

## Rassegna

# È possibile predire il rischio di diabete mediante marcatori genetici?

### A. Doria

Research Division, Joslin Diabetes Center, and Department of Medicine, Harvard Medical School, Boston, MA

Corrispondenza: prof. Alessandro Doria, Section on Genetics & Epidemiology, Joslin Diabetes Center, One Joslin Place, Boston, MA 02215

G It Diabetol Metab 2011;31:205-212

*Pervenuto in Redazione il 10-10-2011*

*Accettato per la pubblicazione il 26-10-2011*

Parole chiave: diabete di tipo 2, genetica, polimorfismi, predizione del rischio

Key words: type 2 diabetes, genetics, polymorphisms, risk prediction

### RIASSUNTO

La recente identificazione di vari loci genetici associati con il diabete di tipo 2 ha sollevato l'ipotesi di poter usare questi marcatori genetici per predire il rischio di diabete in giovane età, in modo da instaurare programmi di prevenzione precoce specificamente rivolti agli individui a rischio. Per aumentare il potere predittivo, sono stati sviluppati *genetic risk score* basati su vari marcatori genetici considerati insieme secondo un semplice modello additivo. I risultati, però, sono stati piuttosto deludenti, dimostrando un modesto valore predittivo degli score genetici usati da soli e un aumento trascurabile delle capacità discriminatorie quando questi vengono aggiunti a modelli predittivi basati su caratteristiche cliniche. È possibile che l'analisi di interazioni tra geni e con fattori ambientali e l'estensione dello studio a varianti genetiche meno frequenti possano portare in futuro allo sviluppo di migliori strumenti predittivi su base genetica. Tuttavia, al momento attuale, caratteristiche cliniche quali il BMI, alcuni indici metabolici e la storia familiare di diabete rimangono i migliori fattori disponibili per predire il rischio di diabete di tipo 2 nella pratica clinica.

### SUMMARY

*Can genetic markers predict diabetes risk?*

*The recent finding of several loci robustly associated with type 2 diabetes has raised the possibility of using genetic markers to predict the risk of this disease at a young age in order to start early prevention programs. In an attempt to increase predictive power, multiple genetic markers have been considered together into genetic risk scores built according to simple additive models. Results, however, have been disappointing, with genetic scores having modest predictive power when used alone and yielding a negligible increase in risk discrimination when added to predictive models based on clinical characteristics. The analysis of gene-gene and gene-environment interactions, and the extension of the study to less frequent genetic variant offer some hope for the development of better genetic tools in the future. However, at this time, clinical characteristics such as BMI, certain metabolic indices, and a family history of diabetes remain the best predictors of type 2 diabetes risk to be used in clinical practice.*

## Introduzione

Una delle promesse della rivoluzione genomica attualmente in corso è quella di identificare marcatori genetici che possano essere usati per predire lo sviluppo futuro di malattie croniche multifattoriali, in modo da identificare soggetti a rischio da sottoporre a efficaci programmi di prevenzione. La disponibilità di tali strumenti predittivi non è importante nel caso vi siano misure preventive facili da implementare. Tuttavia, nel caso di misure preventive onerose, la cui implementazione deve durare per lunghi periodi o addirittura per una vita intera, il fatto di poter individuare soggetti a rischio su cui indirizzare l'intervento preventivo diventa fondamentale per un buon rapporto costo-beneficio.

L'identificazione di tali strumenti predittivi è particolarmente importante nel caso del diabete, dato l'elevato numero di soggetti affetti, la lunga lista di complicanze croniche e i costi sia umani sia finanziari di tale malattia<sup>1-3</sup>. Negli ultimi anni sono stati sviluppati vari algoritmi per predire il rischio di diabete sulla base di caratteristiche cliniche, demografiche e antropometriche. Un esempio è il *Cambridge type 2 diabetes risk score*, basato su età, sesso, terapia in corso, storia familiare di diabete di tipo 2, BMI e uso di nicotina<sup>4</sup>. Un altro è il *Framingham offspring type 2 diabetes risk score*, basato su età, sesso, storia familiare di diabete di tipo 2, BMI e livelli ematici di HDL, trigliceridi e glucosio a digiuno<sup>5</sup>. L'accuratezza predittiva di questi due modelli è abbastanza buona, essendo del 75% e 85%, rispettivamente<sup>4,5</sup>. Tuttavia, questi strumenti predittivi si basano su variabili valutate pochi anni prima dello sviluppo del diabete. Dal momento che la gran parte di queste variabili cambia nel tempo, non è chiaro se la performance di questi score sia altrettanto buona per predire il rischio di diabete con largo anticipo.

Vari studi hanno dimostrato che il diabete, sia di tipo 1 sia di tipo 2, ha una forte componente genetica, tale da determinare un aumento del rischio di malattia tra le 2 e le 6 volte nei soggetti con storia familiare positiva rispetto a quelli con storia familiare negativa<sup>6</sup>. Contrariamente alle caratteristiche cliniche o biomorali, le caratteristiche genetiche di un individuo non cambiano nel tempo. Offrono perciò la possibilità, almeno teorica, di essere usate per stimare il rischio di diabete con vari decenni di anticipo o addirittura alla nascita. Lo scopo di questo articolo è quello di illustrare i progressi che sono stati fatti negli ultimi anni nell'identificare marcatori genetici di diabete e discutere il loro uso per identificare soggetti a rischio. La discussione riguarda in particolare il diabete di tipo 2, vista l'abbondanza di dati al riguardo in letteratura e il fatto che 9 casi di diabete su 10 sono dovuti a questa forma della malattia.

