

---

# Fattori genetici della sclerosi multipla

Filippo Martinelli Boneschi

---

La sclerosi multipla (SM) è definibile dal punto di vista epidemiologico come una malattia complessa, causata dalla combinazione ed interazione fra fattori di rischio genetici e fattori di rischio ambientali.

Il dr. Eric Lander, che è stato per alcuni anni direttore del National Institutes of Health americano (NIH), per spiegare cosa si intende per malattia complessa paragona questo modello di rischio di malattia all'atto di sparare che necessita di due azioni necessarie ma non individualmente sufficienti: caricare la pistola (che corrisponde all'essere portatore di fattori di rischio genetici) e premere il grilletto (che corrisponde all'essere esposto a fattori di rischio ambientali).

Le varianti genetiche conferiscono dunque un rischio aumentato di sviluppare la malattia, ma non la causano da sole: è necessaria, infatti, anche una esposizione a fattori di rischio ambientali predisponenti. Il modello è quindi molto diverso da quello delle malattie monogeniche o mendeliane, ad esempio la distrofia muscolare di Duchenne o la Corea di Huntington, che sono invece causate dalla mutazione di un gene che è responsabile in gran parte dello sviluppo della malattia. Da un punto di vista epidemiologico, possiamo affermare che individui geneticamente predisposti alla malattia, se esposti a fattori di rischio ambientali nel corso della propria vita, hanno una elevata probabilità di contrarre la malattia, mentre è improbabile che sviluppino la patologia individui non geneticamente predisposti, anche se esposti a fattori di rischio ambientali, e individui geneticamente predisposti che nel corso della vita non siano stati esposti a fattori di rischio ambientali specifici della malattia.

Un altro possibile modello di rischio di malattia è il “modello soglia”, secondo il quale ogni individuo sin dall'infanzia entra in contatto con fattori di rischio o

fattori protettivi che lo avvicinano o allontanano da una ipotetica soglia, oltrepassata la quale si innesca la malattia.

In gergo epidemiologico possiamo anche ricondurre il modello a quello di causa sufficiente sviluppato da Rothman,<sup>1</sup> dove per causa sufficiente si intende “un set di condizioni minime ed eventi che inevitabilmente portano alla malattia”. In malattie complesse con eterogeneità eziologica, più di una causa sufficiente è coinvolta nello sviluppo della malattia. Potremo così avere soggetti che contraggono la malattia a causa del contributo di alcuni fattori genetici e della esposizione a 3 fattori di rischio ambientali, ad esempio, ed altri nei quali la patologia si manifesta a causa del contributo di altri fattori genetici associati alla esposizione ad un quarto fattore di rischio ambientale.

Non mi soffermo in questo capitolo sui fattori di rischio ambientali (fumo di sigaretta, obesità, bassi livelli di vitamina D, pregressa infezione da virus di Epstein-Barr), che sono meno numerosi di quanti siano invece i fattori genetici, ma citerò alcuni studi che hanno dimostrato meccanismi di interazione fra fattori genetici e fattori ambientali.

Il contributo dei fattori di rischio genetici è stato inizialmente evidenziato dagli studi di aggregazione familiare, che confrontano il rischio di malattia di persone con familiari con sclerosi multipla con il rischio nella popolazione generale. Questi studi hanno per primi dimostrato come individui con un familiare di primo grado (genitore o fratello/sorella) affetto da sclerosi multipla presentino un rischio di sviluppare la malattia di venti o trenta volte superiore rispetto alla popolazione generale.<sup>2</sup>

Secondo uno studio australiano del 2013,<sup>3</sup> che ha unito dati provenienti da 12 studi internazionali e studiato il

## 4 Sclerosi multipla e malattie demielinizzanti del sistema nervoso centrale

rischio di sviluppare la malattia nel corso della vita in relazione al rapporto di parentela con una persona con sclerosi multipla, possiamo trarre le seguenti indicazioni relativamente al rischio di sviluppare la malattia nel corso della propria vita: rischio nella popolazione generale: 1 su 500 persone; rischio in un gemello monozigote di persona con SM: 1 su 5; rischio in un figlio di 2 genitori affetti da SM: 1 su 5; rischio in un gemello dizigote di persona con SM: 1 su 22; rischio in un fratello o sorella di persona con SM: 1 su 37; rischio in un figlio di genitore affetto da SM: 1 su 67; rischio in genitore di figlio affetto da SM: 1 su 48.

ambientali di rischio ma non il materiale genetico di persone affette.

Se il rischio di sviluppo di sclerosi multipla di una persona con un fratello o una sorella con sclerosi multipla è di 1 su 37, non dissimile dal rischio nella popolazione generale di avere una malattia ben più frequente come il diabete, la indicazione clinica è quella di consigliare a familiari di primo grado di persone con la malattia di prevenire la malattia cercando di adottare uno stile di vita sano che permetta di evitare sin dalla età pediatrica ed adolescenziale la esposizione ai fattori di rischio ambientali noti.

