

Analisi di DNA per la valutazione del rischio cardiovascolare

Gian Antonio Danieli

*Laboratorio di Genetica Molecolare e Genomica
Dipartimento di Biologia, Università di Padova*

Il progresso compiuto dalla Genetica umana molecolare nell'ultimo decennio è stato velocissimo. Il risultato che ha ottenuto maggiore risonanza a livello di opinione pubblica è stato il completamento della sequenza del genoma umano. La notizia della accessibilità alla sequenza nucleotidica di ogni gene ha indotto molti a ritenere che si possano stabilire correlazioni univoche tra specifiche mutazioni e specifici fenotipi clinici e che sia quindi possibile diagnosticare qualsiasi malattia organica mediante analisi di DNA e valutare con grande precisione il rischio nella predisposizione genetica a qualsiasi patologia degenerativa. In particolare, si è rapidamente diffusa tra i pazienti ed i loro familiari la convinzione che i test genetici, essendo sostanzialmente analisi di sequenze nucleotidiche, siano semplici, rapidi ed assolutamente precisi. In realtà, i test di DNA sono invece piuttosto complessi, spesso costosi e possono presentare non trascurabili problemi metodologici.

L'identificazione dei segmenti di genoma implicati nel determinismo genetico di molte malattie con eredità mendeliana ha consentito di scoprire i geni coinvolti, e con quali mutazioni patogene, nella determinazione del difetto primario. Tuttavia si è capito che l'equazione "mutazione - malattia" è possibile soltanto in pochissimi casi di patologie con penetranza completa ed espressività costante. Anche i risultati delle ricerche per individuare associazioni tra specifici polimorfismi nucleotidici e specifiche predisposizioni hanno indotto a ridimensionare le speranze di avere a disposizione efficaci e precisi strumenti di medicina molecolare predittiva. Inoltre rimane il problema di come utilizzare i dati ottenuti da tali test per la valutazione del rischio, in particolare nell'ambito delle malattie cardiovascolari.

Per tutto il ventesimo secolo, la Genetica aveva proceduto identificando l'esistenza di ciascun gene in base alla ricorrenza familiare di un dato fenotipo "anomalo", la cui trasmissione nelle generazioni rispettasse le regole della segregazione mendeliana.

Non era possibile conoscere né il prodotto "normale" del gene in questione, né le mutazioni patogene in grado di determinare la comparsa di tale fenotipo. Lavorando in questa prospettiva limitata e restringendo necessariamente l'analisi genetica ai casi il cui fenotipo clinico corrispondesse pienamente ad un dato standard, raramente ci si è posti il proble-

ma di eventuali contributi di altri geni o di fattori ambientali nella determinazione del fenotipo.

La ricerca genetica attuale è centrata sulla conoscenza della sequenza del DNA del gene, da cui discendono sia la conoscenza del prodotto che di eventuali alterazioni dovute a mutazione, e quindi la comprensione della patogenesi a livello molecolare. Nell'ultimo decennio, la maggior parte dei ricercatori si è affannata a cercare i geni implicati in malattie umane ed a descriverne le mutazioni patogene, trascurando però di considerare che ogni gene si esprime nel contesto genomico individuale e che variazioni in altri geni possono avere effetti non trascurabili sul fenotipo clinico considerato. Perciò oggi, nonostante siano disponibili sia l'intera sequenza del DNA genomico che una serie di metodi per identificare la presenza di eventuali mutazioni in un dato segmento, possono sfuggire altre componenti del genotipo (ossia altri geni) le cui varianti siano in grado di influire in modo correttivo o peggiorativo sulla patogenesi. Sfugge inoltre, nella stragrande maggioranza dei casi, quali siano i fattori "ambientali" (inclusi nel concetto piuttosto vago di "stile di vita") in grado di influire positivamente o negativamente sulla patogenesi.