## La rivoluzione GWAS e l'identificazione di marcatori genetici del diabete di tipo 2

Negli ultimi cinque anni lo studio della genetica del diabete e di altre malattie multifattoriali ha ricevuto un impulso senza precedenti. Due fattori hanno contribuito a questo sviluppo.

Il primo è il completamento del progetto HapMap attraverso il quale è stato definito il *linkage disequilibrium* lungo l'intero genoma umano<sup>7</sup>. Il *linkage disequilibrium* è l'associazione statistica tra due polimorfismi dovuta all'esistenza di un aplo-tipo ancestrale contenente entrambi i polimorfismi e avente alta frequenza nella popolazione<sup>8</sup>. Tale fenomeno fa in modo che due polimorfismi adiacenti possano essere altamente correlati, al punto che uno dei due possa essere usato come marcatore dell'altro. Una volta che tali correlazioni sono state determinate lungo l'intero genoma da HapMap, è diventato possibile selezionare un set di ~500.000 polimorfismi (cosiddetti *single nucleotide polymorphisms* o SNPs) in grado di rappresentare tutti i 10 milioni di SNPs che si stima siano presenti nel genoma umano. Il secondo fattore responsabile per la crescita esponenziale di questo campo di ricerca è stato lo sviluppo di tecnologie basate su *microarray* attraverso le quali è possibile tipizzare un individuo per centinaia di migliaia di SNPs in un solo assay e a un costo ragionevole<sup>9</sup>. Insieme, questi due fattori hanno reso possibile che la ricerca di marcatori genetici per malattie multifattoriali potesse essere estesa all'intero genoma invece di riguardare una manciata di geni scelti sulla base di criteri arbitrari. Tali studi sistematici – denominati *genome-wide association studies* o GWAS – sono rapidamente diventati la norma nel campo<sup>10</sup>, grazie anche all'esistenza di banche di DNA contenenti migliaia di campioni raccolti da casi di diabete e controlli non diabetici.

È importante ricordare che l'approccio GWAS ha i suoi limiti. Primo, perlomeno nel modo in cui è stato condotto finora, l'approccio assume che le varianti genetiche che conferiscono un aumentato rischio di diabete abbiano una frequenza  $\geq 5\%$  nella popolazione. Anche se questo assunto è giustificato da considerazioni teoriche<sup>11</sup>, è stato dimostrato che anche varianti genetiche più rare possono contribuire al rischio di malattie metaboliche<sup>12</sup>. Secondo, la strategia GWAS pone dei seri problemi metodologici dovuti al numero elevatissimo di test (centinaia di migliaia) che vengono condotti in ciascuno studio. In queste condizioni, molti polimorfismi possono raggiungere la significatività nominale ( $p < 0,05$ ) per caso<sup>10</sup>. È quindi necessario usare una soglia di significatività molto più alta ( $p < 5 \times 10^{-8}$ ), la quale a sua volta richiede lo studio di diverse migliaia di casi e controlli in modo da avere power sufficiente per l'identificazione di effetti genetici relativamente modesti. Infine, gli studi GWAS soffrono degli stessi problemi di altri studi genetici per quanto riguarda la definizione del fenotipo. Il diabete, soprattutto quello di tipo 2, è una malattia eterogenea, nel senso che ne esistono varie forme caratterizzate da fisiopatologia diversa (per esempio, forme accompagnate da obesità verso forme normopeso, forme con un difetto prevalentemente secretorio verso forme caratterizzate prevalentemente da insulino-resistenza). In queste condizioni, il risultato di uno studio genetico di associazione dipende moltissimo da come sono stati reclutati i casi e i controlli e dalle loro caratteristiche.

Nonostante questi limiti, l'approccio GWAS ha portato finora all'identificazione di più di 40 SNPs associati con il diabete di tipo 2, alcuni dei quali (quelli identificati dai primi studi<sup>13-23</sup>)

sono riportati in tabella 1. Rimandiamo il lettore alle varie rassegne pubblicate su questo argomento per una descrizione dettagliata di questi risultati<sup>24,25</sup>. In generale, le varianti genetiche identificate finora sono associate con un incremento piuttosto modesto del rischio di diabete, con odds ratio allelici nell'ordine di 1,10-1,20. L'unica eccezione è il segnale di associazione nella regione del gene *TCF7L2*, ma anche in questo caso l'odds ratio allelico non supera l'1,50<sup>26</sup>. Questo effetto modesto è probabilmente da ascrivere al fatto che, con poche eccezioni, le varianti a rischio sono localizzate in regioni non codificanti del genoma e quindi agiscono sull'espressione piuttosto che sulla sequenza aminoacidica delle proteine codificata dai geni interessati. La frequenza di queste varianti è generalmente molto più elevata di quella del diabete di tipo 2 nella popolazione, il che comporta l'esistenza di un numero elevato di portatori di varianti a rischio che finiscono per non diventare diabetici. Tale bassa penetranza suggerisce che le varianti genetiche in questione conferiscono un'aumentata suscettibilità, ma anche altri fattori, presumibilmente ambientali, devono essere presenti perché si sviluppi il diabete.

