizzati per gli studi GWAS di nuova generazione pannelli di referenza sempre più dettagliati basati sul sequenziamento a bassa copertura di migliaia di individui, in ampie casistiche ben caratterizzate. Questo approccio ci ha permesso di individuare oltre a varianti di tipo SNP anche un INDEL associato a questa patologia nella popolazione sarda. E' molto probabile che una delle due varianti strettamente correlate riportate in questo studio sia quella primariamente responsabile dell'effetto biologico alla base dell'associazione. Dal punto di vista genetico, l'analisi condotta ha raggiunto il livello di dettaglio più alto considerando le conoscenze attuali della variabilità genetica.

Gli studi funzionali effettuati in questa tesi hanno messo in evidenza una differenza di espressione genotipo-specifica dell'isoforma principale di *CBL-B* nelle cellule T, con un livelli ridotti del trascritto negli individui con genotipo predisponente. Questo importante risultato suggerisce un ruolo diretto di una delle varianti nella regolazione della trascrizione di *CBL-B*, il cui meccanismo rimane però da chiarire dal punto di vista molecolare. Inoltre, il nostro studio non ha potuto mettere in evidenza in quale sottopopolazione specifica di cellule T CD3+ (ad esempio CD8+, CD4+, T regolatorie) la variante abbia primariamente un effetto. Abbiamo infatti visto che *Cbl-b* gioca un ruolo cruciale nello sviluppo delle cellule T regolatore inducibili nella periferia e nell'espressione del fattore di trascrizione Foxp3, il quale è essenziale per lo

sviluppo e la funzione delle cellule T regolatorie (45), che a loro volta sono critiche per lo sviluppo di autoimmunità. In futuro ci proponiamo specificamente di studiare i livelli di espressione delle isoforme di *CBL-B* sia in cellule CD4+ CD25- (T effettrici) che in altre sottopopolazioni di cellule T regolatorie note (52).

Dal momento che le cellule T regolatorie rappresentano solo il 5% delle cellule T CD4+ circolanti, per isolare un numero di cellule sufficiente per condurre gli esperimenti, per queste come per altre cellule poco rappresentate nel sangue periferico, questi esperimenti verranno condotti utilizzando cellule separate con un cell sorter (FACS ARIA, Becton Dickinson) a breve disponibile nel nostro Istituto.

In particolare, pur non essendo emersa alcuna differenza genotipo-specifica significativa nell'espressione di *CBL-B* nelle cellule B, visto il ruolo cruciale di queste cellule che sta emergendo nell'autoimmunità e nella SM, abbiamo anche pianificato studi più dettagliati nelle varie sottopopolazioni B.

Per meglio comprendere l'effetto delle singole varianti rs9657904 e rs147983726, le regioni genomiche in cui sono posizionate rispettivamente saranno clonate all'interno di un vettore di espressione contenente un minimal promoter (SV40) a monte del gene reporter luciferase. I costrutti contenenti i genotipi protettivo e predisponente delle due varianti verranno usati per transfettare cellule eucariotiche e valutare il loro effetto regolatorio sulla trascrizione genica.

L'analisi del trascrittoma consentirà di esaminare, inoltre, da un punto di vista qualitativo e quantitativo i profili di espressione e correlarli con i profili genetici (eQTL) e con i tratti fenotipici esaminati, fornendo ulteriori informazioni sulle conseguenze funzionali delle associazioni genetiche. Tale metodo consentirà anche una migliore estrazione delle varianti rare, e permetterà, altresì, di valutare variazioni del numero di copie, come inserzioni, delezioni ma anche inversioni e traslocazioni, per comprenderne il loro eventuale ruolo nell'eziologia della SM e di altre patologie autoimmuni comuni in Sardegna.