Il concetto di "determinazione genetica" ha comunque subito negli ultimi decenni una notevole revisione: le (pochissime) malattie genetiche con eredità mendeliana, penetranza completa ed espressività costante sono all'estremo di una distribuzione nella quale troviamo, senza soluzione di continuità, malattie mendeliane con gradi diversi di penetranza e di espressività, malattie poligeniche con uno o qualche gene principale (major genes), malattie tipicamente poligeniche che risentono di effetti ambientali, per giungere infine a malattie multifattoriali nelle quali varianti genetiche potenzialmente patogene non costituiscono causa né necessaria né sufficiente per determinare la manifestazione della malattia. A seconda della malattia considerata ed a seconda del gene e della mutazione potremo avere mutazioni in grado di esercitare una determinazione assoluta del fenotipo, una forte determinazione (ma suscettibile di eventuali modificazioni da parte dell'effetto di altri geni o dell'ambiente) o una determinazione tanto debole da avere effetto soltanto in presenza di altre mutazioni patogene con effetto altrettanto modesto. Questo quadro di riferimento rende impossibile qualsiasi generalizzazione riguardo alla ca-

pacità predittiva di test genetici applicati alla clinica. D'altra parte è però indubbio che una volta identificata in un paziente una mutazione patogena, la ricerca della stessa mutazione nei membri della sua famiglia permette di identificare le persone che potrebbero eventualmente sviluppare la malattia. Il vero problema è definirne il rischio.

Al momento attuale, soltanto per un numero limitato di malattie umane è disponibile l'informazione relativa al gene implicato, tuttavia il progresso in questa direzione è abbastanza rapido. Chi fosse interessato ad una specifica malattia genetica può consultare OMIM¹ per conoscere lo stato di avanzamento delle conoscenze sia a livello genetico che a livello clinico. Verosimilmente ci si può attendere che entro qualche anno siano noti tutti i geni implicati nella determinazione delle malattie ereditate come caratteri mendeliani semplici.

In linea di principio, una volta noto il gene implicato, dovrebbe essere facile identificarne eventuali mutazioni patologiche. In realtà però ciò è vero soltanto per le sequenze codificanti analizzate mediante sequenziamento diretto del DNA o mediante HPLC denaturante (DHPLC) (anche se talvolta il numero elevato di esoni da analizzare, la loro dimensione o peculiari composizioni in nucleotidi possono comportare un carico di lavoro non indifferente). Una volta identificata una mutazione in una sequenza, se il suo ruolo patogeno non può essere desunto direttamente o se non sia stato già stato riportato in letteratura, esso deve essere dimostrato. Molto spesso l'evidenza è indiretta, basata unicamente sul fatto che la variazione in questione non sia stata trovata in una serie costituita da qualche centinaio di individui di controllo. Molto più raramente l'evidenza risulta da vere e proprie prove funzionali.

La ricerca di eventuali mutazioni in sequenze di regolazione è molto più problematica, in quanto non solo comporta il sequenziamento di lunghi segmenti di DNA a monte ed a valle del gene, ma richiede successive prove funzionali per verificare l'effetto della variazione identificata sul livello di espressione del gene.

Eventuali mutazioni in sequenze introniche, che determinino l'attivazione di siti criptici di *splicing*, o che riducano l'efficienza dello *splicing* possono essere individuate soltanto mediante analisi del cDNA.

Queste sommarie informazioni sono sufficienti a chiarire che il procedimento per giungere all'identificazione di una mutazione patogena può risultare molto lungo e complesso. Un eventuale risultato negativo di tale ricerca non significa necessariamente si tratti di una malattia diversa da quella per la quale era stata posta diagnosi; la mutazione potrebbe infatti trovarsi in una sequenza non considerata dall'indagine (es. in una delle sequenze *untranslated* (UTR)) oppure potrebbe non essere stata individuata perchè il metodo impiegato non è abbastanza sensibile (ad esempio, l'analisi mediante DHPLC ha un potere di risoluzione inferiore a quello del sequenziamento diretto ed anche quest'ultimo può avere un tasso di individuazione inferiore al 100% qualora la sequenza di DNA analizzata sia ricca di nucleotidi ripetuti).

