

# Interazione geni-ambiente in relazione al rischio cardiovascolare

Sergio Coccozza, Olga Vaccaro

L'identificazione della componente genetica delle malattie complesse viene unanimamente riconosciuta essere una delle principali sfide della medicina dei prossimi anni. La possibilità di avere test genetici che predicano la suscettibilità al cancro, all'infarto o al morbo di Alzheimer, avrebbe un enorme impatto in termini di salute pubblica. L'identificazione di questi geni pone, però, problemi tecnici e metodologici di non piccola rilevanza. In questo lavoro cercheremo di descrivere, almeno parzialmente, lo "stato dell'arte" sull'argomento, rappresentando i principali contesti culturali, le problematiche aperte e le loro possibili soluzioni.

---

## Sergio Coccozza

Dipartimento di  
Biologia e Patologia  
Cellulare e Molecolare  
"A. Califano"  
IEOS, CNR, Napoli

## Olga Vaccaro

Dipartimento di  
Medicina Clinica  
e Sperimentale,  
Università "Federico II",  
Napoli

## Introduzione

Le malattie ereditarie tendono a essere generalmente classificate in due grandi gruppi. Il primo è rappresentato dalle cosiddette "malattie mendeliane". Esse tendono generalmente a essere rare con frequenze nelle popolazioni inferiori allo 0,05%. Al secondo gruppo appartengono le cosiddette "malattie multifattoriali", anche definite come "malattie complesse". Queste patologie rappresentano un pericolo molto più grave per la salute pubblica poiché possono presentare frequenze pari o superiori all'1%. Infatti, a questa categoria vanno ascritte le patologie più frequenti nel mondo occidentale come il diabete, la schizofrenia e l'obesità. Un gran numero di patologie del sistema cardiovascolare sono presenti in questo gruppo inclusi l'infarto, l'atero-

sclerosi e l'ictus. L'identificazione dei geni responsabili di queste malattie viene considerata come una delle grandi sfide della medicina (1). Una caratteristica comune a tutte le malattie complesse è il fatto di presentare un'aumentata "familiarità" senza però poter riconoscere pattern di trasmissione mendeliana. Classicamente, infatti, in una famiglia di un paziente sono presenti spesso altri casi tra i parenti, senza però riconoscere caratteristiche riconducibili a una dominanza, o recessività o a un tratto legato al sesso. Un'altra caratteristica importante delle malattie complesse è che esse sono determinate dall'intricata interazione di diversi fattori genetici e diversi fattori ambientali. Questi fattori, siano essi genetici o ambientali, vanno sotto il nome di "fattori di suscettibilità". Un altro aspetto da sottolineare è che questi fat-

tori agiscono in differenti combinazioni da individuo a individuo. In altre parole per ottenere la stessa suscettibilità in un individuo possono cooperare alcuni fattori, mentre un differente set di fattori può essere implicato in un altro individuo. In comune i due set di fattori devono avere soltanto il fatto di determinare in entrambi i casi il superamento di una “soglia” di suscettibilità, oltre la quale si determina la malattia.

## La genetica dei tratti complessi

Volendosi soffermare, in questa fase, solo sugli aspetti genetici e tralasciando, almeno per adesso, i pur importantissimi fattori ambientali, quello che si nota immediatamente è che la sola componente genetica presenta un notevole grado di complessità (2). Infatti ancora adesso esistono due teorie in conflitto su questa questione. La prima va sotto la denominazione di *common disease/rare allele* (CD/RA). Secondo questa teoria la componente genetica di una malattia complessa sarebbe costituita da un grande numero di differenti mutazioni, ognuna con una frequenza molto bassa e una penetranza alta. Tali mutazioni sarebbero avvenute relativamente di recente (da cui la bassa frequenza) e in ogni caso prima della divergenza delle popolazioni umane. Una conseguenza di questa teoria è che ci si aspetta un’altissima eterogeneità nei geni di suscettibilità delle malattie complesse. La seconda teoria va sotto il nome di *common disease/common variant* (CD/CV). Secondo questa teoria la componente genetica di una malattia complessa sarebbe costituita da un numero relativamente limitato di mutazioni a bassa penetranza. Ognuna di queste varianti sarebbe abbastanza frequente nella popolazione normale, ma il combinarsi di più forme alleliche “negative” sarebbe la base della suscettibilità alla malattia. In questo contesto teorico gli alleli responsabili si sarebbero generati prima della dispersione degli umani e sarebbero stati, almeno in alcuni casi, selezionati positivamente in ambienti antichi. Il cosiddetto “genotipo frugale” (3) è uno degli esempi più conosciuti di questa teoria. Questo genotipo, costituito dall’insieme di forme alleliche di geni differenti, aumenterebbe la capacità dell’uomo di accumulare energia. La capacità di resistere a periodi di restrizione alimentare potrebbe essere stata selezionata positivamente nel Paleolitico, dove la disponibilità di cibo era molto modesta. Questo stesso genotipo però, in epoca moderna dove le risorse energetiche sono facilmente disponibili, predisporrebbe a patologie come l’o-

