



UNIVERSITÀ DEGLI STUDI DI SASSARI

SCUOLA DI DOTTORATO DI RICERCA IN SCIENZE BIOMEDICHE

Direttore della Scuola: Prof. Andrea Fausto Piana

**INDIRIZZO IN GENETICA MEDICA, MALATTIE METABOLICHE E
NUTRIGENOMICA**

XXVIII CICLO

**Alla ricerca delle basi molecolari del Diabete di Tipo 1:
sequenziamento di ultima generazione in gemelli
monozigoti concordanti e discordanti per la malattia.**

Direttore:

Prof. Andrea Fausto Piana

Tutor:

Prof. Francesco Cucca

Tesi di dottorato di:

Dott. ssa Loredana Marcia

Anno Accademico 2014-2015

La presente tesi è stata prodotta nell'ambito della Scuola di Dottorato in Scienze Biomediche dell'Università degli Studi di Sassari, a.a. 2012/2013 – XXVIII ciclo, con il supporto di una borsa di studio finanziata con le risorse del P.O.R. SARDEGNA F.S.E. 2007-2013 - Obiettivo competitività regionale e occupazione, Asse IV Capitale umano, Linea di Attività 1.3.1.

INDICE

PREMESSA	4
INTRODUZIONE	5
Il “progetto diabete”	8
Studio di associazione su tutto il genoma (Genome-Wide Association Study, GWAS)	8
Studio dell’esoma su grandi famiglie”	9
SCOPO DEL PROGETTO	10
IL DIABETE DI TIPO 1 (DT1)	11
Quadro clinico	11
Prevalenza, incidenza e distribuzione geografica del DT1 nel mondo	12
Età e sesso	13
Aumento dell’incidenza e della stagionalità	13
ALTRI TIPI DI DIABETE	14
COMPONENTE GENETICA DEL DT1	14
Era pre-GWAS	14
Era post-GWAS	15
COMPONENTE AMBIENTALE DEL DT1	18
PATOFISIOLOGIA DEL DT1	19
Storia immunologica del DT1	19
Il ruolo dell’HLA	22
Il ruolo dell’insulina nell’insorgenza della malattia	24
Regolazione del sistema immunitario	25
Storia naturale della malattia	28
L’IMPORTANZA DELLO STUDIO DEI GEMELLI	29
MATERIALI E METODI	30
Descrizione del campione	30
Estrazione di cellule mononucleate da sangue periferico (PBMC)	30
Separazione cellule CD4+ e CD8+	30
Estrazione RNA, DNA e proteine da cellule CD4+ e CD8+	31
Estrazione DNA per la genotipizzazione	32
Selezione RNA polyA	32
Preparazione librerie	33
Sequenziamento con HiSeq 2000	33
Allineamento e controlli di qualità	34
Annotazione, normalizzazione e stabilizzazione della varianza	34

PCA e clusterizzazione gerarchica-“hierarchical clustering”	34
Determinazione dei geni differentemente espressi	35
RISULTATI E DISCUSSIONE	36
Analisi del profilo trascrizionale	36
Mappaggio delle sequenze e controlli di qualità	37
Clusterizzazione gerarchica	40
Analisi dell’espressione differenziale.....	41
RISULTATI PRELIMINARI E PROSPETTIVE FUTURE	45
Analisi del repertorio dei recettori delle cellule T (TCR)	45
Analisi del metiloma.....	46
CONCLUSIONI.....	47
RINGRAZIAMENTI	50
BIBLIOGRAFIA	51

PREMESSA

Il diabete di tipo 1 (DT1) è una patologia autoimmune nella quale le cellule β delle isole pancreatiche, deputate alla produzione dell'insulina, vengono selettivamente distrutte da linfociti T auto-reattivi (Borg et al., 2000).

La malattia insorge frequentemente nell'infanzia, e la prevalenza in individui di età inferiore a 15 anni varia tra 0.03% e 0.5% nella maggior parte delle popolazioni europee e nord americane (Liese et al., 2006).

In molte nazioni, per esempio in Europa, in Medio Oriente ed in Australia, l'incidenza del DT1 in ragazzi di età inferiore ai 15 anni, sta aumentando al ritmo del 2-5% all'anno (Karvonen et al., 2000).

Ad oggi, la patologia è affrontata dal punto di vista terapeutico con la somministrazione di insulina esogena, sebbene si stiano sviluppando anche trattamenti immunologici volti a contrastare la distruzione autoimmune delle cellule β .

E' una malattia multifattoriale dovuta a fattori ambientali permissivi e ancora non conosciuti- se pure ipotizzati- e alla presenza di numerosi fattori di rischio genetici. Nonostante lo studio delle componenti genetiche del DT1 (e di altre malattie a trasmissione ereditaria multifattoriale), nel corso degli ultimi anni abbia subito una rapida accelerazione soprattutto grazie all'avvento degli studi di associazione sull'intero genoma (GWAS), le dinamiche molecolari implicate nell'insorgenza e nella progressione della malattia non sono state ancora completamente chiarite.

Lo studio genetico di gemelli monozigoti (MZ) concordanti e discordanti per il DT1 che presenterò in questa tesi di Dottorato, potrebbe rappresentare, come già dimostrato per altre patologie complesse, un utile approccio scientifico per l'identificazione di fattori genetici ed epigenetici concorrenti alla malattia non ancora descritti.

INTRODUZIONE

Il diabete di tipo 1 (DT1) è una patologia cronica multifattoriale, a carattere autoimmune. Il processo autoimmune alla base della malattia dipende dalla complessa correlazione tra numerose varianti genetiche di suscettibilità distribuite lungo l'intero genoma e fattori ambientali di rischio permissivi ancora in larga parte sconosciuti (Cucca et al.,2001).

La popolazione fondatrice della Sardegna, mostra la seconda più alta incidenza del DT1 al mondo dopo la Finlandia (rivisto da Songini 2010), il che rappresenta un'importante eccezione al gradiente nord-sud descritto per l'incidenza della patologia in Europa, e nel mondo (Fig.1).

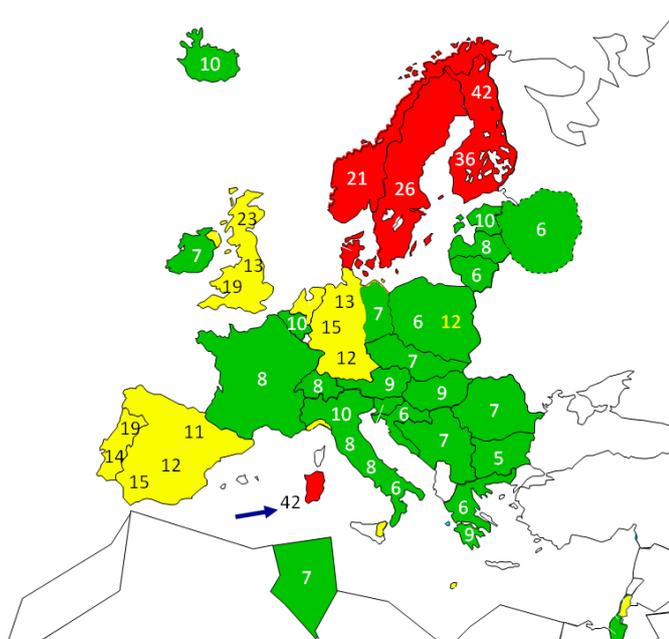


Figura 1. Incidenza del DT1 in Europa e in alcune nazioni del mediterraneo (età 0-14 anni)*. Il numero riportato nelle differenti aree geografiche si riferisce ai nuovi casi/ anno per 100.000 nati vivi. *Adattata da EURODIAB.

La Sardegna è un macro isolato genetico, la cui popolazione ha vissuto un lungo isolamento geografico, con minimo tasso di immigrazione ed elevata percentuale di matrimoni fra consanguinei. L'alto grado di parentela che caratterizza la popolazione sarda, implica che un

numero ridotto di geni di predisposizione o varianti alleliche esprima fenotipi complessi.

Pur collocandosi nell'ambito della variabilità europea, la popolazione sarda manifesta tutta una serie di caratteristiche di unicità: alcune varianti genetiche sono particolarmente frequenti in Sardegna e talvolta rare o assenti in altre popolazioni.

Il DT1 presenta una forte componente genetica. Nella popolazione generale, il rischio di sviluppare il DT1 è dello 0.03-0.5% (Liese et al., 2006), mentre è maggiore tra individui imparentati in cui è presente un soggetto affetto. Infatti, all'aumentare del grado di parentela, aumenta il rischio di sviluppare la patologia. Tra i parenti di primo grado, i fratelli presentano il rischio più elevato (5-10% all'età di 20 anni) rispetto ai figli degli individui affetti. I figli di padri diabetici hanno un rischio maggiore (circa il 12%) rispetto ai figli di madri diabetiche (circa il 6%) (Harjutsalo et al., 2006) ma non è stato ancora chiarito il perché di questa differenza. Finora esiste un solo studio che riporta come la presenza, nella madre, di auto-anticorpi trans-placentali diretti contro le isole pancreatiche, possa diminuire il rischio malattia nel figlio (Koczwara et al., 2004).

Il rischio maggiore è stato osservato tra gemelli MZ e si stima sia compreso tra il 35-42 % (Redondo et al., 2001) e, con un *follow-up* a lungo termine (>30 anni), sale oltre il 60 % (Redondo et al., 2008). Inoltre è stato dimostrato che almeno i 2/3 dei gemelli MZ discordanti, sviluppano la presenza di auto-anticorpi diretti contro le cellule β e/o il diabete (Fig.2). Il tasso di concordanza nei gemelli dizigoti (DT) è invece del 6-10 %, simile a quello descritto per fratelli non gemelli.

La non completa concordanza per la malattia tra gemelli monozigoti, aventi identico patrimonio genetico, e che spesso condividono l'ambiente in cui crescono dal concepimento fino all'esordio della malattia, supporta l'ipotesi che, oltre alla forte componente genetica, ci sia una combinazione di fattori ambientali, epigenetici e stocastici che giocano un ruolo importante nella predisposizione o nella protezione nei confronti del DT1. A supporto di ciò ci sono anche numerosi studi sui topi diabetici non-obesi –NOD–, che seppure geneticamente identici e condividendo negli stabulari lo stesso ambiente, hanno mostrato una condivisione della malattia inferiore al 100% (Makino et al., 1980; Anderson et al., 2005).

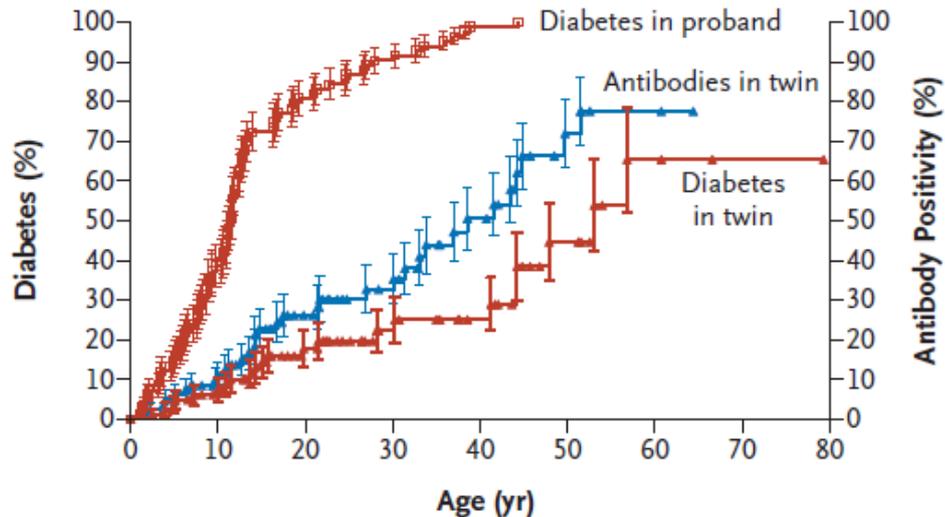


Figura 2. Sviluppo progressivo del DT1 e della presenza di autoanticorpi anti isola. Nel grafico sono mostrate l'incidenza cumulativa del DT1 e della positività verso gli autoanticorpi durante la vita di MZ che inizialmente erano discordanti per la malattia. In grafico le stime Kaplan Meier della progressione verso il diabete e della autoimmunità contro le isole pancreatiche in base all'età. L'analisi della sopravvivenza relativa alla progressione del DT1 in pazienti (proband) e nel loro gemello MZ inizialmente non affetto, sono indicate dalle linee rosse ($P < 0.001$). La progressione verso l'autoimmunità anti isola (ad esempio lo sviluppo di autoanticorpi anti isola, il DT1, o entrambi) nel gemello del paziente è indicata dalla linea blu. Ogni punto indica numero di pazienti ancora seguiti ad un dato momento (Age). Le barre indicano l'intervallo di confidenza del 95% (Redondo et al., 2008).

Un ulteriore supporto al ruolo dei fattori ambientali è dato dal fatto che l'incidenza del DT1 appare costantemente in aumento nei paesi sviluppati a partire dal dopoguerra, e previsioni allarmanti prospettano che entro il 2020, l'incidenza potrebbe addirittura raddoppiare nei bambini sotto i 5 anni di età (Patterson et al, 2009). L'aumento di incidenza sta avvenendo in un arco di tempo troppo breve per poter essere causato da cambiamenti nell'assetto genetico, ed in questo contesto è quindi verosimile che abbiano giocato un ruolo importante le modifiche ambientali, ovvero il cambiamento degli stili di vita, avvenute nel corso degli ultimi anni.

Nell'ambito di questo progetto, come approccio completo allo studio del DT1, abbiamo intrapreso l'analisi del profilo trascrizionale, del repertorio dei recettori delle cellule T (TCR) e del metiloma in 5 coppie di gemelli monozigoti, concordanti e discordanti per il DT1, con l'obiettivo di identificare fattori genetici ed epigenetici non ancora descritti che possano concorrere a migliorare il quadro delle conoscenze finora acquisite sulla patologia.

Il “Progetto Diabete”

L’istituto che mi ha ospitato durante il mio Dottorato di ricerca è l’Istituto di Ricerca Genetica e Biomedica (IRGB) ed è diretto dal Prof. Francesco Cucca, il mio tutor, un esperto di DT1, di cui si è occupato per larga parte della sua carriera.

Dal suo arrivo in qualità di direttore e responsabile scientifico, lo studio del DT1 è diventato un progetto centrale per l’IRGB ed è articolato fondamentalmente in 3 parti:

- Studio di associazione su tutto il genoma (Genome-Wide Association Study, GWAS)
- Studio dell’esoma su grandi famiglie
- La ricerca delle basi molecolari del Diabete di Tipo 1 attraverso il sequenziamento di ultima generazione in gemelli monozigoti concordanti e discordanti per la malattia.

Io ho svolto il mio progetto di dottorato nel contesto dell’ultimo progetto in elenco, che verrà spiegato in dettaglio nell’ambito di questa tesi.

Di seguito invece darò una breve descrizione dei primi due progetti:

Studio di associazione su tutto il genoma (Genome-Wide Association Study - GWAS)

Gli studi di associazione si basano sull’analisi di ampie casistiche, nell’ordine di diverse migliaia di individui testati per milioni di varianti polimorfiche.

L’obiettivo del progetto è quello di identificare le varianti di suscettibilità al DT1 in una casistica di soggetti sardi: 1603 casi di DT1, di cui 69 individui affetti anche da sclerosi multipla, 2904 controlli e 421 AFBAC (*Affected Family-Based Controls*, ossia controlli fittizi costruiti a partire dal genoma dei genitori)

Tutti gli individui sono stati genotipizzati con diversi chip array: due chip ideali per studi GWAS (Affymetrix 6.0 e Illumina-OmniExpress), e un chip custom specifico per patologie immuni e autoimmuni. I dati di genotipizzazione sono stati integrati con i dati di sequenziamento a bassa copertura (*low coverage sequencing*) dell’intero genoma di 2120 individui sardi.

Ad oggi sono stati sequenziati 3514 individui e le sequenze generate saranno utilizzate per identificare non solo SNPs, ma anche altri polimorfismi quali inserzioni, delezioni o varianti strutturali.

Studio dell'esoma su grandi famiglie

Lo studio utilizza analisi congiunte di *linkage* e di sequenziamento degli esoni (Exome-seq) su individui appartenenti a una famiglia multigenerazionale con più membri affetti da DT1. La famiglia è composta da 13 membri suddivisi in 3 generazioni, di cui 9 sono affetti dalla malattia. Tutti i campioni sono stati genotipizzati per circa 700.000 varianti comuni utilizzando l'array OmniExpress della Illumina, e il sequenziamento dell'esoma è stato eseguito in 3 individui affetti (1 per generazione) e un individuo sano.

Lo scopo è la ricerca di varianti rare e a bassa frequenza e di polimorfismi informativi del DNA, presenti nei pazienti col DT1 e non condivisi con i loro parenti sani.

“La ricerca delle basi molecolari del Diabete di Tipo 1 attraverso il sequenziamento di ultima generazione in gemelli monozigoti concordanti e discordanti per la malattia.”

SCOPO DEL PROGETTO

L’obiettivo del progetto è la comprensione dell’impatto relativo delle variabili genetiche ed epigenetiche nella suscettibilità al DT1. Il progetto si focalizza sullo studio del profilo trascrizionale, del metiloma e del repertorio TCR su cellule T CD4+ e CD8+ provenienti da 5 coppie di gemelli monozigoti di cui 3 concordanti e 2 discordanti per la malattia.

IL DIABETE DI TIPO I

Il DT1 è una patologia cronica autoimmune che si sviluppa prevalentemente in bambini e adolescenti, con un picco di incidenza tra i 10 e i 14 anni, per questo viene chiamato anche diabete giovanile.

La sua caratteristica principale è la ridotta o assente produzione di insulina (Borg et al., 2000), dovuta alla distruzione immuno-mediata delle cellule β delle isole di Langerhans nel pancreas, deputate alla sua produzione. L'insulina è un ormone peptidico che consente di regolare i livelli di glucosio ematici mediante l'attivazione di diversi processi metabolici e cellulari. La mancanza o ridotta secrezione d'insulina comporta negli individui diabetici uno stato di iperglicemia (alti livelli di glucosio ematico) che deve essere controllato mediante l'assunzione quotidiana di insulina.

