

GENETICA

SUSCETTIBILITÀ GENETICA ALLE MALATTIE CARDIOVASCOLARI

LAURA D'ERASMO, MARCELLO ARCA**Introduzione**

L'ereditarietà della malattia cardiovascolare aterosclerotica (*atherosclerotic cardiovascular disease, ASCVD*) è stata a lungo postulata sulla base dell'osservazione della presenza di malattia coronarica (MC) in più membri dello stesso nucleo familiare, soprattutto quando la malattia si presentava precocemente (1). Nel Framingham Heart Study Offspring Cohort, la presenza nella famiglia di un genitore con storia di MC prematura si associava ad un rischio circa due volte superiore di sviluppare malattia cardiovascolare nei figli indipendentemente dagli altri fattori di rischio tradizionali, suggerendo così la presenza di un *background* genetico responsabile dell'insorgenza della malattia cardiovascolare (1). L'arricchimento di varianti genetiche deleterie e la condivisione di fattori di rischio ambientali (quali comportamenti legati allo stile di vita, reddito dei genitori, scolarità, e tipo di quartiere) sarebbero pertanto responsabili dell'aggregazione di ASCVD all'interno dei gruppi familiari. Studi condotti su famiglie ad alto rischio o nelle popolazioni di gemelli hanno approssimato che l'ereditarietà della MC, definita specificatamente come la percentuale totale della variazione fenotipica spiegata da fattori genetici, è di circa il 40-60% (2) e, a livello individuale, il contributo ereditario

dei fattori di rischio è maggiore tra gli individui nei quali la MC insorge prematuramente (1). Tuttavia, studi genetici in pazienti con ASCVD o con ipercolesterolemia severa hanno ripetutamente dimostrato che solo una frazione di individui affetti sono portatori di varianti rare monogeniche (3). Al contrario, una grande percentuale del rischio genetico sembra derivare dall'essere portatori di un accumulo di varianti genetiche comuni. Individualmente queste varianti avrebbero uno scarso effetto sul tratto d'interesse, ma aggregandosi sarebbero in grado di esercitare degli effetti clinicamente rilevanti (3).

Grazie allo sviluppo delle tecniche di sequenziamento di nuova generazione (*next generation sequencing o NGS*), si è resa possibile la quantificazione dell'effetto di questi determinanti genetici comuni nella eziopatogenesi della malattia cardiovascolare e dei fenotipi lipidici permettendo la creazione di *score* di rischio genetico (*genetic risk score, GRS*). Con il termine GRS si intende dunque un sistema a punteggio basato sull'accumulo di varianti a singolo nucleotide (SNPs) identificate dagli studi di *genome-wide association (GWAS)* come associate al tratto di interesse (4). Nonostante la valutazione simultanea di diversi SNPs venga spesso chiamata con il termine generico di '*genetic risk score*', recentemente è diventato sempre più frequente

l'utilizzo di *'polygenic risk score (PRS)'* a sottolineare il contributo simultaneo di multiple regioni genomiche. Gli studi *GWAS* hanno recentemente permesso di includere in questi *score* milioni di *SNPs* così da creare dei *PRS* di dimensioni sempre maggiori. Con il termine *'omnigenic risk score (ORS)'* si identificano pertanto quegli *score* di rischio che includono numeri particolarmente elevati di varianti direttamente o indirettamente associate con il tratto fenotipico di interesse (4).

È importante notare come negli ultimi anni i *GRS* vengono sempre più utilizzati per la previsione clinica del rischio, la diagnosi e la stratificazione dell'efficacia delle singole terapie andando a rivestire un ruolo sempre più rilevante nella comprensione dell'architettura genetica che è alla base di diverse patologie, tra cui anche la malattia cardiovascolare.