Molte delle varianti identificate finora sembrano agire sulla secrezione insulinica piuttosto che sulla sensibilità dei tessuti periferici all'insulina<sup>25</sup>. Questo è probabilmente in relazione al fatto che la maggior parte degli studi ha considerato il fenotipo diabete piuttosto che quello dell'insulino-resistenza,

ma potrebbe anche essere dovuto a influenze ambientali più forti su quest'ultimo fenotipo. Un altro aspetto interessante è che quasi nessuno dei geni che sembrano essere responsabili dell'associazione con il diabete (assumendo che questi siano quelli più vicini agli SNPs) è stato precedentemente implicato nella regolazione della secrezione o della sensibilità insulinica. L'identificazione di questi effetti genetici potrebbe quindi portare alla scoperta di meccanismi di regolazione del metabolismo fino a ora completamente sconosciuti, con importantissime implicazioni per lo sviluppo di nuovi agenti terapeutici.

## Metodi per la valutazione del valore predittivo dei marcatori genetici

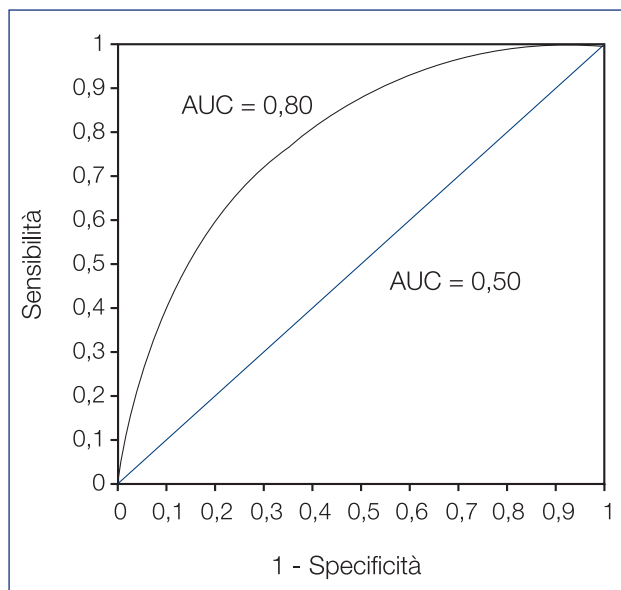
Prima di discutere sull'utilità di questi marcatori per predire il rischio diabete, è opportuno accennare alle metodiche usate per analizzare la performance di modelli predittivi. Mentre l'enfasi degli studi di associazione è su quanto probabile sia l'ipotesi nulla di assenza di associazione tra lo SNP e il diabete, in questo caso l'enfasi è sulla capacità del marcatore genetico di discriminare tra soggetti che svilupperanno il diabete e quelli che non lo svilupperanno. Questa capacità di *discriminazione* viene tradizionalmente espressa dall'area

**Tabella 1** Geni associati con il diabete di tipo 2 identificati dai primi GWAS nel 2007-2008.

| Cromosoma | Gene più vicino                | Funzione cellulare  | OR        | Frequenza dell'allele a rischio |
|-----------|--------------------------------|---|-----------|---------------------------------|
| 1p12      | <b>NOTCH2</b>                  | Regolatore della differenziazione cellulare                           | 1,13      | 0,11                            |
| 2p21      | <b>THADA</b>                   | Sconosciuta   | 1,15      | 0,90                            |
| 3p14      | <b>ADAMTS9</b>                 | Enzima proteolitico regolatore della matrice extracellulare           | 1,09      | 0,76                            |
| 3p25      | <b>PPARG</b>                   | Fattore di trascrizione e recettore di tiazolidinedioni               | 1,17      | 0,85                            |
| 3q28      | <b>IGF2BP2</b>                 | Proteina legante l'mRNA di IGF2                                       | 1,14      | 0,29                            |
| 6p22.3    | <b>CDKAL1</b>                  | Presunto regolatore di ciclini chinasi                                | 1,0-1,20  | 0,31                            |
| 7p15      | <b>JAZF1</b>                   | Proteina zinc-finger di funzione sconosciuta                          | 1,10      | 0,50                            |
| 8q24.11   | <b>SLC30A8</b>                 | Trasportatore dello zinco 8   | 1,18      | 0,65                            |
| 9p21      | <b>CDKN2A</b><br><b>CDKN2B</b> | Inibitori delle chinasi ciclino-dipendenti 2A e 2B                    | 1,20      | 0,83                            |
| 10p13-p14 | <b>CDC123</b>                  | Proteina coinvolta nella regolazione del ciclo cellulare              | 1,11      | 0,18                            |
|           | <b>CAMK1D</b>                  | Mediatore della trasduzione del segnale di chemochine nei granulociti |           |                                 |
| 10q23-q25 | <b>IDE</b>                     | Metallopeptidasi neutra   | 1,13      | 0,53                            |
|           | <b>HHEX</b>                    | Fattore di trascrizione homeobox                                      |           |                                 |
|           | <b>KIF11</b>                   | Proteina dei microtubuli  |           |                                 |
| 10q25.3   | <b>TCF7L2</b>                  | Fattore di trascrizione   | 1,31-1,71 | 0,26                            |
| 11p15.1   | <b>KCNJ11</b>                  | Canale del potassio   | 1,14      | 0,47                            |
| 12q21     | <b>TSPAN8</b>                  | Glicoproteina di superficie   | 1,09      | 0,27                            |
|           | <b>LGR5</b>                    | Recettore di proteina G non nota                                      |           |                                 |
| 16q12.2   | <b>FTO</b>                     | Proteina a localizzazione nucleare di funzione ignota                 | 1,27      | 0,38                            |

OR: odds ratio allelico di diabete di tipo 2.