Quadro clinico

Quando la produzione d'insulina da parte delle cellule β ancora funzionanti non è più sufficiente per regolare l'ingresso e l'utilizzo del glucosio all'interno delle cellule dell'organismo, compare il quadro sintomatologico tipico della malattia: iperglicemia plasmatica e concomitante squilibrio metabolico.

Il soggetto diabetico manifesta minzione frequente, poliuria (cioè aumento della quantità di urina emessa), enuresi o emissione involontaria di urina, polidipsia o sensazione di sete intensa che porta ad un frequente bisogno di bere, diminuzione della massa muscolare, stanchezza e dimagrimento.

La causa di questa sintomatologia è una cheto-acidosi metabolica, legata allo scompenso metabolico di base, associata ad una disidratazione iper-osmolare, in quanto la concentrazione osmotica del plasma aumenta per un incremento della concentrazione del glucosio extra-cellulare. Ciò determina richiamo di acqua dalle cellule al plasma con conseguente disidratazione intra-cellulare e sofferenza di diversi organi, fra i quali il sistema nervoso centrale, che in situazioni estreme determina il coma diabetico.

Prevalenza, incidenza, e distribuzione geografica del DT1 nel mondo

Il termine “prevalenza”, relativo ad una patologia, descrive la proporzione di individui di una popolazione che presenta la malattia in un dato momento, ed è determinata dall’incidenza e dal tasso di sopravvivenza, che può variare marcatamente nelle diverse popolazioni. L’incidenza indica invece i nuovi casi della malattia ed è solitamente espresso come il numero di nuovi casi per 100.000 persone nate vive l’anno.

Il DT1 ha una prevalenza più alta nelle popolazioni di origine caucasica e più bassa in quelle di origine orientale. La prevalenza del DT1 in individui di età inferiore a 15 anni varia tra 0,05 e 0,3% nella maggior parte delle popolazioni europee e nord americane (Liese et al., 2006).

L’incidenza del DT1 varia per zona geografica, etnicità, età, sesso, e tempo storico. Una delle caratteristiche più importanti del DT1 è la grande variabilità geografica osservata per l’incidenza, che varia da 0,73/100.000 in Cina a 60/100.000 in Finlandia. Questo significa che un bambino in Finlandia ha 100 volte più probabilità di sviluppare la malattia di un bambino in Cina (rivisto da Steck e Rewers , 2015).

E’ stato inoltre descritto un gradiente di incidenza nord-sud a livello mondiale: le nazioni sotto l’equatore presentano infatti una bassa incidenza a differenza delle regioni situate sopra l’equatore, dove la malattia invece è più comune (Fig 3).

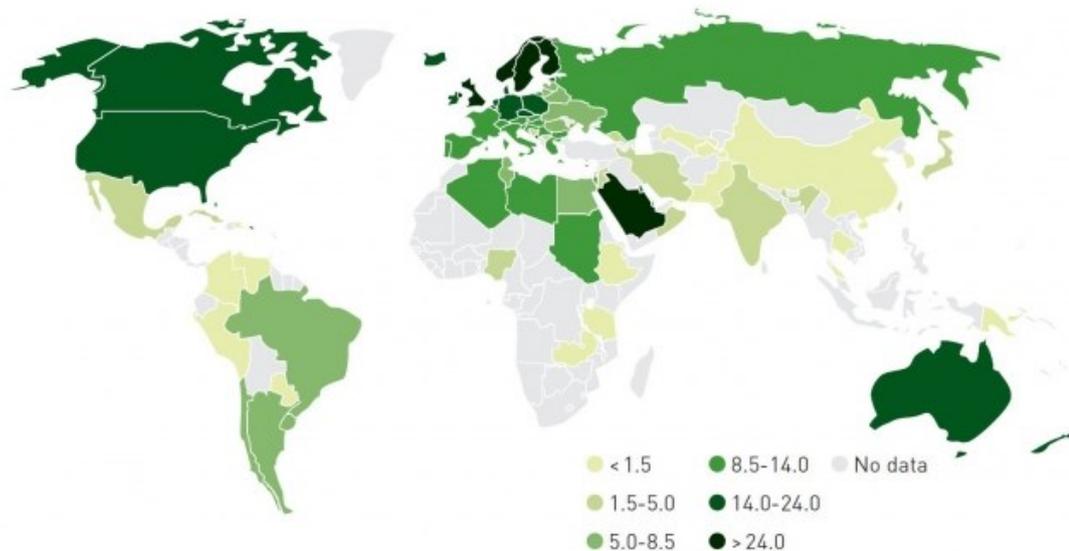


Figura 3. Gradiente di incidenza nord-sud a livello mondiale. Le nazioni sotto l’equatore presentano una bassa incidenza a differenza delle regioni situate sopra l’equatore, dove la malattia invece è più comune.

L'Europa è il continente con il rischio maggiore al mondo, ed è caratterizzata da un marcato gradiente del rischio malattia, inoltre presenta la più grande variazione intra-continentale. Più in dettaglio, è stata osservata la più alta prevalenza in Finlandia, rispetto all'area mediterranea, dove però la Sardegna rappresenta l'eccezione, presentando una sorprendente e insolita alta incidenza -42 affetti DT1 ogni 100.000 nati vivi- (Fig.1).

Le variazioni del rischio del DT1 nella distribuzione geografica ed etnica possono riflettere la presenza di *pool* diversi di geni di suscettibilità, di differenze nei fattori ambientali scatenanti o della combinazione di entrambi.

Età e sesso

I picchi d'incidenza del DT1 sono a 2, 4-6, e 10-14 anni, forse dovuti ad alterazioni nel pattern d'infezioni o all'aumento dell'insulino-resistenza. Il picco d'incidenza principale si ha durante la pubertà, e nelle femmine questo picco è anticipato di circa un anno rispetto ai maschi. In generale, il rischio di sviluppare la malattia è simile tra maschi e femmine (Gale e Gillespie, 2001)

In Europa, le popolazioni con l'incidenza generale maggiore di 20/100.000 (Italia e in particolare Sardegna, Regno Unito, Finlandia, Norvegia) mostrano un eccesso d'incidenza maschile, mentre nelle popolazioni con incidenza generale inferiore (Paesi Baltici, Macedonia, Jugoslavia, Romania) è stata descritta una maggiore incidenza femminile (Serban et al., 2001).

La Sardegna è in linea con l'Italia, pur avendo un maggior rischio generale, per quanto riguarda la maggiore incidenza maschile. Il rapporto maschi-femmine è 1,4 ed è ancor maggiore nell'intervallo di età 10-14 (1,8) (EURODIAB, periodo 1989-1999).

Aumento dell'incidenza e della stagionalità

L'incidenza del DT1 sta aumentando a livello mondiale sia nelle popolazioni con bassa che in quelle con alta incidenza (TEDDY study, 2008; Liese et al., 2006). Ci sono evidenze di marcate variazioni nell'incidenza di DT1 nel tempo, sia in relazione alla stagionalità che negli anni. Nello studio EURODIAB ACE Study Group, che coinvolge la maggior parte delle nazioni europee e Israele, i tassi d'incremento erano del 6,3%. Più in dettaglio, sono del 4,1–8,5% per bambini dai 0 ai 4 anni, del 3,1% (1,5–4,8%) per bambini dai 5 ai 9 anni, e 2,4% (1,0–3,8%) per quelli di età compresa tra i 10 e i 14 anni (EURODIAB 2000).

ALTRI TIPI DI DIABETE

In alcuni pazienti il diabete autoimmune è diagnosticato in età adulta. La maggior parte degli studi nell'adulto è confinata alla tipologia nota come "LADA- *latent onset diabetes of the adult*" (LADA), la cui classificazione è controversa. In generale, tali soggetti sono di età compresa fra i 30 e i 50 anni, non insulino-dipendenti o il cui trattamento con insulina inizia dopo 3-6 mesi dalla diagnosi, e che abbiano positività per gli auto-anticorpi GAD (GADAs) che sono associati con il DT1 (www.actionlada.org). La diagnosi è per lo più basata sulla positività di anticorpi anti-GAD, un parametro usato anche per la diagnosi di DT1.

E' stato proposto che il LADA possa essere fondamentalmente una forma clinica del DT1 caratterizzata da un'insorgenza tardiva e/o da una progressione lenta, probabilmente attribuibili ad una minore suscettibilità genetica (Leslie et al., 2008; Leslie et al., 2010; Howson et al., 2011).

La forma di diabete tipica dell'adulto è il diabete di tipo2 (DT2), chiamato anche "diabete mellito non insulino-dipendente" (NIDDM), ed è caratterizzato da glicemia alta in un contesto di insulino-resistenza e insulino-deficienza relativa, e non è una malattia autoimmune.

Oltre al DT1 e DT2 esistono anche altre forme minori e monogeniche che sottolineano l'eterogeneità della malattia. La tipologia diabetica nota come MODY (*Maturity Onset Diabetes of the Young*) è una forma di diabete familiare a trasmissione autosomica dominante, che insorge tipicamente prima dei 25 anni di età. E' monogenica, e sono stati identificati ad oggi 13 geni causativi (OMIM #606391). Spesso non richiede trattamenti con insulina.

COMPONENTE GENETICA DEL DT1

Era pre-GWAS

Fino a pochi anni fa, a causa dei limiti tecnologici e metodologici, è stato molto difficile ottenere risultati certi ed inequivocabili nell'identificare le componenti genetiche delle malattie multifattoriali o complesse e per qualunque variante localizzata al di fuori della regione HLA. Storicamente, prima dell'avvento del GWAS, furono scoperti solo 6 loci pienamente associati con DT1. La ricerca dei fattori genetici coinvolti nella malattia è iniziata nei primi anni '70 attraverso una lunga serie di studi di *linkage* basati sulla ricerca di regioni cromosomiche candidate per la malattia e successivo mappaggio fine.

Dott.ssa Loredana Marcia. Titolo tesi: Alla ricerca delle basi molecolari del Diabete di Tipo 1: sequenziamento di ultima generazione in gemelli monozigoti concordanti e discordanti per la malattia. Tesi di Dottorato in Scienze Biomediche. Indirizzo: Genetica Medica, Malattie Metaboliche e Nutrigenomica. Ciclo XXVIII. Università degli Studi di Sassari.

La regione genomica contenente il locus del complesso maggiore di istocompatibilità/antigene leucocitario umano (MHC-HLA) localizzato nella regione cromosomica 6p21.3 fu il primo candidato fortemente associato nel DT1 già nel 1970 (Singal & Blajchman, 1973; Nerup et al., 1974; Cudworth & Woodrow, 1975). Il complesso maggiore di istocompatibilità è un insieme di geni che codifica antigeni specifici la cui funzione è la regolazione delle risposte immuni.

La regione HLA-MHC rappresenta il locus con il maggior effetto genetico sulla suscettibilità alla malattia. Si stima, infatti, che circa il 45% della suscettibilità genetica sia da attribuire a tale regione (Levin et al., 2003).

In seguito sono state identificate altre varianti di rischio, in particolare alcuni polimorfismi nella regione promotrice del gene che codifica per l'insulina (INS), che influenzano l'espressione del peptide pre-pro-insulina nel timo (Bell et al., 1984) e rappresenta il locus di suscettibilità con più alto effetto dopo la regione HLA (Fig. 4).

Altre varianti di rischio sono state descritte nel gene codificante per la proteina CTLA-4 -Antigene 4 dei linfociti T citotossici- (Ueda et al., 2003) e nel gene codificante la proteina PTPN22 -tirosina fosfatasi N22 espressa nei linfociti- (Bottini et al., 2004), geni che inibiscono la risposta delle cellule T, e nel gene codificante l' IL2RA -recettore α dell'interleuchina-2- (Vella et al., 2005), che determina una riduzione di una particolare sottoclasse di cellule T memoria (CD45-CD25hi cellule T non regolatrici) protettive nei confronti della malattia.

Era Post-GWAS

Nel corso degli ultimi anni, il sequenziamento dell'intero genoma umano e la sua caratterizzazione, insieme con la realizzazione del progetto HapMap (International HapMap Consortium, 2003, 2005), lo sviluppo di tecnologie innovative per la genotipizzazione su larga scala e di nuovi potenti metodi statistici, hanno portato ad una rapida accelerazione dello studio delle componenti genetiche del DT1 e di altre malattie a trasmissione ereditaria multifattoriale, grazie agli studi di associazione sull'intero genoma (GWAS).

Lo studio GWAS si basa inizialmente sulla caratterizzazione di varianti genetiche polimorfiche, sparse in tutto il genoma, in un'ampia casistica, composta a volte da decine di migliaia di individui. La casistica solitamente è composta da un gruppo di individui affetti da una specifica malattia, e da un gruppo composto da individui sani. La prima fase di caratterizzazione genetica è seguita poi dalla ricerca di quelle varianti con differenza di

frequenza statisticamente significativa tra i due gruppi. Queste varianti sono poi considerate come indicative della regione in cui è probabile che si trovi anche la mutazione che causa la malattia.

I GWAS hanno consentito l'identificazione di oltre 60 loci di suscettibilità nel DT1 (www.immunobase.org). I loci principali sono mostrati in Figura 4, ordinati per *Odds Ratio*, e nella mappa concettuale in Figura 5.

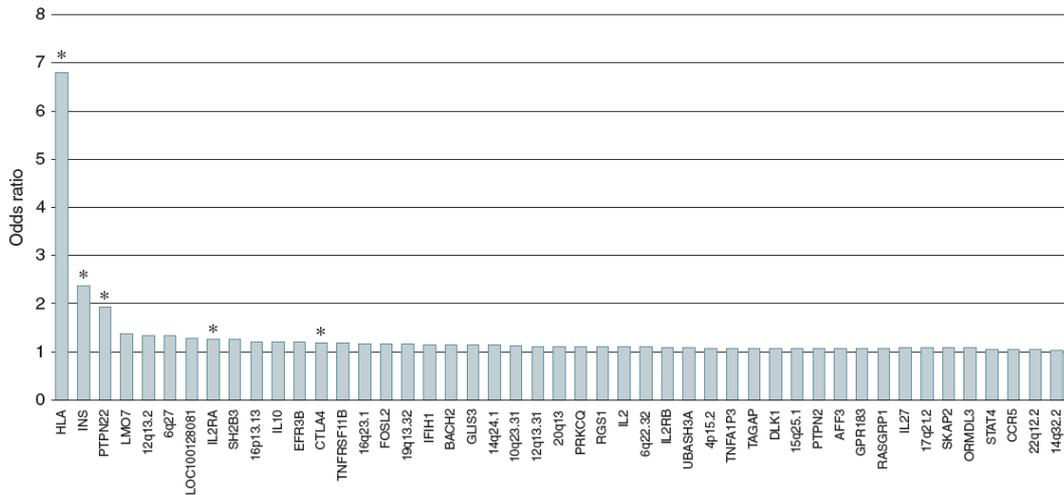


Figura 4. Geni associati al DT1 con Odds ratio >1. L'asterisco evidenzia i geni scoperti nell'era pre-GWAS. Figura presa dal capitolo 30 del libro "International Textbook of Diabetes Mellitus, 4° Edizione. Edito da Ralph A. De Fronzo, Ele Ferrannini, Paul Zimmet, e K. George M. M. Alberti. © 2015 John Wiley & Sons, Ltd. Pubblicato nel 2015 da John Wiley & Sons, Ltd.

Coerentemente con la natura autoimmune della malattia, la maggior parte dei loci di suscettibilità mappa in geni che influenzano la risposta immunitaria; ma ad ogni modo, ciascuno di questi loci contribuisce solo in minima parte al rischio totale di sviluppare la malattia (*Odds Ratio*, OR ~ 1,2), ad eccezione di *HLA*, *INS*, *PTPN22*.

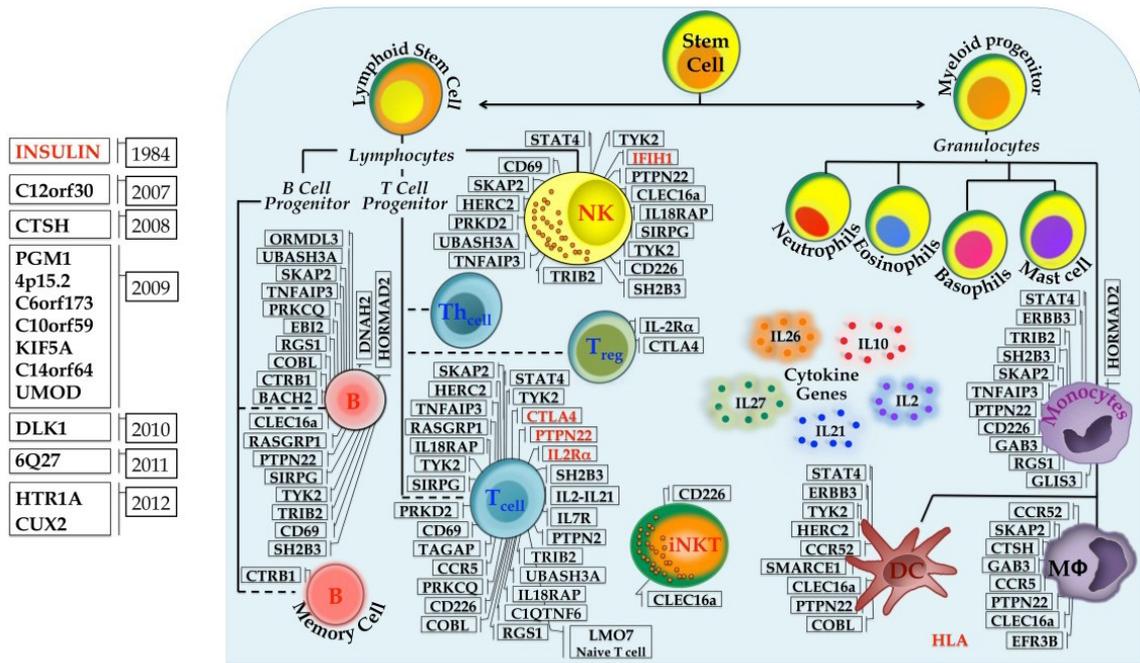


Figura 5. Geni implicati nel DT1. Dal 1974 al 2006, nell'era pre-GWAS, sono stati identificati i 6 geni evidenziati in rosso. Al 2012, i GWAS hanno permesso di identificare circa 60 geni associati alla malattia. <http://www.intechopen.com/books/type-1-diabetes>

Nonostante gli evidenti progressi degli ultimi anni, una porzione di ereditabilità per il DT1, come per altre malattie complesse, rimane comunque inspiegata. Ad esempio è stato stimato che, per il DT1, la regione HLA, il gene *INS* e i circa 60 loci fino ad ora descritti spieghino insieme solo circa il 60% dell'ereditabilità (Polychronakos Li, 2011). Uno studio per calcolare il ruolo della componente genetica verso i fattori ambientali ha dimostrato che l'88% della varianza fenotipica è dovuta a fattori genetici e il resto-12%- a fattori ambientali (Mehers et al., 2008). Inoltre recentemente è stato stimato che in età pediatrica, il contributo ambientale alla malattia rappresenta circa il 10% (Li et al., 2015).