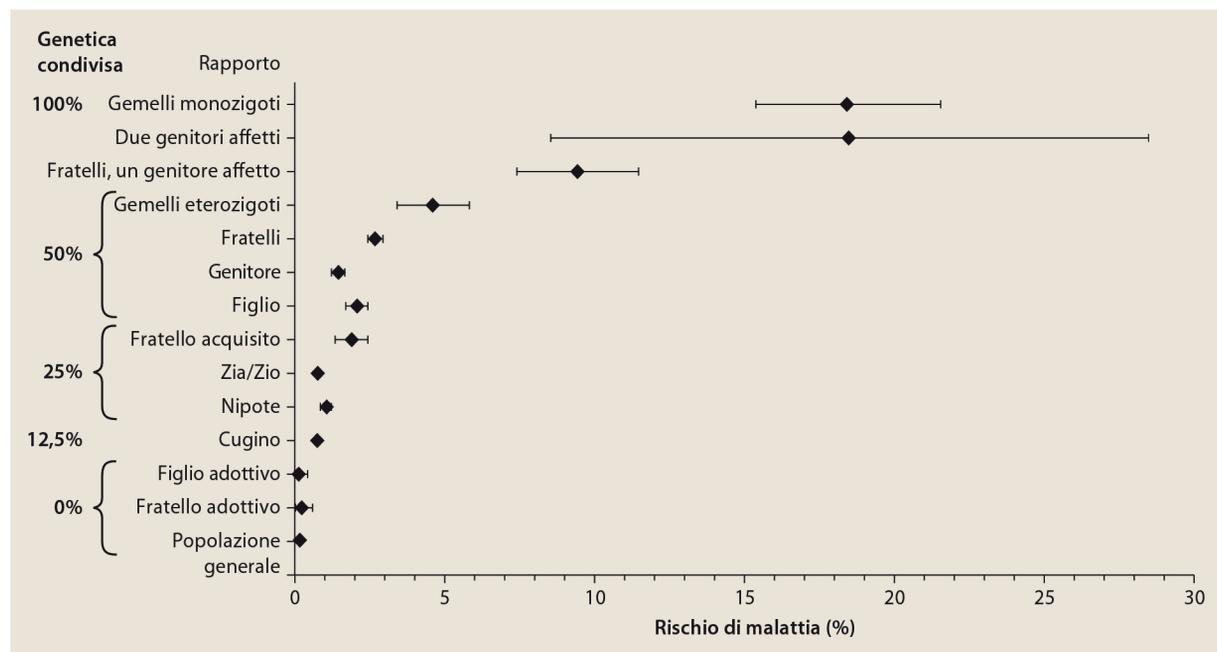


Figura 1a.1 Rischio di SM in rapporto alla relazione di parentela ed alla percentuale di materiale genetico condiviso.<sup>3</sup>

Come si vede dal grafico (Figura 1a.1),<sup>3</sup> che mette in relazione il rischio di sviluppare la malattia nel corso della propria vita con il grado di parentela e dunque con la percentuale di materiale genetico condiviso, la probabilità si riduce dal 20% (che corrisponde ad 1 su 5) in gemelli monozigoti e figli di due genitori entrambi affetti da SM che condividono il 100% di materiale genetico al 2-5% in familiari che condividono il 50% di materiale genetico (fratelli, genitori o figli di 1 persona affetta) ed ulteriormente in familiari che condividono il 25% (fratelli che condividono solo un genitore, oppure zii e nipoti) ed il 12.5% del corredo genetico (cugini di persona affetta da SM). Il grafico mostra anche che il rischio di malattia non è diverso da quello della popolazione generale in figli adottivi e fratelli acquisiti, che rappresentano una popolazione di controllo che condivide la esposizione a fattori

Una interessante recente meta-analisi<sup>4</sup> ha mostrato una prevalenza dell'11.8% (95% di intervalli di confidenza (IC): 10.7-13%) di forme familiari di sclerosi multipla in 16.179 familiari intervistati, che era più frequente nelle forme pediatriche (15.5%) rispetto a quelle adulte (10.8%) e più frequente nelle donne rispetto agli uomini (15.4% vs. 13.7%). Lo studio ha anche mostrato che la prevalenza delle forme familiari tende ad abbassarsi all'aumentare della latitudine e della prevalenza di malattia. Da questa evidenza è iniziata la ricerca dei fattori genetici coinvolti nel rischio di sviluppo di sclerosi multipla.

Nei primi anni 2000 era già nota la associazione del rischio di SM con una variante allelica, chiamata HLA-DRB1\*1501, presente sul braccio corto del cromosoma 6 in una regione del genoma di 4 megabasi (Mb) chiamata HLA (*Human Leukocyte Antigen*), ricca di

geni che codificano proteine con un ruolo nel funzionamento del sistema immunitario. In particolare, l'allele HLA-DRB1\*1501 è una molecola MHC di classe II con un ruolo nel presentare l'antigene, ed è normalmente espresso sulla superficie di cellule presentanti l'antigene specializzate come le cellule dendritiche, i macrofagi, le cellule B ed alcune cellule T attivate. È associato a numerose patologie, fra le quali la sindrome di Goodpasture, l'artrite reumatoide giovanile, il lupus eritematoso sistemico, la sindrome di Sjogren e la uveite intermedia. Conferisce un incremento di rischio di 2.5-3 volte di sviluppare la SM confermato in diversi studi su casistiche diverse, ed è il più forte fattore di rischio genetico nella SM. In particolare, in una meta-analisi che ha analizzato 5401 casi di SM e 7538 controlli l'incremento di rischio associato all'essere portatore dell'allele è stato di 3.35 (95% IC: 2.95-3.80) nella popolazione caucasica.<sup>5</sup> Altri studi hanno mostrato che nelle popolazioni afro-americane l'incremento di rischio è attenuato ad un odds-ratio (OR) di 1.8. L'HLA-DRB1\*1501 è usato come fattore di rischio genetico negli studi di interazione gene-ambiente, come vedremo più avanti nel capitolo.

