

REVIEW

GENOMICA DELLA CARDIOPATIA ISCHEMICA. DAGLI STUDI DI ASSOCIAZIONE GENOME-WIDE ALL'ERA DEL NEXT-GENERATION-SEQUENCING

Genomics of ischemic heart disease. From Genome Wide Association Studies to the Next-Generation Sequencing era

FEDERICA TOSI, NICOLA MARTINELLI, FABIANA BUSTI, ROBERTO CORROCHER, OLIVIERO OLIVIERI, DOMANICO GIRELLI

Dipartimento di Medicina, UOC di Medicina Generale a Indirizzo Immunoematologico ed Emocoagulativo, Università di Verona

SUMMARY

The importance of a positive family history for coronary artery disease (CAD) or myocardial infarction (MI) is a long-known notion to any practicing physician. However, as for any complex disease, our understanding of the genetic basis of CAD/MI has proven very difficult. The scenario has recently changed with the advent of *genomic* techniques able to obtain high-throughput unbiased information on the whole genome, at variance with traditional *genetic* studies focusing on single or few candidate genes. The Genome Wide Association Studies (GWAS) have revealed near sixty genetic loci significantly associated with CAD/MI. While the GWAS have undoubtedly represented a major breakthrough in CAD/MI genomics, suggesting new pathways (i.e. that driven by the 9p21.3 locus) and challenging some dogmas from observational epidemiology (i.e. downsizing of the causal role of HDL-cholesterol levels), they have explained only a fraction of CAD/MI heritability. The complementary application of the so-called Next Generation Sequencing (NGS) technologies is rapidly adding new information to solve the puzzle on CAD/MI "missing heritability". At present, a bottleneck is represented by our capability to interpret correctly sequencing data. This will require a strongest integration among molecular biologists, epidemiologists, statisticians, and, in particular, clinicians.

Keywords: coronary artery disease, Verona Heart Study, GWAS, Next Generation Sequencing.

Indirizzo per la corrispondenza

Domenico Girelli, MD PhD
Medicina Generale a Indirizzo
Immunoematologico ed Emocoagulativo,
Dipartimento di Medicina, Università di Verona,
Azienda Ospedaliera Universitaria
Integrata di Verona
e-mail: domenico.girelli@univr.it

Prevalenza della cardiopatia ischemica e ruolo dell'ereditarietà nel suo determinismo

Sin dalle fondamentali osservazioni dello studio di Framingham negli anni '60 del secolo scorso (1), che individuarono il ruolo dei lipidi plasmatici e degli altri "classici" fattori di rischio per malattie

cardiovascolari (Cardiovascular Diseases, CVD), la mortalità per tali patologie è gradualmente diminuita, ma rimane ancora inaccettabilmente elevata (2). Le ultime statistiche dell'Organizzazione Mondiale della Sanità confermano che le CVD globalmente considerate rappresentano la prima causa di morte sia nei paesi sviluppati che in quelli in via di sviluppo (<http://www.who.int/mediacentre/factsheets/fs317/en/>). Tra esse, i “killer” principali sono senz'altro la malattia coronarica (coronary artery disease, CAD) e la sua principale complicanza, l'infarto del miocardio, myocardial infarction (MI) (3).

Ogni medico pratico conosce per esperienza l'importanza di una storia familiare positiva per CAD o MI, ma il peso effettivo di tale dato anamnestico in relazione a quello determinato dai tradizionali fattori di rischio per CVD è stato definito solo recentemente. Il Framingham Offspring Study ha dimostrato, infatti, che l'aver almeno un genitore con CVD “precoce” (cioè insorta a un'età < 55 anni per il padre, e < 65 anni per la madre), implica un rischio quasi raddoppiato, anche in un modello corretto per tutti i tradizionali fattori di rischio (4). Tale risultato è stato sostanzialmente confermato dallo studio INTERHEART, che ha coinvolto oltre 25.000 soggetti (12.149 casi e 14.467 controlli) provenienti da ben 52 paesi (5). Gli studi su gemelli identici (monozigoti), noti come il migliore “esperimento della natura” per valutare la quota della variabilità fenotipica attribuibile a fattori genetici (comunemente designata con il termine di “ereditarietà”) (6), hanno fornito da altra angolazione ulteriori dati a supporto della stretta relazione tra patrimonio genetico e sviluppo di CAD/MI. Si è così stimato, attraverso la concordanza fenotipica, che l'ereditarietà influisca sull'espressione

della cardiopatia ischemica per una quota non inferiore al 40%, che può giungere in alcuni casi al 60% (7). Ciò peraltro non deve sorprendere, se si pensa che molti dei tradizionali fattori di rischio per CVD mostrano a loro volta una ereditarietà sostanziale, all'incirca del 50% per l'ipertensione (8) e del 60% per i livelli di colesterolo plasmatico (9).

Sebbene CAD e MI siano talvolta determinate da malattie monogeniche relativamente rare, come ad esempio l'ipercolesterolemia familiare (10) esse sono tipicamente “malattie complesse” multifattoriali, con un'architettura poligenica che interagisce con una serie di fattori di rischio ambientali e correlati allo stile di vita (11). Tale complessità ha ostacolato per molto tempo la comprensione delle basi genetiche. Lo scenario è radicalmente cambiato a partire dal 2005-2007 con l'avvento di tecnologie “omiche” (*genomica*), in grado cioè di ottenere grandi quantità di informazioni su tutto il genoma non influenzate da ipotesi a priori (*unbiased*). Si sono così potuti superare i limiti tipici dei tradizionali studi *genetici*, che si concentravano su singoli o pochi geni “candidati”, cioè selezionati preventivamente dai ricercatori. Di seguito riassumiamo brevemente i grandi progressi ottenuti con l'avvento degli studi di “*Genome Wide Association*” (GWAS), nonché le ulteriori prospettive derivanti dall'applicazione delle tecnologie cosiddette di *Next Generation Sequencing* (NGS).