Nel loro insieme questi sviluppi futuri delucideranno i meccanismi cellulari in cui *CBL-B* è coinvolto, e consentiranno una migliore comprensione di come una singola variazione nel DNA può modificare i meccanismi in senso predisponente nei confronti della SM. Infine, ciò permetterà di evidenziare potenziali bersagli terapeutici e aiuterà i ricercatori impegnati nella ricerca farmacologica pre-clinica, che potranno, attraverso un approccio basato sulla conoscenza, selezionare composti utili per la generazione di nuovi e più efficaci farmaci.

9 BIBLIOGRAFIA

1. McFarlin DE, McFarland HF Multiple Sclerosis N Engl J Med. 1982 Nov 11;307(20):1246-51.
2. Sospedra M, Martin R. Immunology of multiple sclerosis Annu Rev Immunol. 2005;23:683-747. Review.
3. David A. Hafler, Multiple sclerosis J Clin Invest. 2004 Mar;113(6):788-94.
4. Hohol MJ, Olek MJ, Orav EJ, Stazzone L, Hafler DA, Khoury SJ, Dawson DM, Weiner HL. Treatment of progressive multiple sclerosis with pulse cyclophosphamide/methylprednisolone: response to therapy is linked to the duration of progressive disease. Mult Scler. 1999 Dec;5(6):403-9.
5. Steinman L. et al. Multiple sclerosis: a two stage disease. Nature Immunol. 2, 762–765 2001
6. Granieri E, Casetta I, Govoni V, Tola MR, Marchi D, Murgia SB, Ticca A, Pugliatti M, Murgia B, Rosati G The increasing incidence and prevalence of MS in a Sardinian province Neurology. 2000 Sep 26;55(6):842-8.
7. Pugliatti M, Rosati G, Carton H, Riise T, Drulovic J, Vécsei L, Milanov I. The epidemiology of multiple sclerosis in Europe. Eur J Neurol. 2006 Jul;13(7):700-22.
8. The Correlation between relatives on the supposition of mendelian inheritance. Transactions of the Royal Society of Edinburgh 52:399-433 (1918)
9. Sadovnick AD, Baird PA. The familial nature of multiple sclerosis: age-corrected empiric recurrence risks for children and siblings of patients. Neurology. 1988 Jun;38(6):990-1.
10. Mumford CJ, Wood NW, Kellar-Wood H, Thorpe JW, Miller DH, Compston DA. The British Isles survey of multiple sclerosis in twins. Neurology. 1994 Jan;44(1):11-5.
11. Ebers GC, Sadovnick AD, Risch NJ. A genetic basis for familial aggregation in multiple sclerosis. Canadian Collaborative Study Group. Nature. 1995 Sep 14;377(6545):150-1.
12. Robertson NP, O'Riordan JI, Chataway J, Kingsley DP, Miller DH, Clayton D, Compston DA. Offspring recurrence rates and clinical characteristics of conjugal multiple sclerosis. Lancet. 1997 May 31;349(9065):1587-90.; Ebers et al 2000
13. Ebers GC, Sadovnick AD, Risch NJ. A genetic basis for familial aggregation in multiple sclerosis. Canadian Collaborative Study Group. Nature. 1995 Sep 14;377(6545):150-1.
14. Sospedra, M., and R. Martin. Immunology of multiple sclerosis. Annu Rev Immunol 23:683-747 2005.