Determinanti monogenici della MC

Come detto in precedenza, le indagini sui determinanti genetici della malattia cardiovascolare hanno evidenziato la presenza di una architettura genetica complessa. Tuttavia, per alcuni individui o famiglie, il rischio genetico segue i classici pattern di trasmissione mendeliana in virtù dei quali una malattia si manifesta precocemente nella vita e il contributo degli altri fattori di rischio ambientali è piuttosto limitato (4). In questi casi, il rischio è mediato da mutazioni monogeniche rare ad alto impatto sul fenotipo, cioè con una frequenza dell'allele minore (*MAF*) nella popolazione generale minore dell'1% (4). Un esempio di queste mutazioni sono quelle responsabili dell'ipercolesterolemia familiare (FH), una sindrome clinica caratterizzata da una marcata elevazione delle concentrazioni sieriche del colesterolo delle lipoproteine a bassa densità (LDL-C) ed una

serie di segni clinici distintivi quali l'arco corneale precoce, la xantomatosi tendinea e soprattutto, la presenza di malattia coronarica (MC) precoce. L'FH segue un pattern di trasmissione autosomico-dominante in cui il numero degli alleli anomali ereditati (0, 1 o 2) determina la gravità del fenotipo clinico (5). Studi di *linkage* condotti in pazienti affetti da FH hanno permesso di identificare i geni responsabili del fenotipo FH ovvero quelli che codificano per il recettore dell'LDL (LDLR), apolipoproteina B (APOB), e la *proprotein convertase subtilisin/kexin type 9* (PCSK9). Mutazioni patologiche rare a carico di questi geni interferiscono tutte, direttamente o indirettamente, con l'*uptake* epatico delle particelle LDL circolanti determinandone un considerevole aumento dei livelli (5). L'FH deve essere sospettata in tutte quelle condizioni di ipercolesterolemia severa cui si associano i segni clinici caratteristici della malattia, facendosi eventualmente guidare nella diagnosi da uno degli *score* predittivi attualmente disponibili in letteratura quali il sistema di punteggio del Dutch Lipid Clinic Network Score (*DLCNS*) (5). La presenza di MC prematura è un fenomeno ormai accertato nei pazienti FH nei quali l'età media di insorgenza dei sintomi coronarici si è dimostrata essere molto più precoce che nella popolazione generale (età media di esordio a 45 anni nell'uomo e 55 anni nella donna) (5). Il rischio e l'età di insorgenza dell'aterosclerosi nelle persone affette da FH monogenica tendono ad essere proporzionali all'entità e alla durata dell'aumento dell'LDL-C, cioè quello che viene attualmente definito come il carico di esposizione cumulativa alla colesterolemia detto anche *cholesterol burden* (5). Ciò viene facilmente spiegato dal fatto che i pazienti FH presentano livelli di colesterolo elevati sin dalla nascita con aumento del *cholesterol burden* e conseguentemente,

del rischio cardiovascolare. Inoltre, sia studi genetici che studi di intervento farmacologici randomizzati-controllati hanno dimostrato in modo consistente che la riduzione della concentrazione di particelle di LDL-C nel plasma riduce il rischio di MC proporzionalmente sia alla loro riduzione assoluta che alla durata di esposizione a bassi livelli, indipendentemente dal meccanismo mediante il quale viene ottenuta (6). Dunque, l'identificazione precoce dei soggetti FH è particolarmente rilevante in quanto l'intervento farmacologico precoce con statine è in grado di ridurre il rischio di incidenza di MC in questa categoria di soggetti cambiando drasticamente la storia naturale della malattia (5). Tuttavia, nonostante si tratti di una patologia relativamente comune (≈ 1 in 300-500 individui), una causa monogenica di FH viene rilevata solo nel 30-80% dei pazienti con FH diagnosticata clinicamente, mentre nella restante parte non vengono riscontrate mutazioni patogenetiche (5). In questi casi il fenotipo FH potrebbe essere causato da:

- 1) mutazioni a carico di un gene responsabile di FH non ancora identificato;
- 2) una fenocopia di FH legata a meccanismi non genetici quali comportamenti deleteri per la salute (es. disturbi del comportamento alimentare);
- 3) meccanismi alternativi di ereditarietà (ipercolesterolemia poligenica, PHC).

