

UNIVERSITÀ DEGLI STUDI DI NAPOLI FEDERICO II



DOTTORATO DI RICERCA IN
MEDICINA CLINICA E SPERIMENTALE
CURRICULUM IN SCIENZE MEDICHE TRASLAZIONALI

XXX Ciclo

(Anni 2014-2017)

Coordinatore: Prof. Gianni Marone

TESI DI DOTTORATO

**GENETICA ED EPIGENETICA DEI GENI CANDIDATI NELLA
MALATTIA CELIACA: ANALISI MOLECOLARE DEI
MECCANISMI DI REGOLAZIONE**

TUTOR

Chiar.mo Luigi Greco

Handwritten signature of Chiar.mo Luigi Greco.

CANDIDATA

Dott. Donatella Cielo

INDICE

1. INTRODUZIONE	3
1.1 La Malattia Celiaca: Definizione ed epidemiologia	3
1.2 Eziopatogenesi	3
1.3 Quadro clinico	6
1.4 La genetica: geni HLA e geni non HLA	7
1.5 Epigenetica	15
1.6 Differenziamento delle cellule intestinali	16
1.6.1 Epitelio	17
1.6.2 Lamina propria	20
2. SCOPO DELLA TESI	23
3. MATERIALI E METODI	25
3.1 Pazienti	25
3.1.1 Pazienti arruolati nello studio longitudinale su cellule di sangue periferico	25
3.1.2 Pazienti arruolati nello studio su cellule intestinali	26
3.2 Isolamento delle cellule mononucleari del sangue periferico	28
3.3 L'isolamento delle cellule intestinali CD326+ da quelle CD326-	28
3.4 Real-Time PCR	29
3.5 Genotipizzazione	31
3.6 Analisi statistica	31
3.7 Studi di metilazione	32
3.7.1 Conversione con bisolfito	32
3.7.2 Methylation Specific PCR	32
3.7.3 Dna Clean Up	34
3.7.4 Sequenziamento	34
4. RISULTATI	35
4.1 Analisi molecolare dei geni candidati in cellule di sangue periferico	35
4.1.1 Genotipizzazione dei geni candidati	35
4.1.2 Analisi di espressione genica dei geni candidati	36
4.1.3 Analisi Discriminante Multivariata	37
4.2 Analisi molecolare dei geni candidati in cellule intestinali	42
4.2.1 Valutazione della purezza del campione	42
4.2.2 Espressione genica in Epitelio e Lamina Propria	43

<i>4.2.3 Analisi di correlazione tra espressione genica dei geni candidati</i>	46
<i>4.2.4 Analisi discriminate multivariata</i>	45
<i>4.2.5 Analisi di Metilazione</i>	46
5. DISCUSSIONE E CONCLUSIONI	49
6. BIBLIOGRAFIA	62

1.INTRODUZIONE

1.1 la malattia celiaca: definizione ed epidemiologia

La malattia celiaca (MC) è una patologia sistemica immunomediata che colpisce prevalentemente il piccolo intestino (*Ludvigsson J, 2013*). È causata dall'ingestione di glutine, una prolamina presente in frumento, orzo, segale e avena e in soggetti geneticamente predisposti, si manifesta con un'ampia varietà di segni e sintomi sia gastrointestinali che extraintestinali.

Studi epidemiologici condotti in regioni geografiche che presentano un'apparente bassa prevalenza di MC (Asia, America meridionale, Africa) hanno dimostrato che la malattia è frequentemente sotto-diagnosticata, fornendo dimostrazione che la celiachia è invece una delle più comuni patologie risultanti dalla combinazione di fattori genetici e ambientali. La prevalenza Europa e negli USA è di 1:130-1:300 (*Mustalahti K, 2010; Rubio-Tapia, 2012; Tucci F, 2014*), i dati più recenti relativi al numero di celiaci diagnosticati in Italia sono quelli forniti dalla Relazione al Parlamento del Ministero della Salute edizione 2013 (*Ministero della Salute 2013*). Il numero dei pazienti effettivamente diagnosticati al 31/12/2013 è 164.500 ma la prevalenza stimata di questa patologia è appunto l'1% della popolazione italiana, e dunque il numero teorico complessivo dei celiaci sarebbe quindi pari a 600.000 persone circa, dei quali 435.500 (73% circa) ad oggi non è ancora consapevole di essere celiaco.

Il rapporto stimato medio di pazienti celiaci MASCHI:FEMMINE è pari a circa 1:2. Il numero atteso di donne celiache è quindi pari a circa 400.000 mentre le donne ad oggi effettivamente diagnosticate sono solo 116.000 e, di conseguenza, oltre 284.000 non hanno ancora ricevuto la diagnosi. Dal 2012 al 2013 (*Ministero della Salute, 2013*) il numero di diagnosticati è cresciuto del 10,6%, confermando un trend in ascesa di diagnosi di MC già verificatosi nel biennio 2011-2012.

1.2 Eziopatogenesi

La celiachia è una patologia multifattoriale e come tale nel determinare la sua insorgenza concorrono diverse componenti:

- *Componente ambientale*: è rappresentata dall'assunzione di *gliadina*, principale frazione del glutine. Oltre all'introduzione della gliadina con la dieta, altri fattori ambientali sono stati descritti come associati all'insorgenza della MC. Tra essi ci sono l'allattamento al seno, che sembrerebbe ritardare l'insorgenza della malattia così come

l'epoca di introduzione del glutine (*Capriati T, 2015*). Anche le infezioni intestinali, in particolare quelle da Rotavirus, secondo alcuni studi potrebbero esercitare un ruolo nella patogenesi della malattia (*Dolcino M, 2013*). Infine, recenti ricerche hanno indicato che la composizione del microbiota intestinale possa avere un ruolo nell'insorgenza della celiachia, e a seconda della sua composizione, favorire o proteggere dalla comparsa della malattia (*Verde EF, 2015*). Tuttavia il reale peso di questi fattori ambientali aggiuntivi resta oggi ancora controverso ed è oggetto di opinioni talvolta contrastanti tra i ricercatori. (*Lionetti E, 2014; Szajewska H, 2016*)

- *Componente genetica*: è dimostrata dalla presenza di una frequenza di circa 20 volte maggiore della malattia nei familiari di primo grado dei pazienti celiaci, rispetto alla popolazione generale (*Singh P, 2015*) e dalla maggiore concordanza tra gemelli omozigoti (>80%) rispetto ai gemelli eterozigoti (10%) (*Greco L, 2002; Nisticò L, 2006; Kuja-Halkola R, 2016*). La componente genetica è costituita prevalentemente dai geni del complesso maggiore di istocompatibilità umano (HLA) di classe II, presenti sul braccio corto del cromosoma 6. Approssimativamente il 90-95% dei pazienti celiaci esprimono l'eterodimero DQ2 mentre il 5-10% l'eterodimero DQ8. Tuttavia i soli geni HLA non sono sufficienti per lo sviluppo della malattia, come dimostrato dal fatto che solo una minima parte delle persone che esprimono il DQ2 e il DQ8 sviluppano la malattia.

I meccanismi patogenetici che sono alla base del danno mucosale sono innescati da una reazione autoimmunitaria diretta contro le prolamine (gliadina, ordeina, secalina), proteine ricche di residui glutamminici e prolinici. Il gran numero di residui prolinici presenti in questi peptidi li rende estremamente resistenti all'azione proteolitica degli enzimi digestivi, pertanto essi arrivano intatti nel lume intestinale. In condizioni fisiologiche gli stessi peptidi, attraversano la mucosa per essere digeriti all'interno degli enterociti dagli enzimi lisosomiali. Si ipotizza che nel soggetto celiaco tale meccanismo di transitosi risulti difettoso pertanto i peptidi raggiungerebbero intatti la lamina propria. Altri meccanismi ipotizzati di passaggio attraverso la mucosa dei peptidi gliadinici sono: la via paracellulare come conseguenza di una aumentata espressione di zonulina, una proteina che regola l'apertura delle tight junction, o per retro-transitosi di immunocomplessi formati tra i peptidi gliadinici e IgA secretorie (sIgA) presenti nel lume intestinale, mediata dal legame al recettore della transferrina CD71. Attraversata la barriera intestinale, i peptidi gliadinici sono in grado di innescare una risposta immune di tipo adattativo. In particolare, i peptidi, principalmente il 33-mer, legandosi alle cellule presentanti l'antigene (APC) determinano l'attivazione delle cellule linfocitarie T

CD4+ della lamina propria. Essi però vengono più facilmente riconosciuti dai linfociti T gliadina-specifici DQ2/DQ8 dopo essere stati deamidati ad opera delle transglutaminasi. I linfociti T, una volta attivati, producono un pattern di citochine di tipo Th1 quali: $IFN\gamma$, IL2, IL6, $TNF\alpha$ indirettamente responsabili delle lesioni mucosali, attraverso la produzione di ossido nitrico ed attraverso un'umentata apoptosi enterocitaria. Le stesse cellule T inducono l'attivazione e l'espansione clonale dei linfociti B con conseguente produzione di specifici anticorpi anti-gliadina, anti-transglutaminasi e anti-endomisio.

Tuttavia sembra che anche la risposta immunitaria di tipo *innato* (innescata da un peptide differente, il 31-43) giochi un ruolo fondamentale nella patogenesi della MC, in particolar modo, un'umentata espressione dell'IL15 da parte degli enterociti, conduce all'attivazione dei linfociti intraepiteliali che esprimono il recettore NKG2D e CDI94 (marker delle cellule natural-killer) e del pathway perforina-granzima con conseguente distruzione degli enterociti (Fig. 1) (Sollid LM & Jabri B, 2013).

L'effetto finale di entrambi i tipi di risposta immune è sempre un danno mucosale con atrofia dei villi intestinali e riduzione della superficie utile all'assorbimento dei nutrienti, dando origine ad un malassorbimento.

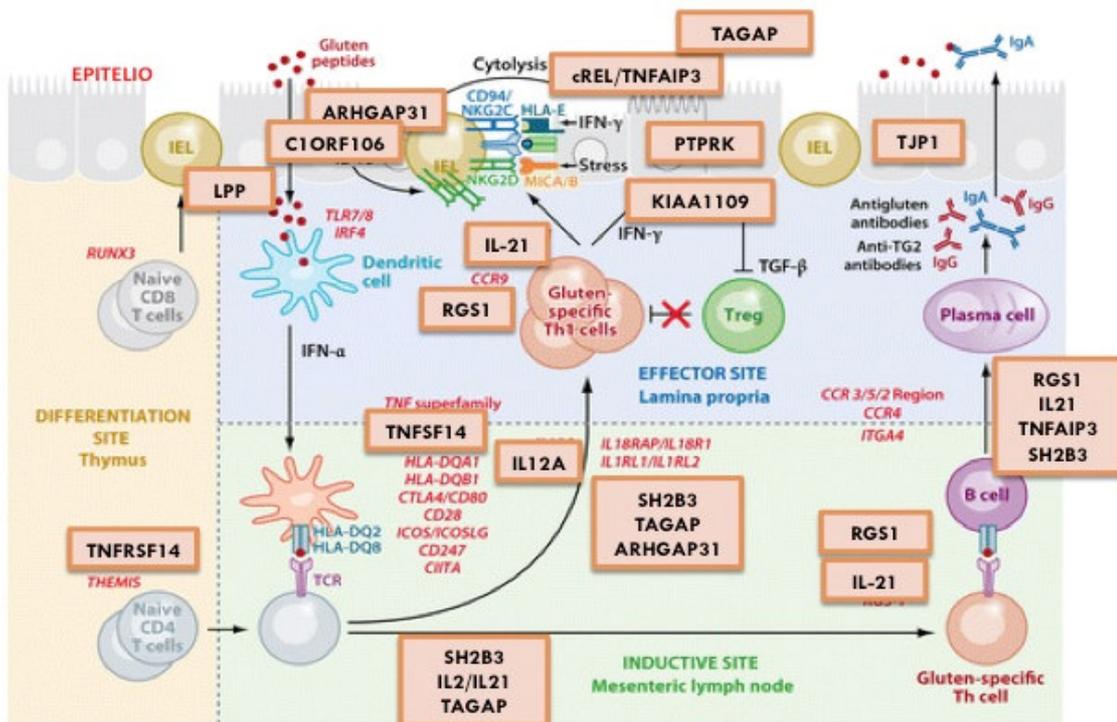


Figura 1: Fasi della patogenesi della celiachia. (Mod. da Sollid LM & Jabri B, 2013)

1.3 Quadro clinico

La Malattia Celiaca interessa prevalentemente la mucosa del piccolo intestino. Prelievi biotipici di mucosa duodenale (Fig.2), eseguiti mediante esofagogastroduodenoscopia (EGDS), mostrano alterazioni caratteristiche quali ipertrofia delle cripte, atrofia dei villi, incremento del numero dei linfociti intraepiteliali (IEL) (Husby S, 2012).

La diagnosi di MC viene eseguita valutando il grado di atrofia dei villi secondo la classificazione di Marsh (Marsh MN, 1990):

- Tipo 1 o infiltrativa, villi normali ma incremento patologico del numero dei linfociti intraepiteliali (> 25/100 cellule epiteliali);
- Tipo 2 o iperplastica, in cui all'incremento del numero dei linfociti intraepiteliali si associa anche una iperplasia delle cripte ghiandolari;
- Tipo 3 o atrofica, in cui all'incremento del numero di linfociti intraepiteliali ed alla iperplasia delle cripte, si associa un'atrofia dei villi di grado variabile (IIIA parziale, IIIB subtotale, IIIC totale).

Il 50-60% dei pazienti celiaci cadono nella categoria Marsh III.

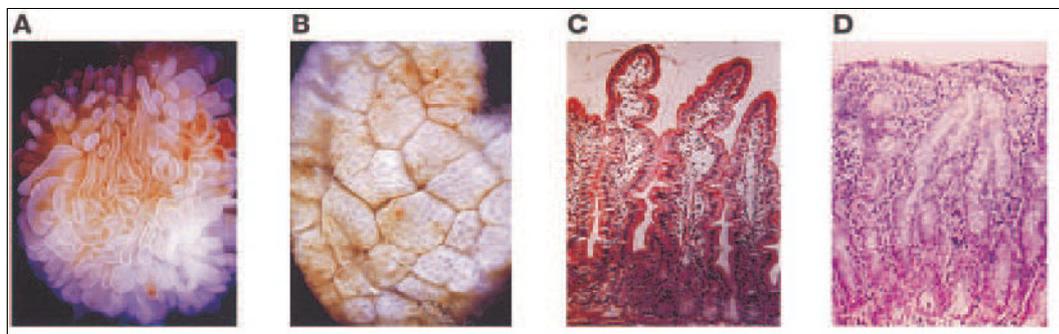


Figura2: Biopsia della mucosa dell'intestino tenue. (A e B) Biopsia di mucosa di intestino tenue vista attraverso un microscopio stereoscopico. L'aspetto normale (A) mostra numerosi villi in superficie mentre la biopsia di un soggetto celiaco si caratterizza per una totale atrofia dei villi (B). Sezione colorata con ematossilina-eosina di mucosa intestinale di un soggetto normale (C) e di un soggetto affetto da celiachia (D), ingrandimento 400X.

Il rilievo di tali alterazioni della mucosa duodenale, accompagnato dalla positività alle indagini sierologiche per la ricerca degli anticorpi anti-transglutaminasi (tTG) ed anti-endomisio (EMA), altamente specifici per la MC, rappresenta il gold-standard diagnostico.

Solo una recente revisione dei criteri diagnostici ESPGHAN sottolinea la possibilità di predire la presenza di un'atrofia intestinale glutine-dipendente in presenza di un elevato titolo sierico di anticorpi anti-transglutaminasi (TG2). La stessa revisione suggerisce che in presenza di

sintomi suggestivi, anticorpi tTG ≥ 10 volte il limite superiore del valore normale (ULN) e un HLA compatibile, per i solo pazienti pediatrici la biopsia intestinale non è necessaria ai fini diagnostici (*Husby S, 2012*).

Dal punto di vista clinico la MC è caratterizzata da una gamma molto ampia di sintomi. Tra i differenti spettri di presentazione della malattia è possibile riconoscerne quattro principali forme (*Greetje J, 2013*):

- **Forma classica:** è la più frequente in età pediatrica, predominano sintomi tipici del malassorbimento quali diarrea, steatorrea, calo ponderale e/o scarsa crescita accompagnati frequentemente da distensione e dolore addominale, anoressia e vomito. Presenta, ovviamente, positività per gli anticorpi specifici e per l'esame biotico.
- **Forma atipica:** nella quale vi è una netta prevalenza di sintomi extraintestinali: anemia sideropenica refrattaria, bassa statura, ritardo puberale, ipoplasia dello smalto dentario, astenia e crampi muscolari, epiessia con calcificazioni bilaterali (*Gobbi G, 1992*).
- **Forma silente:** caratterizzata dalla presenza di positività per gli anticorpi specifici, per HLA DQ2 e/o DQ8 all'esame biotico, ma in assenza di sintomi o segni clinici.
- **Forma potenziale:** i celiaci potenziali sono pazienti, positivi per HLA DQ2 e/o DQ8 che, indipendentemente dalla presenza o meno di sintomi, risultano negativi all'indagine biotica pur in presenza di una sierologia positiva per le tTG (*Kurppa K, 2010; Sperandeo MP, 2011*).

L'unico trattamento di provata efficacia per la MC è la stretta osservanza della dieta senza glutine (GFD). Tuttavia le contaminazioni da tracce di glutine presenti nei prodotti gluten-free non possono essere completamente evitate; i risultati di uno studio placebo-controllo in doppio cieco hanno mostrato che dosi giornaliere fino a 10 mg di glutine sono ancora tollerate, ma la presenza di una forte variabilità individuale non permette di stabilire una soglia universale di sicurezza (*Di Sabatino A, 2009*).

1.4 La genetica: i geni HLA e i geni non HLA

I geni HLA

I principali fattori genetici predisponenti la MC sono i geni del complesso maggiore di istocompatibilità (MHC) di classe II. Tale complesso è localizzato sul cromosoma 6 (regione 6p21), contiene centinaia di geni con funzioni immunologiche ed è responsabile della più forte associazione osservata alle malattie immuno-mediate.

L'associazione tra MHC e sviluppo di MC è stata descritta per la prima volta nel 1970 (Falchuk Z.M, 1972; Stokes PL, 1972). Un decennio più tardi, gli alleli HLA di classe II codificanti per l'antigene DQ2 sono stati identificati come i principali responsabili del rischio genetico conferito dalla regione del MHC (Tosi R, 1983; Sollid LM, 1989) attribuito precedentemente agli alleli di classe I HLA-B8 e HLA-DR3 (Falchuk Z.M, 1972).

Le molecole codificate dai geni di classe II sono degli eterodimeri caratterizzati da due catene transmembrana (α e β) codificate rispettivamente dai geni DQA e DQB, tali geni sono altamente polimorfici; per le molecole DR il polimorfismo è limitato invece alle catene β (codificate dai geni DRB).

Circa il 90% dei pazienti celiaci sono portatori degli alleli che codificano per l'antigene DQ2, in particolare, per l'isoforma DQ2.5: DQA1*05 e DQB1*02, che possono essere localizzati in cis (in associazione con DRB1*03) o in trans (in associazione con DRB1*07 e DRB1*05) e per l'isoforma DQ2.2: DQA1*02 e DQB1*02 (in associazione con DRB1*07). Le molecole HLA codificanti il DQ2.5 in cis e in trans conferiscono un rischio simile e molto elevato (21%) di sviluppare la MC (Bourgey M, 2007).

A livello funzionale è stato dimostrato in vitro il legame preferenziale dei peptidi derivanti dalla digestione della gliadina all'eterodimero DQ2 (Fig.3). Il legame e la presentazione alle cellule T-CD4+ nella mucosa intestinale di peptidi derivati dal glutine da parte delle molecole HLA-DQ sarebbe dunque il meccanismo alla base della suscettibilità.

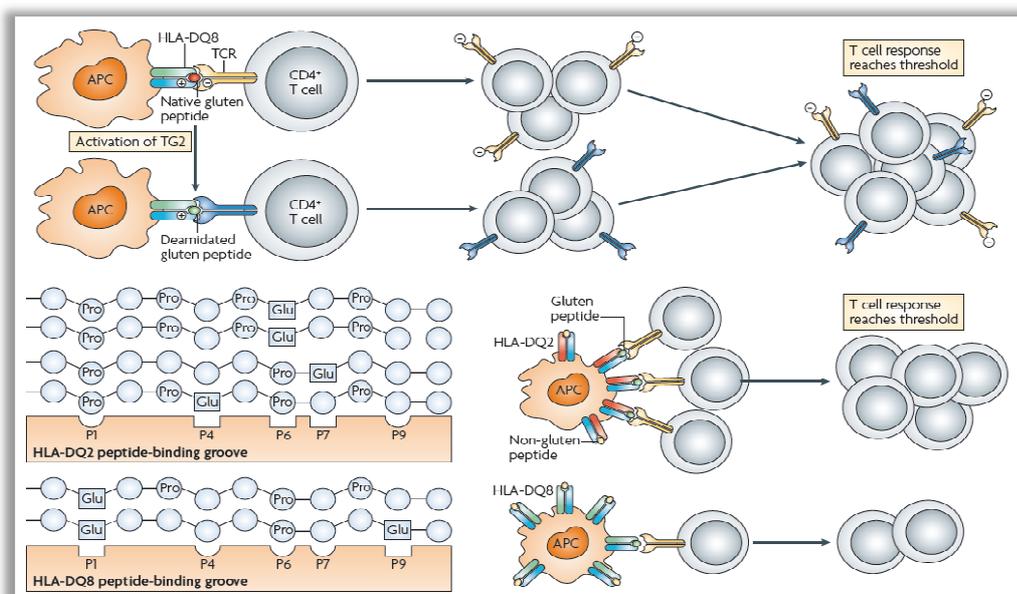


Figura 3. Molecole HLA e risposta glutine specifica delle cellule T CELIACI4+

Inoltre, la transglutaminasi tissutale, presente a livello della membrana basale delle cellule epiteliali intestinali, è in grado di deamidare la gliadina, creando un peptide che si lega alla molecola DQ più efficientemente rispetto al peptide nativo. La deamidazione produce Acido Glutammico, carico negativamente, al posto dell'originale Glutamina con carica neutra. Poiché la molecola DQ2 accomoda preferenzialmente nel suo sito di legame peptidi che recano residui negativi nella loro struttura, i peptidi della gliadina deamidati possono produrre risposte più forti legandosi con maggiore affinità alla molecola DQ2.

Esiste inoltre, un effetto dose del DQB1*02, in quanto la presenza in omozigosi di tale allele determina un rischio di sviluppare la MC superiore rispetto a quando esso è presente in singola copia (*Ploski R, 1983*). La maggior parte dei pazienti senza DQ2.5 sono portatori degli alleli DQA1*03 e DQB1*0302, che codificano la molecola DQ8 (*Spurkland, A 1992*). Anche per il DQ8 è stato proposto un effetto dose. Nei pazienti celiaci che non sono portatori né del DQ2.5 né del DQ8, sono presenti gli alleli DQA1*02 e DQB1*02 codificanti la molecola DQ2.2. Il rischio genetico attribuibile a quest'ultimo aplotipo è inferiore a quello del DQ2.5, ma non può essere tuttavia trascurato in quanto circa il 27% dei celiaci esprime tale etero dimero (*Karell K, 2003*). Infine, recentemente, in uno studio effettuato dal nostro gruppo di ricerca, è stata evidenziata la presenza della molecola DQ7 nel 24% dei soggetti celiaci (*Tinto N, 2015*).

I geni di rischio HLA rendono conto per il 40% dell'ereditabilità della celiachia. La complessità della regione MHC, caratterizzata dall'ospitare un grandissimo numero di geni, dall'alto grado di polimorfismo e dal *linkage disequilibrium*, ha reso molto difficile l'identificazione di varianti di rischio aggiuntive in questa regione. Tuttavia, di recente, mediante una fine mappatura della regione MHC, è stato possibile individuare cinque nuove varianti indipendenti in questa regione. In questo recente studio sono state individuate come nuove varianti di rischio L'HLA-DPβ1 (posizione 9), l'HLA-B (alleli HLA-B*08 e HLA-B*39:06) e due SNP, l'rs1611710, che mostra un effetto diretto sull'espressione del HLA-F, e l'rs2301226, che mostra un effetto anch'esso un effetto diretto sull'espressione di B3GALT4 e del HLA-DPB1 (*Gutierrez-Achury, 2015*). Questi cinque fattori spiegano un ulteriore 2,5-3% della ereditabilità malattia, lasciando una frazione considerevole ancora da spiegare.

I geni non-HLA

Per lungo tempo, le varianti genetiche situate all'interno della regione MHC sono state ampiamente riconosciute come gli unici fattori genetici che contribuiscono al rischio di sviluppare la MC. Tuttavia, è sempre stato evidente che le molecole DQ2 e DQ8 non erano

sufficienti a causare lo sviluppo di MC. Nei paesi occidentali, circa il 30-40% della popolazione generale possiede una o entrambe queste molecole, ma solo l'1% degli individui presenta malattia, rendendo chiaro che anche altri fattori di suscettibilità (ambientali e genetici) devono essere coinvolti nella insorgenza della MC.

Durante l'ultimo decennio, con l'avvento degli studi di associazione genome-wide (GWAS) e delle piattaforme di Immunochip e di NGS, il panorama della genetica della MC è cambiato drasticamente ed è ormai ben noto che sia geni HLA che non-HLA influenzano lo sviluppo della celiachia.

Numerosi sono gli studi effettuati in tal senso, attraverso cui si è giunti all'identificazione di 57 geni non-HLA che ad oggi hanno contribuito a spiegare un ulteriore 6,5% dell'ereditabilità della MC (*van Heel DA, 2007; Hunt KA, 2008; Castellanos-Rubio A, 2008; Romanos J 2009; Trynka G, 2009; Dubois PC, 2010, Trynka G, 2011, Plaza-Izurietta 2011; Kumar V 2015*) (Tab.1).

I “*Geni Candidati*”, sono geni che si presume siano responsabili della ereditarietà mancante della MC. Si tratta di geni di cui è noto in letteratura un coinvolgimento nei meccanismi legati alla malattia o che codificano per proteine coinvolte nel meccanismo patogenetico.

In relazione alla loro funzione i geni precedentemente menzionati possono essere suddivisi in 4 gruppi differenti:

1. Geni direttamente coinvolti nell'infiammazione/danno cellulare
2. Geni candidati “classici” in quanto risultati fortemente associati alla MC da un punto di vista genetico
3. Geni coinvolti nel homing di linfociti e nella regolazione dei processi infiammatori e del danno cellulare
4. Geni coinvolti nell'adesione cellulare ed integrità della barriera intestinale

Parte di questi geni sono coinvolti in altre patologie autoimmuni: IL2, IL21 e SH2B3 sono comuni a celiachia e diabete; IL18RAP a celiachia e malattia Crohn, confermando l'ipotesi del coinvolgimento di meccanismi comuni nell'insorgenza di tali patologie. Alcuni studi hanno dimostrato una differente espressione di questi geni nella malattia celiaca rispetto alla popolazione sana, sia a livello della mucosa duodenale che delle cellule del sangue periferico (*Galatola M, 2013*).

Dei marcatori identificati, pochi sono localizzati in regioni esoniche, infatti la maggior parte di essi si trova in regioni non codificanti suggerendo un ruolo di questi nei meccanismi di regolazione genica. Questo è stato ampiamente confermato dagli studi di espressione genica, che come è noto è tessuto specifica, e dunque, è importante studiare tessuti diversi per

ottenere un quadro più reale dell'influenza delle variazioni di espressione di questi geni. In uno studio preliminare, Plaza-Izurietae colleghi hanno eseguito una analisi cis e-QTL (expression quantitative trait loci) nel piccolo intestino, analizzando 8 regioni GWAS-associate con risultati significativi per i geni IL12A, LPP, SCHIP1 e SH2B3. In un lavoro successivo, gli stessi autori hanno esteso le analisi ad altri 45 geni, 15 dei quali sono risultati differenzialmente espressi tra i pazienti celiaci e controlli, e 11 tra pazienti celiaci e pazienti in remissione. Gli autori hanno osservato una stretta correlazione nell'espressione di 18 geni in pazienti celiaci (attivi e remissioni), ma non nei controlli, tali geni sembravano essere sotto il controllo di tre SNP, uno dei quali localizzato su un lincRNA (long intergenic non-coding RNAs) (Plaza-Izurietae, 2015). Anche in un altro studio, eseguito da Dubois e colleghi, è stato evidenziato come, uno squilibrio nella regolazione della risposta immunitaria possa essere associato allo sviluppo della MC, in quanto la maggior parte dei loci associati al rischio ha funzioni relative al sistema immunitario. Questo lavoro ha evidenziato geni di suscettibilità coinvolti nello sviluppo delle cellule T nel timo, come THEMIS, TNFRSF14.

Il lavoro di Dubois inoltre suggerisce il coinvolgimento anche della risposta immunitaria innata innescata da RNA virale, dalla co-stimolazione delle cellule T-B (o co-inibizione) e di un terzo pathway che coinvolge citochine, chemochine e loro recettori nella patogenesi della celiachia.

Studi di espressione genica hanno fornito ulteriori prove dell'implicazione dei diversi pathways nella malattia celiaca. Fernandez- Jimenez e colleghi hanno studiato l'espressione in biopsie intestinali di 93 geni del pathway di NFkB e confermato l'attivazione di una considerevole parte di questi geni nella celiachia. Gli autori hanno trovato una stretta correlazione nell'espressione dei geni studiati nel gruppo di controllo rispetto ai pazienti, il che suggerisce una stretta regolazione di questo pathway che viene deregolato nella malattia celiaca. Il pathway di NFkB, considerato uno dei principali regolatori sia della risposta immunitaria innata sia adattativa, rappresenta quindi un esempio di come la genetica, l'epigenetica e la patogenesi convergono. Infatti le varianti dei geni REL e TNFAIP3 sono coinvolte nella regolazione del pathway NFkB che è considerato un mediatore dell'aumento di IL-15 e della permeabilità intestinale indotti dal glutine, eventi legati alla patogenesi della CD (Fernandez- Jimenez, 2014).

In uno studio del 2013 hanno stabilito che diversi elementi funzionali quali: la funzionalità delle cellule epiteliali intestinali, la crescita e l'omeostasi energetica e il sistema immunitario innato e adattativo, sono coinvolti nella celiachia (Östensson M, 2013). Sulla base di questo, gli autori propongono un modello patogenetico in cui la combinazione di questi elementi con

l'ingestione di glutine, porterebbe ad uno squilibrio metabolico che, in assenza di un agente patogeno, potrebbe innescare una risposta immunitaria contro gli "auto-antigeni" (TG2) e la distruzione dei tessuti.

Per comprendere al meglio le implicazioni funzionali dei dati genetici, un passo in avanti è stato eseguito da Kumar e colleghi. Seguendo un approccio integrativo (eQTL e dati di espressione tessuto-specifica, le informazioni di elementi regolatori previsti dal progetto ENCODE e i risultati delle analisi di pathways e network), questi autori hanno ottenuto risultati molto interessanti. In primo luogo, hanno rafforzato l'idea che la maggior parte degli SNP funzionali agisce attraverso meccanismi che regolano l'espressione genica. Gli autori hanno ottenuto un elenco di 49 potenziali geni di suscettibilità nella MC situati in 39 loci, suddivisi in diversi sottogruppi funzionali, evidenziando che alla base della malattia celiaca può esserci uno squilibrio in diverse linee cellulari. Alcuni dei geni studiati sono preferenzialmente espressi nelle cellule T, nelle cellule B e cellule NK, dando un ulteriore supporto all'implicazione delle cellule B nella patogenesi della celiachia, oltre a confermare il ruolo centrale già descritto delle cellule T. Infine, gli autori hanno identificato quattro nuovi geni candidati (C1orf106, ARHGAP31, LPP e PTPRK) con ruolo nel mantenimento dell'integrità della barriera intestinale. Kumar e colleghi hanno formulato una interessante ipotesi: le varianti di rischio genetico potrebbero causare piccole variazioni nell'espressione dei geni di suscettibilità, ed il loro effetto potrebbe aumentare se questi vanno ad alterare l'espressione di ulteriori geni coinvolti nello stesso pathway, dando quindi luogo ad un effetto cumulativo.

Tutti questi studi genetici si adattano abbastanza bene con i principali modelli immunologici della patogenesi della malattia celiaca:

1. Alterazione dell'epitelio intestinale. Un effetto tossico, dovuto all'ingestione di glutine, attivato dal sistema immunitario innato potrebbe causare la rottura dei complessi delle giunzioni strette e aumentare così la capacità di permeabilità intestinale, che faciliterebbe il passaggio della gliadina attraverso l'epitelio.
2. Deamidazione di peptidi del glutine nella lamina propria da parte dell'enzima TG2 e la successiva presentazione da parte di cellule presentanti l'antigene alle cellule T CD4+.
3. Risposta immunitaria cellulo-mediata, con produzione di anticorpi e rilascio di citochine. Queste citochine possono perpetuare la risposta in corso e aumentare la permeabilità intestinale.

4. Infiltrazione dei linfociti intraepiteliali (IELS) nell' epitelio come conseguenza dell'ambiente infiammatorio creato dalle cellule T CD4+ nella *lamina propria*. IELS sono i responsabili della distruzione degli enterociti e la conseguente atrofia dei villi.

Questa è una panoramica semplificata dei diversi geni che codificano per molecole con varie funzioni e a più livelli, poiché la natura pleiotropica delle citochine implica diversi e molteplici ruoli relativi alla risposta immunitaria. Gli studi di epigenetica potrebbero contribuire a chiarire questo complesso meccanismo patogenetico che è alla base della malattia.

Studio	Locus	Geni candidati	Funzione	
GWAS 1	4q27	KIAA1109, ADAD1, IL2, IL21	<i>IL2, IL21</i> : B- and T-cell activation and proliferation	
Follow-up del GWAS 1	1q31	RGSI ³	B-cell activation and proliferation. Cell signaling and migration	
	2q12	IL1RL1 ¹ , IL18R1, IL18RAP ^{1,3} , SLC9A4	<i>IL1RL1, IL18R1, IL18RAP</i> : pro-inflammatory role; <i>SLC9A4</i> : maintenance of cellular pH homeostasis	
	3p21	CCR1 ³ , CCR2 ¹ , CCR	Cell differentiation, recruitment and signaling	
	3q25	IL12A ^{2,3} , SCHIP1 ^{2,3}	<i>IL12A</i> : Regulation of Th1 differentiation; <i>SCHIP1</i> : tumor suppression	
	3q28	LPP ^{2,3}	Structural role at sites of cell adhesion in maintaining cell shape and motility	
	6q25	TAGAP ¹	T cell activation and modulation of cytoskeletal changes	
	12q24	SH2B3 ^{1,2,3} , ATXN2	<i>SH2B3</i> : innate immune response and T cell signaling; <i>ATXN2</i> : stabilization of mRNAs	
	2q31	ITGA4, UBE2E3 ¹	<i>ITGA4</i> : cell adhesion, migration, and activation of immune cells; <i>UBE2E3</i> : ubiquitination	
	2p16	REL, PUS10	<i>REL</i> : NF-κB signaling pathway; <i>PUS10</i> : posttranscriptional nucleotide modification	
	6q23	OLIG3, TNFAIP3	<i>OLIG3</i> : neuronal development; <i>TNFAIP3</i> : NF-κB signaling pathway and TNF-mediated apoptosis	
	2q33	CTLA4, ICOS ³ , CD28 ¹	<i>CD28, ICOS</i> : T and B cell activation and proliferation stimulation; <i>CTLA4</i> : T and B cell activation and proliferation stimulationinhibition	
	18p11	PTPN2	T-cell differentiation and activation	
	GWAS 2	1p36	TNFRSF14 ¹ , MMEL1 ¹	<i>TNFRSF14</i> : signal transduction pathways influencing immune response; <i>MMEL1</i> : phosphate metabolism and homeostasis
		1p36	RUNX3	T cell differentiation in the thymus
	1q32	C1orf106	Unknown function	
	2p13	PLEK ^{1,3} , FBXO48	<i>PLEK</i> : cytoskeletal reorganization	
	3p21	CCR4, GLB1	<i>CCR4</i> : cell differentiation, recruitment and signaling; <i>GLB1</i> : hydrolysis of glycoconjugates	
	3q13	ARHGAP31 ³	Protein trafficking and cell growth	
	6q15	BACH2	Regulation of humoral and cellular immunity	
	6q22	PTPRK, THEMIS	<i>PTPRK</i> : barrier function, CD4+ T cell development; <i>THEMIS</i> : thymic differentiation of CD4 ⁺ T cells	
	8q24	PVT1	Participation in recurrent translocations	
	10q22	ZMIZ1 ¹	Regulation of transcription factors activation	
	11q23	ETS1	Differentiation, survival and proliferation of lymphoid cells	
	16p13	CIITA	<i>CIITA</i> : regulation of HLA class II genes transcription	
	16p13	SOCS1 ¹ , PRM1, PRM2	<i>SOCS1</i> : negative regulation of cytokine signaling; <i>PRM1</i> and <i>PRM2</i> : germ cell development	
	21q22	ICOSLG	T and B cell activation and proliferation	
ImmunoChip	1q23	FASLG, TNFSF18	<i>FASLG</i> : survival, activation and co-stimulation of T cells	
	2q32	STAT4	T cell differentiation	
	6p25	IRF4	Lymphocyte specific regulator of TLR signaling. Involved in innate and adaptive immunity	
	7p14	ELMO1 ^{1,2,3}	Involved in phagocytosis and cell migration	
	10p15	PFKFB3, PRKCQ	<i>PFKFB3</i> : regulation of oxidative stress; <i>PRKCQ</i> : T cell activation and NF-κB signaling pathway	
	11q23	POU2AF1, C11orf93(COLCA2)	<i>POU2AF1</i> : B-cell development and function	
	11q23	TREH, DDX6	<i>TREH</i> : hydrolysis of trehalose, stress-response protein; <i>DDX6</i> : mRNA degradation and suppression of protein translation	
	14q22	ZFP36L1	Regulation of the response to growth factors, wound healing	
	15q24	CLK3, CSK ¹ , ULK3 ¹ and other genes	<i>CLK3</i> : regulation of alternative splicing; <i>CSK</i> : regulation of cell growth, differentiation, migration and immune response; <i>ULK3</i> : autophagy and regulation of the Hedgehog signaling pathway	
	21q22	UBASH3A ¹	T-cell signaling and T-cell apoptosis	
	2q11	UBE2L3 ^{1,3} , YDJC	<i>UBE2L3</i> : NF-κB signaling pathway and ubiquitination	
	Xq28	HCFC1, TMEM187 ³ , IRAK1	<i>HCFC1</i> : control of the cell cycle and transcriptional regulation; <i>IRAK1</i> : innate immunity	
Studi di varianti rare	1q25	NCF2	Oxidative stress	
Genome wide linkage study	1q41	DUSP10	Innate immunity and oxidative stress	
Fine mapping MHC	6p21	HLA-B	Peptide recognition and CD8 ⁺ T cell presentation	
	6p21	HLA-DPB1 ¹	Peptide recognition and CD4 ⁺ T cell presentation	
	6p21	HLA-F ¹	Peptide recognition and CD8 ⁺ T cell presentation	
	6p21	B3GALT4 ¹	Synthesis of gangliosides	
Meta-analysis of ImmunoChip data	20q13	ZNF335	Regulation of gene expression, chromatin remodeling	
	1p31	NFIA	Differentiation of myeloid lineages. Central nervous system development	

Tabella 1: Varianti di rischio genetico associate alla celiachia (da Dieli-Crimi R, 2015).