In ambito scientifico sappiamo che sono sufficienti numeri relativamente piccoli per individuare e scoprire fattori di rischio con effetto individuale elevato sulla malattia (ad esempio, l'effetto dell'allele HLA-DRB1\*1501 superiore a 3.0 era già stato scoperto da studi genetici con numerosità campionaria inferiore), mentre sono necessari numeri maggiori per individuare fattori di rischio con effetto individuale minore. Nel caso specifico della sclerosi multipla, classificata come malattia complessa, studi di popolazione, studi sulle famiglie con più casi di SM e studi molecolari hanno fornito un rilevante supporto nell'ipotizzare un modello poligenico di ereditabilità, costituito prevalentemente da varianti alleliche comuni nella popolazione generale e che causassero un incremento di rischio di malattia in un range fra 1.1 e 1.3. Dunque, la ipotesi sperimentale era che ci fossero altre varianti genetiche, ma con un effetto individuale lieve sulla malattia, e che fossero necessari studi con ampie casistiche.

Il primo elemento che ha permesso di studiare casistiche di numerosità elevata è stata la creazione di un consorzio internazionale di genetica (chiamato IMMSGC, *International Multiple Sclerosis Genetics Consortium*; <https://imsgc.net/>), fondato nel 2003 per studiare le basi genetiche della sclerosi multipla, che ad oggi comprende più di duecento laboratori in tutto il mondo che hanno scelto di collaborare e condividere i propri dati, permettendo così di confrontare il

patrimonio genetico di decine di migliaia di persone con la malattia con quello di soggetti senza malattia. In sostanza, è stato possibile eseguire un esperimento scientifico che nessuno dei laboratori coinvolti avrebbe potuto svolgere da solo. Il secondo elemento contributivo è stato l'avanzamento tecnologico degli ultimi anni, che ha portato allo sviluppo di tecnologie sempre più sofisticate, in particolare la tecnologia array implementata dall'azienda Illumina® che permette di tipizzare centinaia di migliaia di varianti genetiche nel genoma, chiamate polimorfismi a singola base (o SNP), e la tecnologia di sequenziamento di nuova generazione, che ha significativamente accelerato i tempi di lettura di ampie sequenze del genoma. La tecnologia array ha permesso di eseguire i cosiddetti studi caso-controllo di analisi su tutto il genoma (chiamati genome-wide association studies, o GWAS). In questi studi viene confrontata la frequenza allelica di centinaia di migliaia di SNPs sparsi in tutto il genoma in persone affette da SM (casi) con persone non affette (controlli), e la associazione con la malattia è causata da una differenza significativa di frequenza allelica in un gruppo rispetto all'altro.

Il primo GWAS in SM è stato condotto dal consorzio IMMSGC, ed ha confrontato il genoma di 931 trio (persona affetta da SM ed i genitori) con 2431 controlli con una successiva fase di replicazione in 609 trio e 2987 controlli, portando alla identificazione di due varianti in geni codificanti due recettori delle citochine, la subunità alfa del recettore dell'interleuchina 2 (*IL2RA*) e la subunità alfa del recettore dell'interleuchina 7 (*IL7RA*), che avevano mostrato una significatività statistica che resisteva all'aggiustamento per test multipli e corrispondente ad una probabilità  $<5 \times 10^{-8}$ .<sup>6</sup> Per quanto riguarda *IL2RA*, anche chiamato CD25 e localizzato a livello del cromosoma 10p15, la variante allelica associata si chiama rs12722489 ( $P=2 \cdot 96 \times 10^{-8}$ ; OR dell'allele C: 1.25) e risiede in un introne del gene. Una recente meta-analisi ha confermato l'associazione della variante nella popolazione caucasica (OR: 1.2, 95% CI: 1.12-1.29) ma non in quella asiatica (OR: 1.1, 95% CI: 0.75-1.63), ed anche la associazione di un'altra variante allelica, chiamata rs2104286, che invece è risultata associata sia nella popolazione caucasica (allele A: OR: 1.9, 95% CI: 1.13-1.25) che asiatica (OR: 1.25, 95% CI: 1.01-1.55).<sup>7</sup> Le varianti associate chiamano in causa il ruolo del recettore di IL-2 (IL-2R), che è una molecola eterotrimerica presente nella forma completa solo in linfociti attivati dall'incontro con l'antigene. Le tre subunità del recettore sono chiamate alfa (CD25), beta (CD122) e

gamma (CD132). In condizioni normali solo le subunità beta e gamma sono presenti, mentre in caso di attivazione della cellula viene sintetizzata la subunità alfa che in tal modo si associa formando un trimero ad elevata affinità. L'IL-2R dimerica è espressa dalle cellule T CD8 + e dalle cellule NK, mentre le cellule T regolatorie e le cellule T attivate esprimono alti livelli di IL-2R trimerica. L'IL-2 è un tipo di interleuchina prodotta dalle cellule T CD4 + e CD8 + attivate in seguito all'interazione con l'antigene, ed è necessaria al differenziamento e all'espansione dei linfociti T. A conferma dello stretto ruolo fra varianti genetiche di rischio e terapie per la cura della malattia, è da segnalare l'uso nella SM di un anticorpo monoclonale, chiamato daclizumab, in grado di prevenire la formazione di recettori ad alta affinità per IL-2.<sup>8</sup> Il farmaco si è rivelato efficace nella cura della malattia, ma ha causato numerosi casi di encefalite e meningoencefalite che ne hanno causato il ritiro dal mercato.