sotto la curva ROC (*receiver operating characteristic*) (Fig. 1). Questo indice, identico in valore alla *c statistics*, esprime la probabilità che il rischio di diabete stimato dal marcatore in esame sia più alto per un individuo che svilupperà il diabete che per un soggetto che non lo svilupperà<sup>27</sup>. Un'area sotto la curva ROC uguale a 1 indica una discriminazione perfetta, nel senso che il rischio stimato dal marcatore per quei soggetti che poi svilupperanno il diabete sarà sempre più alto del rischio stimato per i soggetti che non lo svilupperanno. Un'area sotto la curva ROC uguale a 0,50 indica invece assenza completa di capacità discriminatorie da parte del marcatore, visto che la probabilità che il rischio stimato per coloro che svilupperanno il diabete sia maggiore di quello stimato per coloro che non svilupperanno il diabete è del 50%, cioè uguale alla probabilità di ottenere testa o croce lanciando una moneta. L'area sotto la curva ROC può essere anche usata per valutare la performance di modelli più complessi basati su molteplici predittori genetici e non genetici, e stimare quanto questa venga aumentata dall'aggiunta di un nuovo marcatore al modello predittivo. Va notato, tuttavia, che la sensibilità dell'area sotto la curva è piuttosto limitata<sup>28</sup>. Per tale motivo, sono stati sviluppati altri approcci basati sulla *riclassificazione*, come gli indici *net reclassification improvement* (NRI) e *integrated discrimination improvement* (IDI)<sup>29,30</sup>. Questi metodi valutano come l'aggiunta di una o più variabili migliori la classificazione dei soggetti nelle categorie di rischio a cui dovrebbero appartenere dato il loro rischio di diabete osservato nello studio. Un'altra caratteristica che deve essere valutata è la *calibratura*, cioè la capacità di un modello predittivo di stimare con precisione il rischio assoluto di un evento come per l'appunto l'insorgenza di diabete<sup>31</sup>.



**Figura 1** Esempio di curva ROC. La retta più bassa (area sotto la curva [AUC] = 0,50) corrisponde alla mancanza completa di discriminazione tra soggetti a rischio e non a rischio. La curva più alta (AUC = 0,80) corrisponde invece a una buona discriminazione tra i due gruppi.

Il fatto che un modello predittivo abbia una buona capacità discriminativa, cioè permetta di distinguere accuratamente soggetti a rischio da soggetti non a rischio, non significa che abbia anche una buona calibratura, cioè che fornisca una stima accurata della probabilità che un individuo sviluppi il diabete.

## Sviluppo di score genetici per predire il rischio di diabete di tipo 2

Com'è stato discusso in precedenza, i marcatori genetici identificati finora sono associati con un incremento piuttosto modesto, tra il 10 e il 30%, del rischio di diabete di tipo 2. Quindi, considerati individualmente, tali marcatori non sono utili per predire l'insorgenza di diabete. Per ovviare a questo problema, sono stati considerati più polimorfismi insieme mediante la creazione di *genetic risk score* (GRS) dati dalla somma del numero di alleli a rischio a ciascun sito polimorfo. Per esempio, nel caso di 10 polimorfismi, il valore del GRS va da 0 (nessun allele a rischio) a 20 (2 alleli a rischio a ciascuno dei 10 siti polimorfi). Tale approccio assume che i polimorfismi sui quali si basa il GRS abbiano effetto indipendente (cioè che non varia a seconda di quali altri alleli a rischio siano presenti) e grandezza simile, anche se in alcuni casi il numero di alleli a rischio contribuiti da ciascun polimorfismo è stato assegnato proporzionalmente alla grandezza dell'effetto di ciascun marcatore (*weighted GRS*).

Questo approccio fu inizialmente usato per valutare il valore predittivo complessivo di quei pochi geni candidati per i quali era stata dimostrata un'associazione con il diabete di tipo 2 quali PPAR $\gamma$ , calpain 10, KCNJ11, glucocinasi e TCF7L2<sup>32-34</sup>. L'uso dei GRS è stato quindi esteso allo studio dei numerosi marcatori genetici identificati mediante il metodo GWAS. I primi due studi di questo tipo, entrambi prospettici, sono apparsi nello stesso numero del *New England Journal of Medicine* nel novembre del 2008. Uno di questi era basato sul Framingham Offspring Study<sup>35</sup>, l'altro sul Malmö Preventive Study e il Botnia Prospective Study<sup>36</sup>.

Lo score genetico esaminato dallo studio di Framingham era composto da 18 SNPs identificati dai primi GWAS per il diabete di tipo 2<sup>18,23</sup>. La performance dello score veniva valutata su 2377 soggetti seguiti per 28 anni, durante i quali erano insorti 255 casi di diabete di tipo 2. In questa popolazione, ciascun punto dello score genetico (corrispondente all'aggiunta di un allele a rischio) era associato a un aumento del 12% del rischio di diabete, per cui il rischio era circa due volte e mezza più alto nei soggetti aventi uno score  $\geq 21$  (l'11% della popolazione) rispetto ai soggetti con uno score  $\leq 15$  (il 25% della popolazione). L'aggiunta dello score a un modello preventivo che includeva solo il sesso (corrispondente alla situazione alla nascita) determinava un modesto, ma significativo aumento dell'area sotto la curva da 0,534 a 0,581 ( $p = 0,01$ ), ma non aveva effetti significativi su modelli predittivi basati sulla storia familiare di diabete considerata da sola (incremento dell'area sotto la curva da 0,595 a 0,615) o insieme ai predittori clinici del *Framingham Offspring*

*type 2 diabetes risk score* (incremento dell'area sotto la curva da 0,90 a 0,901).