1.5 Epigenetica

Nell'ambito della genetica della malattia celiaca è ormai chiaro che nonostante il contributo fornito dai geni candidati, ad oggi resta ancora da chiarire circa il 50% dell'ereditarietà della patologia.

Ad oggi, i fattori genetici identificati, da soli, non giustificano l'aumentata incidenza di tale patologia negli ultimi decenni, anche in considerazione della stabilità del genoma umano nell'ultimo secolo. Nella patogenesi delle celiachia, come già menzionato, giocano un ruolo importante fattori ambientali e pertanto i meccanismi epigenetici possono rappresentare il punto di incontro tra genetica e ambiente. L'epigenetica è la branca della genetica che studia le modificazioni ereditabili che alterano la funzionalità del gene, senza alterare la sequenza nucleotidica del DNA. I due meccanismi epigenetici maggiormente studiati nei mammiferi comprendono le modificazioni degli istoni (acetilazione e metilazione) e la metilazione del DNA. Recentemente, è stata posta l'attenzione anche sui microRNA, piccoli RNA non codificanti, che agirebbero da fattori epigenetici (*Buoli Comani G, 2015*). Tali meccanismi agiscono modificando il grado di espressione di un gene, riducendolo o aumentandolo. Ad esempio, la metilazione del DNA all'estremità 5' del promotore di un gene porta al silenziamento del gene stesso, mentre le modificazioni degli istoni portano ad alterazioni del compattamento della cromatina, modulando l'espressione dei geni. Mentre il profilo genetico rimane stabile durante la vita dell'individuo, il profilo epigenetico è più facilmente e più rapidamente influenzato dai fattori ambientali (dieta, stress, agenti chimici, farmaci). Le epimutazioni durano per tutta la vita della cellula e vengono trasmesse alle cellule figlie durante la mitosi, portando ad un nuovo fenotipo acquisito, che potrà anche essere ereditato. Per tale motivo l'epigenoma rappresenta un sistema allo stesso tempo stabile, per la possibilità di trasmissione alle cellule figlie, e dinamico, per la possibilità che fattori casuali e ambientali lo modifichino nel tempo. Normalmente il profilo epigenetico va incontro ad un processo di riprogrammazione durante la gametogenesi. Nel caso in cui tale processo avvenga in maniera incompleta, le modifiche epigenetiche indotte dall'ambiente potranno essere trasmesse alla generazione successiva (ereditarietà epigenetica transgenerazionale). Quando la mutazione persiste per più generazioni, si può verificare il fenomeno dell'anticipazione epigenetica, che consiste nel progressivo aggravamento del fenotipo nelle generazioni successive, con un esordio più precoce ed un decorso più aggressivo della malattia.

Gli GWAS hanno dimostrato che il 90% degli SNP associati alla malattia si trovano in regioni non codificanti del genoma (regioni promotrici, enhancer e geni di RNA non codificanti),

mentre meno del 10% di essi influenzano sequenze codificanti, suggerendo che la maggioranza degli SNP associati alla malattia hanno un ruolo potenzialmente regolatorio (Ricano-Ponce I, 2013).

Lo studio dei meccanismi di regolazione epigenetica potrebbe aiutare a completare le informazioni riguardo alla ereditarietà della MC.

1.6 Differenziamento delle cellule intestinali

La mucosa intestinale è lo strato più interno dell'organo. Come tale, si affaccia direttamente sul lume dell'intestino, a stretto contatto con i prodotti della digestione. Nel tenue, la mucosa intestinale presenta un aspetto vellutato, conferitole dalla diffusa presenza di piccoli rilievi digitiformi chiamati villi intestinali. Quest'ultimi hanno lo scopo di aumentare la superficie assorbente, ulteriormente estesa dal sollevamento della mucosa in pieghe macroscopiche e dall'infossamento dello strato epiteliale: i numerosi ripiegamenti della mucosa intestinale del duodeno e del digiuno vengono detti pliche, mentre le invaginazioni tubulari dell'epitelio, che si portano in profondità nel tessuto connettivo di sostegno, sono dette cripte.

Si calcola che la mucosa intestinale sia ricoperta da circa 40 milioni di villi; ogni villo è costituito da un asse connettivale tappezzato da un singolo strato di cellule assorbenti, la cui membrana - rivolta verso il lume - presenta delle sottili estroflessioni citoplasmatiche a dito di guanto. Questi minuscoli rilievi, visibili soltanto al microscopio elettronico, sono chiamati microvilli e nel loro insieme costituiscono l'orletto a spazzola. La mucosa intestinale ha una straordinaria capacità di rigenerazione, che consente un rapido ripristino delle sue funzioni fisiologiche dopo un trauma. La capacità di fare questo risiede nelle cellule staminali epiteliali situate all'interno di invaginazioni ghiandolari della superficie della mucosa.

La mucosa intestinale è costituita da tre distinte porzioni: un singolo strato di **cellule epiteliali**, un tessuto connettivo che mantiene in sede l'epitelio (**la lamina propria**) ed un esiguo strato di muscolo liscio, chiamato *muscularis mucosae*, che la separa dalla tunica sottostante (Fig. 4).

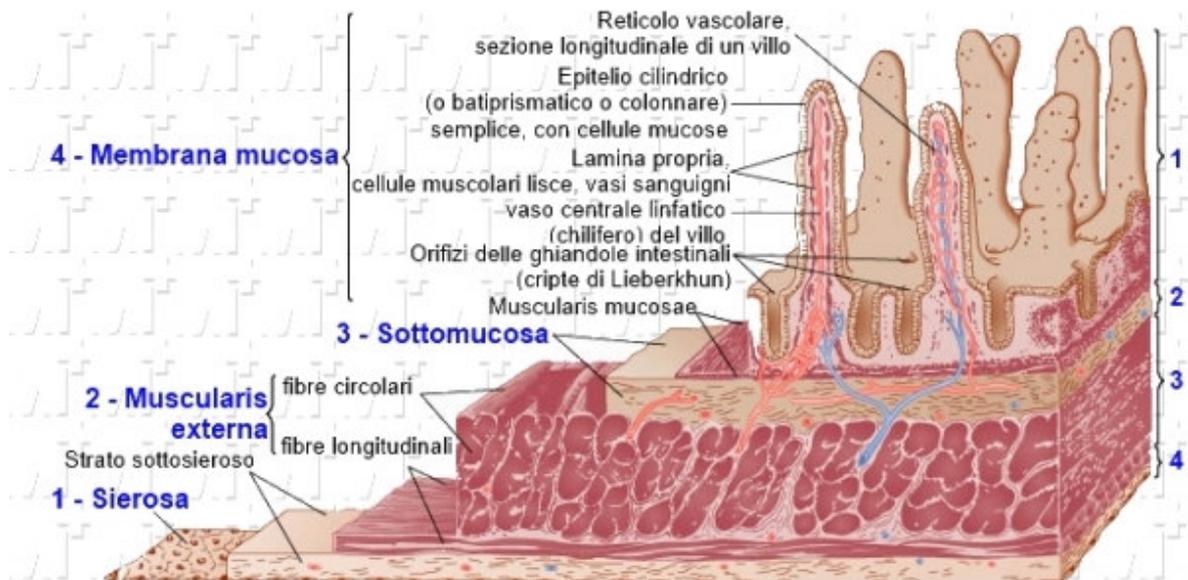


Figura 4: Struttura della mucosa intestinale

1.6.1 Epitelio

L'epitelio dell'intestino è uno strato di cellule confluenti costituito da un certo numero di sottotipi cellulari ciascuno con funzioni specializzate che aiutano a svolgere due importanti funzioni fisiologiche: (1) fornire una barriera efficace tra il contenuto del lume del tratto intestinale e la mucosa sterile sottostante, e (2) assorbire efficacemente acqua e sostanze nutritive dal cibo ingerito. La perdita di una di queste funzioni è associata a vari gradi di morbilità, e nella peggiore delle ipotesi, se lasciata incontrollata, può essere fatale. L'organismo ha dunque evoluto un meccanismo efficace di mantenimento dell'omeostasi nell'epitelio garantendo un flusso costante di cellule disponibili a ripopolare la mucosa. Quindi, mucosa ed epitelio subiscono un rapido turnover continuo per tutta la vita, con l'intero rivestimento epiteliale sostituito ogni 3-5 giorni. Tuttavia la capacità potenziale di questo meccanismo è dimostrata solo in condizioni patologiche quando si verificano gravi danni per l'intestino, per esempio a seguito di trauma o malattia, con conseguente aumento della rigenerazione epiteliale.

L'omeostasi intestinale è strettamente controllata da vie di trasduzione del segnale ben caratterizzate. In particolare le vie di Wnt, Bmp, EGF e Notch sembrano giocare un ruolo importante in questa regolazione. Il segnale Wnt ha un ruolo chiave nel sostenere le cellule staminali e nel guidare la proliferazione sul fondo delle cripte. Topi deficitari per fattori di trascrizione chiave di questa via (Tcf7l2 e Tcf4), perdono la capacità proliferativa delle cripte. Al contrario il cancro del colon è causato dalla perdita del gene *Apc* che codifica per un regolatore negativo della via Wnt. In aggiunta al segnale Wnt, la stimolazione della via di

Notch è richiesta per mantenere le cellule delle cripte in uno stato indifferenziato. Il segnali Bmp e Tgf- β sono attivati nelle cellule dei villi, e si crede regolino il differenziamento. I ligandi di Bmp sono espressi nei villi, mentre antagonisti come Noggin e Gremlins sono espressi nel mesenchima delle cripte per bloccare il segnale *Bmp* (Laurens G. van derFlier, 2009).

Tutte queste vie del segnale interagiscono l'una con l'altra, formando complesse reti molecolari che regolano l'omeostasi tissutale (Fig.5).

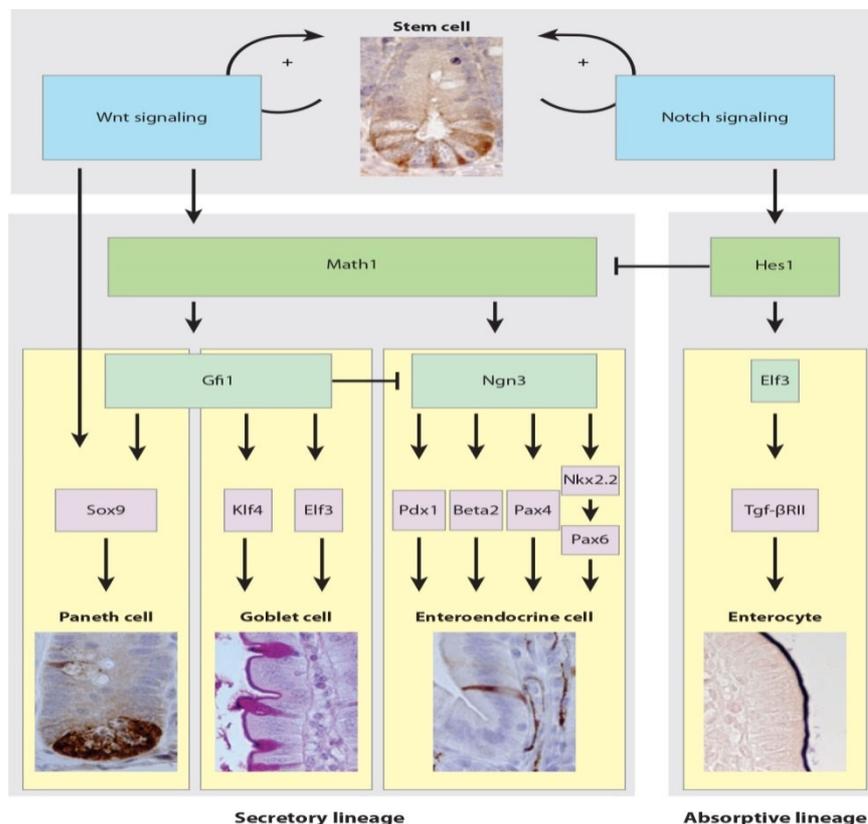


Figura 5: Coinvolgimento del signaling Wnt e Notch nel differenziamento di cellule intestinali.

Nell'epitelio mucosale intestinale sono stati trovati almeno cinque tipi di cellule; esse si trovano sia nelle ghiandole intestinali che sulla superficie dei villi (Fig.6). I principali tipi sono:

- **Enterociti** : sono specializzati per il trasporto di sostanze dal lume intestinale al sistema circolatorio, sono lunghe cellule colonnari con nucleo posto alla base, la cui funzione principale è l'assorbimento di nutrienti.
- **Cellule Goblet (o caliciformi)** : ghiandole unicellulari secernenti di mucina, localizzate sparse tra gli enterociti e caratterizzate da un grande vacuolo.

- **Cellule di Paneth** : si trovano alla base delle cripte nel piccolo intestino e nel colon ascendente e sono caratterizzate da granuli fortemente eosinofili, che contengono lisozima e defensine antibatteriche.
- **Cellule enteroendocrine** : le quali producono vari ormoni paracrini ed endocrini , sparsi in tutti l'epitelio.
- **Cellule M** : enterociti modificati che coprono noduli linfatici ingranditi nella lamina propria.

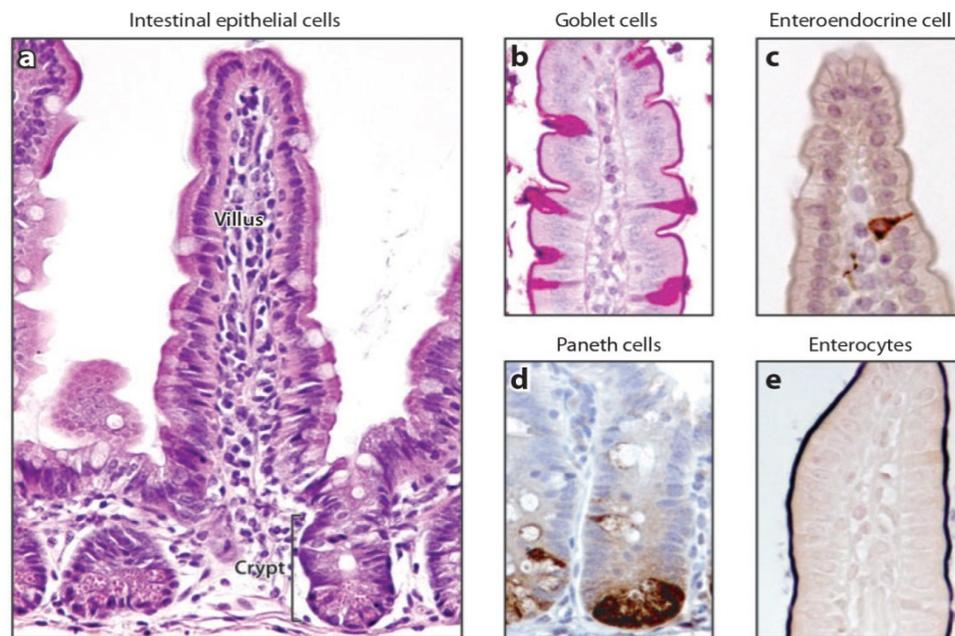


Figura 1

Figura 6: Epitelio intestinale con le diverse cellule che lo rappresentano.

Enterociti, cellule caliciformi e cellule enteroendocrine diventano più differenziate quando maturano e migrano verso l'alto dalla base della cripta verso la punta dei villi. Le cellule Paneth, invece, completano il loro differenziamento terminale quando migrano verso la base della cripta (Fig.7).

Una volta che le cellule epiteliali raggiungono la punta dei villi subiscono apoptosi spontanea e sono perse nel lume intestinale o fagocitate. Nel tratto gastrointestinale sano, la quantità netta di apoptosi è uguale al flusso di cellule epiteliali che migrano sui villi ogni giorno, mantenendo l'omeostasi della popolazione di cellule epiteliali.

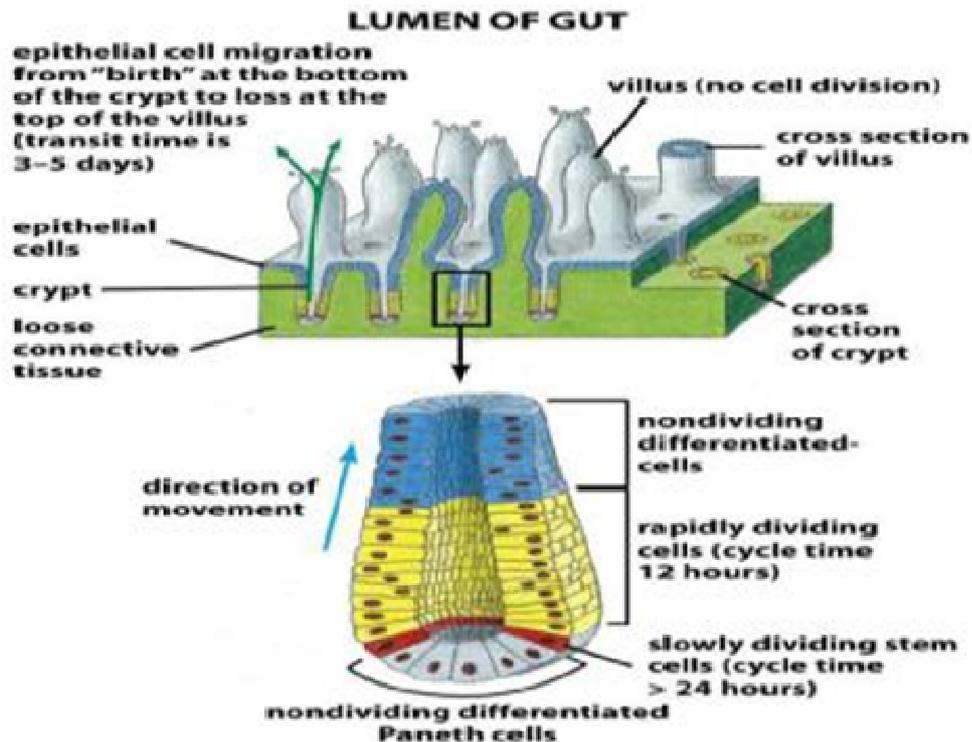


Figura 7:Struttura delle cripte

Un antigene di membrana comune a tutti i tipi di cellule epiteliali è rappresentato dalla Molecola Epiteliale di Adesione Cellulare (EpCAM, CD326), una glicoproteina transmembrana che media l'adesione omotipica cellula-cellula Ca^{2+} indipendente in epitelii. In questo studio EpCAM è stato utilizzato come elemento discriminante per effettuare una procedura di isolamento e separazione tra cellule epiteliali e cellule della lamina propria, presenti a livello del tessuto intestinale.

1.6.2 Lamina propria

Come nello stomaco, la lamina propria circonda le ghiandole intestinali e contiene numerose cellule del GALT (*Gut-Associated Lymphoid Tissue*) come linfociti, cellule plasmatiche, mastociti, macrofagi ed eosinofili; contiene altresì numerosi noduli di tessuto linfatico che rappresentano la componente principale del sistema immunitario intestinale. I noduli sono particolarmente grandi e frequenti nell'ileo, dove sono posti preferibilmente sul lato dell'intestino opposto al fissaggio mesenterico. Queste aggregazioni nodulari sono note anche con il nome di noduli aggregati o placche di Peyer. Quest'ultime sono quindi costituite da noduli linfatici che occupano la lamina propria e la sottomucosa dell'ileo; il nucleo delle placche di Peyer è costituito da grossi Linfociti B in proliferazione, mentre la porzione più

periferica presenta linfociti B in uno stato quiescente. Tra le Placche di Peyer è possibile trovare Linfociti T e una serie di cellule APC dette cellule M.

I linfociti sono continuamente in movimento attraverso il flusso sanguigno e linfatico, i tessuti linfoidei secondari, e i tessuti periferici non linfoidei; le distinte popolazioni di linfociti mostrano differenti pattern di “trafficking” attraverso questi siti. I linfociti “naive” si muoveranno ripetute volte dal flusso sanguigno, dentro i linfonodi, e dal flusso linfatico verso il sangue, finché non incontreranno e riconosceranno un antigene dentro un linfonodo. Questo “traffic pattern”, chiamato ricircolazione linfoitaria, permette al limitato numero di linfociti “naive” che sono specifici per un particolare antigene di cercare quell’antigene attraverso il corpo. I linfociti che hanno reagito con l’antigene e si differenziano in cellule effettrici e cellule della memoria dentro i tessuti linfoidei secondari potrebbero ritornare verso il flusso sanguigno ed in seguito migrare in siti di infezione e/o infiammazione nei tessuti periferici non linfoidei. Qualche linfocita effettrici migrerà preferenzialmente in particolari tessuti, come la lamina propria della mucosa intestinale. Il processo attraverso cui una particolare popolazione di linfociti entra selettivamente nei linfonodi o in alcuni tessuti, viene chiamato homing. L’esistenza di differenti homing assicura che differenti linfociti siano presenti nel microambiente del tessuto dove sono effettivamente richiesti per una risposta immuno-adattativa e che non vengano così “sprecati”. I diversi tipi di linfociti come ad esempio il linfociti T “naive” giungono ai tessuti linfoidei secondari attraverso il flusso sanguigno, lasciano la circolazione e migrano nello stroma dei linfonodi grazie a venule post-capillari modificate chiamate high endothelial venules (HEVs). I linfociti T “naive” prediligono i linfonodi, mentre quelli effettrici escono dai linfonodi ed hanno un homing preferenziale per i tessuti periferici in cui si hanno siti di infezione, dove sono necessari per eliminare i microbi durante la fase attiva della risposta immunitaria adattativa. La presenza dei linfociti naive è resa possibile grazie al legame della L-selectina sui linfociti T, delle addressine dei linfonodi periferici sulle HEVs, e attraverso il recettore CCR7 sulle cellule T che legano le chemochine prodotte nei linfonodi. Le cellule T effettrici e della memoria che sono generate dalla stimolazione antigenica dei linfociti “naive” escono dal linfonodo. Esse diminuiscono la presenza di L-selectina e del recettore CCR7 ma aumentano l’espressione delle integrine e dei ligandi della E-selectina e della P-selectina, che mediano il legame all’endotelio nei siti di infiammazione periferici. Entrambi i tipi di linfociti esprimono anche i recettori per le chemochine prodotte nei tessuti periferici. I linfociti B utilizzano lo stesso meccanismo di base dei linfociti T “naive” per compiere l’homing verso i tessuti linfoidei secondari del corpo, aumentando così la probabilità di poter rispondere alla presenza dei microbi nei differenti siti.

Come sappiamo la maturazione dei linfociti B implica che questi prima di tutto fuoriescano dal midollo osseo ed attraverso il sangue giungano alla milza dove iniziano il primo processo di maturazione (zona follicolare della milza). Quando la maturazione è completata le cellule B “naive” rientrano nella circolazione e raggiungono i linfonodi ed i tessuti linfoide delle mucose. Questo homing coinvolge l’interazione per il legame e il “rolling” sulle HEV, l’attivazione delle integrine mediata dalle chemochine ed un arresto stabile. Le cellule B esprimono le L-selectine, CCR7 e LFA1 che legano PNAs, CXCL19 e CXCL21 e ICAM-1 sulle HEVs dei linfonodi. In più esprimono altri recettori chemotattici chiamati CXCR4 che lega la CXCL12 e questa interazione gioca un ruolo significativo nell’attivazione e nell’homing dei linfociti B. Una volta che queste cellule sono entrate nello stroma dei linfonodi secondari, migrano nei follicoli, i siti dove potranno incontrare gli antigeni e attivarsi; questa migrazione è mediata da CXCL13 la quale è prodotta nei follicoli. L’homing nelle placche di Peyer coinvolge CXCR5 e l’integrina $\alpha 4\beta 7$ (up-regolata in lamina propria), che lega MadCAM1 (mucosal vascular addressin cell adhesion molecule), un addressina coinvolta nell’adesione cellulare. (Aida Habtezion, , 2016)(Fig. 8).

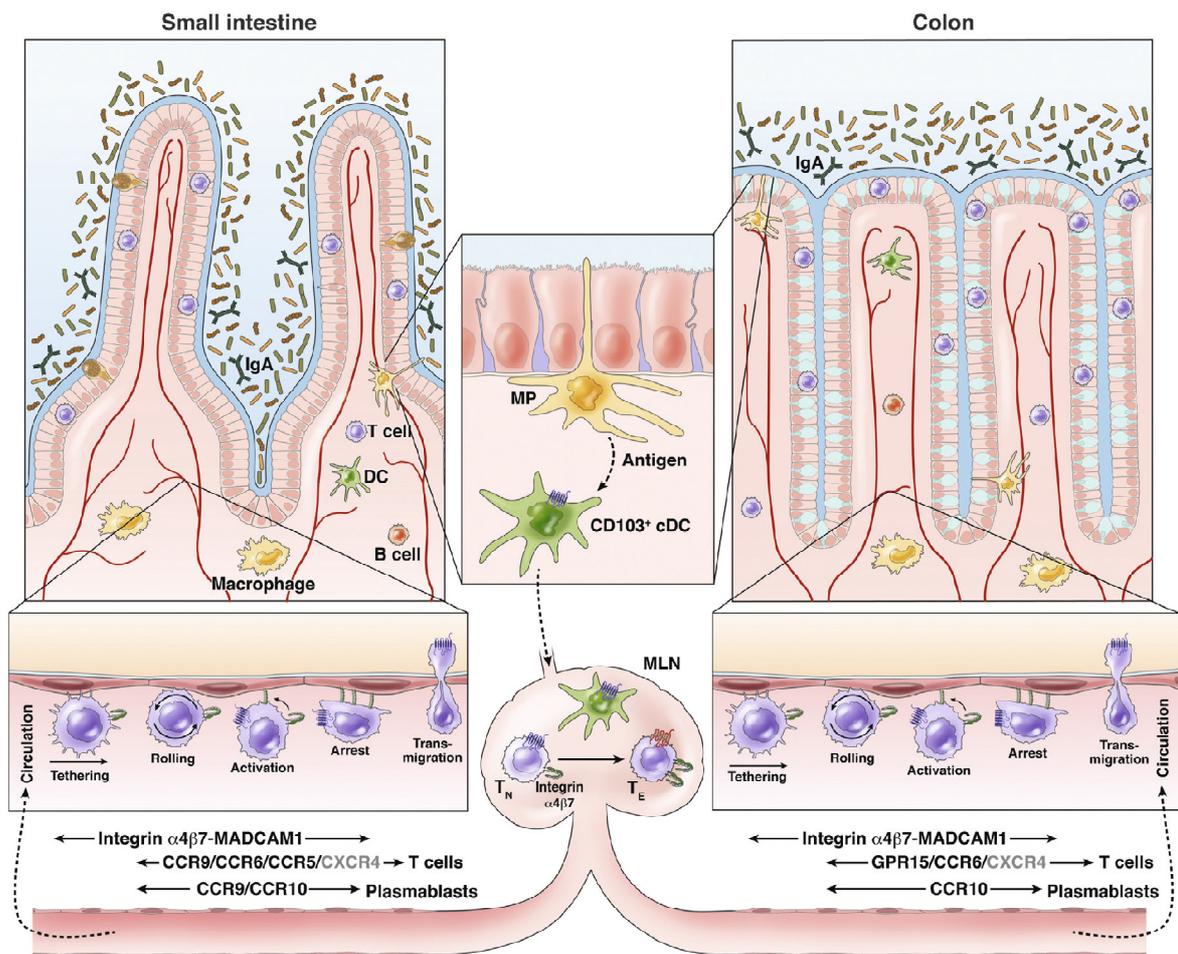


Figura8. Homing dei Linfociti nel piccolo intestino e nel colon

2. SCOPO DELLA TESI

Per quanto concerne la malattia celiaca non è stata identificata una forte associazione con l'alterazione in un singolo gene, nonostante siano stati numerosissimi i lavori condotti in questo senso.

Oltre all'implicazione di meccanismi di regolazione genica di altro tipo (epigenetica, microRNA ecc.), una possibile spiegazione del fallimento di questo tipo di approccio è che la patologia non sia determinata dal "default" di un singolo meccanismo, ma bensì sia il risultato da una serie di alterazioni a carico di processi fisiologici che nell'insieme determinano l'insorgere del "fenotipo celiaco".

Lo scopo del nostro studio è stato quello di valutare le eventuali differenze nei livelli di espressione di alcuni geni candidati nella malattia celiaca in diversi distretti cellulari tra pazienti celiaci e controlli. In primo luogo ci soffermeremo sull'analisi dei livelli di espressione in cellule mononucleate di sangue periferico (PBMC) di soggetti predisposti geneticamente e con una accertata familiarità per la malattia celiaca. La coorte da noi analizzata consiste in bambini predisposti geneticamente (HLA DQ 2 e/o 8 positivi) con almeno un familiare di primo grado affetto da MC, arruolati e seguiti in follow-up dalla nascita al compimento di 6 anni di età. Grazie alle particolari caratteristiche della di questi soggetti, che presentano un rischio molto elevato di sviluppare la celiachia, è stato possibile condurre un' analisi longitudinale, avendo a disposizione campioni raccolti sia prima dello sviluppo della malattia, sia al momento della diagnosi, sia in remissione clinica. L'unicità della nostra popolazione ci ha permesso di indagare l'esistenza di eventuali alterazioni molecolari presintomatiche e di valutarne poi l'evoluzione al momento della diagnosi della patologia e dopo un anno a dieta senza glutine. In aggiunta alla valutazione dei livelli di espressione in questa coorte sarà effettuata anche una analisi genotipo dei geni candidati, in quanto essendo soggetti appartenenti a famiglie a rischio era importante una corretta stratificazione dei fattori di rischio genetico di ciascun componente.

Tutti gli studi che nella MC si sono concentrati sulla mucosa intestinale, hanno analizzato il tessuto senza fare distinzione tra i due compartimenti che lo compongono e che presentano caratteristiche molto diverse: l'epitelio e lamina propria. Ad oggi, non è stato mostrato chiaramente in che modo ciascuna componente della mucosa intestinale sia coinvolta nel processo patogenetico della celiachia. Quindi partendo da queste premesse nella seconda parte dello studio analizzeremo gli stessi geni candidati studiati nella prima parte della ricerca, la

presenza di alterazioni sia geniche che epigenetiche(in particolare la metilazione) in cellule intestinali selezionate sia di natura epiteliale che appartenenti alla *lamina propria*.

Lo scopo finale di questo lavoro di tesi è stato, dunque, la valutazione globale delle alterazioni di espressione genica in geni candidati, in modo da tentare di identificare a livello di singole e distinte popolazioni cellulari, eventuali alterazioni costanti e coerenti con il quadro patogenetico della MC, e in seconda battuta di definire i meccanismi di regolazione epigenetica che sottendono tali alterazioni.

3.MATERIALI E METODI

3.1 Pazienti.

3.1.1 Pazienti arruolati nello studio longitudinale su cellule di sangue periferico

I pazienti analizzati nel nostro studio appartengono alla coorte di famiglie arruolate per lo studio PREVENT-CD un progetto di ricerca europeo che studia l'influenza della nutrizione in bambini appartenenti a famiglie a rischio per MC, e dei fattori genetici, ambientali e immunologici con il rischio di sviluppare la malattia (*Hogen Esch C2010;Vriezinga S.L2014*). In particolare queste famiglie presentano un forte rischio genetico per la MC, sono famiglie in cui è presente almeno un familiare di primo grado celiaco del probando arruolato, che deve presentare HLA-DQ2 e/o -DQ8 positivo, tutti i bambini sono stati monitorati dalla nascita fino ai 6 anni,

I soggetti sono stati monitorati clinicamente e sierologicamente (anticorpi anti-tessuto-transglutaminasi (anti-tTG) ogni 3 - 6 mesi dalla nascita a 6 anni. Nei bambini che hanno sviluppato anticorpi anti-tTG, la diagnosi è stata confermata da una biopsia intestinale. Abbiamo analizzato 9 bambini che hanno sviluppato la malattia celiaca ad una età media di 30 mesi e 13 controlli (bambini della stessa coorte che non hanno sviluppato la celiachia fino a 6 anni). Tra i pazienti che hanno sviluppato la malattia, 5 erano asintomatici, 4 avevano sintomi, 3 hanno mostrato ritardo di crescita, di cui uno è stato associato alla diarrea, il quarto ha avuto ricorrenti dolori addominali.

I pazienti sono stati sottoposti ad un prelievo di sangue almeno 1 volta all'anno durante lo studio per misurare la concentrazione sierica degli anticorpi antitransglutaminasi e antigliadina e in contemporanea hanno effettuato una visita medica in cui sono state monitorate lo stato di salute, le misure antropometriche (altezza, peso e circonferenza cranica), le abitudini alimentari e il consumo di glutine mediante specifici questionari (*Vriezinga S.L 2014*).

Nella coorte PREVENT-CD studiata nel nostro centro 24 bambini sono diventati celiaci prima dei 6 anni di età (*Galatola M,2017*). Purtroppo, il numero di campioni sufficienti per dimensioni e qualità per questo studio era inferiore a causa del progressivo deterioramento dei campioni per il lungo periodo che intercorre dal reclutamento alla divulgazione dei risultati dello studio in doppio cieco (> 6 anni). Pertanto, solo per i 9 pazienti qui analizzati, siamo stati in grado di utilizzare il materiale biologico in tutti i punti di tempo stabiliti in questo lavoro:

- T1: prima della diagnosi

- T2: al momento della diagnosi
- T3: un anno dopo la diagnosi

3.1.2 Pazienti arruolati nello studio su cellule intestinali

Gli studi di espressione dei geni candidati in cellule isolate da mucosa intestinale sono state condotte su campioni biotici di mucosa duodenale di:

- *19 celiaci floridi* con atrofia mucosale
- *21 controlli sani*

Le cui caratteristiche riassunte in tabella 2. I pazienti sono stati arruolati presso il reparto di Gastroenterologia, endoscopia digestiva e Nutrizione Clinica del ospedale Burlo Garofolo di Trieste, i relativi campioni biotici di mucosa duodenale sono stati prelevati mediante EGDS e conservati ad una temperatura di -80°C. Le cellule di questi 40 soggetti sono state isolate dalle biopsie duodenali, e successivamente separate (come descritto nel paragrafo seguente) utilizzando come marcatore discriminante un antigene di membrana, CD326 (EpCAM).

ID	Diagnosi	anti-tTG U/ml (vn>7)	EMA	Istologia	Depositi
CD01	Celiachia	9	+	M3	+
CD04	Celiachia	25	+	M2	+
CD05	Celiachia	222	+	M3	+
CD07	Celiachia	142	+	M3	+
CD08	Celiachia	142	+	M3	+
CD15	Celiachia	128	+	M3	+
CD18	Celiachia	201	+	M3	+
CD19	Celiachia	5	+	M3	+
CD20	Celiachia	50	+	M3	+
CD23	Celiachia	21	+	M0	+
CD28	Celiachia	9	+	M0	+
CD30	Celiachia	50	+	M3	+
CD33	Celiachia	44	+	M3	+
CD34	Celiachia	148	+	M3	+
CD37	Celiachia	208	+	M3	+
CD41	Celiachia	171	+	M3	+
CD42	Celiachia	134	+	M3	+
CD43	Celiachia	191	+	M3	+
CD50	Celiachia	4	+	M1	+
CTR02	Celiachia negata	999	999	999	999
CTR10	Esofagite eosinofila	999	999	999	999
CTR11	Gastrite	999	999	999	999
CTR12	Esofagite da reflusso	999	999	999	999
CTR16	Celiachia negata	999	999	999	999
CTR17	dolore addominale funzionale	999	999	999	999
CTR21	Ingestione caustici	999	999	999	999
CTR24	Esofagite da reflusso	999	999	999	999
CTR25	Celiachia negata	999	999	999	999
CTR27	Celiachia negata	999	999	999	999
CTR29	Gastrite	999	999	999	999
CTR32	Esofagite eosinofila	999	999	999	999
CTR36	Celiachia negata	999	999	999	999
CTR38	Celiachia negata	999	999	999	999
CTR39	Esofagite da reflusso	999	999	999	999
CTR40	sospetta Sindrome Poliendocrina Autoimmune (APECED)	999	999	999	999
CTR44	Gastrite	999	999	999	999
CTR45	Celiachia negata	999	999	999	999
CTR46	Esofagite eosinofila	999	999	999	999
CTR48	Celiachia negata	999	999	999	999
CTR49	Celiachia negata	999	999	999	999

Tabella2: Pazienti arruolati per l'isolamento delle cellule intestinali

3.2 Isolamento delle cellule mononucleari del sangue periferico

Dai campioni di sangue periferico (4-5mL) sono state isolate le PBMCs mediante il *Ficoll*(Histopaque[®]) secondo il protocollo del produttore. Le PBMC sono state poi congelate in contenitori crioprotettivi contenenti il 10% dimetilsolfossido e siero fetale bovino.