besità e il diabete. Numerose evidenze si sono accumulate nel tempo sia a favore della ipotesi CD/RA sia di quella CD/CV. Tra le prime vanno ricordate le mutazioni nei geni *breast cancer* (BRCA) 1 e 2 nel cancro dell’ovaio e della mammella e dei geni dell’ $\alpha$ -sinucleina nel Parkinson. Anche il *maturity onset diabetes in young* (MODY) o il cancro del colon, con il loro limitato numero di geni ad alta penetranza possono essere considerati esempi autorevoli. Per il versante CD/CV, d’altro canto, si possono citare le varianti di ApoE nel morbo di Alzheimer, le varianti del fattore V Leiden nella trombosi oppure le varianti del recettore *CC-chemokine receptor 5* (CCR5) nella resistenza all’AIDS. La **figura 1** riassume le principali caratteristiche delle due ipotesi. Molti altri esempi si potrebbero fare nell’un campo e nell’altro, tanto che l’ipotesi che attualmente sembra la più accettata risulta in una mediazione tra le due. Secondo questa ipotesi “unificante” la maggior parte delle malattie complesse sarebbe dovuta a varianti geniche ad alta frequenza e bassa penetranza, che agirebbero in poligenia. Un minoritario numero di pazienti riconoscerebbe invece una predisposizione genetica in mutazioni rare ad alta penetranza. Un’ulteriore espansione dell’ipotesi CD/CV va sotto il nome di *common variant/multiple disease* (CV/MD). Secondo questa ipotesi esisterebbero varianti genetiche che associate in un certo modo darebbero una predisposizione per una certa malattia. Le stesse varianti, associate in modo differente potrebbero dare altre malattie.

*L’identificazione della componente genetica delle malattie complesse viene unanimemente riconosciuta come una delle principali sfide della medicina dei prossimi anni*

*Esistono due teorie in conflitto su questa questione. La prima va sotto la denominazione di common disease/rare allele (CD/RA). La seconda teoria va sotto il nome di common disease/common variant (CD/CV)*

**Figura 1**  
**Principali caratteristiche dell’ipotesi CD/RA e CD/CV**



### Common disease/rare allele

La maggior parte delle mutazioni che causano malattie complesse sono avvenute recentemente, dopo la divergenza delle popolazioni. La conseguenza è l’attesa di una forte eterogeneità nei geni di suscettibilità per le malattie complesse.



### Common disease/common variant

Varianti alleliche esistenti prima della dispersione degli umani sulla terra o varianti alleliche sottoposte a pressione selettiva rappresentano una proporzione significativa degli alleli di suscettibilità per le malattie complesse. La conseguenza è l’attesa di una minore eterogeneità nei geni di suscettibilità per le malattie complesse.



Questa ipotesi trova fondamento soprattutto per quelle patologie che mostrano comorbidità, ad esempio asma e dermatite atopica. Sono state riconosciute varianti alleliche responsabili della suscettibilità all'asma, altre della dermatite atopica e altre ancora di entrambe le forme patologiche.

## Strumenti di analisi genetica

Negli ultimi anni i geni responsabili di più di 1500 malattie genetiche umane sono stati isolati e riconosciuti. La stragrande maggioranza di queste malattie sono classiche malattie monogeniche mendeliane. Questo grande risultato della ricerca genetica si deve sostanzialmente a una metodica chiamata "clonaggio posizionale" (4). Con questo termine si indica l'isolamento di un gene responsabile di malattia basandosi esclusivamente sulla sua posizione sul cromosoma, senza alcuna conoscenza "a priori" della sua struttura o funzione. Il clonaggio posizionale comincia con l'analisi del *linkage*. Due loci genetici si dicono in *linkage* quando tendono a cosegregare nelle generazioni più frequentemente di quanto ci sia da aspettare in accordo con la seconda legge di Mendel. Uno dei loci può essere il locus di una malattia mendeliana con una posizione sconosciuta nel genoma, mentre l'altro può essere un marcatore genetico di posizione conosciuta. Quindi un *linkage* significativo localizza una malattia su un segmento di cromosoma che può essere ulteriormente studiato fino ad arrivare al gene in esso mutato. Purtroppo l'analisi di *linkage* non può essere facilmente usata nell'ambito delle malattie complesse. Questo per vari ordini di motivi di cui citeremo solo i due principali. Il primo è che un individuo può essere malato per gli effetti non di un solo, ma di più geni localizzati su cromosomi differenti. Il secondo motivo è che l'analisi di *linkage* prevede, in linea generale, che la modalità di trasmissione (dominante, recessivo, *x-linked*) sia conosciuta. Questa caratteristica fa annoverare l'analisi di *linkage* tra quelle "parametriche". Una possibile alternativa è quella di un approccio attraverso il *linkage disequilibrium*. Senza scendere in dettaglio queste metodiche pur mantenendo il primo dei limiti, sono non parametriche e quindi attivamente sperimentate nel campo della ricerca dei geni causa di malattie complesse. La trattazione di questi approcci però non rientra negli scopi di questo lavoro. Accanto a questi metodi, puramente "genetici" nella loro struttura concettuale, hanno sempre convissuto metodi di più intuitiva concezione. Il più conosciuto tra essi è quello del

"gene candidato". Questo approccio prevede una fase in cui si candida un gene la cui alterazione può essere responsabile della malattia sulla base della sua funzione. Ad esempio si candida il gene dell'insulina come potenzialmente responsabile, quando mutato, del diabete mellito. Come secondo step si procede poi a cercare variazioni del gene candidato che siano presenti nella popolazione degli affetti e assenti nella popolazione dei controlli. I principali criteri di eleggibilità sono:

1. conoscenza di una variazione del prodotto genico con significato funzionale;
2. omologia funzionale con altri geni di cui è stato dimostrato il coinvolgimento in patologie analoghe;
3. coinvolgimento in patologie analoghe in altre specie;
4. ruolo in vie metaboliche la cui alterazione è stata dimostrata nella patologia in esame;
5. ruolo in vie metaboliche la cui alterazione è stata ipotizzata nella patologia in esame.