Risultati simili sono stati ottenuti nel Malmö Preventive Study e nel Botnia Prospective Study<sup>36</sup>. In questo caso, lo score era basato su 11 SNPs provenienti dai GWAS per i quali era stata dimostrata un'associazione significativa con il diabete nei due studi. Considerate insieme, le due coorti comprendevano un totale di oltre 18.000 soggetti, con un'incidenza cumulativa di diabete dell'11,7% durante un follow-up di circa 23 anni. Come nel Framingham Offspring Study, ciascun incremento di un punto dello score era associato con un aumento del 12% del rischio di diabete. Anche se quest'associazione era altamente significativa, l'impatto dello score sulla predizione del diabete era abbastanza modesto, con un'area sotto la curva ROC di 0,62 per lo score genetico considerato da solo e un incremento dell'area da 0,74 per il modello basato sulle caratteristiche cliniche a 0,75 per lo stesso modello con l'aggiunta dello score genetico.

In contrasto con questi risultati, un altro articolo pubblicato nello stesso anno e basato su studi *cross-sectional* di soggetti francesi ha dimostrato un ottimo valore predittivo di uno score genetico composto da 15 loci, 8 dei quali erano diversi da quelli inclusi negli score precedenti<sup>37</sup>. L'area sotto la curva ROC di questo score considerato da solo è risultata essere in questo caso uguale a 0,86. Le ragioni per la discrepanza tra questo risultato e quelli dei due studi precedenti non sono del tutto chiare, ma una differenza fondamentale è che in questo studio la popolazione impiegata per valutare la

performance dello score genetico era composta per l'80% dalla popolazione in cui era stata originariamente identificata l'associazione tra 8 dei marcatori inclusi nello score e il diabete. È possibile che ciò abbia causato un *overfitting* del modello e una conseguente sovrastima del potere predittivo dello score genetico.

Vari altri studi riguardanti score genetici si sono succeduti a questi tre, con risultati simili a quelli dei primi due<sup>38-45</sup> (Tab. 2). In generale, è stato riscontrato che uno score genetico può aumentare la discriminazione tra soggetti a rischio e soggetti protetti di 5-10 punti percentuali (cioè da 0,50 a 0,55-0,60) quando è usato da solo. Tuttavia, la sua aggiunta a un modello predittivo basato su caratteristiche cliniche o antropometriche è in grado di migliorare la discriminazione di solo 1 o 2 punti percentuali. L'aggiunta di altri marcatori genetici (fino a un totale di 40) non sembra aumentare il potere discriminatorio, anche perché i nuovi marcatori che vengono via via aggiunti agli score hanno un effetto progressivamente più debole<sup>41-43</sup>. Alcuni studi hanno suggerito che la performance dei marcatori genetici sia migliore in soggetti di età inferiore ai 50 anni<sup>41</sup> o nei soggetti con BMI  $\geq 30$ <sup>38</sup>, ma tali dati devono essere confermati.

## Prospettive future

Nel complesso, i risultati descritti nella sezione precedente sono piuttosto deludenti, soprattutto date le promesse dell'approccio GWAS e la quantità di risorse investite in questi

**Tabella 2** Valore predittivo degli score genetici analizzati finora.

| Studio   | Disegno                         | n di soggetti* | n di SNPs | Area sotto la curva ROC |                 |                                  |
|--|---------------------------------|----------------|-----------|-------------------------|-----------------|----------------------------------|
|  |                                 |                |           | Score genetico          | Modello clinico | Modello clinico + score genetico |
| Botnia Study <sup>34</sup>   | Prospettico                     | 2293           | 2         | -                       | -               | -                                |
| Warren 2 + altri <sup>33</sup>   | Caso-controllo                  | 2332 + 3592    | 3         | -                       | -               | -                                |
| DESIR <sup>32</sup>  | Prospettico                     | 3877           | 3         | 0,56                    | -               | -                                |
| Framingham Offspring Study <sup>35</sup>                                 | Prospettico                     | 2377           | 18        | 0,58                    | 0,90            | 0,901                            |
| Malmö Preventive Project+Botnia Study <sup>36</sup>                      | Prospettico                     | 18.831         | 11        | 0,62                    | 0,74            | 0,75                             |
| DESIR + DIAB2.NEPHRO-GENE+SU.VI.MAX <sup>37</sup>                        | Caso-controllo                  | 4232 + 4595    | 15        | 0,86                    | -               | -                                |
| Rotterdam Study <sup>39</sup>  | Prospettico                     | 6544           | 18        | 0,60                    | 0,66            | 0,68                             |
| GoDARTS <sup>40</sup>  | Caso-controllo                  | 2309 + 2598    | 18        | 0,60                    | 0,78            | 0,80                             |
| Inter99+ADDITION + altri <sup>43</sup>                                   | Caso-controllo                  | 4093 + 5302    | 19        | 0,60                    | 0,92            | 0,93                             |
| CoLaus Study <sup>44</sup>   | Caso-controllo                  | 36 + 5004      | 15        | 0,59                    | 0,86            | 0,87                             |
| Nurses Health Study + Health Professional Follow-up Study <sup>38</sup>  | Caso-controllo<br><i>nested</i> | 2809 + 3501    | 10        | -                       | 0,78            | 0,79                             |
| Whitehall II Study <sup>42</sup>   | Prospettico                     | 5535           | 20        | 0,55                    | 0,72-0,78       | 0,73-0,78                        |
| Study on Nutrition and Health of Aging Population in China <sup>45</sup> | Prospettico                     | 3210           | 17        | 0,62                    | 0,77            | 0,79                             |
| Framingham Offspring Study <sup>41</sup>                                 | Prospettico                     | 3471           | 40        | -                       | 0,903           | 0,906                            |