15. Christensen T, Petersen T, Thiel S, Brudek T, Ellermann-Eriksen S, Møller-Larsen A. Gene-environment interactions in multiple sclerosis: innate and adaptive immune responses to human endogenous retrovirus and herpesvirus antigens and the lectin complement activation pathway. *J Neuroimmunol.* Feb;183(1-2):175-88 (2007)
16. Smolders J, Damoiseaux J, Menheere P, Hupperts R. Vitamin D as an immune modulator in multiple sclerosis, a review. *J Neuroimmunol.* Feb;194(1-2):7-17 (2008)
17. Compston DA, Batchelor JR, McDonald WI. B-lymphocyte alloantigens associated with multiple sclerosis. *Lancet.* 1976 Dec 11;2(7998):1261-5
18. Hillert, J. and Olerup, O. Multiple sclerosis is associated with genes within or close to the HLA-DR-DQ subregion on a normal DR15, DQ6, Dw2 haplotype. *Neurology*, 43, 163-168(1993).
19. Marrosu MG, Muntoni F, Murru MR, Costa G, Pischedda MP, Pirastu M, Sotgiu S, Rosati G, Cianchetti C. HLA-DQB1 genotype in Sardinian multiple sclerosis: evidence for a key role of DQB1 *0201 and *0302 alleles. *Neurology.* 1992 Apr;42(4):883-6.
20. Marrosu MG, Murru MR, Costa G, Cucca F, Sotgiu S, Rosati G, Muntoni F. Multiple sclerosis in Sardinia is associated and in linkage disequilibrium with HLA-DR3 and -DR4 alleles. *Am J Hum Genet.* 1997 Aug;61(2):454-7.
21. Marrosu MG, Murru MR, Costa G, Murru R, Muntoni F, Cucca F. DRB1-DQA1-DQB1 loci and multiple sclerosis predisposition in the Sardinian population. *Hum Mol Genet.* 1998 Aug;7(8):1235-7.
22. Marrosu MG, Murru R, Murru MR, Costa G, Zavattari P, Whalen M, Cocco E, Mancosu C, Schirru L, Solla E, Fadda E, Melis C, Porru I, Rolesu M, Cucca F Dissection of the HLA association with multiple sclerosis in the founder isolated population of Sardinia. *Hum Mol Genet.* 2001 Dec 1;10(25):2907-16.
23. I. Cogato, E. Montanari *La Sclerosi Multipla* (1996)
24. A. Compston, I McDonald, J Noseworthy, H Lassmann, D Miller, C Confavreux, K Smith, *McAlpine Multiple Sclerosis* (2005)
25. Sawcer S, Jones HB, Feakes R, Gray J, Smaldon N, Chataway J, Robertson N, Clayton D, Goodfellow PN, Compston A. A genome screen in multiple sclerosis reveals susceptibility loci on chromosome 6p21 and 17q22. *Nat Genet.* 1996 Aug;13(4):464-8.
26. The Games Collaborative Group, Linkage disequilibrium screening for multiple sclerosis implicates JAG1 and POU2AF1 as susceptibility genes in Europeans. *J Neuroimmunol.*

