



Collegio Ghislieri

Centro per la Comunicazione e la Ricerca

Progetto “Progressi in Biologia e Medicina”

11° Corso di formazione avanzata

**Medicina genomica
e terapia personalizzata
in ematologia/oncologia**

16-20 aprile 2012, Collegio Ghislieri, Pavia

A cura di Carlo Bernasconi

11° Corso di formazione avanzata

**Medicina genomica
e terapia personalizzata
in ematologia/oncologia**



Collegio Ghislieri

Centro per la Comunicazione e la Ricerca

Progetto “Progressi in Biologia e Medicina”

11° Corso di formazione avanzata

**Medicina genomica
e terapia personalizzata
in ematologia/oncologia**

16-20 aprile 2012, Collegio Ghislieri, Pavia

A cura di Carlo Bernasconi

© Copyright 2012 

Edizioni Internazionali srl
Divisione EDIMES - Edizioni Medico-Scientifiche - Pavia

Via Riviera, 39 - 27100 Pavia
Tel. 0382526253 - Fax 0382423120
E-mail: edint.edimes@tin.it

Tutti i diritti sono riservati.
Nessuna parte può essere riprodotta in alcun modo
(compresi i microfilm e le copie fotostatiche)
senza il permesso scritto dell'editore.

Indice

Prefazione pag. IX

Introduzione alla medicina genomica

1. Genoma umano e variabilità genetica » 3
Carlo Alberto Redi
2. Genetica delle malattie comuni » 12
Antonio Pizzuti
3. Il clinico di fronte alla medicina genomica » 15
Carlo Bernasconi

Strumenti di medicina genomica

4. Sequenziamento genico » 31
Elena Ronchetti
5. Microarray technologies » 34
Angela Amato
6. Cancer Biobanking » 40
Alberto Zambelli

Bioinformatica clinica

7. Banche dati biomediche e risorse on-line » 47
Marco Masseroli
8. Metodi di analisi statistica di dati di next generation sequencing » 55
Claudia Angelini
9. Il Progetto ONCO-i2b2: integrazione di dati clinici e biobanche per il supporto alla ricerca in oncologia » 61
Daniele Segagni, Valentina Tibollo, Arianna Dagliati, Alberto Zambelli, Riccardo Bellazzi

Elementi di farmacogenomica

- | | | |
|--|---|----|
| 10. Eredità e risposta ai farmaci..... | » | 69 |
| <i>Maurizio D'Incalci</i> | | |
| 11. Genoma del paziente e tolleranza farmacologica..... | » | 71 |
| <i>Diego Fornasari</i> | | |
| 12. Pharmacogenetics in cancer care: transferring translational
research into clinical practice | » | 75 |
| <i>Giuseppe Toffoli</i> | | |
| 13. La ricerca genetica per programmare lo sviluppo
di nuovi farmaci | » | 78 |
| <i>Giuseppe Recchia</i> | | |

Genomica e attualità di medicina molecolare in emopatie maligne

- | | | |
|---|---|-----|
| 14. Tecnologie genomiche per migliorare la stratificazione
prognostica e impostare terapie personalizzate nelle LAM | » | 87 |
| <i>Paolo Bernasconi, Marina Boni, Paola Maria Cavigliano,
Ilaria Giardini, Barbara Rocca, Rita Zappatore,
Irene Dambruoso, Celeste Calvello, Marilena Caresana</i> | | |
| 15. Alterazioni dello <i>splicing</i> dell'RNA nelle sindromi
mielodisplastiche..... | » | 98 |
| <i>Mario Cazzola</i> | | |
| 16. La terapia di prima linea della LMC Ph1 positiva
con imatinib può essere migliorata?..... | » | 104 |
| <i>Giuseppe Saglio</i> | | |
| 17. Basi molecolari delle malattie mieloproliferative
croniche Ph ¹ negative | » | 107 |
| <i>Alessandro Maria Vannucchi</i> | | |
| 18. Fattori genetici e microambiente nella patogenesi della LLC | » | 114 |
| <i>Federico Caligaris Cappio, Elisa Ten Hacken</i> | | |
| 19. Biological and clinical predictors of chronic lymphocytic
leukemia transformation to Richter syndrome | » | 119 |
| <i>Valeria Spina, Silvia Rasi, Alessio Brusca, Marco Fangazio,
Clara Deambrogi, Sara Monti, Stefania Cresta,
Rosella Famà, Mariangela Greco, Carmela Ciardullo,
Daniela Piranda, Lorenzo De Paoli, Davide Rossi,
Gianluca Gaidano</i> | | |

20. Genomica e proteomica nella gestione dei pazienti con amiloidosi sistemica » 133
Francesca Lavatelli, Giampaolo Merlini
21. Profilo di espressione genica nelle SMD e LAM: attualità e prospettive » 140
Sergio Ferrari

Genomica e attualità di medicina molecolare in tumori solidi

22. Basi molecolari del rischio genetico di tumori » 149
Paolo Radice
23. Caratterizzazione molecolare del tumore della mammella » 154
Umberto Magrini
24. Terapie personalizzate nel carcinoma mammario » 160
PierFranco Conte, Federico Piacentini
25. Caratterizzazione molecolare del tumore del colon-retto » 167
Alberto Bardelli
26. Diagnosi molecolare del tumore del polmone » 169
Federico Cappuzzo
27. Cancro: malattia genetica di cellule somatiche » 173
Lucio Luzzatto

Genomica e attualità di medicina molecolare in emopatie non-neoplastiche

28. Hemoglobinopathies and Thalassemia: origin of molecular medicine and new insights » 179
Maria Domenica Cappellini, Alessia Marcon, Giovanna Graziadei
29. Disordini ereditari del metabolismo del ferro » 183
Clara Camaschella
30. The evolving spectrum of inherited thrombocytopenias » 187
Carlo L. Balduini
31. Genomica e malattie autoimmuni » 205
Eleonora Gambineri

Prefazione

Nella seconda metà del secolo scorso, il principale obiettivo della ricerca genetica è stato quello di identificare i geni che stanno all'origine di "malattie monogeniche", cioè malattie ereditarie (in genere rare) causate da alterazioni di singoli geni (quali ad esempio la fibrosi cistica, l'anemia a cellule falciformi, il favismo). Questo lavoro di ricerca ha dato origine alla "genetica medica" ed a programmi di screening per la prevenzione neonatale; non ha però avuto praticamente alcun impatto sulla conoscenza della predisposizione verso malattie comuni (quali ad esempio il diabete mellito, alcune malattie cardiovascolari, parecchi tumori), che frequentemente hanno una base genetica complessa, non esplorabile con le tecniche genetiche tradizionali.

Successivamente, nella prima decade del 2000, il completamento del Progetto Genoma Umano e i continui progressi tecnologici hanno consentito di iniziare a capire le basi genetiche di un ampio spettro di malattie comuni definite "multifattoriali", perché originate dall'interazione fra molteplici fattori genetici e fattori ambientali. Sono state così poste le basi per una "medicina genomica", caratterizzata da nuovi approcci diagnostici e terapeutici per parecchie malattie comuni, e in alcuni casi anche dalla possibilità di una precoce previsione e potenziale prevenzione. Alcune nuove acquisizioni sono già entrate nella pratica clinica, e comprendono la precisa diagnosi molecolare di una determinata malattia (ad esempio di una leucemia o di un tumore) e la possibile predizione della risposta ad un farmaco (riguardante sia la tollerabilità che l'efficacia). Partendo da queste premesse, è oggi possibile in alcuni casi prescrivere un'efficace terapia mirata, personalizzata.

Sul piano pratico, la medicina genomica rappresenta una modalità completamente nuova nell'esercizio della medicina, sia al letto dell'ammalato, sia nell'organizzazione della sanità pubblica. Essa sta acquistando un ruolo sempre più importante in alcuni settori della pratica clinica, specie in ematologia/oncologia; a tal riguardo basti pensare che i tumori possono venir definiti come "malattie genetiche delle cellule somatiche". L'approccio genomico sta oggi avendo effetti determinanti sulla classificazione dei tumori, sull'identificazione di marcatori prognostici, sulla previsione della risposta ai farmaci, sul monitoraggio di una malattia neoplastica, sullo sviluppo di nuove terapie mirate, sulla possibilità di trattare in alcuni casi la predisposizione a un tumore.

Nell'organizzazione di questo 11° Corso di formazione avanzata ci è parso quindi opportuno fare una rapida rassegna delle più recenti acquisizioni concettuali e tecnologiche in questo settore della ricerca biomedica, che sempre più si sta trasferendo (grazie anche alla rapida riduzione dei tempi e dei costi delle analisi) dalla ricerca di base all'applicazione clinica. Nel programmare la trattazione dei singoli argomenti ci siamo avvalsi della preziosa collaborazione di Docenti particolarmente qualificati, ai quali va un caloroso ringraziamento da parte del Collegio Ghislieri e mio personale. Allo sforzo di aver voluto affrontare insieme un tema veramente innovativo, l'augurio di incontrare un consenso pari a quello riservato ai Corsi precedenti.

Carlo Bernasconi

Pavia, 16 aprile 2012

**INTRODUZIONE
ALLA MEDICINA GENOMICA**

1

Genoma umano e variabilità genetica

CarloAlberto Redi

Laboratorio di Biologia dello Sviluppo, Dipartimento di Biologia Animale, Università di Pavia

La pubblicazione dieci anni orsono della sequenza del genoma umano ha aperto la possibilità di studiare in un modo più sistematico e dettagliato le basi genetiche delle malattie complesse, ad esempio di quelle a maggiore impatto sociale quali Alzheimer, diabete tipo 2 ed infarto del miocardio. È stato infatti possibile da quel momento coniugare la conoscenza delle sequenze di DNA con una tecnica di sequenziamento che si è fatta di giorno in giorno sempre più raffinata e meno costosa al punto da promettere per la metà del corrente anno, al massimo per la primavera del 2013, il sequenziamento di un intero genoma umano in 24 ore al costo di 1000 € (*Illumina Inc* e *Life Technologies Corp*)! La dettagliata conoscenza dell'anatomia del genoma umano con la ricerca di tutti gli elementi funzionali in esso presenti (si ricordi che solo poco più di un 1,5% del genoma umano è codificante proteine) si è rivelata una impresa assai ardua quant'anche necessaria. In effetti stabilire la quantità di tutte le basi nucleotidiche che costituiscono il genoma umano capaci di essere portatrici di "informazione genetica" non è affatto semplice, sia con interventi di tipo sperimentale sia con calcoli computazionali. Ora grazie a sforzi internazionali quali il progetto ENCODE (Encyclopedia of DNA Elements, <http://genome.ucsc.edu/ENCODE/>), che mira ad identificare tutti gli elementi di DNA che contengono "informazione", unitamente alle nuove capacità di potenza di sequenziamento, in altre parole la convergenza delle due capacità tecnologiche, promette di avanzare le conoscenze delle complesse relazioni tra costituzione genetica dell'individuo (genotipo) ed i tratti associati di interesse e risvolto medico-clinico (fenotipo).

Gli studi di associazione genomica

È così possibile indirizzare gli studi verso domande sempre più concettualmente raffinate riguardo alla genesi ed ereditabilità (e dunque sul chiarire la componente di suscettibilità ambientale) delle malattie. Nel corso dei progressi compiuti nella direzione della completa e precisa decifrazione del genoma umano (si badi bene: NON del codice genetico ! errore grossolano da matita blu ! compiuto persino dai direttori di testate come *Nature*! il codice genetico è unico ed universale, dai virus, ai batteri, al lievito ai vegetali ed animali, con pochissime eccezioni per

i retrovirus) nell'ottobre del 2005 è poi stata pubblicata la mappa degli aplotipi umani (International HapMap Consortium, 2005) il che ha poi permesso ben più di 1000 studi di associazione genomica, GWAS (Genome-Wide Association Studies; Bustamante et al., 2011). Questo tipo di studi si basa sul sequenziamento di particolari coorti di soggetti alla ricerca di varianti genetiche associabili a particolari condizioni fenotipiche. Ad esempio, studi GWAS hanno identificato regioni del genoma umano che controllano specifici tratti fenotipici grazie alla identificazione di polimorfismi di singole basi nucleotidiche (Single Nucleotide Polymorphism, SNP, la più comune forma di variante genetica; un tipo di variante genetica dovuta al cambiamento di un singolo nucleotide in una sequenza di DNA). Va subito precisato che sebbene gli studi di tipo GWAS possono stabilire un legame (associare) tra un locus genetico, marcato da un SNP ad esso adiacente, ed il fenotipo (il tratto di interesse medico-clinico) ad esso associato, questo fatto non significa che gli studi GWAS automaticamente identificano la posizione del nucleotide implicato nell'associazione stessa poichè ad oggi abbiamo solo la conoscenza di un limitato repertorio (numero) degli SNP umani. I GWAS vengono impiegati poichè il loro costo (vedi più avanti) è ben inferiore rispetto all'impiego del completo sequenziamento del genoma di tanti pazienti in terapia, metodologia che permetterebbe di disporre di una quantità di informazioni genetiche "personalizzate" ben più elevata. È evidente infatti che il sequenziare il genoma di molti individui permetterebbe di superare il problema di identificare quale nucleotide (o quali nucleotidi) è implicato nel determinare un certo tratto fenotipico di interesse medico-clinico. Dall'intera sequenza genomica di molti individui è infatti possibile determinare con precisione quali varianti di sequenze di DNA, e con quale frequenza, si presentano negli individui portatori di un determinato tratto fenotipico, ad esempio una specifica malattia o altro tratto di interesse medico-clinico, e quindi confrontare questi dati con quelli di individui che non presentano quella malattia. In tal modo sarebbe possibile conoscere ed identificare tutte le varianti genetiche che contribuiscono nella determinazione di quella specifica malattia, o tratto medico-clinico di interesse, permettendo dunque una visione biologica completa del fenomeno. Ad oggi, quasi incredibile a credersi, conosciamo in dettaglio solo il genoma completo di 4 (quattro!) individui: James Watson, Craig Venter, un anonimo africano ed un anonimo asiatico! È di rilievo ricordare che il genoma umano non è semplicemente una collezione di varianti genetiche: la derivazione ancestrale del DNA di ciascun individuo è di fondamentale importanza poichè specifiche associazioni di variante genetiche si sono sedimentate e consolidate nel corso della storia evolutiva dell'uomo. A tale proposito va ricordato che la American Association of Medical Sciences ha recentemente pubblicato sui prestigiosi Proceedings of the National Academy of Sciences un documento ove invita i responsabili politici e le associazioni mediche a riformare i curricula di studio della medicina al fine in rendere obbligatorio lo studio di una disciplina di base e fondamentale per la formazione di un buon medico nell'era della medicina "personalizzata", "genomica", "rigenerativa": la Biologia evolutiva (Nesse et al., 2010)! E dunque la conoscenza di queste associazioni promette di scoprire se e quanto una proteina "aberrante" sia prodotta e

sia la causa del fenotipo osservato nello specifico aplotipo di quel particolare individuo (paziente; per aplotipo si intende proprio quell'insieme specifico di varianti genetiche che caratterizzano uno specifico individuo o insieme, sufficientemente omogeneo sotto il profilo genetico, di individui).

Successi e difficoltà

Prima di esaminare le possibilità, i successi, le difficoltà tecniche e le future promesse dello studio del genoma umano e delle sue varianti genetiche è bene precisare che si è realizzato, per ragioni storiche, un fatto che deve essere ben considerato quando si teorizza e si derivino conclusioni su questi dati: il 96% dei dati ad oggi disponibili sulla conoscenza di sequenze del genoma umano e delle sue varianti genetiche con la loro correlazione, associazione e legame con le malattie è solo e soltanto stato desunto da individui europei (e loro discendenti) e dunque deve essere da un lato valutato con estrema prudenza e dall'altro costituire la base per ampliare al massimo il repertorio di genomi di popolazioni etniche (anche e soprattutto le meno rappresentate numericamente) il più diverso possibile così da poter costituire in pochi anni un database veramente significativo per una popolazione umana ormai avviata ai 9 miliardi di individui. La ragione alla base di questa necessità non è (solo) di natura sociale come si potrebbe pensare in un primo momento, e dunque legata ad un principio di eguaglianza sotteso alla applicazione medica della conoscenza biologica (non è ammissibile sviluppare benefici solo per alcuni privilegiati caucasici), ma squisitamente scientifica per quattro ordini di ragioni così logicamente concatenate:

- poiché la varianza genetica tra individui deriva dalle differenze nelle sequenze di DNA, le varianti genetiche;
- poiché queste sono in gran parte "varianti comuni" (e cioè a dire diffuse in più del 5% degli individui di quasi tutte le popolazioni a livello geografico e di gruppi etnici);
- poiché gli studi GWAS hanno dimostrato una chiara relazione/associazione tra varianti comuni e molte delle comuni malattie;
- poiché queste associazioni spiegano solo dal 5% al 50% della ereditarietà genetica di queste patologie, allora se ne deve concludere che gran parte dei fattori genetici responsabili delle comuni malattie sono delle varianti genetiche "rare" (varianti genetiche che si riscontrano in meno del 5% della popolazione a livello mondiale) e che queste varianti rare costituiscano la gran parte, la quasi totalità, della varianza genetica capace sia di determinare il rischio di sviluppare una malattia da parte di un individuo (Manolio et al., 2009) sia di predirne la sua risposta ad una certa molecola ad azione farmacologica (McClellan e King, 2010).

Le varianti genetiche rare tendono a essere popolazione-specifiche (Gravel et al., 2011) e dunque, poiché giocano un ruolo così importante nel determinare le malattie, la scarsità o assenza della loro rappresentatività negli studi GWAS ne determina la scarsa validità per capire quale di queste varianti è veramente importante. Ad esempio negli individui a derivazione ancestrale dai nativi del sud america una particolare variante di una proteina per il trasporto del colesterolo a livel-

lo intracellulare è fortemente associata con bassi livelli di colesterolo ad alta densità, obesità e diabete di tipo 2. Nelle popolazioni europee, asiatiche ed africane questa variante genetica è assente (Acuna-Alonzo et al., 2010). Al contrario, molti studi hanno dimostrato l'esistenza nelle popolazioni europee di ben 19 comuni SNP fortemente associati al diabete di tipo 2. Inoltre, Waters e collaboratori (2010) in uno studio che ha arruolato ben 6000 individui tra americani di origine europea, africana, giapponese, "latinos" e nativi delle Hawaii, 13 di questi SNP si sono rivelati sempre fortemente associati al diabete di tipo 2, 5 paiono avere diversi effetti ed 1 è rimasto del tutto non chiarito nella sua rilevanza genetica. Diverse sono le ragioni per le quali un risultato valido per una popolazione non trova riscontro in un'altra ma due sono le considerazioni necessarie per spiegare l'impossibilità teorica di compiere generalizzazioni:

- gli studi GWAS identificano dei marcatori genetici associati ad un particolare tratto fenotipico, non identificano la mutazione che causa la malattia;
- le frequenze alleliche di geni associati alla malattia possono variare in maniera sostanziale tra le diverse popolazioni.

Ne consegue che se un certo marcatore è legato ad una miscela di varianti genetiche comuni e rare, alcune o tutte delle rare possono variare significativamente tra le diverse popolazioni arrivando addirittura ad essere completamente assenti in alcune (Gravel et al., 2011). Un'altra considerazione ancora chiede una estrema prudenza nel generalizzare i risultati ottenuti. Si consideri il fatto che le popolazioni africane sono in media più diverse geneticamente rispetto agli europei, agli asiatici o ai nativi americani: ne consegue che ci si dovrà attendere una minore associazione tra marcatori e mutazioni causanti la malattia negli studi che arruolano individui africani e americani-africani. In altre parole, il grado di associazione tra varianti genetiche comuni e mutazioni causanti la malattia varia molto in dipendenza della popolazione in esame. Così Yang et al. (2011) hanno dimostrato che tra i discendenti dei nativi americani vi è un alto rischio di ritorno della malattia (leucemia limfoblastica acuta) dopo la remissione per gli adolescenti con più del 10% di genoma "nativo", necessitando questi ultimi di trattamenti chemioterapici addizionali per rispondere al trattamento.

Sforzi internazionali

Appare del tutto dunque l'evidenza scientifica che imporrebbe un grande progetto di ricerca, simile a quello che ha portato al sequenziamento del genoma umano e poi alla costruzione della mappa degli aplotipi, per giungere ad una mappa delle varianti genetiche in relazione alla ancestralità genetica dei soggetti a cui è rivolta una determinata terapia, in particolare quella farmacologica. Sono altrettanto evidenti le ragioni della riluttanza dei genetisti e delle grandi imprese farmaceutiche (big pharma) ad intraprendere una via caratterizzata da due principali difficoltà, l'una di carattere (paradossalmente) scientifico, l'altra di economia dell'investimento finanziario. Quella scientifica è di propria natura di metodologia della indagine genetica: più popolazioni (etnie, gruppi) vengono incluse in un dato studio e maggiore sarà la probabilità di trovare una associazione positiva tra malattia e una variante di sequenza a causa delle differenze caso/controllo esistenti tra

le diverse etnie (gruppi) considerate piuttosto che a causa delle differenze sullo stato della malattia nei particolari gruppi considerati. Si potrebbe ovviare con un approfondito studio epidemiologico che permettesse di accedere (chiarire) a dati sulla incidenza (rilevanza) di un simile fattore. E qui entra in gioco il fattore economico: non è semplice bilanciare l'investimento economico per avere da un lato una ampia rappresentatività etnica nel campione in studio e dall'altro poter impiegare adeguate procedure di metodologia genetica e statistica capaci di fornire dati per decine o centinaia di migliaia di individui (pazienti/controlli). Sotto il profilo bioetico è evidente che queste difficoltà non posso giustificare il restringere gli evidenti benefici medici derivanti dallo studio delle varianti genetiche associate alle malattie alle sole popolazioni caucasiche (privilegiate) sino ad oggi considerate. Gli studi basati sulle popolazioni debbono essere condotti a livello globale di scala per beneficiare l'intera umanità e questo traguardo può essere raggiunto solo per piccoli passi invogliando (e.g. finanziando) i ricercatori che considerano le minoranze etniche nei loro protocolli e favorendo i piani di sviluppo dei paesi emergenti ed in via di sviluppo affinché anche i ricercatori di quei paesi si dedichino a studi di genomica delle popolazioni.

Il problema della ereditarietà persa

A livello di politica della ricerca è questo quello che viene chiamato il “problema della ereditarietà persa” (“missing heritability problem” Manolio et al., 2009), un fatto di rilievo poichè porta molti ricercatori a non lavorare nella direzione della genomica di popolazioni per le due grandi difficoltà ora ricordate (genetica/metodologica - economico/temporale). Alcuni esempi positivi si possono ricordare con il caso notissimo (anche se molto discusso e discutibile sotto il profilo giurisprudenziale) dell'Islanda, ma anche per Finlandia e Costa Rica. Una riflessione che andrebbe fatta è che se da un lato il costo di nuovi studi di GWAS è ancora altissimo, il replicare uno studio su un diverso gruppo etnico ha un costo che è circa un decimo! e se poi ci si limita alla ricerca di alcuni marcatori (anzichè eseguire una sequenza genomica completa che costerebbe oggi almeno sui 5000€) questa ha un costo attorno ai 250 € per individuo. A livello globale si sta cercando di giungere ad una soddisfacente rappresentatività globale della variazione genetica con il progetto “1000 genomi” (The 1000 Genomes Project Consortium, 2010) capace di fornire un catalogo aperto (che costituisca una sorta di referenza universale per tutti gli studi di GWAS da intraprendere) delle varianti genetiche presenti in più del 1% delle popolazioni a livello mondiale (inclusendo gruppi etnici misti, del nord e sud America, Africa, Europa estremo oriente e Asia meridionale). Per rendere la medicina genomica veramente globale i genetisti debbono sviluppare nuove metodologie statistiche capaci di dissezionare il contributo dei tanti fattori che concorrono nella interazione genoma-ambiente a determinare il fenotipo di un individuo, dissezionare il contributo della genetica (al di fuori di una semplice e riduzionista visione di tipo deterministico), delle condizioni socio-culturali, stili di vita e più in generale dei fattori ambientali che contribuiscono alle malattie (in specie quelle croniche ed infettive). Nello sforzo econo-

mico intrapreso una lode particolare va rivolta, al di là delle doverose iniziative a livello governativo da parte di tanti paesi, alle associazioni filantropiche quali i Wellcome Trust, la “Bill & Melinda Gates Foundation” e la fondazione messicana “Carlos Slim Helú” per citarne alcune.

I costi ed il futuro

Il costo del sequenziamento di un intero genoma umano si sta avvicinando sempre più ai fatidici 1000 €, livello superato il quale verrà infranta una barriera simbolica ancor prima che economica. A quel punto le attuali proiezioni ci dicono che l'era della medicina personalizzata sarà ufficialmente inaugurata. Vale dunque la pena di riflettere sulla base dei genomi conosciuti in dettaglio per almeno quattro persone ed un rappresentante della linea evolutiva degli ominidi (non un nostro antenato nel senso pieno della parola, si legga più avanti), un individuo Neanderthal. Infatti oggi giorno due aspetti sono di rilievo per la piena comprensione del significato della conoscenza della sequenza del genoma umano e delle sue varianti genetiche, il sequenziamento del genoma di James Watson (uno dei quattro più sopra ricordati) e dell'uomo di Neanderthal e le informazioni che siamo in grado di estrarre da tali dati.

Il genoma del dottor James Watson e del signor Neandertal

Il genoma del dottor James Watson

Wheeler e collaboratori (Wheeler et al., 2008) hanno di recente terminato di sequenziare l'intero genoma di uno dei padri della doppia elica di DNA, James Watson. Si pone dunque la domanda: che cosa possiamo aspettarci di imparare da tale conoscenza!? Ebbene il risultato finale è che appare di estrema difficoltà, con gli attuali strumenti bioinformatici e con le attuali conoscenze di biologia e medicina, poter estrarre informazioni di rilievo sotto il profilo biologico predittivo e medico preventivo per il signor James Watson! Si consideri ad esempio la difficoltà di interpretare il significato dei SNP del signor Watson: questi sono all'incirca 3.300.000 quando confrontati con la sequenza di riferimento pubblicata dal progetto genoma umano. Di questi 3.300.000 SNP, ben 82% sono già stati descritti e sono noti per altri individui. Si ritiene che la maggioranza di queste varianti sia di significato neutro (non abbia cioè alcun valore evolutivo o funzionale). Circa 11.000 SNP del sig. Watson ci si attende che siano legati a cambiamenti di sequenze aminoacidiche (di cui 85% precedentemente noti e 15% nuovi) e dunque, forse, legati a cambiamenti di funzione della proteina. Un numero imprecisato di SNP sono capaci di regolare il livello quantitativo di alcune proteine. Questo è quanto emerge dal genoma sequenziato del dottor James Watson. Ora poiché vi sono solo circa 20.000 geni codificanti proteine nel genoma umano è chiaro che se due individui vengono confrontati dettagliatamente ne risulterà che una frazione significativa delle proteine apparirà differire tra i due individui e solo in rarissimi casi si potranno associare tali differenze ad un qualunque effetto biologico della differenza stessa. Esaminando con questa lente di indagine e

con questi strumenti di analisi (quelli attuali!) ne emerge che il livello generale di variazione offerto dal genoma di James Watson non è diverso da quello di altri individui di ancestralità europea ! Questa conclusione è suffragata dall'altro caso di completo sequenziamento di genoma, quello del signor Craig Venter (Levy et al., 2007) che pure corrisponde nella sua generalità a quello di riferimento. Se il signor Watson si presentasse ad un genetista medico per chiedere informazioni predittive o preventive per la sua salute si sentirebbe dire che non vi è alcuna possibilità di predire un alcunchè se non la rara possibilità che dal matrimonio con una signora portatrice di alcune delle medesime varianti vi è una probabilità alta per la nascita di un bimbo portatore della variante in omozigosi ! Dobbiamo così riconoscere l'attuale sottilissimo valore clinico della conoscenza delle sue varianti genetiche, fattore questo che certamente allontana la possibilità che imprese mercantili sovvenzionino questi studi (leggi sopra). La vera sfida oggi è quella di riuscire a legare, associare, queste varianti genotipiche a specifici fenotipi ed in particolare a quelli che sappiamo essere predisponenti a determinate malattie ed a rispondere al trattamento terapeutico farmacologico. Ad oggi la gran parte di ciò che conosciamo a questo proposito lo si deve agli studi di co-ereditarietà dei fenotipi in gruppi famigliari eseguiti sulla base di marcatori genetici (i.e., varianti neutrale di DNA capaci di marcare la comune ancestralità di porzioni di genoma). In altre parole lo dobbiamo a studi "indiretti" (o di "associazione") della relazione genotipo-fenotipo e non "diretti", dovuti cioè al sequenziamento diretto del gene (geni) responsabili del tratto fenotipico in esame. Pare proprio che la attuale pratica di medicina personalizzata basata sull'intero sequenziamento del genoma individuale debba ancora attendere lo sviluppo di potenti metodi di estrazione di informazione genetica dalla gran massa di dati che siamo in grado di accumulare in poche ore grazie alle tecnologie più raffinate di sequenziamento. Per fare un esempio chiaro, sulla base dell'intera sequenza del genoma del signor Watson, oggi non saremmo in grado di predire la sua altezza neppure in modo approssimativo: il più informativo degli SNP che siano noti per influenzare la altezza è in grado di spiegare una variazione di circa 2-4 mm per un tratto fenotipico caratterizzato da una variazione standard di circa 7 cm (Weedon et al., 2007).

Il genoma del Neandertal

Di estremo interesse per quello che i bioinformatici sapranno ricavare è la recente acquisizione della sequenza genomica dell'uomo di Neandertal, il gruppo ominide parallelo all'evoluzione di sapiens (Green et al., 2010). Va subito precisato infatti che il genoma umano (*H. sapiens sapiens*) porta segmenti di genoma di Neandertal come è stato di recente chiarito. In altre parole Neandertal non è un progenitore di sapiens, è un contemporaneo di sapiens, si è incrociato con sapiens (tanto è che portiamo nel nostro genoma dei segmenti di DNA di Neandertal). Il confronto tra le due sequenze ha per ora messo in luce la esistenza di ben 111 Neandertal-specifiche duplicazioni la cui lunghezza media è di 22,321 bp per una lunghezza totale di 1862 kb) ed una maggior presenza di variazioni dovute a numero di copie (CNV, copy number variation, una delle varianti genetiche tra individui). Inoltre, Somel e collaboratori (Somel et al., 2011) hanno identificato

degli specifici micro RNA (fattori capaci di regolare la espressione genica a livello post-trascrizionale) quali miR-92a, miR-454, e miR-320b come probabili regolatori della espressione di alcuni geni chiave per lo sviluppo neurale suggerendo così che una possibile forza evolutrice verso gli ominidi sia stata la rapida evoluzione del sistema nervoso centrale (in particolare della corteccia prefrontale) nella linea di ominidi che ha portato al Neandertal ed al sapiens rispetto ai primati (scimpanzè).

Studi di GWAS

Sulla base degli elementi ora esposti verranno esaminati, nel corso della presentazione, alcuni dei più promettenti studi di GWAS per tumore della mammella, malattia celiaca, tumore gastrico, schizofrenia e leucemia mieloide acuta.

Testo di riferimento

Nell'ambito dei numerosi testi di riferimento che si possono utilmente impiegare per dedicarsi alla genomica medica suggerisco il testo di *Claudia Mikail* della School of Public Health and Health Sciences della Università del Massachussetts (USA) "*public health genomics: The essentials*" pubblicato da Jossey-Bass di San Francisco (USA) nel 2008.

Bibliografia

1. Acuna-Alonzo V, et al. A functional ABCA1 gene variant is associated with low HDL-cholesterol levels and shows evidence of positive selection in Native Americans. *Hum Mol Genet.* 2010; 19: 2877-85.
2. Bustamante CD, Burchard EG, De La Vega FM. Genomics for the world. *Nature.* 2011; 475: 163-165.
3. Gravel S, et al. Demographic history and rare allele sharing among human populations PNAS USA. doi: 10.1073/pnas.1019276108, 2011.
4. Green RE, et al. A draft sequence of the Neandertal genome. *Science.* 2010; 328: 710-22.
5. International HapMap Consortium: A haplotype map of the human genome. *Nature.* 2005; 437: 1299-320.
6. Lander ES, et al. Initial sequencing and analysis of the human genome *Nature.* 2001; 409: 860-921.
7. Levy S, et al. The diploid genome sequence of an individual human. *PLoS Biol.* 2007; 5 (10): e254. doi:10.1371/journal.pbio. 0050254.
8. Manolio TA, et al. Finding the missing heritability of the complex diseases. *Nature.* 2009; 461: 747-53.
9. McClellan J, King MC. Genetic heterogeneity in human disease. *Cel.* 2010; 141: 210-7.
10. Somel M, Liu X, Tang L, Yan Z, Hu H, Guo S, Jiang X, Zhang X, Xu G, Xie G, Li N, Hu Y, Chen W, Pääbo S, Khaitovich P. MicroRNA-Driven

- Developmental Remodeling in the Brain Distinguishes Humans from Other Primates. *PLoS Biol.* 2011; 9 (12): e1001214.
11. The 1000 Genomes Project Consortium. A map of human genome variation from population-scale sequencing. *Nature.* 2010; 467: 1061-73.
 12. Yang JJ, et al. Ancestry and pharmacogenomics of relapse in acute lymphoblastic leukemia. *Nature Genet.* 2011; 43: 237-41.
 13. Waters KM, et al. Consistent Association of Type 2 Diabetes Risk Variants Found in Europeans in Diverse Racial and Ethnic Groups. *PLoS Genet.* 6, e1001078, 2010.
 14. Weedon MN, et al. A common variant of HMGA2 is associated with adult and childhood height in the general population. *Nature Genet.* 2007; 39: 1245-50.
 15. Wheeler DA, et al. The complete genome of an individual by massively parallel DNA sequencing. *Nature.* 2008; 452: 872-6.

Genetica delle malattie comuni

Antonio Pizzuti

Ordinario di Genetica Medica, Dipartimento di Medicina Sperimentale,
Università di Roma "La Sapienza"

La maggior parte delle malattie comuni ha una qualche determinante genetica. Noi definiamo queste malattie come patologie genetiche *complesse* e si distinguono dalle più rare patologie *monogeniche* o *mendeliane* perchè la ricorrenza in famiglie non segue, come queste, le regole della segregazione allelica come vengono dedotte dalle leggi di Mendel. La maggior parte delle malformazioni congenite sono ad eredità complessa, e lo stesso si può dire per le patologie metaboliche comuni (diabete), le malattie psichiatriche e quelle vascolari.

Alcuni frequenti tratti patologici si comportano come un continuo di anomalie, diverse solo dal lato quantitativo, situazioni nelle quali i *fenotipi* patologici non possono essere descritti in categorie distinte ma descrivono una gradazione di situazioni. La distribuzione di questi *tratti continui* (ad esempio il grado di resistenza all'insulina nel diabetico di tipo II, oppure la severità di discostamento della pressione sanguigna dagli standard ritenuti normali) è tipicamente normale, e la maggior parte degli individui hanno un fenotipo intermedio.

L'eredità multifattoriale fu studiata per primo da Galton alla fine dell'800. Analizzando appunto tratti continui come l'altezza o l'intelligenza Galton si accorse di un fenomeno che chiamò *regressione alla media*. Genitori alti tendevano ad avere figli alti ma più bassi di loro e genitori bassi, figli più alti. Ovvero l'altezza non si trasmetteva come i caratteri di Mendel, ma i figli tendevano a regredire verso la media di quel carattere nella popolazione generale. Questo scostamento del comportamento di alcuni tratti fenotipici da quello di tratti semplici descritti dalla genetica Mendeliana venne poi spiegata dal fatto che i tratti complessi sono determinati appunto dalla contemporanea presenza di più fattori genetici con valore additivo, la cui trasmissione "in blocco" nelle generazioni successive ha una probabilità minore di un semplice tratto singolo; e che l'influenza patogenetica di questi fattori genetici è spesso condizionata, anche da fattori ambientali. Quindi i fattori genetici agiscono di concerto ciascuno aggiungendo o detraendo una piccola quantità di attività dal fenotipo, e l'ambiente interagisce con il genotipo per produrre il prodotto finale. Non tutti i tratti complessi a cui si rifanno le comuni malattie hanno una distribuzione gaussiana. Molte condizioni patologiche comuni, sempre ad eredità complessa non vi è alcuna gradazione ma il tratto "patologico" appare *dicotomico* (un individuo ha la spina bifida o non la

ha). Il modello complesso con effetti additivi e sottrattivi (a volte epistatici) tra differenti fattori genetici e non, oltre che per tratti continui come quelli studiati da Galton, è valido anche quando siamo di fronte a tratti dicotomici (la schisi labiale, che è un tratto per di più multifattoriale o c'è o non c'è, non esistono intermedi) con l'ipotesi di un *effetto soglia*. Con l'aumentare dei fattori genetici responsabili di aumentata suscettibilità alla patologia, aumenta il rischio e la malattia si palesa quando una determinata soglia viene superata. In questo caso nonostante il fenotipo si esprima in modo non continuo (sì o no) la suscettibilità è ancora determinata da una distribuzione normale.

Pertanto le malattie non mendeliane sono anche definite *multifattoriali*. In queste patologie, sebbene l'aggregazione familiare sia comune (cioè, vi possono essere casi multipli nella stessa famiglia), il modello di eredità e i rischi di ricorrenza familiari non possono essere predetti in base alle leggi della segregazione degli alleli di Mendel. In queste condizioni l'influenza di fattori genetici sul rischio di sviluppare un fenotipo clinico è meno incidente di quanto avviene nelle malattie mendeliane ed è quasi sempre fortemente influenzato da fattori ambientali; e qui per fattori ambientali vanno considerati sia l'ambiente in senso comune, ovvero le esperienze personali della vita di un individuo (come le abitudini alimentari, la scolarità, l'occupazione, ecc.), sia l'ambiente genetico nel quale coesistono i vari fattori anch'essi genetici di suscettibilità o di protezione rispetto all'instaurarsi di un processo patologico. Dal punto di vista molecolare sappiamo poco dei fattori genetici di suscettibilità per la maggior parte delle patologie comuni, perchè esse, proprio per la loro complessità, sono assai difficili da analizzare con le metodiche analitiche che hanno permesso lo studio ed il riconoscimento del substrato genetico di ormai centinaia di patologie monogeniche. Sappiamo che due individui che non siano gemelli identici hanno genomi che differiscono per circa una base nucleotidica su mille, ed che in queste differenze vi sono anche i fattori genetici responsabili della suscettibilità alle malattie comuni. La difficoltà nell'individuare queste varianti "patogenetiche", tra milioni di varianti nucleotidiche presenti nel genoma di ogni individuo, sono dovute a vari fattori tra i quali l'estrema *eterogeneità genetica* delle determinanti stesse ed il fatto che in molte di queste situazioni genetica-cliniche le determinanti sono tante e con piccolo valore individuale.

Negli anni passati la migliore fonte per studiare il contributo genetico al determinismo dei tratti complessi, compresa la suscettibilità a condizioni patologiche, e le interazioni geni-ambiente, sono stati gli studi sui gemelli. I gemelli *monozigoti* (identici) hanno lo stesso genoma, ma non necessariamente crescono nello stesso ambiente. Il tasso di concordanza per un tratto fenotipico in gemelli monozigoti può essere confrontato con quello in gemelli *dizigoti*, che hanno in comune solo la metà del proprio genoma. Se il tratto fenotipico è al 100% di origine genetica allora la concordanza tra gemelli monozigoti sarà del 100% e più basso nei gemelli dizigoti; ed anche per tratti solo in parte geneticamente determinati la concordanza sarà più elevata nei gemelli monozigoti che in quelli di zigoti, pur non arrivando al 100%.

In molte malattie multifattoriali i due sessi hanno probabilità differenti di essere colpiti. Per esempio, l'artrite reumatoide colpisce molto più frequentemente le fem-

mine dei maschi, il che significa che vi sono livelli di soglia differenti tra i due sessi, e la soglia della femmina per questa patologia è più vicina alla mediana di quella del maschio. Detto in altri termini se la malattia colpisce preferenzialmente le femmine, il carico genetico necessario perchè un maschio sia anche lui affetto deve essere superiore a quello necessario per determinare malattia nella femmina. Tuttavia, siccome il carico genetico del maschio affetto è più importante di quello della femmina, il maschio ha un rischio maggiore che la patologia multifattoriale “artrite reumatoide” ricorra nella sua progenie. È chiaro da quanto detto che non è possibile, diversamente da quello che accade con le malattie monogeniche, derivare con calcolo matematico una figura di rischio in figli di genitori affetti per le malattie complesse. Le figure di rischio che viene di solito somministrato in consulenza genetica ad individui che possano trasmettere o sviluppare patologie geneticamente complesse è totalmente empirico, dove le stime sono determinate da analisi di ricorrenza in popolazioni di pazienti affetti. Ad esempio, la nozione comunemente riconosciuta che il rischio di ricorrenza in fratelli della sindrome di Down è 1% indipendentemente dall’età materna, deriva da studi osservazionali che hanno studiato numerosissime famiglie con più figli affetti.

Il rischio di ricorrenza di un evento patologico legato ad una o più varianti genetiche in una stessa famiglia dipende quindi da quanto è grande il peso del fattore genetico nel determinismo della malattia (fino al 50% nelle forme genetiche mendeliane), che per le patologie complesse è anche proporzionale alla gravità del fenotipo clinico. Quando in una famiglia ricorre realmente un fenotipo patologico complesso questo evento testimonia in quella famiglia l’esistenza di fattori genetici “forti” che determina un aumento ulteriore del rischio di ricorrenza familiare. Ad esempio, se l’incidenza dei difetti di chiusura del tubo neurale nella popolazione generale è di 1 su 500 nati, dopo il primo figlio affetto il rischio diventa decine di volte superiore (3%) e si alza ancora se i figli affetti sono 2 (9%). Questo fenomeno non si osserva nei tratti Mendeliani, dove il rischio di ricorrenza è sempre identico a se stesso, indipendentemente dalla frequenza familiare della malattia e dalle sue caratteristiche cliniche.

Riassumendo, le caratteristiche più importanti della eredità multifattoriale (malattie genetiche complesse) sono:

1. La maggior parte delle volte il malato ha genitori sani.
2. Il rischio di ricorrenza aumenta con l’aumentare degli affetti in una famiglia.
3. Il rischio di ricorrenza aumenta con la severità della malattia.
4. La consanguineità aumenta leggermente il rischio di ricorrenza.
5. Il rischio per consanguinei scende col passare delle generazioni e col ridursi del grado di consanguineità.
6. Se i due sessi hanno una differente probabilità di essere colpiti, allora il rischio di ricorrenza è maggiore per il sesso meno colpito.

Bibliografia

1. Speiche et. al. ed. Vogel and Motulsky’s Human Genetics 4th.
2. Strachan and Read. Human Molecular Genetics, 4th edition.

Il clinico di fronte alla medicina genomica

Carlo Bernasconi

Già Professore Ordinario di Ematologia, Università di Pavia,
Consulente Ematologo presso l'IRCCS Fondazione "Salvatore Maugeri", Pavia

Il completamento del Progetto Genoma Umano (Collins et al., 2003) ha rappresentato il più grande stimolo per l'esecuzione di un numero straordinario di ricerche, che da allora sono state e sono condotte allo scopo di esaminare quale contributo le conoscenze sul genoma umano possono portare alla comprensione dei meccanismi di salute e di malattia. Innanzitutto, scoperte clinicamente utili ottenute dal Progetto hanno prodotto importanti acquisizioni di "medicina genetica", cioè l'uso di conoscenze riguardanti singoli geni per migliorare la diagnosi e il trattamento di malattie monogeniche rare (definite anche "malattie mendeliane" poiché trasmesse secondo le leggi di Mendel). Inoltre, il progresso delle conoscenze sulle interazioni fra l'intero genoma e fattori non genomici, all'origine di condizioni di salute o di malattia, ha aperto la strada verso la "medicina genomica", caratterizzata da nuovi approcci diagnostici e terapeutici per malattie multifattoriali comuni.