3.3 L'isolamento delle cellule intestinali CD326+ da quelle CD326-

L'isolamento delle **cellule epiteliali** dalle **cellule della lamina propria** è stato eseguito mediante l'utilizzo di *CD326 (EpCAM) MicroBeads[®] human MACs (Miltenyi Biotec GmbH, Germany)*. Tale kit consente di ottenere cellule CD326(EpCAM)+ da una sospensione cellulare, allontanando dal campione tutte le altre tipologie cellulari (linfociti B e T, cellule NK, eritrociti, cellule dendritiche etc.) mediante l'utilizzo di piccole sfere di polistirene magnetiche (*Microbeads*) coniugate con il suddetto antigene di membrana. Le Epithelial Cell Adhesion Molecule (EpCAM), conosciute anche come CD326 o antigene epiteliale umano (HEA), sono ampiamente espresse sulle cellule di origine epiteliale e su cellule tumorali di derivazione epiteliale (*Moldenhauer, G 1987*).

3.3 Estrazione di acidi nucleici e retrotrascrizione dell'RNA

L'RNA totale è stato estratto dalle PBMC utilizzando il reagente TRIZOL(Life Technologies). La quantità di RNA è stata misurata usando un Nanodrop spettrofotometro e la qualità è stata analizzata mediante elettroforesi del gel di agarosio. L'RNA (1µg) è stato trascritto retrotrascritto in cDNA utilizzando il kit *High-Capacity celiaciNA Reverse Transcription Kit* della Applied Biosystems[®]. Mentre il DNA è stato estratto da sangue venoso, raccolto in una provetta con EDTA (per ogni paziente) con il metodo del *Salting Out*. Per quanto riguarda le cellule intestinali gli acidi nucleici ci sono stati invece ottenuti mediante l'utilizzo del kit *AllPrep DNA/RNA Mini Kit*(QIAGEN) che consente la simultanea estrazione di DNA e RNA dallo stesso campione biologico con un'unica procedura. Il kit utilizza una separazione su colonna al termine della quale seguendo il protocollo fornito dalla ditta si ottengono mediante due diverse eluizioni i due acidi nucleici.

L'RNA è stato retrotrascritto in cDNA utilizzando ancora una volta il kit *High-Capacity celiaciNA Reverse Transcription Kit* della Applied Biosystems[®] partendo da una quantità di 1µg.

3.4 Real-Time PCR

Gli esperimenti di Real Time PCR sono stati eseguiti mediante metodologia Taqman (7900T) utilizzando sonde TaqMan Gene Expression (Life Technologies). L'espressione relativa di ciascun gene è stata ottenuta utilizzando il metodo $\Delta\Delta\text{Ct}$ e normalizzata con l'espressione di un gene endogeno (GUSb). Tutte le reazioni sono state effettuate in triplicato ed è stata presa in considerazione la media dei tre risultati. Gli *assay on demand*, relativi ai geni candidati selezionati nel nostro studio, sono stati forniti dall'Applied Biosystem e sono elencati nella tabella 3, e come è stato discusso nell'introduzione sono stati distinti in 4 categorie in base alla loro funzione biologica:

1. Geni Direttamente coinvolti nell'infiammazione/danno cellulare
2. Geni candidati "classici" in quanto risultati fortemente associati alla MC da un punto di vista genetico
3. Geni coinvolti nel homing di linfociti e nella regolazione dei processi infiammatori e del danno cellulare
4. Geni coinvolti nell'adesione cellulare ed integrità della barriera intestinale

GENE	Popolazione Cellulare Analizzata	Funzione	Funzione biologica	Assay Taqman
IL12A	Cellule Intestinali	Proinflammatory cytokine	Infiammazione/Danno cellulare	Hs00222327_ml
IL21	Cellule Intestinali	Pleiotropic cytokine	Infiammazione/Danno cellulare	Hs00168405_ml
NFKB1	Cellule Intestinali	Regulation of autoimmunity and inflammation	Infiammazione/Danno cellulare	Hs00765730_ml
C-REL	PBMC Cellule Intestinali	Subunit of the NF-κB transcription complex	Infiammazione/Danno cellulare	Hs00968436_ml
TNFAIP3	PBMC Cellule Intestinali	Negative feedback loop control of NF-κB	Infiammazione/Danno cellulare	Hs00234713_ml
KIAA1109	PBMC Cellule Intestinali	Located in the genomic region associated with CeD	Canidato/Associato	Hs00361070_ml
SH2B3	PBMC Cellule Intestinali	Activates PI3-kinase	Regolazione di infiammazione/Danno cellulare	Hs00193878_ml
RGS1	PBMC Cellule Intestinali	Lymphocytes Homing	Regolazione di infiammazione/Danno cellulare	Hs00175260_ml
TAGAP	PBMC Cellule Intestinali	Negative regulator of the immune response	Regolazione di infiammazione/Danno cellulare	Hs00611823_ml
TNFRSF14	PBMC Cellule Intestinali	Activation of NK intestinal and CD4 + T cells "gut-homing cells"	Regolazione di infiammazione/Danno cellulare	Hs00998604_ml
TNFSF14	PBMC Cellule Intestinali	Activation of NK intestinal and CD4 + T cells "gut-homing cells"	Regolazione di infiammazione/Danno cellulare	Hs00542477_ml
LPP	PBMC Cellule Intestinali	Extracellular matrix and cell-cell contacts homeostasis	Adesione cellulare ed integrità della barriera intestinale	Hs00944352_ml
TJP1	Cellule Intestinali	Proteins of the tight junctions, involved in maintaining the integrity of the intestinal barrier.	Adesione cellulare ed integrità della barriera intestinale	Hs01551861_ml
PTPRK	Cellule Intestinali	Maintenance of cell junctions and in the modulation of EGFR activity, resulting in an inhibition of cell proliferation.	Adesione cellulare ed integrità della barriera intestinale	Hs00267788_ml
ARHGAP31	Cellule Intestinali	Regulation of cell migration, Focal Adhesion size and dynamics	Adesione cellulare ed integrità della barriera intestinale	Hs00393361_ml
C1orf106	Cellule Intestinali	Involved in cell adhesion processes	Adesione cellulare ed integrità della barriera intestinale	Hs01009089_ml

Tabella 3: Geni candidati analizzati nel lavoro con le relative funzioni cellulari e le sonde utilizzate per eseguire gli esperimenti di quantizzazione.

3.5 Genotipizzazione

Tutti i pazienti arruolati nel lavoro sono stati genotipizzati per i geni HLA (*Izzo V,2011; Monsuur AJ;2008*) e raggruppati in 5 classi di rischio HLA, come riportato in precedenti lavori (*Monsuur AJ,2008*).I pazienti appartenenti alla coorte di famiglie a rischio analizzati nello studio longitudinale sono stati genotipizzati anche 9 SNPs non-HLA analizzati, localizzati nei geni candidati.

Le genotipizzazioni sono state effettuate mediante tecnologia TaqMan(Life Technologies), utilizzando TaqMan SNP Genotyping Assays mediante la Real Time PCR (7900T) (Applied Biosystems, Foster City, CA); il volume finale della reazione è di 15 mL, contenente master mix, TaqMan assays and circa 50 ng di DNA genomico da analizzare.

3.6 Analisi statistica

È stato adottato un test non parametrico(Mann-Whitney) per confrontare i livelli di espressione genica a causa della piccola dimensione del campione. Le frequenze sono state confrontate usando il χ^2 -test, con probabilità di ipotesi nulle pari a 0,05.

Per stimare il contributo del genotipo e/o l'espressione di ciascun gene per distinguere i pazienti dai controlli è stata condotta una analisi discriminante multivariata. L'analisi discriminante viene condotta per definire una modalità di assegnazione dei casi a differenti gruppi (nel nostro caso Celiaci e Controlli), in funzione di una serie di variabili fra di loro correlata (rappresentate dai risultati di genotipizzazione e analisi dell'espressione genica).

Lambda di Wilks che consiste in un test multivariato per valutare l'ipotesi che due o più gruppi provengano da popolazioni con la stessa media, detta anche statistica U. Lambda varia tra 0 ed 1. Valori elevati indicano che le medie di gruppo non sono molto differenti (e pari a 1 se le medie sono tutte uguali). Valori bassi indicano differenze nelle medie di gruppo.

Il Lambda di Wilks fornisce quindi una stima della capacità di discriminazione cumulativa tra due gruppi prodotti dalla combinazione multivariata di variabili.

Moltiplicando il valore standardizzato di ciascuna variabile inclusa nell'equazione discriminante "stepwise" con il suo rispettivo coefficiente di regressione, otteniamo un D-Score che esprime la posizione di ciascun individuo a destra o a sinistra: è quindi possibile assegnare l'individuo al gruppo di pazienti o al gruppo di controlli, solo sulla base dei dati genetici. Pertanto, può essere ottenuta la percentuale di classificazione corretta. Le analisi statistiche sono state eseguite con SPSS 17.0 (SPSS Inc, Chicago, IL).

3.7 Studi di Metilazione

Gli studi di metilazione sono stati condotti nel corso del primo anno di dottorato quando ho trascorso tre mesi presso il laboratorio “*Immunogenetics Research Laboratory*” dell’Università dei Paesi Baschi, del Prof. Jose Ramon Bilbao con il quale ancora oggi collaboriamo. Gli studi son stati effettuati su tutti i soggetti arruolati per l’analisi delle cellule intestinali. In particolare dalle cellule epiteliali e della *lamina propria* isolate dai campioni biotipici è stato estratto anche il DNA (come descritto nei paragrafi precedenti).

3.7.1 Conversione con Bisolfito

Diversi metodi sono stati messi a punto per rivelare e quantificare la metilazione del DNA. Essi normalmente si avvalgono di un trattamento del DNA con sodio bisolfito, attraverso il quale le citosine non metilate vengono deamminate e solfonate per essere convertite in uracile, mentre le 5’-metil-citosine rimangono inalterate. Il trattamento consente quindi di discriminare le citosine non metilate da quelle metilate, le quali vengono rivelate da successive analisi. Per il trattamento del DNA con bisolfito di sodio è stato utilizzato il kit *Bisulfite Conversion* della Active Motif®.I campioni così ottenuti vengono successivamente analizzati mediante il saggio denominato Methylation Specific PCR.

Una curva di calibrazione è stata usata per quantizzare lo stato di metilazione dei campioni. La curva è stata costruita combinando diverse concentrazioni di Dna HeLa genomico disponibili commercialmente sia come Dna non metilato sia come Dna CpG metilato.I livelli di concentrazione della curva sono i seguenti:0,12,5, 25, 37.5, 50, 62.5, 75, 87.5 e 100%.

3.7.2 Methylation Specific PCR (MSP)

La Methylation Specific PCR (MSP) è una tecnica ampiamente utilizzata per lo studio della metilazione delle isole CpG. Le differenze generate dopo il trattamento con Na-bisolfito tra citosine metilate e non sono alla base del funzionamento dell’MSP. I primers per la MSP sono stati disegnati con l’ausilio del programma methprimer (<http://www.urogene.org/methprimer/>) (Li Lc,2002) sulla base della sequenza di DNA delle isole CpG da noi identificate attraverso il *Genome Browser* (Tab.4).

Posizione UCSC (GRCh37/hg19)	Chr.	Locus	Lunghezza regione	Isola CpG selezionata	Lunghezza Isola CpG	Met-Primers	Lunghezza Amplicone
61108026-61109316	2	cREL	1291 bp	80	757 bp	FP-GGTTTTTTGATTGATTGATTG RP-CCATATAAAATTTTAAAAACAAAAC	216bp
88706112.88706792	16	IL21	681 bp	30	305bp	FP-TTTAGGAGGATAGGGTAGGGTTAGA RP-AACAAAAACCCCAATATAAAAAAC	247 bp
103421992-103423529	4	NFKB1	1538 bp	100	207 bp	FP-GGAGGTTGATAGTAGTTGAGAGGTATAT RP-CATCCCCCAATAAACTAAAAATTA	205 bp
200859977-200860876	1	C1orf106	900bp	53	248 bp	FP-GGTATTGGGATTAAGAGAGAAGTTT RP-TTTACAACATCCCAATATTAATAATC	300bp
128840059-128842426	6	PTPRK	2368bp	163	847 bp	FP-TTTTTTTATTTTGTTGTTTTTATTTT RP-ACTCCTCTCTCTTAACTCTCTCTA	192 bp
159706391-159707311	3	IL12A	921 bp	52	137 bp	FP-TAAAAATGTGGTTTTTTGGGTTAGT RP-ACAAACTTAAATTTCTACCCACTCTTC	209 bp
111842839-111844692	12	SH2B3	1854 bp	132	796 bp	FP-TATTTTTTTTAGAATTGGGGGATAG RP-CCCTCTACAAAACCTTATCAATAC	260 bp
30114011-30115515	15	TJP1	1505 bp	124	1505 bp	FP-TATGGGGTGAGTTTTATGTATGAGT RP-ATCACCATTTCACCTTAACAATC	262 bp
123300308-123301105	4	KIAA1109	798 bp	45	399 bp	FP-TTATTGGTTTTTTTAGAAGGGTAAG RP-AAATTCACAAAATCAAAAACC	287 bp
187871379-187872654	3	LPP	1276 bp	67	448 bp	FP-TTTTTTTATTTTTATATGGATAAATG RP-TACAACCAACCTTACCTTACAAC	286 bp
137867116-137868004	6	TNFAIP3	889 bp	82	766 bp	FP-TTATTGGTTTTTTTAGAAGGGTAAG RP-AAATTCACAAAATCAAAAACC	300 bp

Tabella 4. Isole CpG e relativi amplicone scelti per l'analisi di metilazione. I geni candidati scelti per l'analisi di espressione per cui non erano disponibili CpG non sono potuti essere oggetto dell'analisi di metilazione

In particolare, due set di primers sono stati disegnati: un set di primers che riconosce il DNA modificato dal trattamento con sodio bisolfito e non metilato, e un set di primers che riconosce il DNA modificato e metilato. Inoltre, con il programma Primer3 è stato disegnato un terzo set di primers specifici per il DNA non modificato, necessari per verificare l'efficienza del trattamento col bisolfito. I campioni di DNA trattati col Na-bisolfito sono stati amplificati con i tre diversi set di primers utilizzando *KAPA2G Fast HotStart ReadyMix PCR Kit* (KAPABIOSYSTEMS®).

I campioni così amplificati sono stati infine analizzati tramite corsa elettroforetica su gel di agaroso al 2%.

3.7.3 DNA Clean up

Per purificare i campioni ho utilizzato il kit *PCR clean-up NucleoMag®96 PCR* della MACHEREY-NAGEL®

3.7.4 Sequenziamento

I campioni così processati sono stati inviati alla piattaforma di sequenziamento high-throughput di ultima generazione, Nextera di Illumina.

4. RISULTATI

4.1 Analisi molecolare dei geni candidati in cellule di sangue periferico

4.1.1 Genotipizzazione dei geni candidati

Nonostante il numero esiguo di campioni, tre geni su nove geni candidati analizzati hanno mostrato una differente distribuzione dell'allele di rischio tra celiaci e non celiaci statisticamente significativa (*Romanos J,2009*). Per i geni c-REL ed SH2B3 l'allele di rischio "A" risultato essere più frequente nei celiaci, rispetto ai controlli, mentre per TNFRSF14 l'allele "A" è meno frequente (Tab.5)

SNP	Chr.	Locus	Alleles*	M.A.F.†	FREQ CD	FREQ Not CD at 6 years	χ^2	P value
rs2816316	1	RGS1	C/A	0.160	AA 88.9% (8/9) AC 11.1% (1/9) CC 0% (0/9)	AA 68.2% (9/13) AC 30.8% (4/13) CC 0% (0/13)	1.869	0.393
rs3748816	1	TNFRSF14	C/A	0.477	AA 33.3% (3/9) AC 66.6% (6/9) CC 0% (0/9)	AA 76.9% (10/13) AC 23.1% (3/13) CC 0% (0/13)	5.743	0.057
rs842647	2	c-REL	A/G	0.482	AA 88.8% (8/9) AG 0% (0/9) GG 11.1% (1/9)	AA 61.5% (8/13) AG 38.5% (5/13) GG 0% (0/13)	9.097	0.056
rs1464510	3	LPP	C/A	0.453	AA 22.2% (2/9) AC 55.6% (5/9) CC 22.2% (2/9)	AA 23.1% (3/13) AC 61.5% (8/13) CC 15.4% (2/13)	0.231	0.994
rs13151961	4	KIAA1109	A/G	0.066	AA 55.6% (5/9) AG 44.4% (4/9) GG 0% (0/9)	AA 84.6% (11/13) AG 15.4% (2/13) GG 0% (0/13)	2.917	0.233
rs2327832	6	OLIG-TNFAIP3	A/G	0.102	AA 66.6% (6/9) AG 33.3% (3/9) GG 0% (0/9)	AA 16.9% (10/13) AG 15.4% (2/13) GG 7.7% (1/13)	2.439	0.565
rs1738074	6	TAGAP	A/G	0.441	AA 0% (0/9) AG 66.7% (6/9) GG 33.3% (3/9)	AA 15.4% (2/13) AG 76.9% (10/13) GG 7.7% (1/13)	5.079	0.279
rs3184504	12	SH2B3	G/A	0.218	AA 100% (9/9) AG 0% (0/9) GG 0% (0/9)	AA 38.5% (5/13) AG 61.5% (8/13) GG 0% (0/13)	14.930	0.001
rs2279627	19	TNFSF14	C/G	0.295	CC 22.2% (2/9) CG 44.4% (4/9) GG 33.3% (3/9)	CC 53.8% (7/13) CG 30.8% (4/13) GG 15.4% (2/13)	3.434	0.488

Tabella 5: Genotipizzazione delle varianti di rischio

4.1.2 Analisi di espressione genica dei geni candidati

T1: (Prima della comparsa dei sintomi o della produzione di anti-tTG) l'espressione di KIAA110, TAGAP ed SH2B3 risulta maggiore nei celiaci prima della diagnosi, mentre TNFSF14 ed RGS1 sono risultati essere down regolati (Fig.9). Le differenze sono statisticamente significative per TAGAP, SH2B3 e TNFSF14 (Mann Whitney test $p < 0.05$), ma non abbastanza significative per KIAA1109 ed RGS1.

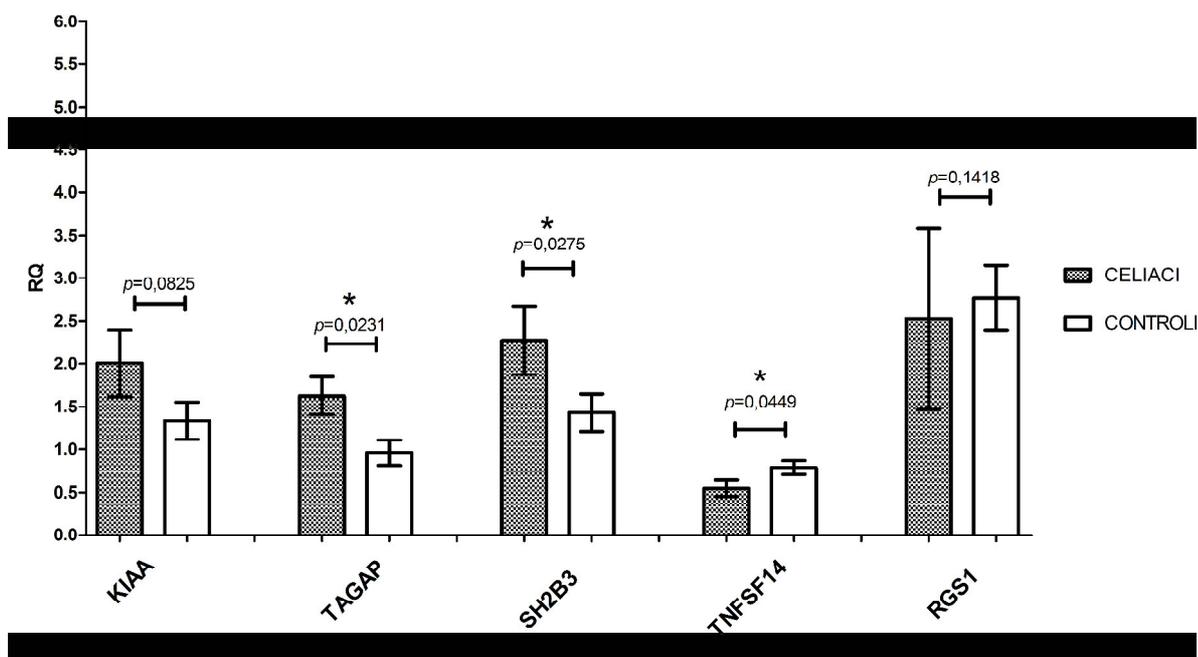


Figura 9. Espressione dei geni prima della diagnosi (T1)

T2 (Diagnosi): Al momento della diagnosi di malattia celiaca, solo due geni, SH2B3 e TNFRSF14 hanno mostrato livelli di espressione significativamente più elevati nei celiaci rispetto ai controlli (Fig 10).

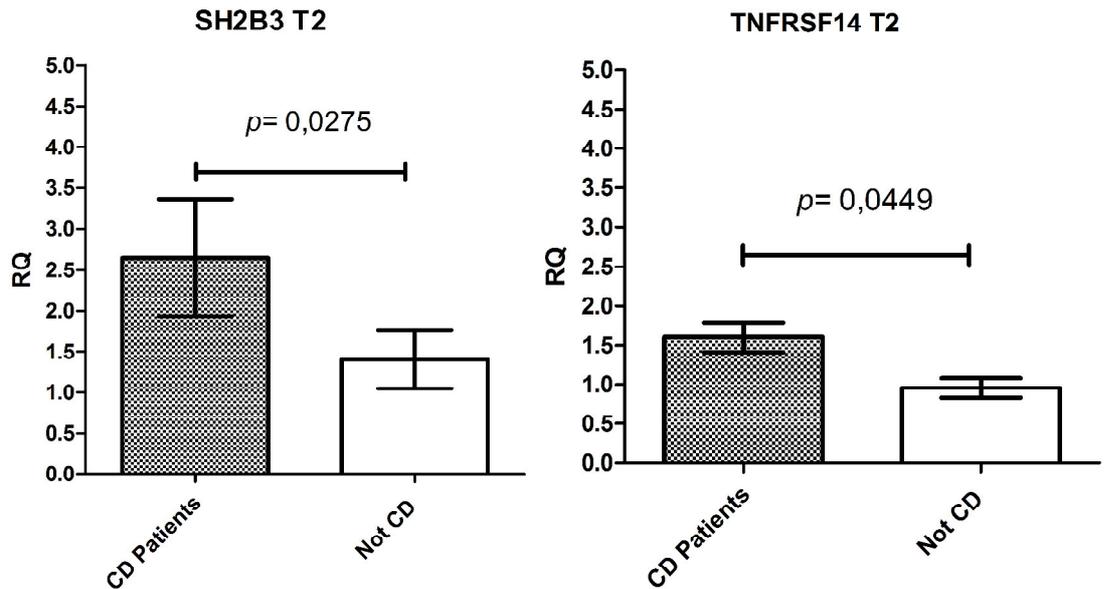


Figura10:Espressione dei geni alla diagnosi di CD (T2).

T3 (*Un anno dopo la diagnosi*): L'espressione dei geni esaminati non mostra differenze significative tra i due gruppi al time point 3 (dati non mostrati).

4.1.3 Analisi Discriminante Multivariata.

Genotipizzazione

Considerando le differenze osservate nella distribuzione dei genotipi nei casi rispetto ai controlli (Tab.5) abbiamo esplorato il profilo combinato dei genotipi dei 9 geni candidati, nei nostri pazienti. Mediante una analisi discriminante multivariata, 5geni sono risultati significativamente capaci di discriminare tra celiaci e controlli, con una classificazione corretta del 91%(Tab.6A-6B).

A

ANALISI DISCRIMINANTE DEL GENOTIPO

Step	Genotype	Wilks Lambda	Variance Ratio F	
			Statistic	p
1	SH2B3	0.604	13.091	0.002
2	RGS1	0.48	10.292	0.001
3	TAGAP	0.429	7.973	0.001
4	cREL	0.35	7.908	0.001
5	LPP	0.307	7.233	0.001

B

CLASSIFICAZIONE DEL GENOTIPOMEDIANTE EQUAZIONE DISCRIMINANTE

Predicted Group				
	Status	CD	Not CD	Total
Original Group	CD	8 (89%)	1 (11.1%)	9
	Not CD	1 (7.7%)	12 (92.3%)	13

Overall Correct Classification = 91%

Tabella 6A-B: Risultati dell'analisi discriminante effettuata utilizzando le genotipizzazioni dei geni candidati.

Espressione dei geni candidati

Poiché l'espressione dei geni all'interno dello stesso tipo di cellula è inter-correlata, l'espressione di ciascun gene candidato non può fornire un quadro preciso di eventuali differenze tra i gruppi. Alcuni geni possono essere sovra-espresi a causa dell'effetto di un altro gene nello stesso pathway. Dai risultati ottenuti dall'analisi discriminante l'espressione combinata di 4 geni (TAGAP, TNFSF14, SH2B3 ed RGS1) discrimina significativamente i celiaci dai controlli in fase presintomatica (Tempo 1). L'analisi ha predetto correttamente 9 celiaci su 9 e 12 controlli su 13 (Tab.7A-B). La figura 9 mostra il grafico dello score discriminante ottenuto dalla combinazione dell'espressione dei 4 geni in ogni singolo individuo. Tutti i bambini con diagnosi di celiachia mostrano un D-score negativo e tutti quelli senza diagnosi di celiachia mostrano un D-score positivo, nessun bambino viene classificato erroneamente come "CD" (Fig.9).

A

ANALISI DISCRIMINANTE DELL'ESPRESSIONE GENICA AL T1

Step	Genotype	Wilks Lambda	Variance Ratio F	
			<i>Statistic</i>	<i>p</i>
1	TAGAP	0.742	6.938	0.016
2	TNFSF14	0.542	8.019	0.003
3	SH2B3	0.446	7.453	0.002
4	RGS1	0.366	7.377	0.001

B

**CLASSIFICAZIONE MEDIANTE EQUAZIONE DISCRIMINANTE
DELL'ESPRESSIONE GENICA AL T1**

		Predicted Group			
		Status	CD	Not CD	Total
Original Group	CD		9 (100%)	0 (0%)	9
	Not CD		1 (7.7%)	12 (92.3%)	13

Overall Correct Classification = 95%

Tabella 7A-B: Risultati dell'analisi discriminante effettuata utilizzando l'espressione genica dei geni candidati

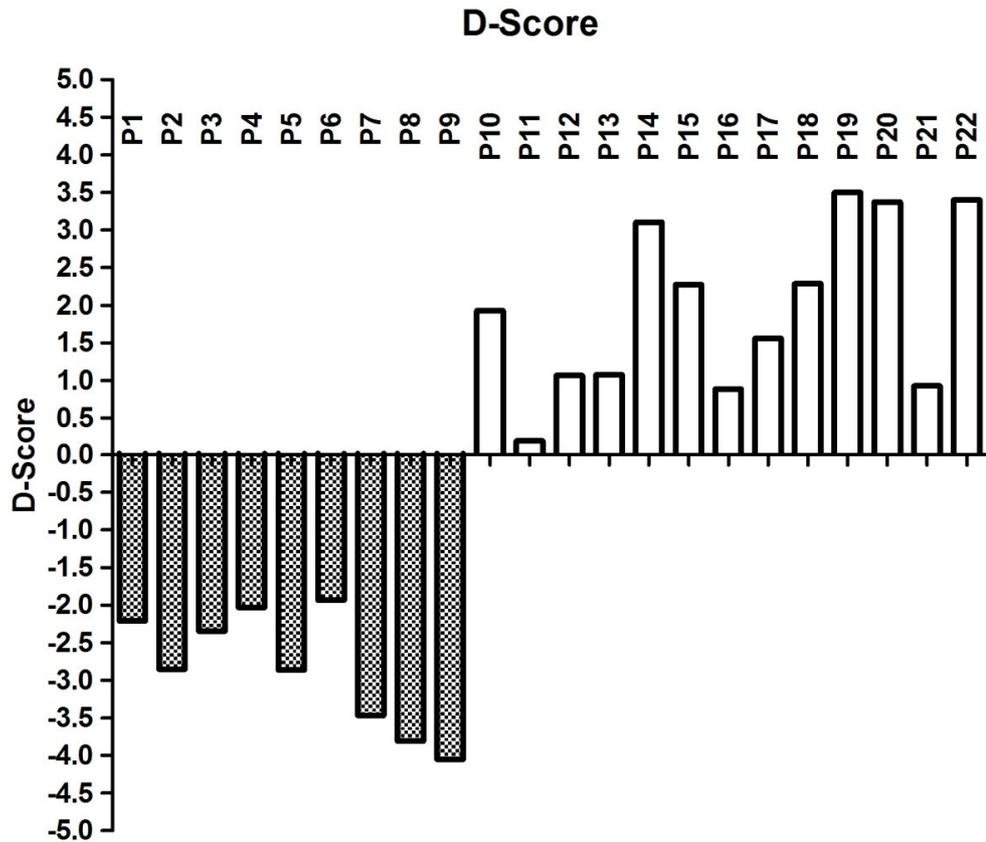


Figura 11: Punteggio discriminante (D-score) ottenuto dalla combinazione dei valori di espressione dei geni selezionati nei singoli individui al T1 (pre-diagnosi): tutti i bambini che hanno sviluppato la celiachia hanno un punteggio negativo, tutti i bambini non celiaci un punteggio positivo. I numeri rappresentano l'età (mesi) dei soggetti analizzati.

Genotipizzazione ed espressione genica

Poiché l'analisi basata su un numero di campioni relativamente piccolo può non essere sufficientemente affidabile ai fini diagnostici, per migliorare la robustezza del modello abbiamo combinato i dati di genotipizzazione con quelli dell'espressione genica al T1, e applicato l'analisi multivariata precedentemente descritta. La tabella 8 mostra i dati relativi a questa combinazione.?

I genotipi dei due geni SH2B3 ed RGS1, rappresentano i parametri con il maggior potere nel discriminare tra celiaci e controlli, a questi segue l'espressione genica del gene cREL che incrementa ulteriormente la capacità discriminante dell'equazione. Il genotipo di TNFRSF14 l'espressione del gene SH2B3, ed infine il genotipo di TNFSF14 (Tab. 8A) completano le 5 variabili selezionate dall'equazione per l'elevato potere discriminante, che ci ha permesso di classificare correttamente il 100% dei soggetti (Tab. 8B).

Per verificare la robustezza del modello, abbiamo adottato una strategia di auto-esclusione (la classificazione di ciascun individuo è stata effettuata utilizzando una equazione ricavata da tutti i casi meno se stesso) per ottenere una stima imparziale della capacità predittiva della funzione discriminante. Utilizzando questo approccio solo 2/9 celiaci non vengono predetti come celiaci, e solo 2/13 controlli vengono erroneamente classificati come CD con l'81.8% dei correttamente classificato (Tab.8C).

A ANALISI DISCRIMINANTE DELLA COMBINAZIONE GENOTIPO ED ESPRESSIONE GENICA DEI GENI AL T1

Step	Genotype	Wilks Lambda	Variance Ratio F	
			Statistic	p
1	SH2B3 genotype	0.604	13.091	0.002
2	RGS1 genotype	0.480	10.292	0.001
3	cREL expression at T1	0.340	11.626	0.000
4	TNFRSF14 genotype	0.260	12.066	0.000
5	SH2B3 expression at T1	0.217	11.523	0.000
6	TNFSF14 genotype	0.190	10.656	0.000

B CLASSIFICAZIONE MEDIANTE EQUAZIONE DISCRIMINANTE DELLA COMBINAZIONE GENOTIPO ED ESPRESSIONE GENICA AL T1 DEI GENI

		Predicted Group		
		CD	Not CD	Total
Original Group	CD	9 (100%)	0 (0%)	9
	Not CD	0 (0%)	13(100%)	13

C CLASSIFICAZIONE MEDIANTE EQUAZIONE DISCRIMINANTE DELLA COMBINAZIONE GENOTIPO ED ESPRESSIONE GENICA AL T1 DEI GENI

		Predicted Group by Jacknife Iterative Exclusion		
		CD	Not CD	Total
Original Group	CD	7 (77.8%)	2 (22.2%)	9
	Not CD	2 (15.4%)	11(84.6%)	13

Tabella8A-C. Analisi discriminante combinando genotipo ed espressione dei geni candidati prima della diagnosi.

4.2 Analisi molecolare dei geni candidati in cellule intestinali

4.2.1 Valutazione della purezza del campione.

L'efficacia di separazione tra i due compartimenti, epitelio e *lamina propria*, è stata valutata mediante Real-Time PCR, e in particolare andando a dosare lo specifico marker utilizzato per la separazione mediante Microbeads (vedi materiali e metodi) EpCAM. Come mostrato dal grafico, è stato raggiunto circa il 98% di purezza globale. In particolare, l'analisi dei livelli di espressione di EpCAM, indicano una percentuale di selezione positiva del 97.8% delle cellule epiteliali (CD326+) nei campioni biotipici dei celiaci, e un 97.5% di positività nei campioni biotipici dei controlli (Fig.12).

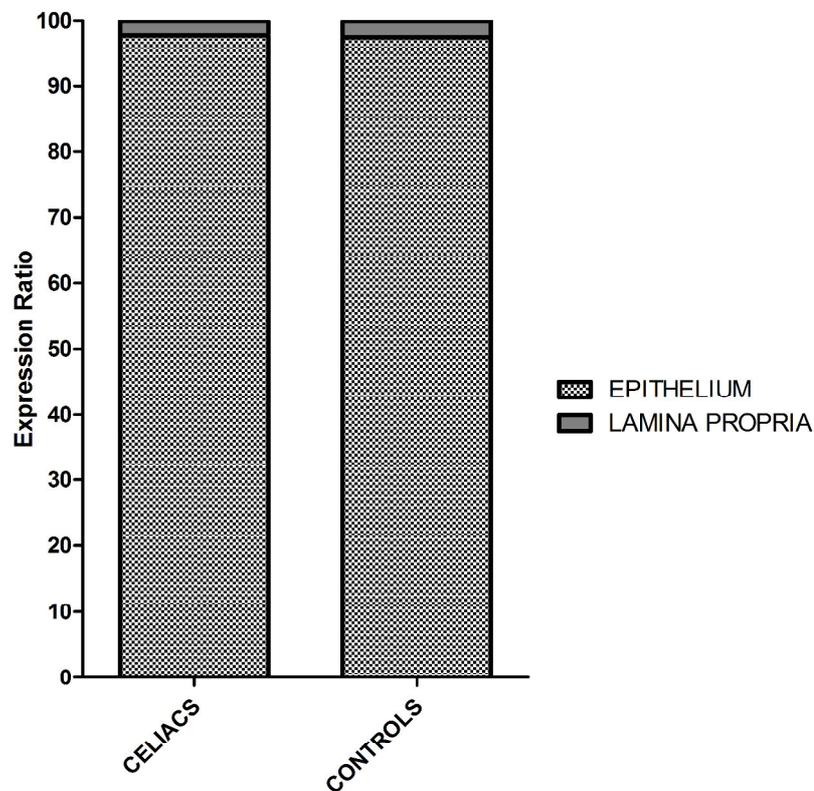


Figura12. Valutazione dell'efficacia di separazione tra i due compartimenti.

4.2.2 Espressione genica in Epitelio e Lamina Propria

L'espressione dei 16 geni associati alla celiachia (elencati in Tab. 3) è stata analizzata nei due compartimenti cellulari (epitelio e lamina propria) al fine di evidenziare eventuali differenze tra celiaci e controlli.

La figura 13A mostra come i geni *IL12A*, *IL21*, *c-REL*, *RGS1*, *TAGAP*, *SH2B3*, tutti direttamente o indirettamente coinvolti nei processi infiammatori, sono significativamente up-regolati nelle cellule epiteliali dei soggetti celiaci rispetto ai controlli.

Allo stesso modo *IL12A*, *IL21*, and *RGS1* sono up-regolati anche nella *lamina propria* dei celiaci rispetto ai controlli. Inoltre *NFKB1*, *TNFSF14* e *PTPRK* risultano differenzialmente espressi nelle cellule della *lamina propria* dei celiaci rispetto ai controlli (Fig. 13B).

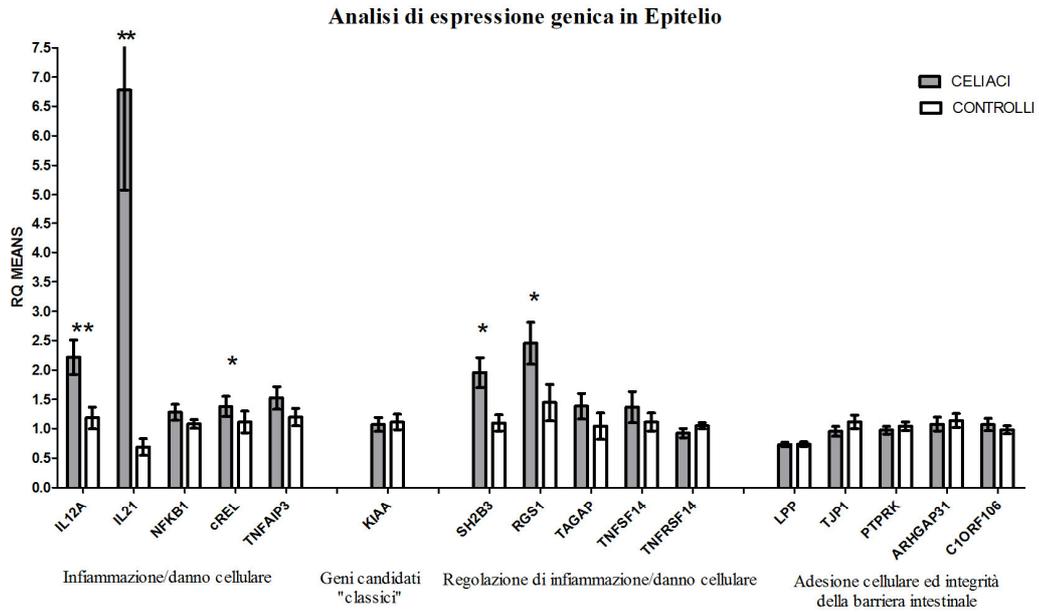
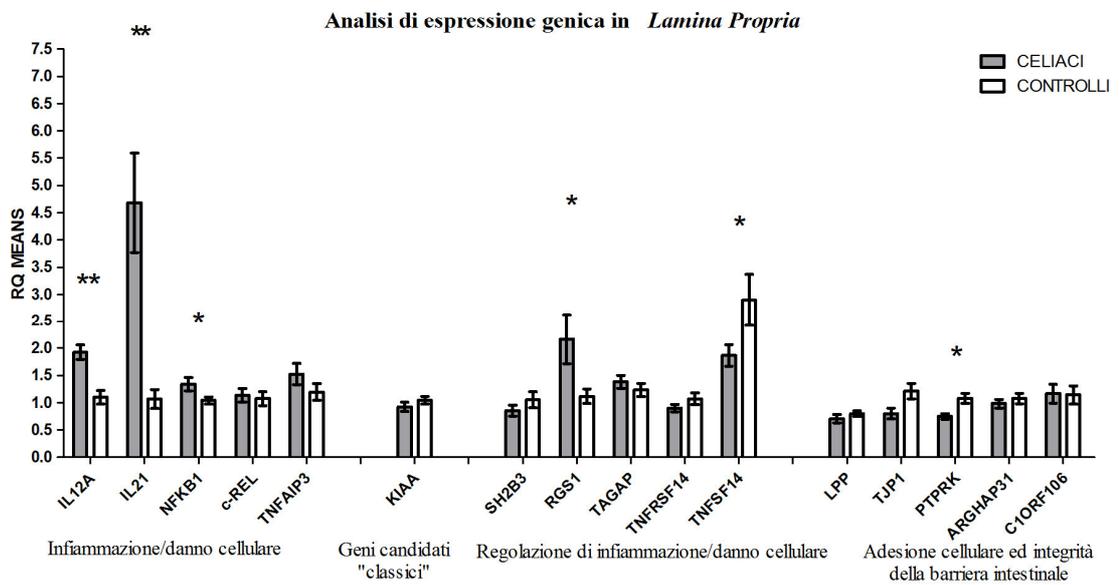
A**B**

Figura 13. Espressione genica in epitelio (A) e in *lamina propria* (B). * $p < 0.05$; ** $P < 0.01$.