La PHC è una causa frequente di ipercolesterolemia severa in assenza di mutazioni patogenetiche (5) (circa 80%) e molto spesso è difficilmente distinguibile fenotipicamente dall'FH. I fattori che complicano la possibilità di distinguere clinicamente l'ipercolesterolemia monogenica e poligenica sono determinati dalla presenza di più geni che hanno un piccolo effetto positivo sulla concentrazione di LDL-C, aumentando nei soggetti PHC i livelli in modo simile a quelli FH, o, al contrario, negli FH la pre-

senza di geni compensatori che riducono il colesterolo LDL al di sotto delle soglie per la diagnosi di una forma monogenica. Alcune evidenze, tuttavia, suggeriscono che gli individui con fenotipo clinico di FH e una causa monogenica sono a maggior rischio di CVD rispetto a quelli in cui non viene trovata alcuna mutazione causale.

In un nostro recente studio su 370 soggetti con sospetto clinico di FH, abbiamo riscontrato la presenza di una forma monogenica in 209 (56,5%) pazienti mentre in 89 (24,1%) abbiamo riscontrato la presenza di una forma poligenica di ipercolesterolemia (definita dalla presenza di un *LDL-rising genetic risk score* $>0,69$) e in 72 (19,5%) la causa dell'ipercolesterolemia è rimasta indefinita. In questo studio abbiamo inoltre evidenziato come una causa monogenica di FH si associava ad una minore risposta alle terapie convenzionali ipolipemizzanti, nonché ad un aumento del *burden* aterosclerotico valutato mediante TAC delle coronarie e del rischio di eventi cardiovascolari maggiori incidenti nel corso dei 31 mesi di osservazione (7). Dunque, distinguere le due forme di ipercolesterolemia è rilevante al fine di ottenere una adeguata stima del rischio di malattia cardiovascolare e valutare così la migliore gestione terapeutica.

Polygenic Risk Score (PRS): piccoli score di rischio basati su SNPs significativi

Dal 2007 al 2013, i grandi studi di GWAS hanno mostrato l'associazione tra diversi SNPs e i livelli della colesterolemia totale, LDL, HDL, e della trigliceridemia totale confermando quindi come le alterazioni del metabolismo lipidico abbiano un'architettura multifattoriale (8). Altri GWAS hanno invece studiato l'associazione tra SNPs e la presenza della malattia cardiovascolare (3). È importante considerare che molti

degli *SNPs* identificati dimostravano una associazione significativa anche dopo correzione per confronti multipli utilizzando un valore soglia di significatività piuttosto stringente ($P \leq 1 * 10^{-8}$). Questo rigore nell'analisi dei risultati ha permesso di identificare poche decine di *SNPs* che avevano una associazione significativa con i tratti fenotipici di interesse (Figura 1) (9). I risultati dei *GWAS* sono stati così perfezionati scegliendo questi *SNPs* "migliori" (*top-ranked*), cioè quelli associati ad un maggiore effetto sul fenotipo (con valori di *p value* più significativi). I *top-ranked SNPs* sono stati successivamente inclusi nella costruzione di *PRS* basati generalmente sull'utilizzo di meno di 100 *SNPs*, spesso anche solo di poche decine. I *PRS* così costruiti sono stati utilizzati in coorti relativamente piccole in quanto ritenuti sufficientemente sensibili e specifici.

Omnigenic Risk Score (ORS): score di rischio basati su grandi numeri di *SNPs*

I *GWAS* successivi, in particolare in campo lipidologico, hanno reso sempre maggiormente chiaro un fatto fondamentale, cioè che più grande è la coorte studiata, maggiore è il numero di *SNPs* che vengono identificati come associati al tratto di interesse. Nel 2010 un *GWAS* condotto su 100.000 individui ha identificato circa 95 *SNPs* associati con diversi lipidi plasmatici. Nel 2013 e nel 2018 quando le dimensioni del campione sono aumentate rispettivamente a 188.000 e a oltre 600.000 individui, rispettivamente ulteriori 62 e 118 *SNPs* si sono mostrati come significativamente associati con i diversi tratti lipidici (4). Tuttavia, è importante notare come rispetto agli *SNPs* precedentemente individuati, il "peso" di questi nuovi *SNPs* identificati sia relativamente piccolo. Infatti, le dimensio-