L'analisi di geni candidati ha fatto la storia della genetica precedente al clonaggio posizionale. Gli esempi più noti sono l'identificazione dei geni responsabili delle emoglobinopatie. Essa è poi diventata di marginale importanza, per le malattie mendeliane, con l'avvento del clonaggio posizionale.

Tale approccio è tornato, in una versione leggermente modificata, dopo il completamento della sequenza del genoma umano, proprio come metodo di approccio alle malattie complesse. Le ragioni di questo ritorno vanno ricercate in nuove possibilità conseguenti a quel sequenziamento (5). In estrema sintesi queste possibilità risiedono nella scoperta dei *single nucleotide polymorphisms* (SNP), nella disponibilità di metodi di analisi di varianti alleliche con elevata produttività e di nuovi strumenti di analisi statistica. Per quanto riguarda le SNP si ricorda che con questo termine si intende definire delle normali differenze genetiche tra gli individui (polimorfismi) la cui base molecolare consista in una variazione di una singola coppia di basi. La scoperta di questi polimorfismi può essere considerata una vera e propria diretta conseguenza dello sforzo di sequenziamento che si è fatto durante il progetto *Genoma Umano*. Variazioni genetiche tra individui, infatti, sono venute fuori dal sequenziamento parallelo di differenti individui. Altre SNP sono state scoperte quando si è confrontata la sequenza di acido ribonucleico (RNA) (*expressed tag sequences*, EST) con altre provenienti da altre persone o con quello dell'acido deossiribonucleico (DNA) genomico. Ulteriori SNP sono

---

*Negli ultimi anni sono stati identificati i geni responsabili di più di 1500 malattie genetiche umane, la stragrande maggioranza delle quali sono classiche malattie monogeniche mendeliane. Nel caso delle malattie complesse un individuo può essere malato per gli effetti non di uno solo, ma di più geni localizzati su cromosomi differenti*

state scoperte sequenziando le regioni vicine a geni causa di malattia. Se a questo si aggiunge lo sforzo di molti ricercatori di cercare attivamente queste variazioni sequenziando soggetti normali, ben si comprende come le SNP rappresentino, al momento, la maggiore fonte di conoscenza sulla variabilità genetica nell'uomo. Per dare alcuni numeri uno dei database più autorevoli di SNP, il dbSNP del *National Center for Biotechnology Information* (NCBI), nella versione 124 datata al momento in cui questo articolo viene scritto, contiene più di 5 milioni di queste variazioni. Siamo certi che questo numero sarà aumentato al momento in cui ci state leggendo. Le SNP possono capitare in differenti parti del genoma e, quindi, avere differenti effetti. Se queste variazioni capitano nelle regioni del gene destinate alla codifica delle proteine, due sono i possibili esiti: esse possono determinare un cambio di aminoacido oppure no. Nel primo caso si parla di SNP "non-sinonime" nel secondo caso si parla di SNP "sinonime". Se le SNP capitano invece nelle regioni del DNA destinate alla regolazione dei geni si parla di SNP "regolative". Se, infine, esse sono localizzate nelle regioni introniche, cioè quelle che separano le varie regioni codificanti e il cui significato funzionale è ancora in discussione, si parla di SNP "introniche". Sebbene tutte le SNP possano avere un impatto funzionale, è maggiormente intuitivo che il possedere proteine con un aminoacido differente possa rivelare una differenza funzionale tra gli individui. Si calcola in circa 40.000 il numero di questo tipo di SNP "non-sinonime". Abbiamo, quindi, un potente strumento per candidare variazioni genetiche potenzialmente in grado di determinare suscettibilità alle malattie complesse. Non rientra tra gli scopi di questo scritto quello di trattare in dettaglio gli aspetti tecnologici che ci permettono, attualmente, di determinare il genotipo di queste variazioni in moltissimi individui in un tempo relativamente ristretto. Va solo sottolineato come queste metodiche permettano, attualmente, di analizzare numerosità impensabili fino a qualche anno fa. Si deve altresì sottolineare come l'innovazione tecnologica prometta sempre maggiori riduzioni di tempi e di costi. Questo argomento ha molte implicazioni per il discorso sul campionamento che faremo nelle parti successive.

## Nuove metodiche di analisi statistiche

### I problemi da affrontare

L'analisi di associazione che si effettua applican-

do la strategia del gene candidato è dal punto di vista squisitamente teorico di grande semplicità. Infatti non si tratterebbe altro che valutare la differenza di frequenza allelica di un gene tra una popolazione di individui affetti e una di controllo. Se quella variante allelica fosse maggiormente presente tra gli affetti in maniera statisticamente significativa, si potrebbe suggerire un suo ruolo nella suscettibilità alla malattia. Purtroppo le cose non sono così semplici. Dei molti problemi metodologici legati a questo tipo di studio tratteremo, in questa sede, solo di alcuni, e in particolare del problema del test multiplo, di quello della stratificazione, di quello dell'epistasi e di quello della dimensionalità.

### Il problema del test multiplo

Quando diversi geni candidati vengono testati c'è una probabilità che la soglia statistica venga superata per caso. Questa probabilità aumenta con l'aumentare dei geni che vengono analizzati. Esistono, ovviamente, metodi che servono a modificare la soglia di significatività in relazione al numero di ipotesi valutate. Il più conosciuto di questi metodi è quello di Bonferroni (6). Un'alternativa a questi metodi si basa su tecniche che utilizzano il computer. Sempre più attenzione, in questo contesto, stanno ricevendo le tecniche di permutazione, attraverso le quali la tabella originale dei dati viene permutata automaticamente un certo numero di volte e per ogni nuova configurazione viene ricalcolata la statistica. Si ottiene così una valutazione empirica della significatività. Queste metodiche liberano il ricercatore dalla necessità di focalizzarsi su misure statistiche le cui proprietà teoriche possono, alcune volte, essere di non intuitiva comprensione.