\*Nel caso di studi caso controllo, il primo numero si riferisce ai casi di diabete di tipo 2, il secondo ai controlli non diabetici.

progetti. Tali studi hanno probabilmente aperto nuove frontiere per quanto riguarda l'eziologia del diabete, ma dal punto di vista della predizione del rischio di malattia, non hanno portato a un reale progresso. Forse i marcatori genetici identificati finora potrebbero essere usati nei primi due decenni di vita, quando le caratteristiche cliniche e antropometriche non sono ancora predittive del rischio di diabete e la storia familiare di diabete è ancora incompleta a causa della giovane età dei genitori. Tuttavia, anche in questo caso, il miglioramento delle capacità discriminative tra soggetti a rischio e soggetti non a rischio determinato dagli score genetici costruiti finora è modesto e di improbabile utilità clinica.

Possiamo aspettarci di più dai marcatori genetici in termini di predizione del rischio di diabete? La risposta a questa domanda deve essere, a mio avviso, cautamente affermativa. La prima considerazione da fare è che tutti gli studi di genetica del diabete di tipo 2 (e di altre malattie multifattoriali) condotti finora non hanno considerato possibili interazioni tra geni, anche se varie osservazioni sperimentali suggeriscono la loro importanza. Per esempio, in un modello murino di diabete di tipo 2 creato dall'inattivazione sia del recettore insulinico sia del suo substrato IRS-1, si osserva una marcata interazione, tale che ciascuno dei due difetti genetici da solo non produce diabete in più del 10% dei topi ma, quando i due difetti sono presenti insieme, più del 50% degli animali sviluppa diabete in giovane età<sup>46</sup>. Inoltre, in questo e in altri modelli, lo sviluppo del diabete dipende in modo marcato dal background genetico del topo, cosa che sottolinea ulteriormente l'importanza delle interazioni tra geni<sup>47</sup>. Che i fattori genetici predisponenti al diabete di tipo 2 identificati finora non sembrano interagire tra loro viene spesso portato come prova del fatto che le interazioni tra geni non siano frequenti nell'uomo. Tuttavia, non è realistico aspettarsi che ci siano interazioni tra quei pochi geni che mostrano associazioni altamente significative senza che debbano essere considerati effetti interattivi. Tali geni andrebbero invece cercati tra quelli che dimostrano livelli modesti di significatività quando vengono considerati da soli. L'aggiunta di effetti interattivi non ancora identificati potrebbe tradursi in una performance migliore degli score genetici di quella riscontrata finora.

La seconda considerazione è che gli score predittivi costruiti finora non considerano possibili interazioni tra geni e caratteristiche cliniche o antropometriche. Per esempio, è possibile che l'effetto di un gene che agisce sulla secrezione insulinica possa essere potenziato dalla presenza di obesità e insulino-resistenza. Alcuni studi, come quello di Cornelis et al., hanno evidenziato una tale interazione, ma poi non hanno considerato questo effetto nella costruzione dello score genetico<sup>38</sup>. È auspicabile che questo tipo di analisi diventi la norma in futuro, approfittando dei nuovi approcci analitici che sono stati sviluppati per incorporare lo studio di interazioni nell'approccio GWAS<sup>48</sup>.

La terza considerazione da fare è che gli studi condotti finora sono stati circoscritti a polimorfismi relativamente frequenti nella popolazione (cioè, aventi una prevalenza maggiore del 5%). Tuttavia, il numero di varianti meno frequenti presente nei database genetici cresce giornalmente sulla spinta di iniziative come il 1000 Genomes Project ([www.1000genomes.org](http://www.1000genomes.org)).

È probabile che lo studio di queste varianti più rare porti all'identificazione di altri marcatori genetici di diabete, com'è suggerito dal fatto che le varianti identificate finora spiegano una parte abbastanza piccola dell'ereditarietà di questa e di altre malattie multifattoriali<sup>49</sup>. La bassa frequenza nella popolazione potrebbe limitare l'utilità di questi altri marcatori genetici, ma il loro numero elevato e il fatto che le varianti più rare hanno generalmente un effetto più marcato<sup>50</sup> di quelle comuni fanno ben sperare.

## Conclusioni

Nonostante il progresso delle metodologie genetiche e gli investimenti fatti in questo campo, e nonostante il marketing di diagnostici genetici da parte di alcune aziende, al momento attuale non sono disponibili marcatori genetici che possano essere utilizzati per predire accuratamente il rischio di diabete di tipo 2. L'utilità di questi marcatori risiede invece negli indizi che questi possono fornire sui meccanismi molecolari della regolazione del metabolismo glucidico e sulle alterazioni che conducono al diabete. È possibile che lo studio di modelli predittivi più complessi, che considerino le interazioni tra geni e con fattori ambientali, e lo spostamento dell'interesse su varianti genetiche meno frequenti possa portare in futuro all'identificazione di predittori genetici più accurati. Tuttavia, al momento attuale, caratteristiche cliniche quali il BMI, indici metabolici, e storia familiare di diabete, rimangono i migliori fattori disponibili per predire il rischio di diabete di tipo 2, anche se il loro valore predittivo è relativamente limitato nel tempo.