La componente genetica del DT1, non spiega quindi interamente l'eziopatogenesi della malattia.

COMPONENTE AMBIENTALE DEL DT1

La non completa concordanza per la malattia tra gemelli MZ e nei topi NOD con stesso patrimonio genetico e stabulati nelle stesse condizioni, il fatto che avere geni di suscettibilità non sia sufficiente a indurre la malattia, l'aumento dell'incidenza osservato negli anni e soprattutto nei paesi industrializzati, ci hanno portato a ipotizzare che, oltre alla forte componente genetica, ci sia una combinazione di fattori ambientali, e di conseguenza anche epigenetici, che giocano un ruolo importante nella predisposizione o nella protezione nei confronti del DT1.

Esistono infatti diversi meccanismi epigenetici che consistono essenzialmente nella variazione dello stato di metilazione del DNA, nelle modifiche istoniche e nell'azione di trascritti di RNA non codificanti che possono verificarsi anche come risposta ad esposizioni ambientali. Queste modifiche epigenetiche potrebbero modulare l'espressione di geni chiave nell'insorgenza del DT1 o di altre patologie autoimmuni e contribuire a spiegare la discordanza dei gemelli monozigoti per la malattia (Kaminsky et al., 2009).

Quando si parla di "fattori ambientali", si intendono gli stili di vita e l'ambiente, quindi l'alimentazione, la carenza o l'eccesso di vitamine o altri nutrienti, le abitudini come il fumo o il grado di attività fisica, l'esposizione a virus e batteri, l'esposizione ad inquinanti ambientali. I fattori ambientali rimangono tuttora speculativi e di difficile dimostrazione.

Nei paesi sviluppati, tra gli anni '40 e '50 c'è stato un netto miglioramento delle condizioni igienico-sanitarie che ha ridotto l'incidenza di alcune malattie e l'esposizione a specie batteriche e parassitarie. È stato quindi ipotizzato, in base alla cosiddetta "teoria igienica" che l'aumentata incidenza di malattia sia legata alla scomparsa o comunque riduzione d'infezioni protettive. Inoltre, durante i ~200.000 anni trascorsi dalla comparsa della nostra specie (Francalacci et al., 2013) il nostro sistema immunitario si è evoluto per far fronte ad infezioni di vario tipo. Questo meccanismo potrebbe aver contribuito alla selezione di varianti genetiche utili per le nostre difese contro i patogeni ma attualmente, in un ambiente molto meno settico, potenzialmente dannose in quanto in grado di determinare una condizione di iperattivazione del sistema immune predisponente nei confronti dell'autoimmunità (Pitzalis et al., 2013).

Da tutte queste considerazioni deriva la necessità di integrare i disegni sperimentali esistenti con nuovi approcci metodologici, sia per identificare nuove associazioni genetiche, che per dare un nesso biologico di tali varianti con le malattia in esame.

PATOFISIOLOGIA DEL DT1

Malgrado la malattia sia ben caratterizzata da un punto di vista clinico, gli eventi genetici e fisiologici che ne controllano l'insorgenza e la progressione non sono stati ancora chiariti.

Durante gli ultimi 25 anni la ricerca scientifica si è avvalsa di due modelli animali fondamentali -il ratto in-breed BioBreeding (BB) (Mordes et al., 1996)- ed il topo diabetico non obeso -topo NOD (Anderson 2005; Makino 1980)- con lo scopo di studiare la genetica, la patofisiologia e l'impatto ambientale sulla forma spontanea della malattia.

I modelli murini presentano molti aspetti in comune con la malattia nell'uomo, inclusi alcuni loci genici di suscettibilità, l'esposizione a determinati fattori ambientali e la patogenesi stessa della malattia. Gli studi sul topo NOD hanno mostrato che la malattia insorge come conseguenza di un disequilibrio della regolazione immunitaria, che comporta l'espansione di cellule CD4+ e CD8+ auto-reattive, di cellule B deputate alla produzione di autoanticorpi, e l'attivazione del sistema immunitario innato che contribuisce alla distruzione delle cellule β producenti l'insulina (rivisto da Bluestone et al., 2010). Queste caratteristiche della malattia sono consistenti con gli studi sul DT1 nell'uomo.

Storia immunologica del DT1

Nonostante la presenza di autoanticorpi nel siero dei pazienti con DT1 sia il primo elemento diagnostico dell'autoimmunità (il che porterebbe a pensare al ruolo primario delle cellule B), c'è una forte evidenza, supportata sia dagli studi sui topi NOD che dalle caratteristiche della malattia nell'uomo, che a giocare in ruolo dominante nell'innescare e nella progressione della malattia siano le cellule T auto-reattive.

Possiamo suddividere la patogenesi della malattia in 3 fasi principali, chiaramente descritte nella Figura 6:

- 1) innescare (*priming*) a livello pancreatico
- 2) “*epitope spreading*” con conseguente attivazione delle cellule T a livello dei linfonodi pancreatici
- 3) invasione delle isole pancreatiche da parte dei linfociti T effettori con conseguente insulite

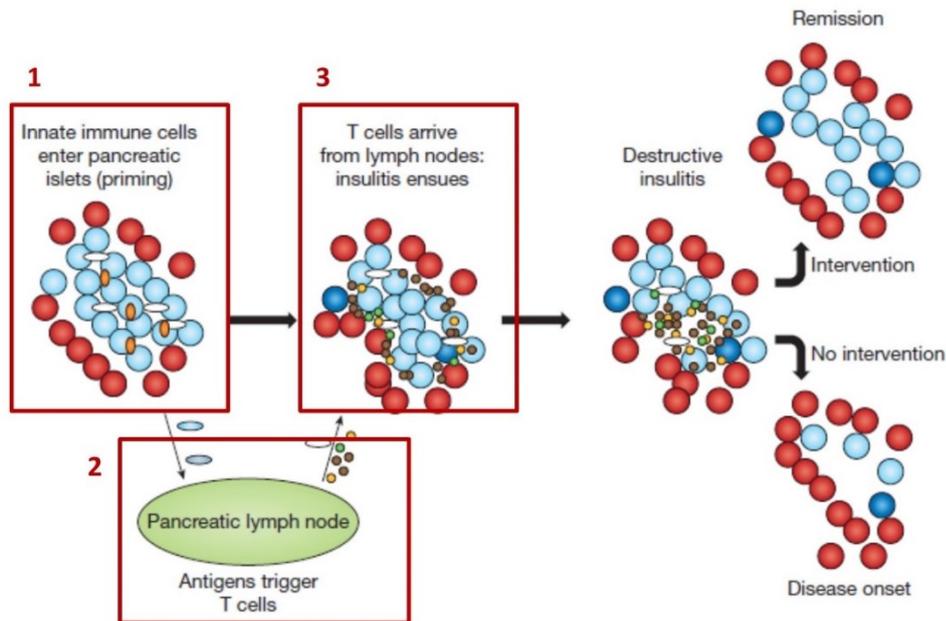


Figura 6. Storia immunologica del DT1 (adattato da Bluestone et al., 2010)

1) La prima fase del processo immunologico avviene nel pancreas, dove, in un individuo con predisposizione genetica, un evento scatenante non ancora ben identificato, innesca un processo infiammatorio cronico di bassa entità (“*priming*”). Gli eventi iniziali coinvolgono l’infiltrazione nelle isole pancreatiche di cellule del sistema immunitario innato come cellule dendritiche convenzionali (cDC), Natural Killer (NK) e di cellule B auto-reattive (Fig.6, ovali arancioni). Nel topo NOD si è visto che le cellule dendritiche CD11c+ e i macrofagi ER-MP23+ sono le prime cellule che si infiltrano nel pancreas, a circa 3 settimane di età (Bluestone et al., 2010). In termini di meccanismi patogenici potenziali, le CD8+ sono la popolazione cellulare più rappresentata nelle lesioni delle isole, seguite da, in ordine decrescente, macrofagi (CD68+), cellule CD4+, linfociti B (CD20+) e plasma cellule (CD138+) (Willcox et al., 2009). Sorprendentemente le cellule FOXP3+ (ad esempio le T reg) e le NK sono rare nelle lesioni.

2) L’accumulo del danno delle isole dovuto alla morte cellulare e/o a infezioni virali o batteriche a sua volta comporta un ulteriore rilascio di auto-antigeni, con la conseguente presentazione, a livello dei linfonodi pancreatici, di nuovi autoantigeni alle cellule T (Fig.6, puntini marroni) ad opera di cellule presentanti l’antigene (APC, ovali bianchi), il che porta all’attivazione di nuove cellule T (“*epitope spreading*”), all’espansione clonale e quindi all’amplificazione del processo infiammatorio.

Gli ovali blu in figura 6 indicano le APC legate agli antigeni delle isole. Le cellule B (Fig.6, puntini verdi) e le cellule dendritiche potrebbero essere le prime cellule responsabili della presentazione dell'antigene. Le cellule T regolatrici (Fig.6, T-Reg-puntini gialli) possono reprimere il processo infiammatorio soprattutto negli stadi precoci, ma non sono capaci di contenere il processo di amplificazione negli stadi tardivi malgrado vi sia un aumento del loro numero.

3) La lesione delle isole pancreatiche, caratterizzata da infiltrazione da parte di cellule CD4⁺ e CD8⁺ e dalla presenza di necrosi cellulare è chiamata "insulite". L'insulite non colpisce tutte le isole pancreatiche, inoltre il processo infiammatorio ha come bersaglio specifico le cellule β deputate alla produzione di insulina (Fig.6, cerchi celesti), mentre le altre cellule endocrine (Fig.6, cerchi rossi) presenti nelle isole non vengono attaccate. Sono stati descritti tentativi di riparazione delle cellule beta sopravvissute presenti nel centro delle isole (Fig.6, cerchi blu scuro).

Il ciclo della presentazione dell'antigene, l'attivazione delle cellule del sistema adattativo, l'attivazione delle cellule T effettrici e l' "*epitope spreading*" contribuiscono così alla perdita progressiva delle cellule β , che comporta a sua volta l'insorgenza di uno stato di iperglicemia.

Si pensa anche che all'interno delle isole si sviluppino organi linfonodi terziari, il che può portare ad un'ulteriore amplificazione della risposta adattativa. La perdita delle funzioni metaboliche all'inizio può essere sia funzionale che anatomica, perché le terapie immunologiche possono ripristinare le cellule che hanno perso la capacità di produrre insulina. Tuttavia senza alcun intervento terapeutico la perdita delle beta cellule continua (Bluestone et al., 2010).

Il ruolo delle cellule B (Fig.6, cerchi verdi) rimane ancora da chiarire. Benché abbiano un ruolo nella produzione di autoanticorpi, le cellule B "antigene specifiche" sono molto efficienti nel presentare l'antigene e nel produrre citochine che possono sia promuovere che sopprimere l'immunità. Sta diventando sempre più chiaro che nel topo NOD e probabilmente nell'uomo, il DT1 è amplificato da componenti immuni dell'infiammazione tissutale, analogamente a quanto descritto nel DT2, nella malattia di Alzheimer, nell'aterosclerosi e in altre malattie.

Cellule T reattive, aventi sia il complesso maggiore di istocompatibilità (MHC) di classe I che di classe II (CD4+ e CD8+) e dirette contro antigeni delle isole sono state identificate sia nel sangue periferico di topi NOD che in quello di pazienti DT1.

In molti casi è stato dimostrato che queste cellule T riconoscono antigeni anti-isola simili a quelli visti attraverso lo studio degli autoanticorpi (come auto anticorpi anti insulina-IAA, anti acido gluatammico decarbossilasi-GAD-, anti trasportatori di zinco - ZnT8). Le cellule T riconoscono anche altri antigeni delle isole, come la glucosio-6-fosfatasi proteina relata alla subunità catalitica (IGRP) e la cromogranina A, sia nei topi NOD e che nell'uomo (Burton et al., 2008; Lieberman et al., 2003; Medarova et al., 2008).

Le cellule T presenti nel sangue periferico dei pazienti possono essere un riflesso indiretto di ciò che sta succedendo nel pancreas, e rappresentano una fonte accessibile per poter eseguire studi di espressione, funzionalità, e per studiare il repertorio TCR.

Il ruolo dell' HLA

La regione HLA è un cluster di geni localizzato nel complesso maggiore di istocompatibilità (MHC) sul cromosoma 6 in posizione 6p21.3 e rappresenta una regione cromosomica molto estesa (circa 4M di paia di basi) e di fondamentale importanza per la risposta immunitaria dell'organismo (Fig.7)

Il *locus* HLA può essere suddiviso in 3 regioni che codificano principalmente per:

- molecole di classe I, rappresentano i classici antigeni da trapianto e sono espresse sulla superficie di tutte le cellule nucleate
- molecole di classe II, da subito identificate come molecole della risposta immunitaria
- molecole di classe III, con diverse funzioni inclusi i geni TNF e alcuni sistemi del complemento, ma non i classici geni HLA

I geni della classe I (A, B e C) e II (DR, DQ and DP) sono strutturalmente simili e i loro prodotti proteici presentano gli antigeni ai recettori delle cellule T: quelli di classe I alle cellule CD8, quelli di classe II alle CD4.

La catena pesante degli antigeni di classe I viene codificata dai geni situati nei *loci* HLA-A, B o C. Le molecole di classe I sono polipeptidi etero-dimerici costituiti dalla catena pesante legata a una molecola di β_2 -microglobulina (Fig. 7B).

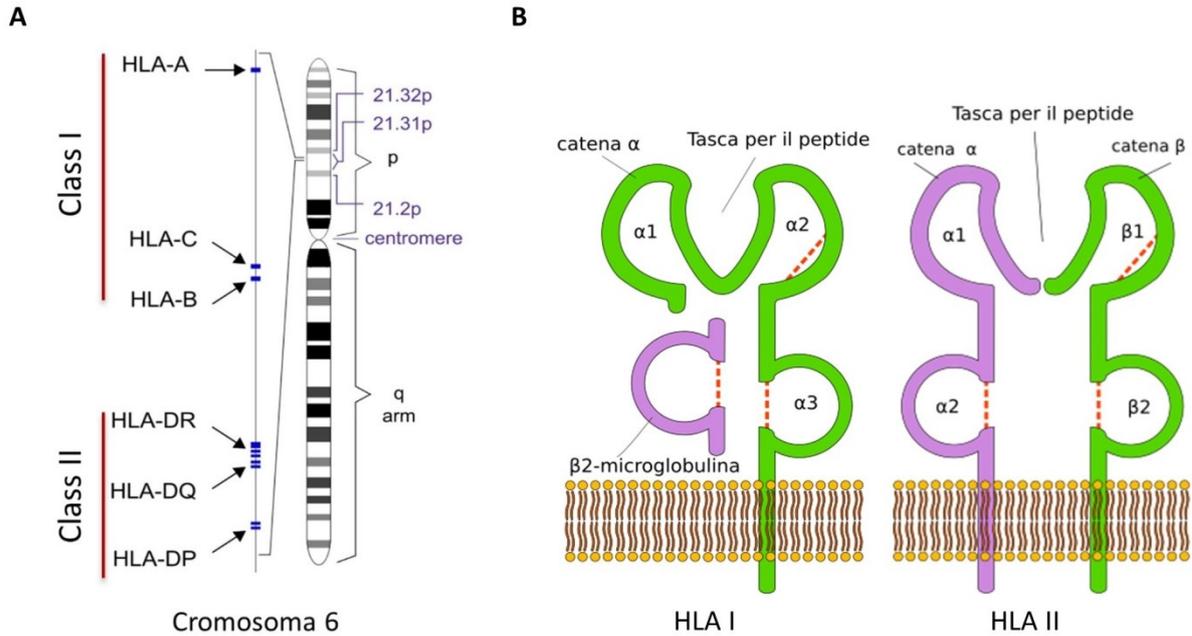


Figura 7. HLA I e II, geni e struttura proteica. (A) Distribuzione dei geni HLA I e II nel cromosoma 6. (B) Le molecole HLA I sono costituite da una catena alfa e da una molecola di β_2 -microglobulina. Le molecole HLA-II, simili alle molecole HLA-I, sono costituite da due catene proteiche, chiamate α e β . Entrambe le catene hanno due regioni extracellulari, una regione trans-membrana che le ancora alla cellula e una corta regione intra-citoplasmatica.

Questi antigeni sono presenti sulla maggior parte delle cellule nucleate dell'organismo, come pure sulle piastrine e sono omologhi agli antigeni di trapianto identificati con metodi sierologici presenti in altre specie.

Gli antigeni di classe II sono costituiti da due catene polipeptidiche α e β (Fig. 7B), le quali sono entrambe codificate da geni situati nella regione HLA-D. La regione HLA-D è divisa in subregioni, ognuna avente geni che codificano sia per la catena α sia per quella β delle diverse molecole di classe II (HLA-DR, DQ e DP). Gli antigeni di classe II vengono espressi prevalentemente sulle cellule presentanti l'antigene (APC) come i linfociti B, i macrofagi, le cellule dendritiche e alcune cellule endoteliali.

I geni MHC sono inoltre specificati da centinaia di varianti alleliche per cui la regione codificante per proteine MHC è sicuramente la più polimorfica del genoma umano (sebbene in un singolo individuo non siano presenti più di una ventina di varianti alleliche del MHC). I geni HLA sono co-dominanti, per cui in un individuo sono espressi sia gli alleli di origine materna che paterna.