L'analisi di mutazioni di singoli geni è oggi utilizzata nella pratica clinica in diverse cardiomiopatie, in diverse malattie aritmiche, in alcuni difetti dei vasi sanguigni ed in alcuni difetti di geni correlati all'aterogenesi. Tuttavia, finora sono disponibili soltanto in pochi casi dati di correlazione genotipo-fenotipo che possano essere utilizzati in senso predittivo.

La scoperta che alcune specifiche varianti di polimorfismi di

DNA appaiono associate ad alcuni fenotipi clinici ha generato notevole interesse, grandi entusiasmi e grandi delusioni. Un esempio per tutti è quello del polimorfismo dell'*angiotensin converting enzyme* (ACE). Un introne di questo gene presenta una variazione interindividuale che riguarda la lunghezza di una ripetizione nucleotidica. Nel 1992 venne riportato che la variante "corta" in omozigosi risultava significativamente più frequente tra gli infartuati che tra i sani²; da ciò si dedusse che tale polimorfismo potesse essere utilizzato come "potent predictor of myocardial infarction". Dopo la pubblicazione di una miriade di studi basati su numeri relativamente modesti di osservazioni, finalmente uno studio basato su circa 5.000 casi e 6.000 controlli⁴ ha stabilito che in realtà si trattava di un predittore molto modesto.

In letteratura compaiono abbastanza spesso segnalazioni dell'esistenza di polimorfismi di DNA associati a specifiche patologie; nella maggior parte dei casi mancano dati riguardo la sensibilità e specificità di tali tests, per cui non dovrebbero assolutamente essere utilizzati nella pratica clinica. Le prospettive di questo settore di ricerca oggi riscuotono modesto interesse. Infatti viene privilegiata la ricerca di varianti funzionali di specifici geni perché potenzialmente in grado di essere messi in relazione, in termini di patogenesi, con il fenotipo clinico.

Una nuova metodologia per identificare alterazioni genetiche patologiche, utilizzabile in prospettiva per valutare il rischio, è costituita dai DNA microarrays.

Mediante questa tecnica è possibile valutare lo schema di espressione di diversi geni in un dato tessuto, rispetto ad uno standard di riferimento. Nel caso di una qualsiasi alterazione genetica patogena (sia essa mono- o poli-genica) lo schema di espressione genomica a livello del tessuto interessato dalla malattia risulterà alterato.

Teoricamente dovrebbe essere possibile tipizzare un dato paziente in base alla serie di geni la cui espressione subisca significative variazioni rispetto al normale e, reciprocamente, utilizzare tali variazioni come elemento diagnostico. I progressi in questo campo sono estremamente incoraggianti⁵, nonostante permangano irrisolti alcuni problemi di standardizzazione del metodo.

Indubbiamente il problema più pressante in cardiologia è rappresentato dalla malattia arterosclerotica e dall'infarto miocardico, data l'elevata prevalenza di queste patologie nella popolazione generale. Sono state finora identificate varianti di alcuni geni associate ad aumentato rischio (Tabella I), ma la ricerca è soltanto all'inizio. Sfortunatamente, trattandosi di malattie multifattoriali, dovute cioè ad una miscela di fattori genetici ed ambientali, diversi sono i geni che contribuiscono al fenotipo clinico e forse ciascuno di essi ha singolarmente un effetto patogeno relativamente modesto.

Per riuscire a determinare la predisposizione genetica a sviluppare la malattia considerata, i metodi di laboratorio dovrebbero consentire l'analisi contemporanea di tali geni, per identificarne eventuali varianti predisponenti.

L'unica prospettiva attuale è costituita da DNACHIPS per effettuare *minisequencing* di specifiche sequenze. Si tratta di metodologie ancora molto costose ed attualmente in grado di analizzare soltanto quattro o cinque geni alla volta, purchè siano di piccole dimensioni. Mancano inoltre per il momento le informazioni su quali sequenze di DNA sia più opportuno scegliere per tale indagine.