## Conflitto di interessi

Nessuno.

## Bibliografia

1. American Diabetes Association. *Economic costs of diabetes in the U.S. In 2007*. Diabetes Care 2008;31:596-615.
2. Cowie CC, Rust KF, Byrd-Holt DD, Eberhardt MS, Flegal KM, Engelgau MM et al. *Prevalence of diabetes and impaired fasting glucose in adults in the U.S. population: National Health And Nutrition Examination Survey 1999-2002*. Diabetes Care 2006;29:1263-8.
3. Krolewski AS, Warram JH. *Epidemiology of late complications of diabetes: A basis for the development and evaluation of preventive program*. In: Kahn CR, Weir GC, King GL, Jacobson AM, Moses AC, Smith RJ, eds. *Joslin's Diabetes Mellitus*. New York: Lippincott, Williams & Wilkins 2005.
4. Rahman M, Simmons RK, Harding AH, Wareham NJ, Griffin SJ. *A simple risk score identifies individuals at high risk of developing type 2 diabetes: a prospective cohort study*. Fam Pract 2008;25:191-6.
5. Wilson PW, Meigs JB, Sullivan L, Fox CS, Nathan DM, D'Agostino RB, Sr. *Prediction of incident diabetes mellitus in*

- middle-aged adults: the Framingham Offspring Study. *Arch Intern Med* 2007;167:1068-74.
6. Weijnen CF, Rich SS, Meigs JB, Krolewski AS, Warram JH. *Risk of diabetes in siblings of index cases with type 2 diabetes: implications for genetic studies.* *Diabet Med* 2002;19:41-50.
  7. Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, Gibbs RA et al. *A second generation human haplotype map of over 3.1 million SNPs.* *Nature* 2007;449:851-61.
  8. Strachan T, Read A. *Human molecular genetics.* Oxford: Garland Science 2010.
  9. Syvanen AC. *Toward genome-wide SNP genotyping.* *Nat Genet* 2005;37(suppl):S5-10.
  10. 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.
  11. Reich DE, Lander ES. *On the allelic spectrum of human disease.* *Trends Genet* 2001;17:502-10.
  12. Brunham LR, Singaraja RR, Hayden MR. *Variations on a gene: rare and common variants in ABCA1 and their impact on HDL cholesterol levels and atherosclerosis.* *Annu Rev Nutr* 2006;26:105-29.
  13. Florez JC, Manning AK, Dupuis J, McAteer J, Irenze K, Gianniny L et al. *A 100K genome-wide association scan for diabetes and related traits in the Framingham Heart Study: replication and integration with other genome-wide datasets.* *Diabetes* 2007;56:3063-74.
  14. Hanson RL, Bogardus C, Duggan D, Kobes S, Knowlton M, Infante AM et al. *A search for variants associated with young-onset type 2 diabetes in American Indians in a 100K genotyping array.* *Diabetes* 2007;56:3045-52.
  15. Hayes MG, Pluzhnikov A, Miyake K, Sun Y, Ng MC, Roe CA et al. *Identification of type 2 diabetes genes in Mexican Americans through genome-wide association studies.* *Diabetes* 2007;56:3033-44.
  16. Rampersaud E, Damcott CM, Fu M, Shen H, McArdle P, Shi X et al. *Identification of novel candidate genes for type 2 diabetes from a genome-wide association scan in the Old Order Amish: evidence for replication from diabetes-related quantitative traits and from independent populations.* *Diabetes* 2007;56:3053-62.
  17. Salonen JT, Uimari P, Aalto JM, Pirskanen M, Kaikkonen J, Todorova B et al. *Type 2 diabetes whole-genome association study in four populations: the DiaGen consortium.* *Am J Hum Genet* 2007;81:338-45.
  18. Saxena R, Voight BF, Lyssenko V, Burt NP, de Bakker PI, Chen H et al. *Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels.* *Science* 2007;316:1331-6.
  19. Scott LJ, Mohlke KL, Bonnycastle LL, Willer CJ, Li Y, Duren WL et al. *A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants.* *Science* 2007;316:1341-5.
  20. Sladek R, Rocheleau G, Rung J, Dina C, Shen L, Serre D et al. *A genome-wide association study identifies novel risk loci for type 2 diabetes.* *Nature* 2007;445:881-5.
  21. Steinthorsdottir V, Thorleifsson G, Reynisdottir I, Benediktsson R, Jonsdottir T, Walters GB et al. *A variant in CDKAL1 influences insulin response and risk of type 2 diabetes.* *Nat Genet* 2007;39:770-5.
  22. Zeggini E, Weedon MN, Lindgren CM, Frayling TM, Elliott KS, Lango H et al. *Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes.* *Science* 2007;316:1336-41.
  23. Zeggini E, Scott LJ, Saxena R, Voight BF, Marchini JL, Hu T et al. *Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes.* *Nat Genet* 2008;40:638-45.
  24. McCarthy MI. *Genomics, type 2 diabetes, and obesity.* *N Engl J Med* 2010;363:2339-50.
  25. Doria A, Patti ME, Kahn CR. *The emerging genetic architecture of type 2 diabetes.* *Cell Metab* 2008;8:186-200.
  26. Grant SF, Thorleifsson G, Reynisdottir I, Benediktsson R, Manolescu A, Sainz J et al. *Variant of transcription factor 7-like 2 (TCF7L2) gene confers risk of type 2 diabetes.* *Nat Genet* 2006;38:320-3.
  27. Hanley JA, McNeil BJ. *The meaning and use of the area under a receiver operating characteristic (ROC) curve.* *Radiology* 1982;143:29-36.
  28. Ridker PM, Buring JE, Rifai N, Cook NR. *Development and validation of improved algorithms for the assessment of global cardiovascular risk in women: the Reynolds Risk Score.* *JAMA* 2007;297:611-9.
  29. Pencina MJ, D'Agostino RB, Sr, D'Agostino RB, Jr, Vasan RS. *Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond.* *Stat Med* 2008;27:157-72.
  30. Cook NR, Ridker PM. *Advances in measuring the effect of individual predictors of cardiovascular risk: the role of reclassification measures.* *Ann Intern Med* 2009;150:795-802.
  31. Hosmer DW, Jr, Lemeshow S. *Applied logistic regression.* New York: John Wiley and Sons, Inc 1989.
  32. Vaxillaire M, Veslot J, Dina C, Proenca C, Cauchi S, Charpentier G et al. *Impact of common type 2 diabetes risk polymorphisms in the DESIR prospective study.* *Diabetes* 2008;57:244-54.
  33. Weedon MN, McCarthy MI, Hitman G, Walker M, Groves CJ, Zeggini E et al. *Combining information from common type 2 diabetes risk polymorphisms improves disease prediction.* *PLoS Med* 2006;3:e374.
  34. Lyssenko V, Almgren P, Anevski D, Orho-Melander M, Sjogren M, Saloranta C et al. *Genetic prediction of future type 2 diabetes.* *PLoS Med* 2005;2:e345.
  35. Meigs JB, Shrader P, Sullivan LM, McAteer JB, Fox CS, Dupuis J et al. *Genotype score in addition to common risk factors for prediction of type 2 diabetes.* *N Engl J Med* 2008;359:2208-19.
  36. Lyssenko V, Jonsson A, Almgren P, Pulizzi N, Isomaa B, Tuomi T et al. *Clinical risk factors, DNA variants, and the development of type 2 diabetes.* *N Engl J Med* 2008;359:2220-32.
  37. Cauchi S, Meyre D, Durand E, Proenca C, Marre M, Hadjadj S et al. *Post genome-wide association studies of novel genes associated with type 2 diabetes show gene-gene interaction and high predictive value.* *PLoS ONE* 2008;3:e2031.
  38. Cornelis MC, Qi L, Zhang C, Kraft P, Manson J, Cai T et al. *Joint effects of common genetic variants on the risk for type 2 diabetes in U.S. men and women of European ancestry.* *Ann Intern Med* 2009;150:541-50.
  39. van HM, Dehghan A, Witteman JC, van Duijn CM, Uitterlinden AG, Oostra BA et al. *Predicting type 2 diabetes based on polymorphisms from genome-wide association studies: a population-based study.* *Diabetes* 2008;57:3122-8.
  40. Lango H, Palmer CN, Morris AD, Zeggini E, Hattersley AT, McCarthy MI et al. *Assessing the combined impact of 18 common genetic variants of modest effect sizes on type 2 diabetes risk.* *Diabetes* 2008;57:3129-35.
  41. de Miguel-Yanes JM, Shrader P, Pencina MJ, Fox CS, Manning AK, Grant RW et al. *Genetic risk reclassification for type 2 dia-*