4.2.3 Analisi di correlazione tra espressione genica dei geni candidati

L'espressione dei geni in ogni compartimento cellulare è il risultato di complesse correlazioni che intercorrono tra loro all'interno di specifici ed pathways molecolari e di vie di segnalazione. E' quindi interessante verificare le correlazioni tra i geni all'interno dello specifico compartimento e le eventuali variazioni nei celiaci e nei controlli.

L'immagine che ne viene fuori mostra la complessità di queste correlazioni, che non solo sono bivariate ma anche multidimensionali. La figura 14 mostra il pattern di correlazione tra l'espressione dei geni in epitelio dei celiaci rispetto a quello dei controlli (Fig14A-B) e nella lamina propria (Fig14C-D).

E' facile osservare (dall'intensità del colore) che alcuni profili metabolici sono altrettanto correlati nel celiaco come nei controlli in entrambi i compartimenti: riflettendo i pathway cellulari "obbligatori". D'altra parte alcune correlazioni, abbastanza forti tra i controlli, vengono perse sia in epitelio che nella lamina propria dei celiaci. Nella lamina Propria si osservano forti correlazioni nei controlli, precisamente 23 rispetto a solo 14 nei celiaci. Nello specifico TNFAIP3 in lamina propria dei controlli è fortemente correlato agli altri geni coinvolti nella regolazione dei processi infiammatori TAGAP, RGS1, cREL, e al gene C1orf106, mentre a livello della lamina propria dei celiaci perde completamente tutte le correlazioni con questi geni e ne acquista solo una lieve con LPP.

Al contrario, in epitelio i geni candidati sembrano essere maggiormente correlati nei pazienti celiaci. SH2B3 nell'epitelio dei controlli non è correlato a nessun gene analizzato, mentre nei celiaci è correlato principalmente con TAGAP, cREL e IL-12A.

A

B

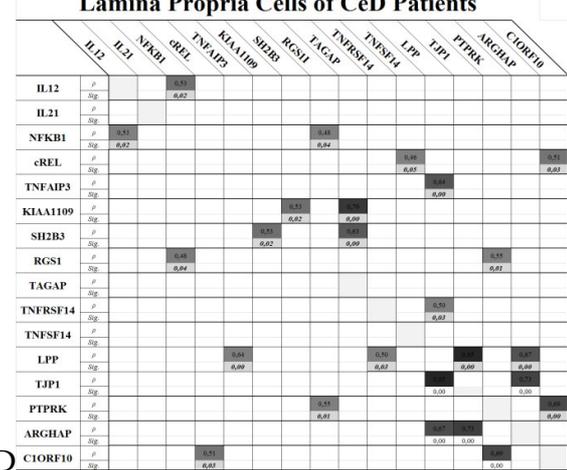
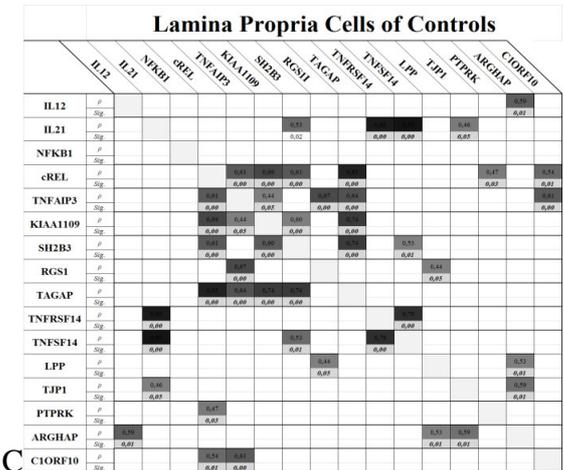
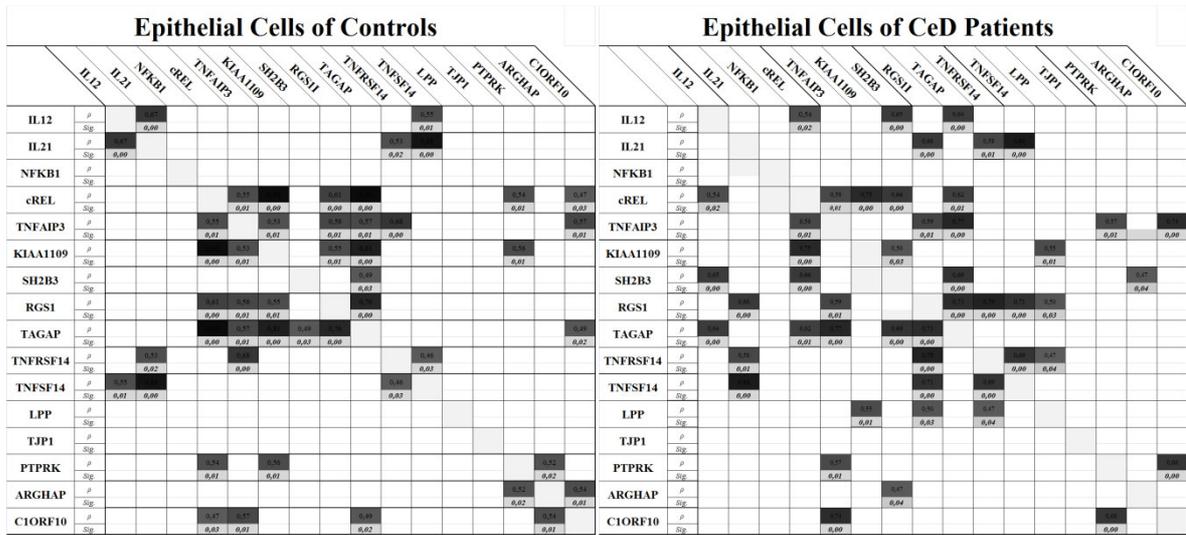


Figura 14. Pattern di correlazione tra l'espressione dei geni in epitelio(A-B) dei celiaci rispetto a quello dei controlli e in lamina propria.

4.2.4 Analisi discriminate multivariata

Un approccio multivariato ci aiuta a capire meglio il coinvolgimento differenziale dei geni candidati sia dei celiaci che dei controlli nei due compartimenti.

Attraverso l'analisi discriminante, abbiamo selezionato il livello di espressione genica che più efficacemente ci permette di discriminare tra celiaci e controlli nei due compartimenti.

I risultati dell'analisi nelle cellule epiteliali sono mostrati nella tabella9, in cui i geni inclusi nel modello aiutano a discriminare correttamente tutti i celiaci (100%), mentre 3 controlli (15,8%) vengono erroneamente classificati come celiaci. Allo stesso modo vediamo che in lamina propria(Tab.10)il 90% degli individui viene correttamente classificato. Anche in

questo studio è stato calcolato il punteggio discriminante (D-Score) dei singoli campioni in seguito all'analisi discriminante, che come si vede nelle figure 15 e 16 distingue visibilmente i soggetti celiaci dai controlli.

Le interleuchine IL21, IL12 ed il gene NFKB risultano essere efficienti discriminatori in entrambi i compartimenti: in epitelio un ulteriore contributo è dato dal gene TNFSF14 con il suo recettore e dal gene SH2B3. In *lamina propria* i geni PTPRK e KIAA1109 vengono aggiunti per migliorare la capacità discriminante (tabella 10).

**ANALISI DISCRIMINANTE DELL'ESPRESSIONE GENICA
NELLE CELLULE EPITELIALI**

Step	Gene	Variance Ratio F		
			Statistic	p
1	IL21 EPI	0,749	11,042	0.000
2	TNFSF14 EPI	0,643	8,878	0.000
3	NFKB1 EPI	0,558	8,183	0.000
4	TNFRSF14 EPI	0,508	7,268	0.000
5	SH2B3 EPI	0,456	6,926	0.000

**CLASSIFICAZIONE MEDIANTE EQUAZIONE DISCRIMINANTE
DELL'ESPRESSIONE GENICA NELLE CELLULE EPITELIALI**

Status	Predicted Group			Total
	CD	Not CD		
Original Group	CD	13 (72,2%)	5 (28,7%)	18
	Not CD	2 (11,8%)	16 (88,9%)	19

Overall Correct Classification = 80,6%

Tabella 9. Analisi discriminante ottenuta dai dati di espressione dei geni candidati in cellule epiteliali

ANALISI DISCRIMINANTE DELL'ESPRESSIONE GENICA IN LAMINA PROPRIA

Step	Gene		Variance Ratio F	
			Statistic	p
1	IL12 LP	0,595	23,100	0.000
2	NFKB1 LP	0,442	20,823	0.000
3	TNFSF14 LP	0,400	16,018	0.000
4	IL21 LP	0,314	16,949	0.000
5	PTPRK LP	0,288	14,814	0.000
6	KIAA LP	0,270	13,040	0.000

**CLASSIFICAZIONE MEDIANTE EQUAZIONE DISCRIMINANTE
DELL'ESPRESSIONE GENICA IN LAMINA PROPRIA**

Status	Predicted Group		Total
	CD	Not CD	
	16	2	
CD	(88,9%)	(11,1%)	18
Original	17		
Group	Not CD	2(10,5%)	(89,5%) 19

Overall Correct Classification = 89,2%

Tabella 10. Analisi discriminante ottenuta dai dati di espressione dei geni candidati in cellule della lamina propria

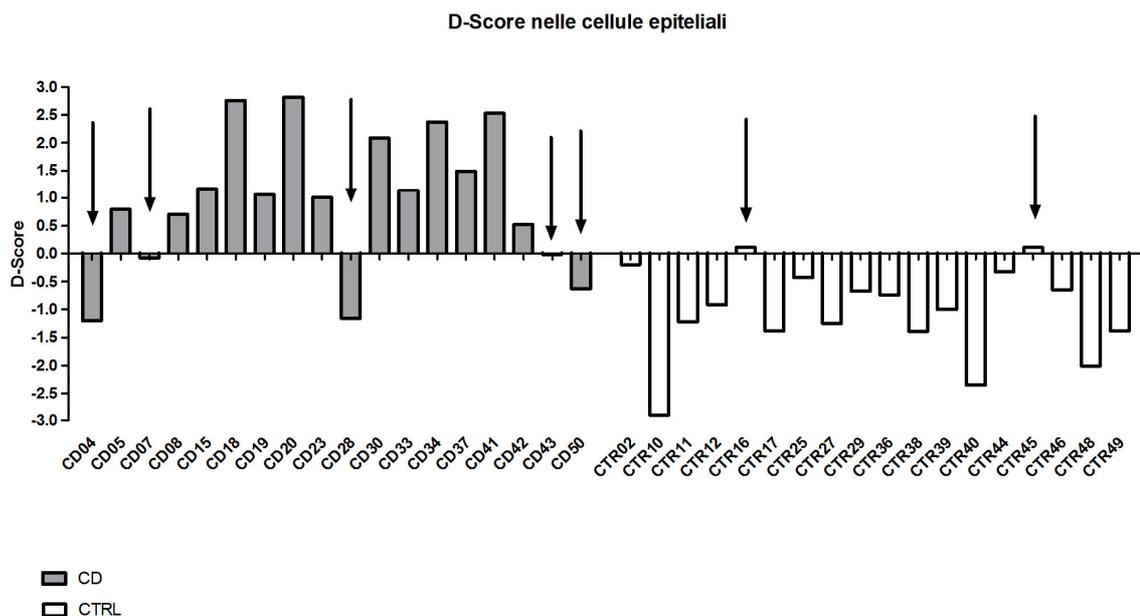


Figura15. Punteggio discriminante (D-score) in cellule epiteliali

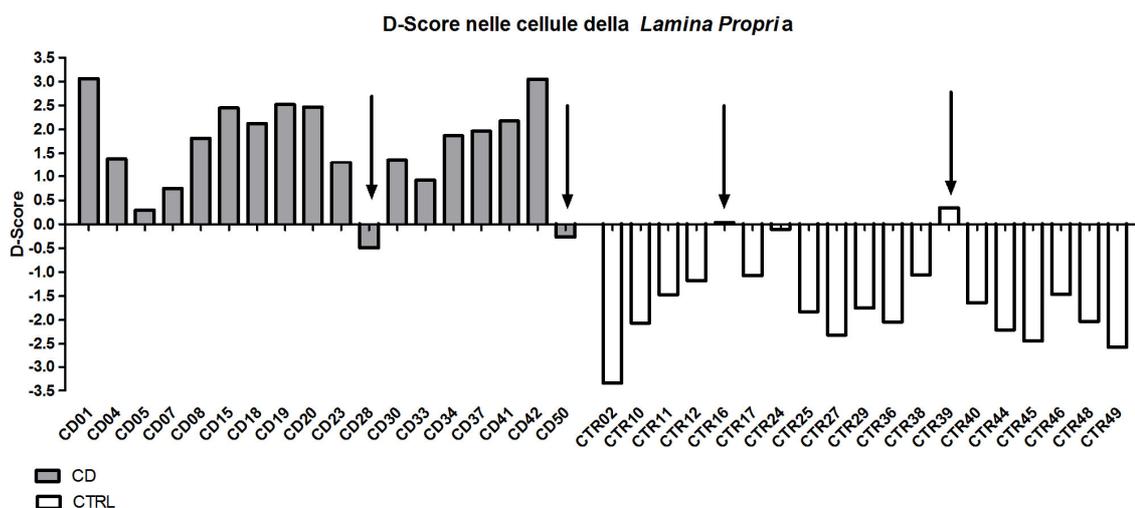


Figura 16: Punteggio discriminante (D-score) in *lamina propria*

4.2.5 Analisi di Metilazione

L'analisi di metilazione ha messo in evidenza alcune differenze significative tra celiaci e controlli nei due compartimenti. Nei grafici seguenti (Fig. 17-18) sono mostrati i livelli medi di metilazione per ciascun gene rispettivamente in epitelio e in *lamina Propria*. Tuttavia al fine di avere una più corretta informazione riguardo le differenze di metilazione tra celiaci e controlli è opportuno andare a visualizzare in quale particolari punti del gene ci sono differenze di metilazione, che a livello di specifiche regioni geniche quali regioni di

regolazione hanno sicuramente un peso maggiore rispetto all' andamento complessivo dei livelli di metilazione.

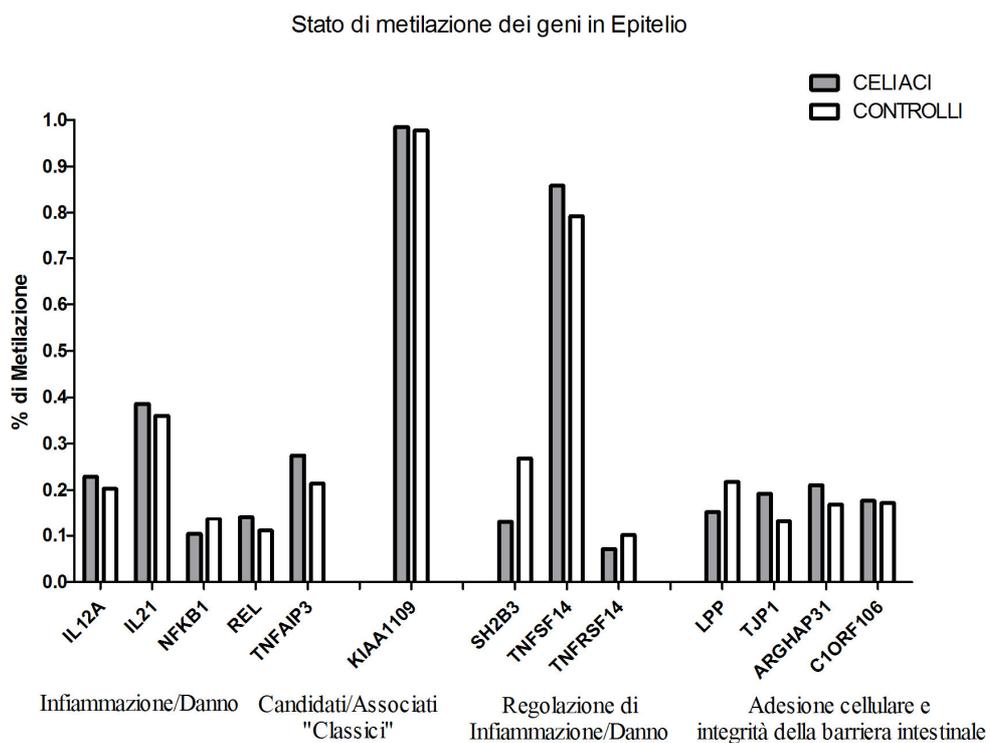


Figura17. Analisi di metilazione dei geni in epitelio

Stato di metilazione dei geni in *Lamina Propria*

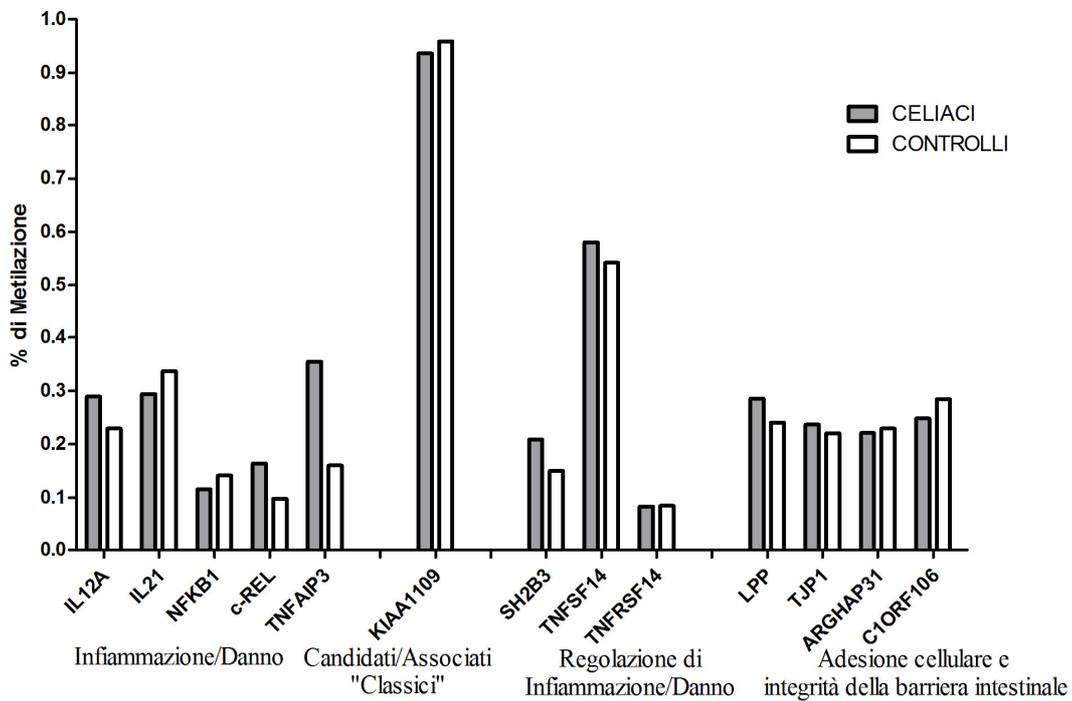


Figura 18. Analisi di metilazione dei geni in *lamina propria*

Una analisi approfondita dei dati ha mostrato che IL21 è più espressa nei celiaci rispetto ai controlli e a questo aumento del livello di espressione coincide una diminuzione dello stato di metilazione del gene riferito in particolare ad alcuni specifici punti del isola CpG analizzata (Fig. 19 A-B). In particolare il grafico seguente (Fig. 19 C) mostra l'andamento dei livelli di metilazione per ogni singolo nucleotide dell'isola CpG (position).

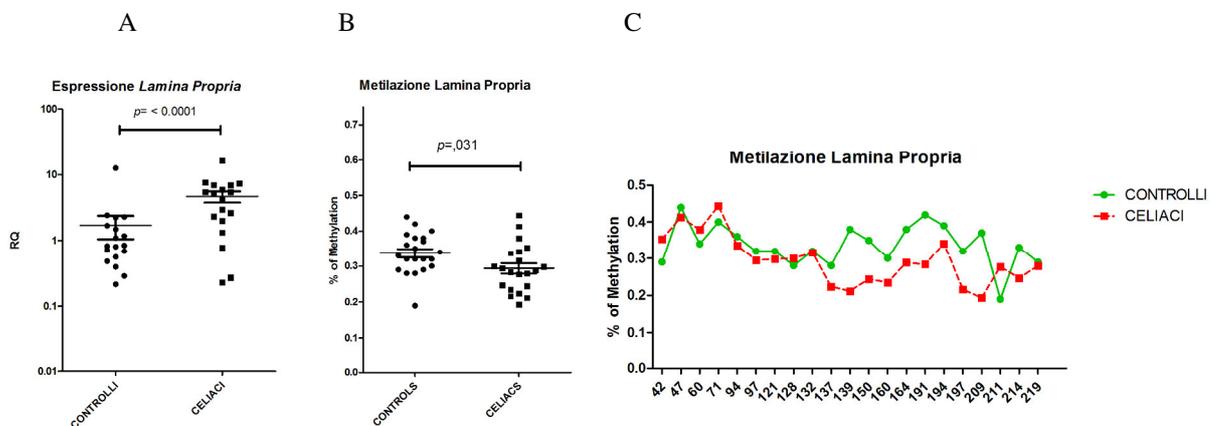


Figura 19. Correlazione fra espressione genica ed livelli di metilazione del gene IL21

Coerentemente a quanto osservato negli studi di espressione, anche l'elevata metilazione del gene SH2B3 nelle cellule epiteliali dei controlli correla con una minore espressione del gene (Fig.20 ABC).

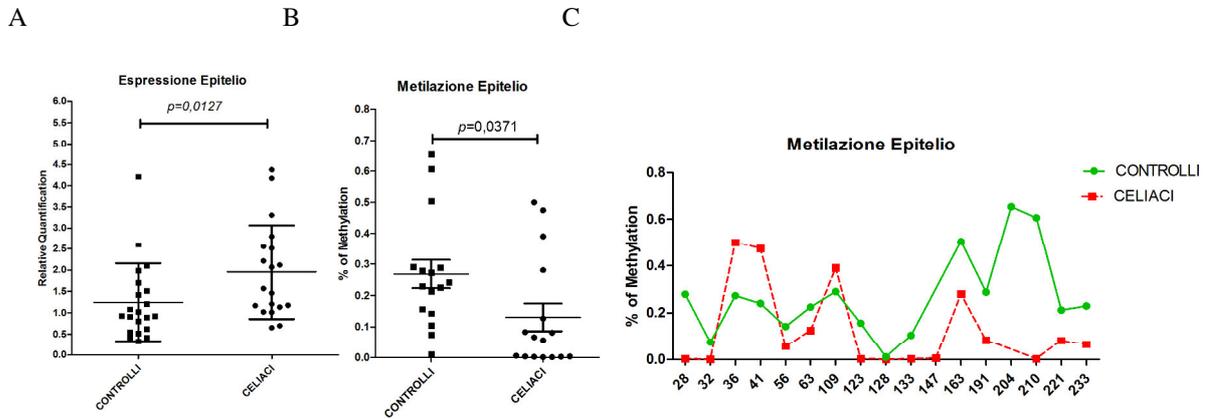


Figura 20. Correlazione fra espressione genica ed livelli di metilazione del gene SH2B3

Andando a visualizzare la regione del gene SH2B3, ma più dettagliatamente della isola CpG 132 selezionata in questo studio in cui vi è una significativa differenza nello status di metilazione tra celiaci e controlli, possiamo notare che proprio in corrispondenza dei nucleotidi in cui abbiamo rilevato la maggiore differenza vi sono degli elementi di regolazione del DNA (Fig 21).

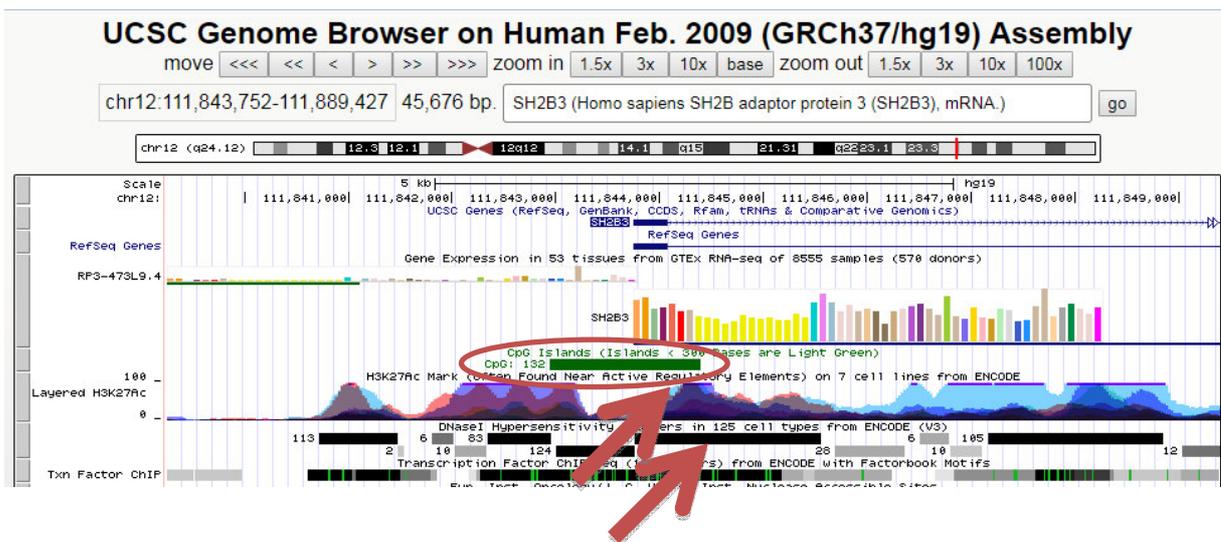


Figura21. Output dell'analisi bioinformatica relativa agli elementi di regolazione del gene SH2B3

Nei geni TNFAIP3 e c-REL abbiamo riscontrato un profilo di metilazione alterato a livello della *Lamina Propria* dei celiaci tuttavia nessuna alterazione significativa nei livelli di espressione genica.(Fig.21 A-C;Fig.22 A-C).

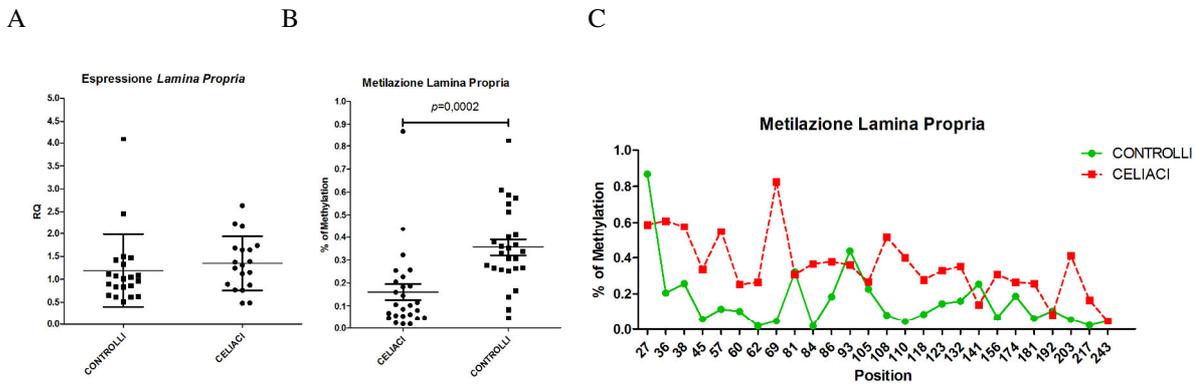


Figura 21.Correlazione fra espressione genica ed livelli di metilazione del gene TNFAIP3

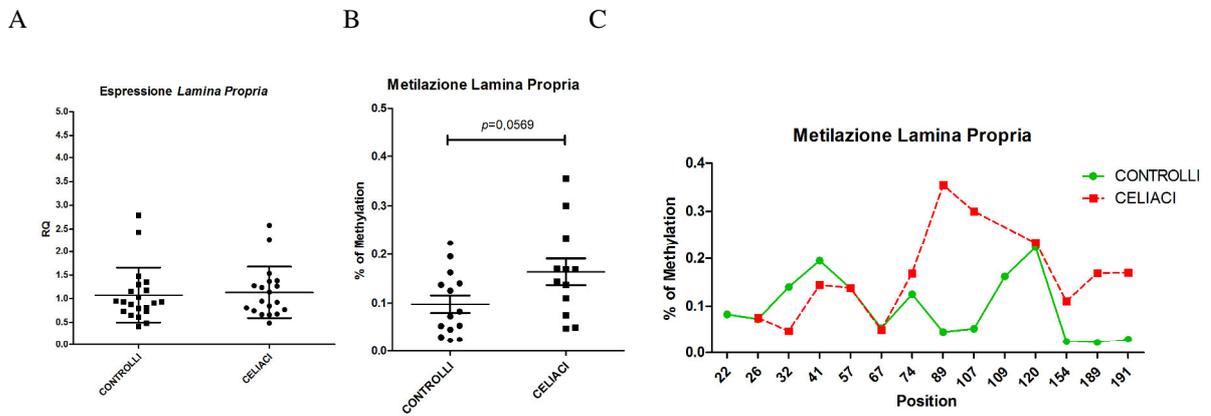


Figura 22.Correlazione fra espressione genica ed livelli di metilazione del gene cREL

5. DISCUSSIONE E CONCLUSIONI

La celiachia ha ormai assunto la connotazione di vera e propria malattia sociale in diversi paesi occidentali, tra cui l'Italia. Sebbene diversi gruppi di ricerca in ambito internazionale si occupino dello studio dei meccanismi molecolari e cellulari che sottendono allo sviluppo di questa patologia, ad oggi restano ancora da chiarire tutti i fattori genetici che ne determinano l'insorgenza. Il coinvolgimento dei geni HLA è ormai stato consolidato come essere una condizione assolutamente necessaria, ma non sufficiente allo sviluppo della malattia e tra le possibili ipotesi di ricerca percorribili, quella che ha inizialmente prodotto il maggior numero di consensi e anche di lavori scientifici è stata la ricerca di altri loci genici. Nell'ultimo decennio sono state numerosissime le pubblicazioni che si sono susseguite sull'argomento (*van Heel DA, 2007; Hunt KA, 2008; Castellanos-Rubio A, 2008; Romanos J, 2009a; Romanos J, 2009b, Trynka G, 2009; Dubois PC, 2010, Trynka G, 2011, Kumar V 2015*), ma, fatta eccezione per qualche recente aggiornamento, ci si è ormai convinti che gli oltre 60 loci genici addizionali individuati, contribuiscono a spiegare non oltre il 6% dell'ereditarietà della celiachia, lasciandone ancora da chiarire oltre il 50%.

Inoltre, un altro fenomeno significativo che non può essere giustificato dalla genetica classica, è l'aumento esponenziale che la malattia ha avuto negli ultimi 20 anni. Questo può essere stato determinato in parte dalla variazione delle abitudini alimentari di alcune nazioni e all'introduzione di nuovi grani sul mercato, ma sicuramente anche dall'azione di meccanismi modificatori dell'espressione genica. Pertanto è ormai chiaro che, per cercare di chiarire al meglio i fattori che concorrono allo sviluppo della malattia celiaca, è necessario indagare altri meccanismi di regolazione a carico dei cosiddetti 'geni candidati'.

Già in passato il nostro gruppo ha studiato l'espressione dei geni candidati, in diversi coorti e in molteplici tessuti quali campioni biotipici di mucosa duodenale, monociti estratti da sangue periferico (PBM) di celiaci attivi, in remissione e controlli, e da cloni T isolati da biopsie duodenali e reattivi alla gliadina. Questo lavoro ha fornito una panoramica interessante delle possibili relazioni tra le espressioni dei geni candidati sia nella mucosa che nei PBM. L'analisi dell'espressione di ogni singolo gene, infatti, non risulta completamente informativa e non ne rivela la funzione specifica. Diversi, degli 11 geni analizzati, (KIAA1109, IL-2, IL-21, LPP, RGS1, cREL, SH2B3, TAGAP, TNFAIP3, TNFSF14, e TNFRSF14) risultavano essere differenzialmente espressi nei pazienti celiaci rispetto ai controlli, ma l'analisi di ogni singolo gene non è riuscita a fornire un quadro completo. L'analisi discriminante multivariata invece ha suggerito che l'espressione di un piccolo set di geni (5 geni nella mucosa duodenale e 4 geni nei monociti) poteva essere utilizzata per distinguere i pazienti celiaci da controlli

sani e da controlli affetti da malattia di Crohn, senza considerare i dati clinici, gli anticorpi HLA o anti-tTG. L'equazione discriminante ottenuta ci permetteva una classificazione diagnostica corretta di tutti i casi di celiaci e del 91 % dei campioni di controllo. Già questo lavoro, quindi, ci indirizzava verso la ricerca di complessi pathways funzionali per far luce sulla complessa risposta indotta dal glutine negli individui geneticamente predisposti (Galatola, 2013).

Per quanto riguarda la prima parte dello studio, che ha poi portato alla pubblicazione del lavoro (Auricchio R, 2017), la coorte del progetto PREVENT-CD, descritta precedentemente, ci ha permesso di fare una analisi presintomatica, ossia di prevedere quali bambini, appartenenti a famiglie a rischio, seguiti dalla nascita fino a 6 anni finivano per sviluppare la MC, molto prima della produzione di auto-anticorpi (anti-tTG) o della comparsa dei sintomi clinici. Abbiamo confermato l'importanza dell'aplotipo HLA e abbiamo dimostrato che il genotipo di 5 geni candidati (SH2B3, RGS1, TAGAP, cREL e LPP) aumenta significativamente (oltre il 90%) il rischio di sviluppare la MC. Nonostante questi risultati però, è chiaro che la maggior parte dei bambini a rischio, condividono i loro genotipi con i familiari affetti, sia per quanto riguarda la positività HLA al DQ2 e/o 8 che per il genotipo dei geni candidati; da qui l'ipotesi che esplorare l'espressione dei geni piuttosto che il solo genotipo può fornire un'informazione più chiara dell'effettivo rischio di questi bambini di sviluppare la malattia.

Da un semplice campione di sangue siamo stati in grado di prevedere quali bambini avrebbero sviluppato la MC entro i 18-24 mesi. Infatti, 5 geni (KIAA, TAGAP, SH2B3, TNFSF14 e RGS1) sono risultati espressi differenzialmente molto prima della comparsa di qualsiasi segno clinico della malattia, e quindi hanno permesso di identificare con notevole precisione i bambini che avrebbero sviluppato eventualmente la malattia. Per aumentare la robustezza della predizione abbiamo combinato il genotipo del bambino, con i livelli di espressione presintomatica dei geni candidati ottenendo una combinazione di genotipi e profili di espressione genica che aumentavano significativamente (fino al 100%) la capacità di predire l'esito del bambino. Poiché l'equazione discriminante, con la quale abbiamo classificato i casi almeno nove mesi prima della diagnosi, è stata sviluppata dai dati ottenuti nella stessa coorte, abbiamo adottato una strategia di autoesclusione ("jack-knifing") in cui abbiamo escluso iterativamente dall'equazione gli individui da classificare. La reversibilità dell'espressione di alcuni geni nei pazienti affetti da MC trattati con una GFD è in linea con la reversibilità dell'espressione genica nelle tight junction in bambini celiaci dopo 2 anni di dieta senza glutine (Fernandez-Jimenes 2014; Plaza-Izurrieta 2015). La reversibilità dell'espressione

genica potrebbe indirettamente confermare una diagnosi di MC, e di conseguenza la valutazione dell'espressione già a livello presintomatico potrebbe essere considerata un marker diagnostico. Tuttavia, gli studi sul potenziale ruolo di queste e di altre espressioni geniche nella diagnosi di MC, indicano che, a differenza dei profili genetici costitutivi, alcuni meccanismi di regolazione epigenetica di questi geni candidati potrebbero svolgere un ruolo fondamentale nell'eziologia della MC (*Fernandez-Jimenes 2014; Zilbauer M, 2016*).

Questo studio ha come limite la scarsa numerosità dei campioni analizzati, che tuttavia è controbilanciato dal disegno di studio longitudinale di una coorte particolare di bambini ad alto rischio di sviluppare la MC nei primi anni di vita.

È importante sottolineare che questo modello viene applicato in questo contesto per la prima volta, dal momento che il follow-up dei bambini a rischio di MC non è ancora una pratica comune. Dobbiamo anche riconoscere che per il reclutamento di neonati con rischio familiare, sottoposti ad una sorveglianza che va dalla nascita ai 6 anni, con una regolare raccolta di campioni di sangue, è stato possibile solo grazie allo studio longitudinale PREVENT-CD. E che questa coorte unica nel suo genere, è stata utilizzata dal nostro gruppo per altri studi di natura epidemiologica (*Auricchio R, 2017*) dimostrando che la frequenza delle infezioni respiratorie nei primi due anni di vita nei bambini con rischio genetico contribuisce a discriminare i celiaci da i controlli.