Il Verona Heart Study (VHS)

La nostra esperienza in tale ambito deriva dal VHS, uno studio tuttora in corso e iniziato nel 1996 (12, 13), che ha arruolato fino ad oggi più di 2.500 individui. Il VHS è caratterizzato da un doppio disegno, caso-controllo e prospettico, ed è basato su

reperti angiografici. Quest'ultima caratteristica rende il VHS particolarmente adatto per studi genetici, laddove è essenziale una definizione rigorosa e oggettiva dei fenotipi di interesse. Ciò è particolarmente vero per la malattia coronarica, notoriamente caratterizzata da una lunga fase preclinica. Il gruppo di controllo del VHS è infatti costituito da soggetti sottoposti a coronarografia per motivi diversi da una sospetta CAD, che presentavano un albero coronarico completamente indenne da lesioni ateromasiche. Tale peculiarità previene il problema dell'erronea inclusione tra i controlli di soggetti con presenza di malattia coronarica, sebbene ancora asintomatica, come può talora avvenire negli studi in cui i controlli sono arruolati tra la popolazione generale apparentemente sana. Ciò ha fatto sì che il VHS sia stato recentemente incluso nei principali consorzi internazionali sorti allo scopo di approfondire la genomica di CAD e MI (14, 15), tra cui il Myocardial Infarction Genetics (MIGEN) (16, 17), il Coronary Artery Disease (O) Genomewide Replication And Meta-analysis (CARDIOGRAM) (18), il Triglyceride Coronary Disease Genetics Consortium (19) e, più recentemente, il MIGEN Exome-Array (MIGEN Ex-A) (20, 21), una sezione del poderoso NHLBI-ESP (National Heart, Lung, and Blood Institute Exome Sequencing Project; (<http://www.nhlbi.nih.gov/resources/exome.htm>)).

Cosa ci hanno insegnato i GWAS

Nella sua definizione più semplice, un GWAS è uno studio che utilizza *microarrays* contenenti fino a oltre 1.000.000 di varianti genetiche comuni (polimorfismi a singolo nucleotide, o SNPs, con frequenza dell'allele minore >5%, cioè almeno 1:20 nella popolazione) sparse sull'intero

genoma che permettono di associare la variabilità di determinate regioni genomiche (loci) ad un dato fenotipo (22). I GWAS, per come sono progettati, possono evidenziare geni verosimilmente implicati nella patogenesi di varie condizioni morbose, ma di regola non sono in grado di identificare le mutazioni effettivamente responsabili delle patologie medesime. In ogni caso, grazie alla possibilità di scoprire in modo *unbiased* nuovi geni, negli ultimi anni si è assistito a un aumento esponenziale di pubblicazioni basate sui GWAS, il cui elenco aggiornato è consultabile sul sito <http://www.genome.gov/gwastudies>.

Per quanto attiene più strettamente alla cardiopatia ischemica, i GWAS hanno sinora permesso l'identificazione di almeno 58 loci (23), dei quali i primi quindici per importanza sono riportati nella *Tabella 1*. È interessante notare che soltanto una frazione minoritaria di tali loci è associata ai tradizionali fattori per CVD (ad es. lipidi plasmatici, diabete mellito, ipertensione), mentre la maggior parte di essi (60%) include regioni del genoma per le quali i meccanismi predisponenti a CAD e/o MI rimangono tuttora poco chiari. Ciò peraltro non appare sorprendente se si considera che, come sopra menzionato, una storia familiare positiva rimane un predittore significativo di CAD/MI anche dopo la correzione per tutti i tradizionali fattori di rischio per CVD (4, 5). Da un punto di vista meccanicistico, questi loci rappresentano dunque un punto di partenza prezioso per ulteriori studi volti a individuare nuove vie patogenetiche coinvolte nella malattia aterotrombotica, con conseguente possibilità di aprire nuove prospettive terapeutiche (24).

Un esempio paradigmatico in tal senso è rappresentato dal locus 9p21.3, il primo a essere identificato dai GWAS (25),

Tabella I - Elenco dei principali IS loci associata CAD/MI identificati mediante GWAS.

Locus	SNP ID	Gene/i	Frequenza allelica (allele di rischio)	OR (95% CI)	Possibile meccanismo	Riferimento bibliografico
1p13.3	rs646776	CELSR2 PSRC1 SORT1	0,78 (A)	1,11 (1,08-1,15)	Regolazione dei livelli di LDL-colesterolo	(16)
1p32.2	rs11206510	PCSK9	0,91 (A)	1,17 (1,13-1,22)	Regolazione di LDLR	(16)
1q41	rs17465637	MIA3	0,74 (C)	1,14 (1,09-1,20)	sconosciuto	(16)
2q33.1	rs6725887	WDR12	0,15 (C)	1,14 (1,09-1,19)	sconosciuto	(16)
3q22.3	rs9818870	MRAS	0,18 (C)	1,12 (1,07-1,16)	Regolazione del signaling di adesione	(18)
6p24.1	rs12526453	PHACTR1	0,67 (C)	1,10 (1,06-1,13)	sconosciuto	(16)
6q26-q27	Haplotype based on 4 SNPs: rs2048327, rs3127599, rs7767084, rs10755578	SLC22A3-LPAL2-LPA Gene cluster	-	-	Modulazione dei livelli di lipoproteina _a	(18)
9p21.3	rs4977574	CDKN2A CDKN2B ANRIL	0,46 (G)	1,29 (1,23-1,36)	Regolazione della proliferazione cellulare	(16)
9q34.2	rs579459	ABO	0,21 (C)	1,10 (1,07-1,13)	sconosciuto	(17)
10q11.21	rs1746048	CXCL12	0,87 (C)	1,09 (1,07-1,13)	Chemochina che regola il traffico delle cellule progenitrici midollari stromali.	(16)
11q23.3	rs964184	ZNF259 APOA5/4/1 APOC3	0,13 (G)	1,13 (1,10-1,16)	sconosciuto	(18)
12q24.12	rs3184504	SH2B3	0,44 (T)	1,07 (1,04-1,10)	Regolazione del numero di eosinofili circolanti, attività anti-infiammatoria e anti-proliferativa	(15)
15q25.1	rs3825807	ADAMTS7	0,57 (A)	1,08 (1,06-1,10)	sconosciuto	(18)
19p13.2	rs1122608	LDLR	0,77 (G)	1,14 (1,09-1,18)	Regolazione dei livelli di LDL-colesterolo	(16)
21q22.11	rs9982601	SLC5A3 MRPS6 KCNE2	0,15 (T)	1,18 (1,12-1,24)	sconosciuto	(16)