### Il problema della stratificazione

In accordo con quanto detto precedentemente sulla natura della componente genetica delle malattie complesse, ci si attende un certo grado di eterogeneità genetica tra gli individui affetti. In altre parole ci si aspetta che non tutti abbiano gli stessi geni di suscettibilità, ma che lo stato di natura sia che una popolazione di affetti in realtà sia composta da differenti gruppi con differente componente genetica di suscettibilità. È ovvio che non siamo in grado di riconoscere questi gruppi, ma è altresì necessario fare dei tentativi per riuscirci. Rischieremmo, infatti, di concludere che un gene candidato non è coinvolto solo perché esso è responsabile della suscettibilità solo di un sottogruppo di pazienti. È necessario, quindi, stratificare il nostro campione nella

---

*Gli strumenti per l'analisi della componente genetica delle malattie complesse sono tutt'ora in evoluzione, giacché nasce la necessità di associare a una malattia non più una sola forma allelica, ma la presenza contemporanea di più forme alleliche. Quando si pensa a un effetto poligenico, la cosa più intuitiva cui si pensa è un effetto additivo, ma questo non è lo scenario più probabile*

maniera più razionale possibile, per ridurlo a sottogruppi, per lo meno ipoteticamente, più omogenei geneticamente. Vari approcci sono stati usati per raggiungere questo obiettivo. Il più tradizionale è quello di cercare sottogruppi della stessa malattia omogenei per fenotipo clinico (esordio, gravità, presenza di complicazioni, ecc.) (7). Anche il fenotipo biochimico, ad esempio particolari valori di un parametro ematochimico, può essere utilizzato. Un altro approccio, invece, può sfruttare differenze genetiche. Si possono suddividere i pazienti, ad esempio, in base al tipo o al grado di aggregazione familiare presente. Questi e altri parametri possono altresì essere utilizzati insieme, attraverso analisi statistiche come la *cluster analysis*. Questo metodo è infatti in grado proprio di creare gruppi omogenei in base ad alcune variabili immesse. Recentemente proprio alla luce della ipotesi CV/MD prima descritta, un certo peso stanno avendo le stratificazioni basate sulla co-morbilità. Si differenziano i pazienti in base al fatto di avere o non avere due patologie associate, ad esempio diabete e obesità, asma e dermatite atopica, epilessia e disordini psichiatrici, ecc. Il primo, importante problema nell'approccio del gene candidato è, quindi, quello della eterogeneità genetica. Razionali sistemi di stratificazione possono aiutare la sua risoluzione.

*I geni interagiscono epistaticamente tra di loro. Col termine epistasi si definisce quel fenomeno per il quale il valore fenotipico di un gene viene influenzato dal valore fenotipico di un altro gene. Questo complica molto l'analisi perché la natura e gli effetti di queste interazioni non sono ancora completamente compresi*

### Il problema dell'epistasi

Col termine epistasi classicamente si definisce quel fenomeno per il quale il valore fenotipico di un gene viene influenzato dal valore fenotipico di un altro gene. Quando si pensa a un effetto poligenico, la cosa più intuitiva cui si pensa è un effetto additivo. Si è portati a credere che ogni gene aggiunga un qualcosa al determinismo di quel carattere senza entrare in relazione con gli

altri geni. Questo scenario, per quanto realistico e dimostrato scientificamente in molti casi, non è però il più probabile. Tutte le conoscenze accumulate negli ultimi anni sul funzionamento dei geni hanno dimostrato, infatti, un elevato grado di interazione gene-gene. Questa interazione può avvenire sia nei complessi macromolecolari di cui si ha sempre maggiore conoscenza sia nella partecipazione a vie metaboliche comuni. Ci sono, quindi, sempre maggiori dimostrazioni che i geni interagiscono epistaticamente l'uno con l'altro e che, quindi, una variante allelica di un gene potrebbe avere influenza nella suscettibilità a una malattia solo se presenti altre varianti di altri geni. È ovvio che l'esistenza di epistasi complica grandemente l'analisi, anche perché la natura e gli effetti di queste interazioni sono ancora lontani dall'essere completamente definiti e, per larga misura, rimangono non prevedibili. In ogni caso nasce la necessità non più di associare a una malattia una forma allelica ma la presenza contemporanea di più forme alleliche. Ad esempio, se una popolazione in base a una sola variante (gene A) potrà essere divisa in tre gruppi (omozigoti AA, eterozigoti Aa e omozigoti aa), analizzando due varianti insieme i gruppi diventano 9, con tre diventano 27 e così via. La **figura 2** illustra questo problema. È ovvio che il numero dei pazienti per gruppo diminuisce, rendendo difficile se non impossibile il raggiungimento di una significatività statistica.

### Il problema della dimensionalità

La necessità di stratificare insieme con quella di tenere in giusto conto un possibile effetto epistatico, concorrono a determinare quella che alcuni chiamano la "maledizione" della dimensionalità. Infatti, la necessità di analizzare sottogruppi è talmente incalzante e progressiva che anche campioni di dimensioni enormi potrebbero essere insoddisfacenti a raggiungere una significatività statistica. Esistono diversi approcci a questo problema.