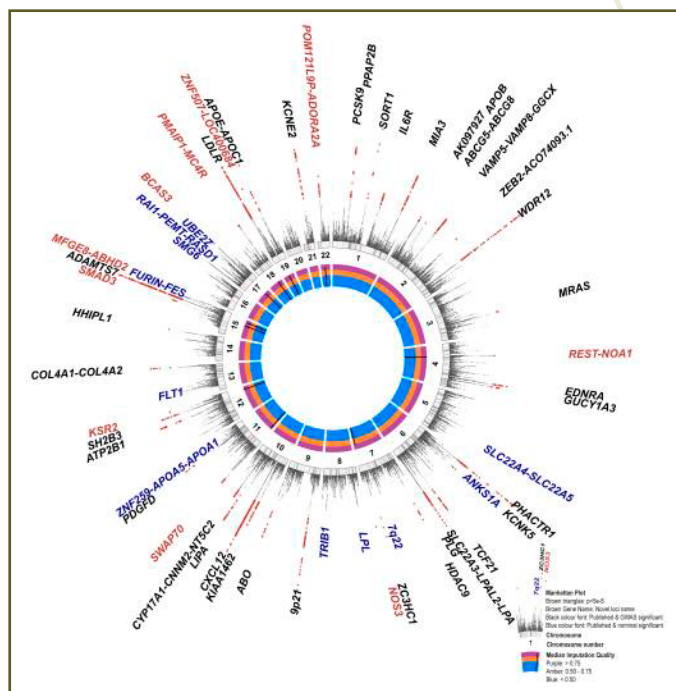


Figura 1 - Manhattan plot che riassume i risultati dello studio di associazione con la CAD ottenuto dal progetto 1000 Genomi [9]

ni più ampie del campione hanno permesso di identificare *SNPs* che non avevano raggiunto la significatività statistica a causa dei limiti metodologici degli studi precedenti. Questi nuovi *GRS* incorporerebbero dunque un numero considerevole di *SNPs* rendendo così possibile calcolare questi nuovi score in coorti sempre più grandi. L'utilizzo di questi nuovi *GRS* ha permesso di sottolineare un punto chiave: grazie allo studio di popolazioni sufficientemente grandi, per esempio miliardi di persone, molti loci apparentemente non significativi nei precedenti *GWAS* sono diventati significativi anche se il loro effetto sul tratto lipidico è prossimo allo zero. Da qui lo sviluppo di questo modello *omnigenico* per la valutazione dell'associazione con patologie a genetica complessa. Secondo questo modello ci sarebbero dunque un numero quasi infinito di piccole varianti (*SNPs* "minori") sparse nel genoma che influenzerebbe

ro il fenotipo non necessariamente in modo diretto bensì anche attraverso meccanismi (*pathway*) indiretti.

Sebbene molto più difficili da determinare e calcolare, gli *ORS* possono valutare molti più determinanti genetici contemporaneamente rispetto ai *PRS* permettendo così di perfezionare la capacità di discriminare i vari individui sulla base del rischio di uno specifico fenotipo di interesse.

Genetic risk score (GRS) e rischio di ASCVD

Numerosi studi hanno valutato il ruolo predittivo degli score *omnigenetici* (Tabella 1). Un esempio è lo studio di Abraham, et al. (10) che ha utilizzato 49310 *SNPs* identificati grazie ad una metanalisi condotta sul database CARDIoGRAMplusC4D al fine di costruire il primo *ORS* per la valutazione del rischio di malattia coronarica (*CHD*). Lo score omnigenico così costruito (*metaGRS*) è stato successivamente testato sulle coorti del FINRISK (N=12.676) e del Framingham Heart Study (N=3406), ove è stato anche possibile effettuare una

correzione per i fattori di rischio tradizionali. I risultati hanno dimostrato che gli individui che avevano uno score nel quintile superiore presentavano un *hazard ratio* (*HR*) di *CHD* incidente rispettivamente del 4,51 (95% CI (3,47-5,85)) per la popolazione del FINRISK e dell'1,84 (95% CI (1,43-2,37)) per quella del Framingham Heart Study. Successivamente, Abraham, et al. (10) hanno provato ad applicare questo *ORS* nella coorte di pazienti FH inclusi nello studio ARGOS (N=464) scoprendo così che anche in questa popolazione un elevato punteggio di *ORS* associava con la *CHD*. In conclusione, nel complesso questo *ORS* si è dimostrato essere in grado di predire il rischio di *CHD* meglio rispetto ai *PRS* sviluppati precedentemente ed in modo indipendente dai fattori di rischio classici per la malattia cardiovascolare.