La seconda variante emersa dallo studio è la rs6897932 ( $p=2.94 \times 10^{-7}$ ; allele C: OR: 1.18), ed è una variante codificante che codifica l'aminoacido treonina o isoleucina in posizione 244 dell'esone 6 di una delle isoforme del gene *IL7RA*. Una recente meta-analisi ha confermato una associazione significativa dello SNP nella popolazione generale e negli Europei, che invece non risulta evidente nella popolazione Asiatica ed Americana.<sup>9</sup> Il recettore di IL-7 (IL-7R) è una molecola eterodimerica costituita da due subunità, la subunità alfa IL-7RA (CD127) ed il comune recettore gamma (CD132) che è condiviso con diverse citochine, quali IL-2, -4, -9 e -15. È un recettore espresso su diverse cellule, includenti le cellule T naïve e le cellule T memoria. L'interleuchina 7 è una molecola con un ruolo importante nello sviluppo e nella maturazione dei linfociti B e T. Numerose malattie sono associate dal punto di vista genetico ad *IL7RA*, oltre la SM,<sup>10</sup> fra le quali la leucemia linfoblastica acuta,<sup>11</sup> l'artrite reumatoide e l'artrite giovanile idiopatica.<sup>12</sup> Questo primo studio di associazione su tutto il genoma è stato importante perché ha per primo individuato una variante genetica al di fuori della regione HLA e mostrato che gli studi di associazione rappresentavano il disegno di studio più efficace per individuare i fattori genetici associati alla SM. Inoltre, la scoperta delle associazioni con il recettore di IL2 ed IL7 ha sottolineato il coinvolgimento del sistema immunitario nella fisiopatologia della SM.

Studi successivi GWAS con progressivo incremento della dimensione campionaria hanno raddoppiato il numero di associazioni di varianti genetiche a 49 in

uno studio pubblicato nel 2011 che aveva confrontato il patrimonio genetico di 9772 casi appartenenti a 15 paesi diversi con quello di 17376 controlli usando la tecnologia array.<sup>13</sup> Le evidenze più importanti di questo secondo GWAS del consorzio sono state la conferma del ruolo del sistema immunitario nello sviluppo della SM ed anche la presenza di una sostanziale sovrapposizione di varianti genetiche che sono risultate associate a malattie autoimmuni diverse fra loro, ad indicare la presenza di una suscettibilità genetica in generale allo sviluppo di autoimmunità. Alla luce di questo dato, il consorzio IMMSGC ha condotto un terzo studio GWAS usando un array di genotipizzazione customizzato, chiamato Immunochip, in grado di studiare in maggiore dettaglio circa 200 regioni del genoma associate a diverse patologie autoimmuni. Il basso costo dell'array ha permesso in questo lavoro di genotipizzare un totale di 80.000 individui che hanno aumentato il numero di varianti con associazione su tutto il genoma a 110.<sup>14</sup> Circa un quinto delle varianti associate si sovrapponeva a quelle di un'altra malattia autoimmune. In particolare, il 9.1% delle varianti SM si sovrapponeva con il rischio di sviluppare la malattia infiammatoria intestinale, includente la malattia di Crohn e la retto-colite ulcerosa, il 4.5% con la celiachia e l'artrite reumatoide, il 2.7% con la tiroidite autoimmune e lo 0.9% con la psoriasi.<sup>14</sup> Il dato genetico riceve conferma anche dal punto di vista epidemiologico. Infatti, un lavoro del mio gruppo ha mostrato che circa un terzo delle persone con sclerosi multipla avevano almeno un familiare con una diagnosi di malattia autoimmune, fra le quali le più frequenti erano la tiroidite autoimmune, la psoriasi e l'artrite reumatoide.<sup>15</sup> Un altro lavoro di revisione della letteratura ha mostrato invece come nel medesimo soggetto le comorbidità più frequenti con la sclerosi multipla fossero la psoriasi, le malattie tiroidee, la malattia infiammatoria intestinale ed altre patologie autoimmuni più rare come l'uveite.<sup>16</sup>

Il più recente e vasto lavoro del consorzio ha infine confrontato il patrimonio genetico di 47.429 soggetti con sclerosi multipla con quello di 68.374 soggetti usando array in grado di coprire tutto il genoma, il sopra citato Immunochip ed un terzo array sviluppato specificatamente per indagare in maggiore dettaglio i loci associati alla SM, chiamato MSchip.<sup>17</sup> Questo studio ha permesso di individuare 233 varianti genetiche con associazione su tutto il genoma, includenti 200 varianti fuori dalla regione HLA, 1 variante sul cromosoma X e 32 segnali indipendenti nella regione HLA.

L'associazione sul cromosoma X (rs2807267; OR per l'allele T: 1.07;  $P=6.86 \times 10^{-9}$ ) è stata la prima riportata in una malattia con una forte preponderanza di donne affette (più di 2 donne affette per ogni uomo). La potenza dello studio ha permesso di individuare varianti di rischio con una frequenza allelica del 2.1%, suggerendo la possibilità che esistano altre varianti di rischio per la SM con una frequenza minore nella popolazione ancora da scoprire.