Come risultato del rapido progresso nelle conoscenze genomiche, oggi un numero sempre crescente di linee guida cliniche suggerisce di includere tests di genetica molecolare nella pratica medica routinaria. In alcuni casi la rapidità della traslazione ha animato una discussione sul livello di evidenza del beneficio clinico necessario per introdurre nuove, e potenzialmente costose, tecnologie mediche (Hunter et al., 2008).

Sebbene l'effetto delle scoperte genomiche sulla pratica medica quotidiana non possa venir bene quantificato, è probabile che sia minore nell'assistenza di base e in un ambiente non accademico rispetto a quanto avviene, ad esempio, in un centro medico oncologico di riferimento. Certamente però, indipendentemente da dove la medicina viene praticata, la genomica ha inesorabilmente cambiato la nostra comprensione della biologia di quasi tutte le malattie (Ferro et al., 2010). Infatti, come può oggi un clinico capire ed esprimersi in modo adeguato sulla diagnosi e terapia di una leucemia o di un tumore senza una rudimentale conoscenza di genomica medica? Appare quindi molto opportuno fare una rapida rassegna delle più recenti acquisizioni concettuali e tecnologiche in questo settore della ricerca biomedica, e considerare poi alcuni dei loro aspetti applicativi nella pratica clinica.

Concetti base della medicina genomica e dogma centrale della biologia molecolare

Il “genoma” può essere definito come l’intera sequenza del DNA di un individuo. Il termine “genoma” è la sintesi dei due termini “gene” (tradizionalmente definito come una unità di eredità) e cromosoma. Circa il 2% del genoma umano è fatto da geni, le unità funzionali di DNA che contengono le istruzioni per produrre proteine. Il resto del genoma possiede la capacità di regolare dove, quando e in che quantità le proteine vengono prodotte (proprietà conosciuta come “espressione genica”).

L’identificazione dei geni e delle variazioni nella sequenza del DNA che stanno alla base della suscettibilità ereditaria per malattie rare e per malattie comuni dell’uomo, è stato il maggior obiettivo delle ricerche condotte dai genetisti negli ultimi anni, e ha portato a fondamentali progressi nella comprensione dei meccanismi molecolari e cellulari di malattia.

L’insieme di tutte queste nuove conoscenze sta all’origine della “medicina genomica”, che può essere definita come l’uso di informazioni e tecnologie genomiche per determinare il rischio individuale a contrarre determinate malattie, per precisarne a livello molecolare la diagnosi e la prognosi, e infine per decidere la scelta e la priorità delle opzioni terapeutiche.

A quest’ultimo riguardo il medico si avvale di nozioni di “farmacogenomica”, intesa come lo studio del modo in cui variazioni del genoma possono influenzare il metabolismo di un farmaco e i suoi effetti, sia nel senso della tollerabilità che dell’efficacia. Ne deriva il concetto di “terapia personalizzata”, che comporta l’impiego di tests di laboratorio per stratificare gruppi di pazienti secondo la loro prevista responsività ad un determinato trattamento. L’uso di medicinali in accordo a tale stratificazione aiuta a migliorare l’efficacia della terapia, utilizzando tests genetici per selezionare i soggetti che risponderanno meglio ad un determinato trattamento, oppure escludere i soggetti predisposti a sviluppare una reazione avversa a un farmaco.

Quando il DNA è stato identificato come la base dell’eredità ed è stato formulato il dogma centrale della biologia molecolare (DNA→RNA→proteina), il gene è stato definito come un segmento di DNA codificante una proteina. Tuttavia, con la scoperta di nuove classi di RNA, si è reso necessario un riesame della definizione tradizionale di un gene. È emerso così un quadro generale di regolazione genica che rappresenta strati interdipendenti e reti di controllo consistenti in interazioni di DNA con proteine regolatorie e molecole di RNA, simili alle interazioni che si osservano nella circuiteria di un computer. Questo sviluppo ha condotto alla formulazione di sofisticati approcci di “biologia dei sistemi” per capire la complessa rete dei meccanismi di regolazione genica (Davidson e Levine, 2008).

Tale complessità è particolarmente operativa nella razza umana, nella quale un singolo gene può dare origine ad un’ampia schiera di prodotti genici, secondo l’ambiente in cui lo stesso gene viene espresso. In tal modo si espande enormemente il repertorio di proteine prodotte dai circa 20.000 geni che compongono il

genoma umano (numero relativamente piccolo, se si pensa che il bue e la pianta di senape ne hanno di più!). Il primo evento dell'espressione genica, cioè la trascrizione, è regolato da un complesso scenario di processi, che coinvolgono la struttura tridimensionale del DNA, modificazioni epigenetiche (covalenze chimiche) del sostegno del DNA, interazioni proteine - DNA e RNA - DNA. Anche il secondo evento, cioè la traduzione, è molto complesso e strettamente regolato da interazioni fra RNA messaggero (mRNA) e proteine. Infatti, l'elaborazione di molecole di un unico precursore RNA (preRNA) può dare origine a molteplici differenti molecole di mRNA, comprese anche molecole di micro-RNA e *small (or short) interfering RNA* (siRNA). Infine, la modificazione post-traslazionale di proteine può anch'essa contribuire marcatamente alla diversità del prodotto finale del genoma umano, attraverso modificazioni di singole proteine immature (ad esempio, con ripiegamenti, fenditure, modificazioni chimiche), che danno origine ad una serie di prodotti chimici terminali (Ferro et al., 2010).

Una delle più importanti conoscenze della biologia molecolare dell'ultimo decennio è il diverso e ubiquitario ruolo delle piccole molecole di RNA nella regolazione genica (Carthew e Sontheimer, 2009). Chiare evidenze suggeriscono che questa classe di molecole contribuisce alla patogenesi di malattie, in particolare dei tumori e dei disordini immunologici.

Tests basati su patterns di espressione di miRNA in tumori possono aggiungersi alle tecniche tradizionali per determinare il tipo cellulare che sta all'origine di una neoplasia. I miRNA sono molecole endogene non-codificanti di RNA (usualmente costituite da 22 nucleotidi) che inibiscono la traduzione dei corrispondenti RNA (Lee e Dutta, 2009).

Un'azione simile a quella dei miRNA viene svolta dai siRNA, che si legano alle molecole dell'mRNA corrispondente, inibendo la traduzione attraverso una loro degradazione. È interessante al riguardo sottolineare che alcune molecole siRNA sintetiche sono già attualmente testate in trial clinici avanzati (Castanotto e Rossi, 2009).

Variazioni genetiche strutturali

Il Progetto Genoma Umano originariamente aveva l'obiettivo di sequenziare i nucleotidi contenuti in un genoma umano aploide di riferimento. Nel portare avanti il progetto si è lavorato sul DNA ottenuto da un certo numero di donatori selezionati con criteri di rappresentatività statistica. Ma il genoma di qualsiasi individuo (tranne quello dei gemelli monozigoti) è "unico", e non esiste una sequenza genomica umana che si possa in senso stretto definire "normale"; attualmente si preferisce utilizzare il termine "*wild type*" che indica la variante più comune presente in una determinata popolazione. In realtà, quindi, mappare il genoma umano significa fare anche il sequenziamento delle variazioni multiple di ciascun gene. Nella sua forma più semplice, una variazione ha due differenti letture (sulle due sequenze dello stesso cromosoma), indicate come "alleli".

Il termine "mutazione" è generalmente utilizzato per indicare cambiamenti nel DNA associati all'insorgenza di malattie genetiche (ad esempio, l'anemia falciforme).

forme e la fibrosi cistica) oppure insorti in cellule somatiche (ad esempio, all'origine di un tumore). La sequenza genomica completa di un individuo (genotipo), agendo in associazione con le influenze ambientali, dà origine alle caratteristiche della sua individualità (fenotipo). L'insieme degli alleli collegati linearmente lungo le molecole di DNA di una persona è indicato come "aplotipo". Gli uomini sono molto simili a livello di sequenza del DNA: circa il 99,6% delle paia di basi sono identiche da persona a persona (Kidd et al., 2008). Vista la dimensione del genoma (circa 6 bilioni di bp in ciascuna cellula nucleata non-germinale), ne rimane però sempre una quota sostanziale per le variazioni genetiche individuali, poiché la differenza fra due persone qualsiasi è circa 24 milioni di bp.

Gli eventi strutturali che contribuiscono a generare variazioni del genoma possono essere classificati in tre categorie:

1. cambiamenti di una singola copia di basi (o mutazioni puntiformi), che altera la normale sequenza dei nucleotidi del DNA;
2. delezioni o inserzioni nel DNA di tratti più o meno lunghi di nucleotidi;
3. riarrangiamenti strutturali che rimescolano la sequenza del DNA, cambiando così l'ordine dei nucleotidi.

L'impatto clinico delle conseguenze di tali variazioni può essere estremamente differente. Per esempio, il cambiamento di una singola copia di basi può avere gravi conseguenze per la salute (ad esempio, la sostituzione di una base timidinica con una base guaninica nel gene della β -emoglobina umana causa l'anemia falciforme), mentre un ampio intervento traslocativo (in cui l'informazione genetica di un intero braccio di cromosoma può scambiarsi con un altro cromosoma) potrebbe non avere conseguenze per la persona affetta.

Prendiamo in considerazione, per ora, i cambiamenti di una singola copia di basi nella sequenza del DNA. Una mappa degli aplotipi del genoma umano (*HapMap Project*) è stata pubblicata nel 2005, e ha fornito una prima informazione sulla distribuzione in tutto il genoma delle più comuni variazioni delle singole copie di basi (conosciute anche come *single-nucleotide polymorphisms*, *SNPs*) in individui appartenenti a vari gruppi etnici. Questo progetto ha dimostrato che gli SNPs sono molto comuni lungo tutto il genoma umano, sono frequentemente correlati con loro SNPs confinanti, e si osservano in media all'incirca ogni 800 bp. Inoltre, ha fornito uno strumento molto valido (gli studi di associazione indicati come *genomewide*) per facilitare l'identificazione di associazioni genetiche con condizioni complesse (per esempio, in serie molto ampie di pazienti affetti da malattie comuni con predisposizione familiare).

Le inserzioni e le delezioni sono cambiamenti abbastanza frequenti del genoma e possono variare in dimensioni: da una a migliaia di copie di basi. Come gli SNPs, possono essere modificazioni benigne e non aver alcun effetto sul fenotipo, o possono comportare un rischio di malattia. Una classe di queste variazioni, indicate come *copy-number variations* (*CNVs*), è associato a un gruppo sempre più numeroso di malattie, specie neuroipsichiatriche (come l'autismo e la schizofrenia) (Cook e Scherer, 2008). Certamente molto rimane ancora da capire su quanto e come queste variazioni possono intervenire sul rischio di ereditabilità di malattie comuni.

Progressi nelle tecnologie genomiche

Nell'ultimo decennio lo scenario della diagnostica molecolare è cambiato rapidamente. Nel periodo pregenomico la diagnosi genetica era focalizzata principalmente su condizioni originate da mutazioni in singoli geni, che richiedevano l'evidenziamento di una o poche mutazioni. L'attenzione è ora trasferita a tecnologie ad ampio spettro, capaci di scoprire con un solo test migliaia (e anche milioni) di varianti.

I laboratori di diagnostica clinica ancor oggi continuano ad utilizzare tests basati sul legame con sequenze specifiche di brevi probes di DNA (oligonucleotidi) a campioni di DNA di pazienti, al fine di scoprire mutazioni associate con disordini di un singolo gene. Generalmente questi tests utilizzano la *polymerase chain reaction* (PCR) per amplificare le regioni di DNA interessanti per un determinato paziente; il prodotto della PCR viene poi analizzato per la presenza o assenza di mutazioni. Tuttavia, negli ultimi anni le nuove tecnologie, che comprendono l'uso di *gene chips* ed il sequenziamento del DNA, hanno rapidamente accantonato i metodi tradizionali di evidenziamento delle variazioni e mutazioni del genoma umano.

La tecnologia *gene chips* (*microarray*) è alla base di molte delle più importanti acquisizioni scientifiche e cliniche degli ultimi anni. I *gene chips* consistono in una matrice microscopica molto ordinata di sequenze nucleotidiche fissate ad una superficie solida, conosciuta come *microarray*. Per condurre l'analisi degli SNPs lungo tutto il genoma, il DNA è isolato da un campione ottenuto da un paziente, tagliato in piccoli frammenti, coniugati con un colorante fluorescente e poi incubati con un chip di silicene. I frammenti si legano ai nucleotidi fissati secondo una specifica sequenza, e sofisticati sistemi di scansione e di elaborazione del segnale consentono poi di analizzare il pattern e l'intensità del segnale fluorescente per determinare le sequenze presenti nel campione. Fra le varie applicazioni cliniche, la tecnologia *microarray* è stata utilizzata anche per lo studio dei profili di espressione genica in campioni di tumori (Sotiriou e Puztai, 2009) e per studi del rischio genetico a determinati tumori (ad esempio, ca. della mammella, del colon-retto, della prostata, ecc.) (Stadler et al., 2010).

Anche le tecnologie usate per determinare la sequenza completa del DNA di un singolo individuo sono progredite molto rapidamente, abbattendo in modo straordinario i costi di tale determinazione.

Le tecnologie tradizionali di sequenziamento, utilizzate anche nella realizzazione del Progetto Genoma Umano, si riferiscono ad un procedimento di clonaggio di frammenti di DNA, selezione e crescita di cloni, e poi sequenziamento delle copie purificate di un singolo frammento. Ciascuna reazione di sequenziamento tradizionale produce circa 1 Kb (1000 bp) di sequenza di DNA; usualmente, circa 100 reazioni vengono fatte in parallelo. Per contro, le tecnologie di sequenziamento di nuova generazione traggono vantaggio dal poter esaminare in parallelo centinaia di migliaia sino a milioni di frammenti di DNA, usando per ogni reazione quantità estremamente piccole di reagenti chimici. Con le nuove tecnologie una singola procedura di sequenziamento può esaminare più di 100 Gb (100 bilioni

di bp) di sequenze di DNA, in un periodo di tempo di ore o di giorni, dipendendo dalla particolare tecnologia utilizzata. Ormai si può pensare di ottenere il completo sequenziamento delle molecole del DNA di un singolo individuo con un costo di 1.000 dollari e in pochi giorni. Per ottenere la stessa informazione dieci anni fa sarebbero stati necessari un costo di 500.000 dollari e mesi di lavoro di un ampio team di laboratori.

Malattie monogeniche rare e malattie comuni geneticamente complesse

Mutazioni genetiche dannose che interessano il DNA di cellule germinali diventano mutazioni ubiquitarie nell'organismo che si sviluppa, cioè presenti in tutte le cellule. Queste mutazioni danno origine al classico pattern mendeliano delle rare malattie genetiche ereditarie. Nuove mutazioni in cellule somatiche non vengono trasmesse da generazione a generazione, e stanno alla base dello sviluppo dei tumori (che possono quindi venir considerati come malattie genetiche di cellule somatiche). La predisposizione familiare ad alcune malattie comuni (come diabete mellito, coronaropatie, tumori, ecc.), disordini complessi nei quali il fenotipo è determinato sia dai geni che dall'ambiente, può essere causata dal concorso di variazioni genetiche sia comuni che rare.

Nel corso del 20° secolo gli studi di familiarità hanno chiarito l'ereditarietà di condizioni patologiche derivate da un singolo gene, o monogeniche, o anche "malattie mendeliane". Migliaia di condizioni causate da mutazioni in singoli geni sono state identificate e catalogate in un volume (McKusick, 1998), e successivamente in un compendio informatico noto come OMIM (*Online Mendelian Inheritance in Man*). Delle singole malattie sono state precisate le modalità di trasmissione ereditaria: autosomica dominante, autosomica recessiva, legata al cromosoma X. Più recentemente sono stati poi descritti altri meccanismi per l'eredità monogenica, fra i quali particolarmente interessante è l'ereditarietà mitocondriale, che si svolge solo attraverso la linea materna.

Le mutazioni possono causare malattia attraverso varie modalità. Fra queste la modalità più frequente è la perdita di funzione, che altera il fenotipo diminuendo la quantità dell'attività funzionale di una proteina. Per esempio, mutazioni nel gene della glucosio-6-fosfato deidrogenasi (*G6PD*) sul cromosoma X diminuiscono l'attività funzionale di questo enzima, causando un'anemia emolitica acuta se un maschio (che ovviamente ha solo una copia di cromosoma X) con la mutazione è esposto a certi farmaci, come sulfamidici, primachina e nitrofurantoina. In condizioni naturali la deficienza di *G6PD* causa una crisi emolitica acuta nei maschi che ne sono affetti quando questi ingeriscono le fave (favismo), poiché l'enzima è importante nella degradazione di una componente delle fave (Luzzatto et al., 2001).

Alcune mutazioni causano malattia attraverso l'acquisizione di una funzione, nei casi in cui la proteina sintetizzata ha la capacità di esercitare una nuova funzione, che si rivela tossica. L'espansione delle triplette esoniche CAG, che si associa alla malattia di Huntington e all'atassia spinocellulare, appare generare disordini neurologici attraverso la produzione di proteine che funzionano in modo anormale a

causa dell'espansione di tratti poligluttammici (CAG codifica infatti per l'acido glutammico; Goldberg et al., 1996). Queste mutazioni che fanno aumentare una funzione sono frequentemente ereditate in modo dominante, poiché una singola copia del gene mutato può alterare la funzione.

Si potrebbe ritenere che le mutazioni che interessano il 98,5% circa del genoma che non codifica per le proteine non influenzino il fenotipo. In realtà non è così, poiché anche le mutazioni regolatorie agiscono alterando l'espressione di un gene. Infatti, una mutazione regolatoria può portare ad una caduta di espressione di un gene, ad una sua inaspettata espressione in un tessuto in cui è usualmente silente, oppure a una modifica del tempo in cui esso viene espresso. Esempi di mutazioni regolatorie associate a malattie sono quelli nella regione vicina al gene per l'insulina (con aumento del rischio per il diabete mellito di tipo 1; Todd, 1999) e in un sito regolatorio del gene del collagene di tipo I (con aumento del rischio per l'osteoporosi; Grant et al., 1996).

Le mutazioni possono anche diminuire il rischio di una malattia. Un esempio di questa possibilità è una delezione 32-bp in un gene recettore per una chemochina, il *CCR5*.

Le persone che sono omozigote per questa delezione sono quasi completamente resistenti all'infezione con HIV di tipo 1, e quelle che sono eterozigote per la delezione presentano una più lenta progressione dall'infezione all'AIDS. Questi effetti derivano dal fatto che il gene *CCR5* svolge un ruolo importante nei meccanismi di entrata del HIV nelle cellule (Dean et al., 1996).

Le malattie monogeniche sono quasi sempre malattie rare. Anche le più frequenti, come l'emocromatosi ereditaria (incidenza approssimativa 1:300 persone), la fibrosi cistica (incidenza approssimativa 1:3000), la deficienza di alfa₁-antitripsina (incidenza approssimativa 1:1 700) o la neurofibromatosi (incidenza approssimativa 1:3 000) colpiscono non più di 1 su centinaia o migliaia di individui in Europa e nel Nord America.

Tuttavia, i risultati complessivi degli studi delle malattie monogeniche sono importanti, riguardando non solo il singolo paziente, ma comportando anche l'acquisizione di informazioni per la salute della popolazione in generale. Infatti, la comprensione dei meccanismi attraverso i quali fattori genetici possono causare malattie monogeniche ha fornito importanti informazioni su processi fisiopatologici basali, che sottostanno a disordini correlati e che si osservano con maggior frequenza rispetto alla malattia genetica. Per esempio, le conoscenze riguardanti l'ipercolesterolemia familiare, una malattia genetica che nella nostra popolazione colpisce all'incirca solo 1 su 500 individui, sono state fondamentali per capire la patogenesi dell'arteriosclerosi, che colpisce un'elevata percentuale di persone; inoltre, le stesse conoscenze hanno portato allo sviluppo di nuovi farmaci (le statine), che sono fra i medicinali oggi più ampiamente prescritti (Goldstein et al., 2001).

Sino ad alcuni anni fa, la maggior parte dei geni coinvolti nei meccanismi di insorgenza di malattie comuni e complesse sono stati identificati grazie alla loro alta penetranza. Alcuni esempi sono le mutazioni in *BRCA1* e *BCRA2* (aumentato rischio di carcinoma della mammella e dell'ovaio; Nathanson et al., 2001),

HNPCC (aumentato rischio di carcinoma del colon-retto senza poliposi ereditaria; Lynch, 1999), *MODY 1*, *MODY 2* e *MODY3* (aumentato rischio di diabete mellito; Froguel e Velho, 1999) e nel gene per l' α -sinucleina (correlato alla malattia di Parkinson; Mouradian, 2002).

Se un individuo possiede una di queste mutazioni, la probabilità di sviluppare la malattia corrispondente è elevata. Tuttavia, si deve osservare che ciascuna di queste mutazioni altamente penetranti associate a malattie comuni ha una prevalenza nella popolazione generale solo di uno su parecchie centinaia o alcune migliaia di individui.

Dal punto di vista della salute pubblica, geni con mutazioni meno penetranti ma molto più prevalenti hanno un maggior effetto sulla popolazione rispetto a geni che sono altamente penetranti ma poco frequenti. Mutazioni di questo tipo sono state segnalate in geni quali l'*APC* (che aumenta il rischio del carcinoma del colon-retto; Fearnhead et al., 2001) e il fattore V Leiden (che aumenta il rischio di trombosi; Major et al., 2000).

Per quanto riguarda la determinazione della predisposizione genetica all'insorgenza di malattie comuni, negli ultimi anni particolare attenzione è stata rivolta ai *genome-wide association studies* (GWASs), che paragonano popolazioni affette da una determinata malattia con gruppi controllo senza la malattia, al fine di identificare differenze genetiche fra i due gruppi. Se particolari varianti genetiche sono più frequenti nei pazienti affetti da una determinata malattia rispetto ai controlli, si dice che queste varianti sono "associate" con quella malattia. È ben noto che la gran parte delle lettere del genoma (circa il 99,9%) non è differente fra i singoli individui; tuttavia, la piccola frazione che è differente (nota come varianti nella sequenza del DNA o polimorfismi) non solo spiega le differenze di forma e funzioni fra i singoli individui, ma serve anche come segnale molecolare utilizzato nei GWASs per indicare l'associazione fra geni e lo sviluppo di varie malattie comuni a patogenesi complessa.

Le tecnologie che si sono sviluppate verso la metà del primo decennio del 2000, hanno poi consentito di leggere centinaia di migliaia sino a milioni di SNPs da un genoma individuale in un unico esperimento, dando la possibilità di costruire e analizzare il "profilo genomico" di un gran numero di individui. Sono quindi iniziati i primi GWASs, con l'analisi statistica della frequenza di SNPs in pazienti con differenti malattie e in soggetti controllo, per dimostrare un'associazione fra il possesso di particolari SNPs e la suscettibilità a determinate malattie (Wang et al., 2005). Per esempio, il GWAS coordinato dal *Wellcome Trust Case Control Consortium* (2007) riporta la localizzazione di 24 geni di suscettibilità per sette malattie comuni (comprendenti coronaropatie, diabete mellito, artrite reumatoide), e per quanto riguarda alcuni tumori (mammella, ovaio, colon-retto, polmone, ecc.) è oggi possibile formulare un'attendibile previsione di rischio (Weitzel et al., 2011).

Si può quindi concludere che gli studi di associazione *genome-wide* possono consentire l'identificazione di fattori genetici di rischio per malattie comuni e complesse, come gli studi di familiarità sono stati importanti per l'identificazione dei geni coinvolti nelle malattie monogeniche rare.

Impatto della genomica in oncologia

Le nuove conoscenze di genomica medica nell'ultimo decennio hanno fatto riconoscere i tumori come "malattie del genoma" ed hanno stimolato la ricerca oncologica in molteplici direzioni. L'identificazione pressoché completa dell'insieme dei geni che codificano per le proteine, completata con la scoperta di nuovi elementi trascritti come i microRNAs, ha portato a numerose indagini (condotte con l'impiego di procedure *microarray*) su modelli di espressione genica in parecchi tipi di tumore. Inoltre, lo sviluppo di approcci sistematici nell'identificazione di mutazioni somatiche ha suggerito l'esecuzione di analisi complessive delle modificazioni nei genomi tumorali, inclusi cambiamenti del numero di copie di basi, riarrangiamenti, piccole inserzioni e delezioni, mutazioni puntiformi (Stratton et al., 2009). In Tabella 1 è schematicamente riportato lo sviluppo dei test diagnostici basati su *microarray* in oncologia clinica.

Recentemente è stato analizzato il sequenziamento dell'intero genoma di un tumore umano, fornendo un catalogo completo delle variazioni osservate (Pleasance et al., 2010). Studi di questo tipo, sistematicamente condotti, possono dare precise indicazioni sui geni che contribuiscono alla trasformazione neoplastica cellulare. Contemporaneamente la caratterizzazione della variabilità ereditaria nella popolazione umana ha suscitato una serie di indagini sulla predisposizione ai tumori, puntando soprattutto sulle varianti di DNA (SNPs e CNVs) che sono comuni nella popolazione generale e che possono conferire un differente incremento di rischio di varie neoplasie (Stadler et al., 2010; Weitzel et al., 2011). Infine, è stata sviluppata una serie di reagenti biologici che interferiscono con le funzioni dei geni nelle cellule viventi. A tal riguardo i più interessanti sono piccole molecole di RNA (microRNA), che vengono impiegati per parecchi scopi; ad esempio, per determinare quali geni sono richiesti per la sopravvivenza cellulare e quali geni conferiscono sensibilità a determinati farmaci.

Alcuni risultati di queste ricerche incominciano ad essere utilizzati in oncologia clinica. In particolare, sono di grande interesse gli effetti che l'approccio genomico sta avendo su: la classificazione dei tumori, l'identificazione di marcatori prognostici, l'analisi di indicatori predittivi della risposta ai farmaci, lo sviluppo di nuove terapie mirate, la messa a punto di strategie per monitorare una malattia neoplastica, la possibilità di trattare la predisposizione a un tumore (McDermott et al., 2011).

Tab. 1 - Sviluppo di test diagnostici basati su *microarray* in oncologia clinica.

-
- Campioni di tumore vengono prelevati in pazienti diversi, congelati e sottoposti a DNA-*microarray* per ottenere il profilo di espressione genica del tumore.
 - Analisi computazionale e statistica vengono utilizzate per identificare marcatori di espressione genica di responsività o resistenza alla chemioterapia.
 - Vengono sviluppati molteplici test di RT-PCR quantitativa per i maggiori marcatori. Questi test vengono applicati retrospettivamente o prospettivamente a sezioni di tessuto fissate in formalina e incluse in paraffina.
 - Sulla base di questi risultati vengono sviluppati modelli diagnostici e predittivi, poi validati su un gran numero di pazienti, in modo da poter essere utilizzati come strumenti diagnostici definitivi.
-

Farmacogenomica e terapia personalizzata

È conoscenza clinica da tutti condivisa che pazienti differenti rispondono in modo diverso allo stesso medicamento. Sebbene parecchi fattori non-genetici influenzino gli effetti dei medicinali (età, funzioni di organi e apparati, natura della malattia, terapie concomitanti, interazioni di farmaci, ecc.), si possono osservare numerosi esempi di casi in cui differenze individuali della risposta ai farmaci sono dovute a variazioni nella codificazione genetica di:

1. enzimi che condizionano il metabolismo dei farmaci;
2. molecole coinvolte nel trasporto dei farmaci;
3. strutture bersaglio dell'azione dei farmaci (Evans e McLeod, 2003).

A differenza degli altri fattori che influenzano la risposta ai farmaci, le determinanti genetiche ereditate generalmente rimangono stabili per tutta la durata della vita di una persona.

Le prime osservazioni cliniche di differenze ereditarie nella risposta a farmaci sono state fatte negli anni 1950, dando inizio ad una nuova disciplina denominata dapprima "farmacogenetica" e più tardi "farmacogenomica". Sebbene i due termini siano sinonimi sul piano pratico, la farmacogenomica utilizza un approccio più ampio - comprensivo di tutto il genoma - per spiegare le basi ereditarie di differenze nella risposta ai farmaci fra persone diverse.

Soprattutto nell'ultimo decennio gli studi di genomica e la loro applicazione alla risposta ai farmaci hanno consentito di raggiungere importanti acquisizioni soprattutto nell'ambito del trattamento di malattie cardiovascolari (basti ricordare gli studi di associazione fra SNPs e dosi dell'anticoagulante orale warfarina, che hanno portato all'identificazione di varianti di *VKORC1*, *CYP2C9* e *CYP4F2* come le principali determinanti genetiche della posologia di questo farmaco (Takeuchi F. et al., 2009), di malattie infettive e di malattie neoplastiche (Wang et al., 2011).

Particolarmente interessante sul piano pratico è la farmacogenomica dei tumori, resa complessa dal fatto che sono coinvolti due genomi: il genoma del paziente ed il genoma del tumore. Il genoma del paziente può intervenire sulla tolleranza al farmaco, e quindi sulla sua posologia; un esempio di tale fenomeno è rappresentato dalle variazioni nel gene *TPMT*, che codifica l'enzima tiopurina S-metiltransferasi e che quindi può aumentare la sensibilità alla 6-mercaptopurina come risultato di un ridotto metabolismo di tale farmaco (Weinshilboum, 2003). Il genoma del tumore svolge poi un ruolo fondamentale nella variazione della risposta alla terapia antitumorale; gli esempi che la clinica offre sono numerosi a tal riguardo; infatti, basta ricordare la sovraespressione di *HER2* in pazienti con cancro della mammella e la risposta di questo tumore all'anticorpo monoclonale trastuzumab (Slamon et al., 2001), oppure le mutazioni nel gene *EGFR* (che codifica per l'*epidermal growth factor receptor*) che correlano con la risposta clinica alla terapia con gefitinib in pazienti con tumore polmonare non a piccole cellule (Paez et al., 2004).

L'acquisizione che determinanti genetiche condizionano la risposta farmacologica nel trattamento di malattie comuni ha portato in questi ultimi anni a sviluppa-

re il concetto di “terapia personalizzata”. È infatti sogno di ogni clinico poter curare i singoli pazienti con farmaci specificamente mirati verso la causa/meccanismo della malattia (*targeted therapy*), somministrati alla dose più appropriata per essere efficaci e ben tollerati. Nella comunità scientifica è ancora aperto il dibattito se la terapia personalizzata rappresenti una reale speranza, o un'esagerazione trionfalistica, o qualcosa di intermedio (Smith et al., 2005; Shurin e Nabel, 2008; Browmann et al., 2011). Certamente si tratta di un concetto che deve ancora essere riempito di contenuti, e la ricerca farmacogenomica sta attualmente procedendo in tal senso molto attivamente; ne è una prova il programma di studi progettato e sostenuto negli USA. congiuntamente dalla *Food and Drug Administration* e dal *National Institutes of Health* (Hamburg e Collins, 2010).

Conclusioni e prospettive

Le più recenti conoscenze ottenute con il sequenziamento del genoma umano e lo sviluppo delle correlate nozioni di biologia molecolare stanno acquistando una sempre maggiore importanza anche nell'ambito di loro possibili applicazioni in campo clinico. Infatti, l'identificazione di determinanti genetiche che svolgono un ruolo nell'insorgenza di malattie comuni (quali malattie cardiovascolari, infezioni, tumori, malattie immunologiche e neorodegenerative, ecc.), la precisazione dei meccanismi molecolari che causano l'insorgenza di alcuni processi patologici (premessa per un loro efficace controllo terapeutico), la possibilità di prevedere la risposta ad un determinato trattamento farmacologico (nei riguardi sia della sua efficacia che della sua tollerabilità) costituiscono interventi veramente innovativi nell'esercizio della pratica medica.

Necessariamente tali nozioni debbono oggi far parte del patrimonio culturale e professionale di un medico. Gli esempi che possono illustrare tale concetto sono già numerosi, con osservazioni cliniche di ordine sia diagnostico (aumentato rischio di eventi tromboembolici per mutazione del fattore V Leiden, profili di espressione genica correlati all'insorgenza di varie malattie comuni, ecc.), sia prognostico (marcatori genici che possono aumentare o diminuire il rischio e la gravità di una malattia infettiva o neoplastica, marcatori genici di responsività o resistenza alla chemioterapia di emopatie maligne o tumori solidi, ecc.), sia terapeutico (test genetici per ottimizzare la posologia di farmaci anticoagulanti o anti-tumorali prevenendo effetti collaterali indesiderabili, grado di espressione o mutazioni geniche correlate con l'efficacia di una terapia con farmaci mirati su un preciso bersaglio molecolare o cellulare, ecc.). La genomica medica deve quindi attualmente costituire parte integrante del piano universitario di studi per la formazione di giovani medici.

Il progresso delle conoscenze in questo settore di studi è molto rapido, sia sul piano delle nuove acquisizioni scientifiche di base che su quello degli aspetti applicativi clinici. Di questi ultimi aspetti un'applicazione pratica già attuata è la nuova modalità di prescrizione per alcuni farmaci mirati, possibile solo sulla base del risultato di adeguati test molecolari. Si tratta di una normativa di sanità pubblica che correttamente indirizza verso una terapia personalizzata, e che corri-

sponde anche a criteri di farmacoconomia. La possibilità di realizzare terapie personalizzate per ammalati affetti da malattie comuni costituisce però ancora un lungo e faticoso percorso, che richiede parecchie energie per la ricerca e una efficace organizzazione della sanità.

Bibliografia

1. Browman G, Hébert PC, Coutts J. Personalized medicine: a windfall for science, but what about patients? *Canadian Med Ass J* published at www.cmaj.ca on July 18, 2011.
2. Carthew RW, Sontheimer EJ. Origins and mechanisms of miRNAs and siRNAs. *Cell*. 2009; 136: 642.
3. Castanotto D, Rossi JJ. The promises and pitfalls of RNA-interference-based therapeutics. 2009.
4. Collins FS, Morgan M, Patrinos A. The Human Genome Project: lessons from large-scale biology. *Science*. 2003; 300: 286.
5. Cook EH Jr, Scherer SW. Copy-number variations associated with neuropsychiatric conditions. *Nature*. 2008; 455: 232.
6. Davidson EH, Levine MS. Properties of developmental gene regulatory networks. *Proc Nat Acad Sci USA*. 2008; 105: 20063.
7. Dean M, Carrington M, Winkler C, et al. Genetic restriction of HIV-1 infection and progression to AIDS by a deletion allele of the *CCR5* structural gene. *Science*. 1996; 273: 1856.
8. Evans WE, McLeod HL. Pharmacogenomics - Drug disposition, drug targets, and side effects. *N Engl J Med*. 2003; 348: 538.
9. Fearnhead NS, Britton MP, Bodmer WF. The ABC of APC. *Hum Mol Genet*. 2001; 721.
10. Feero WG, Guttmacher AE, Collins FS. Genomic medicine - An updated primer. *N Engl J Med*. 2010; 362: 2001.
11. Froguel P, Velho G. Molecular genetics of maturity-onset diabetes of the young. *Trends Endocrinol Metab*. 1999; 10: 142.
12. Goldberg YP, Nicholson DW, Rasper DM, et al. Cleavage of huntingtin by apopain, a proapoptotic cysteine protease, is modulated by the polyglutamic tract. *Nat Genet*. 1996; 13: 442.
13. Goldstein JL, Hobbs HH, Brown MS, et al. Familial hypercholesterolemia. In: Scriver CR, Beauder AL, Sly WS, Valle D, eds. *The metabolic and molecular bases of inherited diseases*. 8th ed. Vol 2. McGraw-Hill, New York, 2001.
14. Grant SFA, Reid DM, Blake G, et al. Reduced bone density and osteoporosis associated with a polymorphic Sp1 binding site in the collagen type I $\alpha 1$ gene. *Nat Genet*. 1996; 14: 203.
15. Hamburg MA, Collins FS. The path to personalized medicine. *N Engl J Med*. 2010; 363: 301.
16. Hunter DJ, Khoury MJ, Drazen JM. Letting the genome out of the bottle - Will we get out wish? *N Engl J Med*. 2008; 358: 105.
17. Kidd JM, Cooper GM, Donahue WF, et al. Mapping and sequencing of struc-

- tural variazion from eight human genomes. *Nature*. 2008; 453: 56.
18. Lee YS, Dutta A. MicroRNA in cancer. *Ann Rev Pathol*. 2009; 4: 199.
 19. Lynch HT. Hereditary nonpolyposis colorectal cancer (HNPCC). *Cytogenet. Cell Genet*. 1999; 86: 130.
 20. Luzzatto L, Mehta A, Vulliamy T. Glucose 6-phosphate dehydrogenase deficiency. In: Scriver CR, Beauder AL, Sly WS, Valle D, eds. *The metabolic and molecular bases of inherited diseases*. 8th ed. Vol 3. McGraw-Hill, New York, 2001.
 21. Major DA, Sane DC, Harrington DM. Cardiovascular implications of the factor V Leiden mutation. *Am Heart. J*. 2000; 140: 189.
 22. McDermott U, Downing JR, Stratton MR. Genomics and continuum of cancer care. *N Engl J Med*. 2011; 364: 340.
 23. McKusick VA. *Mendelian inheritance in man: cataloges of human genes and genetic disorders*. 12th ed. Baltimore: Johns Hopkins University Press, 1998.
 24. Mouradian MM. Recent advances in the genetics and pathogenesis of Parkinson disease. *Neurology*. 2002; 58: 179.
 25. Nathanson KN, Wooster R, Weber BL. Breast cancer genetics: what we know and what we need. *Nat Med*. 2001; 7: 552.
 26. Online Mendelian Inheritance in Man, OMIM™. Baltimore: McKusick-Nathans Institute for Genetic Medicine, Johns Hopkins University. 2000. (Accessed October 15, 2002, at <http://www.ncbi.nlm.nih.gov/omim/>).
 27. Paez JG, Jänne PA, Lee JC, et al. EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. *Science*. 2004; 304: 1497.
 28. Shurin SB, Nabel EG. Pharmacogenomics - Ready for prime time? *N Engl J Med*. 2008; 358: 1061.
 29. Slamon DJ, Leyland-Jones B, Shak S, et al. Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2. *N Engl J Med*. 2001; 344: 783.
 30. Smith GD, Ebrahim S, Lewis S, et al. Genetic epidemiology and public health: hope, hype, and future prospects. *Lancet*. 2005; 366: 1484.
 31. Sotiriou C, Pusztai L. Gene-expression signature in breast cancer. *N Engl J Med*. 2009; 360: 790.
 32. Stadler ZK, Thom P, Robson ME, et al. Genome-wide association studies of cancer. *J Clin Oncol*. 2010; 28: 4255.
 33. Stratton MR, Campbell PJ, Futreal PA. The cancer genome. *Nature*. 2009; 458: 719.
 34. Takeuchi F, McGinnis R, Bourgeois S, et al. A genome-wide association study confirms VKORC1, CYP2C9, and CYP4F2 as principal genetic determinants of warfarin dose. *PLoS Genet*. 2009; 5: e1000433.
 35. The International HapMap Consortium. A haplotype map of the human genome. *Nature*. 2005; 437: 1299.
 36. Todd JA. From genome to aetiology in a multifactorial disease, type 1 diabetes. *Bioessays*. 1999; 21: 164.
 37. Wang WY, Barratt BJ, Clayton DG, et al. Genome-wide association studies. Theoretical and practical concerns. *Nat Rev Genet*. 2005; 6: 109.

38. Wang L, McLeod HL, Weinshilboum RM. Genomics and drug response. *N Engl J Med.* 2011; 364: 1144.
39. Weinshilboum RM. Inheritance and drug response. *N Engl J Med.* 2003; 348: 529.
40. Weitzel JN, Blazer KR, MacDonald DJ, et al. Genetics, genomics, and cancer risk assessment. *CA Cancer J Clin.* 2011; 61: 327.
41. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature.* 2007; 447: 661.

**STRUMENTI
DI MEDICINA GENOMICA**

Sequenziamento genico

Elena Ronchetti

Laboratorio di Oncologia Sperimentale e Farmacogenomica, Fondazione S. Maugeri, Pavia

Il termine **sequenziamento**, in biologia molecolare indica il processo per la determinazione dell'esatta struttura primaria di un biopolimero, e cioè dell'ordine delle basi nel caso di un acido nucleico o degli aminoacidi nel caso di proteine, che rappresentano le strutture sequenziate in prevalenza.

Il **sequenziamento del DNA** è la determinazione dell'ordine dei diversi nucleotidi (Adenina, Citosina, Guanina, Timina) che costituiscono l'acido nucleico. La sequenza del DNA contiene tutte le informazioni genetiche ereditarie del nucleo, plasmidi, mitocondri e cloroplasti che sono alla base per lo sviluppo di tutti gli organismi viventi. All'interno di questa sequenza sono codificati i geni di ogni organismo vivente, nonché le istruzioni per esprimerli nel tempo e nello spazio (regolazione dell'espressione genica). Determinare la sequenza è dunque utile nella ricerca del *perché* e *come* gli organismi vivono.

La conoscenza del genoma risulta quindi utile in ogni campo della biologia e l'avvento di metodi per il sequenziamento del DNA ha accelerato significativamente la ricerca. In medicina, ad esempio, il sequenziamento è usato per identificare e diagnosticare, ad esempio, malattie ereditarie e, in farmacogenomica, per sviluppare nuovi trattamenti mirati in dipendenza della presenza o meno di mutazioni o polimorfismi modulatori della risposta al farmaco. Ancora, il genoma degli agenti patogeni può portare allo sviluppo di medicine contro malattie contagiose.

La rapidità del processo di sequenziamento oggi è stato di grande aiuto per il sequenziamento su larga scala del genoma umano. Allo stesso modo, è stato completato il sequenziamento del genoma di diversi organismi animali e vegetali, nonché di molti microrganismi.

La determinazione di sequenze di DNA è risultata utile anche in diversi campi applicativi, come le scienze forensi e quelle alimentari.

Sono state ideate diverse strategie per ottenere la sequenza nucleotidica del DNA. I primi metodi, tra cui quello ideato da Allan Maxam e Walter Gilbert tra il 1973 ed il 1977 erano piuttosto complicati (basato su modificazioni chimiche del DNA e sul conseguente taglio in posizioni specifiche) (1); una svolta si ebbe nel 1975 con la prima pubblicazione di una strategia enzimatica tuttora diffusissima, sviluppata da Frederik Sanger e collaboratori (il cosiddetto **metodo dei terminato-**

ri di catena, chain termination method o **metodo Sanger**, dal nome del suo scopritore) (2, 3), che ricevette per questo il suo secondo premio Nobel. Il metodo Sanger è stato quello più utilizzato in assoluto per la sua relativa semplicità che ha permesso anche la produzione di kit commerciali e un'elevata automazione degli strumenti di sequenziamento (ABI 310, 3130 e 3130xl, 3730) (**sequenziamento di prima generazione**).

Più recentemente sono stati sviluppati nuovi metodi caratterizzati dalla capacità di sequenziare molti frammenti di DNA contemporaneamente (anche se con efficienza minore in termini di numero di basi sequenziate per frammento) aprendo una nuova era del sequenziamento. Queste metodiche vanno sotto il nome di sequenziamento ad elevato parallelismo o pirosequenziamento fino alla Next Generation Sequencing.

Queste nuove tecnologie rappresentano strategie molto più complesse rispetto a quelle di prima generazione. Si basano infatti su una combinazione di preparazione del template, di sequenziamento, di imaging, di allineamento del genoma e metodi di assemblaggio molto più sofisticati e basati su principi differenti per ogni piattaforma, cosa che rende possibile la coesistenza sul mercato di parecchi sistemi (Illumina/Solexa; Roche/454; Applied Biosystems/Ion Torrent) che possono essere scelti a seconda delle esigenze.