Tabella I. Alcune varianti genetiche che sono state associate ad aumento di rischio di coronaropatie o di infarto miocardio (informazioni dettagliate si possono reperire al corrispondente numero di OMIM, consultando il sito: <http://www.ncbi.nlm.nih.gov/omim/>).

OMIM	Gene	Variante
143890	LDR	Molte
505747	ARH	Molte
107730	ApoB100	Arg3500Glu
107741	ApoE	Many
238600	LPL	Asn291Ser
106150	AGT (Angiotensinogen)	Thr174Met Met235Thr
188040	THBD (Thrombomodulin)	Ala455Val 1bpIns1689T
264900	F7 (Clotting Factor VII)	10bpIns-323 Arg353Gln

Poiché nelle malattie con determinazione multifattoriale la variazione in uno o più geni può essere condizione necessaria, ma non sufficiente per causare la malattia, ogni predizione risulta problematica: Anche ammesso che fossero noti tutti i geni implicati nella determinazione di una data malattia multifattoriale ed ammesso che fossero state identificate in un dato soggetto tutte le varianti di tali geni, sarebbe impossibile stabilire un preciso valore di rischio, per l'ignoranza dei fattori ambientali che hanno agito sul dato individuo nel corso della sua esistenza. E' quindi molto improbabile che, almeno nel prossimo futuro, l'analisi di DNA possa predire con sicurezza i fenotipi clinici. Tuttavia, essa può diventare un utile strumento ancillare per valutare la predisposizione genetica individuale ad una data malattia, sempre che siano disponibili informazioni attendibili sul valore di rischio associato sia a ciascuna variante che alla presenza contemporanea di specifiche varianti nello stesso individuo.

In questa prospettiva, è utile ricordare che i tests genetici utilizzati per identificare preventivamente gli individui a rischio cardiovascolare sono ancora in una fase sperimentale. Come si è già detto, i risultati riportati in letteratura riguardo ad un dato test dovrebbero essere considerati con molta cautela fintanto che non ne sia stata stabilita la sensibilità (% di veri positivi individuati nel campione di ammalati esaminato), specificità (% di veri negativi nel campione di individui sani esaminati, valore predittivo positivo (% di veri positivi tra coloro che sono risultati positivi al test) e valore predittivo negativo (% di veri negativi tra coloro che sono risultati negativi al test). Inoltre, prima di introdurre il test nella pratica clinica, se ne deve valutare attentamente la validità clinica (cioè l'accuratezza del test nell'identificare o predire una data malattia) e l'utilità clinica (ossia l'utilità del test per gli individui cui il test è destinato). Da questo punto di vista, le analisi di DNA per scopo predittivo non differiscono da un qualsiasi test biochimico-clinico.

Indipendentemente dal fatto che la malattia sia ereditata come carattere mendeliano semplice o viceversa come carattere multifattoriale, la valutazione del rischio riguarda due situazioni diverse a) stabilire la probabilità che un dato individuo

(probando) sviluppi una data malattia, e b) stabilire la probabilità che possano manifestarsi nella famiglia del caso indice altri casi della stessa malattia, nella generazione presente o in generazioni future.

In entrambi i casi, un'accurata anamnesi familiare rappresenta il punto di partenza obbligato dell'indagine.

Quando non siano riportati parenti del probando che manifestino la malattia, nè nella presente generazione, nè nelle tre passate generazioni, il rischio di sviluppare la malattia ereditaria in questione corrisponde al tasso di mutazione del gene implicato, se la malattia è ereditata come carattere autosomico dominante; corrisponde invece approssimativamente al tasso di incidenza della malattia qualora l'ereditarietà sia di tipo autosomico recessivo. Nel caso di ereditarietà di tipo recessivo *X-linked* e qualora il probando sia di sesso maschile, la probabilità che sviluppi la malattia è circa uguale al tasso di mutazione del gene; se invece si tratta di una probanda, la probabilità corrisponde al tasso di prevalenza del fenotipo affetto tra le femmine della popolazione generale.