e quello con l'associazione più forte con il fenotipo CAD. È curioso notare che il locus 9p21.3 non codifica per una proteina, bensì per un lungo RNA non codificante, designato con l'acronimo ANRIL (Antisense Non-coding RNA in the INK4 Locus) (23). D'altra parte, la recente pubblicazione dei dati del progetto ENCODE (ENCyclopedia Of Dna Elements) (26) ha dimostrato che una frazione rilevante del DNA non codificante per proteine, precedentemente considerato inerte dal punto di vista funzionale, se non addirittura inutile ("junk" DNA, o DNA "spazzatura"), viene in realtà trascritto dando origine a molteplici elementi regolatori, inclusi appunto i cosiddetti *long noncoding RNA*, in grado di modulare l'espressione dei geni canonici (27). Nonostante il ruolo esatto del locus 9p21.3 nel promuovere la patologia coronarica non sia stato ancora del tutto chiarito, i dati più recenti suggeriscono che ANRIL agisca modulando l'espressione di un gene adiacente (CDKN2B) codificante per una proteina in grado di inibire il ciclo cellulare (cyclin-dependent kinase inhibitor 2B - *cdkn2b*), espressa in numerosi tessuti, incluse le cellule muscolari lisce dei vasi (28, 29). Una modulazione inadeguata di *cdkn2b*, quale si verificerebbe negli individui portatori al locus 9p21.3 degli alleli correlati ad un aumento del rischio (per i quali circa un quinto della popolazione mondiale è omozigote), potrebbe promuovere la proliferazione cellulare nella pareti arteriose contribuendo infine alla formazione dell'ateroma (28).

Un altro esempio di locus identificato dai GWAS che potrebbe sottendere a inediti meccanismi di sviluppo della CAD si trova sul cromosoma 6p24.1, e contiene il gene codificante per PHACTR1 (Phosphatase and ACTin Regulator 1). Infatti, analogamente al 9p21.3, anche il locus 6p24.1 non pare associarsi ai fattori di ri-

schio tradizionali, ed il possibile ruolo nel processo di aterogenesi della proteina che ne deriva è oggetto di intensa attenzione da parte dei ricercatori (23).

Va sottolineato peraltro che, tra i numerosi loci identificati dai GWAS, anche quelli associati con i noti fattori di rischio per CVD hanno fornito nuove e significative informazioni circa i meccanismi fisiopatologici della CAD. A tal proposito, l'esempio più rilevante è rappresentato da PCSK9, un gene il cui prodotto è risultato fortemente implicato nel metabolismo dei recettori delle LDL (16, 18). PCSK9 (Protein Convertase Subtilisin/Kexin type 9) è, infatti, una serin proteasi secreta dal fegato in grado di legare il recettore LDL (LDLR) presente sulle membrane cellulari. Il legame LDLR/PCSK9 previene il riciclo del recettore favorendone la degradazione, esitando perciò in una ridotta clearance del colesterolo LDL (30).

Tali osservazioni hanno portato allo sviluppo di nuovi farmaci in grado di inibire PCSK9, che hanno dimostrato una rimarchevole efficacia terapeutica nel ridurre i livelli di colesterolo in aggiunta alle statine, nonché un profilo di sicurezza favorevole che ne ha determinato l'approvazione accelerata da parte delle agenzie regolatorie per l'imminente utilizzo nella pratica clinica (31-33).

Ulteriori utili informazioni sono derivate dall'utilizzo di punteggi di rischio genetico (*genetic risk scores* - GRS), calcolati contando il numero di alleli "a rischio" associati con elevati livelli di lipidi plasmatici (34). I GRS sono stati testati come predittori di eventi cardiovascolari, partendo dal presupposto che potrebbero stimare in modo più accurato l'esposizione individuale a determinati livelli di lipidi plasmatici, rispetto alla singola determinazione estemporanea con il classico dosaggio laboratoristico. In effetti, un GRS

per il colesterolo LDL ha dimostrato una capacità predittiva superiore ai livelli plasmatici nella coorte di Malmö (34), dato in seguito confermato da una meta-analisi di 17 studi che ha incluso più di 62.000 soggetti (35). Più recentemente, GRS analoghi sono stati proposti per discriminare i soggetti a maggiore probabilità di trarre beneficio dalla terapia con statine, sia in prevenzione primaria che in prevenzione secondaria (36).

La suddetta meta-analisi (35) ha inoltre dimostrato per i GRS associati ai livelli di trigliceridi e di HDL-colesterolo risultati opposti rispetto a quelli derivanti dai classici studi di epidemiologia osservazionale. Infatti, per molto tempo gli studi di epidemiologia classica hanno attribuito un ruolo protettivo indipendente ai livelli di colesterolo HDL, mentre il ruolo patogenetico dei trigliceridi è stato messo in discussione a causa della perdita di significatività statistica dopo correzione per i numerosi possibili fattori confondenti (37).

Al contrario, la meta-analisi di Holmes et al. suggerisce un ruolo causale per i trigliceridi ma non per il colesterolo HDL (35). Ciò appare in linea con i risultati indipendenti ottenuti da uno studio basato sul principio della randomizzazione mendeliana (38), nel quale una variante del gene *LIPG* significativamente associata con elevati livelli di colesterolo HDL, ma senza effetti sulle altre frazioni lipidiche, *non è risultata* associata ad un ridotto rischio di MI in una meta-analisi di 20 studi, con più di 116.000 partecipanti (39). Tali risultati, unitamente al fallimento di un trial clinico con dalcetrapib, un inibitore di CEPT in grado di aumentare i livelli di colesterolo HDL (40), hanno seriamente messo in discussione il “dogma” del cosiddetto “colesterolo buono” (41).