Uno degli aspetti distintivi del DT1 è rappresentato dai gradi estremi di resistenza alla malattia conferiti da specifiche varianti alleliche e aplotipiche, cosiddette protettive (Cucca et al., 1995, Cucca et al., 2001). L'associazione della regione HLA con la malattia mostra un *continuum* che va da aplotipi predisponenti (alleli *DRB1 0401*, *DRB1 0402*, *DRB1 0405*, *DQA1 0301*, *DQB1 0302* o *DQB1 0201*) ad aplotipi fortemente protettivi (*DRB1 1501*, *1401* o *0701* e *DQB1 0602*, *0503* o *0303*).

Le molecole MHC di classe I e II hanno la capacità di creare la tasca di legame (*binding groove*) per gli antigeni presentati ai linfociti T, gli effettori finali della malattia. Nel loro insieme gli studi eseguiti sino ad oggi suggeriscono un modello patogenetico in cui la protezione dalla malattia sarebbe direttamente proporzionale all'affinità ed efficienza del legame fra le molecole MHC, il peptide diabetogenico e il recettore dei linfociti T (TCR) nel timo (Ridgway et al., 1998)

Per quanto riguarda le molecole MHC di classe II rispettivamente con un effetto predisponente e protettivo, sono state descritte marcate similarità tra l'uomo e i topi NOD. I dati provenienti da animali transgenici indicano effetti, mediati direttamente dalle molecole MHC di classe II sul repertorio TCR nel timo, sulla suscettibilità alla malattia.

La suscettibilità può insorgere attraverso la selezione positiva di cellule T auto-reattive attraverso alleli di classe II associati positivamente, mentre la protezione può avvenire tramite la selezione negativa di cellule T diabetogeniche e/o selezione di cellule T regolatrici, mediata da alleli associati negativamente. Questi eventi possono influenzare sia i repertori TCR delle cellule CD4+ che CD8+ nel timo (letteratura rivista da Motzo et al., 2004).

Il DT1 può essere considerato quindi come una malattia causata da un difetto di questo complesso meccanismo volto al mantenimento della tolleranza verso uno specifico autoantigene pancreatico.

Il ruolo dell'insulina nell'insorgenza della malattia

Studi recenti suggeriscono l'esistenza di un limitato numero di autoantigeni primari delle isole, riconosciuti da cellule infiltranti precoci, che può essere responsabile dell'innescamento della malattia (Lennon et al., 2009). Ancora non è stato però chiarito se ci sia un'unica proteina "self" responsabile dell'inizio della malattia. A tale proposito è stato osservato che la concentrazione di auto-anticorpi diretti contro l'insulina (IAA) correla con il tasso di progressione verso la malattia conclamata in bambini seguiti dalla nascita (Steck et al., 2011;

Parikka et al., 2012). Questa evidenza, combinata con altri studi nell'uomo e nel topo, supporta la tesi che la proinsulina sia l'antigene chiave nella malattia (Kent et al., 2005), un concetto che in parte potrebbe spiegare la perdita selettiva delle cellule β nel DT1 (rivisto da Atkinson, Eisenbarth e Michels, 2014).

Ci sono inoltre forti evidenze che i difetti nella selezione negativa delle cellule T a livello del timo siano collegate alla reattività verso l'insulina, la quale rappresenta dunque un fattore chiave della predisposizione genetica e che potrebbe essere la proteina "self" principale per l'innescamento. Questa tesi è supportata dal fatto che:

1) Circa il 20 % degli individui con mutazioni spontanee nel gene regolatore dell'autoimmunità AIRE sviluppano il DT1 e/o altri disordini autoimmuni, che sembra riflettere la loro inability di selezionare negativamente le cellule T auto-reattive durante lo sviluppo (Gardner et al., 2009). L'insulina è una proteina delle isole pancreatiche regolata dall'AIRE ed è anche ectopicamente espressa nelle cellule epiteliali della midollare del timo. Inoltre, polimorfismi nel promotore dell'insulina, che mappano in un allele di suscettibilità, controllano l'espressione dell'insulina nel timo, il che suggerisce un ruolo potenziale nella regolazione del repertorio auto-immunologico.

2) Nel topo NOD è stata osservata un'alta percentuale di cellule patogenetiche isolate da infiltrati insulari precoci in grado di riconoscere l'insulina. Infatti, sembra che ci sia un'unica relazione tra l'espressione di certe regioni variabili della catena alfa ($V\alpha$) dei recettori delle cellule T (TCR) e il riconoscimento del peptide B dell'insulina (B9-B23) (Kobayashi et al., 2008).

Comunque, evidenze precliniche supportano la tesi che si abbia una perdita di tolleranza sequenziale verso epitopi multipli, suggerendo che la specificità delle risposte immunitarie verso auto-antigeni multipli sia probabilmente coinvolta nella progressione della malattia.

Regolazione del sistema immunitario

Malgrado tramite GWAS siano stati identificati molti alleli di suscettibilità che possono contribuire alla distruzione delle cellule delle isole mediata dalle cellule T, non ci hanno aiutato ad identificare *pathways* monogenici che portino direttamente alla malattia.

Sembra piuttosto che la suscettibilità alla malattia sia dovuta al problema globale della regolazione della risposta immunitaria.

Mutazioni di geni che codificano in vari *loci* di suscettibilità - inclusi *IL2*, *IL2ra* (CD25),

Ctla4, *Ptpn22*, e *Pdca1* (PD-1) - in vari modelli animali hanno portato allo sviluppo di una diversa serie di malattie autoimmuni, incluso il DT1. Le proteine codificate da questi geni sono implicate nel mantenimento dell'omeostasi immunologica sia direttamente, attraverso la regolazione della forza del segnale del complesso recettore di cellule T e B, sia indirettamente attraverso la sorveglianza di popolazioni cellulari regolatrici critiche per il controllo dell'autoimmunità.

Il ruolo potenziale di questi *pathways* è importante nel controllo delle cellule T regolatrici Foxp3⁺ (T-Reg), che rappresentano i soppressori essenziali di risposte immunitarie non desiderate. Si pensa che le T-Reg modulino l'attivazione delle cellule T e promuovano la tolleranza attraverso la soppressione dell'immunità adattativa tramite interazioni dirette cellula-cellula e la produzione di citochine immuno-modulatrici come il “*transforming growth factor*” TGF- β e IL-10 (Tang et al., 2008).

Inoltre altre funzioni delle T-Reg hanno come bersaglio meccanismi dell'infiammazione più generalizzati come le reazioni redox, l'utilizzo di ATP, il metabolismo del triptofano e i *pathways* dell'ossido nitrico. Questi risultati suggeriscono che le T-Reg possano giocare un ruolo più generale nella sopravvivenza delle cellule β e che funzionino come regolatori innati del danno tissutale immuno-mediato.

I diversi studi eseguiti sino ad ora, insieme al fatto che le cellule T effettrici cambiano nel tempo diventando resistenti alla soppressione T-Reg mediata sia nei topi NOD che nei pazienti affetti da DT1, suggeriscono un modello che ipotizza che la progressione finale della malattia sia una diretta conseguenza dello sbilanciamento tra cellule T regolatorie ed effettrici (Shneider et al., 2008; D'Alise et al., 2008; Bluestone et al., 2010) (Fig.8).

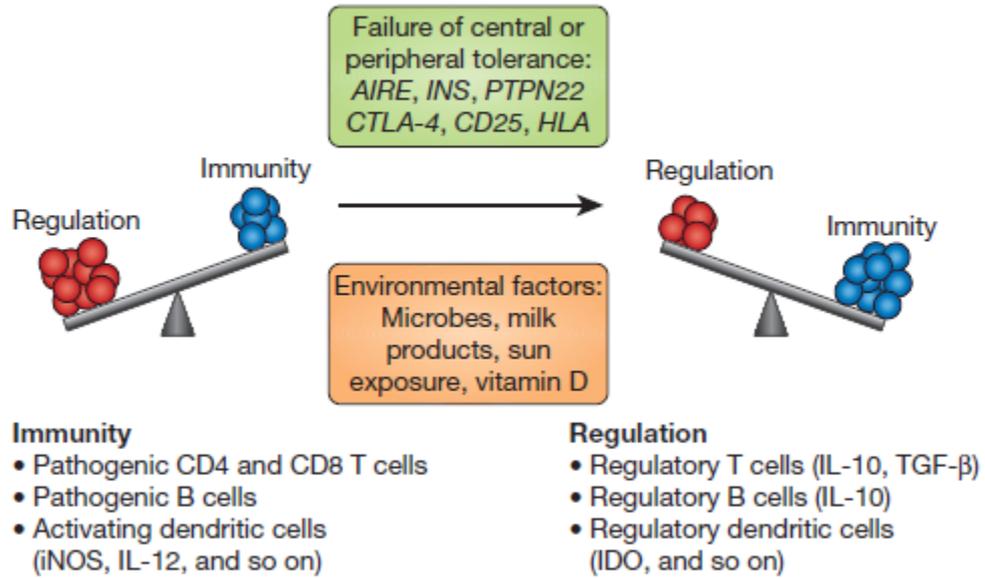


Figura 8. *L'equilibrio del sistema immunitario è la chiave della patogenesi del DT1. La figura illustra il fine equilibrio della regolazione immune versus lo stadio patogeno, evidenziando un numero di geni che pare influenzino l'equilibrio attraverso effetti sulla tolleranza centrale e periferica e i fattori ambientali che controllano l'immunità (Bluestone et al.,2010).*

Storia naturale della malattia

Nel 1986 fu proposto (Eisenbarth et al., 1986) un modello di progressione della malattia che è tuttora considerato valido, se pure con piccole modifiche. L'ultima revisione del 2014 è mostrata in figura 9 (Atkinson et al., 2014).

La malattia insorge in individui geneticamente predisposti. Eventi scatenanti, per esempio influenze ambientali, possono avvenire già in utero e fungere da cosiddetti “inneschi” (*priming*). Eventi fisiologici come lo sviluppo del sistema immunitario (selezione timica negativa e/o positiva), il normale turnover delle beta cellule (che porta al rilascio di auto-antigeni), e de-regolazioni del sistema immunitario possono contribuire al processo patogeno (Atkinson et al., 2011).

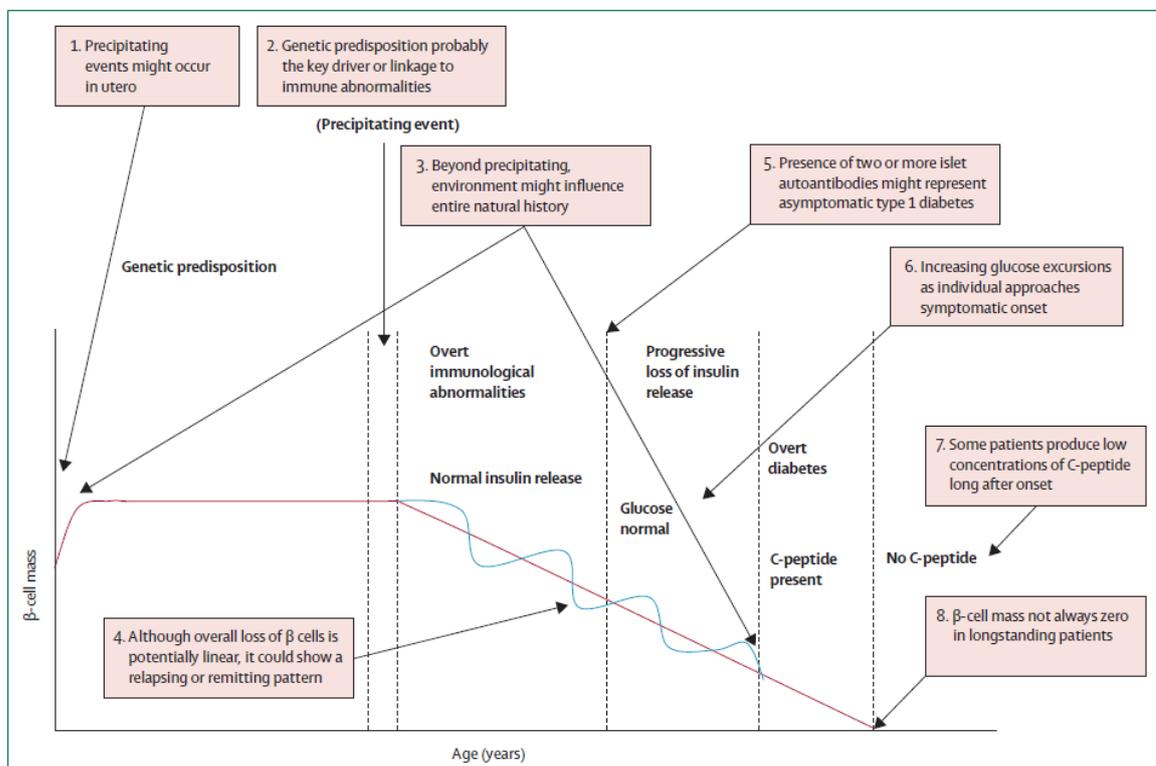


Figura 9. Storia naturale del DM1 - rivisitazione di un concetto di 25 anni fa. Il modello del DM1 è stato ricreato da quello originariamente proposto nel 1986, mostrato in nero. Aggiunte e ipotesi basate sulle scoperte recenti sono state aggiunte in porpora.

L'IMPORTANZA DELLO STUDIO DEI GEMELLI

Gli studi sui gemelli rappresentano un ottimo strumento per stabilire fattori genetici e non in malattie immuno-mediate multifattoriali

I primi studi sui gemelli risalgono al 19esimo secolo, quando Francis Galton, cugino di Charles Darwin, decise di studiare i gemelli per stabilire il ruolo dell'ereditarietà e dell'ambiente sulla vita umana. Galton realizzò che, poiché le coppie di gemelli –identici o non identici– condividono lo stesso ambiente, la maggior concordanza tra gemelli identici comparata a quella di quelli non identici riflette il ruolo di fattori genetici che portano alla malattia. Allo stesso modo, tassi di concordanza simili per una malattia nelle coppie di gemelli- identici o no-, o discordanza per una certa malattia tra gemelli identici, suggeriscono il possibile ruolo di fattori determinanti non genetici.

Le ricerche in campo immunologico nell'ambito di patologie immuno-mediate sono un esempio importante delle potenzialità degli studi sui gemelli. La maggior parte di questi studi sono stati fatti, infatti, su malattie auto-immuni, compreso il DT1.

In queste patologie l'impatto dei fattori genetici è sostanziale, ma l'identità di tutti i geni coinvolti nelle risposte immunitarie è incerta, così gli studi sui gemelli rappresentano l'unico modo, in una popolazione composta da individui non consanguinei, e di identificare le variazioni genetiche che contribuiscono alle differenze immunologiche. Purtroppo ancora le ricerche di questo tipo non hanno dato grandi risultati, forse anche a causa della difficoltà di ottenere numeri sufficienti di coppie di gemelli adatte allo studio.

Studi su gemelli affetti da diverse condizioni immuno-mediate suggeriscono che l'omozigotità gioca un ruolo predominante nel determinare la concordanza del repertorio TCR (Gulwani-Akolkar et al. 1994; Kuwana et al. 2001; Nanki et al. 1996; Somma et al. 2007). Nel nostro caso coppie di gemelli MZ, concordanti e discordanti per DT1, rappresentano lo scenario ideale anche per esplorare fattori stocastici coinvolti nel rimodellamento del repertorio TCR.

Ad ogni modo le malattie immuno-mediate necessitano della presenza sia di fattori genetici che ambientali, e l'evidenza maggiore che siano dovute anche a fattori ambientali proviene dagli studi su gemelli monozigoti.

MATERIALI E METODI

Descrizione del campione

I campioni di sangue dei gemelli utilizzati in questo studio, sardi da almeno tre generazioni, sono stati forniti dal Prof. Fozza che ha già eseguito in precedenza, sugli stessi campioni, studi sul repertorio del TCR (Fozza et al., 2012) con una differente metodica.

Il campione è costituito da 5 coppie di gemelli monozigoti (2 femmine e 8 maschi): 3 coppie concordanti e 2 coppie discordanti per DT1. L'età degli individui è compresa tra i 18 e 45 anni e l'età di insorgenza della malattia varia dai 2 ai 12 anni.

Estrazione di cellule mononucleate da sangue periferico (PBMC)

Il sangue dei soggetti facenti parte dello studio è stato diluito utilizzando Ficoll-Hypaque e una soluzione salina (PBS) tamponata a pH7 in un rapporto 1:1:1 e le cellule sono state separate mediante gradiente di centrifugazione.

Dopo centrifugazione a 1800rpm per 30 minuti, l'anello di PBMC che si colloca nell'interfaccia viene raccolto e lavato per tre volte con PBS arricchito con EDTA2mM ed infine sospeso in Buffer 1 (PBS senza Ca²⁺ e Mg²⁺, 0.1% BSA, 2 mM EDTA, pH 7.4).

La conta delle cellule è stata effettuata tramite il contatore di cellule Scepter™ 2.0 (Merck-Millipore) e utilizzando dei sensori con un apertura di diametro di 40 µm.

Separazione cellule CD4+ e CD8+

Dai PBMC sono state separate due sottopopolazioni di cellule: linfociti T CD4+ e CD8+. Le cellule sono state isolate mediante selezione positiva, utilizzando biglie immunomagnetiche rivestite di anticorpi specifici in grado di riconoscere gli antigeni CD4 o CD8 (Dynabeads; Dynal, Oslo, Norvegia).

Ai PBMC risospesi in Buffer 1 (10⁷ cellule/ml) vengono aggiunte 25µl di biglie/ml. Dopo una incubazione di 20' a +4 °C su un agitatore rotante viene applicato un campo magnetico, quindi viene eliminato il surnatante.

Le biglie, che legano le cellule, vengono lavate diverse volte secondo il protocollo, quindi vengono risospese in Trizol (Invitrogen, Paisley, UK) e conservate a -80 °C fino al momento dell'estrazione. Successivamente abbiamo estratto RNA, DNA e proteine.