Un altro fattore che potrebbe essere considerato come un limite di questo studio è il piccolo numero di geni analizzati; questa tuttavia è stata una scelta strategica poiché abbiamo preferito esplorare i geni replicati in diversi GWAS (*van-Heel DA 2007; Trynka G, 2009*), con una forte implicazione biologica nella risposta immunitaria indotta dal glutine (*Kumar V, 2015*), e già studiati e risultati interessanti in diversi modelli nei lavori precedenti del nostro gruppo (*Galatola M, 2013; Izzo V, 2011*).

Successivamente, nella seconda parte dello studio, siamo andati ad analizzare le differenze di espressione in situ, ma questa volta selezionando le due popolazioni cellulari che compongono la mucosa intestinale, ed utilizzando una coorte di campioni non appartenenti a famiglie con un particolare profilo genetico.

Al giorno d'oggi la maggior parte dei lavori è diretta verso lo studio di una genomica "single-cells", che utilizza marcatori sempre più specifici, al fine di consentire una rappresentazione accurata dell'intera composizione di un tessuto, come un organo specifico o un tumore, o di un processo multicellulare come la risposta del sistema immunitario.

Infatti nel lavoro è stata sperimentata per la prima volta un'analisi differenziale dell'espressione dei geni candidati nella mucosa intestinale dei pazienti celiaci,

specificatamente a livello del compartimento epiteliale e del compartimento della lamina propria. Ad oggi, non è stato pubblicato alcuno studio sulle singole popolazioni cellulari e/o nei compartimenti costituenti la mucosa intestinale dei celiaci, e l'eterogeneità cellulare dei campioni di tessuto è sempre stato un fattore problematico. La maggior parte dei lavori presente in letteratura riguardanti lo studio del pattern di espressione genica nella mucosa intestinale dei celiaci, è stata eseguita isolando le cellule da biopsie intere, mediante omogeneizzazione (Kontaku M. 1995;Beckett CG,1996). Inoltre, alcuni lavori sono stati eseguiti separando il compartimento epiteliale da quello della lamina propria mediante l'utilizzo di agenti chelanti del calcio (EDTA) (Chiba M, 1981). Si evidenzia che l'utilizzo degli agenti chelanti non portava mai ad una separazione completa tra i due compartimenti, generando così delle contaminazioni fra le cellule appartenenti all'epitelio con quelle appartenenti alla lamina propria e viceversa. Tale contaminazione portava ovviamente ad una inaccuratezza dei risultati. La metodica di separazione con le Microbeads utilizzata in questo lavoro è risultata, invece, particolarmente efficiente in quanto come mostrato in figura12 la verifica della purezza del campione da noi effettuata mediante RT-PCR ha fornito una purezza di circa il 98%.

I nostri risultati hanno messo in evidenza che, i geni notoriamente coinvolti nel processo infiammatorio sia direttamente , in quanto citochine pro infiammatorie come IL12 e IL21, sia indirettamente, in quanto geni implicati nel processo di regolazione, tra cui SH2B3 e TAGAP3, risultano essere significativamente up-regolati nei pazienti celiaci ; viceversa, i geni coinvolti nel mantenimento dell'integrità della barriera intestinale o comunque nel processo di adesione intercellulare come PTPRK , sono risultati essere per la maggior parte , coerentemente a quello che ci aspettavamo, down-regolati nei celiaci.

Siamo andati poi ad esplorare le correlazioni tra i geni all'interno dello specifico compartimento e le eventuali variazioni nei celiaci e nei controlli per vedere se le alterazioni osservate influenzano anche l'interazione tra i geni. Coerentemente a quanto riportato in letteratura da Bilbao e colleghi a livello della mucosa intestinale in toto, anche nel nostro studio abbiamo messo in evidenza nei celiaci viene a mancare la coregolazione tra i geni candidati, in particolare, e i geni infiammatori risultano deregolati (Fernandez-Jimenez N, 2014). È interessante notare che nonostante la alterazione di alcuni pathway molecolari, si creano nuove interazioni nella rete di coespressione, in particolare nei soli soggetti celiaci SH2B3 viene sottoposto a coregolazione con gli altri geni infiammatori, evento in linea con il fatto che è un gene attivato dall'infiammazione.

Abbiamo confermato gli studi del gruppo di JR Bilbao che nel lavoro sui geni delle tight junction (*Jauregi-Miguel A,2014*) hanno dimostrato che i livelli di espressione risultano alterati a livello della mucosa intestinale nei pazienti celiaci floridi, riflettendo un pathway deregolato come conseguenza dell'ingestione di glutine, mentre dopo una dieta priva di glutine si osserva una reversibilità dell'espressione. Modelli di coregolazione di diversi geni supportano l'idea di un meccanismo di regolazione che sembra essere interrotto e alterato nel celiaco in fase florida. Il lavoro, dunque, conferma il coinvolgimento dei geni TJ relativi alla permeabilità intestinale, alla polarità apicobasale e alle vie di proliferazione cellulare, nella distruzione a livello epiteliale osservata nei celiaci.

È necessario tenere presente che, i nostri risultati, come quelli dello studio del gruppo spagnolo, sono limitati ai livelli di espressione di mRNA, e che potrebbero non riflettere ciò che in realtà accade a livello delle proteine, per cui i risultati devono essere interpretati con cautela. Ulteriori studi di proteina ed esperimenti di coltura cellulare, sono necessari per confermare questi risultati perché che potrebbero non essere specifici solo per il la MC.

Un altro lavoro molto interessante dello stesso gruppo (*Fenandez –Jimenez N,2014*) è stato lo studio del pathway NFkB che come si detto è costantemente regolato nella MC. L'obiettivo di questo lavoro è stato quello di esaminare i livelli di espressione di alcune delle componenti biologicamente più rilevanti del pathway NFkB in biopsie intestinali di pazienti attivi, trattati, e biopsie coltivate con gliadina. Inoltre, per primi gli spagnoli, sono andati a valutare lo status di metilazione nella mucosa intestinali dei celiaci. In particolare è stata analizzata la metilazione di geni coinvolti nella signaling di NFkB. I diversi gruppi sono stati confrontati e sono state costruite matrici di correlazione per verificare la coespressione e la co-metilazione. Anche in questo lavoro la coespressione tra i geni è regolata nei controlli, mentre l'infiammazione cronica presente nel celiaco non trattato provoca la rottura del normale equilibrio. Al contrario, la co-metilazione si verifica più spesso nel celiaco attivo.

Incoraggiati da questo lavoro, per capire se alla base delle alterazioni mostrate nel nostro studio, ci fossero meccanismi epigenetici di regolazione dell'espressione siamo andati a studiare anche la metilazione del Dna negli stessi geni. Complessivamente, gli studi hanno dimostrato che la metilazione delle isole CpG dei geni non rappresenta, se non in minima parte, il meccanismo responsabile della variazione di espressione osservata nella maggior parte dei geni, tuttavia è comunque molto interessante andare ad analizzare la diversa regolazione dell'interleuchina IL21 e del gene SH2B3, che anche questa volta mostra avere un ruolo importante, infatti, l'elevata espressione del gene nelle cellule epiteliali dei celiaci correla con una maggiore espressione del gene.

Anche nell'analisi discriminante effettuata sugli epitelii abbiamo riscontrato, come ci aspettavamo, che le interleuchine ed i geni infiammatori hanno un enorme peso nella classificazione dei soggetti celiaci e non celiaci, ovviamente trattandosi di tessuto intestinale e poiché abbiamo messo a confronto un tessuto epiteliale infiammato e uno non infiammato, i geni infiammatori hanno ruolo chiave nella classificazione dei pazienti. L'analisi ha evidenziato il ruolo cruciale svolto dal IL21, in associazione alla IL12 che non compare tra le variabili selezionate dall'analisi perché strettamente dipendente dall'altra citochina. Tuttavia anche in questo caso risulta evidente il ruolo di SH2B3.

Dunque i nostri risultati suggeriscono che tutta la catena di geni coinvolti nella regolazione dei processi infiammatori svolge un ruolo chiave a livello delle cellule epiteliali, come evidenziato dall'analisi discriminante e come riscontrato anche nella analisi di correlazione.

Dai risultati ottenuti, sia dalla prima parte che della seconda di questo studio, il gene SH2B3 sembrerebbe assumere un ruolo di primo piano nella caratterizzazione molecolare dei pazienti celiaci, per cui una speculazione più approfondita sul possibile ruolo di questo gene all'interno del complesso mosaico della patogenesi della malattia celiaca appare necessario. Come è stato descritto nei precedenti paragrafi, i livelli di espressione genica di SH2B3 sono significativamente più elevati nei celiaci rispetto ai controlli, sia a livello del sangue periferico (già a livello pre-sintomatico) che livello della mucosa intestinale, e in particolare a livello delle cellule epiteliali. I nostri dati sono perfettamente coerenti con quelli della letteratura (*Hunt KA2008; Galatola 2013, Galatola2015*) in cui si osserva sempre una over-espressione del gene. Da un punto di vista biologico questa aumentata espressione di SH2B3 è giustificabile dai molteplici ruoli che esso svolge, e in particolare dal fatto che questo gene è stato descritto come un importante punto di congiunzione tra infiammazione e immunità (*J. Devalliere, 2011*) Andando ad analizzare in particolare il gene SH2B3 chiamato anche Lnk, mappa sul cromosoma 12, in posizione 12q24 e codifica per una proteina membro della famiglia SH2-B (Src homology 2-B).

Inizialmente SH2B3 è stato descritto come regolatore negativo del recettore delle cellule T, e quindi implicato nel processo di signalling della linea T. (*Huang X., 1995; Takaki S, 1997; Li Y, 2000*), Ad oggi ad SH2B3 sono invece stati attribuiti diversi ruoli tra cui quello di regolatore chiave nelle cellule sia ematopoietiche che non ematopoietiche. Per quanto riguarda il sistema ematopoietico, SH2B3 è stato identificato come un regolatore negativo della linfopoiesi B nelle prime fasi di sviluppo, infatti in alcuni lavori è stato dimostrato che il gene viene espresso nei progenitori ematopoietici (HPC) e nelle cellule staminali ematopoietiche (HSC) e

che le capacità proliferativa di progenitori ematopoietici e delle HSC sono notevolmente aumentati in assenza di SH2B3 (*Takaki S,1997;Velazquez L,2002*) .

SH2B3 è anche coinvolto nel blocco delle tre principali vie di segnalazione indotte sia da eritropoietina (EPO) che da trombopoietina (TPO), attenuando l'attivazione JAK2 e regola la capacità di auto-rinnovamento delle HSC, la megacariopoiesi e l'eritropoiesi. Tutti insieme, questi dati supportano il concetto che SH2B3 è un importante regolatore negativo di linfopoiesi, megacariopoiesi, eritropoiesi ed espansione delle HSC, attraverso la modulazione di fattori di crescita e di recettori delle citochine. (*Seita,2007;Tong W, 2004 Bersenev A, 2008*).

Per quanto riguarda le funzioni di SH2B3 al livello delle cellule non ematopoietiche, è stato dimostrato che esso è espresso in cellule endoteliali e che viene fosforilato dalla citochina proinfiammatoria TNF α e di conseguenza rapidamente up-regolato a livello mRNA e proteina (*Bouldav G, 2002*). La capacità di TNF di regolare SH2B3 nell'endotelio anche se in modo PI3Kinase-dipendente è stato ulteriormente osservato in "human umbilical vein endothelial cell" (*Meifang Wan,2006*).

In un recente lavoro è stato dimostrato che la carenza di SH2B3 diminuisce la migrazione sulla membrana dei precursori della molecola VCAM-1, suggerendone un ruolo nei processi di adesione nella motilità e nelle interazioni cellulari (*Takizawa H,2006*).

Mancano ad oggi evidenze sui meccanismi molecolari in cui è coinvolto SH2B3 a livello delle cellule epiteliali, e cosa generi l'up-regolazione del gene rilevata in queste popolazioni cellulari. Ma come si è detto dai risultati del nostro studio è emerso un dato molto interessante circa il meccanismo di regolazione del gene che sembrerebbe essere mediato dalla diminuita metilazione del gene nelle cellule dei pazienti celiaci e che potrebbe giustificare l'over-espressione dello stesso. Questo dato rappresenta la prima evidenza del diretto coinvolgimento di un meccanismo epigenetico nella regolazione di questa importante proteina. I risultati dell'analisi bioinformatica, circa la presenza di elementi di regolazione del DNA proprio nella zona del gene interessata dalle differenze nello status di metilazione, ha fornito un importante supporto ai nostri dati, e alla nostra ipotesi scientifica, tuttavia la ricerca di ulteriori conferme sperimentali potrebbero fornire uno spunto per un ulteriore approfondimento di questo lavoro, andando ad effettuare, ad esempio, dei saggi di affinità in vitro.

Tutti questi meccanismi molecolari, anche se identificati in altri sistemi cellulari, ben si adattano al ruolo della proteina all'interno della patogenesi della celiachia. La mancata regolazione del sistema immunitario dovuta all'over-espressione di SH2B3, così come la sua

contemporanea induzione del gene ad opera delle citochine pro-infiammatorie potrebbe giustificare un loop di attivazione infiammatoria caratteristico della celiachia e mediato proprio dall'iperespressione di SH2B3. Resta da chiarire come l'azione del glutine entri a far parte di questo meccanismo di iperattivazione infiammatoria.

Concludendo, in questo lavoro abbiamo contribuito a dimostrare come nell'insorgere della malattia celiaca, insieme ad una ormai nota componente genetica, concorrano diversi meccanismi sia puramente ambientali che modificatori quali l'epigenetica, e che quindi l'istaurarsi del "fenotipo celiaco" dipenda strettamente dalla deregolazione di diversi pathways molecolari e cellulari. Pertanto, per chiarire ulteriormente le basi molecolari di questa patologia, sarebbe auspicabile un più moderno approccio orientato verso la system biology e le scienze traslazionali.

6. BIBLIOGRAFIA

1. Auricchio R, Cielo D, de Falco R, Galatola M, Bruno V, Malamisura B, Limongelli MG, Troncone R, Greco L. Respiratory Infections and the Risk of Celiac Disease. *Pediatrics*. 2017 Sep 6. pii: e20164102.
2. Beckett CG, Dell'Olio D, Kontakou M, Przemioslo RT, Rosen-Bronson S, Ciclitira PJ: Analysis of interleukin-4 and interleukin-10 and their association with the lymphocytic infiltrate in the small intestine of patients with coeliac disease. *Gut* 39: 818-823, 1996.
3. Beitnes AC, Ráki M, Brottveit M, Lundin KE, Jahnsen FL, Sollid LM. Rapid accumulation of CD14+CD11c+ dendritic cells in gut mucosa of celiac disease after in vivo gluten challenge. *PLoS One*. 2012;7(3):e33556.
4. Bersenev A, Wu C, Balcerak J, Tong W. Lnk controls mouse hematopoietic stem cell self-renewal and quiescence through direct interactions with JAK2. *J Clin Invest* 2008;118:2832
5. *Bioinformatics*. 2002 Nov;18(11):1427-31.
6. Birney E, J.A. Stamatoyannopoulos, A. Dutta, R. Guigo, T.R. Gingeras, E.H. Margulies, et al. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature*. 2007 Jun 14;447(7146):799-816.
7. Bouliday G, Coulon F, Fraser CC, Soullillou JP, Charreau B. Transcriptional up-regulation of the signaling regulatory protein LNK in activated endothelial cells. *Transplantation*. 2002 Nov 15;74(9):1352-4
8. Bourgey M, Calcagno G, Tinto N, Gennarelli D, Margaritte-Jeannin P, Greco L, Limongelli MG, Esposito O, Marano C, Troncone R, Spampanato A, Clerget-Darpoux F, Sacchetti L. HLA related genetic risk for coeliac disease. *Gut*. 2007 Aug;56(8):1054-9. Epub 2007 Mar 7
9. Buoli Comani G, Panceri R, Dinelli M, Biondi A, Mancuso C, Meneveri R, Barisani D. miRNA-regulated gene expression differs in celiac disease patients according to the age of presentation. *Genes Nutr*. 2015 Sep;10(5):482. doi: 10.1007/s12263-015-0482-2. Epub 2015 Aug 2.
10. Capriati T, Francavilla R, Castellaneta S, Ferretti F, Diamanti A. Impact of the birth's season on the development of celiac disease in Italy. *Eur J Pediatr*. 2015 Dec;174(12):1657-63.

11. Castellanos-Rubio A, Fernandez-Jimenez N, Kratchmarov R, Luo X, Bhagat G, Green PH, Schneider R, Kiledjian M, Bilbao JR, Ghosh S. A long noncoding RNA associated with susceptibility to celiac disease. *Science*. 2016 Apr 1;352(6281):91-5.
12. Castellanos-Rubio A, Martin-Pagola A, Santín I, Hualde I, Aransay AM, Castaño L, Vitoria JC, Bilbao JR. Combined functional and positional gene information for the identification of susceptibility variants in celiac disease. *Gastroenterol*. 2008;134(3):738-46.
13. Chiba M, W Bartnik, S G ReMine, W R Thayer, R G Shorter. Human colonic intraepithelial and lamina propria lymphocytes: cytotoxicity in vitro and the potential effects of the isolation method on their functional properties. *Gut* 1981;22:177-186.
14. Devalliere J, Charreau B. The adaptor Lnk (SH2B3): An emerging regulator in vascular cells and a link between immune and inflammatory signaling. *Biochemical Pharmacology* 82 (2011) 1391–1402].
15. Di Sabatino A, Corazza GR. Coeliac disease. *Lancet*. 2009 Apr 25;373(9673):1480-93.
16. Dieli-Crimi R, Cénit MC, Núñez C. The genetics of celiac disease: A comprehensive review of clinical implications. *J Autoimmun*. 2015 Nov;64:26-41
17. Dolcino M, Zanoni G, Bason C, Tinazzi E, Boccola E, Valletta E, Contreas G, Lunardi C, Puccetti A. A subset of anti-rotavirus antibodies directed against the viral protein VP7 predicts the onset of celiac disease and induces typical features of the disease in the intestinal epithelial cell line T84. *Immunol Res*. 2013 Jul;56(2-3):465-76.
18. Falchuk ZM, G.N. Rogentine, W. Strober. Predominance of histocompatibility antigen HL-A8 in patients with gluten-sensitive enteropathy. *J Clin Invest*. 1972 Jun;51(6):1602-5.
19. Fernandez-Jimenez N, Castellanos-Rubio A, Plaza-Izurieta L, Irastorza I, Elcoroaristizabal X, Jauregi-Miguel A, Lopez-Euba T, Tutau C, de Pancorbo MM, Vitoria JC, Bilbao JR. Coregulation and modulation of NFκB-related genes in celiac disease: uncovered aspects of gut mucosal inflammation. *Hum Mol Genet*. 2014 Mar 1;23(5):1298-310.
20. Frisullo G, Nociti V, Iorio R, Patanella AK, Marti A, Assunta B, Plantone D, Cammarota G, Tonali PA, Batocchi AP. Increased CD4+CD25+Foxp3+ T cells in

- peripheral blood of celiac disease patients: correlation with dietary treatment. *Hum Immunol.* 2009 Jun;70(6):430-5
21. Galatola M, Cielo D, Panico C, Stellato P, Malamisura B, Carbone L, Gianfrani C, Troncone R, Greco L, Auricchio R. Presymptomatic Diagnosis of Celiac Disease in Predisposed Children: The Role of Gene Expression Profile. *J Pediatr Gastroenterol Nutr.* 2017 Sep;
 22. Galatola M, Izzo V, Cielo D, et al. Gene Expression Profile of Peripheral Blood Monocytes: A Step towards the Molecular Diagnosis of Celiac Disease? *PlosOne* 2013;8(9): doi: 10.1371/journal.pone.0074747.
 23. Gobbi G1, Bouquet F, Greco L, Lambertini A, Tassinari CA, Ventura A, Zaniboni MG. Coeliac Disease, epilepsy and cerebral calcification. *Lancet* 1992; 340: 439-443.
 24. Greco L1, Romino R, Coto I, Di Cosmo N, Percopo S, Maglio M, Paparo F, Gasperi V, Limongelli MG, Cotichini R, D'Agate C, Tinto N, Sacchetti L, Tosi R, Stazi MA. The first large population based twin study of coeliac disease. *Gut.* 2002 May;50(5):624-8.
 25. Greetje J. Tack, Wieke H. The spectrum of celiac disease: epidemiology, clinical aspects and treatment. *Nature Reviews Gastroenterology and Hepatology* 2010 Apr; 204-213
 26. Gutierrez-Achury J, Zhernakova A, Pulit SL, Trynka G, Hunt KA, Romanos J, Raychaudhuri S, van Heel DA, Wijmenga C, de Bakker PI. Fine mapping in the MHC region accounts for 18% additional genetic risk for celiac disease. *Nat Genet.* 2015 Jun;47(6):577-8.
 27. Habtezion A, Nguyen LP, Hadeiba H, Butcher EC. Leukocyte Trafficking to the Small Intestine and Colon. *Gastroenterology.* 2016 Feb;150(2):340-54.
 28. Harris KM, Fasano A, Mann DL. Monocytes differentiated with IL-15 support Th17 and Th1 responses to wheat gliadin: implications for celiac disease. *Clin Immunol.* 2010 Jun;135(3):430-9.
 29. HogenEsch C E, Rosén A, Auricchio R, Romanos J, Chmielewska A , Putter H, Ivarsson A, Szajewska H, Koning F, Wijmenga C, Troncone R and Mearin ML. For The Prevent CD study group. The Prevent CD study design: towards new strategies for the prevention of celiac disease. *European Journal of Gastroenterology and Hepatology*,2010,22; 1424-1430.
 30. Huang X, Li Y, Tanaka K, Moore KG, Hayashi JI. Cloning and characterization of Lnk, a signal transduction protein that links T-cell receptor activation signal to

- phospholipase C gamma 1, Grb2, and phosphatidylinositol 3-kinase. *Proc Natl Acad Sci U S A* 1995;92:11618–22.
31. Huang XL, Zhang L, Li JP, Wang YJ, Duan Y, Wang J. MicroRNA-150: A potential regulator in pathogens infection and autoimmune diseases. *Autoimmunity*. 2015 Aug 14;1-8.
 32. Hunt KA, Zhernakova A, Turner G, Heap GA, Franke L, Bruinenberg M, Romanos J, Dinesen LC, Ryan AW, Panesar D, Gwilliam R, Takeuchi F, McLaren WM, Holmes GK, Howdle PD, Walters JR, Sanders DS, Playford RJ, Trynka G, Mulder CJ, Mearin ML, Verbeek WH, Trimble V, Stevens FM, O'Morain C, Kennedy NP, Kelleher D, Pennington DJ, Strachan DP, McArdle WL, Mein CA, Wapenaar MC, Deloukas P, McGinnis R, McManus R, Wijmenga C, van Heel DA. Newly identified genetic risk variants for celiac disease related to the immune response. *Nat Genet*. 2008 Apr;40(4):395-402.
 33. Husby S, Koletzko S, Korponay-Szabó IR, Mearin ML, Phillips A, Shamir R et al. European Society for Pediatric Gastroenterology, Hepatology, and Nutrition guidelines for the diagnosis of coeliac disease. *J Pediatr Gastroenterol Nutr*. 2012;54:136–160.
 34. Husby S1, Koletzko S, Korponay-Szabó IR, Mearin ML, Phillips A, Shamir R, Troncone R, Giersiepen K, Branski D, Catassi C, Lelgeman M, Mäki M, Ribes-Koninckx C, Ventura A, Zimmer KP; ESPGHAN Working Group on Coeliac Disease Diagnosis; ESPGHAN Gastroenterology Committee; European Society for Pediatric Gastroenterology, Hepatology, and Nutrition. European Society for Pediatric Gastroenterology, Hepatology, and Nutrition guidelines for the diagnosis of celiac disease. *J Pediatr Gastroenterol Nutr*. 2012 Jan;54(1):136-60.
 35. Izzo V, Pinelli M, Tinto N, Esposito MV, Cola A, Sperandeo MP, Tucci F, Coccozza S, Greco L, Sacchetti L. Improving the estimation of celiac disease sibling risk by non-HLA genes. *PLoS One*. 2011;6(11):e26920
 36. *J Med Genet*. 2011 Jul;48(7):493-6.
 37. Jauregi-Miguel A, Fernandez-Jimenez N, Irastorza I, Plaza-Izurieta L, Vitoria JC, Bilbao Alteration of tight junction gene expression in celiac disease. *J Pediatr Gastroenterol Nutr*. 2014 Jun;58(6):762-7.
 38. Karell K, Louka AS, Moodie SJ, Ascher H, Clot F, Greco L, Ciclitira PJ, Sollid LM, Partanen J; European Genetics Cluster on Celiac Disease. HLA types in celiac disease

- patients not carrying the DQA1*05-DQB1*02 (DQ2) heterodimer: results from the European Genetics Cluster on Celiac Disease. *Hum Immunol.* 2003 Apr;64(4):469-77.
39. King BC, Esguerra JL, Golec E, Eliasson L, Kemper C, Blom AM. CD46 Activation Regulates miR-150-Mediated Control of GLUT1 Expression and Cytokine Secretion in Human CD4+ T Cells. *J Immunol.* 2016 Feb 15;196(4):1636-45.
40. Kontakou M, Przemioslo RT, Sturgess RP, Limb GA, Ciclitira PJ: Expression of tumournecrosisfactor- α , interleukin-6, and interleukin-2 mRNA in the jejunum of patients with coeliac disease. *Scand J Gastroenterol* 30: 456-463, 1995.
41. Kuja-Halkola R, Lebwohl B, Halfvarson J, Wijmenga C, Magnusson PK, Ludvigsson JF Heritability of non-HLA genetics in coeliac disease: a population-based study in 107 000 twins. *Gut.* 2016 May 20.
42. Kumar V, C. Wijmenga, S. Withoff. From genome-wide association studies to disease mechanisms: celiac disease as a model for autoimmune diseases *Semin. Immunopathol.*, 34 (2012), pp. 567–580
43. Kumar V, Gutierrez-Achury J, Kanduri K, Almeida R, Hrdlickova B, Zhernakova DV, Westra HJ, Karjalainen J, Ricaño-Ponce I, Li Y, Stachurska A, Tigchelaar EF, Abdulahad WH, Lähdesmäki H, Hofker MH, Zhernakova A, Franke L, Lahesmaa R, Wijmenga C, Withoff S. Systematic annotation of celiac disease loci refines pathological pathways and suggests a genetic explanation for increased interferon-gamma levels. *Hum Mol Genet.* 2015 Jan 15;24(2):397-409.
44. Kurppa K1, Ashorn M, Iltanen S, Koskinen LL, Saavalainen P, Koskinen O, Mäki M, Kaukinen K. Celiac Disease without Villous Atrophy in Children: A Prospective Study. *J Pediatr.* 2010; Apr 16.
45. Li LC, Dahiya R. MethPrimer: designing primers for methylation PCRs. *Eur J HumGenet.* 2015 Aug;23(8):1100-5.
46. Lionetti E, Castellaneta S, Francavilla R, Pulvirenti A, Tonutti E, Amarri S, Barbato M, Barbera C, Barera G, Bellantoni A, Castellano E, Guariso G, Limongelli MG, Pellegrino S, Polloni C, Ughi C, Zuin G, Fasano A, Catassi C; SIGENP (Italian Society of Pediatric Gastroenterology, Hepatology, and Nutrition) Working Group on Weaning and CD Risk. Introduction of gluten, HLA status, and the risk of celiac disease in children. *N Engl J Med.* 2014 Oct 2;371(14):1295-303.
47. Ludvigsson JF, Leffler DA, Bai JC, Biagi F, Fasano A, Green PH, Hadjivassiliou M, Kaukinen K, Kelly CP, Leonard JN, Lundin KE, Murray JA, Sanders DS, Walker

- MM, Zingone F, Ciacci C. The Oslo definitions for coeliac disease and related terms. *Gut* 2013;62:43–52.
48. Marsh MN. Grains of truth: evolutionary changes in small intestinal mucosa in response to environmental antigen challenge. *Gut* 1990; 31 : 111-458.
49. MeifangW, YoushengLiHuaXue, QiurongLi, JieshouLiM. TNF alpha induces Lnk expression through PI3K-dependent signaling pathway in human umbilical vein endothelial cells. *J Surg Res* 136 2006
50. Ministero della Salute - Relazione annuale al parlamento della Celiachia - Anno 2013.
51. Moldenhauer G1, Momburg F, Möller P, Schwartz R, Hämmerling GJ. Epithelium-specific surface glycoprotein of Mr 34,000 is a widely distributed human carcinoma marker. *Br J Cancer*. 1987 Dec;56(6):714-21.
52. Monticelli S, Ansel KM, Xiao C, Socci ND, Krichevsky AM, Thai TH, Rajewsky N, Marks DS, Sander C, Rajewsky K, Rao A, Kosik KS. MicroRNA profiling of the murine hematopoietic system. *Genome Biol*. 2005;6(8):R71.
53. Mustalahti K, Catassi C, Reunanen A, Fabiani E, Heier M, McMillan S, Murray L, Metzger MH, Gasparin M, Bravi E, Mäki M; Coeliac EU Cluster, Project Epidemiology. The prevalence of celiac disease in Europe: results of a centralized, international mass screening project. *Ann Med* 2010;42. 587–595
54. Nisticò L, Fagnani C, Coto I, Percopo S, Cotichini R, Limongelli MG, Paparo F, D'Alfonso S, Giordano M, Sferlazzas C, Magazzù G, Momigliano-Richiardi P, Greco L, Stazi MA. Concordance, disease progression, and heritability of coeliac disease in Italian twins. *Gut*. 2006 Jun;55(6):803-8. Epub 2005 Dec 14.
55. Östensson M, Montén C, Bacelis J, Gudjonsdottir AH, Adamovic S, Ek J, Ascher H, Pollak E, Arnell H, Browaldh L, Agardh D, Wahlström J, Nilsson S, Torinsson-Nalwai Å. A possible mechanism behind autoimmune disorders discovered by genome-wide linkage and association analysis in celiac disease *PLoS One*. 2013 Aug 2;8(8):e70174.
56. Perry AS, Baird AM, Gray SG5, Epigenetic Methodologies for the Study of Celiac Disease. *Methods Mol Biol*. 2015;1326:131-138
57. Plaza-Izurieta L, Castellanos-Rubio A, Irastorza I, Fernández-Jimenez N, Gutierrez G; CEGEC, Bilbao JR. Revisiting genome wide association studies (GWAS) in coeliac disease: replication study in Spanish population and expression analysis of candidate genes.

58. Plaza-Izurieta L, Fernandez-Jimenez N, Irastorza I, Jauregi-Miguel A, Romero-Garmendia I, Vitoria JC, Bilbao JR. Expression analysis in intestinal mucosa reveals complex relations among genes under the association peaks in celiac disease.
59. Ploski R, J. Ek, E. Thorsby, L.M. Sollid. On the HLA-DQ (alpha 1*0501, beta *0201)-associated susceptibility in celiac disease: a possible gene dosage effect of DQB1*0201. *Tissue Antigens*. 1993 Apr;41(4):173-7.
60. Ráki M, Tollefsen S, Molberg Ø, Lundin KE, Sollid LM, Jahnsen FL. A unique dendritic cell subset accumulates in the celiac lesion and efficiently activates gluten-reactive T cells. *Gastroenterology*. 2006 Aug;131(2):428-38.
61. Ricano-Ponce, C. Wijmenga. Mapping of immune-mediated disease genes. *Annu. Rev. Genom. Hum. Genet.* 2013 (14): 325–353
62. Romanos J, Barisani D, Trynka G, Zhernakova A, Bardella MT, Wijmenga C. Six new coeliac disease loci replicated in an Italian population confirm association with coeliac disease. *J Med Genet* 2009;(46):60–63.
63. Romanos J, van Diemen CC, Nolte IM, Trynka G, Zhernakova A, Fu J, Bardella MT, Barisani D, McManus R, van Heel DA, Wijmenga C. Analysis of HLA and non-HLA alleles can identify individuals at high risk for celiac disease. *Gastroenterology*. 2009 Sep;137(3):834-40, 840.
64. Rubio-Tapia A1, Ludvigsson JF, Brantner TL, Murray JA, Everhart JE. The prevalence of celiac disease in the United States. *Am J Gastroenterol* 2012;107:1538-44.
65. Seita J, Ema H, Ooehara J, Yamazaki S, Tadokoro Y, Yamasaki A, et al. Lnk negatively regulates self-renewal of hematopoietic stem cells by modifying thrombopoietin-mediated signal transduction. *Proc Natl Acad Sci U S A* 2007;104:2349–54.
66. Singh P, Arora S, Lal S, Strand TA, Makharia GK. Risk of Celiac Disease in the First- and Second-Degree Relatives of Patients With Celiac Disease: A Systematic Review and Meta-Analysis. *Am J Gastroenterol*. 2015 Nov;110(11):1539-48.
67. Sollid LM & Jabri N. Triggers and drivers of autoimmunity: lessons from coeliac disease. *Nature Reviews Immunology* 2013,13, 294-302
68. Sollid, L.M., Markussen, G., Ek, J., Gjerde, H., Vartdal, F., and Thorsby, E. Evidence for a primary association of celiac disease to a particular HLA-DQ alpha/beta heterodimer. *J. Exp. Med.* 1989, 169, 345–350.

69. Sperandeo MP1, Tosco A, Izzo V, Tucci F, Troncone R, Auricchio R, Romanos J, Trynka G, Auricchio S, Jabri B, Greco L. Potential celiac patients: a model of celiac disease pathogenesis. *PLoS One*. 2011;6(7):e21281. Epub 2011 Jul 8
70. Spurkland A, Sollid LM, Polanco I, Vartdal F, Thorsby E. HLA-DR and -DQ genotypes of celiac disease patients serologically typed to be non-DR3 or non-DR5/7. *Hum Immunol*. 1992 Nov;35(3):188-92.
71. Stachurska A, Zorro MM, van der Sijde MR, Withoff S. Small and Long Regulatory RNAs in the Immune System and Immune Diseases. *Front Immunol*. 2014 Oct 20;5:513.
72. Stokes PL, Asquith P, Holmes GK, Mackintosh P, Cooke WT. Histocompatibility antigens associated with adult coeliac disease. *Lancet*. 1972 Jul 22;2(7769):162-4.
73. Szajewska H, Shamir R, Mearin L, Ribes-Koninckx C, Catassi C, Domellöf M, Fewtrell MS, Husby S, Papadopoulou A, Vandenas Y, Castillejo G, Kolacek S, Koletzko S, Korponay-Szabó IR, Lionetti E, Polanco I, Troncone R. Gluten Introduction and the Risk of Coeliac Disease: A Position Paper by the European Society for Pediatric Gastroenterology, Hepatology, and Nutrition. *J Pediatr Gastroenterol Nutr*. 2016 Mar;62(3):507-13.
74. Takaki S, Watts JD, Forbush KA, Nguyen NT, Hayashi J, Alberola-Ila J, et al. Characterization of Lnk. An adaptor protein expressed in lymphocytes. *J Biol Chem* 1997;272:14562-70.
75. Takizawa H, Kubo-Akashi C, Nobuhisa I, Kwon SM, Iseki M, Taga T, Takatsu K, Takaki S. Enhanced engraftment of hematopoietic stem/progenitor cells by the transient inhibition of an adaptor protein, Lnk. *Blood*. 2006 Apr 1;107(7):2968-75.
76. Tinto N, Cola A, Piscopo C, Capuano M, Galatola M, Greco L, Sacchetti L. High Frequency of Haplotype HLA-DQ7 in Celiac Disease Patients from South Italy: Retrospective Evaluation of 5,535 Subjects at Risk of Celiac Disease. *PLoS One*. 2015 Sep 23;10(9):e0138324. doi: 10.1371/journal.pone.0138324
77. Tong W, Zhang J, Lodish HF. Lnk inhibits erythropoiesis and Epo-dependent JAK2 activation and downstream signaling pathways. *Blood* 2005;105: 4604-12. Tong W, Lodish HF. Lnk inhibits Tpo-mpl signaling and Tpo-mediated megakaryocytopoiesis. *J Exp Med* 2004;200:569-80.
78. Tosi R, Vismara D, Tanigaki N, Ferrara GB, Cicimarra F, Buffolano W, Follo D, Auricchio S. Evidence that celiac disease is primarily associated with a DC locus allelic specificity. *Clin Immunol Immunopathol*. 1983 Sep;28(3):395-404

79. Trynka G, Hunt KA, Bockett NA, Romanos J, Mistry V, Szperl A, Bakker SF, Bardella MT, Bhaw-Rosun L, Castillejo G, de la Concha EG, de Almeida RC, Dias KR, van Diemen CC, Dubois PC, Duerr RH, Edkins S, Franke L, Fransen K, Gutierrez J, Heap GA, Hrdlickova B, Hunt S, Plaza Izurieta L, Izzo V, Joosten LA, Langford C, Mazzilli MC, Mein CA, Midah V, Mitrovic M, Mora B, Morelli M, Nutland S, Núñez C, Onengut-Gumuscu S, Pearce K, Platteel M, Polanco I, Potter S, Ribes-Koninckx C, Ricaño-Ponce I, Rich SS, Rybak A, Santiago JL, Senapati S, Sood A, Szajewska H, Troncone R, Varadé J, Wallace C, Wolters VM, Zhernakova A; Spanish Consortium on the Genetics of Coeliac Disease (CEGEC); PreventCD Study Group; Wellcome Trust Case Control Consortium (WTCCC), Thelma BK, Cukrowska B, Urcelay E, Bilbao JR, Mearin ML, Barisani D, Barrett JC, Plagnol V, Deloukas P, Wijmenga C, van Heel DA. Multiple common variants for celiac disease influencing immune gene expression. 295-302. *Nat Genet.* 2010; 42.
80. Trynka G, Hunt KA, Bockett NA, Romanos J, Mistry V, Szperl A, Bakker SF, Bardella MT, Bhaw-Rosun L, Castillejo G, de la Concha EG, de Almeida RC, Dias KR, van Diemen CC, Dubois PC, Duerr RH, Edkins S, Franke L, Fransen K, Gutierrez J, Heap GA, Hrdlickova B, Hunt S, Plaza Izurieta L, Izzo V, Joosten LA, Langford C, Mazzilli MC, Mein CA, Midah V, Mitrovic M, Mora B, Morelli M, Nutland S, Núñez C, Onengut-Gumuscu S, Pearce K, Platteel M, Polanco I, Potter S, Ribes-Koninckx C, Ricaño-Ponce I, Rich SS, Rybak A, Santiago JL, Senapati S, Sood A, Szajewska H, Troncone R, Varadé J, Wallace C, Wolters VM, Zhernakova A; Spanish Consortium on the Genetics of Coeliac Disease (CEGEC); PreventCD Study Group; Wellcome Trust Case Control Consortium (WTCCC), Thelma BK, Cukrowska B, Urcelay E, Bilbao JR, Mearin ML, Barisani D, Barrett JC, Plagnol V, Deloukas P, Wijmenga C, van Heel DA. Dense genotyping identifies and localizes multiple common and rare variant association signals in celiac disease. 1193-1201. *NatGenet.* 2011;43(12).
81. Trynka G, Zhernakova A, Romanos J, Franke L, Hunt KA, Turner G, Bruinenberg M, Heap GA, Platteel M, Ryan AW, de Kovel C, Holmes GK, Howdle PD, Walters JR, Sanders DS, Mulder CJ, Mearin ML, Verbeek WH, Trimble V, Stevens FM, Kelleher D, Barisani D, Bardella MT, McManus R, van Heel DA, Wijmenga C. Coeliac disease-associated risk variants in TNFAIP3 and REL implicate altered NF-kappaB signalling. *Gut.* 2009 Aug;58(8):1078-83.