Nonostante l'importanza dei risultati

raggiunti, i GWAS presentano alcuni limiti non trascurabili, primo fra tutti il fatto che l'insieme delle varianti finora scoperte appare in grado di spiegare soltanto all'incirca poco più del 10% dell'ereditarietà dalla CAD (42). Ciò ha condotto a una profonda revisione della teoria cosiddetta “*common disease - common variants*”, inizialmente proposta quale spiegazione migliore per l'architettura genetica di fenotipi complessi quali CAD e MI (43). Sempre più numerose infatti sono le dimostrazioni che sia mutazioni mendeliane rare (con *minor allele frequency* - MAF <1:1000), che varianti a bassa frequenza (MAF >1: 1000 e <1:20), contribuiscano a determinare fenotipi cardiovascolari complessi insieme alle varianti comuni (MAF > 1:20) (10). Come accennato in precedenza, solo queste ultime varianti comuni sono potenzialmente identificabili dai GWAS, mentre le altre, specie le più rare, richiedono tecnologie di sequenziamento del genoma.

L'era del Next Generation Sequencing (NGS)

Negli ultimi anni abbiamo assistito a un progresso esponenziale delle tecnologie di sequenziamento del DNA, che, di pari passo con una sostanziale riduzione dei costi, pone tali metodiche nella condizione di poter diventare parte integrante della pratica clinica in un prossimo futuro. Non è possibile in questa sede una descrizione esaustiva delle numerose tecnologie di sequenziamento genomico, per le quali si rimanda il lettore a un paio di ottime recenti rassegne (44, 45). Paradossalmente, il problema principale al momento non riguarda le tecnologie di NGS *per sé*, ma piuttosto gli strumenti analitici per interpretare correttamente l'enorme mole di dati genomici così ottenuti. È infatti ormai sempre più chiaro che la va-

riazione individuale del genoma umano è molto più elevata di quanto inizialmente ipotizzato, e che la nostra capacità di distinguere le varianti che contribuiscono a causare le varie malattie è da considerarsi in una fase poco più che iniziale (45). Gli studi che utilizzano il sequenziamento dell'intero genoma (*whole genome sequencing* - WGS) o quello del cosiddetto "esoma" (*whole exome sequencing* - WES), cioè "limitato" alle sequenze codificanti dei ≈ 23.000 geni codificanti proteine (corrispondenti a $\approx 1\%$ del genoma umano totale) hanno stimato che ogni singolo individuo "sano" presenti circa 20.000-25.000 varianti genetiche. Quasi la metà di queste varianti sono "non-sinonime", e ben 40-100 di esse con probabili effetti funzionali significativi (incluse mutazioni deleterie del tipo "non-senso", dei siti di splicing, e inserzioni/delezioni determinanti frameshift) potrebbero potenzialmente causare una patologia mendeliana (45). Da ciò ne consegue l'urgente necessità di affinare l'interpretazione dei dati di sequenziamento, in particolare in ambito clinico, mediante l'utilizzo di linee-guida sempre più stringenti (46).

La ricerca di varianti rare associate a malattie complesse è ancora più difficile che per le malattie mendeliane, richiedendo non solo una corretta interpretazione del significato biologico di una determinata variante, ma anche complessi e sofisticati approcci dal punto di vista statistico e bioinformatico (47, 48). Per quanto riguarda la cardiopatia ischemica, un buon punto di partenza si è rivelato essere l'analisi approfondita dei geni per i quali varianti umane "knock-out" sono state precedentemente dimostrate in popolazioni isolate.

Un paradigma in questo senso è rappresentato dalla mutazione nulla (R19X) nel gene *APOC3* riscontrata negli Amish Lancaster (49). L'apolipoproteina C3 è nota

per essere uno dei maggiori regolatori in senso negativo della lipoproteina lipasi (LPL), l'enzima limitante la clearance dei trigliceridi plasmatici attraverso l'idrolisi delle lipoproteine ricche in trigliceridi, quali VLDL e chilomicroni (50). Portatori eterozigoti per la variante *APOC3* R19X (circa il 5% della popolazione Amish) presentavano livelli sub-normali di trigliceridi a digiuno e post-prandiali, così come minori evidenze di aterosclerosi subclinica (misurata mediante TC coronarica) rispetto alla restante popolazione. Ciò ha rinforzato i sospetti sul possibile ruolo causale di trigliceridi nella CAD (v. sopra per quanto riguarda i GWAS), suggerendo la via metabolica LPL/APOC3/APOA5 (*Figura 1*) quale bersaglio prioritario per gli studi di sequenziamento. In effetti, uno studio collaborativo coordinato dal National Heart, Lung and Blood Institute (NHLBI) ha documentato non solo un'associazione significativa tra mutazioni rare di *APOC3* (in particolare 3 mutazioni *loss of function* - R19X, IVS2+1G->A e IVS3+1G->T - e una mutazione *missense* - A43T) e ridotti livelli plasmatici di trigliceridi e apolipoproteina C3, ma anche che i portatori eterozigoti di almeno una di queste mutazioni (circa 1:150 nella popolazione studiata di 110.970 soggetti) risultavano efficacemente protetti dallo sviluppo di CAD (51).

Ciò è anche in accordo con precedenti studi del nostro gruppo, che indicavano i livelli circolanti di apolipoproteina C3 come un forte e indipendente predittore di mortalità totale e cardiovascolare nella popolazione del VHS (52).