La regressione logistica multipla è senza dubbio la metodica più classica per l'analisi delle possibili associazioni tra geni candidati e malattie complesse. Essa è orientata a trovare la probabilità di un oggetto di essere classificato all'interno di una categoria, in relazione a una combinazione di fattori di predizione. Questa metodica è stata utilizzata con successo in molti casi come, ad esempio, la componente genetica della retinopatia diabetica (8) e il rischio per lo sviluppo di disordini del linguaggio (9). Purtroppo questa metodica soffre di gravi limitazioni. Prima di tutto essa è dipendente dalla numerosità del

**Figura 2**  
**Modulo di interazione gene-gene**

1 Locus		
AA	Aa	aa

  

2 Loci		
AA BB	AA Bb	AA bb
Aa BB	Aa Bb	Aa bb
aa BB	aa Bb	aa bb

3 Loci			
AABBCC	AABBCc	AABBcc	AA
AABbCC	AABbCc	AABbcc	
AAbBCC	AAbBCc	AAbBcc	Aa
AaBBCC	AaBBCc	AaBBcc	
AaBbCC	AaBbCc	AaBbcc	
AaabbCC	AaabbCc	Aaabbcc	aa
aaBBCC	aaBBCc	aaBBcc	
aaBbCC	aaBbCc	aaBbcc	
aabbCC	aabbCc	aabbcc	

campione, e per gli argomenti portati prima, risulta evidente quanto questo sia limitante. In aggiunta essa impone relazioni fisse tra genotipo e fenotipo. Anche questo scenario potrebbe non essere realistico alla luce delle riflessioni sull'effetto epistatico. In conclusione, per quanto ancora molto usata, tale metodica non risulta perfettamente adeguata agli scopi dell'analisi dei geni responsabili di malattie complesse.

Una promettente alternativa potrebbe risiedere nelle reti neurali artificiali (*artificial neural networks*, ANN). In questo contesto, questo strumento di calcolo potrebbe essere interpretato come una generalizzazione della regressione logistica in uno spazio non lineare. Sono state descritte interessanti applicazioni di questa metodica (10). Nello sforzo di ricercare nuovi e più idonei strumenti di analisi, sono state adattate metodiche in passato usate per altri scopi. È il caso dell'algoritmo "a priori". Questo algoritmo, basato su metodi Bayesiani, era stato sviluppato per identificare le preferenze dei consumatori dei supermercati. Esso analizzava i codici a barre dei vari acquirenti al fine di trovare un pattern di associazione tra i prodotti che venivano acquistati. In sostanza si tratta, quindi, di uno strumento di *pattern recognition* che è stato sperimentato nell'associazione tra SNP e malattie complesse (11–13). La sua funzione è sostanzialmente quella del cosiddetto *data mining*, cioè l'esplorazione di dati complessi alla ricerca di associazioni, da validare in una seconda fase con altre metodiche statistiche.

Il *multifactor dimensionality reduction* (MDR) è stato proposto quale metodo in grado di affrontare il problema della dimensionalità, che, come si è visto, all'aumentare dei fattori analizzati diviene repentinamente insuperabile. L'MDR agisce attraverso una strategia di riduzione della dimensionalità che si basa sull'identificazione di particolari associazioni tra fattori ad "alto rischio" o a "basso rischio". Questa nuova variabile mono-dimensionale è, poi, valutata per la sua capacità di predire lo stato di salute o di malattia. L'MDR è stato pensato, quindi, per essere in grado di svelare le interazioni non lineari gene-gene e gene-ambiente utilizzando un approccio non-parametrico e *model-free*. Questa analisi consiste in una modifica del *combinational partition method*. Nella metodica sono previsti sistemi che consentano una convalida interna. Inoltre, attraverso tecniche di permutazione, possono essere derivati dei valori di probabilità. L'MDR è stato adoperato con successo in alcuni studi (14–17) e molta attenzione è riservata alle possibili applicazioni future (18).

## Interazione gene-ambiente

Un problema aggiuntivo di non facile approccio metodologico nello studio della genetica delle malattie complesse è rappresentato dall'interazione gene-ambiente (GxE, *gene x environment*). Lo studio dell'associazione tra varianti genetiche e tratti complessi è, infatti, ulteriormente complicato dall'effetto modulante dei fattori ambientali che può esercitare un ruolo confondente difficile da stimare e che potrebbe in parte spiegare la frequente mancata replicazione dei risultati degli studi di associazione (19).

Alcune malattie a etiologia multifattoriale come il diabete, l'obesità, l'ipertensione arteriosa, l'aterosclerosi e il cancro si sono diffuse in maniera epidemica negli ultimi cinquant'anni, inoltre le manifestazioni cliniche di queste malattie si presentano in epoche sempre più precoci della vita, basti pensare al problema dell'obesità e del diabete tipo 2 (DT2) nei bambini (20, 21). La diffusione di queste malattie riproduce molto da vicino l'andamento geografico dei processi di urbanizzazione e industrializzazione. Poiché non è possibile immaginare variazioni sostanziali nel patrimonio genetico umano in un arco di tempo così limitato, queste osservazioni epidemiologiche fanno ipotizzare che, sebbene la suscettibilità genetica abbia un ruolo importante nella etiologia di queste malattie, tuttavia essa può essere modulata da fattori non genetici largamente legati allo stile di vita che si è rapidamente modificato nell'ultimo secolo come conseguenza della progressiva industrializzazione (20).