Estrazione RNA, DNA e proteine da cellule CD4+ e CD8+

Il Trizol è una soluzione monofasica di fenolo e guanidina isotiocianato che mantiene integro l'RNA durante l'omogeneizzazione o lisi del campione. L'aggiunta di cloroformio (1/10 del volume di Trizol), seguita da centrifugazione (11000 rotazioni per minuto -rpm, 15' +4°C), consente la separazione della soluzione in due fasi: una acquosa contenente l'RNA, e l'altra organica contenente le proteine, separate da un'interfaccia contenente il DNA.

Dopo il trasferimento della fase acquosa in un nuovo tubo, l'RNA viene recuperato tramite precipitazione con un'egual volume di etanolo al 100% seguita da centrifugazione (11.000 rpm, 10', +4°C). Il *pellet* viene lavato con etanolo al 75%, asciugato e risospeso in H₂O priva di RNase o buffer TE (Tris-EDTA).

Dopo la rimozione della fase acquosa, il DNA e le proteine del campione possono essere recuperate tramite precipitazioni sequenziali: la prima, con etanolo al 100% con una proporzione di 300 µl ogni 750 µl di Trizol, permette il recupero del DNA dall'interfase e la seconda, con alcol isopropilico secondo la proporzione di 1.5 ml ogni 750 µl di Trizol che garantisce il recupero delle proteine dalla fase organica.

L'RNA ottenuto è stato sottoposto a vari controlli di qualità che prevedono:

- la valutazione visiva dopo corsa elettroforetica su gel d'agarosio all' 1%, che mette in evidenza l'integrità delle bande 28S, 18S e quando presente della 5S. La presenza delle 3 bande indica l'assenza di degradazione
- la stima della concentrazione e la valutazione della purezza (rapporti OD₂₆₀/280 e 260/230), tramite l'analisi spettrofotometrica utilizzando il Nanodrop 1000
- l'analisi della qualità mediante il Bioanalyzer2100 (Agilent Technologies) usando il kit RNA 6000 LabChip (Agilent #5067-1511). Il Bioanalyzer è un dispositivo bio-analitico basato su chip e costituito da microfluidiche, dove avviene una elettroforesi capillare

I risultati vengono elaborati tramite un algoritmo che assegna un valore di qualità chiamato RNA Integrity Number (RIN) che permette la classificazione dell'RNA totale basandosi su un sistema numerico da 1 a 10, dove 1 rappresenta la massima degradazione e 10 l'integrità totale dell'RNA. Solo i campioni con un RIN > 7,5 sono validi poter procedere con la preparazione delle librerie ed il successivo sequenziamento. I nostri campioni hanno mostrato tutti un RIN > 8, indicante alta qualità, condizione ideale e necessaria per poter procedere con la successiva preparazione per il sequenziamento del trascrittoma.

Il DNA ottenuto è stato sottoposto a vari controlli di qualità che prevedono:

- la valutazione visiva dopo corsa elettroforetica su gel d'agarosio al 1% che mette in evidenza l'integrità delle bande e l'assenza di degradazione
- la stima della concentrazione e una valutazione della purezza (rapporto OD260/280 e 260/230), con analisi spettrofotometrica utilizzando il Nanodrop 1000

Estrazione DNA per la genotipizzazione

I campioni di DNA utilizzati per la genotipizzazione con l'Array Human Omni Express della ditta Illumina sono stati estratti da sangue intero trattato con l'anticoagulante acido etilendiamminotetracetico (EDTA), utilizzando il kit PureLink Genomic DNA Mini della Invitrogen (Cat. N°K1820-01).

Dopo aver preparato il lisato cellulare, partendo da 200 µl di sangue intero, il DNA genomico è rapidamente purificato con colonnine mediante processi di centrifugazione.

Il kit si basa su un legame selettivo del DNA ad una membrana a base di silice, in presenza di sali caotropici. Il lisato è sottoposto a digestione con Proteinase K (20 µl) a 55 °C. Il campione viene trattato con l'enzima RNase (20µl per 2minuti a temperatura ambiente) per consentire l'eliminazione dell'RNA. Successivamente verrà miscelato con 200µl di etanolo e con 200 µl di "PureLink Genomic Binding Buffer" quindi il tutto verrà fatto passare nella colonna, dove il DNA presente si legherà alla membrana.

La purezza del DNA è ulteriormente garantita dai diversi lavaggi a cui viene successivamente sottoposta la membrana, secondo le istruzioni del produttore. Infine il DNA verrà eluito con 30µl di Pure Link Genomic Elution Buffer

Il DNA genomico ottenuto è stato sottoposto ai controlli di qualità descritti nel paragrafo "Estrazione RNA, DNA e proteine da cellule CD4+ e CD8+ "

Selezione RNA polyA

L'RNA polyA è stato isolato dall'RNA totale tramite selezione positiva, attraverso due cicli di purificazione sequenziali con sonde costituite da sequenze di Timidine (poly-dT) attaccate a biglie magnetiche, come indicato nel manuale di preparazione delle librerie per RNA (TruSeq RNA Sample Preparation manual, Illumina # 15015050).

L'RNA è stato utilizzato per produrre la libreria di cDNA, attraverso il "TruSeq RNA Sample Preparation kit" (Illumina# FC-122-1001).

Preparazione librerie

Gli RNA sono stati processati utilizzando il kit TruSeq RNA-Seq Library Preparation (Illumina# FC-122-1001). Il campione è stato sottoposto a frammentazione chimica casuale mediante l'utilizzo di uno specifico buffer contenente cationi bivalenti a temperature elevate.

Il primo filamento di cDNA è stato sintetizzato usando oligonucleotidi casuali e la trascrittasi inversa SuperScript II (Invitrogen# 18064-014).

La sintesi del secondo filamento di cDNA è stata fatta in maniera sequenziale utilizzando DNA Polymerase I, quindi è stata utilizzata l'RNase H per eliminare l'RNA non retrotrascritto.

Successivamente i frammenti di cDNA sono stati selezionati per dimensione utilizzando le biglie di purificazione Agencourt AMPure XP (Beckman#A63882). I frammenti di DNA sono stati adenilati nella loro estremità 3', quindi sono legati a degli adattatori oligonucleotidici contenenti anche un index, tramite il kit Illumina TruSeq PE. In seguito alla ligazione, il campione è stato sottoposto ad una doppia purificazione e successivamente ad un arricchimento selettivo mediante PCR con il cocktail "Illumina PCR primer cocktail". Il prodotto della PCR costituisce la libreria stessa, e viene prima purificato tramite le biglie AMPure XP, poi quantificato attraverso lo strumento Bionalyzer 2100 e l'utilizzo del kit Agilent DNA 1000 assay (Agilent#5067-1504)

Le singole librerie saranno quindi identificate da un index. Prima di procedere con la generazione dei cluster e con il sequenziamento è stato creato un pool di librerie presenti in concentrazioni equimolari per ciascun campione.

Sequenziamento con HiSeq 2000

I *pool* di librerie sono stati ibridizzati e amplificati sulla superficie di uno specifico vetrino, detto *flow-cell*, mediante un processo di amplificazione a ponte "*bridge amplification*", formando i *clusters*, attraverso il cBot System (Illumina) e il kit TruSeq PE (Paired End) and Cluster Generation kit v3 (Illumina# PE-401-3001). Successivamente le librerie sulla *flow cell* sono state sequenziate tramite il sequenziatore l'HiSeq 2000 utilizzando i reagenti TruSeq SBS v3 reagents (Illumina# FC-401-3001).

La tecnologia TruSeq supporta sequenziamenti massivi paralleli e si basa sul rilevamento di singole basi nel momento in cui sono incorporate nel filamento di DNA in formazione. Alla fine della corsa vengono generate delle sequenze di 51bp *paired-end*, ovvero per ogni frammento di circa 200 nt che amplificato costituisce il cluster, vengono sequenziate 51bp

all'inizio e 51bp alla fine. Infine, attraverso la Pipeline Illumina, sono stati generati i file FASTQ per ogni libreria, dopo un processo chiamato “*demultiplexing*” che consente, sfruttando gli index di ogni campione, di suddividere le sequenze in modo appropriato.

Allineamento e controlli di qualità

Le librerie sono distribuite in più di una lane in ogni *flow-cell* e i file FASTQ ottenuti dalla stessa libreria derivanti dalle differenti lane vengono successivamente integrati in un unico file FASTQ. La grandezza dell'insero in ogni coppia di *read* di sequenza è stabilita allineando le prime 5 milioni di *reads* del file FASTQ sul trascrittoma annotato su ENSEMBL utilizzando bwa-0.5.9 (Li H, Durbin R.,2009) riconoscendo i duplicati di PCR e calcolando le metriche per le dimensioni dell'insero con PICARD - 1.57. Tutte le *reads* sono allineate sul genoma di riferimento (h37d5) (<http://www.1000genomes.org>) utilizzando i software STAR-2.2.0c (Dobin A. et al.,2009). Le sequenze *spike-ins* (ERCC) e il database di riferimento delle giunzioni esone-esone sono generate utilizzando l'annotazione di GENCODE.v14. I controlli di qualità sono eseguiti utilizzando SAMtools (samtools-0.1.18; Li, H. et al., 2009), EVER-seq(everseq-1.0.4, <http://code.google.com/p/ever-seq/>), eRSeQC (Wang, L. et al., 2012).

Annotazione, normalizzazione e stabilizzazione della varianza

I livelli di espressione genica sono stati determinati assegnando ad ogni gene la conta delle *reads* per mezzo del software HTSeq (Simon Anders et al., 2010) e l'annotazione GENCODE V14. Le conte sono poi state normalizzate trasformandole in FPKM, cioè la versione per *reads paired-end* della normalizzazione per RPKM (Mortazavi, A. et al., 2008). L'acronimo RPKM sta per *Reads Per Kilobase of transcript per Million mapped reads*, in quanto le *reads* mappanti su un trascritto vengono divise per la lunghezza dello stesso trascritto (espressa in migliaia di basi) e divise per il numero totale di *reads* sequenziate per il campione (espresso in milioni di *reads*). Nel caso di un esperimento *paired-end*, vengono sequenziate due *reads* per ogni frammento di RNA, ed in questo caso, le *reads* appartenenti allo stesso frammento vengono contate una sola volta, per cui si parla di “*fragments*” invece che *reads*, da cui il nome FPKM (*Fragments Per Kilobase of transcript per Million mapped reads*). In alternativa sono state anche normalizzate e stabilizzate per varianza con il software DESeq2 (Simon Anders et al., 2010).

PCA e clusterizzazione gerarchica- “*hierarchical clustering*”

La PCA è stata calcolata tramite la funzione `prcomp` di R. Gli “*hierarchical clustering*” sono stati determinati con la funzione `hclust` di R. Sono stati generati diversi tipi di clusterizzazione, differenti sia per la distanza che per i metodi di *linkage*. In particolare si è usata come distanza sia la distanza euclidea che $d = 1 - \text{correlazione di pearson}$. I metodi di *linkage* sono stati “Ward” e “*complete*”.

Determinazione dei geni differentemente espressi

Per la determinazione dei geni differentemente espressi è stato utilizzato il software DESeq2 (Simon Anders et al., 2010). Per determinare i geni che fossero differentemente espressi nei gemelli concordanti ma non nei discordanti, sono stati confrontati i modelli con e senza la condizione patologica.

Modello 1: Livello di espressione ~ famiglia + condizione patologica

Modello 2: Livello di espressione ~ famiglia

Il fattore bloccante “famiglia” consente di tener conto della relazione familiare tra gli individui. Il confronto tra i due modelli è stato fatto con il LRT (Likelyhood Ratio Test) che restituisce un *p-value* che indica quanto significativamente i due modelli differiscono. Il *p-value* è stato successivamente corretto per test multiplo con il metodo di Benjamini-Hochberg.

RISULTATI E DISCUSSIONE

Il progetto portato avanti nei 3 anni di Dottorato, ha lo scopo di indagare l'impatto relativo delle variabili genetiche ed epigenetiche nella suscettibilità al DT1.

Il disegno sperimentale si avvale della disponibilità di cellule CD4+ e CD8+ provenienti da 5 coppie di gemelli monozigoti di cui 3 concordanti e 2 discordanti per la malattia. Per soddisfare il disegno dello studio, il progetto è dunque stato strutturato in maniera multidisciplinare ed ha tre obiettivi principali:

1. lo studio del profilo trascrizionale
2. lo studio del repertorio TCR
3. lo studio del metiloma

Attualmente abbiamo completato il primo obiettivo, relativo allo studio del profilo trascrizionale. L'analisi del metiloma e del repertorio TCR sono invece in fase di completamento.

Analisi del profilo trascrizionale

Per lo studio del profilo trascrizionale abbiamo deciso di utilizzare il sequenziamento massivo di ultima generazione (*Next Generation Sequencing*) della frazione di RNA poliadenilato -PolyA(+)-, corrispondente in larga parte ai geni codificanti proteine e a lunghi RNA non codificanti (lncRNA).

Il sequenziamento dell'RNA (RNA-seq) permette di identificare e quantificare tutti i trascritti espressi in un campione biologico. Nelle analisi dei dati RNA-seq, l'abbondanza di un gene o di un trascritto è frequentemente espressa come "frammenti per kilobase di esone per milione di reads mappate" (FPKM). Il metodo FPKM permette la normalizzazione delle conte delle *reads* sia sulla lunghezza del gene che sulla dimensione della libreria (vedi METODI), rendendo così ogni gene meno dipendente da tali variabili. Ciò consente la comparazione sia tra i geni in uno stesso campione, che tra campioni differenti.

L'RNA-seq permette inoltre di poter identificare nuovi trascritti, di analizzare l'espressione differenziale e gli *splicing* alternativi. Il metodo fornisce una valutazione imparziale di tutti i trascritti, con un *range* dinamico, inteso come intervallo di livelli di espressione quantificabili, maggiore rispetto ai microarrays di espressione (Sims et al., 2014).

Mappaggio delle sequenze e controlli di qualità

L'RNA Poly(A)+ è stato utilizzato per la preparazione delle librerie, successivamente sequenziate utilizzando la tecnologia Illumina. La metodica ha permesso di sequenziare le 51 paia di basi (pb) di entrambe le regioni terminali (*ends*) dei frammenti lunghi circa 200 pb che costituivano la libreria e che amplificati sulla *flow cell* hanno generato dei “*cluster*”. Le sequenze (*reads*) così ottenute, che hanno superato tutti i controlli di qualità, sono state mappate sulla sequenza del genoma umano.

Per ogni libreria sono stati sequenziati circa (60 M) di clusters (Fig.10A), un livello di coverage paragonabile con la media degli esperimenti di espressione differenziale presenti in letteratura (Lappalainen et al., 2013).

Il consorzio “*Encyclopedia of DNA elements*” (ENCODE) ha infatti fornito i dati che hanno consentito di determinare il numero di *reads* necessario per quantificare accuratamente i geni nei vari *range* dinamici di espressione calcolati in FPKM, nelle cellule umane (ENCODE 2004). Hanno determinato che per geni espressi con valori maggiori di 10 FPKM, più dell'80% dei geni può essere accuratamente quantificato, nel 10% dei *dataset* al loro disposizione, usando 36M di *reads* mappate. Per geni espressi a bassi livelli (meno di 10 FPKM), sono invece necessarie circa 80M di *reads*.

La *coverage* di 60M rappresenta quindi un ottimo compromesso per l'identificazione della maggior parte dei geni anche meno espressi.

Il “*mapping rate*” è la frazione di *reads* che si allineano correttamente sul genoma, sul totale delle *reads* che hanno passato i controlli di qualità. Il *mapping rate* dei campioni sequenziati oscilla tra il 93 e il 95%. (Fig.10B)

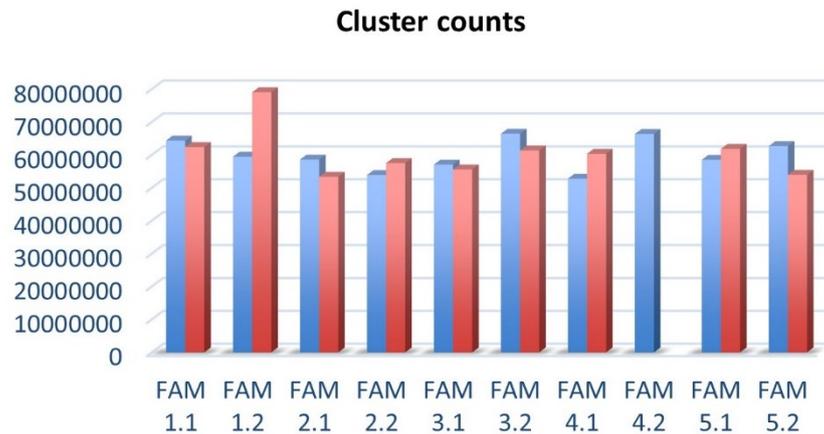
Le *reads* correttamente allineate sul genoma di riferimento sono state quindi classificate (FIG. 11A) in:

1. UsR: mappano in un'unica posizione e non presentano interruzioni (*splicing*)
2. MsR: mappano in diversi siti e non presentano interruzioni
3. USR: mappano in un'unica posizione e presentano interruzioni
4. MSR: mappano in diversi siti e presentano interruzioni

La composizione delle *reads* mostra l'andamento previsto per librerie provenienti da campioni RNA PolyA(+), con una percentuale di 81% di *reads* che mappano in un'unica posizione e che non presentano interruzioni (UsR), del 6% di *reads* che mappano in diversi siti e che non

presentano interruzioni (MsR) e del 1% di *reads* che mappano in diversi siti e che presentano interruzioni (MSR). Le *reads* che mappano in un'unica posizione e che presentano interruzioni (USR) sono il 12% permettono di identificare le giunzioni esone-esone e consentono di identificare le isoforme alternative per un dato gene (Fig.11B,C)

A



B

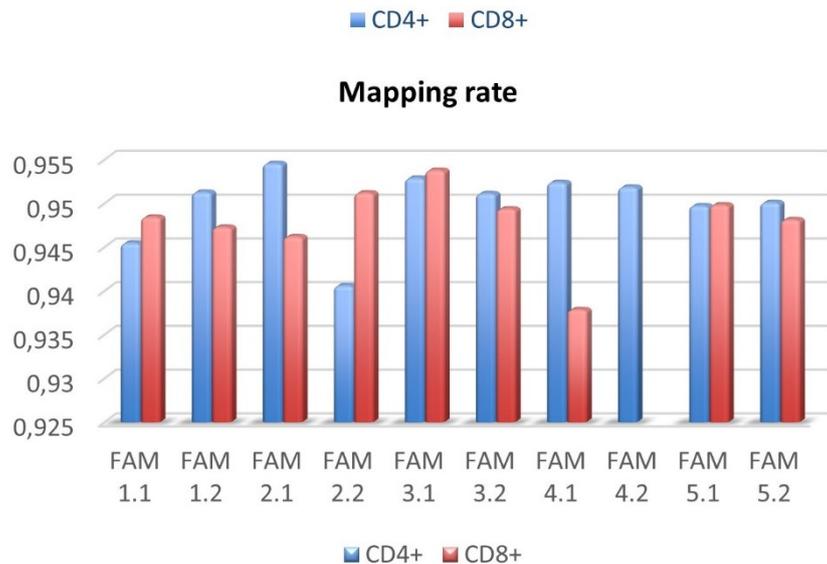


Figura 10. Conta dei cluster e mapping rate. In (A) sono mostrati i numeri di cluster ottenuti per ogni campione. In (B) è riportata la frazione di *reads* mappanti correttamente sul genoma sul totale delle *reads* ottenute, per ogni campione. Per quanto riguarda l'individuo FAM 4.2, non era disponibile la frazione CD8+.