Inoltre, questo studio ha stimato che le varianti genetiche individuate, sia quelle con associazione su tutto il genoma che quelle nella regione HLA che quelle con suggestiva associazione che non superano però la soglia genome-wide ( $5 \times 10^{-8} < P < 10^{-5}$ ), sono in grado di spiegare fino al 48% della ereditabilità della SM, ovvero della proporzione di variabilità fenotipica spiegata dai fattori genetici, a significare che ci siano altre varianti ancora da individuare. A questo 48% contribuisce la regione HLA con il 20.2%, i segnali GWAS con il 18.3%, ed i segnali suggestivi di associazione con il 9%.

Per quanto riguarda la regione HLA, lo studio ha individuato 32 segnali indipendenti, che includono la conferma di un ruolo protettivo da parte dell'allele HLA-A\*0201.<sup>13, 17</sup> Al tempo stesso è da segnalare quanto non sia ancora chiaro l'esatto ruolo della regione MHC nella SM, nonostante i numerosi studi GWAS e di sequenziamento della regione HLA effettuati.

Il lavoro ha confermato che circa un quinto dei 200 loci fuori dalla regione HLA hanno una associazione anche con un'altra malattia autoimmune, similmente a quanto era già stato riportato dal lavoro con l'ImmunoChip.<sup>14</sup> Avendo accesso solo a dati a livello aggregato di SNPs, un recente lavoro ha usato una tecnica, chiamata score di regressione basata sul punteggio LD fra patologie diverse, per calcolare la correlazione genetica fra patologie diverse.<sup>18</sup> Applicando questo score, si è evidenziato che la SM ha mostrato una correlazione significativa che oscillava fra 0.16 e 0.35 con quasi tutte le malattie autoimmuni tranne la celiachia.<sup>19</sup> Un altro lavoro ha applicato lo stesso metodo per stimare la correlazione fra SM ed altre malattie neurologiche, quali la Malattia di Alzheimer, epilessia, emicrania, ictus, e psichiatriche, e non ha trovato alcuna correlazione con la SM.<sup>20</sup>

Alcuni esempi di varianti e geni associati a più malattie autoimmuni è il locus del gene *TAGAP* (*T-cell activation RhoGTPase activating protein*), che è risultato associato ad SM, diabete di tipo 1 ed artrite reumatoide con un effetto di associazione ma direzione oppo-

sta sul rischio di celiachia.<sup>21</sup> Un altro locus riguarda un altro gene, *TYK2*, il cui snp rs34536443 mostra una consistente direzione di aumento di rischio in chi è portatore dell'allele C di sviluppare la SM (OR: 1.2), la cirrosi biliare primitiva (OR: 1.91), la psoriasi (OR: 1.88), il diabete di tipo 1 (OR: 1.88) e l'artrite reumatoide (OR: 1.38).<sup>17</sup> *TYK2* è una tirosin-chinasi non recettoriale, con un ruolo nel signaling di IL-12 e dell'Interferone di tipo 1.

Possiamo quindi concludere che l'approccio GWAS è stato estremamente efficace nell'individuare varianti comuni di rischio per SM con frequenza nella popolazione superiore all'1%. Purtroppo, gli array GWAS non sono ottimizzati per studiare varianti a bassa frequenza o varianti rare, definite come varianti con frequenza inferiore al 5% nella popolazione generale. Per studiare il ruolo di queste varianti, sono stati usati due approcci: la tecnologia array attraverso gli exome-array che contengono varianti genetiche a bassa frequenza contenute all'interno delle sequenze codificanti dei geni, ed approcci di sequenziamento di esoma e genoma.

Un lavoro del consorzio IMSSC condotto su 32.367 casi e 36.012 controlli di origine europea che ha usato l'exome array ha individuato 5 varianti codificanti associate alla SM, di cui 3 non evidenziate nello studio GWAS, e solo 2 con frequenza allelica inferiore al 2%.<sup>22</sup> Si tratta di una variante missenso (rs148755202; R166H) nel gene *HDAC7* con una frequenza dell'1.4% ed una variante missenso (rs61734100; I942M) nel gene *NLRP8* con una frequenza dello 0.2%, entrambe con un ruolo protettivo sulla malattia. *HDAC7* codifica per l'istone deacetilasi 7 di classe 2 che è in grado di potenziare l'effetto repressivo di *FOXP3*, regolatore della maturazione delle cellule CD4+ naive in cellule T regolatorie,<sup>23, 24</sup> mentre *NLRP8* è un recettore intracellulare con un ruolo nelle risposte del sistema immunitario innato. La mutazione riscontrata è stata rilevata solo in individui con origini europee secondo ExAC, consistente con l'alta prevalenza della SM nelle popolazioni di origini europee.

Le altre 3 varianti scoperte dallo studio sono presenti nel gene *GALC*, o galattosilceramidasi, che è un enzima che rimuove il galattosio dai derivati della ceramide, nel gene *TYK2* del quale abbiamo già parlato, e nel gene *PRF1* che codifica per la perforina, che è una proteina citolitica presente nei granuli dei linfociti T citotossici e delle cellule Natural Killer. Il lavoro ha anche mostrato che le varianti rare individuate spiegano fino al 5% dell'ereditabilità, valori inferiori rispetto alle varianti comuni riscontrate negli studi GWAS.