Mentre il metodo convenzionale di Sanger è basato sul sequenziamento diretto di porzioni di genoma, singolarmente amplificate per polymerase chain reaction (PCR) o clonate in plasmide, e consente di sequenziare fino a 800 bp circa per singola reazione, le tecniche di ultima generazione consentono invece di sequenziare parallelamente più regioni del genoma, da ~400 a 3.000 Mb in una singola corsa dello strumento. Il genoma viene dapprima frammentato in molecole di circa 35-400 bp, che sono successivamente amplificate clonalmente in parallelo fino a generare circa 50 copie identiche per singola molecola. Successivamente, avviene il sequenziamento vero e proprio, per il quale possono essere utilizzati due differenti approcci: il sequenziamento "per sintesi", basato sul metodo di Sanger oppure sulla reazione di pirosequenza e il sequenziamento "per ligazione". Con il NGS, l'acquisizione dei dati di sequenza avviene contestualmente per tutte le reazioni, con un notevole risparmio di tempo rispetto alla tecnologia precedente. L'enorme mole di dati generati, opportunamente processata e analizzata per via bioinformatica, viene poi confrontata con un genoma di riferimento (4). È proprio quest'ultima parte e cioè l'elaborazione del dato a costituire l'unico collo

Tab. 1 - Sequenziamento di Sanger vs. NGS.

Sanger		NGS
3h	Tempo necessario per una corsa	10h-6 gg
Max 96	Sequenze analizzate per corsa	Milioni
80 Kb	Totale basi lette per corsa	500 Mb - 50 Gb
70 milioni \$*	Costo per sequenziamento di un intero genoma umano	5000 \$

*La cifra si riferisce al costo stimato per il sequenziamento del primo genoma umano, quello di Craig Venter, realizzato dalla Celera nel 2001. Il costo è comprensivo dell'assemblaggio delle sequenze: essendo stata la prima sequenza genomica umana ad essere stata generata non si avvale dell'allineamento con una sequenza esistente, come avviene adesso (5).

di bottiglia nel sequenziamento di nuova generazione. In termini invece di costi il sistema NGS è sensibilmente più economico rispetto a quello di prima generazione a parità di numero di basi sequenziate. Inoltre il numero di basi lette e il numero di sequenze allineate per corsa sono di gran lunga superiori (Tab. 1). Da ciò si può capire come l'avvento delle tecnologie di Nuova Generazione sul mercato abbia cambiato in modo determinante il nostro modo di approcciarci alla genetica non solo nella ricerca di base, ma anche nella ricerca applicata e clinica. Il grande passo in avanti offerto dalla Next Generation Sequencing è essenzialmente la capacità di produrre un enorme volume di dati a basso costo, capacità aumenta notevolmente l'ambito di sperimentazione. La possibilità di sequenziare l'intero genoma di organismi correlati ha permesso, ad esempio, studi comparativi ed evolutivi inimmaginabili solo fino a pochi anni fa. L'applicazione di maggior impatto della NGS è sicuramente la possibilità di poter risequenziare il genoma umano in tempi brevi dandoci uno strumento rapido e completo per capire meglio la predisposizione individuale ad alcune patologie invece che ad altre. L'applicazione in farmacogenomica è direttamente collegata a quest'aspetto. Le malattie croniche e le patologie oncologiche costituiscono un grande onere per il sistema sanitario in quanto coprono una sempre più crescente percentuale in termini di morbilità e mortalità facendo nascere la necessità di selezionare e di mirare quanto più possibile i trattamenti farmacologici. La possibilità di creare, per ogni individuo analizzato, un profilo di polimorfismi e/o mutazioni che attivano o disattivano la risposta ai farmaci, permette, da un lato una scelta più specifica del farmaco in termini di efficacia e, dall'altro, un vantaggio economico per il Sistema Sanitario Nazionale che arriverebbe a poter finanziare solo le terapie veramente utili (6).

Bibliografia

1. Maxam AM, Gilbert W. A new method for sequencing DNA, *Proc Natl Acad Sci USA*. 1977; 74 (2): 560-4.
2. Sanger F, Coulson AR. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *J Mol Biol*. 1975; 94 (3): 441-8.
3. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA*. 1977; 74 (12): 5463-7.
4. Metzker ML. Sequencing technologies. The next generation. *Nat Rev Genet*. 2010; 11: 31-46.
5. Piccolo P, Brunetti-Pierri N. Il sequenziamento dell'esoma per l'identificazione di geni-malattia e la diagnosi molecolare. In: *Prospettive in pediatria*. Società Italiana di Pediatria. Luglio-settembre 2011; 41, 163: 156-64.
6. Seeley ES, Shi RZ, Jannetto P. Overview of pharmacogenomics and application for the modern clinical laboratory. In: *Molecular diagnostics: techniques and applications for the clinical laboratory*. 401-7.

Microarray technologies

Angela Amato

Laboratorio di Oncologia Sperimentale e Farmacogenomica, Fondazione S. Maugeri, Pavia

La tecnologia del DNA Microarray rappresenta una rivoluzione della biologia e del modo di fare ricerca permettendo di estendere lo studio dei geni all'intero genoma. Questa tecnologia permette infatti la determinazione di profili di espressione genica globale in un campione (test) rispetto a un campione di riferimento (reference).

La rivoluzione rappresentata dal DNA Microarray consiste nella possibilità di analizzare migliaia di geni simultaneamente. Diverse sono gli ambiti sperimentali in cui il DNA Microarray trova un'applicazione, permettendo:

- L'identificazione di geni espressi in maniera differenziale in condizioni diverse (es. geni espressi da un patogeno all'interno della cellula ospite saranno diversi a quelli che esprime quando viene coltivato in vitro).
- L'individuazione di geni markers per alcune patologie alle quali la loro espressione risulta essere correlata.
- L'identificazione di gruppi di geni che svolgono funzioni correlate o la cui espressione viene regolata in maniera concomitante.
- L'identificazione di sottotipi diversi di una stessa malattia (la scoperta di fenotipi diversi ha permesso la distinzione di sottogruppi nel tumore alla mammella), conoscenza che permette di scegliere in maniera opportuna la terapia anti-tumorale.

Questo rende la tecnologia del DNA Microarray versatile per l'applicazione in diversi ambiti della ricerca: microbiologia, biologia cellulare, oncologia e farmacogenomica.

Il DNA Microarray permette di realizzare un'analisi dell'espressione genica di un campione rispetto ad un controllo, ma a differenza dei convenzionali metodi di analisi dell'espressione genica (Southern Blot, Northern Blot, FISH), i microarray sfruttano una tecnica di ibridazione inversa, che consiste nel fissare tutti i segmenti di DNA (probes) su un supporto e nel marcare invece l'acido nucleico del test e del reference.

La tecnologia del DNA microarray è racchiusa in un chip, un supporto solido di vetro, plastica o silicio nel quale vengono legate migliaia di sonde di DNA (Fig. 1). Diverse piattaforme esistono che differiscono nel tipo di probe e nelle tecniche di preparazione del chip, le principali piattaforme utilizzate sono rappresentate dai cDNA Array e dagli array ad alta densità.

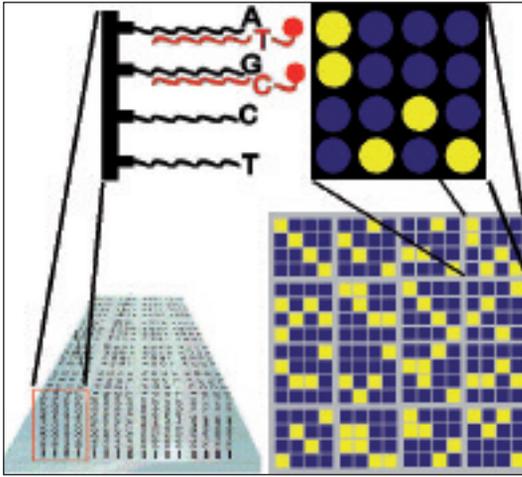


Fig. 1 - Un chip può contenere 10000-20000 spots su un'area di 3.6 cm². Ogni spot rappresenta il probe (cDNA o sequenze oligonucleotidiche) specifico per un gene (immagine tratta da SM Carr et al. 2008. *Comp. Biochem. Physiol. D*, 3: 1).

La prima consiste nell'utilizzo di sonde di cDNA prodotte per PCR che vengono legate covalentemente al chip, in genere di vetro o nylon (Scheda et al., 1995).

La seconda metodologia sviluppata dalla Affimetrix Inc. ha generato dei chip (Gene Chip) in cui diverse sequenze oligonucleotidiche rappresentanti un unico gene vengono direttamente sintetizzate in situ, mediante tecniche fotolitografiche (Chee et al., 1996).

Con la prima tecnologia i campioni (test e reference) vengono marcati ciascuno con un fluorocromo diverso e ibridati allo stesso chip. Infine l'abbondanza relative del test rispetto al controllo viene misurata.

Con la seconda tecnologia l'analisi del dato diventa più complicata. I due campioni vengono marcati con lo stesso fluorocromo ma ibridati a due array differenti. Questo permette di poter valutare la quantità assoluta del contenuto di mRNA, cosa che richiede un processo di normalizzazione del dato (Liu et al., 2003).

I vantaggi della prima tecnologia sono: l'alta riproducibilità, l'automazione nelle procedure sperimentali e il costo relativamente basso, mentre il secondo presenta un'alta densità e versatilità dal momento che si presta alla sintesi di chip custom.

Design e preparazione dell'esperimento di DNA Microarray

La preparazione del target prevede l'isolamento dell'mRNA totale dal test e dal reference, pari concentrazioni di entrambe le popolazioni di mRNA totale vengono retrotrascritte per la produzione di cDNA, molecola più stabile dell'mRNA. Durante la retrotrascrizione le molecole di cDNA provenienti dai due diversi campioni vengono marcate con fluorocromi diversi Cy5 (rosso) e Cy3 (verde) e mescolate dopo la marcatura e ibridate allo stesso chip, nel caso in cui si utilizzi una piattaforma a due colori, oppure la marcatura dei due campioni viene effettuata con lo stesso fluorocromo e i campioni fatti ibridare a due differenti chip.

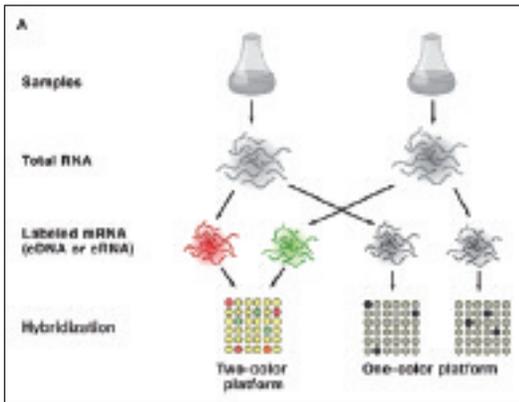


Fig. 2 - Schematizzazione delle fasi di preparazione del Chip per l'analisi Microarray: isolamento della popolazione di mRNA da test e reference, retroscrittione e marcatura del cDNA con fluorocromi, ibridazione al chip. La marcatura sarà diversa a seconda della piattaforma utilizzata (diversa per entrambi i campioni nella piattaforma a due colori, identica per la piattaforma ad un colore).

Durante l'ibridazione le molecole di cDNA si legano ciascuna al probe specifico in una precisa area del chip.

A questo punto il chip è pronto per la lettura allo scanner capace di catturare la fluorescenza emessa dal chip, ovvero da ciascuno spot del chip fluorescente, dopo eccitazione con un raggio laser.

Lo scanner catturerà la fluorescenza emessa per ciascun colore separatamente.

Se il campione test viene marcato con Cy5 e il campione reference con il Cy3 la fluorescenza rossa fornirà una informazione relativa all'espressione genica nel test, mentre la fluorescenza verde darà un'informazione relativa all'espressione genica nel reference. Tuttavia l'analisi del dato prevede la sovrapposizione delle due immagini monocromatiche e l'elaborazione mediante software appositi per convertire i segnali in una gamma di colori dipendente dalla loro intensità (Fig. 2).

Il risultato sarà l'immagine di un chip contenente spots che variano dal rosso al verde passando per diverse sfumature di colore (giallo) intermedio. Ciascuno spot rosso/verde/giallo corrisponde a un singolo cDNA e dunque all'espressione di un singolo gene, che potrà essere up-regolata se maggiore rispetto al reference, down-regolata se minore, oppure potrebbe risultare invariata (Fig. 3).

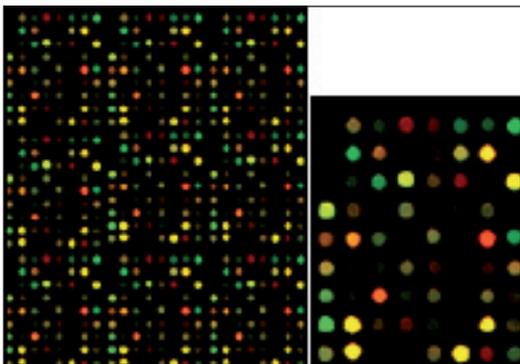


Fig. 3 - Immagine del chip ottenuta dopo la lettura allo scanner, sulla destra un particolare dell'immagine è stato ingrandito.

Analisi del Chip

L'analisi del chip basata sull'immagine ad alta risoluzione acquisita dallo scanner prevede diverse fasi:

1. Gridding (grigliatura): identifica la localizzazione dello spot sul chip;
2. Segmentazione: distingue il background dal foreground;
3. Estrazione dell'intensità: per la sottrazione del background.

La grigliatura localizza nella immagine scannerizzata la posizione degli spots che corrispondono ai singoli probes.

L'estrazione di intensità calcola l'intensità della fluorescenza rossa, della fluorescenza verde e l'intensità del background.

La segmentazione permette invece di separare il segnale emesso dai marcatori fluorescenti (foreground) rispetto al rumore di fondo (background), in modo da isolare il segnale di interesse. Affinché dati ottenuti su array distinti possano essere paragonati è necessaria la rimozione di alcune distorsioni sistematiche che vengono introdotte durante la fase di preparazione dell'array, durante l'esperimento, durante la fase di ibridazione del cDNA e nella scansione del laser.

Il principale problema dei microarray è la mancanza di standardizzazione, che causa difficoltà nel confronto di dati portando spesso a dei falsi positivi, questo accade anche a causa del fatto che alcuni cDNA possono cross-ibridare altre sonde (che avrebbero dovuto rilevare altri geni). Un altro problema è presentato dai fluorofori, che nonostante siano molto simili fra loro presentano delle differenze problematiche. Esiste una diversa efficienza di fluorescenza e di incorporazione tra Cy3 e Cy5 che deve essere standardizzata dai software di rilevazione. Altre variabili possono essere introdotte e dipendere dal processo di preparazione dei campioni (volumi di reazione, errori di pipettaggio, variabilità nella PCR), o dalla fase di ibridazione del chip. Per ovviare a queste problematiche e per creare un minimo di standardizzazione si effettua il dye swap, utile metodo di standardizzazione nel DNA microarray a due colori (cDNA Array) che consiste nell'effettuare un secondo microarray scambiando l'uso dei fluorofori, ovvero se nel primo microarray Cy3 è stato usato per marcare il cDNA del test, nel secondo microarray si userà Cy3 per marcare il cDNA del reference, e viceversa per Cy5. La normalizzazione risulta più complessa nello studio mediante DNA microarray a singolo colore, in cui il segnale rilevato dallo scanner viene poi sottoposto ad algoritmi di filtrazione e di pulizia e convertito in valori numerici.

La normalizzazione prevede che si calcolino fattori di standardizzazione per ciascuno dei tre effetti sopra citati e può usare tre principi diversi.

La prima si basa sul fatto che i geni differenzialmente espressi sono rappresentati da un piccolo sottoinsieme e che in realtà la maggioranza di gene ha una espressione costante per cui la normalizzazione tiene conto dei geni sull'intero chip.

Il secondo metodo si basa sul fatto che i geni housekeeping hanno una espressione invariata nei due campioni test e reference e usa la loro espressione per la normalizzazione, tuttavia questi geni sono difficili da identificare e inoltre non sono rappresentativi dell'intera popolazione di cDNA perché sono molto espressi e quindi rappresentati dagli spot più fluorescenti.

Il terzo metodo invece si basa sulla lettura doppia di ciascun campione marcato separatamente con i due fluorocromi, ottenendo così la stessa intensità degli spots per il rosso e per il verde, le differenze verranno utilizzate per la normalizzazione.

Infine la filtrazione, permette di eliminare dall'analisi quei geni la cui misurazione non è sufficientemente accurata, o che presentano un livello di espressione molto piccolo (prima o dopo la normalizzazione). In pratica tutti gli spot la cui differenza fra foreground e background non supera un valore soglia di 1,4 fold (nell'intensità di fluorescenza) vengono eliminati. Questa osservazione si basa sull'assunto che livelli di espressione inferiori a 1,4 fold sono, in genere, frutto di errore di misura.

Molti articoli scientifici e reviews sono stati pubblicati sui vari approcci e metodi analitici usati nell'analisi dei dati DNA microarray (e.g. Quackenbush 2001, Ringner et al., 2002, Simon et al. 2003, Ahmed & Brenton 2005, Allison et al., 2006), in aggiunta svariati e sofisticati databases per microarray, software di analisi che permettono non solo l'analisi del dato ma anche l'elaborazione per l'identificazione di pathways e networks di regolazione genica sono disponibili commercialmente e in ambito accademico permettendo una migliore interpretazione biologica dei complessi dati ottenuti dall'analisi DNA Microarray, fra questi Gene Ontology (Ashburner et al. 2000); GenMAPP (Dahlquist et al. 2002); SMD (Sherlock et al., 2001); TM4 (Saeed et al., 2003); BASE (Saal et al., 2002); GSEA (Subramanian et al., 2005); Ingenuity Systems (www.ingenuity.com); MetaCore (www.genego.com); and PathwayStudio (www.ariadnegenomics.com).

Bibliografia

1. Schena M, Shalon D, Davis RW, Brown PO. *Science*. 1995; 270 (5235): 467-70.
2. Chee M, Yang R, Hubbell E, Berno. A, Huang XC, et al. *Science*. 1996; 274: 610-14. 13.
3. Liu L, Hawkins DM, Ghosh S, Young SS. *Proc Natl Acad Sci USA*. 2003; 100 (23): 13167-72.
4. Quackenbush J. *Nat Rev Genet*. 2001; 2 (6): 418-27.
5. Ringnér M, Peterson C. *Biotechniques*. 2003 (Suppl); 30-5.
6. Simon R, Radmacher MD, Dobbin K, Lisa M. McShane Ahmed & Brenton 2005, *JNCI J Natl Cancer Inst*. 2003; 95 (1): 14-18.
7. Allison DB, Cui X, Page GP, Sabripour M. *Nat Rev Genet*. 2006; 7 (1): 55-65.
8. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G. *Nat Genet*. 2000; 25 (1): 25-9.
9. Dahlquist KD, Salomonis N, Vranizan K, Lawlor SC, Conklin BR *Nat Genet*. 2002; 31 (1): 19-20.
10. Sherlock G, Hernandez-Boussard T, Kasarskis A, Binkley G, Matese JC,

- Dwight SS, Kaloper M, Weng S, Jin H, Ball CA, Eisen MB, Spellman PT, Brown PO, Botstein D, Cherry JM. *Nucleic Acids Res.* 2001; (1): 152-5.
11. Saeed AI, Sharov V, White J, Li J, Liang W, Bhagabati N, Braisted J, Klapa M, Currier T, Thiagarajan M, Sturn A, Snuffin M, Rezantsev A, Popov D, Ryltsov A, Kostukovich E, Borisovsky I, Liu Z, Vinsavich A, Trush V, Quackenbush J. *Biotechniques.* 2003; 34 (2): 374-8.
 12. Saal LH, Troein C, Vallon-Christersson J, Gruvberger S, Borg A, Peterson C. *Genome Biol.* 2002; 3 (8): SOFTWARE0003. Epub 2002 Jul 15.
 13. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP.; *Proc Natl Acad Sci USA.* 2005; 102 (43): 15545-50.

Cancer biobanking

Alberto Zambelli

Divisione di Oncologia 1, Fondazione Salvatore Maugeri, Pavia

Introduction

In the postgenomic era of biological research, tumor biobanks are an essential resource to improve knowledge on cancer biology. Although the specific mission of a biorepository may require the use of different procedures for collection and processing, common principles may be applied to all specimen types. In the oncology field, high-quality biobanks are now a clinical necessity because of the emerging molecular classifications of tumors as well as of the increasing number of targeted treatments requiring the identification of molecular markers specific to the action mechanism of these treatments. The mandate of a biobank is purely for research in the field of cancer. In this context biobanks are integrated service infrastructures that provide biospecimens and data critical for understanding the long-term benefits and/or side effects of drugs and various treatment options. Biobanks' collections and databases can also reduce time for the development of drugs or diagnostics by providing better logistical or operational support and providing researchers with well-organized molecular and clinical data. Because of the nature of a cancer-oriented biobank, the policies are designed in ways to gather consecutive collections, meaning that there is no specific aim at time of collection. According to this banking model, in fact, samples of potential interest are collected and stored until needed. The advantage is that it immediately provides investigators with all the samples and data required. Samples must be "fit for use" and this also encourages the collection of specimens even when variables exceed recommended limits, with the prospective to match requirements for future application.

Legal and ethical issues

Biological samples cannot be collected and stored in a biobank unless the sample donor is informed of the reasons and objectives for which the samples may be used. The donor then may or may not provide consent. In common practice, a

procedure that provides the sample donor with fairly basic information on the nature of a biobank and the regulations it follows should be sufficient for the donor to provide or refuse consent for biological material storage for certain generally described objectives. Rules governing practices within the healthcare system are produced by country councils and other responsible authorities within the healthcare system. Consent can be withdrawn entirely or partially at any time, and when referred to any type of use, the sample and/or data must be destroyed or all identification labels removed. In Italy, the Privacy Guarantor, who is responsible of surveillance for the application of the legislation on personal data (law 675/96 and subsequent modifications 9/5/97 and 28/7/97), has authorized data processing only for scientific research purposes. The results obtained may be diffused only in an anonymous and/or aggregated form.

Standard operating procedures

The absence of a uniform tissue collection procedure and standardized protocols for specimen handling may certainly limit the usefulness of many specimens. SOPs are an essential tool for minimizing preanalytical handling variability that can affect biospecimen quality. Several international organizations have provided written guidelines for biorepositories, allowing biobanks to develop harmonized operating procedures that comply with good practice and regulations.

Storage

Solid tissue specimens can be preserved in 4 ways:

1. Flash frozen immediately with liquid nitrogen.
2. Embedded in optimal cutting temperature medium and frozen on dry ice.
3. Formalin-fixed and embedded in paraffin (FFPE).
4. RNAlater.

The various issues concerning each of the above preservation methods have been widely discussed in literature. First-hand flash freezing can be considered, but FFPE is the choice when sample quantities are limited. It is preferable to place specimens in cryotubes prior to flash freezing to avoid contamination. For long-term storage of specimens, -80°C mechanical freezers are recommended, although liquid nitrogen isothermal freezers would be more appropriate but require higher running costs. Issues have been reported when using RNAlater because of problems concerning more mechanical degradation occurring during cell disruption; enzyme activity is completely inhibited, precluding any trypsin treatment; RNAlater is unsuitable for comparative proteomics analyses.

Biospecimen identification

A unique local identifier should be assigned to each aliquot and depending on the nature of the donation process and terms of utilization consented by the donor; this can be done according to procedures provided by the International Conference on Harmonization of Technical Requirements (ICH, June 1996). These procedures contemplate 4 levels of sample identification:

1. “Identified” data and samples are labeled with personal identifiers such as name or identification numbers.
2. “Coded” data and samples are labeled with at least 1 specific code and do not carry any personal identifiers.
3. “Anonymized” data and samples are initially single or double coded but the link between the subjects’ identifiers and the unique code(s) is subsequently deleted. Once the link has been deleted it is no longer possible to trace the data and samples back to individual subjects through the coding key(s).
4. “Anonymous” data and samples are never labeled with personal identifiers when originally collected, neither is a coding key generated. Therefore, there is no potential to trace back genomic data and samples to individual subjects.

Minimum dataset

The definition of a communal minimum dataset is fundamental for sample and data sharing. In Italy, the RIBBO network of biobanks (the network of biobanks of Alleanza Contro il Cancro) has developed a common minimum dataset for cancer biobanks.

The dataset consists of a number of variables, some of which are compulsory. The dataset is organized with variables (attributes) associated to the entity “donor” and some associated with the entity “sample.” Donor variables include information about age, sex, vital status, and the donor local code. Sample variables include a section on general sample information (date of collection, local code, organ description, cold ischemia time, etc.) and specific sample data (diagnosis, topography, TNM, grade, etc.). Conservation information and quantity are also requested. Moreover, the single biobank has to flag its samples, indicating the type of usage allowed (no restriction, restricted to certain types of projects, depending on a collaboration agreement). One important figure of this minimum dataset is that it is not required to give detailed information about demographic, clinical, or epidemiological information or on the genotypic end/or phenotypic characterization of the samples, but the biobank has to indicate for every sample their availability. In this way, the scientist querying the communal database will be able to identify samples of interest and the associated information can be retrieved on a case-to-case basis.

Collecting and managing clinical data

The Information Technology (IT) is essential for collecting and managing biological and clinical data.

Extensive annotation of tissue specimens is crucial to the overall usefulness of a biorepository as a resource for scientific research. In particular, the data recorded by tumor biobanks, which are utilized for a variety of purposes, including target treatment discovery and validation, prevention research, research on early detection, and genetic and epidemiologic analyses, will depend on the types of collected samples and the studies they support. Therefore, IT, becoming increasingly critical to the research purposes of tumor biobanks, must be robust and reliable

and able to meet changing needs while remaining interoperable. 3 main fields must be covered:

1. Data management
2. Inventory control
3. Tracking

An IT system should support all aspects of biorepository operations, including patient enrolment and consent; biospecimen collection, processing, storage, and distribution; quality assurance and quality control; collection of patient data; validation documentation; and management reporting functions. The IT system should also manage specimen clinical annotations and, where possible, provide patient follow-up information in compliance with ethical considerations and appropriate regulations. The IT system should also be interoperable with different endpoint data (eg, proteomics and genomics) to ensure that integration of data from multiple sources is possible.

The University of Pavia, the Collegio Ghislieri and the IRCCS Fondazione Salvatore Maugeri of Pavia (FSM), has recently started an IT initiative to support clinical and translational research in oncology and particularly in the managing biological/clinical data of the cancer tissue Biobank at S. Maugeri Institute of Pavia. The IT project, called ONCO-i2b2, has been funded by the Lombardia region, grounds on the software developed by the Informatics for Integrating Biology and the Bedside (i2b2) NIH project. Using i2b2 and new software modules purposely designed, data coming from multiple sources are integrated and jointly queried. The core of the integration process stands in retrieving and merging data from the Biobank management software and from the FSM hospital information system. The integration process is based on a ontology of the problem domain and on open-source software integration modules. A Natural Language Processing module has been implemented, too. This module automatically extracts clinical information of oncology patients from unstructured medical records. The system currently manages more than two thousands patients and will be further implemented and improved in the next years.

The biobank network

Research studies require not only high-quality samples but also adequate numbers. The rarer the disease, or specific subtype of interest, the more difficult this is to provide. Often, cancers are too small or too rare to permit a proper sampling for research. To increase the sample numbers available for research studies, tumor banks need to collaborate and pool their collections. Building the capacity of biological resources is one of the main goals of national and international networks along with exchange of data and material at the international level. Again this means that SOPs are mandatory in order to obtain comparable results when using samples delivered from multiple collection sites. Data associated with research samples and patients from whom they are donated must be structured, comparable, and machine readable. Therefore, one of the key issues in national and international biobanking organizations is to define and apply standardized annotation

and management of clinical data. Data integration is important on multiple levels. As the situation exists today, data integration tools and appropriate methodologies are important to maximize scientific value from the multiple datasets at biobanks. In this context biobanks should agree on a minimum dataset to be interchangeable between biobanks and identify the complete ontology and multilingual definition for this dataset. Sample distribution and utilization are guaranteed owing to supervision of the ethical and the scientific committees within the biobank institution.

Conclusion

A cancer tissue biobank, as described earlier, is an extremely complex structure that can present issues at every level of activities. These issues must be addressed in the outset of the biobank, but in the case of biobanks deriving from the evolution of a repository collection, “adjustments” are necessary in order to provide the biobank with the correct instruments for conducting activities with high-quality standards. The advantage of being able to monitor and control every process during collection, storage, and distribution of samples and data represents a high-level quality control tool that can rely on procedures carried out in the same facility.

BIOINFORMATICA CLINICA

Banche dati biomediche e risorse on-line

Marco Masseroli

Dipartimento di Elettronica e Informazione, Politecnico di Milano

La moderna ricerca genomica e proteomica e i numerosi progetti volti a decifrare la struttura e le funzionalità di vari genomi, primo fra tutti il Progetto Genoma Umano, hanno prodotto e continuano a generare una crescente quantità di risultati pubblicamente disponibili, aprendo, nell'attuale "era post-genomica", nuove prospettive per lo studio delle patologie ereditarie. Grazie anche ai progressi delle moderne tecnologie biomolecolari, i tanti esperimenti che vengono realizzati per l'analisi del genoma di vari organismi continuano a produrre una grande quantità di dati e informazioni, relative a molti geni e ai prodotti proteici da loro codificati, che continuano ad accumularsi in diverse numerose "banche dati genomiche biomolecolari". Tali banche dati hanno principalmente la funzione di conservare e organizzare dati biomolecolari e informazioni biomediche a loro associate, consentendone un facile accesso ed estrazione. Ne consegue che tali banche dati sono oggi uno strumento fondamentale nell'ambito della ricerca genomica e biomedica per un approccio globale e completo alle informazioni e conoscenze disponibili (2, 3).

Banche dati genomiche biomolecolari e dati forniti

Le banche dati genomiche biomolecolari raccolgono dati genomici e/o proteomici di diversi organismi e forniscono un'interfaccia per la loro interrogazione, costituendo un ambiente per gestire la grande quantità eterogenea di informazione strutturale e funzionale relativa a geni e proteine e a fenotipi patologici noti a loro associati. Queste banche dati rappresentano la maggior fonte di informazione su sequenze biomolecolari (DNA, RNA, proteine), dati sperimentali, referenze bibliografiche e conoscenze biomediche e cliniche su geni e proteine di vari organismi. Pertanto costituiscono la fonte primaria d'informazioni e conoscenze a priori per applicazioni bioinformatiche di analisi volte a supportare l'interpretazione di risultati sperimentali e ad ampliare la conoscenza biomedica.

Relativamente alla quantità di informazioni disponibili (Fig. 1), al febbraio 2012 si conosceva la sequenza genomica completa di più di 5'533 specie (inclusi 2'673 virus, 800 fagi, 1'802 batteri, 122 archea e 151 eucarioti). I genomi più studiati sono quelli dell'uomo, del moscerino della frutta (*Drosophila melanogaster*), del

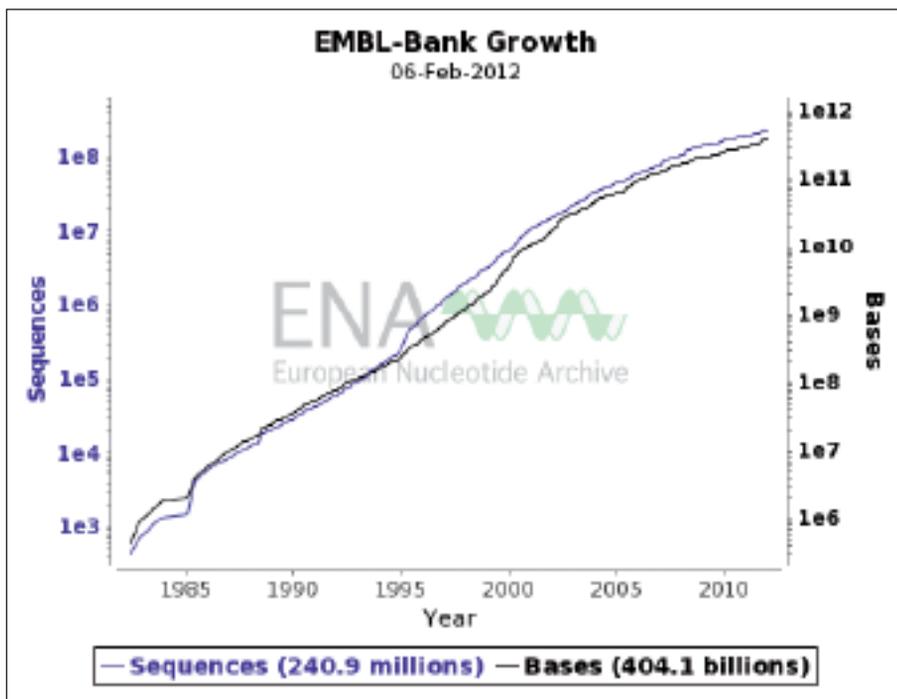


Fig. 1 - Andamento dal 1982 al 2011 del numero di sequenze e basi nucleotidiche presso la banca dati EMBL (<http://www3.ebi.ac.uk/Services/DBStats/>).

topo, del ratto, del pesce zebra, della piantina da fiore *Arabidopsis thaliana*, del batterio *Escherichia coli*, del lievito, del pisello, del mais e del grano.

Di vari di tali organismi, oltre alla sequenza genomica, sono disponibili anche numerose altre informazioni strutturali e funzionali. La maggior parte di tali informazioni è contenuta in numerose banche dati biomolecolari pubblicamente e liberamente accessibili via internet, ed è raggruppabile in:

- sequenze nucleotidiche,
- dati di mappatura genomica,
- profili di espressione (2D-SDS PAGE, DNA chips),
- sequenze aminoacidiche,
- strutture 3D di acidi nucleici e proteine,
- dati metabolici,
- annotazioni funzionali,
- informazioni bibliografiche.

Dato il crescente aumento e specializzazione delle banche dati genomiche biomolecolari disponibili, dal 1994 ogni anno la rivista scientifica "Nucleic Acids Research" (<http://nar.oupjournals.org/>) pubblica un numero speciale sulle banche dati di biologia molecolare, che include una lista di tali banche dati pubblicamente accessibili, assieme a una loro breve descrizione e all'URL ove sono acces-

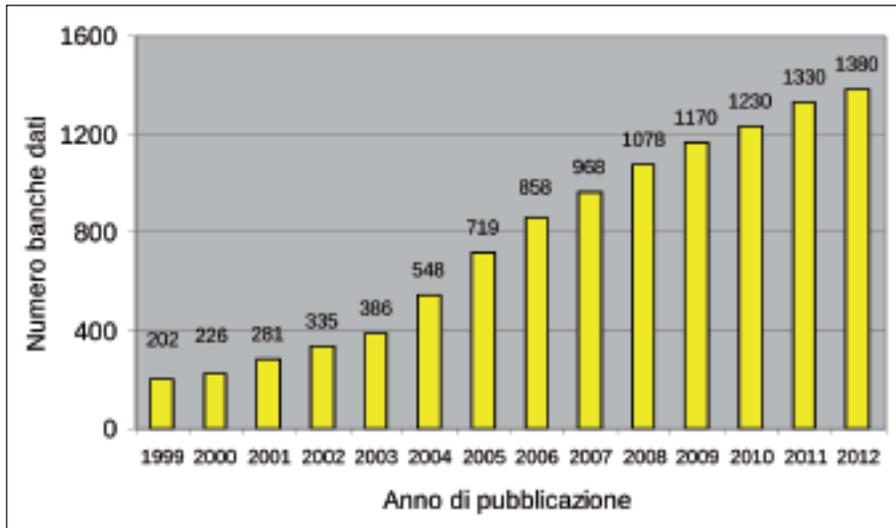


Fig. 2 - Istogramma raffigurante il numero di banche dati genomiche disponibili per anno, dal 1999 al 2012.

sibili. Il numero del 2012 riporta una lista di 1.380 banche dati, 50 in più del precedente anno (Fig. 2) (1).

Tipi di banche dati

Dato l'elevato e crescente numero di banche dati biomediche, in particolare genomiche biomolecolari, pubblicamente disponibili e contenenti informazioni in parte sovrapponibili, è utile raggruppare e classificare tali banche dati in gruppi omogenei in base alle loro principali caratteristiche, i tipi di dati che contengono, e la forma in cui questi sono memorizzati e resi successivamente disponibili.

La principale suddivisione che può essere fatta delle banche dati biomediche è di distinguerle in: banche dati genomiche e banche dati proteomiche in funzione del tipo di informazioni che esse contengono. Dipendendo poi dal tipo di dati contenuti, queste possono essere suddivise in:

- primarie,
- derivate o specializzate.

Le banche dati primarie contengono solo le sequenze e strutture note di acidi nucleici e aminoacidi, le informazioni indispensabili per descrivere e identificare tali sequenze, e le loro principali caratteristiche.

Le principali banche dati primarie di acidi nucleici sono:

- la banca europea European Molecular Biology Laboratory (EMBL) Data Library (<http://www.ebi.ac.uk/embl/>) presso l'European Bioinformatics Institute (EBI),
- la banca americana GenBank (<http://www.ncbi.nlm.nih.gov/Genbank/>) presso il National Center for Biotechnology Information (NCBI),

• la banca giapponese DNA Data Bank of Japan (DDBJ) (<http://www.ddbj.nig.ac.jp/>), Queste tre banche dati (EMBL, GenBank, e DDBJ) hanno definito un accordo (the International Nucleotide Sequence Database Collaboration - <http://www.ncbi.nlm.nih.gov/collab/>) volto a facilitare lo scambio e l'aggiornamento di dati biomolecolari in base ai seguenti due progetti: il Taxonomy Project per utilizzare una tassonomia unificata nelle tre banche dati, e il Feature Table per identificare univocamente le informazioni da associare ad ogni sequenza biomolecolare e per trovare un efficiente meccanismo di scambio giornaliero di dati tale che le informazioni fornite dalle tre banche dati siano le stesse e sempre allineate.

La banca dati primaria di aminoacidi è UniProt (Unified Protein Resource) (<http://www.pir.uniprot.org/>); è stata creata come integrazione non ridondante delle principali banche dati di aminoacidi: Swiss-Prot/TrEMBL (<http://www.expasy.org/sprot/>) e PIR (Protein Identification Resource) (<http://pir.georgetown.edu/>). UniProtKB/Swiss-Prot è la parte curata (supervisionata) e annotata manualmente della UniProt Knowledgebase (UniProtKB), mentre UniProtKB/TrEMBL è la parte annotata automaticamente e non supervisionata.

Le banche dati derivate o specializzate raccolgono informazioni molto utili provenienti dalle banche dati primarie, o loro complementari (quali dati tassonomici, biologici, fisiologici, funzionali), al fine di offrire informazioni di valore aggiunto e un quadro migliore del ruolo biologico della singola sequenza genomica. Tali banche dati raccolgono set di dati omogenei tassonomicamente e/o funzionalmente e possono essere:

- curate da personale specializzato (es. NCBI RefSeq <http://www.ncbi.nlm.nih.gov/RefSeq/>, Entrez Gene <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=gene>),
- ottenute computazionalmente in modo completamente automatico (UniGene <http://www.ncbi.nlm.nih.gov/UniGene/>),
- una combinazione dei due precedenti (NCBI Entrez Genome Project <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=genomeprj>).

Esistono diversi tipi di banche dati specializzate, ognuno dei quali può essere identificato da una delle seguenti definizioni:

- *un subset di dati derivati da una banca dati primaria*, omogeneo dal punto di vista biologico, accuratamente rivisto e arricchito di informazioni biologiche specifiche per il subset considerato. Un esempio è la banca dati PIR Sequence-Structure (PIR-NRL3D), una banca dati di proteine, derivata dalla banca dati Protein Information Resource (PIR), e delle relative strutture 3D note, le cui coordinate atomiche sono memorizzate nella banca dati Protein Data Bank (PDB) (<http://www.pdb.org/>),
- *un set di sequenze omologhe multiallineate*, come nella banca dati Ribosomal RNA Database (rRNA),
- *un set di informazioni complementari a quelle nelle banche dati primarie e specifiche per una classe di sequenze*, come l'Eukaryotic Promoter Databank (EPD) (<http://www.epd.isb-sib.ch/>),
- *banche dati genomiche propriamente dette*, rappresentative di un intero set di informazioni derivanti da progetti di mappatura e sequenziamento del Genoma Umano e di altri organismi selezionati come organismi modello. Un esempio è

il Genome Data Base (GDB) (<http://www.ncbi.nlm.nih.gov/sites/genome/>),

- *banche dati integrazionali*, più recentemente create per riunire informazioni disperse in diverse banche dati specializzate. Le più rappresentative sono le banche dati GeneCards (<http://bioinformatics.weizmann.ac.il/cards/>) e SOURCE (<http://source.stanford.edu/>).

Tra le banche dati specializzate, due in particolare forniscono informazioni cliniche:

- Online Mendelian Inheritance in Man - OMIM (<http://www.omim.org/>): è un completo e autorevole compendio di geni umani e fenotipi genetici che si concentra in particolare sul rapporto tra fenotipo e genotipo; contiene informazioni su tutte le patologie ereditarie note e su oltre 12.000 geni ad esse associati,
- Genetic Association Database - GAD (<http://geneticassociationdb.nih.gov/>): è un archivio di dati riassuntivi estratti da articoli scientifici relativi a studi di associazione genetica genome-wide (GWAS) di patologie umane complesse.

Accessibilità on-line delle banche dati e recupero di informazioni

Le banche dati biomediche sono accessibili tramite diverse modalità (4):

- accesso diretto: è raramente permesso per motivi di sicurezza,
- accesso tramite server FTP: consente di scaricare file in diversi formati con tutti o parte dei dati contenuti nella banca dati; richiede significative risorse tecnologiche e umane per reimplementare e interrogare localmente la banca dati,
- accesso tramite server web (pagine HTML o XML): questa è la modalità di accesso offerta più frequentemente. Generalmente fornisce informazioni non strutturate e interfacce web eterogenee, tramite le quali permette, generalmente, l'interrogazione di una singola sequenza biomolecolare per volta, i cui risultati sono per lo più forniti in formato HTML,
- accesso tramite servizi web: permette l'accesso programmatico (ovvero realizzato automaticamente da programmi informatici) alla banca dati, consentendo la creazione di workflow automatizzati di estrazione ed elaborazione dati.

Per ricercare adeguatamente nelle molte banche dati biomediche le informazioni di interesse in esse raccolte è innanzitutto necessario conoscere le banche dati disponibili, i dati forniti, e le modalità di accesso, in modo da poter scegliere la banca dati più adeguata per la ricerca che si vuole realizzare. Ad esempio, se si cercano specifiche informazioni su di una sequenza proteica, va individuata la banca dati di proteine che contenga le informazioni richieste, e che sia eventualmente specializzata su tali informazioni. Se per tale banca dati esiste un accesso via server web, come in generale accade, ci si può collegare alla home page della banca dati ed eseguire un'interrogazione per ricercare i dati di interesse. L'interrogazione viene eseguita mediante una o più parole chiave o identificativi che identifichino in modo il più possibile univoco le informazioni ricercate e le sequenze biomolecolari a cui si riferiscono. A seconda della banca dati, si possono usare diverse chiavi di ricerca. Le più comuni sono:

- *GenBank Accession number*, l'identificativo principale delle sequenze nucleotidiche,

- *nome o simbolo del gene o della proteina,*
- *Sequence entry ID,* ovvero il codice alfanumerico usato come identificativo della sequenza biomolecolare nella specifica banca dati (es. Entrez Gene ID),
- *termini controllati* che descrivono il coinvolgimento della sequenza biomolecolare in processi biologici, funzioni molecolari, pathway biochimici, patologie ereditarie note, ecc.,
- *nome dell'organismo* di appartenenza,
- *fonti bibliografiche* di riferimento.

Una volta identificata/e la/le sequenza/e biomolecolare/i a cui si riferiscono le informazioni di interesse ricercate, per ogni sequenza identificata è possibile visionare, all'interno di una pagina HTML, tutte le informazioni relative a quella sequenza contenute nella banca dati in cui si è realizzata la ricerca.