Una dieta ricca in calorie e in grassi e una ridotta attività fisica sono tipici dei paesi industrializzati. È stato ampiamente dimostrato che questo stile di vita è causalmente associato allo sviluppo di malattie metaboliche e cardiovascolari e che la modifica di queste abitudini è una misura di prevenzione molto efficace nel ridurre il rischio di malattia (21, 22). Tuttavia è osservazione comune che non tutti i soggetti esposti allo stesso ambiente sfavorevole si ammalano; per ogni singolo individuo il rischio di malattia è il risultato di una complessa interazione tra fattori genetici e fattori ambientali. L'espressione genica può avere più influenza delle variazioni genetiche sulle differenze interindividuali e può essere regolata da diversi fattori inclusa l'esposizione ad alcuni stili di vita. Un modello classico è rappresentato dai gemelli monozigoti: evidentemente il patrimonio genetico dei gemelli monozigoti è identico, ma essi hanno la potenzialità di differire in termini di come e quando i geni vengono espressi, pertanto non è raro osservare che i

---

*Lo studio della genetica dei tratti complessi è ulteriormente complicato dall'effetto modulante dei fattori ambientali. Per interazione gene-ambiente si intende il diverso effetto fenotipico prodotto da una stessa esposizione in soggetti con genotipo differente*

gemelli monozigoti possono differire biologicamente e fenotipicamente come risultato di esposizioni ambientali diverse.

Genericamente il termine GxE si riferisce al differente effetto fenotipico che esposizioni diverse hanno su soggetti con lo stesso genotipo oppure al differente effetto fenotipico di una stessa esposizione in soggetti con genotipo diverso. Alcuni autori preferiscono usare piuttosto che questo termine il termine più ampio di “effetto genetico contesto-dipendente” includendo in questa espressione sia il concetto di GxE sia quello dalla interazione gene-gene (epistasi) (23, 24). La GxE oltre a spiegare in parte la variabilità interindividuale nella suscettibilità a molte malattie, potrebbe anche aprire la strada a un nuovo modo di fare prevenzione. Infatti l'altro lato della medaglia è che soggetti sottoposti a uno stesso intervento, ad esempio dietetico, rispondono in maniera diversa (25) ed è pertanto ipotizzabile che in un futuro non lontano le misure terapeutiche preventive possano essere più mirate (“personalizzate”).

Sebbene il concetto di interazione tra geni e ambiente fosse stato di fatto intuito già all'inizio del secolo, solo recentemente, con gli enormi progressi della genetica, si è iniziato a esplorarne le basi molecolari. C'è un crescente interesse della comunità scientifica per lo studio del ruolo della GxE e soprattutto nella interazione tra dieta e geni, nella etiologia delle malattie complesse, tanto che sono nati i termini “nutrigenetica” e “nutrigenomica” per definire nuove branche della genetica che studiano gli effetti combinati di geni e fattori nutrizionali nella etiologia delle malattie complesse. La nutrigenetica studia l'effetto delle variazioni genetiche (essenzialmente SNP), sull'interazione tra dieta e specifici fenotipi con l'obiettivo di valutare i rischi e i benefici per l'individuo di determinate componenti della dieta. La nutrigenomica studia l'effetto dei nutrienti sul genoma, proteoma e metaboloma con l'obiettivo di chiarire come la dieta possa influenzare le vie metaboliche e la loro omeostasi; comporta pertanto la caratterizzazione dei prodotti genici e della loro funzione.

### Alcuni esempi

I lipidi sono i più studiati determinanti delle malattie cardiovascolari e la comprensione dei meccanismi molecolari alla base dei disordini del metabolismo lipidico è di grande importanza per la prevenzione delle malattie cardiovascolari. I geni coinvolti nella regolazione del metabolismo lipidico finora identificati sono numerosissimi e l'elenco non è ancora completo (26), inol-

tre è noto che l'omeostasi del metabolismo lipidico è regolata anche da diversi fattori ambientali non genetici come abitudine al fumo di sigaretta, consumo di alcool, composizione della dieta e attività fisica. Sebbene allo stato attuale delle conoscenze non sia possibile stabilire con precisione il contributo dei fattori genetici e dei fattori ambientali nella etiologia delle dislipidemie, tuttavia sono molto rari i casi in cui una dislipidemia si manifesti per effetto di una alterazione genetica in assenza di un contesto ambientale predisponente, e anche in questi casi i fattori ambientali sono comunque in grado di modulare la severità del disordine metabolico e di influenzare l'età in cui questo si manifesta.

Sotto il profilo della salute pubblica la nutrizione è il fattore ambientale più importante che interagisce con i nostri geni nel modulare la comparsa di disordini del metabolismo lipidico. È noto che la concentrazione dei lipidi plasmatici è molto influenzata dal contenuto di grassi saturi della dieta e che a livello di popolazione le concentrazioni medie di colesterolo sono più elevate in quei paesi che consumano diete ricche di grassi saturi e più basse nei paesi con diete povere di grassi e ricche di vegetali e fibre. Tuttavia a livello dell'individuo la variazione dei lipidi plasmatici in risposta a modifiche dietetiche è variabile, alcuni soggetti rispondono molto bene, mentre altri sono relativamente insensibili. Uno studio americano controllato (27) riporta che in risposta a una stessa dieta ipocolesterolemizzante, le variazioni osservate nella concentrazione di colesterolo delle *low density lipoproteins* (LDL-C) variavano da +3% a -55% negli uomini, e da +11% a -39% nelle donne partecipanti all'esperimento. Queste differenze di risposta non si spiegavano con una diversa adesione dei pazienti alla dieta, ma riflettevano una reale variabilità interindividuale nella risposta. Poiché la dieta testata in questo esperimento era la stessa dieta raccomandata dal *National Cholesterol Education Program* per la prevenzione delle malattie cardiovascolari (26), i risultati suggeriscono che, pur rimanendo le raccomandazioni generalmente valide a livello di popolazione, tuttavia per alcuni individui sono più appropriate che per altri.