Il VHS è stato incluso come coorte di replica anche nel grande NHLBI-Exome Sequencing Project, che comprendeva una fase preliminare di WES su circa 1.000 pazienti con infarto miocardico precoce e altrettanti controlli con fenotipo

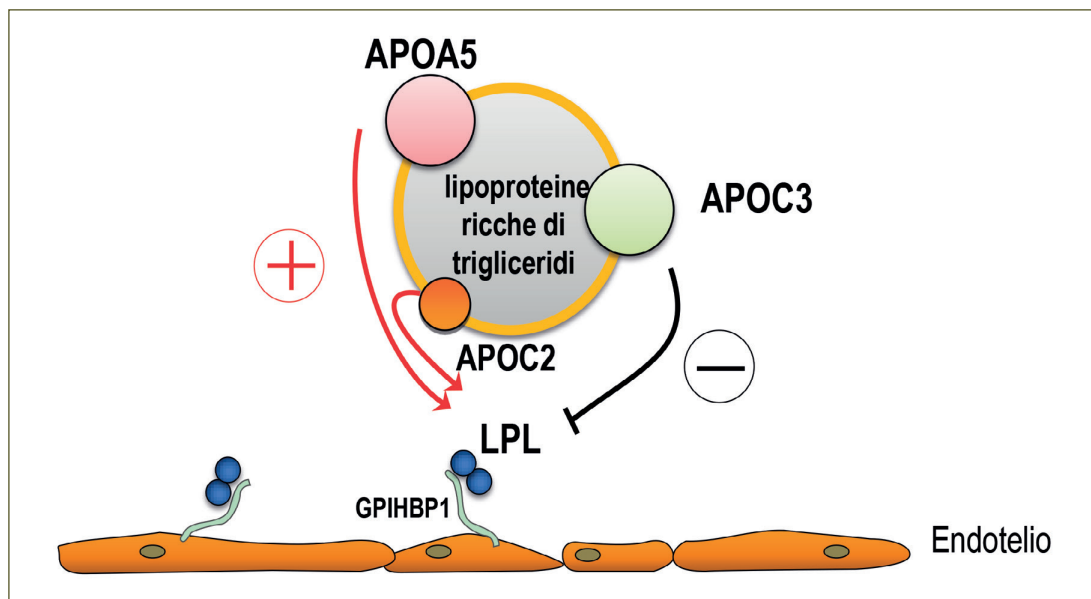


Figura 1 - Rappresentazione schematica del pathway LPL-APOC3-APOA5. La lipoproteinlipasi (LPL), un enzima chiave nel ridurre i trigliceridi plasmatici, è legata all'endotelio vascolare attraverso la proteina GPIHBP1. LPL agisce idrolizzando i trigliceridi contenuti nelle lipoproteine ricche in trigliceridi (VLDL e chilomicroni), ed è controllata principalmente da APOC3 e APOA5. APOC3 inibisce LPL, mentre all'opposto APOA5 ne stimola l'attività. Mutazioni con perdita di funzione di APOC3 sono associate a bassi livelli plasmatici di trigliceridi, e, di conseguenza, risultano protettive verso lo sviluppo della CAD (50). Viceversa, mutazioni di APOA5 esitano in una riduzione di attività della LPL, aumentati livelli di trigliceridi, e aumentato rischio per CAD/MI (vedi testo). Anche APOC2 può attivare LPL ed le mutazioni di APOC2, come anche il deficit di LPL, sono storicamente noti come cause delle iperlipoproteinemie di tipo I (iperchilomicronemie).

“opposto”, cioè privi di coronaropatia in età avanzata.

Il progetto, recentemente completato, sta fornendo i primi risultati che appaiono utilmente complementari a quelli generati dai GWAS.

Un esempio è rappresentato dal locus APOA5, per il quale i GWAS avevano identificato varianti comuni associate con i livelli di trigliceridi e con storia di MI precoce (16).

Il sequenziamento “massivo” di APOA5 in 13.432 soggetti ha evidenziato anche in questo caso la presenza di mutazioni rare in grado di influenzare in modo opposto a quanto osservato per APOC3 i livelli di trigliceridi ed il rischio di MI precoce (entrambi aumentati nei portatori) (53). Ana-

logo è, infine, il caso del locus *DYRK1B*, inizialmente individuato dai GWAS condotti sul diabete mellito di tipo 2 e sulla sindrome metabolica (54).

Recentemente, Keramati et al. hanno identificato tre grandi pedigree caratterizzati da CAD precoce, e franca sindrome metabolica per compresenza di obesità centrale, ipertensione e diabete (55).

Attraverso l'uso combinato di analisi di *linkage* e di WES, gli Autori hanno identificato mutazioni patogenetiche di *DYRK1B*, dimostrando che tale gene codifica per una proteina coinvolta nell'adipogenesi e nell'omeostasi del glucosio attraverso la via di signaling *Wnt* e la regolazione dell'enzima glucosio-6-fosfatasi (55).

Conclusioni

L'applicazione di moderne metodologie per lo studio del genoma quali GWAS e NGS ha consentito negli anni più recenti un grande avanzamento delle conoscenze delle basi genetiche della CAD, con ricadute pratiche finanche in ambito terapeutico (si veda l'esempio di PCSK9). Ciò nonostante, a tutt'oggi siamo ancora lontani da una piena comprensione del problema, come dimostrato dal fatto che siamo in grado di spiegare solo una frazione minore dell'ereditarietà complessiva. La delucidazione della cosiddetta "*missing heritability*" rappresenta ancora una sfida molto impegnativa, che richiede non solo importanti risorse tecnologiche e computazionali (56), ma anche campioni

di notevoli dimensioni (decine di migliaia, come minimo) e, in particolare, una migliore fenotipizzazione (57). Si pensi, ad esempio, che la maggior parte degli studi pubblicati finora ha considerato insieme la coronaropatia aterosclerotica e l'infarto miocardico. Tale approssimazione, seppure strumentalmente utile per aumentare la numerosità delle casistiche ai livelli richiesti dai GWAS, non è intrinsecamente in grado di distinguere tra i meccanismi coinvolti nello sviluppo dell'ateroma (CAD) e quelli che conducono alle complicanze trombotiche (MI), verosimilmente solo in parte sovrapponibili. Una più forte integrazione tra i biologi molecolari, epidemiologi, statistici e clinici, rappresenterà probabilmente l'elemento fondamentale per risolvere il problema in futuro.