L'altro approccio per individuare varianti rare usato è stato quello degli studi di sequenziamento dell'esoma e del genoma. Fra i vari lavori, cito il lavoro del mio gruppo e del gruppo del dr. Vilarino in Canada che, usando un approccio di sequenziamento dell'esoma in casi familiari di SM, hanno rilevato e replicato un ruolo per un gene chiamato UBR2, in particolare guidato dalla variante rs62414610, che è un gene che codifica per la E3 ubiquitin-protein ligasi ed è coinvolto nel riconoscimento e poliubiquitinazione di proteine con alterato residuo N-terminale, permettendone la degradazione nei proteasomi.<sup>25, 26</sup> Un altro lavoro ha individuato, attraverso il sequenziamento del genoma di 38 casi di SM primariamente progressiva (SMPP) e di 81 controlli sani di origine europea, la presenza di 3 varianti patogenetiche causative di disturbi neurologici simili alla SM a livello dei geni *KIF5A*, *MLC1* e *REEPI*, ed un significativo arricchimento di mutazioni legate a forme di paraparesi spastica nei casi di SMPP rispetto ai controlli sani che non era confermato nel confronto fra forme recidivanti-remittenti (RR) rispetto a controlli a supporto della ipotesi che varianti genetiche rare possano contribuire alla forma SMPP e non a quella RR.<sup>27</sup>

Di sicuro interesse è anche lo studio degli endofenotipi, ovvero di quelle variabili quantitativamente misurabili che possono fungere da connessione tra la variabilità genetica legata alla malattia e le manifestazioni cliniche del disturbo. Fra questi annoveriamo le misure di severità di malattia, la presenza di bande oligoclonali specifiche su liquor ed assenti nel siero, il numero di ricadute di malattia, ed i parametri di Risonanza Magnetica Nucleare (RMN) associati alla severità della malattia.

Riguardo alla severità di malattia, a breve saranno pubblicati i dati del consorzio IMSGC che hanno evidenziato la presenza di varianti genetiche in grado di influenzare la severità di malattia calcolata con l'ARMMS (age-related MS severity score). È nota, inoltre, la associazione fra l'essere portatore dell'allele HLA-DRB1\*1501 ed una più precoce età di esordio di malattia, con una anticipazione quantificata nel secondo GWAS del consorzio di 10.6 mesi (+1.9 mesi).<sup>13</sup> Per quanto riguarda la RMN, ci sono studi che hanno indagato il ruolo delle varianti genetiche sulla concentrazione di glutammato a livello cerebrale, sulla distribuzione spaziale delle lesioni nel cervello e sullo spessore corticale cerebrale, senza però evidenza di alcun segnale che superasse la soglia di significatività genome-wide a causa della ridotta dimensione campionaria. Un lavoro ha invece trovato una associazione fra uno

score genetico calcolato a partire dalle varianti associate nella regione HLA con l'atrofia della sostanza grigia sottocorticale nelle donne con forma recidivante-remittente di malattia ( $p_{2.3 \times 10^{-4}}$ ), con un segnale prevalentemente guidato dall'aplotipo HLA-DRB1\*15:01.<sup>28</sup> Il quesito che si pone è quale sia il valore clinico della scoperta dei fattori genetici coinvolti nel rischio di sclerosi multipla, che cercherò di riassumere:

1. Una migliore conoscenza della eziopatogenesi della malattia. In primo luogo, è importante segnalare che la maggior parte delle varianti comuni individuate come associate alla SM risiedono in regioni intergeniche, ovvero in regioni del genoma al di fuori dei geni, o in regioni introniche, o non codificanti, di geni. La maggior parte di queste varianti sembra avere un ruolo regolatorio, ovvero di effetto sulla espressione di geni del genoma in vicinanza (effetto cis) o lontananza (in trans) spaziale. Nell'ultimo lavoro del consorzio<sup>17</sup> sono stati usati diversi approcci per identificare il ruolo funzionale delle varianti genetiche scoperte. In particolare, si è data priorità a geni la cui espressione genica era modulata dalle varianti trovate associate (ovvero, geni in cis-eQTL, che significa cis-expression quantitative trait loci), geni che avessero una variante esonica in linkage disequilibrium ( $r^2 > 0.1$ ) con uno dei 200 segnali, geni che avessero alti punteggi usando un'analisi di network con i geni trovati associati, e/o geni che avessero un pattern di co-espressione simile ai geni trovati associati usando il programma bioinformatico DEPICT. Usando questo approccio è stato possibile attribuire un ruolo funzionale delle 200 varianti genetiche associate alla SM a 551 geni. Si è poi scoperto che questo set di 551 geni è arricchito nelle cellule del sistema immunitario adattativo, ovvero le cellule B e T e Natural Killer, che sono cellule citotossiche importanti nel riconoscimento e nella eliminazione di cellule tumorali o infette da virus, e nelle cellule del sistema immunitario innato, quali le cellule mieloidi e le cellule microgliali che sono in grado di sondare il sistema nervoso centrale in cerca di neuroni danneggiati, placche demielinizzanti ed agenti infettivi da rimuovere. Questo dato richiama il ruolo della componente B cellulare nella malattia, ed al tempo stesso sottolinea il possibile ruolo delle cellule microgliali nello sviluppo della malattia.