Necessità dell'efficace utilizzo on-line dei dati

Le diverse modalità disponibili di accesso alle varie banche dati biomediche spesso non sono immediatamente funzionali all'efficiente utilizzo delle informazioni fornite per le attuali necessità dei ricercatori medici e biologi di analisi di liste di geni e proteine. Infatti, in genere i ricercatori biomedici hanno bisogno di avere, in forma aggregata, i soli dati genomici-biomedici di interesse per lo specifico esperimento o caso in esame, in modo da poterli visualizzare, "sfogliare", e riorganizzare facilmente, e realizzare su di essi interrogazioni articolate per evidenziare eventuali informazioni nascoste e per cercare di rispondere a complesse domande biologiche che permettano di scoprire nuova conoscenza biomedica. Le principali difficoltà derivano dal fatto che i dati raccolti sono distribuiti all'interno di numerose banche dati eterogenee per schema, piattaforme software utilizzate, contenuti e formati di dati, e che generalmente possono essere interrogate pubblicamente solo via web, per una singola sequenza genomica alla volta, e i cui risultati d'interrogazione sono spesso disponibili solo in modo non strutturato all'interno di pagine HTML. Questo genera difficoltà soprattutto per il fine di valutare le informazioni di molte sequenze molecolari contemporaneamente, per esempio per interpretare risultati sperimentali ottenuti mediante tecnologie biomolecolari high-throughput di analisi di molte sequenze molecolari in parallelo. Per ovviare a tali difficoltà sono stati proposti vari approcci di integrazione dati, tra cui la costruzione di data warehouse integrazionali che riuniscono le informazioni più rilevanti fornite in modo sparso da diverse banche dati eterogenee. Il gruppo di Bioinformatica del Politecnico di Milano ha creato e mantiene uno di questi, il Genomic and Proteomic Knowledge Base - GPKB (<http://www.bioinformatics.dei.polimi.it/GPKB/>), che integra dati di sequenze biomolecolari e loro annotazioni strutturali, funzionali e fenotipiche fornite da 13 banche dati (Entrez Gene, Homologene, IPI, MINT, IntAct, ExPasy Enzyme, GO, GOA, BioCyc, KEGG, Reactome, eVOC and OMIM). Attualmente contiene dati di quasi 8.400.000 geni di più di 8.500 organismi, di quasi 51.000.000 di proteine di circa 372.500 specie e un totale di circa 103.100.000 annotazioni geniche (6.500.000 umane) e 183.200.000 proteiche (3.700.000 umane) espresse mediante 14 diver-

se terminologie biomediche controllate, tra cui quelle della Gene Ontology. Il GPKB costituisce quindi una risorsa unica estremamente importante per esplorare in modo integrato tali informazioni, che rappresentano l'attuale conoscenza biomedica-molecolare, e su cui applicare algoritmi di data mining che sfruttino tali informazioni integrate per aiutare a rispondere alle attuali complesse domande biomediche e a inferire nuova conoscenza.

Tools bioinformatici on-line per il recupero e l'analisi di informazioni

Per l'efficace utilizzo dei dati raccolti nelle banche dati biomediche-molecolari esistono e si stanno sviluppando vari strumenti bioinformatici implementati per eseguire ricerche ed estrarre dai dati immagazzinati l'informazione più significativa per successive analisi (5). Un tipo di questi tool sono i Database Search Tools, che possono essere suddivisi in:

- Sequence based tools: permettono di realizzare confronti tra una sequenza biomolecolare fornita dall'utilizzatore e ognuna delle sequenze registrate all'interno di una banca dati, al fine di trovare tutte le sequenze che hanno con la sequenza data una similarità superiore a una certa soglia. Si basano principalmente sui seguenti tre algoritmi:
 - *Smith-Waterman* (<http://www.ebi.ac.uk/Tools/fasta33/>)
 - *FASTA* (<http://www.ebi.ac.uk/fasta33/>)
 - *BLAST* (<http://www.ncbi.nlm.nih.gov/BLAST/>), che è divenuto l'algoritmo di ricerca di similarità di sequenze più utilizzato nelle banche dati di sequenze biomolecolari poiché offre i seguenti vantaggi:
 - . maggiore rapidità d'esecuzione,
 - . risultato fornito che include uno spettro di soluzioni,
 - . ogni matching (sequenza trovata simile alla sequenza fornita dall'utilizzatore) è accompagnato da una stima statistica della significatività della similarità trovata.
- Text-based tools: permettono di realizzare ricerche per parole chiave di annotazioni di sequenze biomolecolari presenti in numerose banche dati specializzate e forniscono all'utilizzatore liste di link a pagine HTML contenenti tutte le informazioni in una certa banca dati su di una sequenza che ha nella banca dati annotazioni contenenti le parole chiave ricercate. Esempi di tali strumenti sono: *SRS 6* (<http://srs6.ebi.ac.uk/>) e *DBGET/LinkDB*.

Seppur molto efficaci, questi strumenti permettono solo di "sfogliare" e "leggere" le informazioni trovate, selezionandole visivamente tra altre delle stesse sequenze. Non sono quindi sufficienti quando vi è la necessità di confrontare varie informazioni di molte diverse sequenze biomolecolari contemporaneamente, o a maggior ragione di diversi campioni biologici o pazienti, come avviene quando si vogliono interpretare i risultati di esperimenti e screening biomolecolari high-throughput, per esempio ottenuti con microarray, che forniscono liste numerose di geni o prodotti genici candidati coinvolti nei processi biologici analizzati. In tali casi le liste identificate devono essere arricchite con le rilevanti informazioni strutturali, funzionali e fenotipiche note di ogni elemento (gene o

prodotto genico) nella lista, per poter fare emergere significative caratteristiche comuni, o diverse, esistenti tra tali elementi. Per questo fine fondamentale è il ruolo delle annotazioni basate su vocabolari controllati o ontologie, che permettono di raggruppare i diversi geni o prodotti genici annotati in categorie funzionali che possono essere analizzate statisticamente (7). Principali esempi di tali strumenti che usano informazioni genomiche pubblicamente disponibili sono: DAVID (<http://david.abcc.ncifcrf.gov/>), Babelomics (<http://babelomics.bioinfo.cipf.es>) e GFINDER. Tra questi, *GFINDER* (Genome Function INtegrated Discoverer) è l'unico che permette di analizzare statisticamente, oltre a varie annotazioni genetiche funzionali espresse tramite vocabolari controllati, anche annotazioni fenotipiche (segni e sintomi) a supporto dell'interpretazione biologica di liste di geni (5, 6).

Bibliografia

1. Galperin MY, Fernández-Suárez XM. The 2012 Nucleic Acids Research Database Issue and the online Molecular Biology Database Collection. *Nucleic Acids Res.* 2012; 40 (Database issue): D1-D8.
2. Maffezzoli A, Masseroli M. Banche genomiche. In: Pinciroli F, Masseroli M, editors. *Elementi di Informatica BioMedica*. ISBN 88-7398-0171. Milano, IT: Polipres. 2005; 165-213.
3. Martin-Sanchez F, Iakovidis I, Norager S, Maojo V, de Groen P, Van der Lei J, Jones T, Apweiler R, Babic A. Synergy between medical informatics and bioinformatics: facilitating genomic medicine for future healthcare. *BIOIN-FOMED Study: prospective analysis of the relationships and synergy between Medical Informatics and Bioinformatics*. EC-IST 2001-35024. 2003.
4. Masseroli M, Stella A, Meani N, Alcalay M, Pinciroli F. MyWEST: My Web Extraction Software Tool for effective mining of annotations from web-based databanks. *Bioinformatics*. 2004a; 20 (18): 3326-3335.
5. Masseroli M, Martucci D, Pinciroli F. GFINDER: Genome Function INtegrated Discoverer through dynamic annotation, statistical analysis, and mining. *Nucleic Acids Res.* 2004b; 32 (Web Server issue): W293-W300.
6. Masseroli M, Galati O, Pinciroli F. GFINDER: genetic disease and phenotype location statistical analysis and mining of dynamically annotated gene lists. *Nucleic Acids Res.* 2005; 33 (Web Server issue): W717-W723.
7. Masseroli M, Pinciroli F. Using Gene Ontology and genomic controlled vocabularies to analyze high-throughput gene lists: three tool comparison. *Comput Biol Med.* 2006; 36 (7-8): 731-47.

Metodi di analisi statistica di dati di next generation sequencing

Claudia Angelini

Istituto per le Applicazioni del Calcolo, CNR, Napoli

Sommario

In questa lezione si fornirà una panoramica sui concetti alla base dell'analisi statistica e computazionale dei dati di sequenziamento massivo (più noti con il nome di Next Generation Sequencing). Sebbene i sistemi di sequenziamento massivo siano stati introdotti solo nel corso negli ultimi anni, essi stanno rivoluzionando il mondo della genomica. I vantaggi che tali strumenti offrono rispetto ai metodi tradizionali consistono nella enorme quantità di dati che vengono prodotti, nella loro alta qualità e riproducibilità, nei minori costi sperimentali, nonché nella elevata versatilità dei suddetti sistemi, i quali consentono di affrontare con uno stesso strumento molteplici quesiti scientifici. Di contro richiedono analisi molto più sofisticate. Nel corso della presente lezione ci occuperemo principalmente di applicazioni di ChIP-Seq ed RNA-Seq e loro analisi.

Introduzione

Fin dalla pubblicazione delle prime sequenze nucleotidiche è apparso chiaro che le tecnologie per il sequenziamento di acidi nucleici avrebbero rappresentato uno strumento fondamentale per la ricerca nel campo della genetica e della biologia molecolare. Negli ultimi 30 anni, l'automazione del sequenziamento ha consentito di compiere notevoli passi avanti verso un'approfondita conoscenza di interi genomi di diversi organismi, dai virus/batteri patogeni fino ai genomi complessi, incluso l'uomo. Sono iniziati così i grandi progetti di sequenziamento tra cui il ben noto progetto Genoma Umano, partito all'inizio degli anni '90 e completato soltanto nel 2003 dopo oltre 10 anni di intenso lavoro. Da allora ad oggi le tecnologie sono cambiate in modo rapidissimo grazie alle piattaforme per il sequenziamento massivo in parallelo introdotte intorno al 2005 (8, 11, 16) che hanno avuto un enorme sviluppo e diffusione. Tale introduzione ha segnato definitivamente l'inizio di quella che è tuttora definita l'era del Next-generation Sequencing (NGS) ed ha dato il via a molteplici nuove applicazioni (6). La possibilità di ottenere contemporaneamente sequenze di decine, poi centinaia di

milioni, adesso anche miliardi, di piccoli frammenti di DNA, con lunghezze fino a qualche centinaio paia di basi (bp), ha contribuito notevolmente all'ampliamento delle conoscenze dei genomi degli organismi viventi e dei meccanismi molecolari alla base della loro complessità. L'introduzione di tali tecnologie ha permesso inoltre di studiare ed interpretare la variabilità interspecie e inter-individuale a diversi livelli, dallo studio delle variazioni nucleotidiche a quello della regolazione dell'espressione genica, sia in condizioni fisiologiche che patologiche. Ed è proprio nell'ambito della cosiddetta ricerca di base ed in campo biomedico che le numerose possibilità d'indagine molecolare offerte da tali tecnologie si sono rivelate, e continueranno ad farlo in futuro, di grande utilità per lo studio di numerose malattie, incluso il cancro (1, 3, 15, 21). Tuttavia, la grandissima quantità di dati prodotti dalle piattaforme di NGS e le diverse applicazioni scientifiche che possono essere affrontate hanno anche lanciato nuove sfide alla comunità scientifica (5, 7, 10, 13). Infatti tali tecnologie richiedono un notevole sforzo nello sviluppo di nuovi strumenti bioinformatici, computazionali e statistici per la gestione e l'analisi dell'enorme mole di dati prodotti al fine di fornirne un'interpretazione che possa essere utile in chiave bio-medica.

In questa lezione saranno introdotte le principali caratteristiche delle piattaforme per il sequenziamento massivo disponibili e saranno descritte alcune delle sfide computazionali e bioinformatiche l'analisi dei dati prodotti con particolare riferimento ad esperimenti di ChIP-Seq ed RNA-Seq (14).

Sequenze ed algoritmi di allineamento ad un genoma di riferimento

Il risultato di un esperimento next generation sequencing consiste nella produzione di files contenenti alcune decine o centinaia di milioni (adesso anche miliardi) di brevi sequenze di nucleotidi (la cui lunghezza e le cui caratteristiche dipendono dal particolare strumento di sequenziamento utilizzato e dalla libreria preparata). Tali sequenze sono chiamate "short reads". Dal punto di vista l'analisi dei dati tali sequenze costituiscono le osservazioni dirette. Il primo passo di qualsiasi "Seq-analisi dei dati" consiste nell'allineamento o nell'assemblaggio di suddette sequenze in un segnale o, più in generale, in una forma di osservazione indiretta ma facilmente trattabile ed interpretabile da un punto di vista quantitativo sulla quale verrà successivamente effettuata l'inferenza statistica. Ad esempio, nel caso di studi su uomo, la procedura consiste nell'allineare le sequenze prodotte al cosiddetto genoma di riferimento (i.e., hg19). Questo rappresenta un passo cruciale, dato che tutti i risultati successivi verranno ottenuti sulla base delle sequenze allineate. A causa dell'enorme numero di sequenze e della loro limitata lunghezza, l'uso di algoritmi di allineamento tradizionali risulta impraticabile poiché richiederebbe centinaia o migliaia di ore di calcolo, di conseguenza, sono stati sviluppati nuovi algoritmi di allineamento (17) e si prevede che altri anche più efficienti saranno sviluppati a breve. Ogni algoritmo di allineamento è un equilibrio tra specificità e sensibilità, velocità di calcolo, memoria richiesta e flessibilità. Non esiste un algoritmo adatto per tutte le applicazioni. In generale a seconda del tipo di algoritmo utilizzato si determineranno tre sottoinsiemi di

sequenze, i.e., quelle che mappano univocamente sul genoma, quelle che a causa della ripetitività del genoma risultano avere più di una possibile collocazione e quelle che a causa di errori di lettura o a causa di forti variazioni strutturali non trovano una collocazione sul genoma di riferimento. La fase di allineamento risulta essere anche la più costosa da un punto di vista computazionale e spesso richiede l'utilizzo di strumenti di calcolo avanzati. Solo dopo aver effettuato la fase di allineamento l'analisi dati può essere condotta con modalità specifiche in base al tipo di esperimenti in esame.

RNA-Seq

Gli esperimenti di RNA-seq consentono di analizzare contemporaneamente l'intero trascrittoma. Recenti studi su larga scala del trascrittoma umano hanno rivelato un livello di complessità dell'espressione genica molto maggiore di quanto fosse prima immaginabile (9, 12, 19). Nell'uomo, a fronte di un relativamente esiguo numero complessivo di geni, attualmente stimato nell'ordine di 20,000-25,000, si è osservato che la complessità del trascrittoma, è almeno di un ordine di grandezza più elevata. L'utilizzo di siti alternativi per l'inizio e la terminazione della trascrizione, e lo splicing alternativo sono i meccanismi chiave alla base di questa significativa espansione del potenziale di espressione del genoma. La specificità dell'espressione genica nei diversi tipi cellulari e tessuti, nei processi di sviluppo e differenziamento o nelle condizioni patologiche è finemente modulata a livello trascrizionale e post-trascrizionale. Studi recenti hanno dimostrato che l'RNA-Seq fornisce una misura molto più precisa dei livelli trascrizionali e delle isoforme di splicing rispetto ad altri metodi, tra i quali il microarray. Tuttavia, modellare ed analizzare dati di RNA-Sequencing è più complesso dal punto di vista statistico (2, 19, 20, 14) e consiste in diversi aspetti tra cui:

1. stima del livello di espressione dei geni noti;
2. identificazione e quantificazione delle diverse isoforme;
3. individuazione di trascritti o parti regioni codificanti non ancora annotate;
4. individuazione di espressioni differenziali legate a condizioni diverse, come ad esempio patologie quali il cancro.

A partire delle prime metodologie proposte in (9, 12) sono state sviluppate tecniche statistiche e computazionali sempre più sofisticate ed il relativo software, come ad esempio (4, 18), tuttavia l'analisi necessita ancora di miglioramenti metodologici. Nel corso della lezione verrà fornita una breve panoramica riguardo gli approcci principali utilizzati per analizzare dati di RNA-seq.

ChIP-seq

Gli esperimenti di ChIP-seq consentono di sia individuare regioni di interazione tra proteine e DNA (i.e., siti di binding o fattori di trascrizione) sia di individuare regioni soggette a modifiche istoniche o altre alterazione della cromatina spesso associate a fenomeni patologici. Tali regioni di interazione possono essere classificate in diverse categorie: regioni che coprono da decine a poche centinaia

di coppie di basi di forma solitamente ben appuntita solitamente corrispondenti siti di interazione proteina-DNA; regioni sufficientemente localizzate ma più ampie e lunghe anche qualche kilobasi, regioni molto vaste e poco definite in forma che possono arrivare fino a diverse centinaia kilobasi. Le ultime due categorie sono solitamente corrispondenti a modifiche istoniche o a RNA polimerasi II. Dal punto di vista dell'analisi dei dati l'idea chiave è di identificare le regioni che si arricchiscono nel campione immunoprecipitato rispetto al controllo (DNA di ingresso) con significatività statistica. Pertanto, gli algoritmi (genericamente chiamati peaks callers) (13, 14) possono essere concettualmente suddivisi nelle seguenti fasi:

1. una stima del profilo del segnale lungo ciascun cromosoma;
2. una stima del rumore di fondo (background);
3. scelta di un criterio di chiamata di picco;
4. rimozione post-chiamate dei picchi artificiali;
5. annotazione dei picchi individuati.

Esistono molti metodi disponibili, soprattutto per quanto concerne l'individuazione picchi legati a fattori di trascrizioni e ogni software implementa un algoritmo diverso. Ogni algoritmo è stato costruito facendo diverse ipotesi sulla struttura stocastica dei dati osservati, il rumore di fondo. Pertanto, non tutti gli algoritmi sono adatti in tutti i casi, e spesso l'analisi richiede di combinare molti di loro al fine di ottenere le migliori prestazioni. A differenza delle tipiche interazioni fattore di trascrizione DNA, ampiamente studiate e per le quali la letteratura comprende già una discreta mole di approcci algoritmici e statistici alternativi (13, 14), lo studio di dati provenienti da ChIP-Seq per l'analisi di modificazioni istoniche e stato della cromatina non ha ancora prodotto applicazioni in grado di affrontare tutta l'ampia casistica che analisi di questo tipo prevedono. Nel corso della lezione verrà fornita una breve panoramica riguardo gli approcci principali utilizzati per analizzare dati di ChIP-seq.

Altre applicazioni e conclusioni

L'analisi di dati di ChIP-seq o RNA-seq rappresenta solo una parte delle tantissime applicazioni possibili con i sequenziatori massivi (6). In generale ogni applicazione richiede strumenti di analisi diversi ed altamente specifici. La scelta della particolare applicazione è legata al problema medico-biologico che si deve risolvere. Ad esempio, quando si studiano patologie come il cancro può assumere enorme importanza la caratterizzazione delle variazioni strutturali come SNPs, riarrangiamenti, variazione del numero di copie, così come lo studio dei microRNAs e degli altri piccoli RNA non codificanti. Per ciascuna di queste applicazioni si stanno sviluppando metodologie di analisi opportune. La risoluzione e l'ottimizzazione di ciascuno di questi problemi è l'argomento principale della ricerca bio-computazionale riguardo i sequenziatori massivi. Tuttavia il salto di qualità verso la reale medicina personalizzata consiste nell'integrazione tra i diversi livelli e le diverse tipologie di dati NGS, come ad esempio (5), l'integrazione tra dati di RNA-seq e di dati di ChIP-seq corrispondenti a più fattori e

più modifiche istoniche. Infatti, sebbene la maggior parte degli approcci attualmente disponibili consenta di analizzare un singolo tipo di esperimenti alla volta, l'integrazione di diversi insiemi e tipologie di dati sembra essere l'unica in grado di fornire una caratterizzazione dei meccanismi che regolano l'espressione genica anche a livello individuale.

Bibliografia

1. Baylin SB, Jones PA. A decade of exploring the cancer epigenome -biological and translational implications. *Nat Rev Cancer*. 2011.
2. Costa V, Angelini C, De Feis I, Ciccodicola A. Uncovering the complexity of transcriptomes with RNA-Seq. *J Biomed Biotechnol*. 2010
3. Ding L, Wendl MC, Koboldt DC, Mardis ER. Analysis of next-generation genomic data in cancer: accomplishments and challenges. *Hum Mol Genet*. 2010.
4. Garber M, Grabherr M, Guttman M, Trapnell C. Computational methods for transcriptome annotation and quantification using RNA-seq. *Nature Methods*. 2011.
5. Hawkins RD, Hon GC, Ren B. Next-generation genomics: an integrative approach. *Nat Rev Genet*. 2010.
6. Kahvejian A, Quackenbush J, Thompson JF. What would you do if you could sequence everything? *Nat Biotechnol*. 2008
7. Koboldt DC, Ding L, Mardis ER, Wilson RK. Challenges of sequencing human genomes. *Brief Bioinform*. 2010.
8. Mardis ER. Next-generation DNA sequencing methods. *Annu Rev Genomics Hum Genet*. 2008.
9. Marioni JC, Mason CE, Mane SM, Stephens M, Gilad Y. RNA-Seq: an assessment of technical reproducibility and comparison with gene expression arrays. *Genome Research*. 2008.
10. McPherson JD. Next generation gap. *Nature Methods*. 2009.
11. Metzker ML. Sequencing technologies - the next generation, *Nat Rev Genet*. 2010.
12. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods*. 2008.
13. Park PJ. ChIP-seq: advantages and challenges of a maturing technology. *Nat Rev Genet*. 2009.
14. Pepke S, Wold B, and Mortazavi A. Computation for ChIP-seq and RNA-Seq studies. *Nature Methods*. 2009.
15. Ross JS, Cronin M. Whole cancer genome sequencing by next-generation methods. *Am J Clin Pathol*. 2011.
16. Shendure J, Ji H. Next-generation DNA sequencing. *Nature Biotechnology*. 2008.
17. Trapnell C, Salzberg SL. How to map billions of short reads onto genomes. *Nature Biotechnology*. 2009.
18. Trapnell C, Williams BA, Pertea G, Mortazavi AM, Kwan G, van Baren MJ,

- Salzberg SL, Wold B, Pachter L. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature Biotechnology*, 2010
19. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics*, 2009.
 20. Wilhelm BT, Landry J.R. RNA - Seq quantitative measurement of expression through massively parallel RNA-sequencing. *Methods*. 2009.
 21. Wong KM, Hudson TJ, McPherson JD. Unraveling the genetics of cancer: genome sequencing and beyond. *Annu Rev Genomics Hum Genet*. 2011.

Il progetto ONCO-i2b2: integrazione di dati clinici e biobanche per il supporto alla ricerca in oncologia

Daniele Segagni¹, Valentina Tibollo¹, Arianna Dagliati³, Alberto Zambelli¹, Riccardo Bellazzi^{1,1}

¹IRCCS Fondazione S. Maugeri, Pavia;

²Dipartimento di Informatica e Sistemistica, Università di Pavia;

³Istituto di Studi Superiori, Pavia

Onco-i2b2 è un progetto finanziato dalla Regione Lombardia, che mira a sostenere la ricerca traslazionale in oncologia. Il progetto sfrutta le soluzioni software implementate dal centro di ricerca “Informatics for Integrating Biology and the Bedside” (i2b2), un’iniziativa finanziata dall’NIH Roadmap per l’informatica biomedica e guidata dal Partners HealthCare Center di Boston (1). L’obiettivo di i2b2 è di fornire un’infrastruttura software che permetta di utilizzare i dati raccolti durante la pratica clinica a scopo di ricerca, al fine di stabilire un percorso virtuoso di integrazione fra clinica e ricerca, e di accelerare il trasferimento dei risultati di quest’ultima alla pratica, con particolare riferimento all’ambito della medicina molecolare. A questo scopo, il progetto i2b2 ha sviluppato un data warehouse e una serie di applicativi software che si basano su un’architettura chiamata “alveare”. L’alveare ha come “celle” diversi programmi dedicati all’estrazione delle informazioni, alla loro manipolazione o all’analisi dei dati (2). Il software i2b2 è stato sviluppato in Java e sfrutta la tecnologia dei web-services per la comunicazione tra le celle.

Nell’ambito di Onco-i2b2, l’Università di Pavia e l’ospedale IRCCS Fondazione S. Maugeri (FSM) hanno integrato l’infrastruttura i2b2 con il Sistema Informativo Ospedaliero (SIO) della FSM e con la biobanca “Bruno Boerci”, che gestisce sia plasma sia tessuti tumorali. L’integrazione con il SIO fornisce l’accesso a tutte le cartelle cliniche elettroniche dei pazienti oncologici. La maggior parte dei dati raccolti nel SIO è contenuta in referti testuali. È stato quindi necessario sviluppare e integrare all’interno dell’architettura ICT un modulo di Natural Language Processing (NLP), al fine di estrarre informazioni importanti e i risultati dei test clinici, quali i referti istologici dei pazienti (3). La biobanca oncologica fornisce campioni provenienti da una raccolta di sangue e tessuti, ottenuta con il consenso informato di soggetti sani e pazienti oncologici.

Lo scopo di questa memoria è di descrivere i passaggi fondamentali del processo di integrazione e di presentare lo stato attuale del progetto Onco-i2b2.

Metodi

Il software Onco-i2b2 implementato presso l'ospedale FSM è stato progettato per integrare dati provenienti da fonti diverse e raccolti per scopi diversi, al fine di permettere ai ricercatori di interrogare e analizzare la grande quantità di informazioni provenienti dalla pratica clinica. Le principali fonti di dati che sono state integrate nel data warehouse i2b2 provengono dalla banca dati dell'unità di anatomia patologica, dal software di gestione della biobanca e dal SIO. La Figura 1 mostra le varie fasi del processo di integrazione. La fase 1 riguarda l'estrazione semi-automatizzata di dati dall'unità di anatomia patologica; la fase 2 prevede il processo di anonimizzazione dei campioni prima che questi vengano immagazzinati in biobanca. La fase 3, invece, si occupa di popolare il data warehouse i2b2 con dati provenienti dalla Biobanca. Infine, la fase 4 si occupa di aggiungere ad i2b2 ulteriori informazioni prelevate dal SIO.

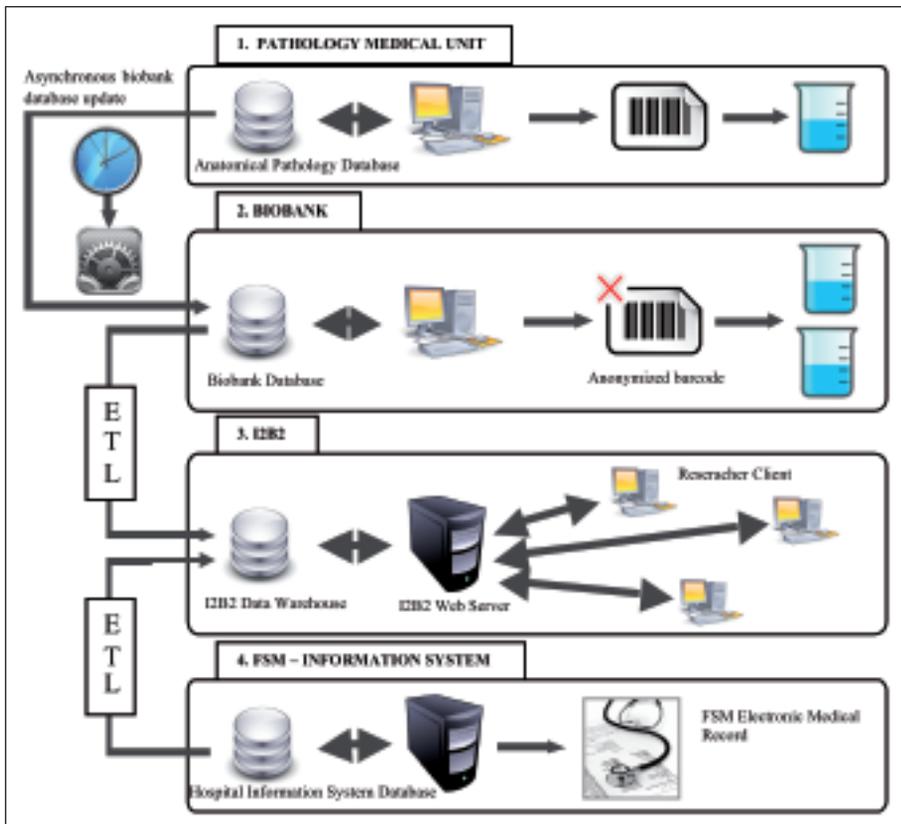


Fig. 1 - L'architettura ICT progettata per integrare le informazioni nel sistema Onco-i2b2.

Tab. 1 - Contenuto della Biobanca per unità operativa ospedaliera di origine e tipo. I dati della tabella si riferiscono al periodo 1/12/2010 - 31/01/2012.

Unità	Pazienti	Tessuto	Plasma-Sangue Intero
Senologia	430	1520	1290
Chirurgia	155	1125	465
Totale	585	2645	1755

In seguito si descriveranno con maggiore dettaglio alcuni aspetti del processo di integrazione.

Integrazione fra i dati della biobanca e dell'anatomia patologica

I dati associati ai campioni biologici memorizzati all'interno della Biobanca sono caricati in modo semi-automatico dal software di gestione dell'unità ospedaliera di anatomia patologica, al fine di diminuire il tempo di inserimento e di ridurre la possibilità di errore umano. Ogni campione è anonimizzato grazie alla creazione di un codice a barre bidimensionale che non include alcun riferimento diretto al paziente donatore. I tessuti tumorali o i campioni di plasma sono selezionati dai ricercatori e raccolti in nuove provette etichettate con il nuovo codice a barre. Solo alcuni utenti con privilegi speciali possono recuperare le informazioni riguardanti i donatori attraverso un software specializzato, che mostra anche il consenso informato del paziente. Il database della Biobanca viene periodicamente sincronizzato (più volte durante il giorno) in modo da garantire un costante aggiornamento.

I dati dei campioni biologici contenuti nella Biobanca sono quindi caricati nel data warehouse di i2b2 attraverso una serie di passi di Extract, Transform, Load (ETL) che comportano l'estrazione dei dati, la loro elaborazione e la mappatura nel data warehouse (4, 5). L'attività di ETL è svolta basandosi su KETTLE, un software sviluppato nell'ambito del progetto Pentaho (<http://www.pentaho.com/>). La Tabella 1 mostra il numero di pazienti e dei campioni biologici attualmente disponibili nella biobanca diviso per tipologia e unità mediche ospedaliere coinvolte.

L'integrazione dei dati SIO in i2b2 e il modulo di NLP

Le informazioni raccolte nel SIO della FSM sono estratte e inserite in i2b2 attraverso un complesso processo di ETL che esegue una corrispondenza fra i dati clinici e i "concetti" dell'ontologia di i2b2. I concetti saranno in seguito utilizzati dai ricercatori per interrogare il data warehouse e quindi consultare i dati e reperire i campioni dalla biobanca rilevanti per specifiche attività di ricerca. Alcuni dei concetti di Onco-i2b2 si riferiscono ai dati raccolti nei referti di anatomia patologica, che sono disponibili solo in formato testuale. Per questo motivo è stato sviluppato un modulo di NLP, che analizza i referti di anatomia patologica di ciascun paziente che abbia almeno un campione biologico memorizzato nella biobanca. Il modulo software NLP è stato configurato per elaborare testi medici scritti in italiano e inclusi nell'ontologia di dominio. In particolare, poiché il referto testuale contiene informazioni sui codici SNOMED dei campioni e altre infor-

mazioni numeriche, il modulo NLP, basato sul sistema GATE, è stato progettato per cercare concetti in questo limitato set di dati. Il sistema è stato validato internamente da parte degli esperti medici coinvolti nello studio su un sottoinsieme di 100 casi con il 100% di accuratezza.

L'architettura del sistema i2b2

Il data warehouse di i2b2, detto Clinical Research Chart (CRC), è stato progettato per gestire dati provenienti da studi clinici, cartelle cliniche e sistemi di laboratorio, insieme a molte altre informazioni contenute in fonti eterogenee (6). Il CRC memorizza questi dati in tre tabelle: “pazienti”, “visite” e “osservazioni”. Le tre tabelle dei dati, insieme a due tabelle di ricerca (“concetti” e “provider”), sono i componenti principali del cosiddetto “star-schema” del data warehouse. L'aspetto più importante della costruzione di uno star-schema è l'identificazione dei “fatti”. In sanità, un fatto logico è un'osservazione su un paziente. Le tabelle delle dimensioni contengono ulteriori informazioni descrittive e analitiche sugli attributi nella tabella dei fatti. Una tabella “dimensione” può contenere informazioni su come i dati sono organizzati, come ad esempio una gerarchia che può essere usata per categorizzare o riepilogare i dati.

L'infrastruttura i2b2 installata presso FSM fornisce un accesso via web al sistema descritto nei paragrafi precedenti. Le informazioni estratte possono essere analizzate attraverso il client web i2b2 anche grazie a dei plug-in dedicati e appositamente configurati (7,8).

Risultati e discussione

Il sistema Onco-i2b2 è operativo alla FSM dal Dicembre 2010. Attualmente, l'installazione di i2b2 presso la FSM consiste in 6713 pazienti (585 di loro hanno almeno un campione biologico nella biobanca), 47140 visite, 960 concetti (26 di loro correlate a diagnosi, 24 correlate a misurazioni cliniche e 31 correlate a rapporti istologici) e 106930 osservazioni.

Al fine di migliorare costantemente la facilità d'uso di i2b2 per i ricercatori dell'ospedale, abbiamo aggiunto al web client di i2b2 alcuni plug-in per l'esportazione dei dati e per l'esplorazione dei fenotipi (9). L'architettura IT realizzata presso la FSM è un esempio concreto di come l'integrazione tra le diverse informazioni provenienti da fonti eterogenee possa essere attuata in modo efficace e possa quindi mettere a disposizione questi dati a supporto della ricerca scientifica.

Per sfruttare pienamente il potenziale dell'architettura IT, i prossimi passi del progetto prevedono l'estensione del set di dati importati dal SIO, nonché la gestione dei dati provenienti dal laboratorio di genomica. Abbiamo anche in programma di continuare a estendere le funzionalità dell'architettura di i2b2 mediante la realizzazione di nuovi plug-in dedicati all'analisi dei dati (10): in particolare, stiamo lavorando su un'estensione del motore di query i2b2 con l'aggiunta di funzionalità di query temporali. Infine, un altro importante aspetto che considereremo per lo sviluppo futuro del sistema riguarderà l'integrazione dei dati di genotipo del

paziente, che richiede un'attenta valutazione sia in termini di rappresentazione e memorizzazione dei dati sia in termini di gestione della sicurezza e della privacy.

Bibliografia

1. Murphy SN, Mendis M, Hackett K, et al, Architecture of the open-source clinical research chart from Informatics for Integrating Biology and the Bedside, *AMIA AnnuSymp Proc.* 2007; 548-52.
2. Murphy SN, Weber G, Mendis M, et al. Serving the enterprise and beyond with informatics for integrating biology and the bedside (i2b2), *J Am Med Inform Assoc.* 2010; 124-30.
3. Jurafsky D, Martin JH. *Speech and language processing, an introduction to natural language processing, computational linguistics, and speech recognition.* Second Edition, Prentice Hall. 2008.
4. Bouman R, van Dongen J. *Pentaho Solutions*, Wiley, 2009.
5. Kimball R, Ross M, Thornthwaite W, Mundy J, Becker B. *The data warehouse lifecycle toolkit.*
6. Partners Health Care Systems, i2b2 software (v.1.5) documentation, 2008.
7. Mendis M, Wattanasin N, Kuttan R, et al. Integration of Hive and cell software in the i2b2 architecture. *AMIA AnnuSymp Proc.* 2007; 1048.
8. Murphy SN, Churchill S, Bry L, et al. Instrumenting the health care enterprise for discovery research in the genomic era. *Genome Res.* 2009; 1675-81.
9. Bellazzi R, Segagni D, et al. R engine cell: integrating R into the i2b2 software infrastructure. *J Am Med Inform Assoc.* 2011.
10. Daniel WW. *Biostatistics: a foundation for analysis in the health sciences.* John Wiley & Sons Inc. 2005.

**ELEMENTI
DI FARMACOGENOMICA**

Eredità e risposta ai farmaci

Maurizio D'Incalci

Dipartimento di Oncologia, Istituto di Ricerche Farmacologiche "Mario Negri", Milano

La familiarità di alcuni tumori è ben dimostrata da moltissimi anni. Negli ultimi 20 anni sono stati fatti passi avanti notevoli nella definizione più precisa dei meccanismi molecolari che sono alla base della predisposizione genetica all'insorgenza di alcuni tumori.

La domanda a cui si cercherà di rispondere nella lezione è se alcune di queste alterazioni siano anche importanti nel conferire una particolare sensibilità a determinate terapie nel caso si sviluppi un tumore

Nel caso del tumore della mammella e dell'ovaio le mutazioni di BRCA1 e di BRCA2 sono principalmente responsabili della predisposizione genetica ed in questo caso ci sono molte evidenze che vi sia una particolare sensibilità ad alcune terapie specifiche.

Cellule con mutazione di BRCA1 o di BRCA2 hanno un difetto nel meccanismo di ricombinazione omologa che è essenziale per la riparazione di rotture del doppio filamento del DNA. È un meccanismo di riparazione che avviene dopo la fase S del ciclo cellulare, prima della mitosi ed è "error free", cioè non porta nel processo della riparazione ad altri possibili errori. Anche se le mutazioni di queste due proteine costituisce il principale meccanismo che è alla base della predisposizione genetica del tumore della mammella e dell'ovaio si è notato che per alcuni tumori anche in assenza di mutazioni non vi sia espressione di queste proteine. Questa mancanza di espressione è dovuta a meccanismi epigenetici che portano all'inibizione della trascrizione dei geni che codificano per queste proteine. Si parla di fenotipo "braceness" per definire complessivamente i casi in cui vi sia un deficit funzionale di BRCA1 o di BRCA2.

Si discuterà in dettaglio l'importanza del meccanismo di ricombinazione omologa per la riparazione dei danni al DNA indotti da agenti fisici e chimici che danneggiano il DNA. Queste conoscenze hanno avuto una ricaduta importante per la terapia delle pazienti che mostrano il fenotipo "braceness". Ad esempio in casi di carcinomi della mammella con fenotipo "braceness" il cisplatino, che normalmente non viene utilizzato per la terapia del carcinoma della mammella, è un farmaco attivo.

Studi preclinici, ampiamente confermati in clinica, hanno mostrato come la deficienza di BRCA1 o BRCA2 porti le cellule tumorali ad una selettiva sensibilità

ad inibitori dell'enzima Poli-ADP-ribosio-polimerasi (PARP), anche utilizzati da soli. Si descriverà questa classe di composti e le loro caratteristiche farmacologiche discutendo nel dettaglio il meccanismo che porta a questa selettiva citotossicità e le applicazioni cliniche che sono state sviluppate. In studi clinici gli inibitori di PARP si sono dimostrati attivi e privi di tossicità apprezzabili in pazienti con fenotipo "bracanes". Si sono anche dimostrati meccanismi di resistenza a questi farmaci, che saranno discussi. Il tentativo di combinare questi inibitori con altri farmaci antitumorali è in corso, ma i risultati ottenuti fino ad ora non sono promettenti in quanto gli inibitori di PARP aumentano fortemente la tossicità di altri antitumorali, soprattutto quelli che danneggiano il DNA, presumibilmente per la loro capacità di inibire i meccanismi di riparazione del DNA che vengono attivati in seguito ad un danno.

Genoma del paziente e tolleranza farmacologica

Diego Fornasari

Dipartimento di Farmacologia, Chemioterapia e Tossicologia Medica, Università di Milano

È ben noto che pazienti diversi possano rispondere in maniera diversa ad un medesimo farmaco, generando uno spettro continuo di risposte cliniche in cui gli effetti terapeutici e le reazioni avverse sono diversi e variamente combinati nei diversi individui. In passato questa variabilità inter-individuale veniva principalmente spiegata alla luce di fattori non genetici quali l'età, lo stato della funzionalità renale ed epatica, lo stile di vita, con particolare riferimento alla dieta e all'abuso di alcool e fumo, o la concomitante assunzione di altri farmaci. Sebbene tutti questi fattori possano certamente essere coinvolti ed in specifiche circostanze giustificare da soli la diversa risposta clinica al medesimo trattamento farmacologico, attualmente si ritiene che la risposta individuale ai farmaci abbia sempre una componente genetica, la quale rende conto, in alcune circostanze, fino al 95% della variabilità nelle risposte cliniche ad un determinato farmaco.

La variabilità individuale nella risposta al trattamento farmacologico costituisce uno dei problemi più rilevanti nella pratica clinica essendo responsabile di fallimenti terapeutici e di reazioni avverse che a loro volta, oltre a danneggiare gravemente il paziente, causano prolungati ricoveri ed aumentati costi sanitari.

La farmacogenetica è la disciplina che studia come l'assetto genetico degli individui influenzi l'azione dei farmaci ad essi somministrati, con l'obiettivo ultimo di predire e quindi prevenire reazioni avverse e fallimenti terapeutici, ma anche di favorire l'identificazione del farmaco giusto e del suo corretto dosaggio per ogni singolo paziente.

I diversi progetti di sequenziamento del genoma umano hanno dimostrato che la gran parte dei geni umani contiene differenze di sequenza tra individui, note come polimorfismi, che causano l'esistenza di forme alternative di un determinato gene, note come alleli. I polimorfismi possono essere di diverso tipo, ma i più numerosi sono i polimorfismi a singolo nucleotide, che nella letteratura scientifica internazionale vengono identificati con la sigla SNPs (single nucleotide polymorphisms).

Gli SNP possono influenzare la quantità o la qualità del prodotto di un determinato gene, cioè della proteina da esso codificata. Per esempio uno SNP nella regione codificante di un gene può causare la sostituzione di un aminoacido nella

relativa proteina. Se la proteina fosse un enzima responsabile del metabolismo epatico di alcuni farmaci, potrebbe metabolizzarli meglio o peggio, se fosse un recettore, potrebbe legare meglio o peggio un determinato farmaco..

I geni che determinano la risposta ai farmaci possono essere distinti in due grandi classi: i geni codificanti per il bersaglio terapeutico primario, come per esempio recettori, canali ionici o enzimi e i geni codificanti per proteine coinvolte nell'assorbimento, distribuzione, metabolismo ed escrezione del farmaco. I polimorfismi a carico dei diversi membri della famiglia del citocromo P450 (CYP450), responsabili del metabolismo di numerosi farmaci, sono quelli maggiormente studiati sia a livello molecolare, valutandone gli effetti sulla funzionalità enzimatica, sia a livello clinico, valutandone l'impatto sulla risposta terapeutica.

Per esempio è stata dimostrata l'esistenza di varianti alleliche del *CYP2C9* (note come *CYP2C9*2* e *CYP2C9*3*) codificanti per enzimi con ridotta funzionalità. In individui portatori di questi varianti, ed in particolare modo negli omozigoti, la somministrazione di dosi "standard" di warfarin è più frequentemente accompagnata da emorragie, soprattutto del tratto-gastroenterico. La frequenza degli alleli *CYP2C9*2* e *CYP2C9*3* è rispettivamente del 8-13% e 6-9% tra i Caucasicci, e questo ha indotto le autorità regolatorie europea e americana a chiedere la segnalazione in scheda tecnica del ruolo di queste varianti nella risposta terapeutica al warfarin e ad approvare l'uso del test genetico per identificare preventivamente i portatori di queste varianti alleliche.

In alcuni casi i farmaci sono introdotti nell'organismo come pro-farmaci, cioè come molecole inattive che necessitano di essere metabolizzate in forma attiva proprio da alcuni degli enzimi della famiglia del CYP450. Nel caso dei pro-farmaci le conseguenze terapeutiche dei polimorfismi sono opposte a quelle osservate con i farmaci. Il clopidogrel è un profarmaco che dopo somministrazione orale viene rapidamente inattivato all'85% a livello epatico per l'azione di esterasi. Il restante 15% deve subire due fasi di reazioni di attivazione dove giocano un ruolo fondamentale diversi CYP, con particolare riferimento al CYP2C19. Numerosi studi clinici hanno dimostrato che i portatori di varianti non funzionanti di questo enzima possiedono una ridotta concentrazione plasmatica di metabolita attivo, hanno una maggiore reattività piastrinica e vanno incontro più frequentemente ed in maniera statisticamente significativa ad eventi cardio-vascolari, quali la morte, l'infarto e l'ictus.