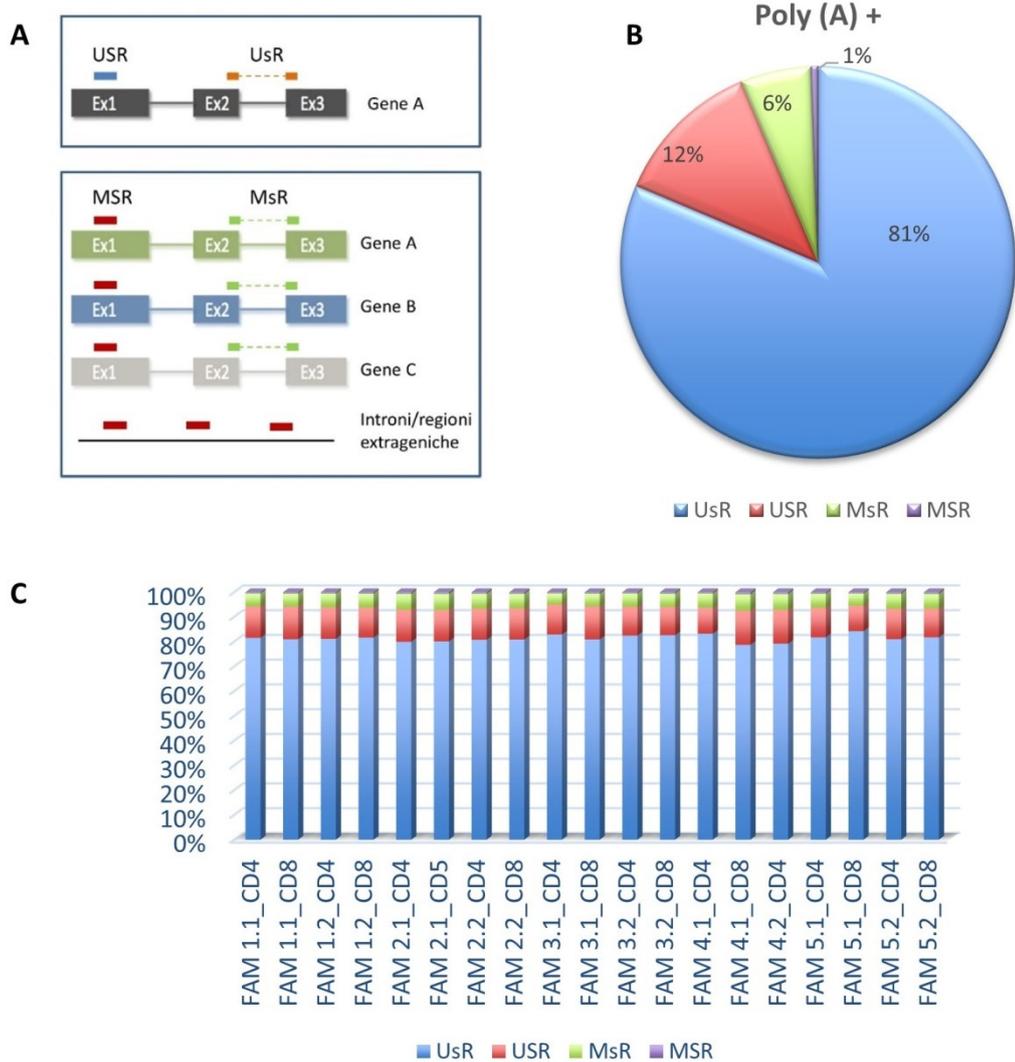


Figura 11. Distribuzione delle reads mappate. In (A) sono schematizzate le reads *UsR*, *UsR*, *MSR*, *MsR* secondo la classificazione data nel testo. Il rettangolo descrive la reads, i quadratini corrispondenti a metà rettangolo, indicano la reads “spezzata” in quanto mappante in punti separati nell’allineamento sul genoma dalla presenza dell’introne. *UsR*, unico rettangolo blu, *UsR*, unico rettangolo arancione spezzato e collegato da tratteggio, le *MSR* multipli rettangoli rossi, *MsR* multipli rettangoli verdi spezzati verdi. In (B) sono rappresentati le percentuali di *UsR*, *UsR*, *MSR*, *MsR* calcolate sulla media di tutti i campioni sequenziati. In (C) è rappresentata la distribuzione delle reads per ogni campione, espressa come percentuale sul totale di reads mappanti per campione (in ordinate).

Clusterizzazione gerarchica

La “clusterizzazione” o analisi dei gruppi (dal termine inglese “*cluster analysis*” introdotto da Robert Tryon nel 1939) costituisce un insieme di tecniche di analisi multivariata dei dati, volta alla selezione e al raggruppamento di elementi omogenei in un insieme di dati. Le tecniche di clusterizzazione si basano su misure relative alla somiglianza tra gli elementi. Per il calcolo della clusterizzazione gerarchica sui livelli di espressione genica, sono stati utilizzati metodi e distanze diversi.

Di seguito riportiamo i risultati della clusterizzazione gerarchica ottenuti con il metodo “Ward” ampiamente usato sin dalla sua prima descrizione pubblicata nel 1963.

Tale metodo, usato per variabili quantitative, è diretto alla minimizzazione della varianza all’interno dei gruppi. Ad ogni passo questo algoritmo tende ad ottimizzare la partizione ottenuta tramite l’aggregazione di due elementi. Una partizione si considera tanto migliore quanto più le classi risultano omogenee al loro interno e differenti l’una dall’altra. In altri termini, quanto più è elevata la varianza tra le classi, tanto più è bassa la varianza interna (alle classi).

La distanza è stata calcolata usando la metrica Euclidea.

I risultati dell’analisi mostrano che i campioni clusterizzano in due gruppi: la Famiglia 4 e la famiglia 1 clusterizzano in un gruppo, la Famiglia 2 and 5 clusterizzano in un altro gruppo, e la Famiglia 3, invece, risulta divisa tra i due sottogruppi (Fig.12).

Il risultato ottenuto indica come improbabile che la clusterizzazione dei campioni sia attribuibile allo *status* malattia. Inoltre suggerisce che l’impatto della patologia sull’espressione genica non sia maggiore del background genetico/familiare o della composizione cellulare.

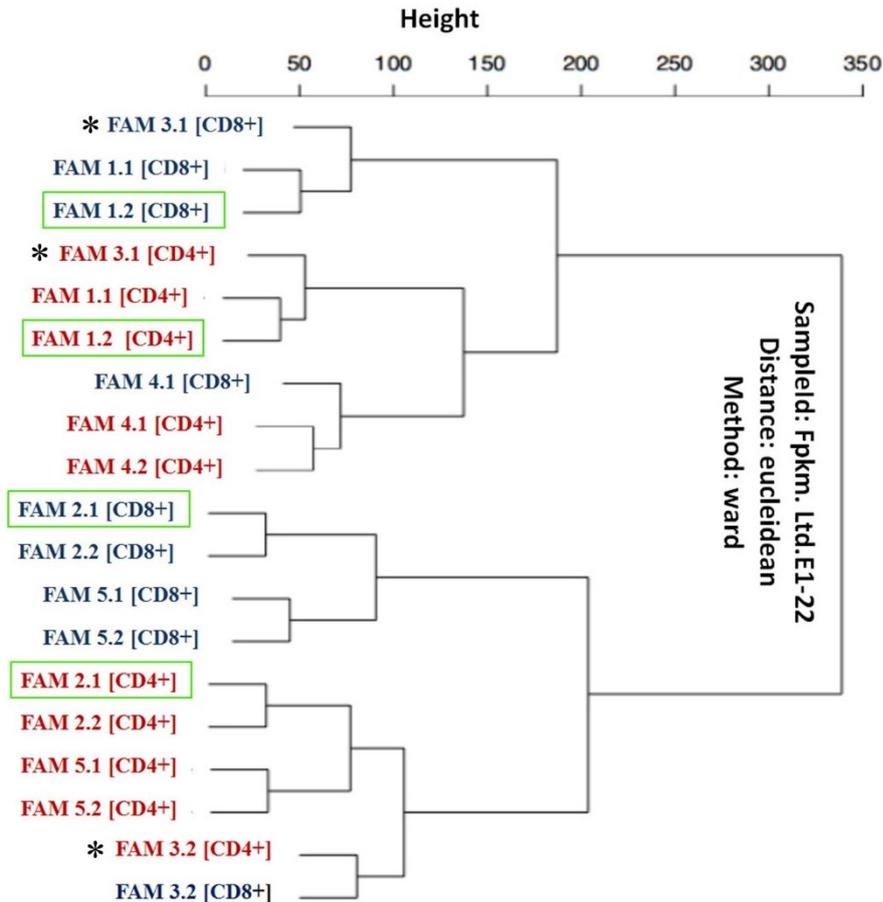


Figura 12. Clusterizzazione gerarchica. I box verdi evidenziano i campioni provenienti da individui sani. In rosso sono mostrati i campioni CD4+, in azzurro i CD8+. L'asterisco indica gli individui che nella clusterizzazione mostrano un comportamento anomalo.

Analisi dell'espressione differenziale

I livelli di espressione genica (somma di tutti i segnali di espressione provenienti da tutte le rispettive isoforme per ogni gene) sono stati determinati attraverso la conta delle *reads* che mappano nei geni (mediante il software HTSeq) seguita dalla normalizzazione dei livelli di espressione tra i vari campioni.

Una volta stabiliti i livelli di espressione dei vari trascritti, siamo andati a selezionare i geni che risultavano espressi in modo differenziale tra i gemelli nell'ambito di ogni coppia. Tra i geni così ottenuti, siamo andati a selezionare i geni differenzialmente espressi nelle coppie discordanti per malattia (FAM 1 e FAM 2), la cui differenza di espressione fosse maggiore che nelle coppie concordanti (FAM 3, 4, 5)

Per spiegare meglio questo approccio riportiamo nella Figura 13 l'esempio per il gene CD86. In questo approccio, come misura dell'espressione sono stati utilizzati, per ogni campione, i dati normalizzati e stabilizzati per la varianza tramite Deseq (VSD).

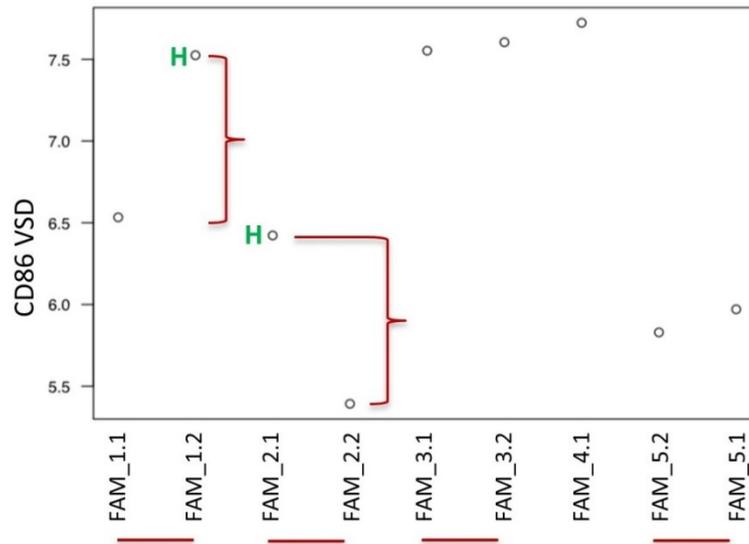


Figura 13. Differenze di espressione rappresentate in VSD per il gene CD86, per ogni individuo. Le differenze sono maggiori tra gli individui delle coppie discordanti, come evidenziato dalle graffette rosse.

Abbiamo identificato, solo nelle cellule CD8+, 9 geni che rispettassero tali requisiti. Le differenze di espressione per ogni gene sono riportate, nella Tabella 1, in “*Fold Change*”, e il segno negativo indica una ridotta espressione negli individui affetti. Per quanto riguarda le cellule CD4+, non abbiamo identificato nessuna differenza significativa di espressione.

Gene	Fold Change	pvalue
VENTX	-2.08	3.83E-07
HLA-DRB1	-1.05	2.67E-06
EPB41L3	-1.15	1.24E-05
RNF217	-1.76	1.21E-05
HLA-DRB5	-0.96	1.29E-05
HLA-DRA	-1.22	6.97E-06
PTPN13	1.69	1.62E-05
DERL3	-1.27	2.99E-05
CD86	-1.26	2.85E-05

Tabella 1: Geni differenzialmente espressi nelle coppie discordanti per DT1, ma non nelle coppie concordanti.

Abbiamo quindi consultato il catalogo GWAS (Welter et al., 2014) per controllare se esistessero associazioni già descritte dei nostri geni con il DT1 o con altre patologie autoimmuni.

E' emerso che gli unici geni trovati associati al DT1 tra i 9 identificati dal nostro studio, sono il CTLA-4 e l'HLA-DRB1. Il gene CD86 risulta invece associato ad un'altra importante malattia autoimmune, la sclerosi multipla. Per i geni rimanenti non sono state descritte associazioni con alcun tipo di patologia.

Abbiamo quindi compiuto una ricerca bibliografica e, di seguito, riportiamo una breve descrizione per ogni gene identificato:

CD86 (# MIM 601020): In letteratura, polimorfismi nel gene CD86 sono stati associati alla sclerosi multipla (Sawcer et al., 2011; Patsopoulos et al., 2011)

Il gene codifica per una proteina di membrana di tipo I che è un membro della superfamiglia delle immunoglobuline. Il CD86 è espresso sulle cellule che presentano l'antigene (APC) e fornisce segnali di co-stimolazione necessari nei primi eventi di attivazione delle cellule T e nella loro sopravvivenza. E' un ligando per due diverse proteine presenti sulla superficie delle cellule T: CD28 e CTLA-4.

CTLA-4 (# MIM 123890): Il *CTLA-4* (*Cytotoxic T Lymphocyte associated Antigen 4*) è un gene associato al DT1. Il gene si trova nel locus IDDM12 in posizione 2q33, già identificato nel 1995 come associato alla malattia (Copeman et al., 1995, Cucca et al. 1995; Nistico et al., 1996). Il suo ruolo nella suscettibilità del DT1 e di altre malattie autoimmuni è stato confermato nel 2003 (Ueda et al., 2003) E' un inibitore della proliferazione dei linfociti T, promuove l'apoptosi delle cellule T, attraverso l'incremento dell'espressione dell'interleuchina 2 (IL2).

DERL3 (# MIM 610305): In studi GWAS, il gene *DERL3* (*Der1-like domain family, member 3*) è stato trovato associato alla glicosilazione delle Immunoglobuline G (IgG) (Lauc et al., 2013). Si ritiene che la glicosilazione delle immunoglobuline influenzi il comportamento delle cellule immunitarie, e in particolare la discriminazione tra cellule self e non self. Una diminuzione della glicosilazione delle IgG è una caratteristica riscontrata già in letteratura come implicata in patologie autoimmuni come l'artrite reumatoide (Parekh et al., 1985). La proteina Derlin-3, codificata da questo gene, sembra essere coinvolta nella degradazione di glicoproteine mal ripiegate nel reticolo endoplasmatico (Oda et al., 2006).

PTPN13 (# MIM 600267): PTPN13 (*protein-tyrosine phosphatase, nonreceptor-type, 13*) è un membro della famiglia delle proteine tirosina fosfatasi (PTP), note per essere molecole di segnalazione che regolano una varietà di processi cellulari tra cui la crescita e la differenziazione delle cellule, il ciclo mitotico, e la trasformazione oncogenica. Diversi studi GWAS riportano l'associazione di questo gene in letteratura, con le variazioni dei livelli sierici di dimetilarginina, un prodotto derivato dalla proteolisi (Seppälä et al., 2014; Lüneburg et al., 2014). Appare quindi plausibile che il gene PTPN13 moduli i livelli della stessa dimetilarginina, la cui attività è a sua volta negativamente correlata con iperglicemia in età pediatrica (Takaya J et al., 2015) e correlata con insulino-resistenza in pazienti con artrite reumatoide (Dimitroulas et al., 2015). E' espressa nella maggior parte dei tessuti, ad eccezione del fegato e del muscolo scheletrico, ed è più abbondante nel polmone, nel rene e nel cervello fetale. Elevati livelli di espressione PTPN13 sono stati segnalati in vari tipi di tumori umani, tra cui del colon, del pancreas, epatocellulare, ematologico e in tumori ovarici.

VENTX (#MIM 607158): La proteina codificata dal gene *VENTX* (*vent homeobox*) è un repressore trascrizionale ed è coinvolta nella formazione del mesoderma e nel mantenimento delle cellule staminali emopoietiche.

Inoltre, VENTX è un regolatore trascrizionale fondamentale nel processo di differenziazione e maturazione delle cellule dendritiche, e un potenziale target di regolazione e terapia immunitaria (Wu et al., 2014).