Inoltre, in alcuni casi esperimenti funzionali hanno permesso di spiegare il meccanismo biologico dell'associazione genetica. Ad esempio, l'associazione con il gene *TNFRSF1A* che codifica per il recettore TNF1 (TNFR1) a livello dello SNP rs1800693

il cui allele G aumenta il rischio di sviluppare la SM (OR 1.12,  $P=4.1 \times 10^{-14}$ ).<sup>29, 30</sup> Studi funzionali hanno mostrato che l'allele G di rischio porta alla espressione di un trascritto del TNFR1 che non ha l'esone 6 e si traduce in una forma solubile della proteina con effetto antagonista sul TNF, analogo a quello causato dai farmaci anti-TNF (infliximab, adalimumab, certolizumab, golimumab) che sono stati studiati in ambito SM ma interrotti nel loro utilizzo a causa dell'effetto di incremento della frequenza di ricadute ed attività di malattia in chi li assumeva.<sup>30, 31</sup>

Un'altra associazione coinvolge lo snp rs11810217, il cui allele A aumenta il rischio di SM (OR: 1.15;  $P=5.85 \times 10^{-15}$ ).<sup>32</sup> Studi di meta-analisi e fine caratterizzazione della regione genomica 1p22.1 dove si trova lo SNP hanno confermato che il segnale di associazione era guidato primariamente da questa variante, presente in un introne del gene EVI5 (sito 5 di integrazione virale exotropa).<sup>32</sup> Attraverso ulteriori analisi di proteomica dell'interattoma di EVI5 si è scoperto che una variante funzionale non sinonima del gene EVI5 in LD con la variante rs11810217, rs11808092, interagiva con la sfingosina-1-fosfato liasi (SGPL1), che è un'enzima chiave per il funzionamento del recettore della sfingosina-1-fosfato (S1P) che è anche il target di farmaci orali usati nella SM come il fingolimod, siponimod e ozanimod.

Altro gene di interesse è l'*NFKB1*. Uno studio ha mostrato che varianti prossimali al gene *NFKB1* ed una variante intronica del gene TNFRSF1A erano associate ad un incremento dell'attivazione del fattore di trascrizione NFkB (Nuclear factor kB, NFkB), che rappresenta il primo fattore di trascrizione individuato con un ruolo centrale nei processi infiammatori ed anche nella attivazione degli astrociti.<sup>33</sup> Un ulteriore studio ha analizzato la variante rs7665090 in prossimità di *NFKB1*, nota per essere associata alla SM (allele G: OR: 1.09,  $P=1.00 \times 10^{-8}$ ), mostrando che era associata ad un aumento del signaling di NFkB e ad una upregolazione di geni target di NFkB in astrociti umani in coltura ed in astrociti presenti all'interno di lesioni encefaliche di pazienti SM, collegando questa variante al funzionamento degli astrociti.<sup>34</sup> Altri esempi di varianti funzionali riguardano lo SNP rs719316 presente in un introne del gene *ATXN1* che secondo studi funzionali sarebbe in grado di regolare la trascrizione ed il signaling di una popolazione di cellule B, chiamate B-1a, ed il locus del gene *MERTK* che è molto espresso nelle cellule microgliali segnalate come implicate nella malattia dall'ultimo lavoro del consorzio.<sup>17</sup>

2. Scoperta di nuovi bersagli terapeutici della malattia. Come si evince dal precedente punto, c'è una forte e stretta relazione fra targets genetici e molecole in grado di curare la malattia. Un recente studio ha individuato, a partire dai dati di associazione genetica, venti proteine associate alla SM che possono essere il bersaglio di agenti farmacologici, dei quali 13 sono già studiati in studi clinici.<sup>35</sup>
3. Predizione del rischio di malattia. Le varianti associate alla SM determinano un modesto aumento di rischio individuale di malattia, ma il loro contributo può essere aggregato per sviluppare score genetici di rischio di malattia. Esistono pertanto score genetici aggregati che tengono conto e sommano il contributo individuale delle varianti genetiche sul rischio di malattia, chiamati weighted genetic risk score (wGRS). Il primo utilizzo è stato in un lavoro del gruppo di De Jager, che ha calcolato uno score basato su 16 varianti genetiche di rischio, mostrando che chi aveva uno score di più di 1.25 deviazioni standard dalla media aveva un più elevato rischio di SM.<sup>36</sup> Studi più recenti hanno usato score poligenici includendo un numero più elevato di varianti associate alla SM fino alla quasi totalità, mostrando una associazione con il rischio di SM sia nella popolazione adulta che in quella pediatrica. Uno studio americano ha sviluppato uno score poligenico che includeva circa 160.000 varianti genetiche, mostrando una associazione fra un più elevato score genetico per SM ed alcune metriche di risonanza magnetica dell'encefalo quali una più elevata alterazione della micro-struttura della sostanza bianca in un campione di soggetti sani fra gli 8 e i 12 anni.<sup>37</sup> Il lavoro ipotizza che esista una finestra temporale durante lo sviluppo del sistema nervoso durante la quale le varianti associate alla SM agiscano sulla maturazione cerebrale, in particolare della sostanza bianca.
4. Modelli di interazione fra gene ed ambiente. I primi studi di interazione fra genetica ed ambiente provengono dalle coorti di pazienti svedesi ed americane, che hanno potuto analizzare dati genetici e dati sulla esposizione a fattori di rischio ambientali in ampie casistiche di pazienti con SM. Riguardo all'obesità, uno studio ha mostrato che l'obesità a vent'anni aumenta il rischio di malattia soprattutto nei soggetti portatori dei fattori genetici di rischio della malattia evidenziati dall'avere l'allele di rischio HLA-DRB1\*1501 e dal non avere l'allele protettivo HLA-A\*02, con un aumento di rischio di sviluppo di malattia di più di dieci volte nei soggetti