RIASSUNTO

La malattia coronarica (coronary artery disease, CAD) e l'infarto miocardico (myocardial infarction, MI), sua principale complicanza, rappresentano tuttora le principali cause di morte nel mondo. Ogni clinico ha ben presente l'importanza di una storia familiare positiva per CAD quale fattore di rischio e studi su gemelli monozigoti hanno stimato che l'ereditarietà incide per il 40-60% nel determinismo della CAD. Ciò nonostante, le basi genetiche della CAD sono rimaste in gran parte oscure sino a pochi anni fa, analogamente alle altre cosiddette "malattie complesse".

Lo scenario è tuttavia sostanzialmente cambiato negli ultimi 5-7 anni con l'avvento delle tecnologie "*genomiche*", in grado cioè di scansionare l'intero genoma umano e di fornire una messe di dati indipendenti da ipotesi a priori. Gli studi di associazione "*genome-wide*" (GWAS) hanno individuato almeno una sessantina di regioni genomiche (loci) associate alla CAD con evidenza statistica mai raggiunta dai precedenti studi su "geni candidati".

Più della metà di tali loci non sono riconducibili ai fattori di rischio tradizionali di CAD/MI, e pertanto suggeriscono possibili nuove vie patogenetiche finanche all'individuazione di nuovi bersagli terapeutici. Paradigmatico in tal senso è il locus 9p21.3, che rappresenta, a oggi, il fattore genetico per CAD maggiormente replicato e con livello indiscutibile di evidenza. È importante rilevare che il locus 9p21.3 non contiene alcun gene codificante proteine, apparendo bensì legato alla produzione di un lungo RNA appartenente alla categoria dei "*long noncoding RNAs*", e perciò dotato di proprietà regolatorie su geni collocati a varia distanza nel genoma. Studi recenti suggeriscono che il locus 9p21.3 codifichi per ANRIL (*antisense noncoding RNA in the INK4 locus*), in grado di controllare il gene CDKN2B, a sua volta codificante per l'inibitore di una chinasi ciclino-dipendente espressa nelle pareti arteriose e implicata nella regolazione della proliferazione cellulare.

D'altra parte, anche i loci associati a fattori di rischio tradizionali, quali ad es. i lipidi plasmatici, hanno fornito informazioni assai importanti. Studi di "*randomizzazione mendeliana*", basati su punteggi genetici associabili alle diverse frazioni lipidiche, hanno infatti messo in dubbio il ruolo patogenetico dei livelli di colesterolo HDL, mentre hanno concretamente rivalutato i trigliceridi. Questa apparente

sconfessione di alcuni “dogmi” derivanti dalla classica epidemiologia osservazionale è stata recentemente confermata da trial clinici che hanno dimostrato la futilità di farmaci disegnati per incrementare i livelli di colesterolo HDL.

I GWAS hanno tuttavia almeno un paio di importanti limitazioni:

- 1) per disegno sperimentale, individuano soltanto varianti comuni nel genoma (frequenza allelica, f.a. >5%), di impatto funzionale generalmente modesto;
- 2) i loci sinora scoperti sono in grado di spiegare soltanto una frazione minimale dell'ereditarietà della CAD.

La cosiddetta ipotesi delle “varianti comuni nelle malattie comuni”, formulata una decina d'anni fa, è stata pertanto sostanzialmente ridimensionata, e attualmente si pensa che le basi genetiche della CAD, come nel caso delle altre “malattie complesse”, siano dovute al contributo collettivo di varianti comuni (f.a. >5%), a bassa frequenza (f.a. compresa tra 0,01% e 5%), e rare (f.a. <0,01%). Queste ultime, in particolare, non possono che essere rilevate mediante sequenziamento del DNA. Pertanto, le tecnologie di sequenziamento di ultima generazione (*next generation sequencing*, NGS) contribuiranno in modo sostanziale nei prossimi anni allo svelamento delle basi genetiche della CAD. Va peraltro rimarcato che, a fronte di uno straordinario progresso nelle tecnologie NGS, la capacità di interpretare correttamente il significato clinico dei dati da esse prodotte è ancora ad uno stadio poco più che iniziale. A tal proposito, è più che mai urgente il recupero di un ruolo centrale dei clinici in ricerche di questo tipo, quali componenti integranti di *team* multidisciplinari in collaborazione con epidemiologi, biotecnologi, bioinformatici, e statistici.

Parole chiave: *cardiopatia ischemica, Verona Heart Study, GWAS, Next Generation Sequencing.*