Si è già detto che molti geni sono presenti nella popolazione in forma polimorfica, ossia con una elevata variabilità, senza che questo abbia di per sé effetti rilevanti sullo stato di salute. I polimorfismi conosciuti di geni coinvolti nel metabolismo lipidico sono numerosissimi, uno dei più studiati è il polimorfismo del gene dell'ApoE. Questo gene regola la sintesi di una proteina che

---

*Sono conosciuti diversi polimorfismi genetici che condizionano la risposta fenotipica ad alcune esposizioni ambientali (es. dieta ricca in grassi, dieta povera in vitamina B12 e acido folico)*

partecipa al trasporto del colesterolo, ne esistono tre varianti, ApoE2, 3 e 4, che modulano l'impatto della dieta sulla concentrazione dei lipidi plasmatici che a loro volta condizionano il rischio cardiovascolare. In particolare la variante ApoE4 si associa con una spiccata sensibilità a sviluppare livelli elevati di colesterolo totale e LDL-C in presenza di un'alimentazione ricca in colesterolo, tipica dei paesi industrializzati. Pertanto i portatori di ApoE4 in questo contesto sono svantaggiati, tuttavia questi soggetti sono anche quelli che rispondono meglio quando sottoposti a diete con ridotto contenuto di grassi, mentre i portatori delle varianti ApoE2 e 3 presentano risposte variabili come è dimostrato da diversi studi di intervento (28, 29).

Un altro esempio è rappresentato dal metabolismo della omocisteina. Livelli plasmatici anche moderatamente elevati di omocisteina sono un fattore di rischio per lo sviluppo di eventi cardiovascolari (30, 31). Il fumo di sigaretta e l'apporto dietetico di folati e vitamina B12 sono tra i principali determinanti delle concentrazioni plasmatiche di omocisteina (32). Negli ultimi anni sono stati identificati numerosi polimorfismi di geni che codificano per enzimi chiave nel metabolismo della omocisteina. In particolare la mutazione 667CT dell'enzima metilene tetraidrofolato reduttasi (MTHFR), uno degli enzimi chiave nel metabolismo della omocisteina, è associata a una forma enzimatica termolabile, a ridotta attività (33, 34). Questo enzima è necessario per la conversione del 5,10-metilene tetraidrofolato (5,10-MTIF) (derivato dal folato introdotto con la dieta) in 5-metilene tetraidrofolato (5-MTIF) che è la forma attiva di folato necessaria per la conversione dell'omocisteina a metionina. Questo polimorfismo, tuttavia, non ha conseguenze sui livelli di omocisteina se il contenuto di folati della dieta è elevato, ma si associa a iperomocisteinemia se il contenuto di acido folico della dieta è scarso. Allo stesso modo i soggetti portatori della variante 677CT sono anche quelli che rispondono meglio quando sottoposti a una supplementazione della dieta con folati.

## Conclusioni

Il rapido sviluppo di nuovi marcatori genetici come le SNP e la possibilità di avere a disposizione metodiche di genotipizzazione rapide e poco costose ci pone nella possibilità, almeno teorica, di affrontare la sfida della ricerca dei geni responsabili di malattie complesse. A questa sfida gli statistici stanno rispondendo proponen-

do strumenti di analisi sempre più interessanti che sono stati rapidamente visitati in questo scritto. Va subito detto che, tra i metodi descritti non c'è ancora "il migliore". La maggior parte di essi, infatti, va ancora convalidato e ulteriormente sperimentato. È possibile che in alcuni casi un metodo risulti migliore di un altro per ragioni legate alla numerosità del campione, al numero dei geni coinvolti o alla modalità di trasmissione. Anche le metodologie per la valutazione della GxE, non sono ancora completamente adeguate per lo studio di fenomeni biologici tanto complessi. Spesso vengono utilizzati modelli basati sull'interazione tra dieta e un singolo gene, il che rappresenta certamente una eccessiva semplificazione della realtà. Tuttavia essi sono in continua evoluzione e siamo certi che in un futuro non lontano potremo disporre di strumenti più adeguati.