Nelle leucemie, l'espressione del gene VENTX è significativamente repressa nella leucemia linfocitica cronica e indotta nella leucemia mieloide acuta umana (Rawat et al., 2010)

EPB41L3 (#MIM 605331): Il gene *EPB41L3* codifica per una proteina di membrana chiamata "*Erythrocyte membrane protein band 4.1-like 3*", che localizza in regioni di contatto cellula-cellula e che funziona come una proteina adattatrice, collegando la membrana plasmatica al citoscheletro e a effettori citoplasmatici, e facilitando la loro attività in vari ambiti, tra i quali quelli della motilità e dell'adesione cellulare. EPB41L3 svolge un ruolo regolatore nella crescita e nella differenziazione cellulare, e nella creazione di strutture simili a epitelii ("epithelial-like") (Wang et al., 2014).

RNF217 (#MIM non disponibile): Il gene *RNF217* codifica per una E3 ubiquitin-protein ligase. La sua over-espressione è stata riscontrata in alcune leucemie umane (Fontanari et al., 2014).

HLA-DRB1, HLA-DRB5 e HLA-DRA: I tre geni *HLA* differenzialmente espressi nello studio sono già stati descritti in letteratura come implicati nell'insorgenza di malattie autoimmuni.

In particolare, alcune varianti nel gene *HLA-DRB1* sono associate ad un aumentato rischio di sviluppare malattie autoimmuni, tra cui la malattia autoimmune di Addison, la sclerosi multipla e l'artrite reumatoide (MacLaren et al., 1986, International Multiple Sclerosis Genetics Consortium et al., 2011, Kochi et al., 2010). Al contrario, altre varianti sembrano diminuire il rischio di sviluppare artrite reumatoide (Kochi et al., 2010).

E' inoltre noto che combinazioni di variazioni nel gene *HLA-DRB1* e altri geni *HLA* influenzano il rischio di DT1 (Hu et al., 2015). Tenendo conto del ruolo che il gene esercita a livello immunitario, esistono in letteratura diverse correlazioni già descritte con varie malattie autoimmuni, tra le quali la malattia di Graves (Kula et al., 2006), la tiroidite di Hashimoto (Kong et al., 1996), l'artrite giovanile idiopatica (Yanagimachi et al., 2011) e l'artrite psoriasica (Ho et al., 2008). Il gene *HLA-DRA* agisce come unica catena α per le proteine codificate dai geni *DRB1*, *DRB3*, *DRB4* e *DRB5*. Ad esso risultano associate malattie come includono la sindrome "Graham little-piccardi-lassueur" (Rodriguez et al., 2007) e la MHC class II deficiency (Hauber et al., 1995).

RISULTATI PRELIMINARI E PROSPETTIVE FUTURE

Analisi del repertorio dei recettori delle cellule T (TCR)

Per l'analisi del repertorio TCR, abbiamo allestito un protocollo sviluppato dal gruppo diretto dal prof. Dmitry Chudakov (Mamedov et al., 2013).

Il metodo si basa sull'amplificazione dell'mRNA delle catene α e β del TCR prima del sequenziamento delle librerie. Questo approccio permette l'amplificazione efficiente e imparziale (*unbiased*) di milioni di molecole di mRNA dei TCR, fornendo quindi materiale a sufficienza per il sequenziamento tramite la tecnologia NGS. L'analisi dei dati sarà eseguita utilizzando il metodo MIGEC (*Molecular Identifier Groups-based Error Correction*), un

software basato su una tecnologia studiata appositamente per l'analisi dei dati di sequenziamento ad alta produttività (*high-throughput*). MIGEC dispone di una strategia per la correzione degli errori tecnici preservando la naturale diversità e complessità dei repertori immunologici (Shugay et al, 2014).

Analisi del metiloma

Per lo studio del profilo di metilazione abbiamo scelto di utilizzare la metodica “*Reduced Representation Bisulfite Sequencing*” (RRBS) che consente di analizzare circa il 10% del genoma metilato, con risoluzione a livello di singolo nucleotide (Gu et al, 2011).

La metodica RRBS è tuttora nella fase di messa a punto. Le prove tecniche condotte ci hanno permesso di ottimizzare e standardizzare la metodica proposta nel lavoro scientifico pubblicato da Boyle et al. nel 2012.

Un punto chiave nella messa a punto è il calcolo dell'efficienza di conversione da parte del bisolfito. Una bassa efficienza di conversione infatti produrrebbe falsi positivi per la metilazione, poiché la base azotata non metilata risulterebbe, in fase di sequenziamento, non convertita così come una base metilata, e quindi verrebbe letta in entrambi i casi come citosina. Finora abbiamo eseguito il calcolo dell'efficienza di conversione valutando lo stato di conversione del nucleotide citosina, non metilato, aggiunto nella fase di riparazione del DNA successiva alla digestione enzimatica. I livelli di conversione minimi finora riscontrati superano ampiamente il 99%.

CONCLUSIONI

Nonostante gli studi GWAS, il sequenziamento dell'intero genoma e la sua caratterizzazione, lo sviluppo di metodi statistici sempre più potenti abbiano permesso di identificare oltre 60 *loci* di suscettibilità per il DT1, una grossa parte dell'ereditabilità della malattia rimane inspiegata.

Lo studio proposto in questa tesi ha dimostrato che l'utilizzo di coppie di gemelli monozigoti concordanti e discordanti per il DT1, unito all'utilizzo di tecniche di sequenziamento di ultima generazione (NGS) e in particolare al sequenziamento dell'RNA (RNA-seq), costituiscono un potenziale enorme per lo studio delle malattie complesse immuno-mediate.

La "clusterizzazione" o analisi dei gruppi ha mostrato che il raggruppamento dei campioni non è dovuto allo *status malattia* e che l'impatto della patologia sull'espressione genica non è maggiore rispetto a quello dovuto al background genetico/familiare o alla composizione cellulare. Sono tuttavia necessarie ulteriori analisi volte alla determinazione delle reali cause della clusterizzazione dei campioni, escludendo eventuali *bias* tecnici o biologici.

Inoltre lo studio del profilo di espressione ci ha consentito di identificare in cellule CD8+, 9 geni differenzialmente espressi tra coppie di gemelli discordanti ma non tra coppie concordanti.

Tra questi geni, il ruolo nel DT1 era stato già descritto per l'*HLA-DRB1*. Geni come il *CD86*, o *HLA-DRB5* e *HLA-DRA* hanno dei chiari ruoli nel sistema immunitario, quindi non è sorprendente averli trovati deregolati in questo studio. Gli altri geni apparentemente non hanno alcun legame con la malattia, ma dopo aver confermato la loro espressione differenziale in una casistica più ampia sarà importante procedere alla loro caratterizzazione funzionale nel contesto del DT1.

Studi di questo tipo, dove i dati vengono generati e analizzati in maniera completamente "oggettiva", hanno il vantaggio di permettere di accendere una luce su *pathways* e meccanismi difficilmente ipotizzabili grazie alle relativamente poche conoscenze sulla grande complessità biologica alla base dei meccanismi che possono portare all'insorgenza, alla progressione o a determinati fenotipi della malattia.

Quando avremo a disposizione anche i risultati dell'analisi del metiloma e del repertorio TCR, potremo combinarli con i risultati dell'analisi del profilo trascrizionale.

Questo rappresenterà un approccio sistematico utile per la comprensione della patogenesi del DT1, e sarà il primo studio di questo tipo sul DT1.

In conclusione, il nostro studio potrebbe contribuire ad approfondire aspetti ancora sconosciuti della patologia verso una migliore conoscenza della sua eziologia e dello sviluppo di metodi sempre più efficaci per la diagnosi, il trattamento e la prevenzione della stessa.

RINGRAZIAMENTI

Desidero ringraziare il Prof. Francesco Cucca direttore dell'Istituto di Ricerca Genetica e Biomedica, relatore di questa tesi, per la disponibilità e cortesia dimostrata in questi tre anni di Dottorato e l'aiuto fornito durante la stesura della tesi.

Un ringraziamento particolare va, alla Dott.ssa Mara Marongiu, che ha sempre messo a mia completa disposizione la sua conoscenza ed esperienza dimostrandosi un'eccellente guida per la mia formazione scientifica e non solo: co-tutor di questa tesi di Dottorato, collega e amica con cui ho condiviso gioie e dolori in questo percorso e che mi ha sempre incoraggiato a dare il meglio di me stessa.

Un grazie va a tutti i colleghi con cui ho collaborato a questo progetto e in particolare alla Dott.ssa Francesca Crobu sempre e totalmente disponibile ad aiutarmi e a sostenermi moralmente.

Inoltre intendo ringraziare tutti i colleghi del Laboratorio 1: Laura mia co-tutor durante il primo anno di Dottorato, Manuela, Manila, Angela e Cristian che mi hanno sempre saputo consigliare e consolare nei momenti più difficili.

Un ultimo ringraziamento, ma non meno importante, va alla mia famiglia che ha sempre creduto in me e che mi ha aiutato a raggiungere questo importante traguardo.

Loredana

BIBLIOGRAFIA

1. Anderson MS, Bluestone JA. The NOD mouse: a model of immune dysregulation. *Annu Rev Immunol*. **23**:447-85 (2005)
2. Atkinson MA, Bluestone JA, Eisenbarth GS, Hebrok M, Herold KC, Accili D, Pietropaolo M, Arvan PR, Von Herrath M, Markel DS, Rhodes CJ. How does type 1 diabetes develop?: the notion of homicide or beta-cell suicide revisited. *Diabetes*; **60**: 1370–79 (2011)
3. Atkinson MA, Eisenbarth GS, Michels AW. Type 1 diabetes. *Lancet*. **4**; 383(9911):69-82 (2014).
4. Bell GI, Horita S, Karam JH. A polymorphic locus near the human insulin gene is associated with insulin-dependent diabetes mellitus. *Diabetes*. **33(2)**:176-83 (1984)
5. Bluestone JA, Herold K, Eisenbarth G. Genetics, pathogenesis and clinical interventions in type 1 diabetes. *Nature*. **464(7293)**:1293-300 (2010)
6. Borg WP, Sherwin RS. Classification of diabetes mellitus. *Adv Intern Med*. **45**:279-95 (2000)
7. Bottini N, Musumeci L, Alonso A, Rahmouni S, Nika K, Rostamkhani M, MacMurray J, Meloni GF, Lucarelli P, Pellecchia M, Eisenbarth GS, Comings D, Mustelin T. A functional variant of lymphoid tyrosine phosphatase is associated with type I diabetes. *Nat Genet*. **36(4)**:337-8 (2004)
8. Boyle P, Clement K, Gu H, Smith ZD, Ziller M, Forster JL, Holmes L, Meldrum J, Kelley F, Gnirke A, Meissner A. Gel-free multiplexed reduced representation bisulfite sequencing for large-scale DNA methylation profiling. *Genome Biol*. **13(10)**:R92 (2012)
9. Burton AR, Vincent E, Arnold PY, Lennon GP, Smeltzer M, Li CS, Haskins K, Hutton J, Tisch RM, Sercarz EE, Santamaria P, Workman CJ, Vignali DA. On the pathogenicity of autoantigen-specific T-cell receptors. *Diabetes* **57**, 1321–1330 (2008).
10. Copeman JB, Cucca F, Hearne CM, Cornall RJ, Reed PW, Rønningen KS, Undlien DE, Nisticò L, Buzzetti R, Tosi R, et al. Linkage disequilibrium mapping of a type 1 diabetes susceptibility gene (IDDM7) to chromosome 2q31-q33. *Nat Genet*. **9(1)**:80-5 (1995)
11. Cucca F, Lampis R, Congia M, Angius E, Nutland S, Bain SC, Barnett AH, Todd JA. A correlation between the relative predisposition of MHC class II alleles to type 1 diabetes and the structure of their proteins. *Hum Mol Genet*. **10(19)**:2025-37 (2001)
12. Cucca F, Lampis R, Frau F, et al. The distribution of DR4 haplotypes in Sardinia suggests a primary association of type I diabetes with DRB1 and DQB1 loci. *Hum Immunol*. **43**:301-8 (1995)
13. Cudworth AG & Woodrow JC (1975). Evidence for HL-A-linked genes in "juvenile" diabetes mellitus. *Br Med J*. **3(5976)**, 133-5 (1975)
14. Dahlquist GG, Nyström L, Patterson CC; Swedish Childhood Diabetes Study Group; Diabetes Incidence in Sweden Study Group. Incidence of type 1 diabetes in Sweden among individuals aged 0-34 years, 1983-2007: an analysis of time trends. *Diabetes Care* **34(8)**:1754-9 (2011)
15. D'Alise AM, Auyeung V, Feuerer M, Nishio J, Fontenot J, Benoist C, Mathis D. The defect in T-cell regulation in NOD mice is an effect on the T-cell effectors. *Proc. Natl Acad. Sci. USA* **105**, 19857–19862 (2008)
16. Dimitroulas T, Hodson J, Sandoo A, Smith J, Kitas GD. Symmetric dimethylarginine (SDMA) serum levels in rheumatoid arthritis: correlations with insulin resistance and disease activity scores. *Amino Acids*. **47(9)**:1995-2004 (2015)
17. Dobin, A. et al. STAR: ultra fast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2012)
18. Eisenbarth GS. Type I diabetes mellitus. A chronic autoimmune disease. *N Engl J Med*. **22**;314(21):1360-8 (1986)
19. ENCODE Project Consortium. The ENCODE (ENCyclopedia of DNA elements) project. *Science* **306**, 636–640 (2004)
20. EURODIAB ACE Study Group: Variation and trends in incidence of childhood diabetes in Europe. *Lancet* **355(9207)**:873–876 (2000)
21. Fontanari Krause LM, Japp AS, Krause A, Mooster J, Chopra M, Müschen M, Bohlander SK. Identification and characterization of OSTL (RNF217) encoding a RING-IBR-RING protein adjacent to a translocation breakpoint involving ETV6 in childhood ALL. *Sci Rep*. **4**:6565 (2014)
22. Fozza C, Contini S, Corda G, Virdis P, Galleu A, Bonfigli S, Pacifico A, Maioli M, Mastinu F, Pitzalis M, Cucca F, Longinotti M. T-cell receptor repertoire analysis in monozygotic twins concordant and discordant for type 1 diabetes. *Immunobiology*. **217(9)**:920-5 (2012)

Dott.ssa Loredana Marcia. Titolo tesi: Alla ricerca delle basi molecolari del Diabete di Tipo I: sequenziamento di ultima generazione in gemelli monozigoti concordanti e discordanti per la malattia. Tesi di Dottorato in Scienze Biomediche. Indirizzo: Genetica Medica, Malattie Metaboliche e Nutrigenomica. Ciclo XXVIII. Università degli Studi di Sassari.

23. Francalacci P, Morelli L, Angius A, Berutti R, Reinier F, Atzeni R, Pilu R, Busonero F, Maschio A, Zara I, et al., Low-pass DNA sequencing of 1200 Sardinians reconstructs European Y-chromosome phylogeny. *Science*. **341**(6145):565-9 (2013)
24. Gale EA, Gillespie KM: Diabetes and gender. *Diabetologia* **44**(1):3–15 (2001).
25. Galton, F. (1875) *Frazer's Magazine* 12, 566–576 (1875)
26. Gardner JM, Fletcher AL, Anderson MS, Turley SJ. AIRE in the thymus and beyond. *Curr. Opin. Immunol.* **21**, 582–589 (2009)
27. Gu H, Smith ZD, Bock C, Boyle P, Gnirke A, Meissner A. Preparation of reduced representation bisulfite sequencing libraries for genome-scale DNA methylation profiling. *Nat Protoc.* **6**(4):468-81 (2011)
28. Gulwani-Akolkar B, Shalon L, Akolkar PN, Fisher SE, Silver J. Analysis of the peripheral blood T-cell receptor (TCR) repertoire in monozygotic twins discordant for Crohn's disease. *Autoimmunity* **17**(3):241-8 (1994)
29. Harjutsalo V, Reunanen A, Tuomilehto J. Differential transmission of type 1 diabetes from diabetic fathers and mothers to their offspring. *Diabetes* **55**(5):1517-24 (2006)
30. Hauber I, Gulle H, Wolf HM, Maris M, Eggenbauer H, Eibl MM. Molecular characterization of major histocompatibility complex class II gene expression and demonstration of antigen-specific T cell response indicate a new phenotype in class II-deficient patients. *J Exp Med.* **181**(4):1411-23 (1995)
31. Ho PY, Barton A, Worthington J, Plant D, Griffiths CE, Young HS, Bradburn P, Thomson W, Silman AJ, Bruce IN. Investigating the role of the HLA-Cw*06 and HLA-DRB1 genes in susceptibility to psoriatic arthritis: comparison with psoriasis and undifferentiated inflammatory arthritis. *Ann Rheum Dis.* **67**(5):677-82 (2008)
32. Howson JM, Rosinger S, Smyth DJ, Boehm BO; ADBW-END Study Group, Todd JA. Genetic analysis of adult-onset autoimmune diabetes. *Diabetes.* **60**(10):2645-53 (2011)
33. Hu X, Deutsch AJ, Lenz TL, Onengut-Gumuscu S, Han B, Chen WM, Howson JM, Todd JA, de Bakker PI, Rich SS, Raychaudhuri S. Additive and interaction effects at three amino acid positions in HLA-DQ and HLA-DR molecules drive type 1 diabetes risk. *Nat Genet.* **47**(8):898-905 (2015)
34. International HapMap Consortium. The International HapMap Project. *Nature* **426**(6968):789-96 (2003)
35. International Multiple Sclerosis Genetics Consortium; Wellcome Trust Case Control Consortium 2, Sawcer S, Hellenthal G, Pirinen M, Spencer CC, Patsopoulos NA, Moutsianas L, Dilthey A, Su Z, et al. A. Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. *Nature* **476**(7359):214-9 (2011)
36. Kaminsky ZA, Tang T, Wang SC, Ptak C, Oh GH, Wong AH, Feldcamp LA, Virtanen C, Halfvarson J, Tysk C, McRae AF, Visscher PM, Montgomery GW, Gottesman II, Martin NG, Petronis A. DNA methylation profiles in monozygotic and dizygotic twins. *Nat Genet.* **41**(2):240-5 (2009)
37. Karvonen M, Viik-Kajander M, Moltchanova E, Libman I, LaPorte R, Tuomilehto J. Incidence of childhood type 1 diabetes worldwide. Diabetes Mondiale (DiaMond) Project Group. *Diabetes Care* **23**(10):1516-26 (2000)
38. Kent SC, Chen Y, Bregoli L, Clemmings SM, Kenyon NS, Ricordi C, Hering BJ, Hafler DA. Expanded T cells from pancreatic lymph nodes of type 1 diabetic subjects recognize an insulin epitope. *Nature* **435**: 224–28 (2005).
39. Kobayashi M, Jasinski J, Liu E, Li M, Miao D, Zhang L, Yu L, Nakayama M, Eisenbarth GS. Conserved T cell receptor alpha-chain induces insulin autoantibodies. *Proc. Natl Acad. Sci. USA* **105**, 10090–10094 (2008)
40. Kochi Y, Suzuki A, Yamada R, Yamamoto K. Ethnogenetic heterogeneity of rheumatoid arthritis-implications for pathogenesis. *Nat Rev Rheumatol.* **6**(5):290-5. (2010)
41. Koczwara K, Bonifacio E, Ziegler AG. Transmission of maternal islet antibodies and risk of autoimmune diabetes in offspring of mothers with type 1 diabetes. *Diabetes.* **53**(1):1-4. (2004)
42. Kong YC, Lomo LC, Motte RW, Giraldo AA, Baisch J, Strauss G, Hämmerling GJ, David CS. HLA-DRB1 polymorphism determines susceptibility to autoimmune thyroiditis in transgenic mice: definitive association with HLA-DRB1*0301 (DR3) gene. *J Exp Med.* **184**(3):1167-72 (1996)
43. Kula D, Bednarczuk T, Jurecka-Lubieniecka B, Polanska J, Hasse-Lazar K, Jarzab M, Steinhof-Radwanska K, Hejduk B, Zebracka J et al., Interaction of HLA-DRB1 alleles with CTLA-4 in the predisposition to Graves' disease: the impact of DRB1*07. *Thyroid.* **16**(5):447-53 (2006)
44. Kuwana M, Feghali CA, Medsger TA Jr, Wright TM. Autoreactive T cells to topoisomerase I in monozygotic twins discordant for systemic sclerosis. *Arthritis Rheum.* **44**(7):1654-9 (2001)