Quando l'anamnesi familiare risulta positiva ed è nota la modalità di trasmissione ereditaria del difetto, la probabilità *a priori* di avere ereditato il genotipo che determina la manifestazione della malattia viene calcolata semplicemente applicando le regole della segregazione dei cromosomi in ogni generazione. Nella Figura 1 è riportato un albero genealogico che riguarda la trasmissione di un carattere patologico ereditato come carattere autosomico dominante, che si manifesta nell'età adulta, dopo i 35 anni.

Nell'esempio citato, il padre degli individui III,4 e III,5 è ammalato; quindi deve essere portatore dell'allele patogeno. La probabilità di aver ereditato tale allele dal padre sarà 0,5 (50%) per ciascuno dei due fratelli. Nel caso di caratteri con penetranza incompleta è necessario apportare una correzione. Per esempio, se l'allele patogeno avesse una penetranza del 70%, la probabilità di ereditare l'allele in questione e di manifestare la malattia sarebbe per ciascun probando (70%) x (50%) = 35%. Essi quindi avrebbero una probabilità di 65% di non manifestare la malattia nel corso della vita.

Un calcolo più preciso del rischio si può utilizzare la logica Bayesiana.

Consideriamo ad esempio un individuo dell'età di 55 anni, la cui madre ha manifestato a 50 anni una malattia genetica ereditata come carattere autosomico dominante e che si manifesta tra i 40 ed i 60 anni.

La probabilità di aver ereditato l'allele patogeno è definita *a priori* poiché deriva dalla conoscenza dell'anamnesi familiare. In questo caso la probabilità *a priori* di aver ereditato l'allele patogeno sarà 0.5 (50%). Supponiamo di conoscere che il 90% (0.9) dei casi manifesti sintomi tipici della malattia prima dei 55 anni. Quindi la probabilità di non mostrare sintomi di malattia a 55 anni, pur avendo ereditato l'allele patogeno sarà 0.10 (1-0.9). Questa probabilità viene detta "probabilità condizionale". La probabilità condizionale di non manifestare sintomi essendo non-affetto sarà ovviamente 1. Moltiplicando la probabilità *a priori* per la probabilità condizionale si ottiene il valore della "probabilità congiunta". Si devono calcolare due probabilità congiunte, una per l'ipotesi che il probando abbia ereditato l'allele patogeno e l'altra nell'ipotesi che non l'abbia ereditato.

Una volta calcolate le probabilità congiunte, la probabilità *a posteriori* (finale) si ottiene dal rapporto tra la probabilità con-

giunta di aver ereditato l'allele patogeno e la somma delle due probabilità congiunte (rispettivamente di avere e di non-avere ereditato l'allele patogeno).

Nell'esempio considerato si avrebbe:

Probabilità	"Affetto"	"Non-Affetto"
<i>A priori</i> (genetica)	0.5	0.5
Condizionale	0.1	1
Congiunta	0.05	0.5
<i>A posteriori</i>	0.09	

e quindi un rischio inferiore al 10%.

Spesso sono disponibili altre informazioni riguardanti il fenotipo clinico o risultati di tests. Si può quindi calcolare una probabilità condizionale per ciascuna informazione disponibile, aumentando così la precisione della stima di rischio. Bisogna però ricordare che le probabilità condizionali devono riguardare eventi indipendenti. Consideriamo ad esempio il caso di un probando di 45 anni di età, il cui padre è deceduto per CAD (Coronary Atherosclerotic Disease). Il probando presenta un livello di colesterolo superiore a 240 mg/dl. Supponiamo che la sensibilità del test sia 0.75 e la sua specificità 0.9. Ipotizziamo inoltre che il soggetto abbia il genotipo DD del polimorfismo ACE (in questo caso la sensibilità del test è 0.45 e la specificità 0.72, secondo i dati riportati da Keavney *et al.*⁴).