In ambito oncologico, la situazione è più complessa in quanto accanto ai polimorfismi del genoma del paziente, si aggiungono le mutazioni del tumore che possono influenzare, positivamente o negativamente, la risposta terapeutica. Per tale motivo, in ambito oncologico, si preferisce distinguere una farmacogenetica germinale, che riguarda il genoma del paziente, da una farmacogenetica somatica, che riguarda il genoma del tumore.

Farmacogenetica germinale: Tiopurine e TPMT

La 6-mercaptopurina (6-MP) è un analogo purinico impiegato nel trattamento di tumori linfoidi e come immunosoppressore in alcune malattie auto-immuni. Il

farmaco è il principale agente impiegato nella terapia di mantenimento della leucemia linfoblastica acuta del bambino.

La 6-MP è convertita enzimaticamente in nucleotidi tioguanilici che vengono incorporati nel DNA e interferiscono con la sua replicazione. Inoltre i metaboliti metilati della 6-MP inibiscono la sintesi *de novo* delle purine, contribuendo agli effetti citotossici osservati.

Nei tessuti linfoidi la 6-MP è principalmente inattivata dall'enzima tiopurina-metiltrasferasi (TPMT), poiché l'attività dell'altro enzima coinvolto nel metabolismo delle purine, la xantina ossidasi, ha un'attività molto bassa in queste cellule. La TPMT è codificata da un gene che presenta varianti alleliche inattive le quali producono proteine con attività enzimatica scarsa o assente. Si stima che nel mondo gli omozigoti per le varianti inattive oscillino da 1 su 180 a 1 su 3.700 individui, a seconda del gruppo etnico di appartenenza; gli eterozigoti rappresentano dal 3 al 14%.

Ai regimi terapeutici convenzionali di 6-MP, gli omozigoti sviluppano una tossicità, potenzialmente mortale, con una penetranza del 100% a causa dell'accumulo di tioguanine; il 30-60% degli eterozigoti non sopporta la dose piena di 6-MP per gli stessi motivi. L'identificazione delle varianti alleliche inattive mediante genotipizzazione "mirata" del paziente prima dell'inizio della terapia consente di ridurre preventivamente la dose di 6-MP, mentre il dosaggio di tutti gli altri farmaci rimane invariata. È stato dimostrato che in tutti i protocolli di trattamento della leucemia linfoblastica acuta, l'aggiustamento della dose di 6-MP sulla base dell'assetto genetico del paziente consente di ottenere nei soggetti privi o parzialmente privi di TPMT risultati terapeutici comparabili a quelli ottenuti nei soggetti "normali", con la medesima frequenza e intensità di effetti avversi (Paugh, et al., 2010).

La genotipizzazione della TPMT è stata raccomandata da FDA prima dell'inizio di qualsiasi trattamento con analoghi purinici.

Farmacogenetica somatica: KRAS e gli inibitori chinasi nei tumori del retto-colon

Cetuximab e panitumumab sono anticorpi monoclonali umanizzati diretti verso il recettore dell'EGF e sono impiegati nel trattamento del tumore metastatico del retto-colon, in monoterapia o in associazione. KRAS è coinvolto nella trasduzione del segnale a valle di diversi recettori per fattori di crescita; studi clinici hanno dimostrato che pazienti con mutazioni di KRAS non rispondono ai 2 monoclonali, probabilmente perché la proliferazione del tumore diventa indipendente dall'attivazione del EGFR.

L'ASCO, la Società americana di oncologia clinica, ha raccomandato che la ricerca delle mutazioni nel gene codificante per EGFR venga sempre eseguita prima dell'inizio del trattamento con anticorpi anti-EGFR e che pazienti positivi per le mutazioni non siano trattati con questi monoclonali (Allegra et al., 2009). La raccomandazione dell'ASCO è stata successivamente recepita dalle autorità regolatorie di Europa e Stati Uniti.

Conclusioni

I progressi della farmacogenetica stanno lentamente modificando i criteri di appropriatezza terapeutica, dimostrando, accanto alle altre variabili note, il ruolo fondamentale dell'assetto genotipico del paziente. I vantaggi delle applicazioni cliniche della farmacogenetica sono evidenti: una migliore qualità dell'intervento terapeutico, con una più razionale scelta della dose, un uso più sicuro dei farmaci da parte del Medico, che potrà richiedere la genotipizzazione del paziente in relazione a determinati trattamenti terapeutici così come oggi richiede i comuni esami del sangue, una più razionale e motivata ricerca per sviluppare strategie terapeutiche alternative al fine di curare coloro i quali non possano assumere uno specifico principio attivo, un risparmio di denaro dovuto all'eliminazione di terapie inefficaci e di ricoveri causati dagli effetti avversi dei farmaci.

Bibliografia

1. Allegra CJ, et al. ASCO provisional clinical opinion: testing for KRAS gene mutations in patients with metastatic colorectal carcinoma to predict response to anti-EGFR monoclonal antibody therapy. *J Clin Oncol.* 2009; 27: 2091-96.
2. Crettol S, Petrovic N, Murray M. Pharmacogenetics of phase I and phase II drug metabolism *Curr Pharm Des.* 2010; 16: 204-19.
3. Paugh SW, Stocco G, Evans WE. Pharmacogenomics in pediatric leukemia. *Curr Opin Pediatr.* 2010; 22: 703-10.
4. Wang L, McLeod HL, Weinshilboum RM. Genomics and drug response. *N Engl J Med.* 2011; 364: 1144-53.

Pharmacogenetics in cancer care: transferring translational research into clinical practice

Giuseppe Toffoli

Direttore U.O. Complessa "Farmacologia Sperimentale e Clinica",
Dipartimento di Oncologia Molecolare e Ricerca traslazionale,
Centro di Riferimento Oncologico, Istituto Nazionale Tumori, Aviano (PN)

Antitumoral drugs are characterized by a low therapeutic index, with a modest difference between toxic and efficient dose: this could determine in some individuals a number of toxic effects, to which a real benefit in terms of antitumoral activity does not always correspond. These compounds are also characterized by a sometimes very evident interindividual variability, and the same drug can determine in different individuals a different toxic effect or a different antitumoral activity. It is then important to personalize the antitumoral activity, that is to give each patient the right drug at the right dose. Therapy personalization represents one of the most important future challenges in the therapy of oncological patients. On this ground it is crucial to identify the biomolecular specificities of the neoplastic cells towards which the antitumoral drugs must be vehiculated, or to identify the genetic specificities of the patient that can explain the reason for the differential effect of antitumoral drugs. Pharmacogenetics, the study of the genetic basis of the interindividual differences in response to drugs results from the intersection between genetics and pharmacology is a notably interesting field for the possible implications in the clinical practice. Genetic polymorphisms, that is the structural alterations of genes involved in the metabolism, transport or drug interaction with the cellular target, have been described both for the drugs traditionally employed in the oncological patient, and for the so-called targeted molecules recently employed in clinic. Pharmacogenetics investigates the individual genetic specificities which are relatively common among the population and that are involved in antitumoral drugs action. The most common form of genetic polymorphism is represented by the SNPs (single nucleotide polymorphisms). Other forms of polymorphisms are represented by abnormal nucleotide sequence repetitions (microsatellites or midisatellites), genomic duplications, pseudogenes etc. At present, more than 37 million SNPs of human genome have been described: it is then important to individuate those which have an effective impact on drug

action, modulating their pharmacogenetics and pharmacogenomics (toxicity and response) and that can influence the ultimate result of the antitumoral therapy, influencing the patients survival. Sometimes the pharmacologic effect cannot be explained by the action of a single polymorphism. The complex metabolic or intercellular transport ways that often characterize antitumoral drugs require the involvement of more genes and only the precise definition of polymorphisms of all these genes results of some utility for prognosis. In the study of interactions between genome and drug, we can aim at a single polymorphism, applying in this case a pharmacogenetic approach or we can analyze the combined effect of more polymorphisms, employing in that case a pharmacogenomic approach. Basically, the strategies adopted at present in pharmacogenetics/omic studies are three: "candidate gene approach", "gene pathway approach" and "genome wide analyses". In the so-called "candidate gene approach" the polymorphisms of a single gene considered critical for the action of the drug are analyzed. This strategy presupposes that the action of the gene in question is actually important for the pharmacologic action and that the polymorphism or the polymorphisms of interest have an effective prognostic power in the clinical practice. It is also important to establish what is the combined effect of the polymorphisms present in a single gene and the modality of their distribution, that is if the contemporary presence of more polymorphisms in a single gene is casual or not (linkage disequilibrium). At present, the kits recommended by Food and Drug Administration (FDA) in the oncological field for single polymorphisms are relatively few (UGT1A1*28 (irinotecan) (invader assay) and TPMT (6-MP, azatioprine) (pro-Predict Py), nevertheless, the data recently produced in literature are encouraging for further clinical development.

In the so-called "gene pathway approach", the polymorphism of the genes involved in the complex metabolic ways of antitumoral drugs are analyzed. The rationale of this approach is that the pharmacologic effect is a multifactorial event and that the pleiotropic effect could depend on polymorphisms located in different genes. In this context, it is important to define for each antitumoral drug all polymorphisms with prognostic potentiality, and the attention of pharmaceutical companies is to design pharmacogenetic kits including all polymorphisms that are relevant for the action of a specific drug.

The "genome wide analysis" represents an innovative strategy in the field of pharmacogenomics obtained by the recent advances in the technological field. With this approach, the entire human genome is analyzed, in an attempt to define clusters of polymorphisms predictive of the pharmacological effect. At present, this strategy is employed principally in the research field, to individuate those polymorphisms whose real meaning will be successively validated, but the possibility to adopt this strategy in the clinical practice represents a stimulating perspective, analogous to what is being done with the microarrays of gene expression in human tumors.

Gene polymorphisms that could influence the biochemistry of the most common tumoral drugs have been described. Among the polymorphisms of enzymes involved in the metabolism of phase I of antitumoral compounds, the attention is

pointed at the polymorphic variants of cytochrome isoforms, in particular 3A4 and 3A5 (irinotecan, taxanes, alkilant agents). For what concerns phase II enzymes, we particularly focus on the polymorphisms in uridindifosfogluconosiltransferasi and in particular on UGT1A1*28 polymorphism in the gene that inactivates SN38, the active principle of irinotecan. We have recently published (Toffoli et al., *J. Clin. Oncol.* 2010) a phase I pharmacogenetic study to define the maximum tolerated dose of irinotecan associated with FOLFIRI regimen based on patients genetic characteristics. The study has demonstrated that the patients carrying the wild type genotype (UGT1A1) can tolerate higher drug dose with respect to patients carrying UGT1A1*28 polymorphism. We observed an increased response rate in patients receiving the higher dose of irinotecan.

Other polymorphic variants of phase II enzymes concern the isoforms of glutathione S-transferase gene (GST). Numerous polymorphic variants of these genes that include allelic deletion or point mutations have been described. Recently, it was found how polymorphic variants of isoforms of GST gene can influence toxicity, and in particular neurotoxicity to chemotherapeutic regimens with oxaliplatin.

Numerous polymorphisms have been described in genes involved in the action of fluoropyrimidines (thymidylate sintetasi, metilenetetrahydrofolate reductase and dihydropyrimidine dehydrogenase). In particular, IVS 14+1G>A variant dihydropyrimidine dehydrogenase can be associated with severe toxicity, even lethal, by fluorouracil. This polymorphism is extremely rare in our population (<1%) but it should be attentively evaluated before starting a therapy with fluoropyrimidines.

Finally, polymorphic variants have been described for the genes involved in the transmembrane transport of antitumoral drugs and in particular in MDR1 gene that codes for Pgp and for other "multidrug resistant proteins" MRPs. The real impact of these polymorphisms on toxicity and tumor response as well as on drug pharmacokinetics of these transporters needs further investigation.

In conclusion, pharmacogenetic/omics represents a potential instrument for personalizing cancer therapy. This approach seems extremely interesting to reduce toxic effects of antitumoral drugs and for increasing their efficacy. Nevertheless, it is still necessary to validate this innovative perspective, both in analytical terms (sensitivity, specificity and reproducibility of tests) and for what concerns the precise relations among genotype, pharmacokinetics and pharmacodynamics (toxicity and response). Finally, it is crucial to define the clinical utility of this approach. In particular, if based on a pharmacogenetic test the therapeutic regimen could be modified, what pharmacologic alternatives can be used and especially what will be the effect of these modifications in terms of toxicity and response to therapy.

La ricerca genetica per programmare lo sviluppo di nuovi farmaci

Giuseppe Recchia

Direttore Medico GlaxoSmithKline Spa, Verona

Le attese di pazienti, medici e società civile per nuovi interventi in grado di risolvere bisogni di salute senza soluzione terapeutica o solo in parte soddisfatti sono in costante aumento. Farmaci e vaccini rappresentano sia beni per la salute che prodotti industriali e devono pertanto soddisfare i requisiti che la società pone per garantire la propria salute ed il proprio sviluppo sociale ed economico. Il processo di ricerca e sviluppo (*Research & Development - R&D*) di un nuovo farmaco è caratterizzato da un livello di rischio di tipo imprenditoriale assai superiore rispetto a quello della maggior parte degli altri prodotti industriali e per tale ragione viene condotto in via pressoché esclusiva all'impresa del farmaco, che si assume in tale modo responsabilità nei confronti sia della società civile che dei propri azionisti.

L'obiettivo principale della R&D farmaceutica è l'innovazione della terapia attraverso la scoperta e lo sviluppo di nuovi composti in grado di assicurare un beneficio incrementale per il paziente e/o la sanità. In termini quantitativi, da oltre un decennio il numero dei nuovi composti resi disponibili dalla R&D per la terapia si è progressivamente ridotto fino a raggiungere in questi ultimi anni livelli giudicati allarmanti dalla *Food and Drug Administration* americane e dalla Commissione Europea, che hanno avviato rispettivamente il programma *From Stagnation to Innovation* e *Innovative Medicines Initiatives - IMI* per stimolare e promuovere la innovazione farmaceutica.

L'applicazione al processo di R&D del farmaco delle conoscenze acquisite dalla ricerca sul genoma umano rappresenta la maggiore tra le iniziative per la promozione della innovazione farmaceutica ed una area di rilevante investimento per le imprese.

I motivi di tale interesse sono da una parte rendere più efficace il processo di scoperta di nuovi farmaci, mediante l'identificazione di nuovi bersagli biologici (“*il bersaglio più appropriato per ciascuna malattia*”) e dall'altra migliorare l'efficacia e la tollerabilità del farmaco attraverso la personalizzazione del trattamento preventivo o terapeutico sulla base delle caratteristiche genetiche del paziente (“*il farmaco più appropriato per ciascun paziente*”).

Con il termine di *Medicina Personalizzata* s'intende oggi il modello di medicina caratterizzato dalla possibilità di identificare la suscettibilità di una singola persona alle malattie comuni, di misurarne il livello di rischio, di personalizzare la terapia in base alla costituzione genetica del paziente e di offrire nuove opzioni terapeutiche basate sulla interazione dei farmaci con nuovi bersagli molecolari al fine di assicurare il miglior esito possibile in termini di salute.

Il raggiungimento di tali obiettivi non dipende solamente dallo sviluppo della conoscenza scientifica, ma anche dalla appropriata gestione delle implicazioni etiche, legali e sociali associate alla ricerca ed alla applicazione delle nuove tecnologie.

Agli inizi del nuovo millennio, la pubblicazione dei risultati del progetto *Human Genome*, aveva alimentato aspettative per una rivoluzione genomica le cui applicazioni avrebbero potuto modificare la pratica della medicina già nel corso dei successivi 10 anni. Ad una prima analisi i risultati finora ottenuti sembrerebbero avere almeno in parte disatteso tali aspettative e generato una certa delusione nella pubblica opinione e negli stessi operatori della sanità. Diviene pertanto necessario considerare in modo critico lo stato dell'arte della conoscenza e condurre una riflessione sulle effettive potenzialità e della genomica e della genetica in relazione alla scoperta, allo sviluppo ed all'utilizzo di nuove tecnologie sanitarie.

Identificazione di nuovi bersagli biologici e scoperta di nuovi farmaci

L'innovazione della terapia dipende principalmente dalla comprensione dei meccanismi biologici che stanno alla base della malattia e dalla identificazione di strutture molecolari, coinvolte in tali meccanismi, che possano rappresentare nuovi bersagli molecolari per nuovi approcci terapeutici.

Nonostante l'elevato numero di specialità medicinali in commercio e di composti in sviluppo clinico, i bersagli biologici utilizzati per la terapia medica sono meno di 600, dei quali il 50% rappresentato da recettori ed il 25% da enzimi. Considerando che il patrimonio proteico dell'uomo ammonterebbe ad alcune centinaia di migliaia di diverse proteine, solo per una minima frazione di esse è stato finora compreso il ruolo nella patogenesi delle malattie al punto tale da rendere possibile l'utilizzo a fini terapeutici.

Da oltre una decina di anni la applicazione della genomica al processo di scoperta dei farmaci, spesso indicata con il termine generale di farmacogenomica, è considerata la tecnologia con la maggior potenzialità per aumentare in modo esponenziale la capacità di identificare nuove molecole coinvolte nei processi patologici. Tali molecole rappresentano bersagli biologici per nuovi farmaci di natura chimica o biologica che legandosi ad essi possano modificare tali processi patologici e ripristinare uno stato di salute.

In modo parallelo alla possibilità di identificare i bersagli molecolari, nell'ultimo decennio è evoluta anche la sintesi di nuove sostanze grazie alle applicazioni della informatica e della robotica alla biologia ed alla chimica. La chimica combinatoriale e lo *screening* computerizzato hanno comportato un progressivo tra-

sferimento delle attività di ricerca dal laboratorio al computer e migliaia di prodotti vengono oggi ottenuti virtualmente, catalogati in librerie virtuali e testati in modo virtuale grazie all'integrazione delle conoscenze di proteomica e bioinformatica, all'ampliarsi delle banche dati ed alle nuove tecnologie

Sebbene l'inizio della ricerca farmacogenomica venga riferito al 1993, con l'avvio della collaborazione tra la società farmaceutica *SmithKline Beecham* e la società di ricerca genomica *Human Genome Science*, sono state nuove piccole imprese sorte nella seconda metà degli anni 90 - quali *Myriad*, *deCODE genetics*, *Illumina* oltre alla stessa *Human Genome Sciences* - a guidare lo sviluppo di nuovi modelli e tecnologie per la scoperta dei farmaci, diversi da quelli usati dai grandi gruppi farmaceutici multinazionali e basati integralmente sull'approccio farmacogenomico.

Tali modelli, anche attraverso la sottoscrizione di accordi di collaborazione con società di genomica, sono poi rapidamente recepiti a livello tecnologico e culturale anche dai grandi gruppi farmaceutici, che oggi devono all'applicazione dell'approccio genomico una parte rilevante delle loro scoperte.

Il modello genomico di R&D del farmaco si articola in diverse fasi:

- l'isolamento dei geni di suscettibilità ad una malattia attraverso ricerca genetica (di popolazione, su famiglie o di altra natura) o genomica (studi di espressione);
- l'identificazione delle proteine codificate da tali geni, lo studio delle loro funzioni e del loro ruolo nella patogenesi della malattia;
- la sintesi mirata o l'identificazione di nuove sostanze, chimiche o biologiche, in grado di legarsi alle proteine bersaglio e di modificare i processi coinvolti nella patogenesi della malattia.

La prima ricerca genetica di ampia portata per identificare geni di suscettibilità alle malattie ai fini della scoperta di nuovi farmaci è stata avviata, non senza polemiche, da *deCODE genetics* in Islanda, attraverso uno studio di popolazione. Questa ricerca ha fornito ad oggi i maggiori risultati nell'isolamento di geni di suscettibilità a malattie che hanno consentito l'identificazione di proteine divenute bersagli per nuovi composti (11).

L'approccio degli studi di popolazione si basa sullo sviluppo di database contenenti tre diversi set di dati - genetici, clinici e genealogici - tra loro correlati e riferiti ad una popolazione quali *Island-Biobank* in Islanda (11), *UK Biobank* nel Regno Unito (12), *Estonian Genome Project Foundation* in Estonia (13) ed altri. Le mappe così ottenute potrebbero contenere frammenti di particolari cromosomi condivisi da un elevato numero di pazienti tra loro imparentati in modo tale da risultare superiore rispetto a quello atteso casualmente. Analisi dettagliate di queste regioni del cromosoma con gruppi più densi di marcatori genetici permettono quindi di isolare i geni chiave che conferiscono il rischio di malattia come pure le varianti o gli aplotipi altamente correlati con la malattia.

Altre modalità spesso utilizzate per l'isolamento dei geni di suscettibilità alle malattie coinvolgono il reclutamento di individui geneticamente correlati, quali famiglie, coppie di fratelli entrambi affetti dalla malattia, genitori non affetti e figli affetti.

Il DNA (o gli RNA) delle persone che hanno partecipato allo studio viene poi analizzato spesso usando il metodo dei “*geni candidati*” che consiste nell’analizzare geni che possano essere ragionevolmente implicati nel fenomeno studiato (ad esempio geni che codificano per proteine con un ruolo noto o ipotizzato nel meccanismo patogenetico di una certa malattia o di malattie simili).

Accanto a questa metodologia sono più recentemente evolute la analisi del genoma con mappe di *SNPs* (polimorfismi di singolo nucleotide) e l’analisi dell’espressione genica tramite *gene chips* (o *microarray*).

L’analisi dell’intero genoma (*Whole Genome Genotyping*) con mappe di *SNPs* ad alta densità è un metodo molto promettente perché permette di scandagliare l’intero genoma senza richiedere che siano noti a priori geni potenzialmente coinvolti nel fenomeno in studio.

Questo metodo potrebbe rendere superato il metodo dei geni candidati, ma ad oggi non è ancora possibile applicarlo in maniera routinaria per i costi elevati e la necessità di risolvere alcuni problemi di natura tecnologica. Un ulteriore strumento per la identificazione dei genotipi correlati al rischio di malattia è rappresentato dalla *HapMap* o mappa degli aplotipi, realizzata dall’*International HapMap Project* (17), che ha sequenziato le più comuni differenze genetiche nell’intero genoma di 269 persone e catalogato 1,3 milioni di variazioni nella prima fase completata nel 2005 e 3,1 milioni nella successiva fase del progetto, completata nel 2007 (18).

Le informazioni fornite dalla prima versione della *HapMap* hanno portato all’identificazione di oltre 50 geni associati a comuni malattie come diabete, ipercolesterolemia, sclerosi multipla e cancro della prostata, e alla messa a punto di tecniche per facilitare la loro individuazione. Il progetto *HapMap* ha generato le informazioni per costruire chip contenenti 500.000 SNP che coprono tutto il genoma umano. Con i chip di SNP si possono determinare aplotipi genetici su tutto il genoma umano nell’ambito dei quali identificare poi il gene o la variante genetica di interesse.

I *gene chips* sono costituiti da un supporto di vetro più piccolo di 2 centimetri quadrati, suddiviso in aree di 0,0005 centimetri quadrati, su ognuna delle quali è possibile fissare circa 100 sequenze di DNA a singola elica di circa 20 paia di basi. Possono essere costruite migliaia di sequenze diverse, specifiche per singoli geni. Ad esempio sono stati già prodotti *gene chips* contenenti sequenze specifiche per riconoscere selettivamente tutti i circa 27.000 geni umani. Con questa tecnica si possono misurare cambiamenti nella regolazione genetica, indotti ad esempio da un processo patologico, come quello tumorale, con analisi rapide e computerizzate. La differenza fondamentale rispetto ad altre metodiche sta nel fatto che con il *gene chip* si possono effettuare analisi sistematiche su un gran numero di geni simultaneamente, anche migliaia, evitando di analizzare i geni uno alla volta come avveniva in precedenza.

Human Genome Sciences ha identificato, attraverso l’isolamento del gene, la Lipoproteina associata alla Fosfolipasi A2 (*Lp-PLA2*), un enzima coinvolto nella formazione della placca aterosclerotica che è divenuta il bersaglio per *Darapladib*, una piccola molecola con azione di inibitore attualmente in fase II di

sviluppo clinico da parte di GlaxoSmithKline. Altre molecole della medesima classe sono in fase più precoce di sviluppo clinico.

Individualizzazione della terapia

Lo sviluppo di un farmaco è un processo complesso, oneroso, impegnativo e scarsamente efficiente. Solo una molecola su 10.000 sintetizzate completa il processo di sviluppo ed entra in commercio, con un elevato attrito nella fase di sviluppo clinico. Il costo complessivo di R&D per ogni farmaco immesso sul mercato è di circa € 850 milioni (27).

In genere un farmaco indicato per il trattamento di una malattia cronica presenta una efficacia compresa tra 60 e 80%. Sebbene le reazioni avverse siano in genere di limitato significato clinico, in taluni casi sono tali da limitarne l'uso e talora da comportare la interruzione del suo sviluppo clinico o il ritiro dal commercio. A livello sanitario, le reazioni avverse ai farmaci rappresentano uno tra i maggiori problemi e si stima che negli Stati Uniti siano la quarta causa di ospedalizzazione e la sesta causa di morte.

L'attività farmacologia e di conseguenza l'effetto terapeutico di un farmaco dipendono dalla interazione della molecola con diverse proteine che ne condizionano la farmacocinetica e la farmacodinamica: alcune sono coinvolte nell'assorbimento, altre nel trasporto ematico, nella sua metabolizzazione ed eliminazione, altre ancora rappresentano i bersagli del farmaco.

È difficile stimare quale percentuale di farmaci possa presentare variazioni tali da giustificare lo sviluppo di un test genetico predittivo di risposta. Le stime disponibili ad oggi si rivelano semplici supposizioni non basate su dati di fatto e solo l'esperienza dei prossimi anni potrà chiarire questo aspetto.

Il miglioramento della efficacia e della tollerabilità di questi farmaci attraverso la selezione dei pazienti da sottoporre al trattamento rappresenta il razionale dello sviluppo della farmacogenetica. La selezione della terapia in base al profilo genetico del paziente consente da una parte di prescrivere il farmaco più appropriato per ciascun paziente e pertanto di migliorare la qualità e l'appropriatezza del processo di cura e dall'altra di migliorare l'efficienza del processo di sviluppo dei

Tab. 1 - Evoluzione della ricerca farmacogenetica.

Fase della Ipotesi - I	Fase della Verifica - II	Fase dello Sviluppo - III
Ricerca della relazione tra il fenotipo (rappresentato dalla risposta al trattamento farmacologico) ed il genotipo (rappresentato da uno o più geni codificanti proteine coinvolte nella farmacocinetica o nella farmacodinamica del farmaco) attraverso studi clinici esploratori con disegno retrospettivo o trasversale	Conferma dell'associazione e talvolta verifica della relazione causale della associazione tra le variabili fenotipica e genotipica attraverso studi clinici confirmatori con disegno randomizzato e prospettico	Modifica dell'attuale modello di sviluppo clinico del farmaco attraverso la incorporazione del test predittivo a partire dalla ricerca clinica pilota

farmaci, riducendone l'attrito. Sono ormai trascorsi una decina di anni dall'avvio dei primi studi di farmacogenetica, che oggi rappresentano una componente presente in modo sistematico nello sviluppo dei nuovi farmaci. L'evoluzione della farmacogenetica nel corso di tale periodo può essere descritto attraverso la successione di tre fasi, distinte da specifici obiettivi: ricerca di ipotesi, verifica delle ipotesi, nuovo modello di sviluppo del farmaco.

Seppure l'avvio di queste fasi nella evoluzione della farmacogenetica sia stato sequenziale, gli studi oggi condotti possono essere classificati in una qualsiasi di queste diverse fasi e per uno specifico farmaco le fasi II e III possono anche essere condotto in concomitanza.

Conclusioni

I risultati prodotti dalla ricerca nell'ultimo decennio ha sostanzialmente consentito lo sviluppo della Medicina Personalizzata e l'introduzione nella pratica clinica delle prime applicazioni basate su nuove tecnologie in grado di identificare la suscettibilità alle malattie comuni e di predire la risposta al trattamento farmacologico.

L'appropriata introduzione delle nuove tecnologie proprie della Medicina Personalizzata richiede la valutazione di prove di efficacia necessarie per orientare le decisioni relative alla loro impiego. Tali prove devono essere generate attraverso studi in grado di verificare la relazione causale tra le associazioni di genotipo e fenotipo (sia esso malattia o risposta alla terapia) che oggi con sempre maggior frequenza sono identificate grazie alla disponibilità di tecnologie di scansione del genoma sempre più efficaci. Studi esploratori e studi confirmatori coesistono attualmente nei centri di ricerca e nelle pubblicazioni: è necessario poter distinguere queste due tipologie diverse di studi ed il significato dei rispettivi risultati, allo scopo di prevenire il rischio che sulla base di risultati che dimostrano mere associazioni possano essere offerte a medici e pazienti test senza provata utilità clinica.

I tempi di sviluppo richiedono diversi anni e pertanto solo negli ultimi 2 anni i primi test hanno completato il proprio sviluppo e sono stati resi disponibili per l'uso clinico.

La frequenza di introduzione in commercio è destinata ad aumentare nei prossimi mesi ed anni, con il progressivo completamento degli studi confirmatori attualmente in corso, mentre la diffusione dei sistemi di scansione del genoma e lo sviluppo di banche dati permetterà la identificazione di un numero elevato di associazioni genotipo - fenotipo da sottoporre a successiva verifica.

Tale identificazione, anticipata rispetto alle fasi di sviluppo clinico, ha consentito di modificare la fase 3 di un farmaco sperimentale. L'ulteriore anticipo della disponibilità delle informazioni potrà a breve modificare anche la fase 2, che potrebbe diventare la fase di confirmatoria delle associazioni. Tale sviluppo renderà a breve disponibili le prime terapie mirate (*targeted therapy*) per malattie diverse da quelle oncologiche, per le quali rappresentano strumenti di cura consolidata da alcuni anni.

È tuttavia prevedibile che le applicazioni della farmacogenetica ed in particolare dello sviluppo clinico del farmaco interessino solo una quota di farmaci pipeline delle imprese farmaceutiche la cui entità sarà possibile stimare solo dopo l'esperienza dei prossimi anni.

Sebbene più limitata nella entità e con un qualche ritardo nei tempi di sviluppo, derivati entrambi da una eccessiva sovrastima iniziale delle sue potenzialità, sia le premesse che le promesse della Medicina Personalizzate sono state finora sostanzialmente rispettate.

L'ulteriore sviluppo della ricerca genomica e farmacogenetica, anche nel nostro paese, richiede la necessità di una appropriata gestione delle implicazioni etiche, sociali e giuridiche e soprattutto una adeguata formazione dei medici, dei comitati etici, degli operatori e degli amministratori sanitari, dei cittadini e dei pazienti. Le raccomandazioni elaborate da esperti di diverse discipline permettono di gestire la maggior parte delle criticità poste dalla ricerca di genetica clinica e di valutare e bilanciare i benefici ed i rischi per massimizzare i primi e a minimizzare i secondi. Tuttavia la conoscenza di tali raccomandazioni è ancora troppo limitata ad un ambito troppo esclusivo e specialistico.

Bibliografia

1. PricewaterhouseCoopers. Pharma 2020: The Vision - Which Path Will You Take? <http://www.pwc.de/fileserver/RepositoryItem/PHARMA%202020%20FINAL.pdf?itemId=1972678> consultato il 10.10.2007
2. European Commission. The Innovative Medicines Initiative in. http://ec.europa.eu/research/health/imi/index_en.html consultato il 30.9.2007 consultato il 10.10.2007.
3. Burke W, Psaty BM. Personalized Medicine in the Era of Genomics. *JAMA*. 2007; 298 (4): 1682-4.
4. Pirazzoli A, Recchia G. Pharmacogenetics and pharmacogenomics: are they still promising? *Pharmacol Res*. 2004; 49 (4): 357-61.

**GENOMICA E ATTUALITÀ
DI MEDICINA MOLECOLARE
IN EMOPATIE MALIGNI**

Tecnologie genomiche per migliorare la stratificazione prognostica e impostare terapie personalizzate nelle LAM

Paolo Bernasconi, Marina Boni, Paola Maria Cavigliano, Ilaria Giardini, Barbara Rocca, Rita Zappatore, Irene Dambruoso, Celeste Calvello, Marilena Caresana

Clinica di Ematologia, Fondazione IRCCS Policlinico San Matteo, Università degli Studi di Pavia

La genomica delle LAM è stata ed è tuttora un'area di intensa attività di ricerca che si prefigge tre principali obiettivi a rilevante impatto biologico e clinico: una sempre più approfondita conoscenza dei meccanismi molecolari responsabili della patogenesi della malattia, l'identificazione di nuovi marcatori molecolari che essendo realmente predittivi del decorso clinico della malattia consentano una sempre più precisa stratificazione prognostica dei pazienti, l'identificazione di nuovi e sempre più precisi bersagli di terapia molecolare.

L'analisi cromosomica convenzionale è stata il primo passo verso l'analisi dell'intero genoma ed è tuttora indispensabile per un'accurata definizione diagnostica e prognostica di ogni paziente affetto da LAM. All'inizio degli anni '70 sono state individuate e caratterizzate sul piano molecolare alcune traslocazioni cromosomiche come la t(8;21), l'inv(16) e la t(15;17). Questi riarrangiamenti hanno in comune alcune sorprendenti caratteristiche: generano proteine di fusione chimeriche, almeno uno dei due geni coinvolti nel riarrangiamento cromosomico svolge un ruolo rilevante nella ematopoiesi normale, identificano pazienti a prognosi favorevole. Negli anni seguenti la citogenetica convenzionale ha permesso di individuare ulteriori anomalie cromosomiche ricorrenti come la delezione (5q) o la monosomia 5, la monosomia 7, l'inv(3) o la t(3;3), le traslocazioni della banda q23 del cromosoma 11 e altre.

L'analisi cromosomica convenzionale presenta però uno scarso potere di risoluzione, dell'ordine delle 3-4 megabasi. Per questo motivo l'esame convenzionale del cariotipo è destinato ad essere sempre meno utilizzato e ad essere sostituito da altre metodiche sviluppatesi recentemente come la "microarray-based comparative genomic hybridization" (aCGH) ed il "single nucleotide polymorphism (SNP) array karyotyping" (SNPa) che, potendo eseguire un esame ad alta risoluzione dell'intero genoma ("high-resolution genome scan"), sono in grado di dimostrare nuove e ricorrenti alterazioni del "copy number" nel 65% delle LAM

(1-3). Anche l'applicazione di metodiche di biologia molecolare ha permesso di dimostrare che nelle LAM un sempre maggior numero di geni è colpito da mutazione. Tra i geni più coinvolti bisogna ricordare quello per la "fms-like tyrosine kinase 3" (FLT3), quello per la nucleofosmina 1 (NPM1) e quello per la "CCAAT/enhancer-binding protein α " (CEBPA). La "internal tandem duplication" (ITD) della regione iuxtamembrana e le mutazioni del dominio kinasico, "activation loop (AL) mutations", del gene FLT3 si osservano rispettivamente nel 20-27% e nel 5-7% delle LAM (4, 5). Mentre solo la prima mutazione ha un impatto prognostico sfavorevole essendo associata ad un alto rischio di ricaduta e ad una bassa probabilità di ottenere una seconda remissione completa (RC), entrambe possono essere bersaglio di piccole molecole ad azione inibitoria impiegate in trials clinici tuttora in corso (6). Le inserzioni "in-frame" nel gene della nucleofosmina 1 (NPM1) causano un'alterata posizione della proteina che dal nucleolo si sposta nel citoplasma, hanno un'incidenza del 30% circa ed in assenza della ITD del gene FLT3 determinano una prognosi favorevole (7). Il gene CEBPA è bersaglio di due tipi di mutazioni eterozigoti. La mutazione nonsense nella regione amino-terminale (N-terminali) codifica per un'isoforma tronca che svolge un'azione dominante negativa nei confronti della proteina normale, mentre la mutazione "in-frame" nella regione basica carbossi-terminale (C-terminale) dominio a cerniera di leucina codifica per proteine dotate di una minor capacità di legarsi al DNA e di formare dimeri (8). In circa i due terzi dei pazienti le mutazioni C- e N-terminali sono bi-alleliche, la maggior parte (il 90% circa) sono doppie eterozigosi (cioè un allele mostra una mutazione C-terminale, l'altro una mutazione N-terminale), le restanti sono omozigoti forse dovute a disomia uniparentale. Le mutazioni di CEBPA si osservano nel 10-18% dei pazienti a cariotipo normale e nel 40% dei pazienti con del(9q) senza cariotipo complesso; solo quelle bi-alleliche sono a prognosi favorevole.

Tuttavia, nonostante queste rilevanti informazioni il 40% delle LAM viene ancora considerata a rischio intermedio perché non presenta specifici marcatori citogenetici o molecolari. Questo dato unitamente al fatto che, salvo rare eccezioni (9), le singole mutazioni sopra elencate non sono da sole capaci di indurre una leucemia acuta in modelli murini geneticamente modificati indica che le nostre conoscenze riguardanti la genomica delle LAM sono ancora del tutto insufficienti. Inoltre, le informazioni a nostra disposizione suggeriscono che questi disordini clonali dell'emopoiesi sono caratterizzati da una profonda eterogeneità genetica determinata dalla progressiva acquisizione di diverse mutazioni somatiche in progenitori ematopoietici che perciò presentano un alterato funzionamento dei meccanismi che ne regolano l'automantenimento, la proliferazione e la differenziazione. Pertanto, molto recentemente è stato ipotizzato che la metodica di "Next-generation sequencing" (NGS), già impiegata per lo "Human Genome Sequencing", potesse essere l'unica in grado di risolvere questa estrema eterogeneità genetica garantendo l'analisi precisa e approfondita dell'intero genoma della cellula leucemica. Dati recentissimi indicano che questa nuova metodologia in combinazione con l'analisi bioinformatica dei dati ottenuti sta realmente rivoluzionando le nostre conoscenze. Inoltre, il continuo miglioramento delle tecno-

logie di sequenziamento con progressiva riduzione dei tempi e dei costi necessari all'esecuzione della procedura fa prevedere per il prossimo futuro un rapido trasferimento del NGS dalla ricerca alla clinica.

Metodiche di “Next-generation sequencing”

Si tratta di una procedura che permette di produrre simultaneamente milioni di piccole sequenze di DNA delle dimensioni di 50-100 nucleotidi. Tali sequenze vengono poi ricercate e mappate all'interno del genoma umano di riferimento in modo tale da ottenere il genoma della cellula neoplastica. Bisogna però sottolineare che è assolutamente necessario analizzare il DNA tumorale ed il DNA normale di ogni individuo. Infatti il genoma umano contiene 3-4 milioni di sequenze varianti e centinaia di variazioni del “copy number”. La maggior parte di queste varianti è ereditata e quindi non costituisce una mutazione acquisita. I metodi per studiare il genoma delle cellule neoplastiche sono sottoelencati. Spesso è necessario impiegare più metodiche simultaneamente.

Whole-genome sequencing

Lo scopo di questa metodica è il sequenziamento dell'intero genoma. La procedura immobilizza in una camera microfluidica “random libraries” composte da piccoli frammenti preparati per “end ligation” di adattatori sintetici piattaforma-specifici sottoposti ad amplificazione. Ogni strumento di sequenziamento legge simultaneamente ed in successione la sequenza di milioni di frammenti usando un sistema di rilevazione a fluorescenza ed un sistema di elaborazione delle immagini (11). I sequenziatori di ultima generazione producono “short reads” di 75-150 paia di basi circa, il genoma viene poi nuovamente sequenziato con un alto grado di ridondanza (di solito 30-40 volte l'ampiezza di 3 bilioni di paia di basi di genoma aploide o 10-120 bilioni di paia di basi) per essere sicuri che tutte le regioni del genoma siano state sottoposte a campionatura. Il sequenziamento di entrambe le estremità di un frammento di DNA (“paired-end sequencing”) rende più preciso il riallineamento dei “short reads” rispetto alla loro sede all'interno del genoma e facilita quindi la scoperta di varianti strutturali (traslocazioni, inversioni, alterazione del numero di copie). Tutte le fasi di questo processo, dall'allineamento genomico all'identificazione della mutazione sono regolate da robusti algoritmi. I maggiori limiti del WGS sono i costi elevati e la complessità dei dati prodotti. Tuttavia, i continui progressi di tale tecnologia consentirà di superare nel prossimo futuro tali problematiche.

Exome sequencing

Lo scopo di questa procedura è il sequenziamento selettivo dell'1-2% del genoma. Le aree genomiche sequenziate contengono sequenze geniche codificanti, microRNA e RNA non codificante. Il DNA genomico viene ibridizzato ad arrays contenenti “probes” disegnati per catturare approssimativamente 200.000 esoni. Il DNA che ibridizza viene poi sequenziato. I maggiori vantaggi di questa procedura sono il basso costo ed un abbastanza accurato sequenziamento visto che

viene analizzato solo l'1-2% del genoma. Tuttavia, la metodica non individua mutazioni che si sviluppano al di fuori delle aree esoniche che costituiscono più del 98% del genoma e varianti strutturali come traslocazioni con punti di rottura in sequenze introniche.

Transcriptome sequencing

Lo scopo di questa metodica è quello di sequenziare tutti i geni che vengono trascritti. La metodica permette di quantificare i livelli di espressione di determinati geni, di individuare cambiamenti post-trascrizionali come quelli prodotti da uno "splicing" alternativo, di identificare trascritti di fusione prodotti da una traslocazione cromosomica. Come l'"exome sequencing" anche questa metodica non individua le mutazioni che coinvolgono aree non codificanti del genoma, quelle causate dalla perdita di una o di entrambe le copie di un gene e quelle che accelerano il turnover del RNA. Inoltre, il "transcriptome sequencing" tende erroneamente a favorire il sequenziamento di trascritti espressi ad alti livelli mentre quello di trascritti espressi a bassi livelli è scarso o assente.

Altre metodiche di sequenziamento

Tra queste bisogna ricordare le procedure che sequenziano aree dell'intero genoma in cui si sono verificate modificazioni epigenetiche. In particolare, sono state sviluppate metodiche che consentono di eseguire la mappatura ad alta risoluzione (a livello di singola base) dello stato di metilazione dell'intero DNA e di definire la struttura della cromatina dell'intero genoma grazie al mappaggio di istoni sede di modificazioni post-translazionali.

Risultati

Nuove mutazioni

Sino ad ora la ricerca di mutazioni eventualmente presenti nella popolazione leucemica veniva condotta andando a sequenziare un singolo gene o un insieme di geni la cui mutazione era già stata osservata in altre neoplasie. Questo approccio metodologico pur fornendo rilevanti contributi nella chiarificazione del processo di leucemogenesi, presenta importanti limiti: richiede che i geni da analizzare siano stati in precedenza selezionati e non fornisce informazioni su una possibile cooperazione di più geni nel processo neoplastico.

Le metodiche di WGS consentono di superare questi limiti perché forniscono informazioni sull'intero genoma consentendo di passare da un "single gene approach" ad un "multiple gene approach". Diventa così possibile individuare nuove mutazioni e soprattutto stabilire quali siano le mutazioni che cooperano nel processo di leucemogenesi.

Nel 2008 Ley et al. aveva per la prima volta esaminato il genoma di cellule leucemiche di un paziente di età inferiore a 60 anni affetto da LAM classificata come M1 in base a criteri morfologici ed immunofenotipici. Il paziente era privo di lesioni all'esame citogenetico convenzionale, all'esame molecolare e all'aCGH. Viceversa, il NGS aveva dimostrato una mutazione con possibili conseguenze

traslazionali della regione codificante di dieci geni, che includevano FLT3, colpito da ITD, e NPM1. Gli altri otto geni non erano mai stati trovati mutati nelle LAM e non erano mutati in altri centoottantasette pazienti analizzati dallo stesso studio. Venne perciò ipotizzato che le mutazioni potessero essere alleli rari a significato patologico o più verosimilmente mutazioni somatiche a significato non patologico acquisite da una cellula staminale ematopoietica durante il processo di “self-renewing” prima della trasformazione neoplastica.