Lo score di Abraham, che comprendeva 49000 *SNPs*, è stato presto oscurato dall'avvento di due ulteriori *SNPs* comprendenti più di 1 milione di *SNPs*. Il più piccolo dei due comprende l'utilizzo di 1745180 *SNPs* (*metaGRS*) ed è stato applicato in 22242 casi di soggetti con malattia coronarica

Tabella 1 - Sommario dei valori di rischio relativo (Hazard Ratio) di ASCVD associato agli score omnigenici.

Coorte	Numerosità	Hazard Ratio (95% CI)	Referenza
FINRISK 1992 FINRISK 1997 FINRISK 2002 Health 2000	24.124	1,27 (1,20-1,35)	Arterioscler Thromb Vasc Biol. 2013; 33: 2261-2266
Malmö Diet and Cancer (MDC)	23.595	1,23 (1,18-1,28)	Eur Heart J. 2016; 37: 561-567
WOSCOPS CARDIA Biobank	4.910 1.154 4.392	1,32 (1,04-1,68)	Circulation. 2017; 135: 2091-210
FINRISK Framingham Heart Study	12.676 3.406	1,66 (1,55-1,78)	Eur Heart J. 2016; 37: 3267-3278
UK Biobank	22.242 casi di CAD 460.387 controlli	1,71 (1,68-1,73)	J Am Coll Cardiol. 2018; 72: 1883-1893

(MC) e 460387 controlli inclusi nella UK Biobank (11). Con un *HR* di 1,71 (95% CI 1,68-1,73) per ogni aumento nella deviazione standard dello score, questo *ORS* ha chiaramente dimostrato superiorità rispetto a *PRS* di minori dimensioni nel predire il rischio di MC (18-21) dimostrando che i soggetti che si trovavano nel quintile superiore del *GRS* avevano un rischio di 4,17 volte (95% CI 3,97-4,38) maggiore rispetto agli individui con punteggi nel quintile più basso. Nel 2018, Khera et al. (12), utilizzando un *dataset* di validazione basato sui 120280 partecipanti alla fase uno della genotipizzazione della UK Biobank, hanno identificato 6630150 *SNPs* associati a MC incorporandoli nel loro score. Nonostante la maggior parte degli *SNPs* inclusi in questo *ORS* non presentasse una significatività a livello dei *GWAS*, mostravano comunque una associazione molto debole con la MC e circa 6.000 tra questi *SNPs* erano direttamente correlati a *pathway* coinvolti direttamente nella sua patogenesi (4). Dopo aver validato le prestazioni di questo *score* in un campione indipendente di soggetti composto dai 288.978 partecipanti alla fase due della UK Biobank, lo studio dello *score* è stato ulteriormente esteso ad una coorte di 409258 individui rilevando così che coloro che presentavano un punteggio maggiore del 5° e del 1° percentile avevano rispettivamente un rischio di sviluppare MC di 3,34 volte (95% CI (3,12-3,58)) e 4,83 volte (95% CI (4,25-5,46)) maggiore rispetto ai soggetti con score più bassi. Per sottolineare ulteriormente l'utilità di questo score di rischio omnigenico, Khera et al. hanno effettuato anche un confronto testa a testa tra il *GRS* 50-*SNP* (13), quello a 49000-*SNP* (10) e il loro *GRS* con 6,6 milioni di *SNPs* per valutare la capacità di predizione del rischio di MC. I risultati di questa comparazione hanno dimostrato chiaramente come lo score basato su un