82. Tucci F, Astarita L, Abkari A, Abu-Zekry M, Attard T, Ben Hariz M, Bilbao JR, Boudraa G, Boukthir S, Costa S, Djuricic V, Hugot JP, Irastorza I, Kansu A, Kolaček S, Magazzù G, Mičetić-Turk D, Misak Z, Roma E, Rossi P, Terzic S, Velmishi V, Arcidiaco C, Auricchio R, Greco L. Celiac disease in the Mediterranean area. *BMC Gastroenterol* 2014;11:14:24. doi: 10.1186/1471-230X-14-24.
83. van der Flier LG, Clevers H. Stem cells, self-renewal, and differentiation in the intestinal epithelium. *Annu Rev Physiol.* 2009;71:241-60
84. van Heel DA, Franke L, Hunt KA, Gwilliam R, Zhernakova A, Inouye M, Wapenaar MC, Barnardo MC, Bethel G, Holmes GK, Feighery C, Jewell D, Kelleher D, Kumar P, Travis S, Walters JR, Sanders DS, Howdle P, Swift J, Playford RJ, McLaren WM, Mearin ML, Mulder CJ, McManus R, McGinnis R, Cardon LR, Deloukas P, Wijmenga C. A genome-wide association study for celiac disease identifies risk variants in the region harboring IL2 and IL21. *Nat Genet.* 2007 Jul;39(7):827-9.
85. ve Li Y, He X, Schembri-King J, Jakes S, Hayashi J. Cloning and characterization of human Lnk, an adaptor protein with pleckstrin homology and Src homology 2 domains that can inhibit T cell activation. *J Immunol* 2000;164:5199–206].
86. Velazquez L, Cheng AM, Fleming HE, Furlonger C, Vesely S, Bernstein A, et al. Cytokine signaling and hematopoietic homeostasis are disrupted in Lnkdeficient mice. *J ExpMed* 2002;195:1599–611].
87. Verdu EF, Galipeau HJ, Jabri B. Novel players in coeliac disease pathogenesis: role of the gut microbiota. *Nat Rev Gastroenterol Hepatol.* 2015 Sep;12(9):497-506.
88. Vriezinga S.L: Auricchio R, Bravi E, Castillejo G, Chmielewska A, Crespo Escobar P, Kolaček S, Koletzko S, Korponay-Szabo I, Mummert E, Polanco I, Putter H, Ribes-Konickx C, Shamir R, Szajewska K, Werkstetter K, Greco L et al. Randomized Feedind Intervention in Infants at High Risk for Celiac Disease. *N Engl J Med* 2014;371; 1304,1305.

Altre attività di ricerca intraprese nel corso del Dottorato

Respiratory Infections and the Risk of Celiac Disease

Background ed obiettivi : l'incremento dell'insorgenza della malattia celiaca (MC) suggerisce che le comuni infezioni che si manifestano prima della comparsa della malattia autoimmune potrebbero essere molto importanti ai fini della modulazione della risposta immunitaria. In questo studio siamo andati ad esplorare la relazione tra i primi eventi clinici e lo sviluppo della malattia celiaca in soggetti geneticamente predisposti.

Metodi: per questo lavoro sono stati arruolati 373 neonati dai 0-3 mesi d'età , predisposti geneticamente (preventivamente tipizzati per le molecole HLA di classe seconda, al fine di valutare la presenza di un aplotipo predisponente alla MC) e con almeno un familiare di primo grado celiaco; durante il follow-up sono state effettuate valutazioni cliniche e sierologiche. Per valutare il rischio associato alle singole variabili sono state utilizzate tavole di contingenza e odds ratio; per determinare le variabili che contribuiscono al rischio di sviluppare la MC è stata utilizzata un' analisi di regressione logistica. Infine è stata effettuata un' analisi discriminante per determinare le variabili che distinguono i casi dai controlli prima della diagnosi.

Risultati: l'incidenza cumulativa di MC in questa coorte è risultata essere del 6% a 3 anni e il 13.5% a 5 anni; 34 bambini (14%) hanno sviluppato la MC entro il sesto anno di vita. L' analisi degli eventi ha mostrato un' elevata frequenza di infezioni del tratto respiratorio nei bambini celiaci durante i primi 24 mesi di vita. Il risultato dell'analisi discriminante che include sesso e classe di rischio HLA, è che le infezioni respiratorie nel primo e nel secondo anno di vita contribuiscono significativamente a discriminare i casi dai controlli.

Conclusioni: un modello multivariato di analisi discriminante ha mostrato che la frequenza delle infezioni respiratorie nei primi due anni di vita potrebbe discriminare bambini che svilupperanno la malattia celiaca da quelli che non la svilupperanno mai.

Respiratory Infections and the Risk of Celiac Disease

Renata Auricchio, MD, PhD,^{a,b} Donatella Cielo, MSc,^a Renato de Falco, MD,^a Martina Galatola, MSc, PhD,^a Valentina Bruno, MD,^a Basilio Malamisura, MD,^c Maria Giovanna Limongelli, MD,^d Riccardo Troncone, MD, PhD,^{a,b} Luigi Greco, MD, PhD^{a,b}

abstract

BACKGROUND AND OBJECTIVES: The increasing incidence of celiac disease (CD) suggests that common infections before the onset of autoimmune diseases could be an important factor in switching the immune response. We aimed to explore the relationship between early clinical events and the development of CD in genetically predisposed infants.

METHODS: In this study, 373 newborns from families with at least 1 relative with CD were recruited, and human leukocyte antigen DQ2- or DQ8-positive infants were followed up with clinical and serological evaluations. Cross tabulation and odds ratios were used to explore the risk associated with single variables, and logistic regression analysis was performed to determine the variables that contributed to the risk of developing CD. Stepwise discriminant analysis was used to determine which variables could distinguish case patients from controls before diagnosis.

RESULTS: The cumulative incidence of CD in this cohort was 6% at 3 years and 13.5% at 5 years of age, and 134 children (14%) developed CD before the sixth year of life. An analysis of adverse events showed a higher frequency of respiratory tract infections among CD patients during the first 24 months of life. In a stepwise discriminant analysis, which included sex and human leukocyte antigen risk class, only respiratory infections in the second and first years of life significantly contributed to discrimination of case patients versus controls.

CONCLUSIONS: A multivariate model of discriminant analysis showed that the frequency of respiratory infections in the first 2 years of life could distinguish children who developed CD from those who did not.



^aDepartment of Translational Medical Science and ^bEuropean Laboratory for the Investigation of Food Induced Diseases, University of Naples Federico II, Naples, Italy; ^cDepartment of Pediatrics, University Hospital of Salerno, Salerno, Italy; and ^dAzienda Ospedaliera Gaetano Rummo Via dell' Angelo, Benevento, Italy

Dr Auricchio obtained funding, supervised the study, conceptualized and designed the study, drafted the initial manuscript, analyzed and interpreted the data, and performed critical revision of the manuscript for important intellectual content; Ms Cielo acquired data, analyzed and interpreted the data, conducted the initial analyses, and drafted the initial manuscript; Dr de Falco acquired data, conducted the initial analyses, and drafted the initial manuscript; Dr Galatola acquired data, analyzed and interpreted the data, drafted the initial manuscript, and performed critical revision of the manuscript; Dr Malamisura recruited the patients, performed clinical evaluation of the subjects, drafted the initial manuscript, and performed critical revision of the manuscript; Dr Bruno recruited the patients, obtained the samples for the experiments, and drafted the initial manuscript; Dr Limongelli obtained the samples for the experiments and performed critical revision of the manuscript for important intellectual content; Dr Troncone obtained funding, supervised the study, analyzed and interpreted the data, and performed critical revision of the manuscript for important intellectual content; Dr Greco obtained funding, supervised the study, conceptualized and designed the study, drafted the initial manuscript, analyzed and interpreted the data, performed statistical analysis, and performed critical revision

WHAT'S KNOWN ON THIS SUBJECT: The incidence of celiac disease (CD) is increasing globally, and its transmission is only partially explained by genetic and nutritional factors. Early infections might contribute to the risk of developing CD.

WHAT THIS STUDY ADDS: This longitudinal study of CD screening found that respiratory infections during the first 2 years of life are associated with a higher risk of CD in high-risk infants.

To cite: Auricchio R, Cielo D, de Falco R, et al. Respiratory Infections and the Risk of Celiac Disease. *Pediatrics*. 2017; 140(4):e20164102

The incidence of celiac disease (CD) is epidemic in communities that consume gluten. The incidence of CD is increasing globally, especially in emerging countries, where it has become a major public health burden.¹ At least 5 million new cases of CD are expected to be diagnosed in Mediterranean countries in the next 10 years.² Familial cases represent a large proportion of this increase because a significant increase has occurred in the prevalence of CD in infants from at-risk families. We estimated that the familial recurrence is ~13% to 15% during the first 6 years of life and more than 20% later in life.³⁻⁵ When we consider human leukocyte antigen (HLA) DQ2- or DQ8-positive infants from families with a proband, the risk of developing CD is not homogenous across the progeny.⁶ Only up to 40% of the heredity of CD is explained by the HLA haplotype and the 54 other CD-associated genes.⁷ Nutritional factors do not appear to confer additional risk for the development of CD.^{3,8} Before any clinical signs of CD appear and at least 1 year before anti-tissue transglutaminase antibodies (anti-tTG) are produced, infants who develop CD within the first 6 years of life show distinct expression patterns of immune-related and non-immune-related CD-associated genes.⁹

We are interested in understanding whether the process of autoimmunity leading to CD is stimulated or switched on by the occurrence of common infections during the period preceding its onset. The authors of several studies have suggested that early infections might contribute to the risk of developing CD. This connection was proposed in 2013 when a study revealed an association between rotavirus infections and the recognition of gluten peptides.¹⁰ Anti-rotavirus VP7 antibodies are present in the majority (81%) of serum samples from patients with diabetes mellitus type 1 and CD but are also

detectable in a fraction (27%) of children with diabetes mellitus type 1 without CD. Moreover, it was found that anti-rotavirus VP7 antibodies are present in the sera before the onset of CD and the detection of anti-tTG and anti-endomysium antibodies.¹⁰ In the Norwegian Mother and Child Cohort Study,¹¹ 581 children (0.8%) were diagnosed with CD. Children with ≥ 10 infections before 18 months of age had a significantly higher risk of developing CD later in life than children with ≤ 4 infections. The added risk of CD was associated with upper respiratory tract infections, lower respiratory tract infections, and gastroenteritis. In an Italian prospective cohort study, 1227 children developed CD among 203 000 infants; in this study, gastrointestinal infections were significantly associated with a subsequent diagnosis of CD (incidence rate ratio, 2.04).¹² Furthermore, a Swedish retrospective case-control study investigated 3835 children diagnosed with CD before the age of 2 years and found that 36 (0.9%) had a previous diagnosis of respiratory syncytial virus compared with 117 (0.6%) out of 19 102 matched controls (odds ratio [OR], 1.46).¹³ The highest risk was observed in children who developed CD before 1 year of age (OR, 1.82). The authors also described a previous incidence of viral bronchiolitis in 3.4% (132 out of 3835) of the case patients with CD and in 2.0% (390 out of 19 102) of the matched controls (OR, 1.60).¹³

However, epidemiologic studies are difficult to interpret because it is difficult to control for confounding variables and recall bias. Therefore, a prospective study of a cohort of at-risk infants, stratified for HLA genetic risk, is ideal to study risk factors for the development of CD.

The aim of this study was to explore the relationship between early clinical events (including infections) and the development of CD in a

prospective cohort of genetically predisposed infants.

METHODS

Patients

A cohort of 373 Italian newborns with 1 first-degree relative with a confirmed case of CD was enrolled in this study from 2007 to 2015 (191 female participants). Until 2011, the newborns were enrolled and followed up with the PreventCD European project,¹⁴ and after 2011, newborns were enrolled and followed up with the PreventCD protocol. Among the newborns, 248 (124 boys and 124 girls) were genetically predisposed to CD by the presence of the HLA DQ2 or DQ8 haplotype. The predisposed individuals were strictly monitored from birth through the end of their sixth year of life.

The follow-up protocol included monthly clinical and serological assessments during the first 6 months of life, followed by assessments every 3 months until 1 year of age, every 6 months between the first year and the third year of life, and then annually until the sixth year of life.

The mothers were encouraged to breastfeed for at least 6 months, and all children were gradually weaned and introduced to gluten between 6 and 9 months of age.¹⁴

Sample Size Estimation

This prospective cohort study included 248 infants. To estimate the incidence of CD (the percentage of the enrolled participants who developed CD within the first 6 years of life), we compared an expected incidence of 10% in the infants not exposed to infections with a predicted incidence of 25% in the infants exposed to infections, and by using a 95% confidence interval, we estimated a minimum sample of 37 cases of CD.

Recording Growth and Health

The growth and health of each infant was regularly reported at each visit (4, 6, 9, 12, 18, 24, 36, 48, 60, and 72 months) in a personal logbook that was kept at home, shared with the family pediatrician, and systematically copied at the study center. All visits to the family pediatrician and drug prescriptions were reported. Infections were recorded by the family pediatrician or at the center only if they had a duration of >24 hours and required medical attention, according to the usual well-infant clinic protocol adopted regionally by the family pediatrician.

The reported clinical events (gastroenteritis, diarrhea, vomiting, fever, constipation, and upper and lower respiratory tract infections) were grouped into bins of 6 months (eg, 0–6, 7–12, 13–18 months, etc). Incidences of diarrhea and gastroenteritis that occurred in the 3 months preceding the diagnosis of CD were considered related to the diagnosis and, therefore, were excluded from the analysis.

During the 6 years of observation, 35 subjects (14.1%) were diagnosed with CD (14 boys, 21 girls) to be further referred to as the case patients. The children who did not develop CD by the age of 6 years will be referred to as controls. The diagnosis of CD was based on the repeated production of elevated anti-tTG and the development of severe mucosal damage (Marsh type 3c). Of the 35 CD case patients, 14 out of 35 children showed relevant clinical symptoms, whereas 21 out of 35 produced repeated positive anti-tTG antibodies without any clinical signs. To avoid selection bias, we did not exclude any case patient with missing data, and we obtained full log books of all clinically relevant events from 238 children (31 who developed CD and 207 who did not).

TABLE 1 Comparison of CD-Affected Relatives (Family Proband) in the Case Patients (With CD) and Controls (Without CD)

Affected Proband	Controls (%)	Case Patients (%)	Total (%)
Father ± siblings	22 (11)	9 (26.5) ^a	31 (13.2)
Mother ± siblings	106 (53)	14 (41.2)	120 (51.3)
Siblings	72 (36)	11 (32.4)	83 (35.5)

^a Statistical difference observed in case patients with father affected compared with controls ($P = .04$).

Statistics

The OR and the χ^2 test were used to explore the risk associated with single variables, whereas a logistic regression analysis was performed to determine the variables that contribute to the risk of developing CD. To obtain logistic regression scores, we grouped the number of events for each child and each time bin into a binomial classification (present or absent).

A survival analysis was used to observe the occurrence of events over time, with a Wilcoxon rank test used to compare groups.

Stepwise discriminant analysis was adopted to determine which variables were capable of distinguishing case patients from controls before the case patients were diagnosed. Wilks' λ (ranging from 1 to 0) was used to estimate the distance between the 2 groups (case patients and controls) produced by each variable entered into the discriminant function.

Ethical Approval

The study was approved by the medical ethics committee of University of Naples Federico II. The authors vouch for the veracity and completeness of the data and analyses reported and for the adherence of the study to the protocol.

RESULTS

Risk at Birth

The cumulative incidence of new cases of CD was 6% at 3 years and 13.5% at 5 years of age. CD was more frequent in girls (female to male ratio = 21:14 for case patients and 101:106 for controls), but the differences in the sex distribution were not statistically significant ($P = .47$).

No differences were observed in the age of the mother, weeks' gestation, or birth weight between the case patients and controls (Supplemental Table 4). Conversely, significant differences were observed for the type of delivery, with 23 out of 32 of the case patients (72%) and 96 out of 175 of the controls (55%) born by a cesarean delivery ($P = .05$). The proband within the family was also different; 9 out of 34 of the case patients (26.5%) and 22 out of 200 (11%) of the controls had a father that was affected by CD ($P = .04$). The frequency of siblings affected by CD was similar among the case patient and control groups (Table 1).

No significant difference was observed in the mean duration of breastfeeding (case patients, 6.93 months; controls, 5.93 months). Gluten introduction (see "per protocol" described ref 15) was also similar in both groups.

Risk Related to Events

Table 2 shows the incidence (new events, child, year) of symptoms and infections for the case patients and the controls during the first, second, and third years of life. The log books of reported events were complete for 31 case patients who developed CD and 207 controls who did not.

During the first year, when no child produced anti-tTG antibodies, respiratory infections (upper and lower tract) were more common among the case patients than among the controls (58.1% vs 40.1%).

In the second year of life, respiratory infections were again more frequent among the case patients than controls (51.6% vs 32.4%).

TABLE 2 Incidence of Clinical Events (Symptoms and Infections) in Case Patients and Controls in the First 3 Years of Life

	No. of Incident Events		OR (CI)
	31 Case Patients (%)	207 Controls (%)	
First year			
Gastroenteritis	6 (19.4)	46 (22.2)	0.84 (0.32–2.17)
Vomiting	5 (16.1)	35 (16.9)	0.94 (0.34–2.63)
Fever	7 (22.6)	40 (19.3)	1.22 (0.49–3.02)
Constipation	5 (16.1)	13 (6.3)	2.87 (0.94–8.70)
Respiratory infections	18 (58.1)	83 (40.1)	2.07 (0.96–4.44)
Second year			
Gastroenteritis	10 (32.3)	43 (20.8)	1.81 (0.79–4.14)
Vomiting	1 (3.2)	8 (3.9)	0.89 (0.10–6.88)
Fever	4 (12.9)	25 (12.1)	1.07 (0.35–3.33)
Constipation	3 (9.7)	12 (5.8)	1.74 (0.46–6.55)
Respiratory infections	16 (51.6)	67 (32.4)	2.23 (1.04–4.77)
Third year			
Gastroenteritis	6 (19.4)	21 (10.1)	2.12 (0.78–5.77)
Vomiting	1 (3.2)	6 (2.9)	1.11 (0.13–9.60)
Fever	3 (9.7)	18 (8.7)	1.12 (0.31–4.06)
Constipation	1 (3.2)	13 (6.3)	0.49 (0.63–3.94)
Respiratory infections	11 (35.5)	54 (26.1)	1.55 (0.70–3.46)

CI, confidence interval.

In the third year of life, which is when most of the case patients were diagnosed with CD, no clinical event was more frequent in the case patients than in the control group. Only gastroenteritis was marginally increased among the case patients.

Multivariate Analysis

Because clinical events are related each other (eg, fever is likely to be associated with infection, vomiting is associated with gastroenteritis, etc), we conducted a multivariate analysis to explore the differences in the occurrence of risk-related events between the case patients and the control group.

Once all the clinical events were grouped into a present or absent binomial classification, a logistic regression was adopted to estimate the crude OR for developing CD

according to the clinical events during the first 3 years of life (Table 3). Only respiratory infections in the second year of life were associated with a twofold increase in the risk of developing CD (OR, 2.25; confidence interval: 1.03–4.88; $P = .04$). The second variable, which was not included in the model, was respiratory infections in the first year of life, which had a score of 1.58.

We performed a stepwise discriminant analysis to determine which of the clinical events discriminated between the case patients and the controls (Supplemental Tables 5 and 6). For this analysis, we used the full information for the number of clinical events in each time bin (0–6 months, 7–12 months, etc, on a scale from 0 to 5 events) for each child. We included sex and HLA risk class in the model. As shown in Supplemental Table 4, only respiratory infections in the second

and first years of life significantly contributed to the lower Wilks' λ , the index of discrimination. Sex and HLA risk class did not significantly contribute to the discrimination of this model once clinical events were included (data not shown in the tables).

DISCUSSION

Children from families with at least 1 first-degree relative with CD have an estimated probability of 13% of developing CD within the first 6 years of life.

Confirming previous studies, we showed that newborns who were homozygous for the HLA DQ2 haplotype were at a higher risk of developing CD.^{3,5,15} A single copy of HLA DQ2 conferred the second highest risk, whereas the HLA DQ8 haplotype was found more often in the controls. Despite the relatively small numbers in the cohort, we observed significant differences in the survival curve of each HLA haplotype risk group ($P = .02$, data not shown).

A higher rate of cesarean delivery, which was indeed high in our population, was observed in the children who developed CD within the first 6 years of life, contrary to what was suggested by Lionetti et al.¹⁶ We observed that more newborns that had a father affected by CD were subsequently diagnosed with CD (compared with those with a mother or sibling affected). This suggests that the genetic risk transmitted by the father is higher than the risk transmitted by the mother. The authors of several studies have shown that a father with CD may confer a higher risk to the progeny than

TABLE 3 Logistic Regression of the Clinical Events

	B	ES	Wald	Degrees of Freedom	Significance	OR	95% CI	
							Lower	Upper
Respiratory infection	0.812	0.396	4.211	1	0.04	2.252	1.037	4.888
second year constant	−2.244	0.281	63.723	1	0.000	0.106	NA	NA

The Wald statistic is the square of the ratio of the logistic regression coefficient to its SE and has a χ^2 distribution. B, unstandardized regression weight; ES, error standard; NA, not applicable.

affected mothers because of genetic risk factors and not to environmental prenatal factors.^{17,18}

Breastfeeding and gluten introduction could not be analyzed in this longitudinal study because of the study protocol, which strongly encouraged breastfeeding and set the timeline for the introduction of gluten at 6 months.

Compared with gastroenteritis, respiratory infections during the first 2 years of life conferred a twofold increase in the risk of developing CD. According to a multivariate model, the variance explained by respiratory infections was higher than that explained by sex or HLA haplotype. It should be noted that this cohort was selected by the presence of an at-risk HLA haplotype; therefore, no infants in this study were without an at-risk HLA haplotype.

Other longitudinal studies failed to reveal an association between early infections and the development of CD in genetically predisposed infants.³ However, this finding may be because of the complexity of the study design¹⁹ and not because of the absence of a risk effect. When each event is reported with its time point, without grouping into 6-month bins, the flat spread of the events over time causes an aliasing effect that may undermine the observation of significant results. The aliasing effect is produced by listing the mean time of events at different time points without setting a defined time interval. To explore the relationship between an event that may have occurred at any time and the subsequent appearance of CD, we stratified the reported events into 6-month bins, which reduced the time-related bias. Therefore, we could estimate the risk of developing CD in relation to the number of events that occurred during a set time interval.

CD has a strong genetic component, and ~40% of the transmitted genetic risk is explained by specific HLA haplotypes. In addition, 54

CD-associated gene polymorphisms confer up to an additional 10% of the genetic variance.²⁰ Therefore, nearly half of the heredity cannot be explained by known genetic risk factors. Additionally, most of the polymorphisms associated with CD do not produce changes in the amino acid sequence of the gene and are likely to have regulatory and epigenetic functions. To date, the search for environmental factors that explain the epidemic of CD has produced conflicting results.²¹ Neither breastfeeding nor gluten introduction, which have consistently been shown to be associated with a delay in the onset of CD, have been confirmed as factors that contribute to the development of CD.^{3,8,22} In addition, familial and perinatal factors that have been observed in large population cohorts have not been confirmed to be associated with CD.²³

In this study, we report that early infections significantly contribute to the risk of developing CD. It is possible that the exposure to early infection stimulates a genetically predisposed immune profile, which contributes to the switch from tolerance to intolerance to gluten, which is a common food antigen.

In recent studies, researchers have suggested that viruses that infect the intestines can disrupt intestinal immunity at inductive and effector sites of oral tolerance by the suppression of regulatory T cells and the promotion of a type 1 T helper cell response. Indeed, the activation of the type 1 T helper cells immune cascade after exposure to dietary antigens is likely related to the interferon regulatory factor 1, which is not related to the stop to regulatory T cell conversion.²⁴

The numerous reports on the protective effect of early infections against the development of autoimmune and allergic diseases,²⁵ described as the hygiene hypothesis, contradict the previously proposed hypothesis. Indeed, the immune

response to infections may modulate natural immunity via mechanisms that can drive tolerance as well as intolerance, according to the pathways involved.

Environmental factors that occur from the prenatal period to early infancy may actively modulate the immune response to gluten in individuals with peculiar genetic profiles and contribute to the gene–environment interactions by modulating gene expression through epigenetic mechanisms.^{9,20,26}

CONCLUSIONS

The strength of this study is the carefully planned design for a controlled longitudinal cohort study. Each child was monitored at home by the family pediatrician using a personal logbook of clinical events. The production of anti-tTG was strictly monitored at fixed time points. However, there are some limitations to this study. The time span of this study was limited to 6 years of life. Although 13.3% of at-risk infants develop CD by 6 years of age, further diagnoses may appear later in life. Furthermore, no laboratory tests could be used to support the reported incidence of clinical events.

The sample size of the cases is marginally smaller than the optimal predicted, and this may have limited our ability to show the significance of risk factors (type II error); a more precise estimation could be achieved by increasing the sample size.

ACKNOWLEDGMENT

We thank Luisa Mearin for critically revising the manuscript.

ABBREVIATIONS

anti-tTG: anti-tissue transglutaminase antibodies
CD: celiac disease
HLA: human leukocyte antigen
OR: odds ratio

of the manuscript for important intellectual content; and all authors approved the final manuscript as submitted and agree to be accountable for all aspects of the work.

DOI: <https://doi.org/10.1542/peds.2016-4102>

Accepted for publication Jul 20, 2017

Address correspondence to Donatella Cielo, MSc, Department of Translational Medical Science, University of Naples Federico II, Via S. Pansini 5, 80131 Naples, Italy. E-mail: dona_cielo@yahoo.it

PEDIATRICS (ISSN Numbers: Print, 0031-4005; Online, 1098-4275).

Copyright © 2017 by the American Academy of Pediatrics

FINANCIAL DISCLOSURE: The authors have indicated they have no financial relationships relevant to this article to disclose.

FUNDING: Funded by the PreventCD project (EU-FP6-2005-F00D4B-contract 036383). The authors thank the European Laboratory for Food Induced Diseases. The funders had no role in the study design, data collection or analysis, decision to publish, or preparation of the manuscript.

POTENTIAL CONFLICT OF INTEREST: The authors have indicated they have no potential conflicts of interest to disclose.

REFERENCES

1. Tucci F, Astarita L, Abkari A, et al. Celiac disease in the Mediterranean area. *BMC Gastroenterol.* 2014;14:24
2. Greco L, Timpone L, Abkari A, et al. Burden of celiac disease in the Mediterranean area. *World J Gastroenterol.* 2011;17(45):4971–4978
3. Vriezinga SL, Auricchio R, Bravi E, et al. Randomized feeding intervention in infants at high risk for celiac disease. *N Engl J Med.* 2014;371(14):1304–1315
4. Ågårdh D, Lee HS, Kurppa K, et al; TEDDY Study Group. Clinical features of celiac disease: a prospective birth cohort. *Pediatrics.* 2015;135(4):627–634
5. Lionetti E, Castellaneta S, Francavilla R, et al; SIGENP (Italian Society of Pediatric Gastroenterology, Hepatology, and Nutrition) Working Group on Weaning and CD Risk. Introduction of gluten, HLA status, and the risk of celiac disease in children. *N Engl J Med.* 2014;371(14):1295–1303
6. Izzo V, Pinelli M, Tinto N, et al. Improving the estimation of celiac disease sibling risk by non-HLA genes. *PLoS One.* 2011;6(11):e26920
7. Romanos J, Rosén A, Kumar V, et al; PreventCD Group. Improving coeliac disease risk prediction by testing non-HLA variants additional to HLA variants. *Gut.* 2014;63(3):415–422
8. Szajewska H, Shamir R, Chmielewska A, et al; PREVENTCD Study Group. Systematic review with meta-analysis: early infant feeding and coeliac disease—update 2015. *Aliment Pharmacol Ther.* 2015;41(11):1038–1054
9. Galatola M, Izzo V, Cielo D, et al. Gene expression profile of peripheral blood monocytes: a step towards the molecular diagnosis of celiac disease? *PLoS One.* 2013;8(9):e74747
10. Dolcino M, Zanoni G, Bason C, et al. A subset of anti-rotavirus antibodies directed against the viral protein VP7 predicts the onset of celiac disease and induces typical features of the disease in the intestinal epithelial cell line T84. *Immunol Res.* 2013;56(2–3):465–476
11. Mårild K, Kahrs CR, Tapia G, Stene LC, Størdal K. Infections and risk of celiac disease in childhood: a prospective nationwide cohort study. *Am J Gastroenterol.* 2015;110(10):1475–1484
12. Canova C, Zabeo V, Pitter G, et al. Association of maternal education, early infections, and antibiotic use with celiac disease: a population-based birth cohort study in northeastern Italy. *Am J Epidemiol.* 2014;180(1):76–85
13. Tjernberg AR, Ludvigsson JF. Children with celiac disease are more likely to have attended hospital for prior respiratory syncytial virus infection. *Dig Dis Sci.* 2014;59(7):1502–1508
14. Hogen Esch CE, Rosén A, Auricchio R, et al; PreventCD Study Group. The PreventCD Study design: towards new strategies for the prevention of coeliac disease. *Eur J Gastroenterol Hepatol.* 2010;22(12):1424–1430
15. Bourgey M, Calcagno G, Tinto N, et al. HLA related genetic risk for coeliac disease. *Gut.* 2007;56(8):1054–1059
16. Lionetti E, Castellaneta S, Francavilla R, Pulvirenti A, Catassi C; SIGENP Working Group of Weaning and CD Risk. Mode of delivery and risk of celiac disease: risk of celiac disease and age at gluten introduction cohort study. *J Pediatr.* 2017;184:81–86.e2
17. Ludvigsson JF, Ludvigsson J. Coeliac disease in the father affects the newborn. *Gut.* 2001;49(2):169–175
18. Greco L. The father figure in coeliac disease. *Gut.* 2001;49(2):163–164
19. Lönnrot M, Lynch K, Larsson HE, et al; TEDDY Study Group. A method for reporting and classifying acute infectious diseases in a prospective study of young children: TEDDY. *BMC Pediatr.* 2015;15:24
20. Dubois PC, Trynka G, Franke L, et al. Multiple common variants for celiac disease influencing immune gene expression. *Nat Genet.* 2010;42(4):295–302
21. Withoff S, Li Y, Jonkers I, Wijmenga C. Understanding celiac disease by genomics. *Trends Genet.* 2016;32(5):295–308
22. Andrén Aronsson C, Lee HS, Koletzko S, et al; TEDDY Study Group. Effects of gluten intake on risk of celiac disease: a case-control study on a Swedish birth cohort. *Clin Gastroenterol Hepatol.* 2016;14(3):403–409.e3
23. Saccone G, Berghella V, Sarno L, et al. Celiac disease and obstetric complications: a systematic review and metaanalysis. *Am J Obstet Gynecol.* 2016;214(2):225–234
24. Bouziat R, Hinterleitner R, Brown JJ, et al. Reovirus infection triggers inflammatory responses to dietary

- antigens and development of celiac disease. *Science*. 2017;356(6333):44–50
25. Strisciuglio C, Giugliano F, Martinelli M, et al. Impact of environmental and familial factors in a cohort of pediatric patients with inflammatory bowel disease. *J Pediatr Gastroenterol Nutr*. 2017;64(4):569–574
26. Capuano M, Iaffaldano L, Tinto N, et al. MicroRNA-449a overexpression, reduced NOTCH1 signals and scarce goblet cells characterize the small intestine of celiac patients. *PLoS One*. 2011;6(12):e29094

Presymptomatic Diagnosis of Celiac Disease in Predisposed Children: The Role of Gene Expression Profile

Obiettivi: La prevalenza della malattia celiaca (MC) ha avuto un incremento significativo negli ultimi anni, e la predizione del rischio e la diagnosi precoce sono diventate indispensabili soprattutto nelle famiglie a rischio. In uno studio precedente, abbiamo identificato individui con MC in base al profilo di espressione di un insieme di geni candidati nei monociti di sangue periferico. In questo lavoro abbiamo valutato l'espressione di un pannello di geni candidati nella MC in cellule mononucleate di sangue periferico di bambini, predisposti geneticamente, molto tempo prima della comparsa di qualsiasi sintomo o produzione di anticorpi.

Metodi: Abbiamo analizzato l'espressione di un set di 9 geni candidati, associati alla MC, in 22 bambini predisposti geneticamente alla MC e provenienti da famiglie con almeno un soggetto celiaco. I soggetti sono stati monitorati clinicamente e sierologicamente (anticorpi anti-tessuto-transglutaminasi (anti-tTG) ogni 3 - 6 mesi dalla nascita a 6 anni. Nei bambini che hanno sviluppato anticorpi anti-tTG, la diagnosi è stata confermata da una biopsia intestinale. Abbiamo analizzato 9 bambini che hanno sviluppato la malattia celiaca ad una età media di 30 mesi e 13 controlli (bambini della stessa coorte che non hanno sviluppato la celiachia fino a 6 anni). Abbiamo analizzato l'espressione a tre differenti time point: prima della diagnosi, al momento della diagnosi e dopo il primo anno di dieta senza glutine. Per i 13 controlli sono stati analizzati time point simili.

Risultati: tre dei geni analizzati (KIAA, TAGAP [T-cell Activation GTPase Activating Protein], and SH2B3 [SH2B Adaptor Protein 3]) sono risultati essere overespressi nei soggetti celiaci, rispetto ai controlli, almeno 9 mesi prima della diagnosi. Dall'analisi discriminante, 4 geni (RGS1 [Regulator of G-protein signaling 1], TAGAP, TNFSF14 [Tumor Necrosis Factor (Ligand) Superfamily member 14], and SH2B3) sembrano differenziare i soggetti celiaci dai controlli molto prima della comparsa della produzione anticorpale e dei sintomi clinici. Inoltre l'equazione discriminante ottenuta ci permette una classificazione diagnostica corretta nel 95.5% dei pazienti.

Conclusioni: questo studio ci ha permesso di fare un'analisi presintomatica, ossia di prevedere quali bambini, appartenenti a famiglie a rischio, seguiti dalla nascita fino a 6 anni sviluppano la MC, molto prima della produzione di auto-anticorpi (anti-tTG) o della comparsa dei sintomi clinici.

Presymptomatic Diagnosis of Celiac Disease in Predisposed Children: The Role of Gene Expression Profile

*[†]Martina Galatola, *[†]Donatella Cielo, *Camilla Panico, *Pio Stellato, [‡]Basilio Malamisura, *Lorenzo Carbone, ^{†§}Carmen Gianfrani, *[†]Riccardo Troncone, *[†]Luigi Greco, and *[†]Renata Auricchio

ABSTRACT

Objective: The prevalence of celiac disease (CD) has increased significantly in recent years, and risk prediction and early diagnosis have become imperative especially in at-risk families. In a previous study, we identified individuals with CD based on the expression profile of a set of candidate genes in peripheral blood monocytes. Here we evaluated the expression of a panel of CD candidate genes in peripheral blood mononuclear cells from at-risk infants long time before any symptom or production of antibodies.

Methods: We analyzed the gene expression of a set of 9 candidate genes, associated with CD, in 22 human leukocyte antigen predisposed children from at-risk families for CD, studied from birth to 6 years of age. Nine of them developed CD (patients) and 13 did not (controls). We analyzed gene expression at 3 different time points (age matched in the 2 groups): 4–19 months before diagnosis, at the time of CD diagnosis, and after at least 1 year of a gluten-free diet. At similar age points, controls were also evaluated.

Results: Three genes (*KIAA*, *TAGAP* [T-cell Activation GTPase Activating Protein], and *SH2B3* [SH2B Adaptor Protein 3]) were overexpressed in patients, compared with controls, at least 9 months before CD diagnosis. At a stepwise discriminant analysis, 4 genes (*RGS1* [Regulator of G-protein signaling 1], *TAGAP*, *TNFSF14* [Tumor Necrosis Factor (Ligand) Superfamily member 14], and *SH2B3*) differentiate patients from controls before serum antibodies production and clinical symptoms. Multivariate equation correctly classified CD from non-CD children in 95.5% of patients.

Conclusions: The expression of a small set of candidate genes in peripheral blood mononuclear cells can predict CD at least 9 months before the appearance of any clinical and serological signs of the disease.

Key Words: celiac disease, celiac disease first-degree relatives, gene expression, presymptomatic diagnosis, risk factors

(*JPGN* 2017;65: 314–320)

Received August 9, 2016; accepted January 10, 2017.

From the *Department of Translational Medical Science, University of Naples Federico II, Naples, Italy, the [†]European Laboratory for Food-Induced Disease (ELFID), University of Naples Federico II, Naples, Italy, the [‡]Pediatric Unit, University Hospital of Salerno, Cava de' Tirreni, and the [§]Institute of Protein Biochemistry-CNR, Naples, Italy. Address correspondence and reprint requests to Martina Galatola, M.Sc., PhD, Department of Translational Medical Science, University of Naples Federico II, Via S. Pansini 5, 80131 Naples, Italy (e-mail: martinagalatola@hotmail.it).