Successivamente l’analisi del genoma di un altro paziente affetto da LAM-M1 aveva fornito un analogo risultato: era stata identificata la mutazione di dieci geni tra cui NPM1, RAS, IDH1 e del gene mitocondriale ND4. Questo studio aveva per la prima volta dimostrato che nelle LAM il gene IDH1 era mutato a livello del codone 132. IDH1 codifica per la isocitrato deidrogenasi citoplasmatica di tipo 1 ed è mutato nel 70% dei pazienti con glioma. In questi ultimi prevale però la mutazione R132H, mentre nelle LAM la mutazione R132C (50% dei pazienti) (12, 13). La scoperta di questa mutazione è stata il punto di partenza per dimostrare che nelle LAM anche il gene IDH2, omologo mitocondriale di IDH1, può essere colpito da mutazione. Le mutazioni di IDH1/IDH2, mai associate tra loro e mai presenti in combinazione con una mutazione acquisita del gene TET2, si osservano nel 12-17% delle LAM, specie a citogenetica normale (incidenza 22-30%) (14, 15). Nei pazienti con mutazione del gene NPM1 e senza ITD del gene FLT3 la mutazione di IDH1 determina una prognosi sfavorevole. L’enzima IDH1 mutato produce il metabolita 2-idrossiglutarato che impedisce l’idrossilazione mediata da TET2 dei residui di metilcitosina (16).

In un secondo paziente un ulteriore miglioramento delle metodiche di NGS avevano permesso di individuare una mutazione frameshift del gene DNMT3A (17). Quest’ultimo codifica per una metiltransferasi che catalizza la metilazione *de novo* dei residui di citosina. Un’estensione dello studio ad altri 281 pazienti affetti da LAM *de novo* aveva individuato mutazioni “missense, nonsense, frameshift”, alterazioni del sito di “splicing” o delezioni di DNMT3A nel 22% dei casi. Nessun paziente a citogenetica favorevole aveva mostrato la mutazione che era invece presente nel 30% dei pazienti a citogenetica intermedia. La mutazione era correlata ad un alto valore di globuli bianchi e aveva determinato una prognosi sfavorevole indipendentemente da età e ITD di FLT3.

Nel 2010 Yamashita et al aveva eseguito il “targeted exome sequencing” in settantadue pazienti affetti da LAM ed aveva dimostrato che tre pazienti presentavano una mutazione acquisita della DNMT3A a livello dell’amminoacido R882. Successivamente, la stessa mutazione era stata identificata in ventitre pazienti (20%) affetti da LAM-M5 (18). A tuttoggi il ruolo patogenetico di queste mutazioni è ancora sconosciuto: la mutazione R882, presente nella maggior parte dei pazienti, si associa ad una “gain of function”, mentre le delezioni di DNMT3A causano una “loss of function”. Inoltre, nessuna mutazione è strettamente correlata ad un particolare pattern di metilazione del DNA o ad particolare alterazione dell’espressione genica. Dati recenti ottenuti in modelli murini hanno suggerito che cellule staminali ematopoietiche prive di DNMT3A possano direttamente contribuire al processo di leucemogenesi visto che vanno incontro ad una signifi-

cativa espansione se impiegate in procedure trapiantologiche seriate (Challen et al., 2010).

Questi primi studi che sono stati più sopra elencati hanno già fornito alcune importati delucidazioni dei meccanismi molecolari responsabili del processo di leucemogenesi:

1. Le LAM non sono causate da instabilità genetica bensì da un numero relativamente piccolo di mutazioni (“driver mutations”). Infatti, i blasti leucemici acquisiscono un numero relativamente piccolo di variazioni a livello di singolo nucleotide (“Single nucleotide variants”, SNV o mutazioni puntiformi somatiche acquisite): in quattro pazienti sottoposti a sequenziamento il numero di SNV all’interno delle aree codificanti del genoma era compreso tra 10 e 26 (12, 17, 19, 20).
2. La maggior parte delle SNV acquisite nelle LAM sono personali, sono cioè “passenger mutations” che non possiedono alcun ruolo nella patogenesi della malattia. Nei quattro pazienti esaminati le SNV situate all’interno delle regioni codificanti erano state 59, ma nessun gene era mutato in due o più pazienti (12, 17, 19, 20).

Suscettibilità allo sviluppo di LAM

Stabilire quali geni siano responsabili dello sviluppo di una LAM è una sfida estremamente affascinante che avrebbe come ricaduta una più accurata prevenzione, una più precoce diagnosi e scelte terapeutiche sempre più mirate a colpire il meccanismo molecolare della malattia. Si ritiene che i geni responsabili della patogenesi di una neoplasia siano circa un centinaio (21). Per il momento analizzare tutti questi geni sarebbe costoso ed improduttivo visto molti tra questi geni sono tuttora sconosciuti, in futuro però il NGS potrebbe fornire interessanti informazioni.

Lo scorso anno è stata riportata la storia clinica di una paziente che all’età di 37 anni aveva sviluppato un tumore al seno, all’età di 39 un tumore ovarico e all’età di 42 anni una LAM secondaria fatale (19). L’anamnesi aveva escluso una familiarità neoplastica, ma il così precoce sviluppo di varie neoplasie aveva fatto ipotizzare che la paziente potesse presentare una sindrome che la esponeva ad un aumentato rischio di contrarre una neoplasia. La ricerca delle mutazioni dei geni BRCA1 e BRCA2 era risultata negativa.

L’analisi del DNA della cute della paziente mediante WGS aveva identificato una delezione eterozigote di circa 3kb che determinava la perdita degli esoni 7-9 del gene TP53 (20). Il DNA del tessuto ematopoietico mostrava la stessa mutazione allo stato omozigote forse a seguito di un processo di disomia uniparentale. Tale mutazione doveva essere insorta spontaneamente ed essere ricercata nei figli della paziente visto che non era stata identificata nel DNA della madre sottoposta ad una nuova analisi con WGS. Infatti, le linee guida del “National Comprehensive Cancer Network” stabiliscono che i portatori di una mutazione eterozigote del gene TP53 hanno un rischio di sviluppare una neoplasia nel corso della loro esistenza pari al 90% e debbano quindi essere annualmente analizzati per certe neoplasie (22).

Definizione di genomi complessi e di traslocazioni criptiche

La citogenetica convenzionale e la FISH non riescono a definire riarrangiamenti complessi e criptici del cariotipo che possono invece essere dimostrati dal NGS. A questo proposito è stato proprio recentemente riportato il caso di una paziente affetta da LAM-M3 che all'analisi cromosomica convenzionale aveva mostrato un cariotipo complesso con numerosi marcatori, la perdita di vari cromosomi e numerose traslocazioni non usuali, ma non la classica traslocazione 15;17. Anche la FISH non era riuscita a dimostrare il riarrangiamento PML-RARA. Dopo un ciclo di chemioterapia d'induzione secondo un protocollo standard la paziente aveva raggiunto la remissione completa e la tipizzazione HLA aveva individuato una sorella identica.

Siccome il cariotipo complesso in assenza della traslocazione 15;17 esponeva la paziente ad un alto rischio di recidiva e la candidava al trapianto allogenico, era a questo punto assolutamente necessario escludere che la presenza del gene di fusione PML-RARA. Fu quindi stabilito di analizzare il DNA delle cellule leucemiche con il WGS e nel frattempo di consolidare la remissione con un ciclo di terapia con triossido di arsenico.

L'analisi mediante WGS permise di individuare il gene di fusione PML-RARA e di stabilire che quest'ultimo era prodotto dall'inserzione di un frammento di 77kb corrispondente al locus LOXL1-PML nel locus RARA. La RT PCR o altre metodiche molecolari che richiedevano l'impiego di "primers" personalizzati avrebbero potuto fornire lo stesso risultato ma non era stato possibile eseguirle per mancava di materiale. In sintesi, l'esecuzione del WGS avendo dimostrato il riarrangiamento PML-RARA aveva permesso di evitare le complicanze ed i costi della procedura trapiantologica (19).

Evoluzione genomica nelle LAM

Un altro dato molto interessante prodotto dal NGS è costituito dal fatto che molto raramente il campione in esame contiene un unico genoma leucemico. Al contrario, la popolazione leucemica è di solito un mosaico genomico, è cioè costituita da diversi genomi che identificano vari sub-cloni cellulari attivamente proliferanti perché dotati di un vantaggio proliferativo. Tali sub-cloni sono in competizione tra loro e con popolazioni clonali ancestrali per la sopravvivenza. Questo dato potrebbe spiegare perché nelle LAM la chemioterapia induce la remissione nella maggior parte dei pazienti ma meno della metà di questi guarisce e perché dopo un'iniziale risposta alcuni pazienti recidivano e possono diventare resistenti alla chemioterapia.

L'esecuzione del NGS potrebbe individuare quei geni che garantiscono una sensibilità alla chemioterapia distinguendoli da quelli che determinano chemioresistenza e potrebbe permettere di stabilire se la chemioterapia sia effettivamente capace di "modellare" il genoma della cellula leucemica. Anche il confronto tra LAM de novo e secondarie potrebbe essere molto interessante perché potrebbe permettere di individuare somiglianze e differenze che avrebbero come possibile ricaduta un miglioramento degli attuali algoritmi diagnostici, prognostici e l'identificazione di nuovi possibili bersagli di terapia molecolare.

Problematiche connesse con il WGS

La capacità del WGS di individuare variazioni di sequenza dipende dall'accuratezza della lettura, cioè dal numero di volte con cui un singolo nucleotide è campionato.

Attualmente, la metodica tende a coprire ciascun nucleotide con una media di 30-40 letture separate. Tuttavia, in questo modo il 50% circa delle sequenze varianti individuate sono falsi positivi prodotti da errori di mappatura, da errori di polimerasi, e dalla bassa frequenza di varianti somatiche nel campione. Perciò, una successiva conferma delle varianti è tuttora necessaria per tutte le procedure di sequenziamento. Inoltre, gli attuali algoritmi rendono problematica l'identificazione di piccole inserzioni e delezioni (indels) e, siccome il 50% del genoma contiene sequenze ripetitive, la mappatura delle sequenze lette risulta spesso imperfetta così da rendere possibile la mancata identificazione di una mutazione (17). Tutti questi problemi tecnici potranno essere migliorati con il sequenziamento di ulteriori genomi umani e con una riduzione dei costi che permetterà un sequencing più accurato. Tuttavia, solo un progressivo miglioramento della tecnologia renderà possibile la mappatura delle sequenze all'interno delle regioni ripetitive del genoma.

Un altro problema che si verifica quando il WGS viene impiegato per l'analisi del DNA di un campione di cellule leucemiche è costituito dal fatto che quando il campione in esame viene paragonato alla sequenza "reference" normale si possono osservare centinaia di alterazioni nel numero di copie e 3-4 milioni di varianti a livello di singoli nucleotidi.

La stragrande maggioranza di queste varianti sono ereditate. Siccome nei database oggi disponibili varianti rare o paziente specifiche sono sottostimate, il confronto ("filtering") delle varianti osservate con quelle contenute in tali database è del tutto inadeguato per distinguere in modo sicuro tra mutazioni somatiche e varianti ereditate. Per identificare nel modo più preciso possibile eventuali mutazioni somatiche è perciò assolutamente necessario che in ogni paziente affetto da LAM il genoma delle cellule leucemiche venga confrontato con il genoma di cellule sane. Una volta escluse le varianti ereditate ed i falsi positivi, il genoma delle cellule leucemiche può contenere sino a 500 mutazioni somatiche, non tutte però, come già riportato, svolgono un ruolo importante nella patogenesi della malattia. Quelle a rilevante impatto patogenetico sono contenute nelle regioni codificanti del genoma. Di solito l'analisi di queste regioni individua almeno 10 mutazioni per genoma leucemico.

La maggior parte di queste lesioni non sono ricorrenti quando vengono analizzate ampie casistiche, si sviluppano spontaneamente in una determinata linea cellulare nel corso della vita del paziente e non forniscono alcun vantaggio proliferativo. Per questo motivo non dovrebbero essere identificate se contenute in una popolazione cellulare policlonale normale. Tuttavia, la popolazione cellulare che acquisisce una mutazione "trasformante" e quindi un vantaggio proliferativo diventa dominante e presenta quindi anche quelle mutazioni che non svolgono alcun ruolo nello sviluppo della malattia. Questo concetto è stato recentemente

dimostrato dallo studio di popolazioni cellulari ottenute da soggetti sani. Diventa quindi difficile distinguere le mutazioni a rilevante ruolo patogenetico (“driver mutations”) da quelle prive di tale ruolo (“background mutations”). Per risolvere questo problema sono necessari algoritmi statistici che identifichino geni la cui mutazione avviene ad una frequenza significativamente superiore a quella determinata dal caso, sono necessarie rigorose dimostrazioni circa il significato funzionale di tali mutazioni ed è necessario valutare il ruolo delle mutazioni non ricorrenti. Altra problematica connessa con il WGS è che esso non dà informazioni su modificazioni epigenetiche e su alterazioni del “gene expression profile”. Pertanto per definire correttamente i cambiamenti genomici avvenuti in un paziente con LAM sarà indispensabile combinare il WGS con una metodica che analizzi l’espressione del RNA (ad esempio il “transcriptome sequencing”) e con una metodica che valuti i cambiamenti epigenetici (ad esempio il “bisulfate sequencing”).

Oltre a questi limiti di natura squisitamente tecnica, esistono altri limiti di natura economica che attualmente impediscono di introdurre il WGS nella pratica clinica. Tra questi bisogna ricordare le infrastrutture richieste, l’“expertise” del personale ed il fatto che il tempo necessario a completare l’analisi è tuttora abbastanza lungo. Tuttavia, il progressivo aumento del numero di pazienti affetti da LAM sottoposti a WGS renderà le infrastrutture bio-informatiche e tecniche più accessibili. Sarà inoltre necessario generare “reports” che possano essere correttamente interpretati da clinici che spesso mancano di nozioni e “training” in genetica medica e fornire ai pazienti non solo quelle informazioni che hanno una diretta influenza sulla loro possibilità di cura ma anche quelle scoperte incidentalmente ma utili per loro ed i loro figli (23).

Prospettive future

Siccome il WGS identifica in modo molto preciso tutte le possibili modificazioni genomiche (mutazioni di singoli nucleotidi, delezioni, amplificazioni, traslocazioni cromosomiche e disomie uniparentali), questa metodica avrà un sicuro impatto sulla classificazione delle LAM che, come già avvenuto con l’avvento di citogenetica convenzionale, FISH e RT PCR, diverrà più accurata. Anche la terapia di questi disordini onco-ematologici potrà divenire più selettiva utilizzando piccole molecole ad azione inibitoria quando sia stata ad esempio individuata la mutazione di una tirosina kinasi, sfruttando l’alterato funzionamento di una particolare via di segnale per indurre l’apoptosi della popolazione leucemica, e infine colpendo nuovi bersagli molecolari.

La possibilità di individuare varianti geniche che influenzano il metabolismo, la tossicità e l’efficacia terapeutica dei vari chemioterapici permetterà di adattare i protocolli di chemioterapia ad ogni singolo paziente. Infine, la possibilità di identificare quelle varianti geniche che espongono il soggetto ad un maggior rischio di contrarre una malattia neoplastica potrà essere sfruttata non solo per consulenza genetica ma anche per guidare le scelte terapeutiche e per migliorare la stratificazione prognostica dei pazienti.

Bibliografia

1. Tiu RV, Gondek LP, O'Keefe CL, et al. New lesions detected by single nucleotide polymorphism array-based chromosomal analysis have important clinical impact in acute myeloid leukemia. *J Clin Oncol.* 2009; 27: 5219-26.
2. Grimwade D, Hills RK, Moorman AV, et al. Refinement of cytogenetic classification in acute myeloid leukemia: determination of prognostic significance of rare recurring chromosomal abnormalities among 5876 younger adult patients treated in the United Kingdom Medical Research Council trials. *Blood.* 2010; 116: 354-65.
3. Yi JH, Huh J, Kim HJ, et al. Adverse prognostic impact of abnormal lesions detected by genome-wide single nucleotide polymorphism array-based karyotyping analysis in acute myeloid leukemia with normal karyotype. *J Clin Oncol.* 2011; 29: 4702-8.
4. Kottaridis PD, Gale RE, Frew ME, et al. The presence of a FLT3 internal tandem duplication in patients with acute myeloid leukemia (AML) adds important prognostic information to cytogenetic risk group and response to the first cycle of chemotherapy: analysis of 854 patients from the United Kingdom Medical Research Council AML 10 and 12 trials. *Blood.* 2001; 98: 1752-9.
5. Schnittger S, Schoch C, Dugas M, et al. Analysis of FLT3 length mutations in 1003 patients with acute myeloid leukemia: correlation to cytogenetics, FAB subtype, and prognosis in the AMLCG study and usefulness as a marker for the detection of minimal residual disease. *Blood.* 2002; 100: 59-66.
6. Wiernik PH. FLT3 inhibitors for the treatment of acute myeloid leukemia. *Clin Adv Hematol Oncol.* 2010; 8: 429-36.
7. Falini B, Martelli MP, Bolli N, et al. Acute myeloid leukemia with mutated nucleophosmin (NPM1): is it a distinct entity? *Blood.* 2011; 117: 1109-20.
8. Marcucci G, Haferlach T, Döhner H. Molecular genetics of adult acute myeloid leukemia: prognostic and therapeutic implications. *J Clin Oncol.* 2011; 29: 475-86.
9. Yan M, Kanbe E, Peterson LF, Boyapati A, Miao Y, Wang Y, Chen IM, Chen Z, Rowley JD, Willman CL, Zhang DE. A previously unidentified alternatively spliced isoform of t(8;21) transcript promotes leukemogenesis. *Nat. Med.* 2006; 12: 945-9.
10. Human Genome Sequencing C. Finishing the euchromatic sequence of the human genome. *Nature* 2004; 431: 931-45.
11. Mardis ER. A decade's perspective on DNA sequencing technologies. *Nature* 2011; 470: 198-203.
12. Ley TJ, Mardis ER, Ding L, et al. DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. *Nature* 2008; 456: 66-72.
13. Mardis ER, Ding L, Dooling DJ, et al. Recurring mutations found by sequencing an acute myeloid leukemia genome. *New Engl J Med.* 2009; 361: 1058-66.
14. Yan H, Parson DW, Jin G, et al. IDH1 and IDH2 mutations in gliomas. *New Engl. J. Med.* 2009; 360: 765-73.

15. Abbas S, Lugthart S, Kavelaars FG, et al. Acquired mutations in the genes encoding IDH1 and IDH2 both are recurrent aberrations in acute myeloid leukemia: prevalence and prognostic value. *Blood*. 2010; 116: 2122-6.
16. Marcucci G, Maharry K, Wu YZ, et al. IDH1 and IDH2 gene mutations identify novel molecular subsets within de novo cytogenetically normal acute myeloid leukemia: a Cancer and Leukemia Group B study. *J Clin Oncol*. 2010; 28: 2348-55.
17. Figueroa ME, Abdel-Wahab O, Lu C, et al. Leukemic IDH1 and IDH2 mutations result in a hypermethylation phenotype, disrupt TET2 function, and impair hematopoietic differentiation. *Cancer Cell*. 2010; 18: 553-567.
18. Ley TJ, Ding L, Walter MJ, et al. DNMT3A mutations in acute myeloid leukemia. *New Engl. J. Med*. 2010; 363: 2424-33.
19. Yamashita Y, Yuan J, Suetake I, et al. Array-based genomic resequencing of human leukemia. *Oncogene*. 2010; 29: 3723-31.
20. Yan XJ, Xu J, Gu ZH, et al. Exome sequencing identifies somatic mutations of DNA methyltransferase gene DNMT3A in acute monocytic leukemia. *Nat Genet*. 2011; 43: 309-15.
21. Welch JS, Westervelt P, Ding L, et al. Use of whole-genome sequencing to diagnose a cryptic fusion oncogene. *JAMA*. 2011; 305: 1577-84.
22. Link DC, Schuettpelz LG, Shen D, et al. Identification of a novel TP53 cancer susceptibility mutation through whole-genome sequencing of a patient with therapy-related AML. *JAMA*. 2011; 305: 1568-76.
23. Garber JE, Offit K. Hereditary cancer predisposition syndromes. *J Clin Oncol*. 2005; 23: 276-92.
24. Huang SJ, Lozano G, Amos CI, et al. Germline p53 mutations in a cohort with childhood sarcoma: sex differences in cancer risk. *Am. J. Hum Gen*. 2003; 72: 975-83.
25. Graubert TA, Mardis ER. Genomics of acute leukaemia. *The Cancer Journal*. 2011; 487-91.

Alterazioni dello *splicing* dell'RNA nelle sindromi mielodisplastiche

Mario Cazzola

Direttore Clinica di Ematologia, IRCCS Policlinico San Matteo, Università di Pavia

Le sindromi mielodisplastiche sono neoplasie ematologiche caratterizzate da ematopoiesi inefficace, deficitaria produzione di cellule ematiche e rischio elevato di evoluzione in leucemia mieloide acuta (1). La loro incidenza annua è superiore a 20/100.000 nella popolazione al di sopra dei 70 anni, ed è in aumento nei paesi occidentali per il progressivo invecchiamento demografico (2).

Si tratta di disordini primitivi del midollo osseo emopoietico caratterizzati da proliferazione clonale di cellule staminali emopoietiche che hanno acquisito alterazioni genetiche.

In particolare, le sindromi mielodisplastiche sono un modello interessante per studiare l'effetto dell'invecchiamento sull'emopoiesi. Per quanto le cause della proliferazione dei cloni mielodisplastici non siano attualmente definite, è plausibile che riflettano il danno progressivo che il DNA delle cellule staminali subisce con l'invecchiamento.

La conseguenza sarebbe una senescenza prematura delle cellule staminali normali e la loro deplezione, alla quale si associa l'emergenza di cloni che hanno un vantaggio selettivo. Il genoma delle cellule mielodisplastiche porterebbe dunque i segni dell'esposizione ambientale ed i danni al DNA acquisiti dalle cellule staminali nel corso della vita.

Il danno midollare determina tipicamente anemia (refrattaria al trattamento), neutropenia e piastrinopenia persistenti (o varie combinazioni di queste citopenie). La storia naturale della malattia, in assenza di trattamento, è caratterizzata da un progressivo aggravamento dell'emopoiesi inefficace, con sviluppo di insufficienza midollare e dei sintomi ad essa correlati, e con un rischio sempre più elevato di evoluzione in leucemia mieloide acuta.

L'unica terapia che può determinare la guarigione completa è attualmente il trapianto allogenico di cellule staminali emopoietiche, che è tuttavia praticabile in una minoranza di pazienti relativamente giovani. Per contro, l'unica terapia che può essere attualmente praticata nella maggioranza dei pazienti è la terapia di supporto, basata sulle trasfusioni di globuli rossi e/o di piastrine e sull'impiego di antibiotici e di farmaci chelanti del ferro.

Classificazione delle sindromi mielodisplastiche e valutazione del rischio mediante modelli prognostici

La diagnosi di sindrome mielodisplastica si basa sulla presenza di citopenia (anemia, neutropenia, piastrinopenia), displasia midollare ed anomalie citogenetiche. La displasia viene valutata sulla base di criteri morfologici, recentemente definiti da esperti della *World Health Organization* (WHO) nella classificazione dei tumori dei tessuti emopoietici e linfoidi (3). La classificazione WHO 2008 è riportata nella Tabella 1. Mentre alcuni criteri morfologici sono semplici, oggettivi e riproducibili (ad esempio, i sideroblasti ad anello), altri sono molto soggettivi e scarsamente riproducibili (ad esempio, la displasia multilineare). Vi è dunque la necessità di definire criteri diagnostici più affidabili, soprattutto nei pazien-

Tab. 1 - Classificazione WHO 2008 delle sindromi mielodisplastiche.

Condizione morbosa	Reperti ematologici (sangue periferico)	Reperti midollari
Anemia refrattaria (RA), neutropenia refrattaria(RN), piastrinopenia refrattaria(RT), citopenia refrattaria con displasia unilineare (RCUD)	Citopenia isolata (anemia, neutropenia, o piastrinopenia), blasti assenti o rari (<1%), talvolta presenza di bi-citopenia.	Displasia unilineare (interessante ≥10% delle cellule di una linea midollare), <5% di blasti, <15% di sideroblasti ad anello.
Anemia refrattaria con sideroblasti ad anello (RARS)	Anemia, assenza di blasti.	Isolata displasia eritroide, <5% di blasti ≥15% di sideroblasti ad anello.
Citopenia refrattaria con displasia multilineare (RCMD)	Citopenia isolata o multipla, blasti assenti o rari (<1%), assenza di corpi di Auer, monociti <1x10 ⁹ /L.	Displasia in ≥10% delle cellule in almeno 2 linee cellulari (eritroide, mieloide, megacariocitaria), <5% di blasti, assenza di corpi di Auer (la percentuale di sideroblasti ad anello è irrilevante).
Anemia refrattaria con eccesso di blasti di tipo 1 (RAEB-1)	Citopenia isolata o multipla, <5% di blasti, assenza di corpi di Auer, monociti <1x10 ⁹ /L (i casi con corpi di Auer e <5% di blasti nel sangue periferico e <10% di blasti nel midollo vanno classificati come RAEB-2).	Displasia unilineare o multilineare, blasti variabili dal 5% al 9%, assenza di corpi di Auer (i casi con corpi di Auer e <5% di blasti nel sangue periferico e <10% di blasti nel midollo vanno classificati come RAEB-2).
Anemia refrattaria con eccesso di blasti di tipo 2 (RAEB-2)	Citopenia isolata o multipla, 5-19% di blasti, occasionali corpi di Auer, monociti <1x10 ⁹ /L.	Displasia unilineare o multilineare, blasti variabili dal 10% al 19%, occasionali corpi di Auer.
Sindrome mielodisplastica non classificabile (MDS-U)	Citopenia isolata o multipla, blasti assenti o rari (<1%).	Displasia in meno del 10% delle cellule di una o più linee cellulari (eritroide, mieloide, megacariocitaria) associata ad anomalia citogenetica ritenuta tipica di una sindrome mielodisplastica, blasti <5%.
Sindrome mielodisplastica associata ad isolata delezione 5q	Anemia, conta piastrinica normale o aumentata, blasti assenti o rari (<1%).	Numero di megacariociti con nuclei ipolobati normale o aumentato, blasti >5%, assenza di corpi di Auer, isolata del(5q).

ti che non hanno anomalie citogenetiche. Per quanto riguarda la definizione prognostica, esistono attualmente diversi modelli prognostici: i più importanti sono il cosiddetto IPSS (4) ed il WPSS (5, 6). Quest'ultimo è riportato nella Tabella 2.

Base molecolare delle sindromi mielodisplastiche

La base molecolare delle sindromi mielodisplastiche è rimasta sostanzialmente sconosciuta fino a poco tempo fa, ma ultimamente sono stati fatti notevoli passi in avanti.

La sindrome mielodisplastica associata a delezione 5q deriva verosimilmente da "haploinsufficienza" di geni che mappano sul cromosoma 5q32-q33, in particolare da ridotta espressione di *RPS14*, miR-145 e miR-146a, e da eccessiva espressione di *TP53* (7-11).

Più recentemente sono stati identificate mutazioni geniche ricorrenti nelle sindromi mielodisplastiche: (12-24) i geni mutanti più importanti sono riportati nella Tabella 3. Questi geni mutanti sono importanti sia per la diagnosi sia per la definizione prognostica.

Varie osservazioni indicano che *TET2* è in grado, per mutazione somatica, di determinare dominanza clonale di cellule staminali emopoietiche e può essere quindi un buon marcatore di clonalità (14). Altri geni mutanti, in particolare

Tab. 2 - Definizione prognostica basata sullo score WPSS.

Calcolo dello score WPSS				
Variabile	Peso delle variabili			
	0	1	2	3
Categoria WHO	RA, RARS, sindrome mielodisplastica con isolata delezione 5q	RCMD	RAEB-1	RAEB-2
Cariotipo*	Favorevole	Intermedio	Sfavorevole	-
Anemia grave (Hb <9 g/dL nei maschi o <8 g/dL nelle femmine)	Assente	Presente	-	-
Definizione del gruppo di rischio del singolo paziente				
Gruppo di rischio	Score	Sopravvivenza mediana attesa	Rischio atteso di evoluzione leucemica	
Molto basso	0	>10 anni	<10% a 15 anni	
Basso	1	>5 anni	10-20% a 5 anni	
Intermedio	2	~4 anni	30-40% a 5 anni	
Alto	3-4	~2 anni	30% a 3 anni	
Molto alto	5-6	~1 anno		
	>50% ad 1 anno			

*Favorevole: cariotipo normale, -Y, del(5q), del(20q); sfavorevole: cariotipo complesso, alterazioni del cromosoma 7; intermedio: altre anomalie cromosomiche.

Tab. 3 - Geni mutanti di rilevanza diagnostica e prognostica nelle sindromi mielodisplastiche.**Geni che codificano proteine con funzioni diverse nel metabolismo cellulare**

- TET2, tet oncogene family member 2, cromosoma 4q24
- ASXL1, additional sex combs like 1, cromosoma 20q11.1
- CBL, Cas-Br-M (murine) ecotropic retroviral transforming, cromosoma 11q23.3
- ETV6, ETS translocation variant 6, cromosoma 12p13
- EZH2, enhancer of zeste homolog 2, cromosoma 7q35-q36
- RUNX1, runt-related transcription factor 1, cromosoma 21q22.3
- TP53, tumor protein p53, cromosoma 17p13.1

Geni che codificano fattori di splicing dell'RNA

- SF3B1, splicing factor 3b, subunit 1, 155kDa, cromosoma 2q33.1
- SRSF2, serine/arginine-rich splicing factor 2, cromosoma 17q25.1
- U2AF1, U2 small nuclear RNA auxiliary factor 1, cromosoma 21q22.3
- ZRSR2, RNA-binding motif and serine/arginine rich 2, cromosoma Xp22.1

EZH2, *RUNX1* e *TP53*, sono invece responsabili di progressione di malattia, ed in particolare dell'evoluzione delle sindromi mielodisplastiche in leucemia mieloida acuta (16).

Mutazioni somatiche di fattori di splicing dell'RNA

Utilizzando un approccio di “*whole exome sequencing*” abbiamo recentemente identificato mutazioni somatiche di *SF3B1*, un gene che codifica per un fattore di *splicing* dell'RNA, in diverse categorie WHO di sindrome mielodisplastica, ma in prevalenza in pazienti con sideroblasti ad anello (18). Uno studio parallelo di autori giapponesi ha documentato l'elevata incidenza di mutazioni di geni che regolano lo *splicing* dell'RNA in pazienti con sindrome mielodisplastica o neoplasia/mielodisplastica mieloproliferativa (17). Queste osservazioni indicano che anomalie dello *splicing* dell'RNA giocano verosimilmente un ruolo importante nella patogenesi di questi disordini mieloidi.

È inoltre interessante il fatto che i diversi geni mutanti dello *splicing* dell'RNA possono associarsi a fenotipi distinti ed avere significato clinico diverso. Le mutazioni somatiche di *SF3B1* sono strettamente correlate ai sideroblasti ad anello, si associano in analisi multivariata a una prognosi favorevole e migliorano il potere predittivo dell'IPSS (19). Per contro le mutazioni somatiche di *SRSF2* si riscontrano soprattutto nei pazienti con il quadro clinico della leucemia mielomonocitica cronica e si associano a prognosi sfavorevole (17, 22, 24). Infine, le mutazioni somatiche di *U2AF1* si associano ad un rischio elevato di evoluzione leucemica (21, 24).

Conclusioni

Nell'era della medicina genomica non è sorprendente che nel giro di un paio di anni si sia passati dalla pressoché totale ignoranza dei meccanismi patogenetici delle sindromi mielodisplastiche ad un ricchissimo bagaglio di conoscenze molecolari. Il passo in avanti più significativo è indubbiamente rappresentato dall'identificazione di mutazioni somatiche di geni che regolano lo *splicing* dell'RNA,

una via metabolica che non si pensava potesse essere interessata da mutazioni geniche ed essere coinvolta nella patogenesi delle neoplasie.

Le ricerche in corso dovrebbero consentire alla comunità scientifica ematologica di definire ulteriormente il ruolo dei geni mutanti, sia per quanto concerne il fenotipo delle sindromi mielodisplastiche, sia per quanto riguarda la loro progressione verso la leucemia mieloide acuta. Dovrebbero inoltre consentire di definire il significato diagnostico e prognostico delle mutazioni geniche identificate e di integrare queste informazioni nei processi diagnostici e nei modelli prognostici attuali, migliorando il “*clinical decision making*”.

Bibliografia

1. Cazzola M, Malcovati L. Myelodysplastic syndromes - coping with ineffective hematopoiesis. *N Engl J Med.* 2005; 352 (6): 536-8.
2. Tefferi A, Vardiman JW. Myelodysplastic syndromes. *N Engl J Med.* 2009; 361 (19): 1872-85.
3. Swerdlow SH, Campo E, Harris NL, Jaffe ES, Pileri SA, Stein H, et al. WHO classification of tumours of haematopoietic and lymphoid tissues. Lyon: IARC, 2008.
4. Greenberg P, Cox C, LeBeau MM, Fenaux P, Morel P, Sanz G, et al. International scoring system for evaluating prognosis in myelodysplastic syndromes. *Blood.* 1997; 89 (6): 2079-88.
5. Malcovati L, Germing U, Kuendgen A, Della Porta MG, Pascutto C, Invernizzi R, et al. Time-dependent prognostic scoring system for predicting survival and leukemic evolution in myelodysplastic syndromes. *J Clin Oncol.* 2007; 25 (23): 3503-10.
6. Della Porta MG, Malcovati L. Clinical relevance of extra-hematologic comorbidity in the management of patients with myelodysplastic syndrome. *Haematologica.* 2009; 94 (5): 602-6.
7. Ebert BL, Pretz J, Bosco J, Chang CY, Tamayo P, Galili N, et al. Identification of RPS14 as a 5q- syndrome gene by RNA interference screen. *Nature.* 2008; 451 (7176): 335-9.
8. Pellagatti A, Hellstrom-Lindberg E, Giagounidis A, Perry J, Malcovati L, Della Porta MG, et al. Haploinsufficiency of RPS14 in 5q- syndrome is associated with deregulation of ribosomal- and translation-related genes. *Br J Haematol.* 2008; 142 (1): 57-64.
9. Starczynowski DT, Kuchenbauer F, Argiropoulos B, Sung S, Morin R, Muranyi A, et al. Identification of miR-145 and miR-146a as mediators of the 5q- syndrome phenotype. *Nat Med.* 2010; 16 (1): 49-58.
10. Barlow JL, Drynan LF, Hewett DR, Holmes LR, Lorenzo-Abalde S, Lane AL, et al. A p53-dependent mechanism underlies macrocytic anemia in a mouse model of human 5q- syndrome. *Nat Med.* 2010; 16 (1): 59-66.
11. Pellagatti A, Marafioti T, Paterson JC, Barlow JL, Drynan LF, Giagounidis A, et al. Induction of p53 and up-regulation of the p53 pathway in the human 5q- syndrome. *Blood.* 2010; 115 (13): 2721-3.

12. Delhommeau F, Dupont S, Della Valle V, James C, Trannoy S, Masse A, et al. Mutation in TET2 in myeloid cancers. *N Engl J Med*. 2009; 360 (22): 2289-301.
13. Langemeijer SM, Kuiper RP, Berends M, Knops R, Aslanyan MG, Massop M, et al. Acquired mutations in TET2 are common in myelodysplastic syndromes. *Nat Genet*. 2009; 41 (7): 838-42.
14. Smith AE, Mohamedali AM, Kulasekararaj A, Lim Z, Gaken J, Lea NC, et al. Next-generation sequencing of the TET2 gene in 355 MDS and CMML patients reveals low-abundance mutant clones with early origins, but indicates no definite prognostic value. *Blood*. 2010; 116 (19): 3923-32.
15. Bejar R, Levine R, Ebert BL. Unraveling the molecular pathophysiology of myelodysplastic syndromes. *J Clin Oncol*. 2011; 29 (5): 504-15.
16. Bejar R, Stevenson K, Abdel-Wahab O, Galili N, Nilsson B, Garcia-Manero G, et al. Clinical effect of point mutations in myelodysplastic syndromes. *N Engl J Med*. 2011; 364 (26): 2496-506.
17. Yoshida K, Sanada M, Shiraishi Y, Nowak D, Nagata Y, Yamamoto R, et al. Frequent pathway mutations of splicing machinery in myelodysplasia. *Nature*. 2011; 478 (7367): 64-9.
18. Papaemmanuil E, Cazzola M, Boultonwood J, Malcovati L, Vyas P, Bowen D, et al. Somatic SF3B1 mutation in myelodysplasia with ring sideroblasts. *N Engl J Med*. 2011; 365 (15): 1384-95.
19. Malcovati L, Papaemmanuil E, Bowen DT, Boultonwood J, Della Porta MG, Pascutto C, et al. Clinical significance of SF3B1 mutations in myelodysplastic syndromes and myelodysplastic/myeloproliferative neoplasms. *Blood*. 2011; 118 (24): 6239-46.
20. Patnaik MM, Lasho TL, Hodnefield JM, Knudson RA, Ketterling RP, Garcia-Manero G, et al. SF3B1 mutations are prevalent in myelodysplastic syndromes with ring sideroblasts but do not hold independent prognostic value. *Blood*. 2012; 119 (2): 569-72.
21. Graubert TA, Shen D, Ding L, Okeyo-Owuor T, Lunn CL, Shao J, et al. Recurrent mutations in the U2AF1 splicing factor in myelodysplastic syndromes. *Nat Genet*. 2012; 44 (1): 53-7.
22. Makishima H, Visconte V, Sakaguchi H, Jankowska AM, Abu Kar S, Jerez A, et al. Mutations in the spliceosome machinery, a novel and ubiquitous pathway in leukemogenesis. *Blood*. 2012.
23. Damm F, Kosmider O, Gelsi-Boyer V, Renneville A, Carbuccia N, Hidalgo-Curtis C, et al. Mutations affecting mRNA splicing define distinct clinical phenotypes and correlate with patient outcome in myelodysplastic syndromes. *Blood*. 2012.
24. Thol F, Kade S, Schlarmann C, Loeffeld P, Morgan M, Krauter J, et al. Frequency and prognostic impact of mutations in SRSF2, U2AF1, and ZRSR2 in patients with myelodysplastic syndromes. *Blood*. 2012.

La terapia di prima linea della LMC Ph1 positiva con imatinib può essere migliorata?

Giuseppe Saglio

Direttore Divisione di Medicina Interna e Ematologia,
Azienda Sanitaria Ospedaliero Universitaria San Luigi, Orbassano (TO)

La Leucemia Mieloide Cronica (LMC) è un disordine mieloproliferativo cronico che colpisce la cellula staminale emopoietica (1-3). Clinicamente la LMC è caratterizzata nel sangue periferico da una leucocitosi di grado variabile, talora anche molto elevata, con la presenza di progenitori mieloidi immaturi quali promielociti, mielociti e metamielociti. La LMC è una malattia tipicamente multifasica. La storia naturale della patologia prevede tuttavia che, dopo un tempo variabile da pochi mesi a molti anni (in media 3-4 anni), la fase cronica si trasformi inevitabilmente in una fase acuta, denominata crisi blastica, che possiede le caratteristiche del blocco maturativo tipico delle leucemie acute. Fin dal 1960 la malattia è stata associata alla traslocazione t(9;22), che da origine al cromosoma Philadelphia (Ph) e, nei primi anni '80, è stato possibile identificare i due geni coinvolti nella traslocazione: il gene ABL sul cromosoma 9 e il gene BCR sul cromosoma 22 (4). I due geni vanno a formare un nuovo gene di fusione, BCR/ABL, trascritto in mRNA e tradotto in una nuova proteina di 210 kd, denominata P210, dotata di attività tirosin-chinasica costitutiva (5). Diversi modelli sperimentali hanno permesso di confermare il ruolo fondamentale del prodotto di BCR/ABL nella patogenesi della LMC (6). Negli ultimi 20-30 anni la terapia della LMC ha subito almeno tre momenti di grande trasformazione, dovute prima all'introduzione del trapianto di midollo allogenico, successivamente dell'interferone ricombinante e soprattutto, negli anni più recenti, dell'imatinib, inibitore della tirosin-chinasi BCR/ABL, che è considerato anche il capostipite dei farmaci della cosiddetta terapia "molecolarmente mirata" (targeted therapy) (3). Questo ultimo farmaco ha infatti permesso di cambiare radicalmente la prognosi dei pazienti con LMC e di portare addirittura le aspettative di vita di questi pazienti molto vicine a quelle della popolazione di controllo di pari età (7).

Nello studio IRIS erano stati originariamente randomizzati 553 pazienti nel braccio dei pazienti trattati con 400 mg di imatinib al giorno. La percentuale di MCyR e CCyR ad un anno sono state rispettivamente del 85 e del 69% e a 8 anni sono state rispettivamente dell'89% e dell'83% (percentuali cumulative) (7, 8). La sopravvivenza libera da malattia a 8 anni è stata dell'81% e la sopravvivenza libera da progressione a fase accelerata o crisi blastica è stata del 92%; la sopravvi-

venza globale è stata dell'85% e del 93%, escludendo i decessi per cause non correlate alla malattia. La percentuale di progressione a fase accelerata o crisi blastica che si è verificata in circa l'8% dei casi nei primi quattro anni di follow-up, si è nel tempo progressivamente ridotta.⁷

Nonostante l'indubbia efficacia della terapia con imatinib, una percentuale ancora abbastanza consistente di pazienti in trattamento non risponde adeguatamente alla terapia o perde la risposta che aveva guadagnato. Per circa un terzo di questi pazienti si apre tuttavia la possibilità di migliorare sostanzialmente i risultati ottenuti con la terapia standard (9). Le strategie recentemente esplorate in ambito internazionale con studi di fase II o III, prospettici a braccio singolo o randomizzati, includono regimi basati su imatinib ad alte dosi o in associazione con chemioterapia o interferone, oppure la somministrazione in prima linea degli inibitori di seconda generazione (imatinib, nilotinib, bosutinib).

Per quanto riguarda gli inibitori di seconda generazione, sulla scorta dei risultati promettenti ottenuti come terapia di seconda linea, dasatinib e nilotinib sono stati valutati sia in studi di fase II che in studi di fase III.

È stato condotto uno studio in aperto (ENESTnd), randomizzato, multicentrico di fase III per determinare l'efficacia di nilotinib verso imatinib in 846 pazienti adulti con LMC Ph+ di nuova diagnosi in fase cronica (10). I pazienti sono stati randomizzati 1:1:1 per ricevere sia nilotinib 300 mg due volte al giorno (n=282), nilotinib 400 mg due volte al giorno (n=281) o imatinib 400 mg/die (n=283). L'endpoint primario di efficacia era la risposta molecolare maggiore (MMR) a 12 mesi. La MMR è stata definita come percentuale di Bcr-Abl/Abl $\leq 0,1\%$ secondo la scala internazionale mediante RQ-PCR. Il tasso di MMR a 12 mesi era superiore per nilotinib 300 mg due volte al giorno e per nilotinib 400 mg due volte al giorno rispetto a imatinib 400 mg/die (44,3% verso 22,3%, $p < 0,0001$ e 42,7% verso 22,3%, $p < 0,0001$, rispettivamente). Le differenze si sono mantenute statisticamente significative anche dopo 36 mesi di follow-up. Le differenze tra nilotinib ed imatinib sono state statisticamente significati ve a tutti i time-points stabiliti, dimostrando che nilotinib anticipa il raggiungimento della remissione molecolare maggiore. Sulla base delle conoscenze fino ad ora acquisite il precoce raggiungimento della remissione molecolare maggiore è un obiettivo estremamente importante, in quanto associato ad un minor rischio di progressione della malattia, e, in alcuni studi, addirittura ad una maggiore sopravvivenza globale. Infatti, c'è stata una differenza statisticamente significativa nella progressione alla fase accelerata o alla crisi blastica tra nilotinib e imatinib, a favore di nilotinib.