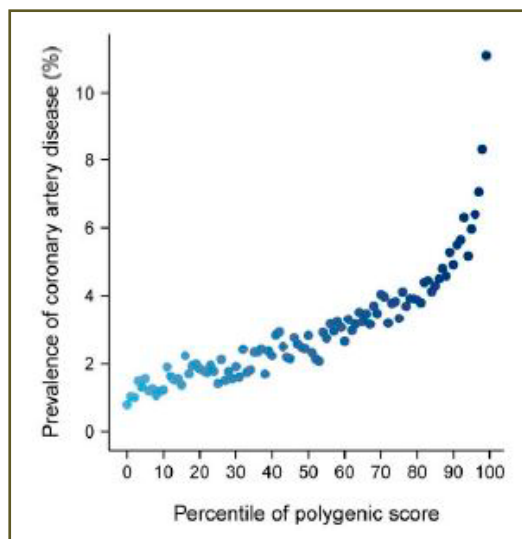


Figura 2 - Prevalenza della CAD secondo i percentili di *PRS* [12].

più ampio numero di *SNPs* sviluppato da Khera et al, fornendo molte informazioni aggiuntive, sia superiore rispetto agli altri due *ORS* nel predire il rischio di MC (*Figura 2*) (12).

Utilità clinica dei *GRS*

I risultati di questi recenti studi hanno chiaramente evidenziato le possibili implicazioni cliniche e l'utilità dei *GRS*, in particolare gli score di rischio omnigenici, nel predire il rischio di MC indipendentemente dai fattori di rischio tradizionali (4, 11). Una questione pratica emergente riguarda il tipo di tecnologia e i costi per determinare questi score di rischio omnigenici. Un altro punto sostanziale consiste nel fatto che idealmente i risultati dovrebbero essere comunicati da un genetista (ad es. un consulente genetico o clinico esperto nella segnalazione del rischio poligenico). Un altro dato su cui è necessario riflettere è quello etico nonché farmaco-economico in quanto poiché i determinanti genetici sono stabiliti sin dalla nascita, questi *GRS* po-

trebbero essere utilizzati in epoche molto precoci nella vita di un individuo per identificare coloro che sono a più alto rischio di sviluppare un determinato tratto in fase pre-sintomatica. Questo è particolarmente rilevante per un fenotipo come la MC in cui un intervento precoce, in particolare sul controllo dei fattori di rischio modificabili quali il fumo o l'obesità ma anche non modificabili come l'ipercolesterolemia, potrebbe contribuire a ridurre il rischio di andare incontro ad un evento. Infatti, una diagnosi genetica potrebbe motivare gli individui ad adottare modifiche dello stile di vita, così come uno stile di vita sano potrebbe ridurre il rischio di MC in soggetti con *GRS* elevato (4).

Nonostante ciò, è importante notare che sebbene *GRS* poligenici o omnigenici possano fornire indubbiamente alcuni benefici clinici, vi sono alcuni problemi che rimangono attualmente irrisolti (e.g. quale punteggio considerare come predittivo, la

replica degli score in studi prospettici). Inoltre, vi sono dei problemi tecnici tra cui, molto importante, la valutazione degli *SNPs* sulla base della etnia.

Conclusioni

La costruzione di uno score di rischio poligenico ha il potenziale di aiutare a capire in maniera più approfondita l'architettura genetica delle patologie complesse, tra cui le alterazioni del metabolismo lipidico e la malattia cardiovascolare. Il più notevole progresso in tal senso è stata l'evoluzione nel calcolo degli *score* di rischio da poligenici a omnigenici. Infatti, questi *score* più grandi sembrano fornire una migliore predizione del rischio cardiovascolare rispetto ai precedenti *PRS* di più piccole dimensioni (*Figura 3*) (11).

In conclusione, l'uso delle informazioni genetiche di un individuo mediante l'uso dei *GRS* si potrebbe integrare con gli approcci tradizionali comunemente utilizzati per la stratificazione del rischio cardiovascolare (es carte del rischio *SCORE*, anamnesi familiare e personale, esami strumentali come ecodoppler TSA o angio-TAC delle coronarie) migliorando potenzialmente l'accuratezza della diagnosi e permettendo così la messa in atto di un diverso percorso di cura individualizzato e personalizzato sulle necessità del paziente (Medicina di Precisione).