- 2006 Oct;179(1-2):108-16. Epub 2006 Aug 24. Erratum in: J Neuroimmunol. 2007 Sep;189(1-2):175-6. multiple author names added.
27. Henry F McFarland and Roland Martin. Multiple sclerosis: a complicated picture of autoimmunity. *Nature Immunology* 8, 913 - 919 (2007)
 28. Disanto G, Morahan JM, Barnett MH, Giovannoni G, Ramagopalan SV. The evidence for a role of B cells in multiple sclerosis. *Neurology*. 2012 Mar 13;78(11):823-32.
 29. International Multiple Sclerosis Genetics Consortium, Hafler DA, Compston A, Sawcer S, Lander ES, Daly MJ, De Jager PL, de Bakker PI, Gabriel SB, Mirel DB, Iverson AJ, Pericak-Vance MA, Gregory SG, Rioux JD, McCauley JL, Haines JL, Barcellos LF, Cree B, Oksenberg JR, Hauser SL. Risk alleles for multiple sclerosis identified by a genomewide study. *N Engl J Med*. 2007 Aug 30;357(9):851-62.
 30. Vigiotta V, Baecher-Allan C, Weiner HL, Hafler DA. Loss of functional suppression by CD4+CD25+ regulatory T cells in patients with multiple sclerosis. *J Exp Med*. 2004 Apr 5;199(7):971-9.
 31. Britt A. Johnson, Joanne Wang, Elise M. Taylor, Stacy J. Caillier, Joseph Herbert, Omar A. Khan, Anne H. Cross, Philip L. De Jager, Pierre-Antoine F. Gourraud, Bruce C.A. Cree, Stephen L. Hauser, Jorge R. Oksenberg, Multiple sclerosis susceptibility alleles in African Americans *Genes Immun*. 2010 June; 11(4): 343–350.
 32. Jakkula E, Leppä V, Sulonen AM, Varilo T, Kallio S, Kempainen A, Purcell S, Koivisto K, Tienari P, Sumelahti ML, Elovaara I, Pirttilä T, Reunanen M, Aromaa A, Oturai AB, Söndergaard HB, Harbo HF, Mero IL, Gabriel SB, Mirel DB, Hauser SL, Kappos L, Polman C, De Jager PL, Hafler DA, Daly MJ, Palotie A, Saarela J, Peltonen L. Genome-wide association study in a high-risk isolate for multiple sclerosis reveals associated variants in STAT3 gene. *Am J Hum Genet*. 2010 Feb 12;86(2):285-91. doi: 10.1016/j.ajhg.2010.01.017.,
 33. Sanna S, Pitzalis M, Zoledziewska M, Zara I, Sidore C, Murru R, Whalen MB, Busonero F, Maschio A, Costa G, Melis MC, Deidda F, Poddie F, Morelli L, Farina G, Li Y, Dei M, Lai S, Mulas A, Cuccuru G, Porcu E, Liang L, Zavattari P, Moi L, Deriu E, Urru MF, Bajorek M, Satta MA, Cocco E, Ferrigno P, Sotgiu S, Pugliatti M, Traccis S, Angius A, Melis M, Rosati G, Abecasis GR, Uda M, Marrosu MG, Schlessinger D, Cucca F. Variants within the immunoregulatory CBLB gene are associated with multiple sclerosis. *Nat Genet*. 2010 Jun;42(6):495-7. doi: 10.1038/ng.584. Epub 2010 May 9

34. Nischwitz S, Cepok S, Kroner A, Wolf C, Knop M, Müller-Sarnowski F, Pfister H, Roeske D, Rieckmann P, Hemmer B, Ising M, Uhr M, Bettecken T, Holsboer F, Müller-Myhsok B, Weber F. Evidence for VAV2 and ZNF433 as susceptibility genes for multiple sclerosis. *J Neuroimmunol.* 2010 Oct 8;227(1-2):162-6. doi: 10.1016/j.jneuroim.2010.06.003.
35. Briggs FB, Shao X, Goldstein BA, Oksenberg JR, Barcellos LF, de Jager PL. (IMSGC) Genome-wide association study of severity in multiple sclerosis. *Genes Immun.* 2011 Dec;12(8):615-25. doi: 10.1038/gene.2011.34.
36. De Jager PL, Jia X, Wang J, de Bakker PI, Ottoboni L, Aggarwal NT, Piccio L, Raychaudhuri S, Tran D, Aubin C, Briskin R, Romano S; International MS Genetics Consortium, Baranzini SE, McCauley JL, Pericak-Vance MA, Haines JL, Gibson RA, Naeglin Y, Uitdehaag B, Matthews PM, Kappos L, Polman C, McArdle WL, Strachan DP, Evans D, Cross AH, Daly MJ, Compston A, Sawcer SJ, Weiner HL, Hauser SL, Hafler DA, Oksenberg JR. Meta-analysis of genome scans and replication identify CD6, IRF8 and TNFRSF1A as new multiple sclerosis susceptibility loci. *Nat Genet.* 2009 Jul;41(7):776-82. doi: 10.1038/ng.401.
37. IMSGC Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. *Nature.* 2011 Aug 10;476(7359):214-9. doi: 10.1038/nature10251.
38. IMSGC Analysis of immune-related loci identifies 48 new susceptibility variants for multiple sclerosis. *Nat Genet.* 2013 Nov;45(11):1353-60. doi: 10.1038/ng.2770.
39. Corrado L, Bergamaschi L, Barizzone N, Fasano ME, Guerini FR, Salvetti M, Galimberti D, Benedetti MD, Leone M, D'Alfonso S. Association of the CBLB gene with multiple sclerosis: new evidence from a replication study in an Italian population. *J Med Genet.* 2011 Mar;48(3):210-1. doi: 10.1136/jmg.2010.081380.
40. Cucca F, Lampis R, Frau F, Macis D, Angius E, Masile P, Chessa M, Frongia P, Silveti M, Cao A, De Virgiliis S, Congia M. The distribution of DR4 haplotypes in Sardinia suggests a primary association of type I diabetes with DRB1 and DQB1 loci. *Hum Immunol.* 1995 Aug;43(4):301-8.
41. Willer CJ, Dymant DA, Risch NJ, Sadovnick AD, Ebers GC; Canadian Collaborative Study Group. Twin concordance and sibling recurrence rates in multiple sclerosis. *Proc Natl Acad Sci U S A.* 2003 Oct 28;100(22):12877-82.
42. Thrower BW. Clinically isolated syndromes: predicting and delaying multiple sclerosis. *Neurology.* 2007 Jun 12;68(24 Suppl 4):S12-5. Review.