Analogamente, è stato condotto uno studio clinico internazionale, denominato Dasision, in pazienti con LMC in fase cronica di nuova diagnosi, randomizzati per ricevere dasatinib 100 mg una volta al giorno o imatinib 400 mg una volta al giorno (11). L'endpoint primario di questo studio era il tasso di CCyR confermate a 12 mesi. Un totale di 519 pazienti sono stati randomizzati in un gruppo di trattamento: 259 con dasatinib e 260 con imatinib. Le percentuali di CCyR nei gruppi di trattamento dasatinib e imatinib, rispettivamente, entro 3 mesi (54% e 30%), 6 mesi (70% e 56%) e 9 mesi (75% e 63%) sono state statisticamente molto significative a favore del dasatinib e consistenti con l'endpoint primario. Anche le percentuali

di MMR e di CMR nei gruppi di trattamento sono state molto maggiori con dasatinib che con imatinib e la trasformazione verso la fase accelerata o blastica si è verificata meno frequentemente con pazienti trattati con dasatinib. Questo dato conferma che, quando impiegati come terapia di prima linea, gli inibitori di seconda generazione si dimostrano più efficaci dell'imatinib e sono in grado di prevenire una quota consistente di quegli episodi di progressione precoce che possono ancora verificarsi in particolare nei primi mesi di terapia con imatinib.

Sulla base di questi risultati e in base al fatto che il trattamento con questi inibitori si è dimostrato ben tollerato, sia dasatinib che nilotinib sono stati approvati dall'FDA e dall'EMA per l'utilizzo in prima linea nel trattamento dei pazienti con LMC in fase cronica e sono in via di commercializzazione con questa indicazione in Italia e in molti paesi del mondo.

Bibliografia

1. Tura S, Baccarani M, Zaccaria A. Chronic myeloid leukemia. *Haematologica*. 1986; 71: 169-76.
2. Melo JV, Hughes TP, Apperley JF. Chronic myeloid leukemia. *Hematology Am Soc Hematol Educ Program*. 2003; 132-52.
3. Cortes JE, Talpaz M, Kantarjian H. Chronic myelogenous leukemia: a review. *Am J Med*. 1996; 100: 555-70.
4. Groffen J, Stephenson JR, Heisterkamp N, de Klein A, Bartram CR, Grosveld G. Philadelphia chromosomal breakpoints are clustered within a limited region, bcr, on chromosome 22. *Cell*. 1984; 36: 93-9.
5. Konopka JB, Watanabe SM, Witte ON. An alteration of the human c-abl protein in K562 leukemia cells unmasks associated tyrosine kinase activity. *Cell* 1984; 37: 1035-42.
6. Daley GQ, Van Etten RA, Baltimore D. Induction of chronic myelogenous leukemia in mice by the P210bcr/abl gene of the Philadelphia chromosome. *Science*. 1990; 247: 824-30.
7. Deininger M, O'Brien SG, Guilhot F, et al. International Randomized Study of Interferon Vs STI571 (IRIS) 8-Year Follow up: Sustained Survival and Low Risk for Progression or Events in Patients with Newly Diagnosed Chronic Myeloid Leukemia in Chronic Phase (CML-CP) Treated with Imatinib. *Blood*. 2009; 114: 462.
8. Druker BJ, Guilhot F, O'Brien SG, et al. Five-year follow-up of patients receiving imatinib for chronic myeloid leukemia. *N Engl J Med*. 2006; 355: 2408-17.
9. Jabbour E, Cortes JE, Kantarjian HM. Suboptimal response to or failure of imatinib treatment for chronic myeloid leukemia: what is the optimal strategy? *Mayo Clin Proc*. 2009; 84: 161-9.
10. Saglio G, Kim DW, Issaragrisil S, et al. Nilotinib versus imatinib for newly diagnosed chronic myeloid leukemia. *N Engl J Med*. 2010; 362: 2251-9.
11. Kantarjian H, Shah NP, Hochhaus A, et al. Dasatinib versus imatinib in newly diagnosed chronic-phase chronic myeloid leukemia. *N Engl J Med*. 2010; 362: 2260-70.

Basi molecolari delle malattie mieloproliferative croniche Ph¹ negative

Alessandro Maria Vannucchi

Unità Funzionale di Ematologia di Ematologia, Azienda Ospedaliera Universitaria Careggi, Firenze

Le neoplasie mieloproliferative croniche (MPN) negative per il cromosoma Philadelphia/riarrangiamento BCR/ABL - termine con cui, nella classificazione WHO del 2008 (1, 2), sono identificate le malattie mieloproliferative originariamente descritte da William Dameshek nel 1951 (3) - comprendono la policitemia vera (PV), la trombocitemia essenziale (ET) e la mielofibrosi primaria (PMF). Si tratta di patologie che derivano da una cellula staminale/progenitore multipotente che, a seguito di uno o più eventi trasformanti, ha acquisito indipendenza e/o ipersensibilità ai normali meccanismi di regolazione ad opera delle citochine emopoietiche (4). Le MPNs si manifestano con un'iperproduzione di cellule mature delle tre serie mieloidi, prevalentemente la serie eritroide nella PV e quella megacariocitaria nella ET, mentre nella PMF si possono avere alterazioni variabili delle cellule ematiche; la presenza di un difetto maturativo a carico dei megacariociti è probabilmente alla base delle caratteristiche alterazioni del microambiente midollare con fibrosi, osteosclerosi e neoangiogenesi (5), e della abnorme circolazione di cellule CD34+ nel sangue circolante (6).

La patogenesi molecolare delle MPN è rimasta sconosciuta sino al 2005, quando quattro gruppi di ricerca, indipendentemente e con approcci diversi, hanno identificato la presenza di una mutazione puntiforme, valina → fenilalanina nel codone 617 (V617F; G>T, nt 1849) del gene *JAK2* (7-10), dopo che R. Kralovics e R. Skoda avevano dimostrato che il 30% dei pazienti con PV presenta una perdita di eterozigotità (LOH) a livello del 9p, regione in cui mappa *JAK2* (11); la condizione di LOH corrisponde alla presenza della mutazione di *JAK2* in forma omozigote, per un processo di ricombinazione mitotica, che è alquanto rara nella ET rispetto a PV e PMF (Fig. 1). La presenza di cellule eterozigoti e omozigoti in variabile combinazione rende ragione delle differenze della carica allelica della mutazione che si osservano nei pazienti con le diverse forme di MPN. La mutazione *JAK2*V617F si ritrova in almeno il 95% delle forme con il fenotipo clinico della PV e circa il 60% delle ET e PMF; il riscontro della mutazione rappresenta pertanto uno dei criteri diagnostici maggiori delle MPN secondo la classificazione WHO (Tab. 1). La mutazione V617F si ritrova anche nel 30-50% dei pazienti con trombosi splancnica e nella anemia refrattaria con sideroblasti ad anello e marcata trombocitosi (RARS-T); è molto rara nelle altre malattie mieloprolifera-

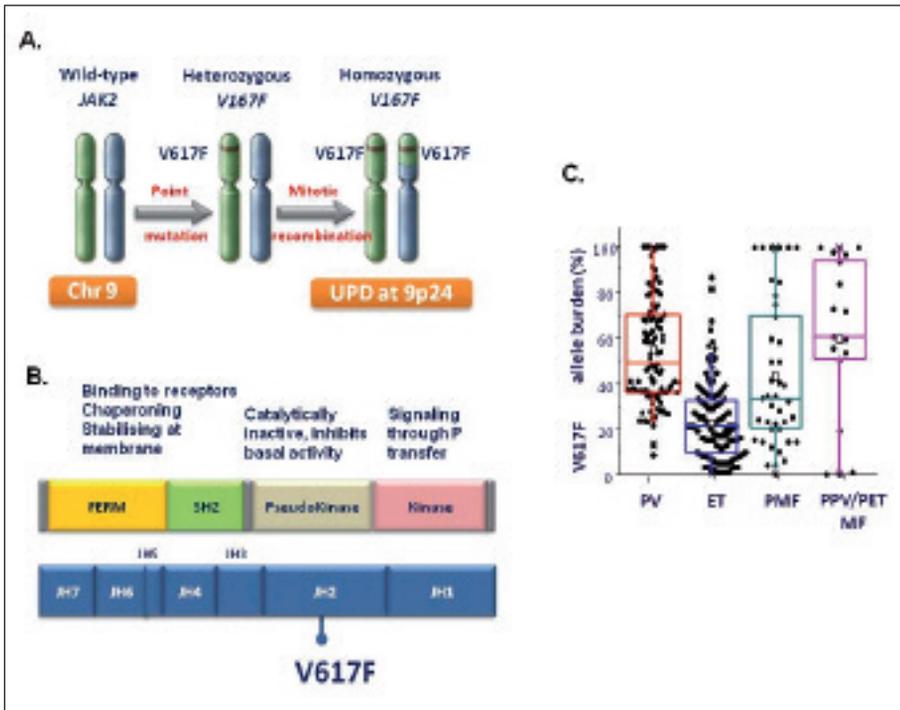


Fig. 1 - La mutazione di JAK2V617F si può riscontrare in forma eterozigote o omozigote a seguito di un processo di ricombinazione mitotica che dà luogo ad una condizione di disomia uni parentale a livello della regione 9p (A). Struttura molecolare di JAK2; la mutazione V617F si localizza nel domain pseudo-cinasi della proteina (B). La carica allelica, che può essere misurata nei granulociti del sangue, si esprime come il rapporto tra l'allele mutato e quello wild-type; pazienti con ET hanno valori mediamente più bassi rispetto a PV e PMF (C).

tive quali la leucemia conica mielomonocitica, mastocitosi sistemica ed ipereosinofilie, o nelle forme di MPN/MDS. Può essere riscontrata in associazione a BCR/ABL nel 2,5% delle leucemie mieloidi croniche (12).

JAK2 è una tirosino-chinasi, parte di una più ampia famiglia di proteine JAKs (JAK1, JAK2, JAK3 e Tyk2), associata alla regione juxta-membrana di numerosi recettori dimerici per citochine emopoietiche, inclusi quelli per l'eritropoietina (Epo), trombopoietina (TPO), G-CSF-R, ed altri quali il recettore per interferone-gamma. A seguito del legame della citochina ai recettori omodimerici, JAK2 attiva la segnalazione intracellulare attraverso le vie di STAT-3 e -5, RAS-MAPK, e PI-3K-Akt, che complessivamente regolano la sopravvivenza e la proliferazione delle cellule mieloidi e ne modulano l'espressione genica. La presenza della mutazione V617F, che si trova nel domain pseudo-chinasi JH2 della proteina, comporta la costitutiva attivazione di JAK2 e delle vie di segnalazione a valle; il meccanismo con il quale la mutazione porta alla perdita del normale controllo auto inibitorio di JH2 sul domain attivo JH1 è da definire ancora con esattezza (13). Recenti dati suggeriscono che, oltre all'attività di segnalazione citoplasmatica, JAK2V617F possa localizzarsi a livello nucleare regolando l'espressione di

Tab. 1 - Criteri diagnostici per le MPN secondo la classificazione WHO del 2008 (2).

	Polycythemia vera	Essential Thrombocythemia	Primary myelofibrosis
Major criteria	1. Hb >18.5 g/dL (men) or 16.5 g/dL (women) or Hb or Hct >99 th percentile of reference range for age, sex, or altitude of residence or Hb >17 g/dL (men) or >15 g/dL (women) if associated with a sustained increase of ≥ 2 g/dL from baseline that can not be attributed to correction of iron deficiency or Elevated red cell mass >25% above mean normal predicted value 2. Presence of JAK2V617F or similar mutation	1. Platelet count $\geq 450 \times 10^9/L$ 2. Megakaryocyte proliferation with large and mature morphology. No or little granulocyte or erythroid proliferation. 3. Not meeting WHO criteria for CML, PV, PMF, MDS or other myeloid neoplasm 4. Demonstration of JAK2V617F or other clonal marker or no evidence of reactive thrombocytosis	1. Megakaryocyte proliferation and atypia accompanied by either reticulin and/or collagen fibrosis, or In the absence of reticulin fibrosis, the megakaryocyte changes must be accompanied by increased marrow cellularity, granulocytic proliferation and often decreased erythropoiesis (i.e. pre-fibrotic PMF) 2. Not meeting WHO criteria for CML, PV, MDS, or other myeloid neoplasm 3. Demonstration of JAK2V617F or other clonal marker or no evidence of reactive marrow fibrosis
Minor criteria	1. BM trilineage myeloproliferation 2. Subnormal serum Epo level 3. EEC growth		1. Leukoerythroblastosis 2. Increased serum LDH 3. Anemia 4. Palpable splenomegaly
Diagnostic combinations	Both major criteria +1 minor criterion or First major criterion +2 minor criteria	All four criteria must be met	All three major criteria + two minor criteria

geni bersaglio attraverso la modulazione dello stato di fosforilazione di HP1a e di PRMT5 (14). L'identificazione della mutazione *JAK2V617F*, che ha richiamato l'attenzione sulle alterazioni della via JAK/STAT, è stata seguita da altre scoperte di mutazioni in proteine che hanno un ruolo in questa stessa cascata di segnalazione. Mutazioni dell'esone 12 di *JAK2* (14), funzionalmente equivalenti alla V617F dell'esone 14, sono state raramente riscontrate in soggetti con fenotipo di PV (16). LNK, che appartiene alla famiglia delle "adaptor proteins", interviene fisiologicamente nella regolazione negativa dell'attivazione di JAK2 attraverso il suo domain SH2, inibendo pertanto le vie di segnalazione attivate da EPO, TPO e altre citochine emopoietiche. Mutazioni nell'esone 2 di LNK sono state descritte per lo più in pazienti con PMF e ET *JAK2V617F*-negative e in alcuni pazienti con eritrocitosi idiopatica (17-19), con forse una maggiore frequenza nelle fasi di trasformazione leucemica (13%) (20).

La proteina c-CBL (Casita B-cell lymphoma) svolge attività di ubiquitin-ligase, ed è quindi coinvolta nella regolazione negativa di recettori tirosino-chinasici ed altri compreso JAK2 e MPL, mediandone l'ubiquitinazione a livello degli endosomi. Il gene *c-CBL* è stato riscontrato mutato in numerose neoplasie mieloidi e

in sindromi mielodisplastiche, ma anche in circa il 6% dei pazienti con PMF, preferenzialmente associato con l'evoluzione di malattia e la trasformazione leucemica (21). Tutte queste mutazioni, pertanto, si aggiungono alla lista di proteine mutate, dopo JAK2, che coinvolgono la via di segnalazione citochinica attivata particolarmente dai recettori di EPO e TPO.

Negli ultimi due anni sta crescendo l'interesse per un altro gruppo di proteine che sono state scoperte essere mutate in una proporzione variabile tra il 3 e il 20% delle MPNs così come in soggetti con MDS, ovvero proteine che intervengono nella regolazione epigenetica dell'espressione genica (Fig. 2) (22, 23). Queste includono TET2, ASXL1, EZH2 e altre proteine del complesso PRC2 (polycomb repressive complex-2), DNMT3A, IDH1 e IDH2. *TET2* è stato scoperto dal gruppo di W. Vainchenker in un paziente con alterazione della banda 4q24 (24). La funzione nota delle proteine TET (ne sono note 3) è di controllare la idrossilazione della 5-metilcitosina, che è concentrata in regioni attivamente trascritte della cromatina, essendo coinvolta anche durante l'embriogenesi. L'analisi sequenziale delle mutazioni di TET2 durante il decorso della malattia, e l'analisi di singoli progenitori, ha indicato che le mutazioni di TET2 possono sia precedere che seguire l'acquisizione della mutazione *JAK2V617F*, e quindi a stretto rigore di termine non rappresentano necessariamente la mutazione iniziante la neoplasia mieloproliferativa (25). Le mutazioni di *TET2* sono generalmente in forma eterozigote.

Il gene *EZH2* codifica per la sub-unità catalitica del PRC2, coinvolto nella trimetilazione della lisina 27 dell'istone H3 (H3K27). H3K27 è un marcatore dello stato inattivo della cromatina. *EZH2* si associa anche con la DNA-metiltransferasi, e mutazioni di *EZH2* e *DNMT3A* appaiono mutualmente esclusive, sebbene i dati si riferiscano a piccole casistiche. Mutazioni di *EZH2* sono state descritte soprattutto in pazienti con PMF e forme secondarie a PV e ET (26). Uno studio recente ha evidenziato un valore prognostico negativo per le mutazioni di *EZH2* per la sopravvivenza totale e libera da leucemia in soggetti con PMF (27).

Mutazioni di *ASXL1*, localizzato alla banda 20q11.21, sono state riscontrate nel 15-20% dei pazienti con PMF, e molto più raramente nelle altre forme mielopro-

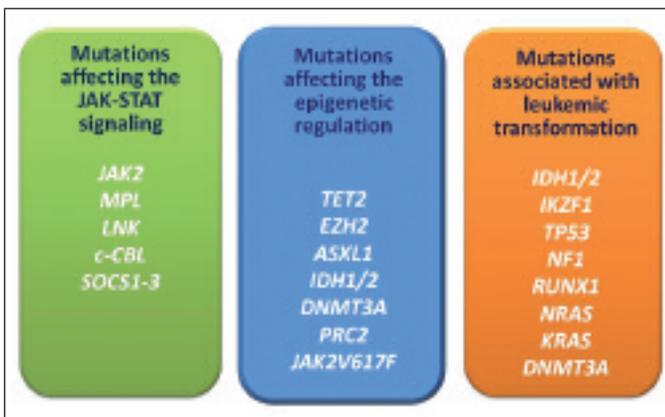


Fig. 2 - Classificazione funzionale delle mutazioni riscontrate in pazienti con neoplasie mieloproliferative croniche. Si rimanda al testo per dettagli.

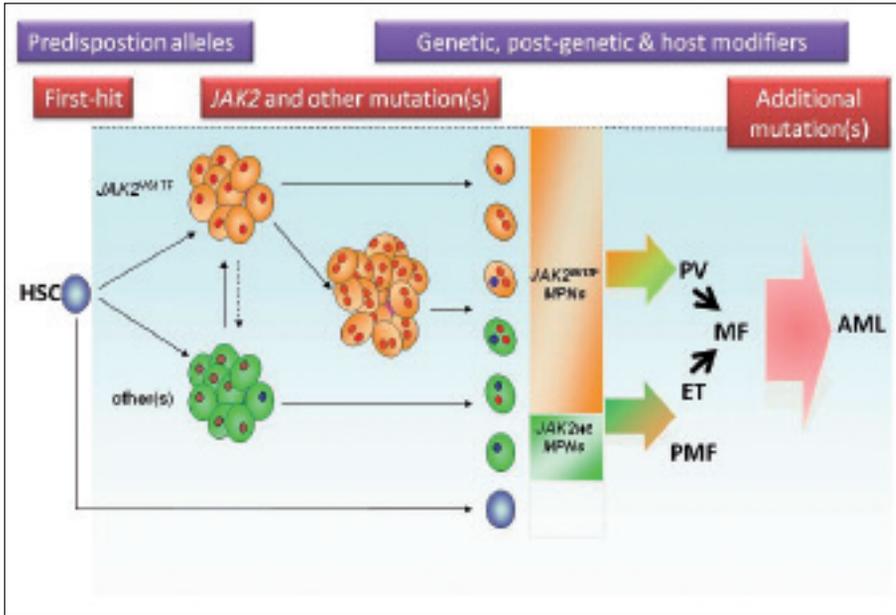


Fig. 3 - La figura illustra le diverse componenti che hanno un ruolo patogenetico nelle MPN. Alleli costituzionali di predisposizione (come l'aplotipo 46/1 in *JAK2*) possono facilitare l'acquisizione della mutazione *JAK2V617F*, *JAK2*-esone 12 e *MPL*, e forse di altre ancora, in una cellula staminale/progenitrice multi potente che ha subito un primo hit mutazionale (che rimane sconosciuto). Alcune delle mutazioni note, come quelle di *JAK2*, danno luogo all'espansione del clone e al fenotipo della malattia, in variabile associazione e combinazione con altri fattori dell'ospite (sesso, riserve di ferro, etc); restano ancora da scoprire le mutazioni responsabili del fenotipo nelle ET e PMF che sono *JAK2/MPL* negative. (modificata da (31)). L'evoluzione di PV e ET a PPV-/PET-MF, e la trasformazione a leucemia acuta, sono eventi ancora non ben definiti sotto il profilo molecolare.

liferative (28-30). *ASXL1* codifica per una proteina detta "Additional Sex Combs-Like protein-1" che è una dei tre omologhi del gene di *Drosophila* Additional Sex Comb (*Asx*) gene; il nome deriva dal fatto che la delezione di *Asx* nella mosca dà luogo a difetto di sviluppo mediato dalla disregolazione di geni homeobox. *ASL1* fa parte di diversi complessi multimolecolari che si legano a specifiche regioni cromatiniche, inducendo o reprimendo la trascrizione a seconda dei complessi nei quali si viene a trovare.

La patogenesi molecolare delle MPN (Fig. 3) appare sempre più complessa, e numerosi sono gli aspetti che ancora necessitano di una definizione e chiarimenti, primo fra tutti se esista, e quale sia, una comune lesione molecolare alla base di queste malattie, come predetto da W. Dameshek. Rimane ancora da comprendere il ruolo delle mutazioni anche nella progressione della malattia, quale sia il loro impatto prognostico e se possano rappresentare utili bersagli per una terapia mirata. Dall'essere malattie cenerentola, le MPN sono divenute un fertile campo di studi, tant che dopo appena 6 anni dalla descrizione della prima anomalia molecolare, la mutazione *JAK2V617F*, l'FDA ha approvato il primo farmaco per il trattamento della mielofibrosi.

Bibliografia

1. Tefferi A, Thiele J, Orazi A, et al. Proposals and rationale for revision of the World Health Organization diagnostic criteria for polycythemia vera, essential thrombocythemia, and primary myelofibrosis: recommendations from an ad hoc international expert panel. *Blood*. 2007; 110 (4): 1092-7.
2. Swerdlow SH, Campo, E, Harris, NL, Jaffe, ES, Pileri, SA, Stein, H, Thiele, J, Vardiman, JW. ed WHO classification of Tumors of Haematopoietic and Lymphoid Tissues. Lyon: International Agency for Research on Cancer; 2008.
3. Dameshek W. Some speculations on the myeloproliferative syndromes. *Blood*. 1951; 6 (4): 372-5.
4. Vannucchi AM, Guglielmelli P, Tefferi A. Advances in understanding and management of myeloproliferative neoplasms. *CA Cancer J Clin*. 2009; 59 (3): 171-91.
5. Lataillade JJ, Pierre-Louis O, Hasselbalch HC, et al. Does primary myelofibrosis involve a defective stem cell niche? From concept to evidence. *Blood*. 2008; 112 (8): 3026-35.
6. Barosi G, Viarengo G, Pecci A, et al. Diagnostic and clinical relevance of the number of circulating CD34(+) cells in myelofibrosis with myeloid metaplasia. *Blood*. 2001; 98 (12): 3249-55.
7. James C, Ugo V, Le Couedic JP, et al. A unique clonal JAK2 mutation leading to constitutive signalling causes polycythaemia vera. *Nature*. 2005; 434 (7037): 1144-8.
8. Baxter EJ, Scott LM, Campbell PJ, et al. Acquired mutation of the tyrosine kinase JAK2 in human myeloproliferative disorders. *Lancet*. 2005; 365 (9464): 1054-61.
9. Levine RL, Wadleigh M, Cools J, et al. Activating mutation in the tyrosine kinase JAK2 in polycythemia vera, essential thrombocythemia, and myeloid metaplasia with myelofibrosis. *Cancer Cell*. 2005; 7 (4): 387-97.
10. Kralovics R, Passamonti F, Buser AS, et al. A gain-of-function mutation of JAK2 in myeloproliferative disorders. *N Engl J Med*. 2005; 352 (17): 1779-90.
11. Kralovics R, Guan Y, Prchal JT. Acquired uniparental disomy of chromosome 9p is a frequent stem cell defect in polycythemia vera. *Exp Hematol*. 2002; 30 (3): 229-36.
12. Pieri L, Spolverini A, Scappini B, et al. Concomitant occurrence of BCR-ABL and JAK2V617F mutation. *Blood*. 2011; 118 (12): 3445-46.
13. Ungureanu D, Wu J, Pekkala T, et al. The pseudokinase domain of JAK2 is a dual-specificity protein kinase that negatively regulates cytokine signaling. *Nat Struct Mol Biol*. 2011; advance online publication.
14. Dawson MA, Bannister AJ, Gottgens B, et al. JAK2 phosphorylates histone H3Y41 and excludes HP1alpha from chromatin. *Nature*. 2009; 461 (7265): 819-22.
15. Scott LM, Tong W, Levine RL, et al. JAK2 exon 12 mutations in polycythemia vera and idiopathic erythrocytosis. *N Engl J Med*. 2007; 356 (5): 459-68.
16. Passamonti F, Elena C, Schnittger S, et al. Molecular and clinical features of

- the myeloproliferative neoplasm associated with JAK2 exon 12 mutations. *Blood*. 2011; 117 (10): 2813-6.
17. Oh ST, Simonds EF, Jones C, et al. Novel mutations in the inhibitory adaptor protein LNK drive JAK-STAT signaling in patients with myeloproliferative neoplasms. *Blood*. 2010; 116 (6): 988-92.
 18. Oh ST, Zahn JM, Jones CD, et al. Identification of novel LNK mutations in patients with chronic myeloproliferative neoplasms and related disorders. *ASH Annual Meeting Abstracts*. 2010; 116 (21): 315.
 19. Lasho TL, Pardanani A, Tefferi A. LNK mutations in JAK2 mutation-negative erythrocytosis. *N Engl J Med*. 2010; 363 (12): 1189-90.
 20. Lasho T, Tefferi A, Vannucchi AM, et al. LNK mutation studies in chronic- and blast-phase myeloproliferative neoplasms and JAK2 mutation-negative erythrocytosis. *ASH Annual Meeting Abstracts*. 2010; 116 (21): 4105-.
 21. Grand FH, Hidalgo-Curtis CE, Ernst T, et al. Frequent CBL mutations associated with 11q acquired uniparental disomy in myeloproliferative neoplasms. *Blood*. 2009; 113 (24): 6182-92.
 22. Vainchenker W, Delhommeau F, Constantinescu SN, Bernard OA. New mutations and pathogenesis of myeloproliferative neoplasms. *Blood*. 2011; 118 (7): 1723-35.
 23. Vannucchi AM, Biamonte F. Epigenetics and mutations in chronic myeloproliferative neoplasms. *Haematologica*. 2011; 96 (10): 1398-402.
 24. Delhommeau F, Dupont S, Della Valle V, et al. Mutation in TET2 in myeloid cancers. *N Engl J Med*. 2009; 360 (22): 2289-301.
 25. Schaub FX, Looser R, Li S, et al. Clonal analysis of TET2 and JAK2 mutations suggests that TET2 can be a late event in the progression of myeloproliferative neoplasms. *Blood*. 2010; 115: 2003-7.
 26. Ernst T, Chase AJ, Score J, et al. Inactivating mutations of the histone methyltransferase gene EZH2 in myeloid disorders. *Nat Genet*. 2010; 42 (8): 722-6.
 27. Guglielmelli P, Biamonte F, Score J, et al. EZH2 mutational status predicts poor survival in myelofibrosis. *Blood*. 2011; 118 (19): 5227-34.
 28. Carbuccia N, Murati A, Trouplin V, et al. Mutations of ASXL1 gene in myeloproliferative neoplasms. *Leukemia*. 2009; 23 (11): 2183-6.
 29. Ricci C, Spinelli O, Salmoiraghi S, Finazzi G, Carobbio A, Rambaldi A. ASXL1 mutations in primary and secondary myelofibrosis. *British Journal of Haematology*. 2011.
 30. Stein BL, Williams DM, O'Keefe C, et al. Disruption of the ASXL1 gene is frequent in primary, post-essential thrombocytosis and post-polycythemia vera myelofibrosis, but not essential thrombocytosis or polycythemia vera: analysis of molecular genetics and clinical phenotypes. *Haematologica*. 2011; 96 (10): 1462-9.
 31. Vannucchi AM, Guglielmelli P. Molecular pathophysiology of Philadelphia-negative myeloproliferative disorders: beyond JAK2 and MPL mutations. *Haematologica*. 2008; 93 (7): 972-6.

Fattori genetici e microambiente nella patogenesi della LLC

Federico Caligaris Cappio, Elisa Ten Hacken

Dipartimento di OncoEmatologia e Divisione di Ricerca di Oncologia Molecolare,
IRCCS Fondazione San Raffaele e Università Vita Salute San Raffaele, Milano

B-cell chronic lymphocytic leukemia (CLL) is a phenotypically homogeneous, yet clinically heterogeneous malignancy that remains at present incurable. The natural history of CLL is determined by the interplay of causal and influential genes and by the relationships that leukemic cells entertain with specific microenvironmental niches (Fig. 1). As at diagnosis the central clinical problem is to define which patients should be treated, when and how, the most compelling challenge is to rapidly move the accumulating innovative biological findings into clinical application. The strategy which is actively pursued is to understand how gene abnormalities concur with microenvironmental stimuli to favour CLL development and progression. This will operatively translate into a fully integrated genotype-phenotype and tumor-microenvironment definition, where all aspects concurring to the pathogenesis of CLL will be specifically addressed and carefully weighted. The goal is will to develop improved risk stratification and more rationale therapeutic strategies for individual patients.

Background and state of the art

The molecular basis of CLL is accounted for by a number of well defined chromosome aberrations, that include 13q, 11q, 17p, trisomy 12 and by mirNome abnormalities, an example being the deletion of miRNA 15 and miRNA 16 located in critical regions on chromosome 13q.

Cytogenetic abnormalities have a prognostic significance and microRNA such as mir221, mir222, mir182b may prove useful biomarkers. More recently next generation sequencing have revealed recurrent though uncommon mutations in the tumour cell clone (the most relevant being, Notch1, MyD88, SF3B1) that influence the disease progression.

Biologically-defined prognostic factors with a widely confirmed clinical value exist including the mutational status of IGHV genes, the expression of CD38 and of ZAP70. All these "e"markers" have directed our attention toward the stimulation of B-cell-receptor (BCR) and microenvironment interactions.

A redirection and reinforcement of microenvironment interactions allow malignant CLL cells to avoid apoptosis and acquire better growing conditions and may be taken to explain why despite major therapeutic advances CLL is still incurable. Though our understanding of the role of T-cells and especially of stromal cells in CLL is just in its infancy, several findings suggest that T cells influence malignant B cell proliferation, while stromal and accessory cells favour the extended survival and accumulation of leukemic cells.

The relevant events occur in tissues

All relevant events of CLL occur in tissue microenvironments where pro-lymphocytes and paraimmunoblasts cluster to form scattered pseudofollicular proliferation centers (PC). The prevailing opinion is that PC are the source of most cellular generation in CLL whose progeny accumulate and may flow into the peripheral blood (PB). Data are gathering to indicate that in terms of proliferation and apoptosis the clone is more dynamic than anticipated.

The key molecule on the surface of CLL B cells is a functional BCR that allows antigen (Ag) interaction. A BCR-mediated stimulation plays a relevant role in the natural history of CLL as documented by the bias in the immunoglobulin heavy variable (IGHV) gene repertoire, the different prognosis of patients with different IGHV gene mutational status and especially the identification of patient subsets

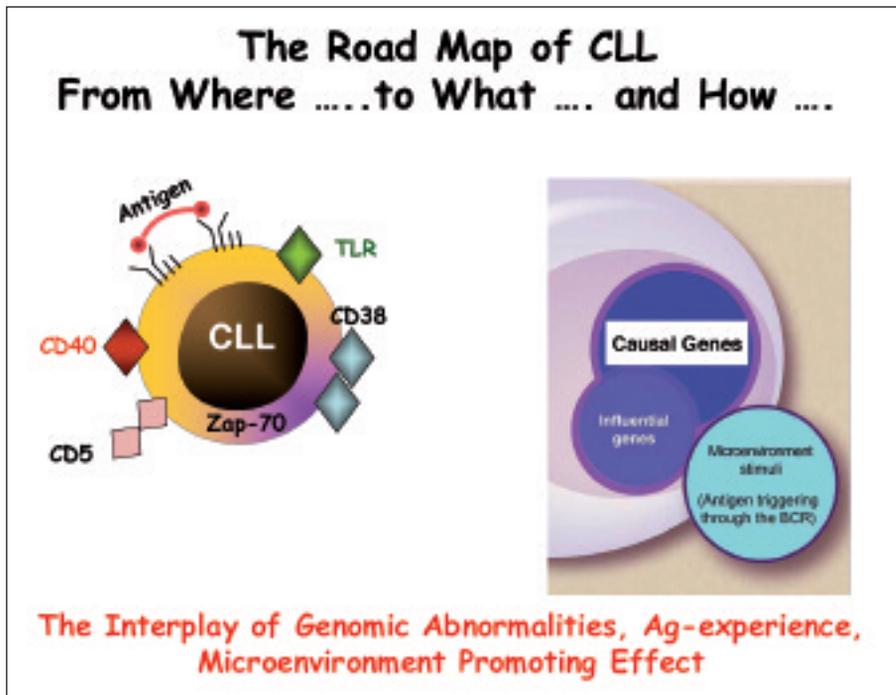


Fig. 1

sharing BCRs with restricted, quasi-identical IG sequences (stereotyped BcRs). Ag receptors normally involved in eliminating cellular debris, scavenging apoptotic cells and providing a first line of defense against pathogenic bacteria may trigger and/or facilitate the onset and evolution of at least some CLL clones.

The issue of Monoclonal B-cell Lymphocytosis (MBL)

Tiny monoclonal B-cell populations phenotypically CLL-like can be detected in the PB of about 3.5% of healthy individuals and are named Monoclonal B Lymphocytosis (MBL). MBL have the same gender and age predisposition as CLL reaching a frequency greater than 10% in people >60 years and have a significantly higher incidence in the relatives of CLL patients. MBL may be further divided into two different general subgroups, *high* and *low-count MBL*. *High count MBL* are usually diagnosed in a clinical setting, are associated with lymphocytosis and have a concentration of clonal B cells greater than ~1500/iL in continuum with CLL. *Low count MBL* are only detected during screening studies of healthy individuals using highly sensitive technical procedures and have generally less than ~50 clonal B cells per il. CLL is always preceded by an MBL state and the potential risk of progression of *high count MBL* into clinically overt CLL has been shown to be about 1.1% per year. It thus appears logical to hypothesize that they might be a precursor state for CLL, reminiscent of the relationship between Monoclonal Gammopathy of Undetermined Significance (MGUS) and Multiple Myeloma. *Low-count MBL* are the most common form in the general population with a prevalence at least 100 fold greater than that of CLL and show very little if any propensity to progress. That notwithstanding they have the cytogenetic hallmarks of CLL including 13q deletions, trisomy 12 and 17p deletion even when the number of circulating monoclonal CLL-like cells is disappointingly small.

Questions to be solved

Despite an unprecedented flurry of novel observations gathered in the last decade we are somehow back “to the start”. Several cytogenetic abnormalities and single-point mutations are well known but still the initiating genetic events remain to be defined. The striking role of exogenous stimuli from the microenvironment is now well established, but this notion has widened the number of avenues of research, limiting the possibility to narrow down a single, easy-to-hit culprit. For all these reasons, a multidisciplinary approach that combines and concentrates different and unrelated expertise is needed to develop unexplored experimental avenues aimed at answering the yet unanswered questions (Fig. 2):

1. Which genetic events favour the onset of CLL?
2. How physiological signals through the BCR or other co-receptors (e.g. CD40, TLR) concur to the expansion and survival of the leukemic clone, how this is initiated, where it occurs, how it translates into cell proliferation (or anergy)?
3. It is likely that both functional and genetic events will all funnel into identical

or parallel molecular pathways that can be affected at different levels and intensity thereby correlating with the heterogeneity of the disease. It is especially important to characterize these pathways as drugs that interfere with intracellular signal transduction are entering the clinical setting in CLL.

4. As the microenvironment is an attractive target for innovative treatment strategies the heterogeneity of bone marrow (BM) and secondary lymphoid organs needs to be better dissected by establishing proper *in vitro* and *in vivo* models to reproduce the different tissue microenvironments.
5. It is important to establish through which mechanisms resting CLL cells flow into the PB and why they are compartmentalized in LN tissues in small lymphocytic lymphoma (SLL) and which rules govern leukemic cells traffic and home to specific tissue microenvironmental niches, such as the BM and, within it, the PC where cell subsets endowed with genetic aberrations exposed to microenvironmental stimuli increase proliferation and avoid apoptosis.
6. At present we do not have a clue as to whether individual subjects with MBL will/will not progress into fully fledged CLL nor when this will occur. It is mandatory to understand which molecular features underlie a potential pre-leukemic condition to focus medical attention onto the rare risky subjects and refrain from useless and prolonged monitoring in the majority of otherwise healthy individuals.
7. The range of prognostic/predictive factors needs to be expanded because

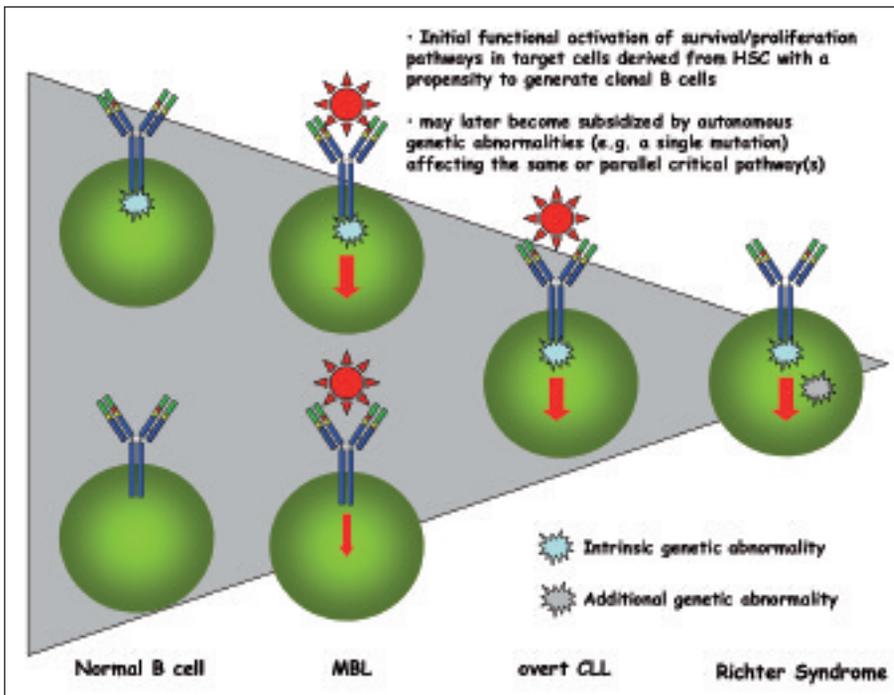


Fig. 2

numerous are the outliers with the existing ones. This may also lead to define a personalized molecular prognostic/predictive card.

References

- Bagnara D, Kaufman MS, Calissano C, et al. *Blood*. 2011; 117: 5463-72.
- Bertilaccio MT, Scielzo C, Simonetti G, et al. *Blood*. 2010; 115 (8): 1605-9.
- Bertilaccio MT, Simonetti G, Dagklis A, et al. *Blood*. 2011; 118: 660-9.
- Bichi R, Shinton SA, Martin ES, et al. *Proc Natl Acad Sci USA*. 2002; 99: 6955-60.
- Burger Jan A, Ghia P, Rosenwald A, Caligaris-Cappio F. *Blood*. 2009; 114 (16): 3367-75.
- Caligaris-Cappio F, Hamblin TJ. *J Clin Oncol*. 1999; 17: 399-408.
- Caligaris-Cappio F, Ghia P. *Journal of Clinical Oncology*. 2008; 26 (27): 4497-503.
- Calin GA, Dumitru CD, Shimizu M, et al. *Proc Natl Acad Sci USA*. 2002; 99: 15524-9.
- Calin GA, Ferracin M, Cimmino A, et al. *N Engl J Med*. 2005; 353: 1793-801.
- Catera R, Silverman GJ, Hatzi K, et al. *Mol Med*. 2008; 14: 665-74.
- Chu CC, Catera R, Zhang L, et al. *Blood*. 2010; 115: 3907-15.
- Crowther-Swanepoel D, Corre T, Lloyd A, et al. *Blood*. 2010; 116 (26): 5957-60.
- Dagklis A, Fazi C, Sala C, et al. *Blood*. 2009; 114: 26-32.
- Damle RN, Wasil T, Fais F, et al. *Blood*. 1999; 94: 1840-7.
- Dohner H, Stilgenbauer S, Benner A, et al. *N Engl J Med*. 2000; 343: 1910-6.
- Fabbri G, Rasi S, Rossi D, et al. *J Exp Med*. 2011; 208: 1389-1401.
- Fazi C, Scarfo L, Pecciarini L, et al. *Blood*. 2011; 118 (25): 6618-25.
- Goldin LR, Lanasa MC, Slager SL, et al. *Br J Haematol*. 151: 152-8.
- Hamblin TJ, Davis Z, Gardiner A, et al. *Blood*. 1999; 94: 1848-54.
- Herishanu Y, Perez-Galan P, Liu D, et al. *Blood*. 2011; 117:563-574.
- Kikushige Y, Ishikawa F, Miyamoto T, et al. *Cancer Cell*. 2011; 20: 246-59.
- Klein U, Tu Y, Stolovitzky GA, et al. *J Exp Med*. 2001; 194: 1625-38.
- Klein U, Lia M, Crespo M, et al. *Cancer Cell*. 2010; 17: 28-40.
- Landgren O, Albitar M, Ma W, et al. *N Engl J Med*. 2009; 360: 659-67.
- Lanemo Myhrinder A, Hellqvist E, Sidorova E, et al. *Blood*. 2008; 111: 3838-48.
- Lia M, Carette A, Tang H, et al. *Blood*. 2011, in press.
- Marti GE, Rawstron AC, Ghia P, et al. *Br J Haematol*. 2005; 130: 325-32.
- Messmer BT, Messmer D, Allen SL, et al. *J Clin Invest*. 2005; 115: 755-64.
- Muzio M, Scielzo C, Bertilaccio MT, et al. *Br J Haematol*. 2009; 144: 507-16.
- Nieto WG, Almeida J, Romero A, et al. *Blood*. 2009; 114: 33-7.
- Puente XS, Pinyol M, Quesada V, et al. 2011; 475: 101-5.
- Quesada V, Conde L, Villamor N, et al. *Nat Genet*. 2012; 44: 47-52.
- Rawstron AC, Green MJ, Kuzmicki A, et al. *Blood*. 2002; 100: 635-639.
- Rawstron AC, Bennett FL, O'Connor SJ, et al. *N Engl J Med*. 2008; 359: 575-83.
- Rossi D, Rasi S, Fabbri G, et al. *Blood*. 2011, in press.
- Scielzo C, Bertilaccio MT, Simonetti G, et al. *Blood*. 2010; 116 (18): 3537-46.
- Stevenson FK, Caligaris-Cappio F. *Blood*. 2004; 103: 4389-95.
- Wang L, Lawrence MS, Wan Y, et al. *N Engl J Med*. 2011; 365: 2497-506.

Biological and clinical predictors of chronic lymphocytic leukemia transformation to Richter syndrome

Valeria Spina, Silvia Rasi, Alessio Brusca, Marco Fangazio, Clara Deambrogi, Sara Monti, Stefania Cresta, Rosella Famà, Mariangela Greco, Carmela Ciardullo, Daniela Piranda, Lorenzo De Paoli, Davide Rossi, Gianluca Gaidano

Dipartimento di Medicina Clinica e Sperimentale e IRCAD, Divisione di Ematologia, Università del Piemonte Orientale "Amedeo Avogadro", Novara

Definition and frequency of Richter syndrome

Over time, a fraction of chronic lymphocytic leukemia (CLL) develop Richter syndrome (RS), representing the clinico-pathologic transformation of CLL to an aggressive lymphoma, most commonly diffuse large B-cell lymphoma (DLBCL). Historically, the definition of Richter syndrome also applies to CLL cases that have developed classical Hodgkin lymphoma, although this event is very rare. Richter syndrome is named after Maurice N. Richter, who described the syndrome for the first time in the 1920s (1). Transformation of CLL to RS should be distinguished from other types of CLL progression recognized by IWCLL-NCI guidelines (2). Histologic documentation is mandatory for diagnosing RS, that otherwise can only be clinically suspected, but not proven (3, 4). Transformation of CLL to DLBCL should also be distinguished from prolymphocytic transformation, CLL "acceleration", and development of several other B-cell malignancies that may arise at an increased rate in CLL patients (4, 5).