- betes by age below or above 50 years using 40 type 2 diabetes risk single nucleotide polymorphisms. *Diabetes Care* 2011;34:121-5.
42. Talmud PJ, Hingorani AD, Cooper JA, Marmot MG, Brunner EJ, Kumari M et al. *Utility of genetic and non-genetic risk factors in prediction of type 2 diabetes: Whitehall II prospective cohort study.* *BMJ* 2010;340:b4838.
43. Sparso T, Grarup N, Andreassen C, Albrechtsen A, Holmkvist J, Andersen G et al. *Combined analysis of 19 common validated type 2 diabetes susceptibility gene variants shows moderate discriminative value and no evidence of gene-gene interaction.* *Diabetologia* 2009;52:1308-14.
44. Lin X, Song K, Lim N, Yuan X, Johnson T, Abderrahmani A et al. *Risk prediction of prevalent diabetes in a Swiss population using a weighted genetic score-the CoLaus Study.* *Diabetologia* 2009;52:600-8.
45. Qi Q, Li H, Wu Y, Liu C, Wu H, Yu Z et al. *Combined effects of 17 common genetic variants on type 2 diabetes risk in a Han Chinese population.* *Diabetologia* 2010;53:2163-6.
46. Bruning JC, Winnay J, Bonner-Weir S, Taylor SI, Accili D, Kahn CR. *Development of a novel polygenic model of NIDDM in mice heterozygous for IR and IRS-1 null alleles.* *Cell* 1997;88:561-72.
47. Almind K, Kulkarni RN, Lannon SM, Kahn CR. *Identification of interactive loci linked to insulin and leptin in mice with genetic insulin resistance.* *Diabetes* 2003;52:1535-43.
48. Kraft P, Yen YC, Stram DO, Morrison J, Gauderman WJ. *Exploiting gene-environment interaction to detect genetic associations.* *Hum Hered* 2007;63:111-9.
49. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorf LA, Hunter DJ et al. *Finding the missing heritability of complex diseases.* *Nature* 2009;461:747-53.
50. Bodmer W, Bonilla C. *Common and rare variants in multifactorial susceptibility to common diseases.* *Nat Genet* 2008;40:695-701.