## Bibliografia

1. Mayeux R. Mapping the new frontier: complex genetic disorders. *J Clin Invest* 115: 1404–1407, 2005.
2. Strauch K, Fimmers R, Baur MP, Wienker TF. How to model a complex trait. 1. General considerations and suggestions. *Hum Hered* 55: 202–210, 2003.
3. Prentice AM, Rayco-Solon P, Moore SE. Insights from the developing world: thrifty genotypes and thrifty phenotypes. *Proc Nutr Soc May* 64: 153–161, 2005.
4. Wicking C, Williamson B. From linked marker to gene. *Trends Genet* 7: 288–293, 1991.
5. Newton-Cheh C, Hirschhorn JN. Genetic association studies of complex traits: design and analysis issues. *Mutat Res* 573: 54–69, 2005.
6. Gordi T, Khamis H. Simple solution to a common statistical problem: interpreting multiple tests. *Clin Ther* 26: 780–786, 2004.
7. Schulze TG, McMahon FJ. Defining the phenotype in human genetic studies: forward genetics and reverse phenotyping. *Hum Hered* 58: 131–138, 2004.
8. Cruickshanks KJ, Vadheim CM, Moss SE, et al. Genetic marker associations with proliferative retinopathy in persons diagnosed with diabetes before 30 yr of age. *Diabetes* 41: 879–885, 1992.
9. Felsenfeld S, Plomin R. Epidemiological and offspring analyses of developmental speech disorders using data from the Colorado Adoption Project. *J Speech Lang Hear Res* 40: 778–791, 1997.
10. Sherriff A, Ott J. Applications of neural networks for gene finding. *Adv Genet* 42: 287–297, 2001.
11. Toivonen HT, Onkamo P, Vasko K, et al. Data mining applied to linkage disequilibrium mapping. *Am J Hum Genet* 67: 133–145, 2000.
12. Flodman P, Macula AJ, Spence MA, Torney DC. Preliminary implementation of new data mining techniques for the analysis of simulation data from Genetic Analysis Workshop 12: Problem 2. *Genet Epidemiol* 21: 390–395, 2001.
13. Cizka WA, Weir BS, Edwards SR, et al. Applying data mining techniques to the mapping of complex disease genes. *Genet Epidemiol* 21: 435–440, 2001.
14. Qin S, Zhao X, Pan Y, et al. An association study of the N-methyl-D-aspartate receptor NR1 subunit gene (GRIN1) and NR2B subunit gene (GRIN2B) in schizophrenia with universal DNA microarray. *Eur J Hum Genet* 13: 807–814, 2005.



15. Williams SM, Ritchie MD, Phillips JA 3rd, et al. Multilocus analysis of hypertension: a hierarchical approach. *Hum Hered* 57: 28–38, 2004.
16. Coffey CS, Hebert PR, Ritchie MD, et al. An application of conditional logistic regression and multifactor dimensionality reduction for detecting gene-gene interactions on risk of myocardial infarction: the importance of model validation. *BMC Bioinformatics* 5: 49, 2004.
17. **Cho YM, Ritchie MD, Moore JH, et al. Multifactor-dimensionality reduction shows a two-locus interaction associated with type 2 diabetes mellitus. *Diabetologia* 47: 549–54, 2004.**
18. Bastone L, Reilly M, Rader DJ, Foulkes AS. MDR and PRP: a comparison of methods for high-order genotype-phenotype associations. *Hum Hered* 58: 82–92, 2004.
19. **Davey Smith G, Ebrahim S. Mendelian randomization: Can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol* 32: 1–22, 2003.**
20. Mann NJ. Paleolithic nutrition: what can we learn from the past? *Asia Pac J Clin Nutr* 13: S17–23, 2004.
21. **Diabetes Prevention Program Research Group: Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. *N Eng J Med* 346: 393–403, 2002.**
22. Tuomilehto J, Lindstrom J, Eriksson JG, et al. Finnish Diabetes Prevention Study Group. Prevention of type 2 diabetes mellitus by changes in lifestyle among subjects with impaired glucose tolerance. *N Engl J Med* 344: 1343–1350, 2001.
23. Kraft P, Hunter D. Integrating epidemiology and genetic association: the challenge of gene-environment interaction. *Philos Trans R Soc Lond B Biol Sci* 360: 1609–1616, 2005.
24. Kramer DA. Commentary: Gene-environment interplay in the context of genetics, epigenetics and gene expression. *J Am Acad Child Adolesc Psychiatry* 44: 19–27, 2005.
25. Katan Mb, Grundy SM, Willett WC. Should a low-fat, high-carbohydrate diet be recommended for everyone? Beyond low fat diets. *N Engl J Med* 337: 563–566, 1997.
26. Schaefer EJ, Lamon-Fava S, Ausman LM, et al. Individual variability in lipoprotein cholesterol response to National Cholesterol Education Program Step 2 diets. *Am J Clin Nutr* 65: 823–830, 1997.
27. **Corella D, Ordovas JM. Single nucleotide polymorphisms that influence lipid metabolism: interaction with dietary factors. *Ann Rev Nutr* 25: 341–390, 2005.**
28. Lokitonov A. Common gene polymorphism and nutrition: emerging links with pathogenesis of multifactorial chronic diseases. *J Nutr Biochem* 14: 426–451, 2003.
29. Masson LF, McNeill G, Avenell A. Genetic variation and the lipid response to dietary intervention: a systematic review. *Am J Clin Nutr* 77: 1098–1111, 2003.
30. Homocysteine studies collaboration. Homocysteine and risk of ischemic heart disease and stroke: a metaanalysis. *JAMA* 288: 2015–2018, 2002.
31. Quinlivan EP, McPartlin J, McNulty H, et al. Importance of both folic acid and vitamin B12 in reduction of risk of vascular disease. *Lancet* 359: 227–231, 2002.
32. Bratstrom L, Widken D, Ohrvik J. Common metilennetrahydrofolate reductase gene mutation leads to hyperhomocysteinemia but not to vascular disease: the result of metaanalysis. *Circulation* 98: 2520–2526, 1998.
33. Zamboni M, DiFrancesco V, Zoico E, et al. Homocysteine and life style in the elderly. *Aging* 13: 437–440, 2001.
34. Schnyder G, Roffi M, Flammer Y, et al. Effect of homocysteine lowering therapy with folic acid, vitamin B12, and vitamin B6, on clinical outcome after percutaneous coronary intervention. The Swiss Heart Study: a randomised controlled trial. *JAMA* 288: 973–979, 2002.

**Nota:** in grassetto sono evidenziati gli “abstract milestones”.