Bibliografia

1. Kannel WB, Dawber TR, Kagan A, Revotskie N, Stokes J, 3rd. Factors of risk in the development of coronary heart disease—six year follow-up experience. The Framingham Study. *Annals of internal medicine.* 1961; 55: 33-50.
2. Nabel EG, Braunwald E. A tale of coronary artery disease and myocardial infarction. *The New England journal of medicine.* 2012; 366: 54-63.
3. Wong ND. Epidemiological studies of CHD and the evolution of preventive cardiology. *Nature reviews Cardiology.* 2014; 11: 276-89.
4. Lloyd-Jones DM, Nam BH, D'Agostino RB, Sr., Levy D, Murabito JM, Wang TJ, et al. Parental cardiovascular disease as a risk factor for cardiovascular disease in middle-aged adults: a prospective study of parents and offspring. *JAMA: the journal of the American Medical Association.* 2004; 291: 2204-11.
5. Chow CK, Islam S, Bautista L, Rumboldt Z, Yusufali A, Xie C, et al. Parental history and myocardial infarction risk across the world: the INTERHEART Study. *Journal of the American College of Cardiology.* 2011; 57: 619-27.
6. Martin N, Boomsma D, Machin G. A twin-pronged attack on complex traits. *Nature genetics.* 1997; 17: 387-92.
7. Zdravkovic S, Wienke A, Pedersen NL, Marenberg ME, Yashin AI, de Faire U. Genetic influences on CHD-death and the impact of known risk factors: comparison of two frailty models. *Behavior genetics.* 2004; 34: 585-92.
8. Levy D, DeStefano AL, Larson MG, O'Donnell CJ, Lifton RP, Gavras H, et al. Evidence for a gene influencing blood pressure on chromosome 17. Genome scan linkage results for longitudinal blood pressure phenotypes in subjects from the framingham heart study. *Hypertension.* 2000; 36: 477-83.
9. Knoblauch H, Busjahn A, Munter S, Nagy Z, Faulhaber HD, Schuster H, et al. Heritability analysis of lipids and three gene loci in twins link the macrophage scavenger receptor to HDL cholesterol concentrations. *Arteriosclerosis, thrombosis, and vascular biology.* 1997; 17: 2054-60.
10. Kathiresan S, Srivastava D. Genetics of human cardiovascular disease. *Cell.* 2012; 148: 1242-57.
11. Watkins H, Farrall M. Genetic susceptibility to coronary artery disease: from promise to progress. *Nat Rev Genet.* 2006; 7: 163-73.
12. Girelli D, Friso S, Trabetti E, Olivieri O, Russo C, Pessotto R, et al. Methylentetrahydrofolate reductase C677T mutation, plasma homocysteine, and folate in subjects from northern Italy with or without angiographically documented severe coronary atherosclerotic disease: evidence for an important genetic-environmental interaction. *Blood.* 1998; 91: 4158-63.
13. Girelli D, Russo C, Ferraresi P, Olivieri O, Pinot-

- ti M, Friso S, et al. Polymorphisms in the factor VII gene and the risk of myocardial infarction in patients with coronary artery disease. *The New England journal of medicine*. 2000; 343: 774-80.
14. Gretarsdottir S, Baas AF, Thorleifsson G, Holm H, den Heijer M, de Vries JP, et al. Genome-wide association study identifies a sequence variant within the DAB2IP gene conferring susceptibility to abdominal aortic aneurysm. *Nature genetics*. 2010; 42: 692-7.
 15. Gudbjartsson DF, Bjornsdottir US, Halapi E, Helgadóttir A, Sulem P, Jonsdóttir GM, et al. Sequence variants affecting eosinophil numbers associate with asthma and myocardial infarction. *Nature genetics*. 2009; 41: 342-7.
 16. Myocardial Infarction Genetics C, Kathiresan S, Voight BF, Purcell S, Musunuru K, Ardissino D, et al. Genome-wide association of early-onset myocardial infarction with single nucleotide polymorphisms and copy number variants. *Nature genetics*. 2009; 41: 334-41.
 17. Reilly MP, Li M, He J, Ferguson JF, Stylianou IM, Mehta NN, et al. Identification of ADAMTS7 as a novel locus for coronary atherosclerosis and association of ABO with myocardial infarction in the presence of coronary atherosclerosis: two genome-wide association studies. *Lancet*. 2011; 377: 383-92.
 18. Schunkert H, König IR, Kathiresan S, Reilly MP, Assimes TL, Holm H, et al. Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nature genetics*. 2011; 43: 333-8.
 19. Triglyceride Coronary Disease Genetics C, Emerging Risk Factors C, Sarwar N, Sandhu MS, Ricketts SL, Butterworth AS, et al. Triglyceride-mediated pathways and coronary disease: collaborative analysis of 101 studies. *Lancet*. 2010; 375: 1634-9.
 20. Peloso GM, Auer PL, Bis JC, Voorman A, Morrison AC, Stitzel NO, et al. Association of low-frequency and rare coding-sequence variants with blood lipids and coronary heart disease in 56,000 whites and blacks. *Am J Hum Genet*. 2014; 94: 223-32.
 21. Stitzel NO, Fouchier SW, Sjouke B, Peloso GM, Moscoso AM, Auer PL, et al. Exome sequencing and directed clinical phenotyping diagnose cholesterol ester storage disease presenting as autosomal recessive hypercholesterolemia. *Arteriosclerosis, thrombosis, and vascular biology*. 2013; 33: 2909-14.
 22. McCarthy MI, Abecasis GR, Cardon LR, Goldstein DB, Little J, Ioannidis JP, et al. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat Rev Genet*. 2008; 9: 356-69.
 23. Holdt LM, Teupser D. From genotype to phenotype in human atherosclerosis—recent findings. *Current opinion in lipidology*. 2013; 24: 410-8.
 24. Girelli D, Martinelli N, Peyvandi F, Olivieri O. Genetic architecture of coronary artery disease in the genome-wide era: implications for the emerging “golden dozen” loci. *Seminars in thrombosis and hemostasis*. 2009; 35: 671-82.
 25. McPherson R, Pertsemlidis A, Kavaslar N, Stewart A, Roberts R, Cox DR, et al. A common allele on chromosome 9 associated with coronary heart disease. *Science*. 2007; 316: 1488-91.
 26. Consortium EP, Bernstein BE, Birney E, Dunham I, Green ED, Gunter C, et al. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 2012; 489: 57-74.
 27. Morris KV, Mattick JS. The rise of regulatory RNA. *Nat Rev Genet*. 2014.
 28. Kojima Y, Downing K, Kundu R, Miller C, Dewey F, Lancero H, et al. Cyclin-dependent kinase inhibitor 2B regulates efferocytosis and atherosclerosis. *The Journal of clinical investigation*. 2014; 124: 1083-97.
 29. Visel A, Zhu Y, May D, Afzal V, Gong E, Attanasio C, et al. Targeted deletion of the 9p21 non-coding coronary artery disease risk interval in mice. *Nature*. 2010; 464: 409-12.
 30. Seidah NG, Awan Z, Chretien M, Mbikay M. PCSK9: a key modulator of cardiovascular health. *Circulation research*. 2014; 114: 1022-36.
 31. Blom DJ, Hala T, Bolognese M, Lillestol MJ, Toth PD, Burgess L, et al. A 52-week placebo-controlled trial of evolocumab in hyperlipidemia. *The New England journal of medicine*. 2014; 370: 1809-19.
 32. Sabatine MS, Giugliano RP, Wiviott SD, Raal FJ, Blom DJ, Robinson J, et al. Efficacy and safety of evolocumab in reducing lipids and cardiovascular events. *The New England journal of medicine*. 2015; 372: 1500-9.
 33. Robinson JG, Farnier M, Krempf M, Bergeron J, Luc G, Averna M, et al. Efficacy and safety of alirocumab in reducing lipids and cardiovascular events. *The New England journal of medicine*. 2015; 372: 1489-99.
 34. Kathiresan S, Melander O, Anevski D, Guiducci C, Burt NP, Roos C, et al. Polymorphisms associated with cholesterol and risk of cardiovascular events. *The New England journal of medicine*. 2008; 358: 1240-9.
 35. Holmes MV, Asselbergs FW, Palmer TM, Drenos F, Lanktree MB, Nelson CP, et al. Mendelian randomization of blood lipids for coronary heart disease. *European heart journal*. 2015; 36: 539-50.
 36. Mega JL, Stitzel NO, Smith JG, Chasman DI, Caulfield MJ, Devlin JJ, et al. Genetic risk, coro-