geneticamente predisposti.<sup>38</sup> Un altro studio americano-svedese ha messo in luce come l'essere fumatore aumenti significativamente il rischio di sclerosi multipla nei soggetti portatori di una variante allelica a livello del gene NAT1, che codifica per un enzima con un ruolo nella detossificazione di agenti chimici.<sup>38</sup> Un ulteriore lavoro ha mostrato che l'aumento di rischio di SM dato dall'essere fumatore era di 13.5 volte nei soggetti geneticamente predisposti alla sclerosi multipla (portatori dell'allele DRB1\*15 e privi dell'allele A\*02) e solo di 1.4 volte in quelli senza predisposizione genetica, mentre era di 4.9 volte nei soggetti non fumatori geneticamente predisposti.<sup>40</sup> Questi dati confermano l'ipotesi che una stimolazione della risposta immunitaria nei polmoni causata dal fumo di sigaretta possa causare la malattia in soggetti geneticamente predisposti.<sup>40</sup>

Altro approccio molto usato sono gli studi di randomizzazione mendeliana, che hanno la capacità di individuare una relazione causale fra fattore di rischio e malattia. Questi studi usano le proprietà di varianti genetiche somatiche nella forma di SNPs fortemente associate ad un fattore di esposizione (ad esempio, obesità o livelli di vitamina D o fumo di sigaretta) Come strumento di misura di quella variabile e ne stimano l'effetto causale sull'outcome di interesse, in questo caso il rischio di SM. Tale approccio ha mostrato la presenza di una relazione causale con la SM di bassi livelli di vitamina D,<sup>41</sup> alto *Body Mass Index*<sup>41</sup> ed obesità durante l'adolescenza,<sup>42</sup> mentre non per il fumo di sigaretta.<sup>43</sup>

Concludo il capitolo parlando del futuro. Le tecnologie di sequenziamento di singola cellula stanno rivoluzionando la nostra capacità di comprendere i fenomeni biologici, ed hanno permesso, attraverso l'analisi di tessuti del sistema nervoso centrale post-mortem, di individuare sottotipi di cellule con un possibile ruolo funzionale sulla malattia quali oligodendrociti, cellule microgliali, astrociti e neuroni.<sup>44</sup> Di pari importanza anche gli studi sul sistema immunitario, in maniera particolare eseguiti sul liquor e sangue di pazienti e soggetti sani, che hanno permesso di aumentare la diversità delle cellule del sistema immunitario e di studiare i processi di espansione delle cellule T citotossiche e la loro interazione con le cellule B.<sup>45</sup>

Altri argomenti di interesse sono rappresentati dalla comprensione dei meccanismi di interazione fra genetica ed ambiente attraverso modelli biologici quali la epigenetica, letteralmente "ciò che sta sopra la genetica". In quanto esseri umani, siamo infatti composti da cento milioni di cellule, che hanno lo stesso codice genetico ma si dividono in duecento tipi, molto diversi gli uni dagli altri. Come questo sia possibile è spiegato appunto dall'epigenetica, che è l'insieme di tutte quelle modificazioni ereditabili in grado di influenzare l'espressione di questi geni e la sintesi di proteine che generano cellule diverse senza modificare il codice genetico. L'epigenetica spiega ad esempio il perché anche gemelli monozigoti, che condividono quasi il 100% di materiale genetico, possano avere destini di vita molto diversi fra loro.



## Bibliografia essenziale

*Scan qui per consultare la bibliografia completa*

- Brainstorm Consortium, Anttila V, Bulik-Sullivan B, Finucane HK, *et al.* Analysis of shared heritability in common disorders of the brain. *Science* 2018;360:eaap8757.
- Gregory AP, Dendrou CA, Attfield KE, *et al.* TNF receptor 1 genetic risk mirrors outcome of anti-TNF therapy in multiple sclerosis. *Nature* 2012;48:508-11.
- International Multiple Sclerosis Genetics Consortium (IMSGC), Beecham AH, Patsopoulos NA, Xifara DK, *et al.* Analysis of immune-related loci identifies 48 new susceptibility variants for multiple sclerosis. *Nat Genet* 2013;45:1353-60.
- International Multiple Sclerosis Genetics Consortium, Hafler DA, Compston A, Sawcer S, *et al.* Risk alleles for multiple sclerosis identified by a genomewide study. *N Engl J Med* 2007;357:851-62.
- International Multiple Sclerosis Genetics Consortium, Wellcome Trust Case Control Consortium 2, Sawcer S, Hellenthal G, Pirinen M, *et al.* Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. *Nature* 2011;476:214-9.
- International Multiple Sclerosis Genetics Consortium. Low-Frequency and Rare-Coding Variation Contributes to Multiple Sclerosis Risk. *Cell* 2018;175:1679-87.
- International Multiple Sclerosis Genetics Consortium. Multiple sclerosis genomic map implicates peripheral immune cells and microglia in susceptibility. *Science* 2019;365:eaav7188.
- Jia X, Madireddy L, Caillier S, *et al.* Genome sequencing uncovers phenocopies in primary progressive multiple sclerosis. *Ann Neurol* 2018;84:51-63.
- Mascia E, Clarelli F, Zauli A, *et al.* *J Neuroimmunol* 2022;362:577760.
- O'Gorman C, Lin R, Stankovich J, *et al.* Modelling genetic susceptibility to multiple sclerosis with family data. *Neuroepidemiology* 2013;40:1-12.