43. Bachmaier K, Krawczyk C, Kozieradzki I, Kong YY, Sasaki T, Oliveira-dos-Santos A, Mariathasan S, Bouchard D, Wakeham A, Itie A, Le J, Ohashi PS, Sarosi I, Nishina H, Lipkowitz S, Penninger JM. Negative regulation of lymphocyte activation and autoimmunity by the molecular adaptor Cbl-b. *Nature*. 2000 Jan 13;403(6766):211-6.
44. Chiang YJ, Kole HK, Brown K, Naramura M, Fukuhara S, Hu RJ, Jang IK, Gutkind JS, Shevach E, Gu H. Cbl-b regulates the CD28 dependence of T-cell activation. *Nature*. 2000 Jan 13;403(6766):216-20.
45. Hori S, Takahashi T, Sakaguchi S. Control of autoimmunity by naturally arising regulatory CD4+ T cells. *Adv Immunol*. 2003;81:331-71. Review.
46. Qiao G, Lei M, Li Z, Sun Y, Minto A, Fu YX, Ying H, Quigg RJ, Zhang J. Negative regulation of CD40-mediated B cell responses by E3 ubiquitin ligase Casitas-B-lineage lymphoma protein-B. *J Immunol*. 2007 Oct 1;179(7):4473-9.
47. Yasuda T, Tezuka T, Maeda A, Inazu T, Yamanashi Y, Gu H, Kurosaki T, Yamamoto T. Cbl-b positively regulates Btk-mediated activation of phospholipase C-gamma2 in B cells. *J Exp Med*. 2002 Jul 1;196(1):51-63.
48. Schwartz RH. T cell anergy. *Annu Rev Immunol*. 2003;21:305-34. Epub 2001 Dec 19. Review.
49. Zhou WB, Wang R, Deng YN, Ji XB, Huang GX, Xu YZ. Study of Cbl-b dynamics in peripheral blood lymphocytes isolated from patients with multiple sclerosis. *Neurosci Lett*. 2008 Aug 8;440(3):336-9. doi: 10.1016/j.neulet.2008.05.089. Epub 2008 Jun 17.
50. Quail MA, Kozarewa I, Smith F, Scally A, Stephens PJ, Durbin R, Swerdlow H, Turner DJ. A large genome center's improvements to the Illumina sequencing system. *Nat Methods*. 2008 Dec;5(12):1005-10. doi: 10.1038/nmeth.1270
51. Pilia G, Chen WM, Scuteri A, Orrù M, Albai G, Dei M, Lai S, Usala G, Lai M, Loi P, Mameli C, Vacca L, Deiana M, Olla N, Masala M, Cao A, Najjar SS, Terracciano A, Nedorezov T, Sharov A, Zonderman AB, Abecasis GR, Costa P, Lakatta E, Schlessinger D. Heritability of cardiovascular and personality traits in 6,148 Sardinians. *PLoS Genet*. 2006 Aug 25;2(8):e132. Epub 2006 Jul 10.
52. Onishi Y, Fehervari Z, Yamaguchi T, Sakaguchi S. Foxp3+ natural regulatory T cells preferentially form aggregates on dendritic cells in vitro and actively inhibit their maturation. *Proc Natl Acad Sci U S A*. 2008 Jul 22;105(29):10113-8. doi: 10.1073/pnas.0711106105.