In questo caso, la probabilità *a priori* non può essere calcolata in base alle leggi di Mendel, in quanto si tratta di una malattia ereditata come carattere multifattoriale. Tuttavia sono disponibili dati empirici secondo i quali il rischio genetico di sviluppare una CAD per una persona con anamnesi familiare positiva per la malattia è circa 1/5 (0.2). La probabilità condizionale di sviluppare una CAD in presenza di colesterolo >240 mg/dL corrisponde alla sensibilità del test (0.75), mentre la probabilità condizionale di non sviluppare CAD sarà: (1- 0.9), ossia (1- specificità).

Lo stesso ragionamento viene ripetuto per calcolare la probabilità condizionale associata al risultato del test ACE.

Quindi otterremo:

Probabilità	"Affetto"	"Non-Affetto"
<i>A priori</i> (genetica)	0.2	0.8
Condizionale 1 (Chol)	0.75	0.1
Condizionale 2 (ACE)	0.45	0.28
Congiunta	0.0675	0.0224
<i>A posteriori</i>	0.75	

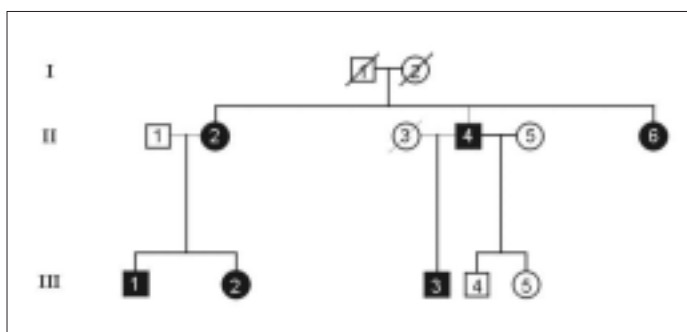
e risulterà dal calcolo che il probando ha un rischio molto elevato (75%) di sviluppare una CAD.

E' interessante notare che la probabilità *a priori* e l'informazione sul livello del colesterolo contribuiscono insieme all'86% del valore finale di rischio.

Qualora il probando fosse risultato positivo ad un test genetico con sensibilità e specificità analoghe a quelle del test per il colesterolo, nella medesima situazione avremmo ottenuto un rischio di 0.93!

Concludendo, le analisi di DNA certamente troveranno sempre maggiore applicazione in cardiologia, sia come supporto

Fig. 1. Albero genealogico di una famiglia nella quale segrega una malattia ereditata come carattere autosomico dominante. Gli individui III,4 e III,5 chiedono che venga valutato il loro rischio di sviluppare la malattia.



diagnostico che come elemento prognostico. Attualmente rappresentano un fertile terreno di ricerca, i cui risultati non possono e non devono essere utilizzati nella pratica clinica se non dopo accurate valutazioni, così come avviene per qualsiasi nuovo marcatore patologico. E' inoltre importante considerare che attualmente le indagini di genetica molecolare continuano ad essere relativamente costose e non sempre semplici.

Bibliografia

1. <http://www.ncbi.nlm.nih.gov/omim> (ultima consultazione 1 aprile 2005)
2. Cambien F, Poirier O, Lecerf L, Evans A, Cambou JP, Arveiler D, *et al.* Deletion polymorphism in the gene for angiotensin-converting enzyme is a potent predictor of myocardial infarction. *Nature* 1992; 359:641-4.
3. Danielli GA. *Genetics and Genomics for the Cardiologist*. Norwell MA, USA: Kluwer Acad. Publ.; 2002
4. Keavney B, McKenzie C, Parish S, Palmer A, Clark S, Youngman L, *et al.* Large-scale test of hypothesized associations between the Angiotensin-Converting-Enzyme insertion/deletion polymorphism and myocardial infarction in about 5,000 cases and 6,000 controls. *Lancet* 2000; 355:434-42.
5. Nanni L. *et al.* Lo studio delle malattie cardiache attraverso l'analisi dei profili di espressione genica. In: Amoroso A, Boncinelli E, Guarnieri G eds. *Cardiologia Molecolare: Nuove frontiere diagnostiche e terapeutiche*. Genova: Forum Service; 2004.