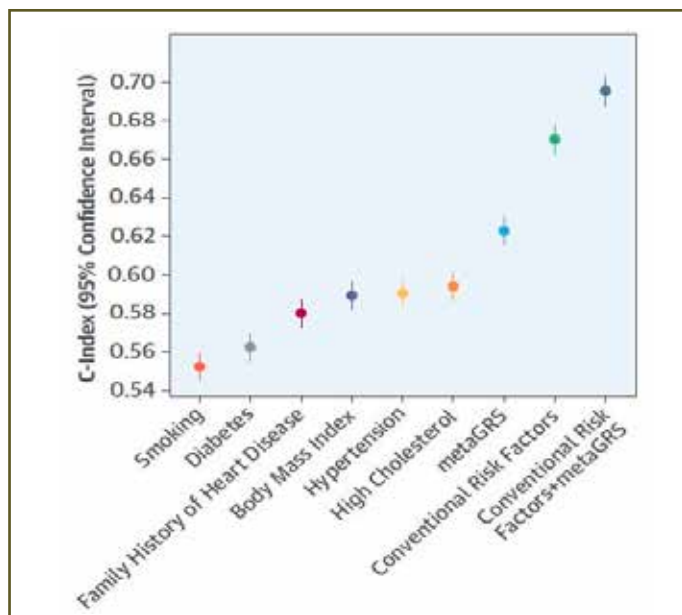


Figura 3 - Gli score genomici si integrano ai fattori di rischio tradizionali per migliorare la stratificazione del rischio cardiovascolare [11].

Bibliografia

1. Chen G, Levy D. Contributions of the Framingham Heart Study to the Epidemiology of Coronary Heart Disease. *JAMA Cardiol.* 2016; 1: 825-830.
2. Martin N, Boomsma D, Machin G. A twin pronged attack on complex traits. *Nature genetics.* 1997; 17: 387-392.
3. Khera AV, Chaffin M, Zekavat SM, et al. Whole-Genome Sequencing to Characterize Monogenic and Polygenic Contributions in Patients Hospitalized with Early-Onset Myocardi-

- al Infarction. *Circulation*. 2019; 139:1593-1602.
4. Dron JS, Hegele RA. The evolution of genetic-based risk scores for lipids and cardiovascular disease. *Curr Opin Lipidol*. 2019; 30: 71-81.
5. Khera AV, Hegele RA. What Is Familial Hypercholesterolemia, and Why Does It Matter? *Circulation*. 2020; 141: 1760-1763.
6. Ference BA, Ginsberg HN, Graham I, et al. Low-density lipoproteins cause atherosclerotic cardiovascular disease. 1. Evidence from genetic, epidemiologic, and clinical studies. A consensus statement from the European Atherosclerosis Society Consensus Panel. *Eur Heart J*. 2017; 38: 2459-2472.
7. D'Erasmo L, Minicocci I, Di Costanzo A, et al. Clinical Implications of Monogenic Versus Polygenic Hypercholesterolemia: Long-Term Response to Treatment, Coronary Atherosclerosis Burden, and Cardiovascular Events. *JAHA in press*.
8. Dron JS, Wang J, Low-Kam C, et al. Polygenic determinants in extremes of high-density lipoprotein cholesterol. *J Lipid Res*. 2017; 58: 2162-2170.
9. Nikpay M, Goel A, Won HH, et al. A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nat Genet*. 2015; 47: 1121-1130.
10. Abraham G, Havulinna AS, Bhalala OG, et al. Genomic prediction of coronary heart disease. *Eur Heart J*. 2016; 37: 3267-3278.
11. Inouye M, Abraham G, Nelson CP et al. UK Biobank CardioMetabolic Consortium CHD Working Group. Genomic Risk Prediction of Coronary Artery Disease in 480,000 Adults: Implications for Primary Prevention. *J Am Coll Cardiol*. 2018; 72: 1883-1893.
12. Khera AV, Chaffin M, Aragam KG, et al. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat Genet*. 2018; 50: 1219-1224.
13. Tada H, Melander O, Louie JZ, et al. Risk prediction by genetic risk scores for coronary heart disease is independent of self-reported family history. *Eur Heart J*. 2016; 37: 561-567.