53. Kozarewa I, Ning Z, Quail MA, Sanders MJ, Berriman M, Turner DJ. Amplification-free Illumina sequencing-library preparation facilitates improved mapping and assembly of (GpC)-biased genomes. *Nat Methods*. 2009 Apr;6(4):291-5.
54. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R; 1000 Genome Project Data Processing Subgroup. The sequence alignment/map format and SAMtools. *Bioinformatics*. 2009 Aug 15;25(16):2078-9. doi: 10.1093/bioinformatics/btp352.
55. Li Y, Willer C, Sanna S, Abecasis G. *Ann. Rev. Genomics Hum. Genetics; Annu Rev Genomics Hum Genet*. 2009;10:387-406.
56. Orrù V, Steri M, Sole G, Sidore C, Viridis F, Dei M, Lai S, Zoledziewska M, Busonero F, Mulas A, Floris M, Mentzen WI, Urru SA, Olla S, Marongiu M, Piras MG, Lobina M, Maschio A, Pitzalis M, Urru MF, Marcelli M, Cusano R, Deidda F, Serra V, Oppo M, Pilu R, Reinier F, Berutti R, Pireddu L, Zara I, Porcu E, Kwong A, Brennan C, Tarrier B, Lyons R, Kang HM, Uzzau S, Atzeni R, Valentini M, Firinu D, Leoni L, Rotta G, Naitza S, Angius A, Congia M, Whalen MB, Jones CM, Schlessinger D, Abecasis GR, Fiorillo E, Sanna S, Cucca F. *Cell*. 2013 Sep 26;155(1):242-56. doi: 10.1016/j.cell.2013.08.041.

10 RINGRAZIAMENTI

Desidero ringraziare il Prof Francesco Cucca, relatore di questa tesi, per la disponibilità e cortesia dimostratemi in questi tre anni di lavoro e per l'aiuto fornito durante la stesura di questa tesi.

Un sentito ringraziamento va alla dott.ssa Silvia Naitza mia co-tutor che si è mostrata sempre cordiale e disponibile nei miei confronti ed è stata una vera e propria guida nella mia formazione scientifica.

Un grazie va a tutti i miei colleghi dell' IRGB CNR di Cagliari. Con collaborazioni costanti, supporto morale e consigli hanno incentivato la mia crescita professionale e non solo. Con loro ho condiviso gioie e dolori di questo percorso lungo e incalzante. Tra loro un pensiero speciale è rivolto a Sonia e Francesca per il sostegno morale che mi hanno rivolto in questi anni passati insieme nel Laboratorio 2.

Inoltre intendo ringraziare i miei genitori ed Eleonora. Mi hanno supportato moralmente e mi hanno saputo consigliare e consolare puntualmente nei momenti più difficili aiutandomi a raggiungere questo importante traguardo.

Un ultimo ringraziamento, ma non meno importante, va ad Alessandro che in questi anni mi ha spronato e incoraggiato a dare sempre il meglio di me stessa.

Grazie infinite a tutti.....

Alessia