This work was funded by the PREVENT-CD project: EU-FP6-2005-FOOD4B-contract no. 036383. The authors thank European Laboratory for Food-Induced Disease (ELFID). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Supplemental digital content is available for this article. Direct URL citations appear in the printed text, and links to the digital files are provided in the HTML text of this article on the journal's Web site (www.jpgn.org).

The authors report no conflicts of interest.

Copyright © 2017 by European Society for Pediatric Gastroenterology, Hepatology, and Nutrition and North American Society for Pediatric Gastroenterology, Hepatology, and Nutrition

DOI: 10.1097/MPG.0000000000001519

What Is Known

- A panel of genes associated to celiac disease have a confirmed role in the pathogenesis of gluten-induced immune response.
- The expression of 12 genes clearly identify celiac from non-celiac disease subjects in mucosa and in peripheral blood monocytes.

What Is New

- At least 9 months before any presence of antibodies in the serum or clinical signs, the expression of a small set of celiac disease candidate genes in peripheral blood mononuclear cells predict the development of celiac disease in at-risk infants.

The estimated prevalence of celiac disease (CD) in first-degree relatives is as high as 10%, which is 10 times higher than in the general population (estimated around 1%) (1–4). Concordance in monozygotic twins is higher than 80%; there is a strong genetic component, which is unusual for a multifactorial disease (5). Recurrence in families is an important source of new patients and an early diagnosis in these at-risk families become an urgent challenge (3,4).

Recently the European Society for Paediatric Gastroenterology, Hepatology and Nutrition (ESPGHAN) recommended to limit the use of small intestinal biopsy for the diagnosis of CD, opening the perspective of non-invasive diagnosis in child or adolescent with typical CD symptoms, tissue-transglutaminase IgA higher than 10 times the reference value and presence of human leukocyte antigen (HLA)-DQ2/-DQ8 genes (6).

Unfortunately, the molecular features of CD genetics are still not fully elucidated. Although 95% of patients carry HLA-DQ2/-DQ8, this trait explains just 35% to 40% of the genetic variance (7,8). Repeated genome-wide association studies identified up to 57 single-nucleotide polymorphisms (SNPs) associated to CD, but they accounted for just a small fraction (approximately 6.5%) of the heredity of CD; together with HLA they explained about half the heredity (9–18).

Given the complexity of CD, the genetic data available are not yet sufficient for disease prediction. In an attempt to obtain functional data, we previously explored the expression of a set of CD-associated genes in peripheral blood monocytes (PBMs) easily available, rather than in the intestinal mucosa tissue (19). Indeed, Ontiveros et al (20) showed that plasma levels of inflammatory cytokines in whole blood collected after acute (3 days) oral gluten

challenge could differentiate patients with CD from controls who have adopted gluten-free diet (GFD) for at least 6 months.

These results encouraged us to explore the adoption of peripheral blood mononuclear cells (PBMCs) as a tool for diagnosis of CD instead of intestinal biopsy. A multicenter randomized study has been performed trying to reduce the risk of CD in children with a first-degree relative affected by CD. This study aimed to identify risk factors associated with CD (breast-feeding, time and quantity of gluten introduction, other environmental and genetic factors) by following them from birth to 5 years of age (21). This study allowed collecting serial blood samples from a subset of infants at risk for CD at fixed time points from birth up to 6 years of age, developing an adequate structure for a longitudinal study.

The aim of the present study is to longitudinally evaluate the expression of a panel of CD-candidate genes directly in PBMCs, in genetically predisposed children with a first-degree relative affected by CD, before the appearance of any clinical or serological markers of CD.

PATIENTS AND METHODS

Population Enrolled in the Study

Participants were infants from the Italian cohort of newborns from at-risk families (ie, families with at least a case of CD), bearing HLA-DQ2 or -DQ8, monitored from birth to 6 years of age, enrolled during the PREVENT-CD study (21,22). Subjects were monitored clinically and serologically (anti-tissue-transglutaminase antibody [anti-tTG]) every 3 to 6 months from birth to 6 years old. In children who developed anti-tTG antibodies, CD was confirmed by a small intestinal biopsy (21,22). We analyzed 9 children who developed CD at a median age of 30 months, and 13 age- and sex-matched controls (children from the same cohort who did not develop CD up to 6 years of age). Among infants who developed CD, 5 were asymptomatic, 4 had symptoms, 3 showed growth retardation, of which one was associated to diarrhea, the 4th had recurrent abdominal pain.

In the PREVENT-CD cohort studied in our center, 24 children became celiac before 6 years of age. Unfortunately, the number of samples of sufficient size and quality required for this study was lower because of progressive deterioration of samples for the long-time span from recruitment to the disclosure of the double-blind design of the study (>6 years). Therefore, only for the 9 patients here analyzed, we were able to use the biological material in all time points established in this work. No selection bias could be envisaged in this subsample.

Management of Samples

Genotyping was performed on DNA obtained from cord blood. Gene expression was assessed on blood sample of 9 subjects (CD patients) in 3 different time points: before antibody production (Time 1), which ranged from 4 to 13 months of age, median 12 months; at the age of anti-tTG positivity and confirmed diagnosis (Time 2), which ranged from 18 to 50 months, median 30 months of age; and at least 1 year after starting GFD (Time 3), which ranged from 36 to 48 months, median 36 months of age. Blood samples were collected at similar time points from 13 controls at Time 1, which ranged from 5 to 19 months of age, median 7 months; at Time 2, which ranged from 18 to 50 months, median 18 months of age; and at Time 3, which ranged from 23 to 62 months, median 36 months of age. None of the infants neither in the group who became celiac nor in the group that did not develop CD produced any measurable level of anti-tTG at the time T1 when Gene Expression was first examined.

In all CD patients, and in 6 of the 13 controls, a duodenal biopsy was performed according to PREVENT-CD protocol based

on clinical or serological suspicion. Of these 6 controls, 1 had diarrhea, whereas the other 5 were asymptomatic but showed at least 1 positive value of CD-associated antibody (anti-gliadin antibodies [AGA], anti-tissue transglutaminase [anti-tTG], endomysial antibodies [EMA]) (22). In all controls, the diagnosis of CD up to 6 years of age was excluded. Features of each subject enrolled in this study are listed in Table 1.

Ethics Statement

The study protocol was approved by the ethics committee of the University of Naples Federico II. Biopsies were taken during routine hospital admission requested for diagnostic purposes: parents or guardian gave their informed consent about the endoscopic procedures and the biopsy. Patients did not undergo specimen sampling over and above those requested by the routine diagnostic procedures according to the ESPGHAN guidelines (6).

Isolation of Peripheral Blood Mononuclear Cells

Blood was obtained while children were in stable clinical conditions without any acute illness in the last 3 weeks. Blood samples (4–5 mL) were collected and PBMCs were isolated by Ficoll gradient according to the manufacturer's protocol. PBMCs were frozen and stored in cryoprotective media containing 10% dimethylsulfoxide and fetal bovine serum.

RNA Extraction and Gene Expression Studies

Total RNA was extracted from PBMCs using TRIZOL Reagent (Life Technologies, Foster City, CA). The quantity of RNA was measured using a Nanodrop spectrophotometer and RNA quality was analyzed using agarose gel electrophoresis in Tris/Borate/ethylenediaminetetraacetic acid buffer. The RNA (1 µg) was reverse transcribed into cDNA using the High Capacity cDNA Reverse Transcription kit (Life Technologies). The experiments were performed with a 7900HT Fast Real-Time polymerase chain reaction (PCR) system using the TaqMan Gene Expression Assay (Life Technologies), and approximately 40 ng of cDNA according to the manufacturer's instructions. The relative expression of each gene was obtained using the $\Delta\Delta C_t$ method, and normalized to an endogenous housekeeping gene Glucuronidase (*GUSb*). We used *GUSb* as reference gene after it was identified as the most stable reference gene of 5 candidates (*β -actin*, *B2M*, *GAPDH*, *GUSb*, and *HPRT1*). In particular, the relative quantification (RQ) was calculated normalizing each sample against the mean value of T2 samples of unaffected subjects. The SDS software (ABI, version 1.4 or 2.4, Life Technologies, Foster City, CA) was used to analyze raw data, and statistical analysis was performed on GraphPad Prism 5.01H. Gene expression experiments were conducted according to minimum information for publication of quantitative real-time PCR experiments (MIQE) guidelines (<http://www.gene-quantification.de/miqe-bustin-et-al-clin-chem-2009.pdf>). The candidate genes evaluated, and their respective assays used for the analysis are listed in Supplemental Digital Content, Table 1, <http://links.lww.com/MPG/A885>.

Genotyping

Patients were genotyped for a set of candidate genes as described elsewhere and for HLA (23,24). Patients and controls were grouped into 5 HLA-haplotype classes, as reported previously (24). Nine non-HLA SNPs, located in the candidate genes and listed in the Supplemental Digital Content, Table 2 (<http://links.lww.com/MPG/A885>), were analyzed by genotyping assay.

TABLE 1. Individual characteristics of patients enrolled in the study

UIC	Sex	Status	T1 Before-CD	Age (mo) T2 At CD-diagnosis	T3 At-GFD	Histology*	At T1† (U/mL)	Anti-tTG T2† (U/mL)	T3† (U/mL)	HLA-DQ	Proband in family
P1	Male	CD	9	18	36	M3a/b/c	0.50	>100.00	5.20	DQ2.5/X	Mother
P2	Male	CD	13	30	47	M3a/b/c	0.10	67.10	0.90	DQ2.5/X	Father
P3	Female	CD	9	20	36	M3a/b/c	0.10	>100.00	0.10	DQ2.5/DQ7	Mother
P4	Female	CD	12	32	48	M3a/b/c	0.10	36.00	0.20	DQ2.2/X	Father + Sib
P5	Female	CD	12	50	NA	M1	0.10	>100.00	0.10	DQ2.5/DQ2.2	Mother
P6	Female	CD	12	25	36	M3a/b/c	0.10	8.00	0.10	DQ2.5/X	Sibs
P7	Female	CD	12	37	NA	M3a/b/c	0.10	>100.00	0.80	DQ2.5/DQ2.2	Mother
P8	Female	CD	9	20	37	M3a/b/c	0.10	>100.00	0.80	DQ2.5/DQ2.2	Father + Sib
P9	Female	CD	4	35	NA	M3a/b/c	0.10	>100.00	0.10	DQ2.2/DQ7	Sibs
P10	Male	No-CD	6	18	NA	M0	0.10	0.10	0.10	DQ2.5/DQ7	Sibs
P11	Male	No-CD	9	24	48	M0	0.10	0.10	0.10	DQ2.2/X	Mother
P12	Female	No-CD	9	25	62	M0	0.10	0.10	0.10	DQ2.5/DQ2.2	Mother+ Sib
P13	Female	No-CD	4	25	48	M1	0.10	0.20	0.50	DQ8/X	Mother
P14	Female	No-CD	6	19	27	M1	0.10	4.50	1.00	DQ2.2/DQ7	Father
P15	Female	No-CD	19	35	42	M1	0.10	4.80	0.60	DQ2.2/DQ7	Sibs
P16	Male	No-CD	9	18	35	NA	0.10	0.10	0.10	DQ2.2/X	Mother+ Sib
P17	Female	No-CD	5	12	23	NA	0.10	0.20	0.10	DQ2.5/DQ7	Mother
P18	Female	No-CD	8	12	23	NA	0.10	0.10	0.10	DQ2.2/DQ7	Sibs
P19	Female	No-CD	9	20	36	NA	0.10	0.10	0.10	DQ8/X	Sibs
P20	Male	No-CD	6	12	23	NA	0.10	0.10	0.10	DQ2.2/X	Mother
P21	Female	No-CD	9	17	61	NA	0.10	0.10	0.10	DQ2.5/DQ8	Sibs
P22	Male	No-CD	6	18	36	NA	0.10	2.40	0.30	DQ2.2/DQ2.2	Sibs

anti-tTG = anti-tissue transglutaminase; CD = celiac disease; GFD = gluten-free diet; T1 = before antibody production; T2 = age at appearance of antibodies and mucosal damage; T3 = at least 1 year after diagnosis on GFD; NA = not available; UIC = unique identification code.

*For the diagnosis of CD Marsh classification has been applied, all biopsied controls have a normal duodenal mucosa with no atrophy (Marsh lesion stage M0-M1).

†Reference values of anti-tTG: Negative: <4.0 U/mL; Doubt: >4.0 and <10.0 U/mL; Positive: >10.0 U/mL.

The SNPs were genotyped using TaqMan technology (Life Technologies) using TaqMan SNP Genotyping Assays on a 7900HT Fast Real-Time PCR System (Applied Biosystems, Foster City, CA); the final volume was 15 μ L, containing master mix, TaqMan assays and about 60 ng of genomic DNA template.

Statistical Analysis

Non-parametric rank sum test was adopted to compare gene expression because of the small sample size. Frequencies were compared using the χ^2 -test, with probability for null hypothesis equal to 0.05. A discriminant analysis was performed to estimate the contribution of the genotype and or the expression of each gene to distinguish patients from controls. Wilks' lambda provides an estimate of the cumulative discriminating capacity between 2 groups produced by the multivariate combination of variables, ranging from 1 = complete overlap to 0 complete distance.

By multiplying the standardized value of each variable included in the stepwise discriminant equation by its respective regression coefficient we obtain a discriminant score D, which express the spatial location of each individual to the right or to the left: it is then possible to assign the individual to the group of patients or the group of controls, only on the basis of the genetic data, blind to the final conclusive diagnosis. Hence, the percentage of correct classification may be obtained. Statistical analyses were performed with the SPSS 17.0 (SPSS Inc, Chicago, IL).

RESULTS

Human Leukocyte Antigen Genotyping

Due to the a priori selection of genetically predisposed children, the distribution of HLA genotypes did not differ

significantly between patients and controls. Three patients but no controls are, however, homozygote for DQ2. Five of the 13 children in the latter group have an HLA low-risk class (DQ8 or heterozygote for DQ2).

Genotyping of Candidate Genes

Although there were few chances to observe significant differences between the 2 groups, because of the small sample size, 3 of 9 candidate genes explored showed a different distribution of the risk allele. For ReticuloEndotheliosis Viral Oncogene homolog (*cREL*) and SH2B Adaptor Protein 3 [*SH2B3*] genes the risk allele "A" is more frequent in patients, than in controls, whereas for Tumor Necrosis Factor Receptor Superfamily member 14 (*TNFRSF14*) the "A" allele is less frequent (Supplemental Digital Content, Table 2 and Fig. 1, <http://links.lww.com/MPG/A885>).

Gene Expression Analysis of Candidate Genes

Time 1 (Before Symptoms or Anti-tTG Production)

KIAA, T-cell Activation GTPase Activating Protein (*TAGAP*), and SH2B3 expression was higher in patients before CD diagnosis, with differences of RQ of 0.6-, 0.5-, and 0.7-fold higher in patients versus controls, respectively. The expression of Tumor Necrosis Factor (Ligand) Superfamily member 14 (*TNFSF14*) and Regulator of G-protein signaling 1 (*RGS1*) was higher in controls (Fig. 1A), with differences of RQ level of 1.5, and 1.2-fold higher in controls versus patients. The differences were statistically significant for TAGAPSH2B3 and TNFSF14 (Mann-Whitney *U* test, $P < 0.05$), but not significant for KIAA and RGS1 (Supplemental Digital Content, Table 3 and Fig. 1, <http://links.lww.com/MPG/A885>).

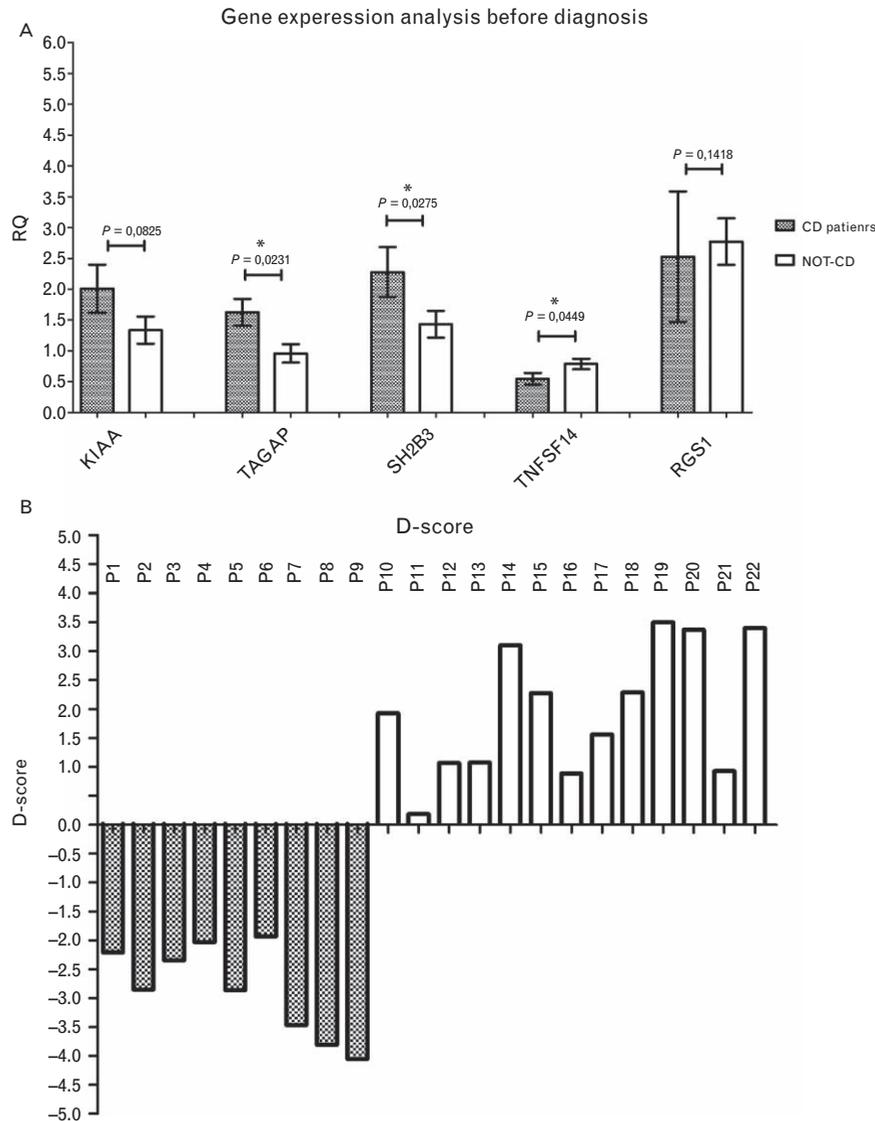


FIGURE 1. (A) mRNA expression of candidate genes in PBMC at least 9 months before the diagnosis (Time 1). KIAA, TAGAP, and SH2B3 expression was higher in 9 patients than in 13 controls, with the differences of RQ level of 0.6, 0.5, and 0.7-fold change, respectively. TNFSF14 and RGS1 expression was lower in patients than in controls, with differences of RQ level of 1.5, and 1.2. *Mann-Whitney U test $P < 0.05$. (B) Distribution of the discriminant score in patients and in controls. The D-score clearly separated the 2 groups of subjects evaluated. Only 9 patients had a negative D-score. PBMC = peripheral blood mononuclear cells; RGS1 = Regulator of G-protein signaling 1; TAGAP = T-cell Activation GTPase Activating Protein; TNFSF14 = Tumor Necrosis Factor (Ligand) Superfamily member 14.

Time 2 (Diagnosis)

At moment of CD diagnosis, only 2 genes, SH2B3 and TNFSF14, showed significant higher expression in patients versus controls, with a differences of RQ level of 0.5, and 0.6-fold (Mann-Whitney test $P = 0.02$ and 0.04 , respectively) (Supplemental Digital Content, Fig. 1, <http://links.lww.com/MPG/A885>).

Time 3 (1 Year After Diagnosis)

The expression of the genes examined did not differ significantly between the 2 groups at time point 3 (data not shown).

Multivariate Analysis of Candidate Genes Genotyping

Considering the observed differences in genotype distribution in patients versus controls (Supplemental Digital Content, Table 2, <http://links.lww.com/MPG/A885>), we explored the combined profile of the 9 candidates' genes in the patients compared with controls. In a stepwise multivariate discriminant analysis, 5 genotypes significantly discriminated between CD and controls, with a correct classification of 91% (Table 1 panels A and B).

TABLE 2. Discriminant analysis of candidate genes genotype

Step	Genotype	Wilks' lambda	Variance ratio F	
			Statistic	P
(A) Stepwise discriminant analysis of genotypes				
1	SH2B3	0.604	13.091	0.002
2	RGS1	0.48	10.292	0.001
3	TAGAP	0.429	7.973	0.001
4	cREL	0.35	7.908	0.001
5	LPP	0.307	7.233	0.001

Status	Predicted group		Total
	CD	Not CD	
Original group	CD 8 (89%)	Not CD 1 (11.1%)	9
	Not CD 1 (7.7%)	CD 12 (92.3%)	13

Overall correct classification = 91%

Step	Genotype	Wilks' lambda	Variance ratio F	
			Statistic	P
(C) Stepwise discriminant analysis of gene expression at T1				
1	TAGAP	0.742	6.938	0.016
2	TNFSF14	0.542	8.019	0.003
3	SH2B3	0.446	7.453	0.002
4	RGS1	0.366	7.377	0.001

Status	Predicted group		Total
	CD	Not CD	
Original group	CD 9 (100%)	Not CD 0 (0%)	9
	Not CD 1 (7.7%)	CD 12 (92.3%)	13

(D) Classification by discriminant equation of gene expression at T1

Original group	CD 9 (100%)	Not CD 0 (0%)	9
	Not CD 1 (7.7%)	CD 12 (92.3%)	13

Overall correct classification = 95%

(A) Five genes (*SH2B3*, *RGS1*, *TAGAP*, *cREL*, and *LPP*) were selected for discriminating capacity, with a *P* value <0.001. (B) Results of the prediction analysis: 92.3% of controls and 89% patients were correctly classified. Panels 1 (C and D) Discriminant analysis of candidate genes gene expression before diagnosis (Time 1). (C) Four genes significantly contributed to lower Wilks' lambda (*P* < 0.001); *TAGAP*, *TNFSF14*, *SH2B3*, and *RGS1* were selected for discriminating capacity. (D) Results of the prediction analysis: 92.3% of controls and 100% of patients were correctly classified.

CD = celiac disease; LPP = Lipoma Preferred Partner; RGS1 = Regulator of G-protein signaling 1; SH2B3 = SH2B Adaptor Protein 3; TAGAP = T-cell Activation GTPase Activating Protein; TNFSF14 = Tumor Necrosis Factor (Ligand) Superfamily member 14.

Multivariate Analysis of Candidate Genes Expression

Since gene expression is inter-correlated within the same cell type; the expression of each candidate gene may not provide an accurate picture of eventual differences between groups. Some genes may be over-expressed due to the effect of another gene in the same pathway. In a stepwise discriminant analysis (Table 2 panels A and B), the combined expression of 4 genes (*TAGAP*, *TNFSF14*, *SH2B3*, and *RGS1*) significantly discriminated between patients and controls before diagnosis (Time 1); in fact, it correctly predicted 9 of 9 patients and 12 of 13 controls (overall 95% correct prediction). Figure 1B shows the discriminant score obtained by

TABLE 3. Discriminant analysis combining genotype and gene expression and of candidate genes before diagnosis (Time 1)

Step	Genotype	Wilks' lambda	Variance ratio F	
			Statistic	P
(A) Stepwise discriminant analysis of combined genotype and gene expression at T1 of candidate genes				
1	SH2B3 genotype	0.604	13.091	0.002
2	RGS1 genotype	0.480	10.292	0.001
3	cREL expression at T1	0.340	11.626	0.000
4	TNFRSF14 genotype	0.260	12.066	0.000
5	SH2B3 expression at T1	0.217	11.523	0.000
6	TNFSF14 genotype	0.190	10.656	0.000

Status	Predicted group		Total
	CD	Not CD	
Original group	CD 9 (100%)	Not CD 0 (0%)	9
	Not CD 0 (0%)	CD 13 (100%)	13

(B) Classification by discriminant equation of combined genotype and gene expression at T1 of candidate genes

Original group	CD 9 (100%)	Not CD 0 (0%)	9
	Not CD 0 (0%)	CD 13 (100%)	13

Predicted group by Jackknife iterative exclusion

Status	Predicted group		Total
	CD	Not CD	
Original group	CD 7 (77.8%)	Not CD 2 (22.2%)	9
	Not CD 2 (15.4%)	CD 11 (84.6%)	13

(C) Classification by discriminant equation of combined genotype and gene expression at T1 of candidate genes

Original group	CD 7 (77.8%)	Not CD 2 (22.2%)	9
	Not CD 2 (15.4%)	CD 11 (84.6%)	13

(A) The genotype or the expression of 5 genes significantly contributed to lowering Wilks' lambda in a stepwise process (*SH2B3* genotype, *RGS1* genotype, *cREL* expression at T1, *TNFRSF14* genotype, and *TNFSF14* genotype) and were selected for discriminating capacity. (B) Computing the relative membership probability by the discriminant score, 100% of patients were correctly classified. (C) Cross validation analysis 84.6% of controls and 77.8% patients were correctly classified.

CD = celiac disease; RGS1 = Regulator of G-protein signaling 1; SH2B3 = SH2B Adaptor Protein 3; TAGAP = T-cell Activation GTPase Activating Protein; TNFSF14 = Tumor Necrosis Factor (Ligand) Superfamily member 14

combining the expression of the 4 selected genes in each individual case. All of the children with CD had a negative D-score and all those without CD had a positive D-score, no child was wrongly classified as "CD."

Multivariate Analysis Combining Genotype With Gene Expression

Although a discriminant function based on a relatively small sample performs well in prediction, it may not be sufficiently reliable for diagnostic purposes. To enhance the robustness of the model, we pooled the genotyping and gene expression data at Time 1, and applied the stepwise multivariate approach reported above. Table 3 shows the performance of genotypes and expression of the candidate genes.

The genotypes of 2 genes (*SH2B3* and *RGS1*) were the most powerful in discriminating between patients and controls. The expression of *cREL* increased the discriminating capacity. The next most powerful discriminator was the *TNFRSF14* genotype, followed by *SH2B3* expression, and finally the *TNFSF14* genotype

(Table 3A). This combination resulted in a low Wilks' lambda (0.604–0.190); consequently, this discriminant equation allowed to correctly classify 100% of subjects (Table 3B).

To verify the robustness of the model, we adopted an auto-exclusion strategy (each individual was predicted by an equation excluding his own data, iteratively) to obtain an unbiased estimate of the predicting capacity of the discriminant function. In this case, only 2 of 9 CD patients were not predicted as CD, and only 2 of 13 controls were wrongly predicted as CD; 81.8% of CD patients and controls were correctly classified (Table 3C).

DISCUSSION

Children from families at risk for CD have a consistent risk of occurrence, from 1% to >20% depending on their HLA (5); they contribute notably to the growth of the celiac population. In fact, in Europe, with a population around 740,000,000, 5 to 7 million CD patients are likely to generate several tens of thousands of affected children each year. Over and above the HLA risk, we recently showed that, within each HLA risk class, the presence of an "at-risk allele" of only 3 candidate genes significantly increases the precision of the estimate of risk of occurrence, with a refined range of risk estimate ranging between 0.04 and 0.08 in DQ8-carrying individuals and between 0.17 and 0.23 in DQB1*02 homozygote (23). The HLA genotype in addition to the genotype of the candidate genes, however, explains less than half the variance of CD heredity. The other half resides in the modulation of the action of genes (expression and regulation). The "missing heredity" of the disease could be explained by the action of epigenetic mechanisms such as DNA methylation, histone modifications, and microRNA regulation, by the interaction between genes and environmental risk factors, but this field of study is still the subject of just a few scientific articles (25,26).

The aim of this study was to predict which children from at-risk families will eventually develop CD from birth to 6 years of age, well in advance of the production of auto-antibodies (anti-tTG) or the appearance of clinical symptoms. We confirmed the importance of the HLA haplotype, and found that the genotype of 5 candidate genes (*SH2B3*, *RGS1*, *TAGAP*, *cREL*, and *LPP*) would seem increases (>90%) the risk of developing CD. But most children in at-risk families share their genotypes with affected relatives, and those selected for HLA-DQ2-DQ8 positivity should have the vast majority of genes equally distributed; hence, it is more informative to explore the expression of genes rather than their genotype.

From a simple blood sample, we were able to predict which children would develop CD within 18 to 24 months in this selected cohort. In fact, 5 genes (*KIAA*, *TAGAP*, *SH2B3*, *TNFSF14*, and *RGS1*) were differentially expressed long before any sign of the disease, and therefore identified with remarkable accuracy the children who did eventually developed CD. To increase the robustness of the prediction we combined the genotype of the child, obtained at birth, with gene expression at 4 to 16 months, and obtained a combination of genotypes and gene expression profiles that significantly increased (up to 100%) the ability to predict the outcome of the child.

Because the discriminant equation by which we classified patients at least 9 months before the CD diagnosis was developed from the data obtained in the same cohort, we adopted an auto-exclusion strategy ("jack-knifing") in which we excluded iteratively from the equation the individuals to be classified. Notably, the differential gene expression (CD vs controls) observed before the appearance of specific antibodies, decrease at moment of diagnosis, and was no longer present 12 or more months after the diagnosis and the start of the GFD. The before/after study design provides the most efficient matched control of the gene expression

profile observed before diagnosis. The reversibility of the expression of some genes in CD patients treated with a GFD is in line with the reversibility of deranged tight junction gene expression in CD children after 2 years of GFD (26,27). The reversibility of gene expression may indirectly confirm a diagnosis of CD, and consequently the evaluation of the expression can be considered a diagnostic marker (26). Studies of the potential role in CD diagnosis of these and other gene expressions, however, indicate that, unlike constitutive genetic profiles, a number of epigenetic regulation mechanisms of these candidate genes, may play a pivotal role in the etiology of CD (26–29).

This study has the limit of a small sample size, which however is counterbalanced by the longitudinal study design of a peculiar cohort of children at high risk of developing CD in the first years of their life. The longitudinal study design allows to observe important modifications of the gene expression in the same individual, which was analyzed 3 times: before diagnosis, at diagnosis, and after diagnosis, multiplying by 3 the power of the study. It is noteworthy that this model is applied in this context for the first time, since the follow-up of babies at risk for CD is not yet a common practice. We have also to acknowledge that to recruit infants from at-risk families, which have to undergo a surveillance lasting 6 years, with regular blood sampling on a scheduled basis, was made possible only by the PREVENT-CD longitudinal study. We can assure that no selection bias was applied to recruit these infants, other than the technical availability of the repeated blood samples.

Another limitation of this study is the small number of genes explored; this was a strategic choice because we preferred to explore genes with robust replication in several Genome Wide Association Studies (10–16), with a strong biological implication in the gluten-induced immune response (17), and explored in previous studies from our group in several models (19,23). Finally, it is important to underline that the selection of this candidate genes is not biased for the present cohort; it was identified in previous studies.

Given that epigenetic mechanisms are highly cell-type specific, analyses conducted with mixed cell tissues (eg, mucosal biopsies) carry the risk to be confounded by differences in cell type composition, particularly when comparing inflamed versus non-inflamed tissues. In our study design, we minimized this bias because we did not compare an inflamed tissue with a healthy one since we worked with PBMCs, when no major sign of antibody production or inflammation was present; and we compared the same cell type in the same individual before and after the occurrence of a major gluten-induced auto-immune event.

In conclusion, our preliminary work may pave the way for a molecular diagnosis of CD, before the onset of antibody production, in infants who carry a genetic risk profile. To consolidate these preliminary results, we are engaged in continuing longitudinal studies of at-risk infants, expanding not only the sample size but also the study of the "regulome" of these subjects. The diagnostic procedures for chronic diseases are slowly shifting from invasive, tissue-based techniques to less invasive molecular methodologies. The time is now ripe for such a common disease as CD to gradually shift to a molecular diagnosis for the vast majority of patients.

Acknowledgments: The authors thank Jean Ann Gilder (Scientific Communication srl, Naples, Italy) for revising and editing the manuscript.

REFERENCES

1. Green PH, Cellier C. Celiac disease. *N Engl J Med* 2007;357:1731–43.
2. Di Sabatino A, Corazza GR. Coeliac disease. *Lancet* 2009;373:1480–93.

3. Mustalahti K, Catassi C, Reunanen A, et al. The prevalence of celiac disease in Europe: results of a centralized, international mass screening project. *Ann Med* 2010;42:587–95.
4. Tucci F, Astarita L, Abkari A, et al. Celiac disease in the Mediterranean area. *BMC Gastroenterol* 2014;14:24.
5. Greco L, Romino R, Coto I, et al. The first large population based twin study of coeliac disease. *Gut* 2002;50:624–8.
6. Husby S, Koletzko S, Korponay-Szabó IR, et al. Working Group of European Society of Paediatric Gastroenterology and Nutr European Society for Pediatric Gastroenterology, Hepatology, and Nutrition guidelines for the diagnosis of coeliac disease. *J Pediatr Gastroenterol Nutr* 2012;54:136–60.
7. Schuppan D, Junker Y, Barisani D. Celiac disease: from pathogenesis to novel therapies. *Gastroenterology* 2009;137:1912–33.
8. Sollid LM, Jabri B. Triggers and drivers of autoimmunity: lessons from coeliac disease. *Nat Rev Immunol* 2013;13:294–302.
9. Romanos J, van Diemen CC, Nolte IM, et al. Analysis of HLA and non-HLA alleles can identify individuals at high risk for celiac disease. *Gastroenterology* 2009;137:834–40.
10. van Heel DA, Franke L, Hunt KA, et al. A genome-wide association study for celiac disease identifies risk variants in the region harboring IL2 and IL21. *Nat Genet* 2007;39:827–9.
11. Hunt KA, Zhernakova A, Turner G, et al. Newly identified genetic risk variants for celiac disease related to the immune response. *Nat Genet* 2008;40:395–402.
12. Castellanos-Rubio A, Martin-Pagola A, Santín I, et al. Combined functional and positional gene information for the identification of susceptibility variants in celiac disease. *Gastroenterology* 2008;134:738–46.
13. Romanos J, Barisani D, Trynka, et al. Six new coeliac disease loci replicated in an Italian population confirm association with coeliac disease. *J Med Genet* 2009;46:60–3.
14. Trynka G, Zhernakova A, Romanos J, et al. Coeliac disease-associated risk variants in TNFAIP3 and REL implicate altered NF- κ B signaling. *Gut* 2009;58:1078–83.
15. Dubois PC, Trynka G, Franke L, et al. Multiple common variants for celiac disease influencing immune gene expression. *Nat Genet* 2010;42:295–302.
16. Trynka G1, Hunt KA, Bockett NA, et al. Dense genotyping identifies and localizes multiple common and rare variant association signals in celiac disease. *Nat Genet* 2011;43:1193–201.
17. Kumar V, Gutierrez-Achury J, Kanduri K, et al. Systematic annotation of celiac disease loci refines pathological pathways and suggests a genetic explanation for increased interferon-gamma levels. *Hum Mol Genet* 2015;24:397–409.
18. Dieli-Crimi R, Cénit MC, Núñez C. The genetics of celiac disease: a comprehensive review of clinical implications. *J Autoimmun* 2015;64:26–41.
19. Galatola M, Izzo V, Cielo D, et al. Gene expression profile of peripheral blood monocytes: a step towards the molecular diagnosis of celiac disease? *Plos One* 2013;8:e74747.
20. Ontiveros N, Tye-Din JA, Hardy MY, et al. Ex-vivo whole blood secretion of interferon (IFN)- γ and IFN- γ -inducible protein-10 measured by enzyme-linked immunosorbent assay are as sensitive as IFN- γ enzyme-linked immunospot for the detection of gluten-reactive T cells in human leucocyte antigen (HLA)-DQ2.5(+)-associated coeliac disease. *Clin Exp Immunol* 2014;175:305–15.
21. HogenEsch CE, Rosén A, Auricchio R, et al. The Prevent CD Study design: towards new strategies for the prevention of coeliac disease. *Eur J Gastroenterol Hepatol* 2010;22:1424–30.
22. Vriezinga SL, Auricchio R, Bravi E, et al. Randomized feeding intervention in infants at high risk for celiac disease. *N Engl J Med* 2014;371:1304–15.
23. Izzo V, Pinelli M, Tinto N, et al. Improving the estimation of celiac diseases sibling risk by non-HLA genes. *PLoS One* 2011;6:e26920.
24. Monsuur AJ, de Bakker PI, Zhernakova A, et al. Effective detection of human leukocyte antigen risk alleles in celiac disease using tag single nucleotide polymorphisms. *PLoSOne* 2008;3:e2270.
25. Bourgey M, Calcagno G, Tinto N, et al. HLA related genetic risk for coeliac disease. *Gut* 2007;56:1054–9.
26. Fernandez-Jimenez N, Castellanos-Rubio A, Plaza-Izurieta L, et al. Co-regulation and modulation of NF κ B-related genes in celiac disease: uncovered aspects of gut mucosal inflammation. *Hum Mol Genet* 2014;23:1298–310.
27. Plaza-Izurieta L, Fernandez-Jimenez N, Irastorza I, et al. Expression analysis in intestinal mucosa reveals complex relations among genes under the association peaks in celiac disease. *Eur J Hum Genet* 2015;23:1100–5.
28. Jauregi-Miguel A, Fernandez-Jimenez N, Irastorza I, et al. Alteration of tight junction gene expression in celiac disease. *J Pediatr Gastroenterol Nutr* 2014;58:762–7.
29. Zilbauer M, Zellos A, Heuschkel R, et al. Epigenetics in paediatric gastroenterology, hepatology and nutrition—current trends and future perspectives. *J Pediatr Gastroenterol Nutr* 2016;62:521–9.