- nary heart disease events, and the clinical benefit of statin therapy: an analysis of primary and secondary prevention trials. *Lancet*. 2015; 385: 2264-71.
37. Emerging Risk Factors C, Di Angelantonio E, Sarwar N, Perry P, Kaptoge S, Ray KK, et al. Major lipids, apolipoproteins, and risk of vascular disease. *JAMA : the journal of the American Medical Association*. 2009; 302: 1993-2000.
 38. Hingorani A, Humphries S. Nature's randomised trials. *Lancet*. 2005; 366: 1906-8.
 39. Voight BF, Peloso GM, Orho-Melander M, Frikke-Schmidt R, Barbalic M, Jensen MK, et al. Plasma HDL cholesterol and risk of myocardial infarction: a mendelian randomisation study. *Lancet*. 2012; 380: 572-80.
 40. Schwartz GG, Olsson AG, Abt M, Ballantyne CM, Barter PJ, Brumm J, et al. Effects of dalcetrapib in patients with a recent acute coronary syndrome. *The New England journal of medicine*. 2012; 367: 2089-99.
 41. Kritharides L. Not so "good" cholesterol. *Bmj*. 2014; 349: g4664.
 42. Consortium CAD, Deloukas P, Kanoni S, Willenborg C, Farrall M, Assimes TL, et al. Large-scale association analysis identifies new risk loci for coronary artery disease. *Nature genetics*. 2013; 45: 25-33.
 43. Becker KG. The common variants/multiple disease hypothesis of common complex genetic disorders. *Medical hypotheses*. 2004; 62: 309-17.
 44. Johnsen JM, Nickerson DA, Reiner AP. Massively parallel sequencing: the new frontier of hematologic genomics. *Blood*. 2013; 122: 3268-75.
 45. Gonzaga-Jauregui C, Lupski JR, Gibbs RA. Human genome sequencing in health and disease. *Annual review of medicine*. 2012; 63: 35-61.
 46. MacArthur DG, Manolio TA, Dimmock DP, Rehm HL, Shendure J, Abecasis GR, et al. Guidelines for investigating causality of sequence variants in human disease. *Nature*. 2014; 508: 469-76.
 47. Liu DJ, Peloso GM, Zhan X, Holmen OL, Zawistowski M, Feng S, et al. Meta-analysis of gene-level tests for rare variant association. *Nature genetics*. 2014; 46: 200-4.
 48. Do R, Kathiresan S, Abecasis GR. Exome sequencing and complex disease: practical aspects of rare variant association studies. *Human molecular genetics*. 2012; 21: R1-9.
 49. Pollin TI, Damcott CM, Shen H, Ott SH, Shelton J, Horenstein RB, et al. A null mutation in human APOC3 confers a favorable plasma lipid profile and apparent cardioprotection. *Science*. 2008; 322: 1702-5.
 50. Kersten S. Physiological regulation of lipoprotein lipase. *Biochim Biophys Acta*. 2014.
 51. Tg, Hdl Working Group of the Exome Sequencing Project NHL, Blood I, Crosby J, Peloso GM, Auer PL, et al. Loss-of-function mutations in APOC3, triglycerides, and coronary disease. *The New England journal of medicine*. 2014; 371: 22-31.
 52. Olivieri O, Martinelli N, Girelli D, Pizzolo F, Friso S, Beltrame F, et al. Apolipoprotein C-III predicts cardiovascular mortality in severe coronary artery disease and is associated with an enhanced plasma thrombin generation. *Journal of thrombosis and haemostasis: JTH*. 2010; 8: 463-71.
 53. Do R, Stitzel NO, Won HH, Jorgensen AB, Duga S, Angelica Merlini P, et al. Exome sequencing identifies rare LDLR and APOA5 alleles conferring risk for myocardial infarction. *Nature*. 2015; 518: 102-6.
 54. Kraja AT, Vaidya D, Pankow JS, Goodarzi MO, Assimes TL, Kullo IJ, et al. A bivariate genome-wide approach to metabolic syndrome: STAMPEED consortium. *Diabetes*. 2011; 60: 1329-39.
 55. Keramati AR, Fathzadeh M, Go GW, Singh R, Choi M, Faramarzi S, et al. A form of the metabolic syndrome associated with mutations in DYRK1B. *The New England journal of medicine*. 2014; 370: 1909-19.
 56. Zuk O, Schaffner SF, Samocha K, Do R, Hechter E, Kathiresan S, et al. Searching for missing heritability: designing rare variant association studies. *Proc Natl Acad Sci USA*. 2014; 111: E455-64.
 57. Robinson MR, Wray NR, Visscher PM. Explaining additional genetic variation in complex traits. *Trends in genetics: TIG*. 2014; 30: 124-32.