45. Lappalainen T, Sammeth M, Friedländer MR, t Hoen PA, Monlong J, Rivas MA, González-Porta M, Kurbatova N, Griebel T, Ferreira PG et al., Transcriptome and genome sequencing uncovers functional variation in humans. *Nature* **501**(7468):506-11 (2013)
46. Lauc G, Huffman JE, Pučić M, Zgaga L, Adamczyk B, Mužinić A, Novokmet M, Polašek O, Gornik O, Krištić J et al Loci associated with N-glycosylation of human immunoglobulin G show pleiotropy with autoimmune diseases and haematological cancers. *PLoSGenet.* **9**(1):e1003225 (2013)
47. Lennon GP, Bettini M, Burton AR, Vincent E, Arnold PY, Santamaria P, Vignali DA. T cell islet accumulation in type 1 diabetes is a tightly regulated, cell-autonomous event. *Immunity.* **31**, 643–653 (2009)
48. Leslie RD, Kolb H, Schloot NC, Buzzetti R, Mauricio D, De Leiva A, Yderstraede K, Sarti C, Thivolet C, Hadden D, Hunter S, Scherthner G, Scherbaum W, Williams R, Pozzilli P. Diabetes classification: grey zones, sound and smoke: Action LADA 1. *Diabetes Metab Res Rev.* **24**(7):511-9 (2008)
49. Leslie RD. Predicting adult-onset autoimmune diabetes: clarity from complexity. *Diabetes.* **59**(2):330-1 (2010)
50. Levin L, Tomer Y: The etiology of autoimmune diabetes and thyroiditis: evidence for common genetic susceptibility. *Autoimmun Rev.* **2**:377 (2003)
51. Li YR, Zhao SD, Li J, Bradfield JP, Mohebnasab M, Steel L, Kobie J, Abrams DJ, Mentch FD, Glessner JT et al., Genetic sharing and heritability of paediatric age of onset autoimmune diseases. *Nat Commun.* **6**:8442 (2015)
52. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009)
53. 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* **15**;25(16):2078-9 (2009)
54. Lieberman, S. M. & Di Lorenzo, T. P. A comprehensive guide to antibody and T-cell responses in type 1 diabetes. *Tissue Antigens* **62**, 359–377 (2003)
55. Liese AD, D'Agostino RB, Jr., Hamman RF, et al.: The burden of diabetes mellitus among US youth: prevalence estimates from the SEARCH for Diabetes in Youth Study. *Pediatrics* **118**(4):1510–1518 (2006)
56. Lüneburg N, Lieb W, Zeller T, Chen MH, Maas R, Carter AM, Xanthakis V, Glazer NL, Schwedhelm E, Seshadri S et al., Genome-wide association study of L-arginine and dimethylarginines reveals novel metabolic pathway for symmetric dimethylarginine. *CircCardiovasc Genet.* **7**(6):864-72 (2014)
57. Maclaren NK, Riley WJ. Inherited susceptibility to autoimmune Addison's disease is linked to human leukocyte antigens-DR3 and/or DR4, except when associated with type I autoimmune polyglandular syndrome. *J ClinEndocrinolMetab.* **62**(3):455-9 (1986)
58. Makino S, Kunitomo K, Muraoka Y, Mizushima Y, Katagiri K, Tochino Y. Breeding of a non-obese, diabetic strain of mice. *JikkenDobutsu.* **29**(1):1-13 (1980)
59. Mamedov IZ, Britanova OV, Zvyagin IV, Turchaninova MA, Bolotin DA, Putintseva EV, Lebedev YB, Chudakov DM. Preparing unbiased T-cell receptor and antibody cDNA libraries for the deep next generation sequencing profiling. *Front Immunol.* **4**:456 (2013)
60. Medarova Z, Tsai S, Evgenov N, Santamaria P, Moore A. In vivo imaging of a diabetogenic CD81T cell response during type 1 diabetes progression. *Magn. Reson. Med.* **59**, 712–720 (2008)
61. Mehers KL, Gillespie KM. The genetic basis for type 1 diabetes. *Br Med Bull.* **88**(1):115-29 (2008)
62. Mordes JP, Bortell R, Doukas J, Rigby M, Whalen B, Zipris D, Greiner DL, Rossini AA. The BB/Wor rat and the balance hypothesis of autoimmunity. *Diabetes Metab Rev.* **12**(2):103-9 (1996)
63. Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *5*, 621–628 (2008)
64. Motzo C, Contu D, Cordell HJ, Lampis R, Congia M, Marrosu MG, Todd JA, Devoto M, Cucca F. Heterogeneity in the magnitude of the insulin gene effect on HLA risk in type 1 diabetes. *Diabetes.* **53**(12):3286-91 (2004)
65. Nanki T, Kohsaka H, Mizushima N, Ollier WE, Carson DA, Miyasaka N. Genetic control of T cell receptor BJ gene expression in peripheral lymphocytes of normal and rheumatoid arthritis monozygotic twins. *J ClinInvest.* **98**(7):1594-601 (1996)
66. Nerup, J, Platz, P, Andersen, O. O, & Svejgaard, A. (1974). HLA antigens and diabetes mellitus. *Lancet* **2**(7885), 864-866 (1974)
67. Nisticò L, Buzzetti R, Pritchard LE, Van der Auwera B, Giovannini C, Bosi E, Larrad MT, Rios MS, Chow CC, Cockram CS et al., The CTLA-4 gene region of chromosome 2q33 is linked to, and associated with, type 1 diabetes. *BelgianDiabetesRegistry. HumMolGenet.* **5**(7):1075-80 (1996)

Dott.ssa Loredana Marcia. Titolo tesi: Alla ricerca delle basi molecolari del Diabete di Tipo I: sequenziamento di ultima generazione in gemelli monozigoti concordanti e discordanti per la malattia. Tesi di Dottorato in Scienze Biomediche. Indirizzo: Genetica Medica, Malattie Metaboliche e Nutrigenomica. Ciclo XXVIII. Università degli Studi di Sassari.

68. Oda Y, Okada T, Yoshida H, Kaufman RJ, Nagata K, Mori K. Derlin-2 and Derlin-3 are regulated by the mammalian unfolded protein response and are required for ER-associated degradation. *J Cell Biol.* **172(3)**:383-93 (2006)
69. Parekh RB, Dwek RA, Sutton BJ, Fernandes DL, Leung A, Stanworth D, Rademacher TW, Mizuochi T, Taniguchi T, Matsuta K, et al. Association of rheumatoid arthritis and primary osteoarthritis with changes in the glycosylation pattern of total serum IgG. *Nature.* **316(6027)**:452-7 (1985)
70. Parikka V, Nääntö-Salonen K, Saarinen M, Simell T, Ilonen J, Hyöty H, Veijola R, Knip M, Simell O. Early seroconversion and rapidly increasing autoantibody concentrations predict prepubertal manifestation of type 1 diabetes in children at genetic risk. *Diabetologia***55**: 1926–36 (2012)
71. Patsopoulos NA; Bayer Pharma MS Genetics Working Group; Steering Committees of Studies Evaluating IFN β -1b and a CCR1-Antagonist; ANZgene Consortium; GeneMSA; International Multiple Sclerosis Genetics Consortium, Esposito F, Reischl J, et al., Genome-wide meta-analysis identifies novel multiple sclerosis susceptibility loci. *AnnNeurol.* **70(6)**:897-912 (2011)
72. Patterson CC, Dahlquist GG, Gyürüs E, Green A, Soltész G; EURODIAB Study Group. Incidence trends for childhood type 1 diabetes in Europe during 1989-2003 and predicted new cases 2005-20: a multicentre prospective registration study. *Lancet* **13;373(9680)**:2027-33 (2009)
73. Pitzalis M, Zoledziewska M, Cucca F. Il contributo della genetica alla comprensione delle basi eziopatogenetiche del diabete di tipo 1. *Frontiere.* Luglio-Settembre 2013 • Vol. 43 • N. 171 • Pp. 179-185
74. Polychronakos C, Li Q. Understanding type 1 diabetes through genetics: advances and prospects. *Nat Rev Genet.* **18**;12(11):781-92 (2011)
75. Rawat VP, Arseni N, Ahmed F, Mulaw MA, Thoene S, Heilmeyer B, Sadlon T, D'Andrea RJ, Hiddemann W, Bohlander SK, Buske C, Feuring-Buske M. The vent-like homeobox gene VENTX promotes human myeloid differentiation and is highly expressed in acute myeloid leukemia. *ProcNatAcadSci U S A.* **107(39)**:16946-51 (2010)
76. Redondo MJ, Jeffrey J, Fain PR, Eisenbarth GS, Orban T. Concordance for islet autoimmunity among monozygotic twins. *N Engl J Med.* **359(26)**:2849-50 (2008)
77. Redondo MJ, Yu L, Hawa M, Mackenzie T, Pyke DA, Eisenbarth GS, Leslie RD. Heterogeneity of type 1 diabetes: analysis of monozygotic twins in Great Britain and the United States. *Diabetologia.* **44**:354-362 (2001)
78. Ridgway WM, Ito H, Fassò M, et al. Fathman CG. Analysis of the role of variation of major histocompatibility complex class II expression on nonobese diabetic (NOD) peripheral T cell response. *J Exp Med* **188**:2267-7 (1998)
79. Rodríguez-Bayona B, Ruchaud S, Rodríguez C, Linares M, Astola A, Ortiz M, Earnshaw WC, Valdivia MM. Autoantibodies against the chromosomal passenger protein INCENP found in a patient with Graham Little-Piccardi-Lassueur syndrome. *J Autoimmune Dis.* **4**:1 (2007)
80. Schneider A, Rieck M, Sanda S, Pihoker C, Greenbaum C, Buckner JH. The effector T cells of diabetic subjects are resistant to regulation via CD41 FOXP31 regulatory T cells. *J. Immunol.* **181**, 7350–7355 (2008)
81. Seppälä I, Kleber ME, Lyytikäinen LP, Hernesniemi JA, Mäkelä KM, Oksala N, Laaksonen R, Pilz S, Tomaschitz A, Silbernagel G et al., Genome-wide association study on dimethylarginines reveals novel AGXT2 variants associated with heart rate variability but not with overall mortality. *Eur Heart J.* **35(8)**:524-31 (2014)
82. Serban V, Timar R, Dabelea D, Green A, McKinney P, Law G; ONROCAD –Romanian National Organization for the Care of Diabetic Children and Adolescents. The epidemiology of childhood-onset type 1 diabetes mellitus in Romania. ONROCAD Study Group. National Romanian Organisation for the Care of Diabetic Children and Adolescents. *J PediatrEndocrinolMetab.* **14(5)**:535-41 (2001)
83. Shugay M, Britanova OV, Merzlyak EM, Turchaninova MA, Mamedov IZ, Tuganbaev TR, Bolotin DA, Staroverov DB, Putintseva EV, Plevova K et al., Towards error-free profiling of immune repertoires. *Nat Methods.* **11(6)**:653-5 (2014)
84. Simon Anders, W. H. Differential expression analysis for sequence count data. *Genome Biology* **11**, R106 (2010)
85. Sims D, Sudbery I, Illott NE, Heger A, Ponting CP. Sequencing depth and coverage: key considerations in genomic analyses. *Nat Rev Genet.* **15(2)**:121-32 (2014)
86. Singal DP & Blajchman MA. Histocompatibility (HL-A) antigens, lymphocytotoxic antibodies and tissue antibodies in patients with diabetes mellitus. *Diabetes.* **22(6)**, 429-432. (1973)

Dott.ssa Loredana Marcia. Titolo tesi: Alla ricerca delle basi molecolari del Diabete di Tipo I: sequenziamento di ultima generazione in gemelli monozigoti concordanti e discordanti per la malattia. Tesi di Dottorato in Scienze Biomediche. Indirizzo: Genetica Medica, Malattie Metaboliche e Nutrigenomica. Ciclo XXVIII. Università degli Studi di Sassari.

87. Somma P, Ristori G, Battistini L, Cannoni S, Borsellino G, Diamantini A, Salvetti M, Sorrentino R, Fiorillo MT. Characterization of CD8⁺ T cell repertoire in identical twins discordant and concordant for multiple sclerosis. *J Leukoc Biol.* **81(3):**696-710 (2007)
88. Songini M, Lombardo C. The Sardinian way to type 1 diabetes. *J Diabetes Sci Technol.* **4(5):**1248-55 (2010)
89. Steck AK, Johnson K, Barriga KJ, Miao D, Yu L, Hutton JC, Eisenbarth GS, Rewers MJ. Age of islet autoantibody appearance and mean levels of insulin, but not GAD or IA-2 autoantibodies, predict age of diagnosis of type 1 diabetes: diabetes autoimmunity study in the young. *Diabetes Care* **34:** 1397–99 (2011).
90. Steck, A. K. and Rewers, M. J. Epidemiology and risk factors for type 1 diabetes mellitus, in International Textbook of Diabetes Mellitus, Fourth Edition, Fourth Edition (eds R. A. DeFronzo, E. Ferrannini, P. Zimmet and K. G. M. M. Alberti), John Wiley & Sons, Ltd, Chichester, UK (2015)
91. Takaya J, Tanabe Y, Kuroyanagi Y, Kaneko K. Asymmetric dimethylarginine is negatively correlated with hyperglycemia in children. *Endocr.J.* **62 (6):**551-6 (2015)
92. Tang, Q. & Bluestone, J. A. The Foxp31 regulatory T cell: a jack of all trades, master of regulation. *Nature Immunol.* **9,** 339–344 (2008)
93. TEDDY Study Group. The Environmental Determinants of Diabetes in the Young (TEDDY) Study. *Ann N Y Acad Sci.* **1150:**1-13 (2008)
94. Ueda H, Howson JM, Esposito L, Heward J, Snook H, Chamberlain G, Rainbow DB, Hunter KM, Smith AN, Di Genova G et al., Association of the T-cell regulatory gene CTLA4 with susceptibility to autoimmune disease. *Nature* **423 (6939):**506-11 (2003)
95. Vella A, Cooper JD, Lowe CE, Walker N, Nutland S, Widmer B, Jones R, Ring SM, McArdle W, Pembrey ME, Strachan DP, Dunger DB, Twells RC, Clayton DG, Todd JA. Localization of a type 1 diabetes locus in the IL2RA/CD25 region by use of tag single-nucleotide polymorphisms. *Am J Hum Genet.* **76(5):**773-9 (2005)
96. Wang Z, Zhang J, Ye M, Zhu M, Zhang B, Roy M, Liu J, An X. Tumor suppressor role of protein 4.1B/DAL-1. *Cell Mol Life Sci.* **71(24):**4815-30 (2014)
97. Wang, L., Wang, S. & Li, W. RSeQC: quality control of RNA-seq experiments. *Bioinformatics* **28,** 2184–2185 (2012)
98. Ward, J. H., Jr. "Hierarchical Grouping to Optimize an Objective Function", *Journal of the American Statistical Association*, **58,** 236–244 (1963)
99. Welter D, MacArthur J, Morales J, Burdett T, Hall P, Junkins H, Klemm A, Flicek P, Manolio T, Hindorff L, Parkinson H. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res.* **42** (Database issue):D1001-6 (2014)
100. Willcox A, Richardson SJ, Bone AJ, Foulis AK, Morgan NG. Analysis of islet inflammation in human type 1 diabetes. *Clin. Exp. Immunol.* **155,** 173–181 (2009)
101. Wu X, Gao H, Bleday R, Zhu Z. Homeobox transcription factor VentX regulates differentiation and maturation of human dendritic cells. *J Biol Chem.* **289(21):**14633-43 (2014)
102. Yanagimachi M, Miyamae T, Naruto T, Hara T, Kikuchi M, Hara R, Imagawa T, Mori M, Kaneko T, Goto H, Morita S, Mizuki N, Kimura A, Yokota S. Association of HLA-A*02:06 and HLA-DRB1*04:05 with clinical subtypes of juvenile idiopathic arthritis. *J Hum Genet.* **56(3):**196-9 (2011)