Post-genomica della crescita e della pubertà

LUIGI GRECO, DONATELLA CIELO

Dipartimento di Scienze Mediche Transazionali, Università Federico II, Napoli

La sorveglianza della crescita del bambino è lo strumento più importante nel bagaglio del pediatra, ma il metodo finora utilizzato, cioè il confronto con i grafici della distribuzione dei percentili dei bambini sani, non considera la peculiarità di crescita e sviluppo di ciascun bambino. Solo recentemente sono iniziati gli studi sulla espressione dei geni che contribuiscono alla crescita delle cellule, dei tessuti, dell'organismo intero. Si affaccia un futuro che renderà obsolete le tabelle dei percentili: sarà infatti possibile, con l'uso dei microchip e le nuove generazioni di analisi molecolare, ottenere un profilo individuale della fase di sviluppo e accrescimento del bambino. Anche per la pubertà non ricorreremo più solo agli stadi di sviluppo del seno o dei peli, avremo bensì un profilo della attivazione dei pathway genomici che iniziano e mantengono la pubertà di quel singolo bambino. È meglio che cominciamo a prepararci: la "precision medicine" calerà massicciamente sul bambino.

GENOMICA DELLA STATURA

Da millenni sappiamo che la crescita in altezza ha una componente genetica dominante in tutte le popolazioni. Io e il mio gemello monozigote ci siamo rincorsi per decenni per vedere chi era più alto, in media 0,5 cm, meno dell'errore di misura (Figura 1). Gli studi longitudinali sulla crescita di gemelli monozigoti mostrano curve di crescita completamente sovrapponibili, con un indice di correlazione tra le due serie di misure di 0,94 vicino all'assoluta identità di 1. Nessun altro carattere ha una componente genetica così forte. Eppure oggi, nel 2016, sappiamo pochissimo sui geni che trasmettono l'eredità della crescita. Lo screening ripetuto del genoma umano ha prodotto risultati molto deludenti per decenni: forse una ventina di polimorfismi di singoli nucleotidi (SNP) di geni candidati sono stati associati all'altezza, ma complessivamente non hanno spiegato più del 3% della trasmissione genetica. È stato calcolato che per spiegare gran parte della ereditarietà sarebbero necessari, a questo passo, decine di migliaia di geni, quasi l'intero genoma!

Un grande passo avanti è stato fatto dal consorzio GIANT (*Genetic Investi-*

POST-GENOMICS GROWTH AND PUBERTY

(*Medico e Bambino 2017;36:31-37*)

Key words

Children, Growth, Gene expression, Molecular auxology, Interactome

Summary

In the last 50 years the surveillance of children's growth was based upon the adoption of the graphs of percentiles of normal growth: there was no chance to explore the pattern of growth of the individual child, more than declaring his normality/abnormality. Recently, the cluster of gene expression related to three phases of growth (infancy, adolescence and adulthood) has been identified. It may now be possible to monitor the growth of the individual child through the next generation analysis of the profile of gene expression of the single child. This information will offer a precise insight of the metabolic processes that regulate the expression of specific cluster of gene. The direction will be from linear or tridimensional measurements to a multidimensional estimation of the growth "regulome".

gation on ANthropometric Traits) che ha analizzato il genoma di più di 200.000 individui, in correlazione con la loro statura da adulti. Hanno identificato 180 loci associati alla statura (e condivisi tra diverse popolazioni): tra questi molti sono polimorfismi (SNP) che controllano l'espressione genica in cis (cis-eQTLs)¹.

Complessivamente questi 180 geni spiegano non più del 15% della ereditarietà dell'altezza, molto meno di quanto spiega l'altezza media dei genitori (40%) proposta da Galton nell'800. Dunque il genoma non basta.

DALLA GENOMICA ALLA POST-GENOMICA

È dunque indispensabile analizzare i geni implicati in percorsi metabolici riconosciuti (*pathway*), nelle varie fasi della crescita. Più che studiare ancora il genoma, è necessario studiare il **trascrittoma**: come funzionano i geni, cominciando a vedere come si esprimono nelle varie fasi della crescita e nei vari individui.

Adam Stevens e la sua squadra di Manchester hanno analizzato delle librerie di espressione genica ottenute



Figura 1. Crescita di gemelli.

dall'analisi di cellule mononucleate del sangue periferico in gruppi modesti (15-30 individui, totale 87) di bambini e ragazzi nelle varie fasi di crescita, e hanno correlato questi dati con l'altezza dell'individuo. Sostanziosi programmi biostatistici hanno permesso di individuare i geni, la cui espressione è correlata con l'altezza raggiunta a quella età².

Nella prima fase del lavoro hanno identificato 688 geni, la cui espressione è correlata all'età del soggetto: in questo gruppo si formano aggregati genici che correlano con la crescita durante l'infanzia, l'età scolare, la pubertà e l'altezza finale. La sorpresa è stata scoprire che questi geni controllano i sistemi molecolari di crescita (*pathway* metabolici) meglio conservati attraverso le specie. La forza di interazione (*interattoma*, *interactome*) all'interno di questi *pathway* è molto energetica nella prima infanzia e decresce progressivamente fino all'età adulta. 175 erano geni già noti, tra i quali IGF2BP3 e STAT5, che regolano la secrezione di ormone della crescita GH. Ma ben il 12% di questi geni sono condizionati da meccanismi epigenetici di

metilazione del DNA che cambiano durante il corso dell'età.

IDENTIFICATI TRE CLUSTER DI ESPRESSIONE GENICA

Attraverso l'analisi incrociata dell'*interattoma* con le banche dati sulla funzione genica sono stati identificati i *pathway* metabolici più rilevanti nelle diverse età:

1. infanzia, < 6 anni: 276 geni;
2. età scolare e adolescenza, 6-17 anni: 207 geni;
3. giovane adulto, > 17 anni: 205 geni.

Nell'infanzia vi è la massima attività di espressione genica, in rapida diminuzione, come mostra la classica curva di velocità di crescita in altezza. Ma la pubertà è caratterizzata, come la curva di crescita, da un rapido ritorno ai sistemi che stimolano la crescita dell'osso, cui si aggiungono i geni specifici di attivazione della pubertà. La giovane età adulta mantiene set di geni, essenziali, ma ne indebolisce molto l'interazione, mostrano un forte "dimagrimento" dell'espressione genica (Figura 2).

QUALI PATHWAY METABOLICI?

Attraverso un complesso incrocio dell'*interattoma* (che comprende le più probabili interazioni tra le espressioni geniche associate alla crescita) con le banche dati che includono i *pathway* metabolici noti, gli Autori hanno assemblato, per ciascuna fase di crescita, dei *minimal essential network* che caratterizzano le fasi della crescita. Questo è infine il messaggio più chiaramente leggibile di questo lavoro.

Conviene dare uno sguardo pittorico a questi percorsi metabolici, che sono altamente conservati nelle specie viventi, tranne che durante la pubertà, che è peculiare della specie umana e non condivisa neanche con i primati.

Ciò che è sorprendente, ma anche rassicurante, è che questi *pathway* metabolici non sono affatto peculiari per la crescita, ma sono proprio quelli comunemente operanti in tutti i processi cellulari, nella proliferazione, nella risposta immunitaria, nel sistema di segnalazione cellulare, nella funzione di ormoni e infine nella patogenesi delle malattie.

Nell'infanzia è ovviamente attivo il sistema di segnalazione del GH con le somatomedine IGF-1, è attivo il *Nerve Growth Factor* (NGF) con la neurotrofina, ma compaiono anche tutti gli altri sistemi: TGFβ^{3,4}, Notch^{5,6}, VEGF^{7,9}, Wnt¹⁰⁻¹².

Durante l'età scolare e puberale si riduce il numero dei *pathway* attivi, rimane l'IGF-1 e gli altri sistemi di proliferazione cellulare.

Nel giovane adulto la gran parte dei processi metabolici descritti non hanno più un'espressione differenziata correlata alla crescita: sono infatti ritornati alle loro funzioni quotidiane di funzionamento di cellule, tessuti e organi.

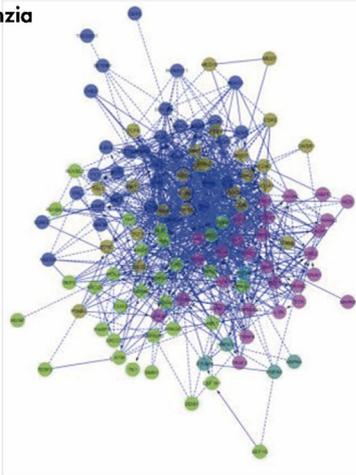
Vi è dunque un sistema centrale di controllo progressivo della funzione (espressione) di geni comuni che agiscono in modo direttamente correlato con l'età del soggetto.

IL RECIPIENTE DEI GLUCOCORTICOIDI

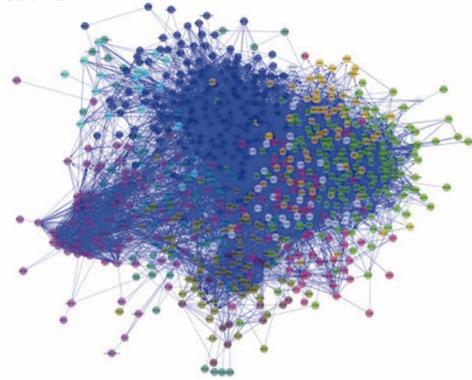
Uno dei fattori centrali di controllo dei *pathway* metabolici descritti (Notch,

MINIMAL ESSENTIAL NETWORK

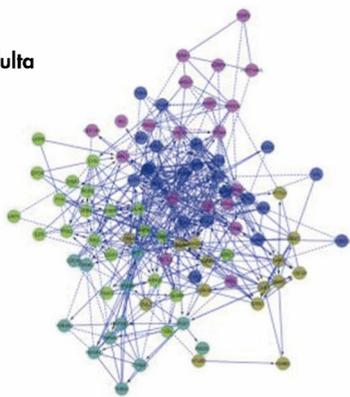
Infanzia



Pubertà



Età adulta



GENE EXPRESSION ASSOCIATED ESSENTIAL PATHWAYS

Pathway	FDR
PDGF Signaling	7.41E-06
VEGF Family Ligand-Receptor Interactions	7.63E-06
VEGF Signaling	7.69E-06
Angiopoietin Signaling	8.62E-06
NGF Signaling	9.09E-06
Melanocyte Development and Pigmentation Signaling	9.35E-06
Renin-Angiotensin Signaling	9.71E-06
NF-kB Signaling	1.28E-05
Notch Signaling	1.45E-05
IGF-1 Signaling	1.89E-05
Erythropoietin Signaling	2.13E-05
Chemokine Signaling	2.27E-05
Wnt/ β -catenin Signaling	3.45E-05
TGF- β Signaling	3.70E-05
BMP Signaling Pathway	5.56E-05
Neurotrophin/TRK Signaling	8.33E-05
Ephrin Receptor Signaling	1.20E-04
Growth Hormone Signaling	4.41E-03
Axonal Guidance Signaling	1.24E-02

Pathway	FDR
IGF-1 Signaling	4.95E-06
Chemokine Signaling	5.24E-06
Actin Cytoskeleton Signaling	5.32E-06
Axonal Guidance Signaling	5.68E-06
Wnt/ β -catenin Signaling	5.92E-06
TGF- β Signaling	7.52E-06
Ephrin Receptor Signaling	8.13E-06
VEGF Signaling	8.47E-06
VEGF Family Ligand-Receptor Interactions	1.01E-05
Renin-Angiotensin Signaling	1.27E-05
NF-kB Signaling	4.17E-05
Notch Signaling	4.76E-05

Pathway	FDR
Ephrin Receptor Signaling	1.72E-05
Erythropoietin Signaling	2.08E-05
Wnt/ β -catenin Signaling	1.25E-04
Chemokine Signaling	1.87E-04
Actin Cytoskeleton Signaling	4.17E-04
Reelin Signaling in Neurons	4.15E-03

Figura 2. Espressione dei geni associati ai pathway correlata con le fasi della crescita umana. I geni di tre cluster correlati con lo sviluppo: ≤ 6 anni (infanzia); $> 6 a \leq 17$ anni (pubertà); > 17 anni (età adulta) (da voce bibliografica 2, modificata).

VEGF, TGF β e Wnt) è, a tutte le età, il recettore dei glucorticoidi (GR) che media la trascrizione genica. Il GR estende la sua azione a molti organi, ma la sua specificità di tessuto è regolata dalla conversione enzimatica da cortisolo a cortisone e da meccanismi epigenetici specifici del tessuto. Questa è la ragione per cui l'eccessiva produzione di corticosteroidi nella sindrome di Cushing o la somministrazione farmacologica interferiscono sempre con i processi di crescita. Studiando l'espressione dei geni che interagiscono con il GR, sono stati identificati 3 set di geni specifici per infanzia, pubertà ed età adulta, che permetteranno di esplorarne le funzioni fisiologiche nelle varie età.

L'ESPRESSIONE GENICA SPECIFICA PER OGNI TESSUTO-ORGANO

Questi dati provengono dall'analisi dell'espressione genica su cellule del sangue periferico, ma Stevens e coll.² hanno fatto un gran lavoro di paragone con l'espressione di 5 altri tessuti umani, verificando che i *minimal essential network* sono sostanzialmente indipendenti dai tessuti, in quanto concernono le funzioni di base di cellule e tessuti. È chiaro che le espressioni geniche della cartilagine sono diverse da quelle del fegato in crescita, ma non tanto quanto si pensi: i geni "postini" agiscono tipicamente con i fattori di trascrizione che permettono l'attivazione specifica di processi legati allo specifico organo. Ma la crescita è in fondo il risultato della sommatoria di questi processi, di cui bisogna comprendere *in primis* i meccanismi di base.

LA PUBERTÀ

La pubertà inizia solo quando i sistemi di rigido controllo trans-sinaptico dei neuroni che producono gonadotropine (GnRH) vengono alleviati. Questa modifica si realizza attraverso la comunicazione trans-sinaptica e quella tra cellule gliali e neuroni. La secrezione di GnRH stimola la produzione di estrogeni dalle ovaie, che ri-stimola la

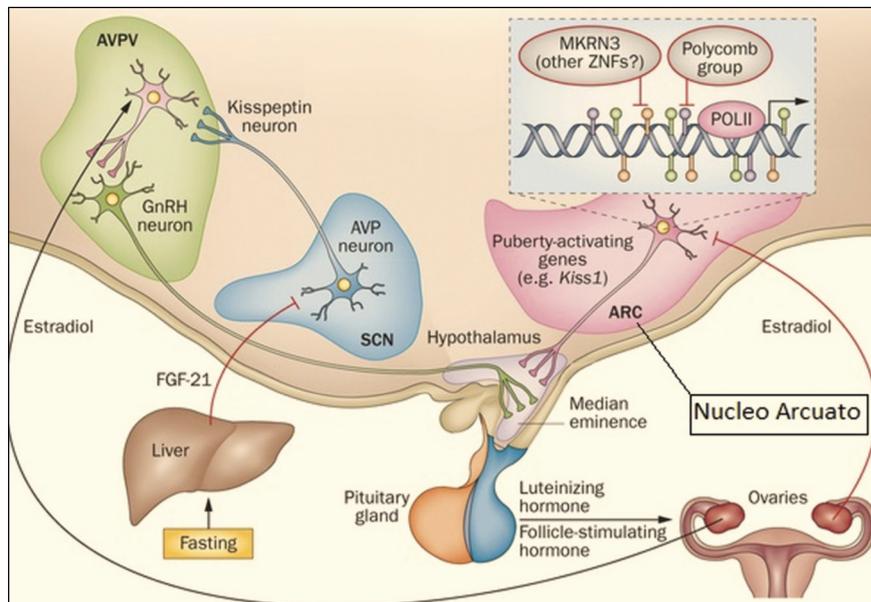


Figura 3. Una nuova visione dei meccanismi centrali e periferici coinvolti nell'inizio della pubertà. Nel nucleo arcuato (ARC) intervengono due sistemi gene-silencing: il complesso Polycomb e MKRN3 (makorin ring finger protein 3). A livello della regione periventricolare anteroventrale il completamento della pubertà è ritardato da FGF-21, il fattore di crescita dei fibroblasti (da voce bibliografica 13, modificata).

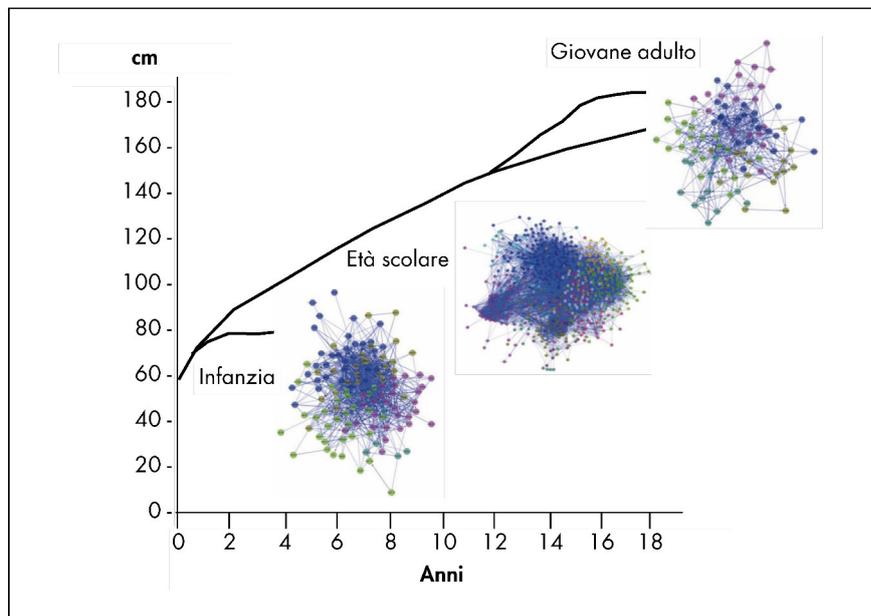


Figura 4. Espressione genica relacionada alla crescita (cm/anno).

produzione di GnRH attraverso i neuroni eccitatori che producono kisspeptina, il peptide neurotrasmettitore specifico. Il sistema di controllo è basato su un bilancio complesso a livello di trascrizione di geni: è un sistema di *switch* elettronici che spengono e poi accendono i geni "postini" che vanno a

scoprire i segmenti di DNA da trascrivere. Sulla trascrizione DNA-RNA messaggero e poi sintesi proteica agiscono i meccanismi epigenetici non controllati dalla genetica.

Ma non basta sollevare il sistema ipotalamico dall'inibizione: è anche necessaria una azione eccitatoria eserci-

tata dai neuroni del nucleo arcuato che producono kisspeptina. Ed è proprio il gene della kisspeptina che viene sottilmente regolato da sistemi di metilazione delle isole CpG del promotore: il gene viene “svolto” dal gomitolo istonico e attivato. L’alterazione dei meccanismi epigenetici di questo gene portano a ritardo puberale e ridotta fertilità.

Il gene MKRN3 dei neuroni del nucleo arcuato serve a fermare l’inizio della pubertà (Figura 3): infatti le mutazioni negative di questo gene sono associate a una forma genetica di pubertà precoce, correlata alla sindrome di Prader-Willi. Questo è il tipico gene che reprime la trascrizione di geni che generano la pubertà. Sappiamo che anche fattori ambientali possono sopprimere la pubertà: una deplezione nutrizionale, il digiuno, anche breve, attiva un fattore di crescita di fibroblasti nel fegato (FGF-21) che va a “spegnere” la produzione di gonadotropine attraverso i neuroni che producono vasopressina¹³.

LO SCATTO PUBERALE

In almeno 2/3 dei bambini si verifica un chiaro aumento della velocità di crescita in altezza all’inizio del processo di sviluppo puberale (Figura 4). Questo scatto puberale ha due componenti: il tempo in cui si verifica e la dimensione dell’incremento in altezza; entrambe condizionano l’altezza finale raggiunta. Il tempo della pubertà ha una forte componente genetica: solo ora cominciamo a comprendere come avviene. I Norvegesi hanno seguito una coorte di 3538 bambini da 0 a 20 anni, rilevando con cura gli incrementi annuali di crescita: hanno analizzato il genoma di questi soggetti per identificare i geni associati allo scatto puberale (l’incremento in altezza tra i 14 anni e l’età adulta)¹⁴.

È stato scoperto così il gene dello scatto puberale: LIN28B, che codifica per una proteina che interagisce con l’RNA. Questo gene agisce diversamente nei due sessi e in tempi diversi, come atteso, ed è fortemente correlato all’altezza finale, ma anche all’insorgere del menarca nella femmina. Sembra essere anche il gene del *tempo puberale*

ben descritto da Tanner. Ma la cosa più interessante è la funzione di questo gene: produce una proteina che si lega all’RNA per interagire e inattivare una famiglia di micro-RNA, che a loro volta controllano una serie di geni associati all’altezza da adulto.

Dunque LIN28B è un gene “intelligente” che influenza una serie di *pathway* metabolici che producono la crescita e la pubertà, non solo attraverso ciò che è scritto nel patrimonio genetico, ma anche attraverso la complessa interazione epigenetica con fattori nutrizionali e ambientali.

In conclusione, la pubertà è regolata da un profilo genetico ereditario sul quale agiscono meccanismi di controllo dell’espressione e traduzione genica di tipo epigenetico (sia ereditari che acquisiti): la pubertà è il risultato del controllo della repressione/attivazione trascrizionale.

IMPLICAZIONI TRANSLAZIONALI?

Dovremmo riscrivere *ex novo* la maggior parte delle peculiarità, ma anche delle anomalie e delle malattie che limitano la crescita dei bambini. Sarà ormai difficile usare il termine *ritardo costituzionale di sviluppo* senza avere un profilo genomico e post-genomico di quello specifico individuo, che permette di comprendere qual è la sua *peculiarità di crescita* rispetto al suo bersaglio genetico.

Anche il concetto di percentile di crescita potrà essere arricchito dalla analisi dell’espressione genica e dunque della funzione di specifici *pathway* metabolici legati alla crescita. Abbiamo ora una serie, complessa ma ragionevole, di bio-marcatore metabolici che caratterizzano le diverse fasi di crescita.

Potremmo forse pensare a un *growth-chip* che permetta di produrre una “mappa della crescita” attraverso l’analisi dell’espressione dei 688 geni associati e identificare i *pathway* che determinano la scarsa crescita nel bambino. Questo ci permetterebbe di individuare quale meccanismo a livello molecolare è coinvolto ad esempio nell’infante prematuro o nel lattante

che “perde centili” permettendo un intervento mirato, magari attraverso la nutrizione.

Certamente gli endocrinologi dovranno considerare la peculiarità del singolo paziente nel gestire le anomalie della crescita: basti pensare che la stima dell’età ossea, oggi strumento preziosissimo di diagnostica differenziale, potrà essere presto sostituita dal profilo metabolico reale della crescita dell’osso.

COME USEREMO I PERCENTILI NEL PROSSIMO FUTURO?

I fattori epigenetici

In pratica ciascuna fase della crescita non dipende *sic et simpliciter* dal patrimonio genetico, che è presente dal giorno del concepimento fino alla morte, bensì dal complesso equilibrio tra i *cluster* di geni espressi e di quelli invece nascosti, attraverso meccanismi di controllo dei fattori di trascrizione dei geni.

Ma su questo equilibrio espressione/repressione agiscono anche una serie di fattori individuali e ambientali, complessivamente ora definiti *epigenetici*.

Una lettura dinamica dei percentili

Sappiamo che i grafici dei percentili sono una immagine “sommatoria” dei profili di crescita di migliaia di bambini che non hanno patologie note. La statura raggiunta a una determinata età è la risultante finale dell’eredità ricevuta dai genitori e dell’espressione genica in quella determinata fase della vita.

Per prudenza diciamo che un bambino con statura sotto al 3° centile è “corto” e va indagato. Ma sappiamo bene che il 3% di 10 milioni di bambini raccoglie almeno 300.000 individui in Italia, e pochi sono affetti da patologie (con un rischio relativo +++). Similmente tra il 3° e il 50° centile potrebbero esserci molte alterazioni della crescita con un basso rischio (+), ma in 5 milioni di individui: dunque potremmo avere più soggetti che, con l’uso acritico dei percentili grafici, ora definiremmo “normali” quando in realtà potrebbero non esserlo.

MESSAGGI CHIAVE

- ❑ La crescita è regolata dalla progressiva attivazione, ed espressione, di *cluster* di geni dedicati alla replicazione cellulare.
- ❑ Attualmente sono stati identificati tre profili genomici della crescita dell'infanzia, dell'adolescenza e dell'età adulta.
- ❑ Sarà possibile, nel prossimo futuro, avere una stima dell'attivazione dei *cluster* genetici della crescita nel singolo bambino.
- ❑ Possiamo prevedere che avremo una stima "più precisa" dello stadio evolutivo dell'accrescimento, non più basato su una misura lineare (la statura) o tridimensionale (il peso), bensì di tipo molecolare.

Guardiamo, ad esempio, la celiachia, che si accompagna ad arresto della statura: l'1-2% della popolazione è affetta da celiachia (su 10 milioni di bambini circa 150.000); di questi celiaci meno del 5% si presenta con "bassa statura" (circa 7500), ma almeno il 30% ha un disturbo del suo normale ritmo di crescita (circa 45.000), che non viene certamente colto dal punto segnato sulla curva dei percentili.

Già ora sorvegliamo i soggetti con statura superiore al 3° centile che hanno un alto potenziale genetico, o quelli che "scendono" di almeno 2 curve di percentili rispetto alla loro curva "normale".

In futuro potremo avere, per ciascun bambino, in una data età e fase di crescita, il suo personale profilo di espressione genica, dunque potremo sapere quali dei suoi geni è attivo e quale è invece represso e questo complesso profilo del trascrittoma individuale potrà essere correlato alla sua posizione nelle curve di crescita. Il problema è che l'espressione genica, diversamente dal genoma, è specifica per ciascun tessuto, ma a oggi è possibile ottenere RNA anche da campioni semplici, come da sangue periferico o da cellule della saliva e quindi non sarà impossibile creare profili di espressione genica individuali con costi contenuti.

Certamente vi sarà da fare un lavoro complesso di elaborazione bioinformatica, ma anche quando abbiamo tracciato

UN CASO CLINICO DA NEW ENGLAND JOURNAL OF MEDICINE DI LUGLIO 2015

Gli *insulin growth factor* 1 e 2 (IGF-1 e IGF-2) regolano la crescita somatica e la proliferazione cellulare attraverso il legame con il recettore dell'IGF-1 (IGF-1R). Entrambi sono espressi durante lo sviluppo fetale, in particolare l'IGF-2 sembra avere un ruolo predominante sullo sviluppo embrionale mentre l'IGF-1 agisce dopo la nascita. Mutazioni del gene IGF-1 e IGF-1R sono implicate in gravi ritardi dello sviluppo intrauterino e post-natale. Inoltre in topi è stato osservato che la delezione dell'allele paterno dell'IGF-2 causa ritardo prenatale e influenza anche lo sviluppo post-natale. Al contrario, la delezione dello stesso allele materno non ha nessun effetto sulla crescita. L'influenza sesso-specifica è dovuta al fatto che il gene dell'IGF-2 è un gene *imprinted*. L'allele paterno, a differenza di quello materno, è espresso in placenta e in molti altri tessuti; al contrario, quello materno viene espresso particolarmente nel fegato. È noto che una ridotta espressione di IGF-2 nel fegato è coinvolta nella patogenesi della sindrome di Silver-Russell.

Il caso clinico presentato su *NEJM* riguarda 4 soggetti, appartenenti a una famiglia multigenerazionale, che hanno ereditato il gene IGF-2 del padre portatore di una mutazione nonsense che riduce l'espressione del gene del 30% rispetto al normale. Tutti hanno avuto un severo ritardo di crescita intrauterina e un fenotipo post-natale tipo Silver-Russell nonostante presentassero normali livelli sierici della proteina IGF-2 prodotta dal fegato dai geni ereditati dalla madre. I figli che invece hanno ereditato solo il gene della madre, e non quello del padre, non hanno alcuna alterazione della crescita. Questa è una prova convincente del *fetal programming*, ossia della programmazione prenatale sulla crescita futura, infatti il gene mutato del padre essendo espresso nella placenta ha condizionato la crescita pre-natale e quella post-natale¹⁵.

IL GENE "SARDO"

Francesco Cucca, pediatra genetista di Cagliari, ha occupato il numero del 15 settembre 2015 di *Nature Genetics*¹⁶ presentando una serie di lavori sui polimorfismi della popolazione sarda nel distretto di Lanusei, che mostrano un preponderante 'effetto fondatore' poiché discendono da pochi soggetti 'fondatori' della comunità. Per la prima volta al mondo sono stati capaci di identificare la mutazione di un gene rarissimo in tutte le popolazioni (< 0,01%), tranne in quella sarda (0,87%), che provoca una alterazione funzionale del recettore dell'ormone della crescita (GHR): questa mutazione riduce la statura media dei portatori di ben 4,5 cm, portando a una altezza adulta di 168,5 cm. Hanno inoltre isolato una seconda mutazione di un gene che controlla un canale di trasporto del potassio (KCNQ1) che riduce l'altezza media di 1,83 cm. Questo è il primo studio che documenta nella specie umana, non affetta da patologia nota, geni che hanno un effetto tanto significativo sulla statura.

le curve dei percentili con il prof. Tanner è stato fatto un lavoro che, negli anni '60-'70, sembrava molto complesso.

In conclusione, potremo seguire il profilo di crescita di un determinato individuo grazie allo studio dei *pathway* molecolari. Oggi nel soggetto con predisposizione alla celiachia ci dedichiamo all'analisi scrupolosa, ogni 3 mesi, della sua velocità di crescita, per identificare eventuali decelerazioni prima che compaiano i sintomi franchi della malattia.

Presto potremo, attraverso un semplice prelievo di saliva, avere la dinamica molecolare individuale della crescita identificando quali geni funzionano e quali sono repressi.

La scienza ci tormenta con le sue complessità, ma ci apre ancora alle speranze.

Indirizzo per corrispondenza:

Donatella Cielo
e-mail: dona_cielo@yahoo.it

Bibliografia

1. Lango Allen H, Estrada K, Lettre G, et al. Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature* 2010;467(7317):832-8.
2. Stevens A, Hanson D, Whatmore A, Destenaves B, Chatelain P, Clayton P. Human growth is associated with distinct patterns of gene expression in evolutionarily conserved networks. *BMC Genomics* 2013;13:547.
3. Augustyniak E, Trzeciak T, Richter M, Kaczmarczyk J, Suchorska W. The role of

- growth factors in stem cell-directed chondrogenesis: a real hope for damaged cartilage regeneration. *Int Orthop* 2015;39(5):995-1003.
4. Gorelik L, Flavell RA. Transforming growth factor-beta in T-cell biology. *Nat Rev Immunol* 2002;2(1):46-5.
5. Chau JF, Leong WF, Li B. Signaling pathways governing osteoblast proliferation, differentiation and function. *Histol Histopathol* 2009;24:1593-606.
6. Engin F, Yao Z, Yang T. Dimorphic effects of Notch signaling in bone homeostasis. *Nat Med* 2008;14:299-305.
7. Gerber HP, Vu TH, Ryan AM, Kowalski J, Werb Z, Ferrara N. VEGF couples hypertrophic cartilage remodeling, ossification and angiogenesis during endochondral bone formation. *Nat Med* 1999;5:623-8.
8. Yao S, Liu D, Pan F, et al. Effect of vascular endothelial growth factor on RANK gene expression in osteoclasts precursors and on osteoclastogenesis. *Arch Oral Biol* 2006;51(7):596-602.
9. Zou L, Lai H, Zhou Q, Xiao F. Lasting controversy on ranibizumab and bevacizumab. *Theranostics* 2011;1:395-402.
10. Miller JR. The Wnts. *Genome Biol* 2013; 1-15.
11. Cadigan KM, Liu YI. Wnt signaling: complexity at the surface. *J Cell Sci* 2006;119:395-402.
12. Westendorf JJ, Kahler RA, Schroeder TM. Wnt signaling in osteoblasts and bone diseases. *Gene* 2004;341:19-39.
13. Ojeda SR, Lomniczi A. Puberty in 2013: Unravelling the mystery of puberty. *Nat Rev Endocrinol* 2014;10(2):67-9.
14. Lettre G. Recent progress in the study of the genetics of height. *Hum Genet* 2011;129(5):465-72.
15. Begemann M, Zirn B, Santen G, et al. Paternally-Inherited IGF-2 Mutation and Growth Restriction. *N Engl J Med* 2015;373(4):349-56.
16. Zoledziwska M, Sidore C, Chiang CW, et al. Height-reducing variants and selection for short stature in Sardinia. *Nat Genet* 2015; 47(11):1352-6.
-



FC FELLOWSHIPS - CALL FOR PROPOSALS 2017

**COMUNICAZIONE UFFICIALE ASSEGNAZIONE FONDI PER LA BORSA DI
STUDIO FC TRIENNALE**

Progetto Fellowship Cielo N. 007_FC_2017

OGGETTO

Bando FC Fellowships 2017

**COMUNICAZIONE UFFICIALE DI ASSEGNAZIONE FONDI PER LA BORSA DI
STUDIO FC TRIENNALE AL PROGETTO FELLOWSHIP Cielo N. 007_FC_2017**

Gent.ma Dott.ssa Cielo

Il **Progetto Fellowship N. 007_FC_2017** da Lei presentato al **Bando FC Fellowships 2017** rientra tra i **progetti Fellowship** che la **Fondazione Celiachia (FC)** ha **deliberato di finanziare** in funzione delle graduatorie di merito ottenute al termine del processo di valutazione Peer Review.

SCHEDA PROGETTO FELLOWSHIP TRIENNALE Cielo N. 007_FC_2017 – Estratto dell'Application Form del Candidato

Candidato Borsista FC	Capo Laboratorio	Titolo Progetto Fellowship	Area di Ricerca
Cielo Donatella	Greco Luigi	Long non-coding RNA Effects on the Development of Celiac Disease	CD - Genetics

Meccanismi di regolazione dei geni candidati nella malattia celiaca

Background

Gli studi di associazione genetica hanno identificato 112 polimorfismi a singolo nucleotide (SNPs) associati alla malattia celiaca (MC) di cui il 90% si trovano in regioni non codificanti del genoma, suggerendo che la maggior parte degli SNP associati alla MC hanno un ruolo potenzialmente regolatorio. Lo studio dei meccanismi di regolazione epigenetica potrebbe aiutare a completare le informazioni riguardo all'ereditarietà della MC. Nel corso degli ultimi anni, è stato scoperto che gli RNA non codificanti (ncRNAs), definiti come piccole molecole di RNA che sono trascritte ma non tradotte in proteine, rappresentano la maggior parte dei trascritti umani. Inoltre, è anche noto che essi sono fortemente coinvolti nella regolazione post-trascrizionale dell'espressione genica che quindi non è controllata solo da fattori di trascrizione, come si pensava. Gli RNA non codificanti che influenzano i processi epigenetici possono essere raggruppati in "short ncRNAs" e "long ncRNAs", questi ultimi saranno oggetto del nostro studio.

Obiettivi

Lo scopo dello studio sarà valutare l'effetto e la localizzazione di 7 SNPs presenti nella sequenza di 7 lncRNA associati alla MC che ne influenzano la loro funzionalità, in sezioni fissate in formalina (FFPE) di biopsie intestinali di pazienti celiaci e in cellule intestinali epiteliali e non epiteliali.

Successivamente saranno esplorati i geni e i pathways molecolari regolati dai 7 lncRNA, nel contesto cellulare più adeguato in base alla loro localizzazione e/o alla loro specifica funzione molecolare.

Infine, analizzeremo l'effetto dei 7 SNP dei lncRNA direttamente in linee cellulari ingegnerizzate, per dimostrare il rapporto causa effetto tra la presenza del singolo polimorfismo e l'alterazione dei lncRNA e dei relativi pathways molecolari.

Metodi

Per la localizzazione e quantizzazione dei lncRNAs regolati da SNP associati alla celiachia, selezioneremo le biopsie intestinali in paraffina di 20 soggetti celiaci, 20 soggetti controllo e 20 pazienti con malattie infiammatorie croniche intestinali (MICI) della nostra banca dati. Osserveremo la localizzazione di lncRNA all'interno di sezioni paraffinate combinando tecniche di ibridazione in situ (ISH) ed immunohistochimica (IHC) sullo stesso campione in una popolazione cellulare specifica. Sui lncRNA che risulteranno deregolati nella popolazione di pazienti celiaci effettueremo un'analisi di predizione bioinformatica dei geni target e dei pathways molecolari regolati dai singoli lncRNA. Per la conferma dei dati sulle cellule intestinali isolate da biopsia intestinale, saranno reclutati 20 pazienti celiaci, 20 con malattie infiammatorie croniche e 20 pazienti controllo affetti da disturbi funzionali (CTR), da cui isoleremo le cellule epiteliali e non epiteliali per analizzare l'espressione dei lncRNA e dei geni da essi regolati.

Risultati attesi

I risultati di questo studio potrebbero essere la prima dimostrazione del diretto contributo dei meccanismi di regolazione epigenetica nella patogenesi della MC e potrebbero fornire un contributo sostanziale per elucidare i meccanismi alla base della cosiddetta "ereditarietà mancante" della MC.

L'impatto dello studio sulla conoscenza e/o sul trattamento della malattia celiaca in termini di innovazione e originalità rispetto a quanto già noto

L' esplorazione dell'intero Genoma Umano in decine di migliaia di soggetti celiaci ha contribuito a chiarire solo una piccola parte della ereditarietà della celiachia. La maggior parte dei polimorfismi trovati in associazione con la malattia non alterano la produzione di proteine, ma, con ogni probabilità, modificano il profilo di funzionamento di geni comuni. Per la prima volta esploreremo gli RNA non codificanti che partono dal nucleo verso il citoplasma della cellula per regolare la funzione di geni comuni. Questo tipo di studio può essere la chiave di volta per conoscere le modificazioni indotte dal glutine in soggetti geneticamente predisposti. La maggior parte degli studi precedenti si è focalizzato sullo studio della genetica classica mentre l'innovazione del nostro studio sta proprio nell'analisi del "reguloma" del celiaco, ossia elementi nuovi mai studiati, che ci potrebbero aiutare a chiarire i meccanismi che sono alla base della MC.

I benefici per il paziente nel medio o lungo termine, nel caso il programma di studi raggiungesse gli obiettivi prefissati

Non vi sono promesse immediate di nuove terapie, ma il progresso nella conoscenza dei meccanismi di regolazione dei geni comuni, coinvolti nella risposta al glutine solo nei soggetti celiaci e non in altri, è la premessa fondamentale per evitare false illusioni terapeutiche e migliorare significativamente la possibilità di modulare l' anomala risposta immune al glutine.