In this review, we shall restrict the term RS solely to CLL transformation to DLBCL, since this clinical scenario represents over 95% of all RS cases. Despite this terminological restriction, the clinical definition of RS maintains a certain degree of heterogeneity, and may be molecularly distinguished into at least two biologically different entities:

1. transformation of CLL cells to a clonally related DLBCL;
2. development of a DLBCL that is unrelated to the CLL clone (3, 6-11).

Transformation of CLL cells to a clonally related DLBCL is by large the most frequent of the two conditions, accounting for approximately 80-90% of all RS cases. In clonally related RS, the pathogenetic link between CLL cells and the emerging DLBCL clone is apparently obvious, and is substantiated by gain of

novel molecular lesions at a certain timepoint of the natural history of the CLL clone, conceivably corresponding to the time of clinico-pathologic transformation. On the other hand, the mechanisms underlying the development of a DLBCL that is clonally unrelated to CLL cells are not fully understood, and may be related, at least in part, to the condition of immune deregulation that frequently characterises patients affected by CLL.

With the advent of novel immunotherapies for CLL, and especially after the introduction of the anti-CD52 antibody alemtuzumab that causes a profound T-cell depletion, occasional CLL patients have been reported to develop clinically aggressive lymphomas characterized by Epstein-Barr virus (EBV) infection (12, 13). These cases of alemtuzumab-associated aggressive lymphomas are clinically and biologically distinct from RS, and should be considered as a novel type of immunodeficiency-related lymphoma developing after T-cell depleting therapies in patients already immunocompromised because of the underlying disease and/or because of previous chemotherapy (12, 13).

The propensity displayed by CLL toward transformation to DLBCL is also shared by other indolent B cell lymphoproliferative disorders, namely follicular lymphoma and marginal zone lymphoma. The risk of transformation varies among the different histologies, being highest in the case of follicular lymphoma. Currently, it is unknown whether one or more common molecular pathways recapitulates transformation to DLBCL in all these different clinical settings, or whether transformation from indolent to aggressive B-cell malignancies is due to different genetic and epigenetic lesions depending on the type of initial disease. Differences in the pattern of genetic lesions accompanying transformation of CLL and of follicular lymphoma to DLBCL suggest that the molecular pathways underlying transformation are not fully overlapping in these clinical contexts (14).

The incidence of RS varies among available series. In particular, since RS diagnosis requires pathological assessment, heterogeneity in biopsy policies among institutions represents a substantial factor influencing the incidence of RS observed in different CLL series (4). Notably, whereas re-biopsy is standard practice at each progression of follicular lymphoma (15), a similar attitude is not commonly followed in the case of CLL, even in patients with substantial enlargement of lymph nodes. The hematologist's perception of RS as a rare complication of CLL may be due, at least in part, to under recognition of RS in the current clinical practice.

Diagnosis and prognosis of Richter syndrome

Classically, RS is pathologically represented by DLBCL (3, 4, 9-11, 16-20). The disease involves most frequently the lymph nodes, but extra-nodal localizations are also not uncommon, and may affect the gastrointestinal tract, liver, tonsil, skin and bone marrow among other sites (3, 4, 9-11, 16-20). The PET characteristics of the lesion, in particular the standardized uptake value (SUV), may guide the choice of the site to be biopsied, since sites affected by RS are expected to have SUVs overlapping those of *de novo* DLBCL (21). RS cases presenting as DLBCL

expressing CD5 should be distinguished from *de novo* CD5 positive DLBCL, that represent a molecularly distinct entity (22). Morphologically, RS may display both immunoblastic and centroblastic features that denote histologic variants of DLBCL occurring *de novo*. Based on the Hans algorithm utilized for immunohistochemical classification of *de novo* DLBCL (23), most (approximately 90%) cases of RS express IRF4 and therefore belong to the post-germinal center phenotype, whereas only 10% of RS display the germinal center profile characterized by CD10 and/or BCL6 positivity in the absence of IRF4 (11).

Prognosis of RS is generally considered highly unfavorable. However, a study of 148 patients with biopsy or fine needle aspiration-proven RS has documented that survival of RS is not uniform among patients, ranging from few weeks to 15 years, and may be predicted on clinical grounds by the RS score (24). The RS score is not intended for predicting RS development in CLL patients, but rather predicts RS prognosis once that transformation has occurred (24). The RS score is a model predicting an individual patient's risk of death, and is based on five adverse factors predicting short survival, namely:

1. Zubrod performance status >1;
2. elevated LDH level;
3. platelet count $\leq 100,000$;
4. tumor size ≥ 5 cm;
5. more than two prior lines of therapy (24).

Patients are assigned to one of four risk groups based on number of presenting risk factors: 0 or 1, low risk; 2, low-intermediate risk; 3, high-intermediate risk; 4 or 5, high risk. In the original report of the RS score, survival was 1.12 years in patients with RS score 0-1; 0.90 years in RS score 2; 0.33 years in RS score 3; and 0.14 years in RS score 4-5 (24). Remarkably, risk factors that are relevant to the International Prognostic Index, namely number of extranodal sites of disease, age and stage were not relevant to the RS score, confirming the notion that RS and *de novo* DLBCL are very different diseases. In contrast, survival prognostication in RS takes advantage of parameters reflecting marrow failure, such as thrombocytopenia, and parameters reflecting immune system exhaustion and selective pressure to chemorefractory clones, such as number of prior lines of treatment.

Very recent data from our group indicate that, in addition to the clinical risk factors included in the RS score, survival post-transformation of RS may also be predicted by the tumor genotype and by the clonal relationship of the RS phase with the pre-existing CLL clone (25).

Clinical and biological tools for predicting Richter syndrome development

Early recognition of RS may be clinically useful in order to avoid the exposure of patients to multiple lines of therapy that, being targeted to CLL progression, are of little efficacy on the transformed clone. This notion prompts the need for a close monitoring of CLL patients harboring clinical and/or biological risk factors of RS development. Recent advancements in the field have disclosed a number of risk factors, both clinical and biological, that might be helpful for RS prediction

(3). It is remarkable that several of the predictors identified to date are specific for CLL transformation to RS, and are not related to other markers predicting CLL clinical progression according to IWCLL-NCI guidelines (2, 3). The existence of predictors for CLL transformation to RS documents that RS transformation and CLL progression without histological transformation are distinct events both clinically and biologically, that require a different, though complementary, diagnostic approach for a comprehensive assessment of the risk category of the individual patient.

The biological risk factors that have been tested for their ability to predict RS include immunoglobulin heavy chain variable gene (IGHV) homology $\geq 98\%$, IGHV gene usage, stereotyped B cell receptor (BCR), FISH karyotype, *TP53* status, *BCL2* genotype, 14q32 translocations, telomere length, and expression of CD38, ZAP70, and CD49 (Table 1, Fig. 1) (3, 26-29). Out of these, the following have been scored as independent risk factors for RS development:

1. absence of del13q14;
2. usage of specific IGHV genes, namely IGHV4-39;
3. usage of a stereotyped BCR;
4. expression of CD38 (Fig. 2);
5. genetic background of the host; and vi) telomere length (3, 26-29).

Among clinical predictors, lymph node size ≥ 3 cm, a parameter of solid disease, was the sole clinical risk factor of RS selected by two models of multivariate analysis (3). Conversely, parameters of leukemic disease, namely lymphocyte count, advanced Rai stage, splenomegaly, and percentage and pattern of bone marrow involvement, were not scored as independent clinical risk factors of RS (3). As expected, parameters of leukemic disease did predict for short time to progression without transformation, further reinforcing the notion that RS transformation and CLL progression according to NCI guidelines are distinct events in the natural history of CLL (3) (Tab. 1).

Tab. 1 - Predictors of CLL progression and/or of RS transformation at the time of CLL diagnosis^a.

	Progression	Transformation
Biological features at diagnosis		
Telomere length	Yes	Yes
CD38 >30%	Yes	Yes
del17p13	Yes	No
del11q22-q23	Yes	No
IGHV homology >98%	Yes	No
Absence of del13q14	No	Yes
IGHV4-39 usage	No	Yes
Stereotyped VH CDR3	No	Yes
CD38 genotype	No	Yes
LRP4 genotype	No	Yes
Clinical features at diagnosis		
Largest node >3 cm	Yes	Yes
Rai stage III-IV	Yes	No
Splenomegaly	Yes	No

^aee refs. (3, 28, 29, 50, 51)

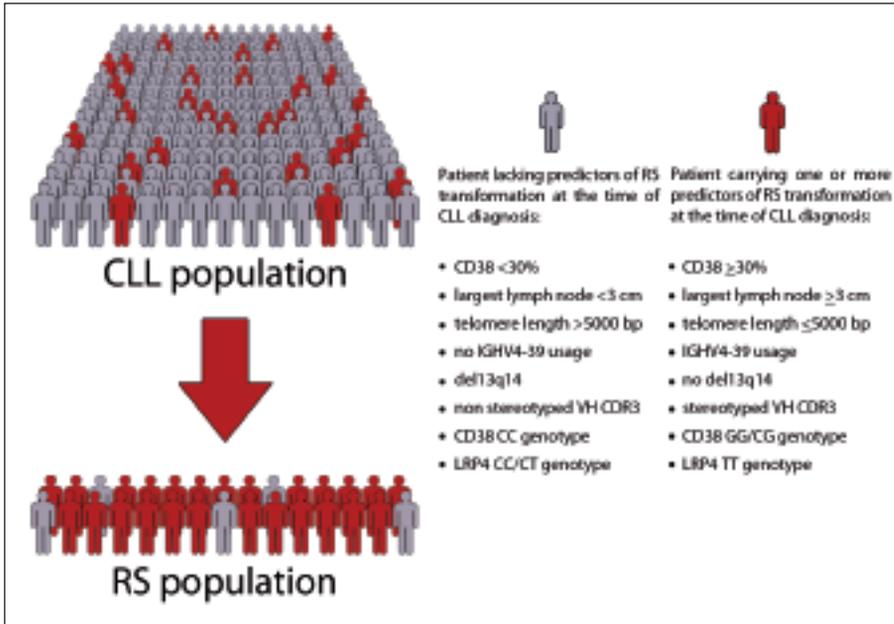


Fig. 1 - Clinical and molecular risk factors that, at the time of CLL diagnosis, predict RS transformation. At the time of CLL diagnosis, several risk factors predict an increased risk of RS transformation, including: expression of CD38, lymph node size, telomere length, absence of del13q14, immunogenetic features of the CLL clone, GG/CG genotypes of the *CD38* rs6449182 SNP, and TT genotype of the *LRP4* rs2306029 SNP.

As stated above, FISH karyotype at CLL diagnosis predicts RS. In particular, absence of del13q14 confers an independent risk of RS development (3). Conceivably, pathogenetic differences between CLL with and without del13q14 underlie the different predisposition of these two CLL subsets toward RS transformation. In fact, del13q14 is thought to play an important role in disease pathogenesis, and its presence activates molecular pathways that are distinct from those of CLL without del13q14 (30-32). The concept that propensity to DLBCL transformation may be affected by the genetic profile of the pre-existing indolent phase is not unique to CLL among B cell malignancies. In fact, transformation to DLBCL of at least another indolent B-cell disorder, i.e. MALT lymphoma, has been documented to depend upon differences in disease pathogenesis of the pre-existing indolent B cell clone (33). Other FISH abnormalities of CLL, including del17p13, del11q23, +12 or 14q32 translocations, do not appear to associate with an increased risk of RS (3, 34).

A sizeable fraction of CLL (30%) show a high degree of structural similarity in the BCR because of homology of the expressed heavy chain complementarity determining region 3 (VH CDR3) (35-38). This phenomenon is known as stereotyped BCR and strongly suggests the role of antigen stimulation in disease development (39). Recently, we have shown that carrying a stereotyped BCR at CLL diagnosis may help in the identification of patients at risk of RS transformation

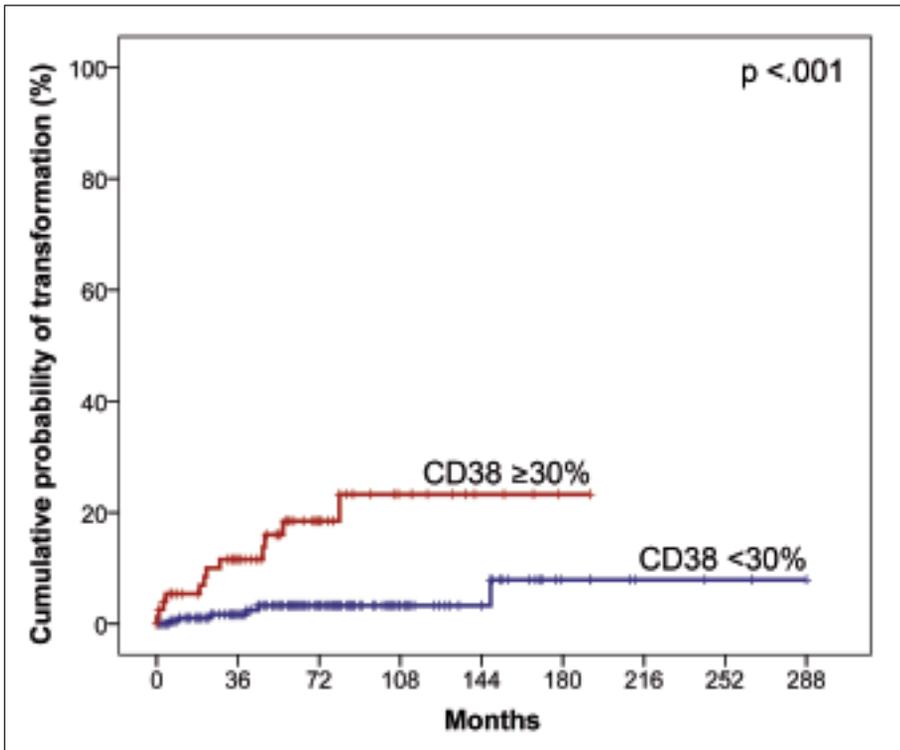


Fig. 2 - Prediction of RS transformation according to CD38 expression at CLL diagnosis. Kaplan-Meier curves showing the value of CD38 in predicting RS transformation in the CLL series (n=360) from the Division of Hematology of the Amedeo Avogadro University of Eastern Piedmont (years 2000-2010).

(28). Comparing the immunogenetic features of 69 cases of RS with those of 714 cases of CLL, the prevalence of stereotyped VH CDR3 was significantly higher in RS than in non-transformed CLL (49.3% vs 21.3%) (28). Of note, applying an actuarial approach to this cohort, stereotyped VH CDR3 at the time of CLL diagnosis was revealed to be an independent predictor of RS transformation (28). The risk of RS transformation carried by stereotyped BCR was independent of the mutation status of *IGHV* genes (28). Among CLL carrying a stereotyped BCR, patients utilizing a specific *IGHV* gene, i.e. *IGHV4-39*, in a stereotyped fashion (so called BCR “subset 8”) had the highest risk of RS transformation (17-fold higher than cases without a stereotyped BCR; 5-year risk 68% vs 4%) (28). In multivariate analysis, usage of *IGHV4-39* along with stereotyped VH CDR3 was an independent risk factor of RS development (28). This observation, beside being of clinical interest, points to the role of antigen stimulation in RS development. The role of antigen stimulation in favoring RS development is further sustained by the observation that the BCR expressed by CLL belonging to subset 8 is characterized by the highest affinity for non-muscle myosin II, an autoantigen that might be involved in CLL promotion (40).

Both CD38 expression and lymph node size have been shown to be independent predictors of RS (3). From a pathogenetic standpoint, this observation suggests that molecular circuits in CLL lymph nodes may be important for determining the risk of RS in individual patients, and may directly contribute to CLL transformation to DLBCL. In fact, a strict association between CD38 expression and lymph node enlargement in CLL is suggested by several lines of evidence (41-44). First, CD38 expression is higher in CLL cells of lymph nodes than in CLL cells of the peripheral blood or bone marrow (41-43). Second, CD38 expression is required for CLL trafficking toward lymphoid tissues (44). Third, CD38 expression associates with predominantly nodal disease (41-43). Additionally, higher levels of CD38 expression in CLL lymph nodes associate with lymphoma-like tumor distribution and diffuse bone marrow infiltration pattern (43). These same features have been reported to predict RS (3). On these basis, the lymph node may thus provide an optimal microenvironment for proliferation and blastic transformation of CD38 expressing CLL cells. In fact, interactions with other cells and cytokines that take place in the lymph node may provide adequate stimulation for CD38 up-regulation and signalling in CLL cells (45-47). Intriguingly, *in vitro* activation of CD38 signalling induces transformation of CLL cells to plasmablasts that are reminiscent of the blasts appearing during RS evolution (45).

Expression of *CD38* is regulated at multiple levels (45). The 5' end of intron 1 of the *CD38* gene is involved in the induction of *CD38* expression by transcription factors in myeloid cells (48). Within the same region maps a well-characterized single nucleotide polymorphism (SNP), leading to a C>G variation at position 184 (*CD38*rs6449182) (49). The presence of the minor G allele, either in a heterozygous or in a homozygous condition, strongly associates with RS (50). Compared to CC homozygotes, GG homozygote CLL patients had a 30.6% increase in the relative risk of developing RS, whereas GC heterozygotes showed an intermediate probability of 12.4% (50). At 5 years, GG homozygotes showed an increase in probability of developing RS that approximately doubled that associated with GC heterozygotes (50). These observations document that the host genetic background may be of relevance in predicting RS development among individuals affected by CLL (50).

The impact of *CD38* SNPs on the risk of RS has prompted further studies aimed at defining the role of the host genetic background in predisposing CLL patients to RS development. Analysis of a large number of SNPs has revealed that also the genetic background of the *LRP4* gene (rs2306029) is an important predictor of RS transformation (51). In particular, patients that are homozygotes for the minor allele of *LRP4* rs2306029 display a significantly higher 5-year cumulative probability of transformation to aggressive lymphoma compared to patients carrying the CT/CC genotypes that contain the wild type allele (51). Notably, *LRP4* encodes for a protein involved in the Wnt/beta-catenin signaling pathway, that is known to be activated in CLL cells (52). The fact that alterations of the LRP4 protein are pathogenetic in clinical context other than CLL points to a possible role of the interaction between LRP4 and Wnt/beta-catenin also in RS pathogenesis (53).

When assessed at the time of diagnosis, telomere length of CLL cells has also been shown to be an independent predictor of RS transformation (29). As already demonstrated for other neoplasms, telomere length recapitulates the cell proliferative history, and short telomere length denotes an aggressive clinical phenotype and chemo-refractoriness (54).

Molecular pathogenesis of Richter syndrome

The molecular mechanisms leading to transformation of CLL to RS are largely unknown. In classic RS represented by DLBCL, immunogenotypic studies have revealed that the majority (approximately 80%), though not the totality, of RS cases are clonally related to the pre-existing CLL clone (3, 4, 6-11). Cases of RS that are clonally related to the pre-existing CLL phase conceivably stem from the progressive accumulation of molecular lesions, either genetic and/or epigenetic, that favor the emergence and predominance of a large cell population causing clinical aggressiveness. In all available RS series, a fraction (10-20%) of RS cases is not related to the pre-existing CLL clone (3, 4, 6-11, 25). The precise pathogenesis of clonally unrelated RS is unclear and requires extensive investigations. Accumulation of molecular lesions in the CLL clone cannot be considered as a relevant factor in the development of clonally unrelated RS. Rather, these RS cases may be favoured by alterations of the host genetic background and immunologic function, or by derangements in the lymph node microenvironment induced by CLL cells. Emergence of DLBCL that are clonally unrelated to CLL may be part of a more generalized phenomenon that, for reasons that are currently obscure, render CLL patients prone to develop a second lymphoid malignancy or, more in general, a second cancer (5).

Scant knowledge is currently available concerning the genetic lesions that associate with transformation from CLL to clonally related DLBCL. Acquisition of *TP53* mutations and/or 17p13 deletion is a very frequent molecular alteration in clonally related RS (3, 25, 55, 56). Interestingly, acquisition of *TP53* disruption by mutation and/or deletion is also common in other types of transformation from indolent to aggressive B cell malignancies (57, 58), suggesting that this molecular pathway is a general feature of clinical progression of mature B cell neoplasia. Curiously, despite their involvement in RS pathogenesis, *TP53* mutations at CLL diagnosis do not increase the subsequent risk of RS development (3).

In *de novo* DLBCL arising in different clinical contexts, aberrant somatic hypermutation (ASHM) of the *c-MYC*, *RhoH/TTF*, *PAX5* and *PIM1* proto-oncogenes represents the most frequent genetic lesion currently known (59-61). ASHM in DLBCL is thought to be caused by malfunctioning of activation-induced cytidine deaminase (AID), an enzyme that is essential for physiological somatic hypermutation of normal B cells (62, 63). Although AID levels increase at the time of CLL transformation to RS (9, 64), investigations of ASHM in RS has yielded controversial results (14, 64). The apparently low incidence of ASHM in RS distinguishes CLL transformation to DLBCL from follicular lymphoma transformation to DLBCL, that, at variance with RS, is frequently accompanied by a high

load of ASHM occurring at the time of histologic transformation (14). Cytogenetic studies have not revealed recurrent abnormalities selectively clustering with RS (65-69). The report of trisomy 12 in RS is not disease-specific, since trisomy 12 occurs in several other B cell malignancies, including CLL and several types of lymphoma (70). Consistent with the role of absence of del13q14 as a risk factor for RS, loss of del13q14 has been reported in RS, and may be due to the emergence of a subclone lacking del13q14 (3). At variance with other types of B cell lymphoma, 14q32 translocations involving immunoglobulin genes do not contribute to the molecular mechanisms that are operative at the time of transformation from CLL to DLBCL (34). Consistently, translocations of *BCL1*, *BCL2*, and *BCL6* are not found in RS (3). Lack of involvement of 14q32 translocations in RS development suggests that RS pathogenesis differs from that of other aggressive B-cell non-Hodgkin lymphomas, including *de novo* DLBCL, in which proto-oncogene deregulation by juxtaposition to immunoglobulin heavy or light chain loci is a relatively common mechanism (34, 71). Molecular cytogenetic studies have shown that RS associates with a higher degree of genomic complexity compared to CLL, although none of the genetic lesions reported to date appear to be specific for RS (72, 73).

The role of EBV infection has been suggested by some studies as a potentially relevant factor for RS pathogenesis. The observation that the overwhelming majority of RS do not carry EBV infection in the malignant cells, however, does not favor this hypothesis (3). The presence of EBV sequences has been documented in some, though not all, cases of RS originating in patients previously treated with fludarabine for their pre-existent CLL (74, 75). EBV infection in these cases has been thought to be related to the immune deregulation caused by purine analogues.

As detailed above, available studies of RS pathogenesis indicate that RS development may not be easily explained by genetic lesions identified in other B cell disorders, namely CLL and *de novo* DLBCL. This lack of knowledge points to the need of novel studies aimed at clarifying the molecular basis of RS. Genome wide studies that are currently in progress will hopefully reveal novel molecular alterations and gene expression abnormalities involved in RS development.

Bibliografia

1. Richter MN. Generalized reticular cell sarcoma of lymph nodes associated with lymphatic leukaemia. *Am J Pathol.* 1928; 4 (4): 285-92.
2. Hallek M, Cheson BD, Catovsky D, et al. International Workshop on Chronic Lymphocytic Leukemia. Guidelines for the diagnosis and treatment of chronic lymphocytic leukemia: a report from the International Workshop on Chronic Lymphocytic Leukemia updating the National Cancer Institute-Working Group 1996 guidelines. *Blood.* 2008; 111 (12): 5446-56.
3. Rossi D, Cerri M, Capello D, Deambrogi C, et al. Biological and clinical risk factors of chronic lymphocytic leukaemia transformation to Richter syndrome. *Br J Haematol.* 2008; 142 (2): 202-15.

4. Tsimberidou A-M, Keating MJ. Richter syndrome. Biology, incidence, and therapeutic strategies. *Cancer* 2005; 103 (2): 216-28.
5. Maddocks-Christianson K, Slager SL, Zent CS, et al. Risk factors for development of a second malignancy in patients with chronic lymphocytic leukaemia. *Br J Haematol*. 2007; 139 (3): 398-404.
6. Cherepakhin V, Baird SM, Meisenholder GW, Kipps TJ. Common clonal origin of chronic lymphocytic leukemia and high grade lymphoma of Richter's syndrome. *Blood*. 1993; 82 (10): 3141-7.
7. Matolcsy A, Inghirami G, Knowles DM. Molecular genetic demonstration of the diverse evolution of Richter's syndrome (chronic lymphocytic leukemia and subsequent large cell lymphoma). *Blood*. 1994; 83 (5): 1363-72.
8. Nakamura N, Kuze T, Hashimoto Y, et al. Analysis of the immunoglobulin heavy chain gene of secondary diffuse large B-cell lymphoma that subsequently developed in four cases with B-cell chronic lymphocytic leukemia or lymphoplasmacytoid lymphoma (Richter syndrome). *Pathol Int*. 2000; 50 (8): 636-43.
9. Timár B, Fülöp Z, Csernus B, et al. Relationship between the mutational status of VH genes and pathogenesis of diffuse large B-cell lymphoma in Richter's syndrome. *Leukemia*. 2004; 18 (2): 326-30.
10. Smit LA, van Maldegem F, Langerak AW, et al. Antigen receptors and somatic hypermutation in B-cell chronic lymphocytic leukemia with Richter's transformation. *Haematologica*. 2006; 91 (7): 903-11.
11. Mao Z, Quintanilla-Martinez L, Raffeld M, et al. IgVH mutational status and clonality analysis of Richter's transformation: diffuse large B-cell lymphoma and Hodgkin lymphoma in association with B-cell chronic lymphocytic leukemia (B-CLL) represent 2 different pathways of disease evolution. *Am J Surg Pathol*. 2007; 31 (10): 1605-14.
12. O'Brien SM, Kantarjian HM, Thomas DA, et al. Alemtuzumab as treatment for residual disease after chemotherapy in patients with chronic lymphocytic leukemia. *Cancer*. 2003; 98 (12): 2657-63.
13. Janssens A, Berth M, De Paepe P, et al. EBV negative Richter's syndrome from a coexistent clone after salvage treatment with alemtuzumab in a CLL patient. *Am J Hematol*. 2006; 81 (9): 706-12.
14. Rossi D, Berra E, Cerri M, et al. Aberrant somatic hypermutation in transformation of follicular lymphoma and chronic lymphocytic leukemia to diffuse large B-cell lymphoma. *Haematologica*: 200691 (10): 1405-9.
15. Montoto S, Davies AJ, Matthews J, et al. Risk and clinical implications of transformation of follicular lymphoma to diffuse large B-cell lymphoma. *J Clin Oncol*. 2007; 25 (17): 2426-33.
16. Armitage JO, Dick FR, Corder MP. Diffuse histiocytic lymphoma complicating chronic lymphocytic leukemia. *Cancer*. 197841 (2): 422-7.
17. Ott MM, Ott G, Roblick U, et al. Localized gastric non-Hodgkin's lymphoma of high-grade malignancy in patients with pre-existing chronic lymphocytic leukemia or immunocytoma. *Leukemia*. 1995; 9 (4): 609-14.
18. Ratnavel RC, Dunn-Walters DK, Boursier L, et al. B-cell lymphoma associat-

- ed with chronic lymphocytic leukaemia: two cases with contrasting aggressive and indolent behaviour. *Br J Dermatol.* 1999; 140 (4): 708-14.
19. Parrens M, Sawan B, Dubus P, et al. Primary digestive Richter's syndrome. *Mod Pathol.* 2001; 14 (5): 452-7.
 20. Omoti CE, Omoti AE. Richter syndrome: a review of clinical, ocular, neurological and other manifestations. *Br J Haematol.* 2008; 142 (5): 709-16.
 21. Bruzzi JF, Macapinlac H, Tsimberidou AM, et al. Detection of Richter's transformation of chronic lymphocytic leukemia by PET/CT. *J Nucl Med.* 2006; 47 (8): 1267-73.
 22. Matolesy A, Chadburn A, Knowles DM. De novo CD5-positive and Richter's syndrome-associated diffuse large B cell lymphomas are genotypically distinct. *Am. J. Pathol.* 1995; 147 (1): 207-16.
 23. Hans CP, Weisenburger DD, Greiner TC, et al. Confirmation of the molecular classification of diffuse large B-cell lymphoma by immunohistochemistry using a tissue microarray. *Blood.* 2004; 103 (1): 275-82.
 24. Tsimberidou A-M, O'Brien S, Khouri I, et al. Clinical outcomes and prognostic factors in patients with Richter's syndrome treated with chemotherapy or chemoimmunotherapy with or without stem-cell transplantation. *J Clin Oncol.* 2006; 24 (15): 2343-51.
 25. Rossi D, Spina V, Deambrogi C, et al. The genetics of Richter syndrome reveals disease heterogeneity and predicts survival after transformation. *Blood.* 2011; 117 (12): 3391-401.
 26. Rossi D, Rasi S, Capello D, Gaidano G. Prognostic assessment of BCL2-938C>A polymorphism in chronic lymphocytic leukemia. *Blood.* 2008; 111 (1): 466-8.
 27. Rossi D, Zucchetto A, Rossi FM, et al. CD49d expression is an independent risk factor of progressive disease in early stage chronic lymphocytic leukemia. *Haematologica.* 2008; 93 (10): 1575-9.
 28. Rossi D, Spina V, Cerri M, et al. Stereotyped B-cell receptor is an independent risk factor of chronic lymphocytic leukemia transformation to Richter syndrome. *Clin Cancer Res.* 2009; 15 (13): 4415-22.
 29. Rossi D, Lobetti Bodoni C, Genuardi E, et al. Telomere length is an independent predictor of survival, treatment requirement and Richter's syndrome transformation in chronic lymphocytic leukemia. *Leukemia.* 2009; 23 (6): 1062-72.
 30. Calin GA, Ferracin M, Cimmino A, et al. A MicroRNA signature associated with prognosis and progression in chronic lymphocytic leukemia. *N Engl J Med.* 2005; 353 (17): 1793-801.
 31. Kienle DL, Korz C, Hosch B, et al. Evidence for distinct pathomechanisms in genetic subgroups of chronic lymphocytic leukemia revealed by quantitative expression analysis of cell cycle, activation, and apoptosis-associated genes. *J Clin Oncol.* 2005; 23 (16): 3780-92.
 32. Klein U, Lia M, Crespo M, et al. The DLEU2/miR-15a/16-1 cluster controls B cell proliferation and its deletion leads to chronic lymphocytic leukemia. *Cancer Cell.* 2010; 17 (1): 28-40.

33. Starostik P, Patzner J, Greiner A, et al. Gastric marginal zone B-cell lymphomas of MALT type develop along 2 distinct pathogenetic pathways. *Blood*. 2002; 99 (1): 3-9.
34. Deambrogi C, Cresta S, Cerri M, et al. 14q32 translocations and risk of Richter's transformation in chronic lymphocytic leukaemia. *Br J Haematol*. 2009; 144 (1): 131-3.
35. Stamatopoulos K, Belessi C, Moreno C, et al. Over 20% of patients with chronic lymphocytic leukemia carry stereotyped receptors: pathogenetic implications and clinical correlations. *Blood*. 2007; 109 (1): 259-70.
36. Bomben R, Dal Bo M, Capello D, et al. Molecular and clinical features of chronic lymphocytic leukaemia with stereotyped B cell receptors: results from an Italian multi center study. *Br J Haematol*. 2009; 144 (4): 492-506.
37. Sutton LA, Kostareli E, Hadzidimitriou A, et al. Extensive intraclonal diversification in a subgroup of chronic lymphocytic leukemia patients with stereotyped IGHV4-34 receptors: implications for ongoing interactions with antigen. *Blood*. 2009; 114 (20): 4460-8.
38. Darzentas N, Hadzidimitriou A, Murray F, et al. A different ontogenesis for chronic lymphocytic leukemia cases carrying stereotyped antigen receptors: molecular and computational evidence. *Leukemia*. 2010; 24 (1): 125-32.
39. Caligaris-Cappio F, Ghia P. Novel insights in chronic lymphocytic leukemia: are we getting closer to understanding the pathogenesis of the disease? *J Clin Oncol*. 2008; 26 (27): 4497-503.
40. Chu CC, CATERA R, Zhang L, et al. Many chronic lymphocytic leukemia antibodies recognize apoptotic cells with exposed nonmuscle myosin heavy chain IIA: implications for patient outcome and cell of origin. *Blood*. 2010; 115 (19): 3907-15.
41. Del Poeta G, Maurillo L, Venditti A, et al. Clinical significance of CD38 expression in chronic lymphocytic leukemia. *Blood*. 2001; 98 (9): 2633-9.
42. Ibrahim S, Keating M, Do KA, et al. CD38 expression as an important prognostic factor in B-cell chronic lymphocytic leukemia. *Blood*. 2001; 98 (1): 181-6.
43. Jaksic O, Paro MM, Kardum Skelin I, Kusec R, Pejisa V, Jaksic B. CD38 on B-cell chronic lymphocytic leukemia cells has higher expression in lymph nodes than in peripheral blood or bone marrow. *Blood*. 2004; 103 (5): 1968-9.
44. Deaglio S, Vaisitti T, Aydin S et al. CD38 and ZAP-70 are functionally linked and mark CLL cells with high migratory potential. *Blood*. 2007; 110 (12): 4012-21.
45. Deaglio S, Capobianco A, Bergui L, et al. CD38 is a signaling molecule in B-cell chronic lymphocytic leukemia cells. *Blood*. 2003; 102 (6): 2146-55.
46. Deaglio S, Vaisitti T, Aydin S, Ferrero E, Malavasi F. In-tandem insights from basic science combined with clinical research: CD38 as both marker and key component of the pathogenetic network underlying chronic lymphocytic leukemia. *Blood*. 2006; 108 (4): 1135-44.
47. Patten PE, Buggins AG, Richards J, et al. CD38 expression in chronic lym-

- phocytic leukemia is regulated by the tumor microenvironment. *Blood*. 2008; 111 (10): 5173-81.
48. Kishimoto H, Hoshino S, Ohori M, et al. Molecular mechanisms of human CD38 gene expression by retinoic acid. Identification of retinoic acid response element in the first intron. *J Biol Chem*. 1998; 273 (25): 15429-134.
 49. Ferrero E, Saccucci F, Malavasi F. The human CD38 gene: polymorphism, CpG island, and linkage to the CD157 (BST-1) gene. *Immunogenetics*. 1999; 49 (7-8): 597-604.
 50. Aydin S, Rossi D, Bergui L, et al. CD38 gene polymorphism and chronic lymphocytic leukemia: a role in Richter syndrome? *Blood*. 2008; 111 (12): 5646-53.
 51. Rasi S, Spina V, Brusca A, et al. A variant of the LRP4 gene affects the risk of chronic lymphocytic leukaemia transformation to Richter syndrome. *Br. J. Haematol.* (in press).
 52. Lu D, Zhao Y, Tawatao R, et al. Activation of the Wnt signaling pathway in chronic lymphocytic leukemia. *Proc Natl Acad Sci USA*. 2004; 101 (9): 3118-23.
 53. Li Y, Pawlik B, Elcioglu N, et al. LRP4 mutations alter Wnt/beta-catenin signaling and cause limb and kidney malformations in Cenani-Lenz syndrome. *Am J Hum Genet*. 2010; 86 (5), 696-706.
 54. Kim NW, Piatyszek MA, Prowse KR, et al. Specific association of human telomerase activity with immortal cells and cancer. *Science*. 1994; 266 (5193): 2011-5.
 55. Gaidano G, Ballerini P, Gong JZ, et al. p53 mutations in human lymphoid malignancies: association with Burkitt lymphoma and chronic lymphocytic leukemia. *Proc Natl Acad Sci USA*. 1991; 88 (12): 5413-7.
 56. Lee JN, Giles F, Huh YO, et al. Molecular differences between small and large cells in patients with chronic lymphocytic leukemia. *Eur. J. Haematol*. 2003; 71 (4): 235-42.
 57. Lo Coco F, Gaidano G, Louie DC, Offit K, Chaganti RS, Dalla-Favera R. p53 mutations are associated with histologic transformation of follicular lymphoma. *Blood*. 1993; 82 (8): 2289-95.
 58. Lossos IS. Higher-grade transformation of follicular lymphoma - a continuous enigma. *Leukemia*. 2005; 19 (8): 1331-3.
 59. Pasqualucci L, Neumeister P, Goossens T, et al. Hypermutation of multiple proto-oncogenes in B-cell diffuse large-cell lymphomas. *Nature*. 2001; 412 (6844): 341-6.
 60. Gaidano G, Pasqualucci L, Capello D, et al. Aberrant somatic hypermutation in multiple subtypes of AIDS-associated non-Hodgkin lymphoma. *Blood*. 2001; 102 (5): 1833-41.
 61. Libra M, Capello D, Ghoghini A, et al. Analysis of aberrant somatic hypermutation (SHM) in non-Hodgkin lymphomas of patients with HCV infection. *J Pathol*. 2005; 206 (1): 87-91.
 62. Klein U, Dalla-Favera R. Germinal centres: role in B-cell physiology and malignancy. *Nat Rev Immunol*. 2008; 8 (1): 22-33.

63. Pasqualucci L, Bhagat G, Jankovic M, et al. AID is required for germinal center-derived lymphomagenesis. *Nat Genetics*. 2008; 40 (1): 108-12.
64. Reiniger L, Böddör C, Bognár Á, et al. Richter's and prolymphocytic transformation of chronic lymphocytic leukemia are associated with high mRNA expression of activation-induced cytidine deaminase and aberrant somatic hypermutation. *Leukemia*. 2006; 20 (6): 1089-95.
65. Han T, Sadamori N, Ozer H, et al. Cytogenetic studies in 77 patients with chronic lymphocytic leukemia: correlations with clinical, immunologic, and phenotypic data. *J Clin Oncol*. 1984; 2 (10): 1121-32.
66. Hébert J, Jonveaux P, d'Agay MF, Berger R. Cytogenetic studies in patients with Richter's syndrome. *Cancer Genet Cytogenet*. 1994; 73 (1): 65-8.
67. Brynes RK, McCourty A, Sun NC, Koo CH. Trisomy 12 in Richter's transformation of chronic lymphocytic leukemia. *Am J Clin Pathol*. 1995; 104 (2): 199-203.
68. Caraway NP, Du Y, Zhang H-Z, Hayes K, Glassman AB, Katz RL. Numeric chromosomal abnormalities in small lymphocytic and transformed large cell lymphomas detected by fluorescence in situ hybridization of fine-needle aspiration biopsies. *Cancer*. 2000; 90 (2): 126-32.
69. Santulli B, Kazmierczak B, Napolitano R, et al. A 12q13 translocation involving the HMGI-C gene in Richter transformation of a chronic lymphocytic leukemia. *Cancer Genet Cytogenet*. 2000; 119 (1): 70-3.
70. Zenz T, Mertens D, Döhner H, Stilgenbauer S. Molecular diagnostics in chronic lymphocytic leukemia - pathogenetic and clinical implications. *Leuk. Lymphoma*. 2008; 49 (5): 864-73.
71. Lenz G, Staudt LM. Aggressive lymphomas. *N Engl J Med*. 2010; 362 (15): 1417-29.
72. Beà S, López-Guillermo A, Ribas M, et al. Genetic imbalances in progressed B-cell chronic lymphocytic leukemia and transformed large-cell lymphoma (Richter's syndrome). *Am J Pathol*. 2002; 161 (3): 957-68.
73. Scandurra M, Rossi D, Deambrogi C, et al. Genomic profiling of Richter's syndrome: recurrent lesions and differences with de novo diffuse large B-cell lymphomas. *Hematol Oncol*. 2010; 28 (2): 62-7.
74. Cohen Y, Da'as N, Libster D, Amir G, Berrebi A, Polliack A. Large-cell transformation of chronic lymphocytic leukemia and follicular lymphoma during or soon after treatment with fludarabine-rituximab containing regimens: natural history or related therapy complication. *Eur J Hematol*. 2002; 68 (2): 80-3.
75. Thornton PD, Bellas C, Santon A, et al. Richter's transformation of chronic lymphocytic leukemia. The possible role of fludarabine and the Epstein-Barr virus in its pathogenesis. *Leukemia Res*. 2005; 29 (4): 389-95.

Genomica e proteomica nella gestione dei pazienti con amiloidosi sistemica

Francesca Lavatelli, Giampaolo Merlini

Laboratorio di Analisi Chimico Cliniche, Centro per lo Studio e la Cura delle Amiloidosi Sistemiche, Fondazione IRCCS Policlinico San Matteo

Le amiloidosi sistemiche

Le amiloidosi sono malattie da alterata conformazione proteica, accomunate dalla presenza di depositi extracellulari di materiale amiloide, costituiti da aggregati fibrillari di proteine (1, 2). Tali depositi possono essere sistemici o localizzati in siti specifici. Nonostante le analogie morfologiche delle fibrille, almeno 28 proteine differenti sono state identificate come agenti eziologici di amiloidosi nell'uomo, 14 delle quali associate a forme sistemiche (3). In queste ultime, le fibrille originano da proteine circolanti, trasportate agli organi bersaglio attraverso il sangue. Accanto alla proteina principale, varie altre specie proteiche sono comunemente associate alle fibrille amiloidi; questo si traduce nel fatto che la composizione dei depositi di amiloide sia complessa ed eterogenea (4-7). Le fibrille formate dalle varie proteine condividono proprietà ultrastrutturali e tintoriali; in particolare, la presenza di birifrangenza verde brillante all'osservazione in luce polarizzata, dopo colorazione con rosso Congo, è specifica ed usata a scopo diagnostico.

Nelle amiloidosi sistemiche, il decorso della malattia, il trattamento e la prognosi dipendono in modo critico dalla natura della proteina che origina le fibrille. Appare pertanto chiaro come la tipizzazione dei depositi di amiloide sia un momento cruciale nella gestione clinica di queste patologie. La prevalenza relativa delle varie forme di amiloidosi varia a seconda delle regioni geografiche. La forma più frequente nei paesi occidentali è quella da catene leggere delle immunoglobuline (AL), malattia sporadica causata dalla deposizione di catene leggere libere immunoglobuliniche monoclonali. Tuttavia, le forme ereditarie (conseguenti a mutazioni in geni codificanti per proteine amiloidogeniche) o l'amiloidosi reattiva (associata a condizioni infiammatorie croniche) si osservano con incidenza particolarmente significativa in determinate aree geografiche o popolazioni di pazienti. Ad esempio l'amiloidosi da transtiretina con la mutazione pVal30Met è presente con elevata prevalenza nel nord del Portogallo, nel nord della Svezia e in alcune regioni del Giappone. Il trattamento differisce radicalmente tra le varie forme di amiloidosi, spaziando dalla chemioterapia nella forma

AL, al trapianto di fegato nella amiloidosi da transtiretina (ATTR). Appare pertanto chiaro come errori nella tipizzazione possano portare ad errori terapeutici catastrofici.

Le varie forme di amiloidosi hanno manifestazioni cliniche sovrapponibili, che rendono impossibile differenziarle senza l'aiuto di analisi di laboratorio e istologiche. Il *workflow* diagnostico richiede l'uso combinato di diversi approcci, al fine di dimostrare la presenza del precursore amiloidogenico, di rilevare mutazioni nel DNA associate a forme familiari della malattia, ed ad identificare la natura delle fibrille di amiloide. Tuttavia, la possibilità di giungere ad una tipizzazione errata è un problema significativo. L'analisi diretta delle fibrille nei tessuti affetti è la strategia conclusiva per una tipizzazione definitiva. Nella realtà clinica, ciò è solitamente ottenuto utilizzando metodi di immunoistochimica. Tuttavia, l'immunoistochimica, in questo contesto patologico, possiede specifiche limitazioni (prevalentemente dovute a scarsa sensibilità e specificità degli anticorpi verso le specie proteiche depositate, con alterata struttura), che rendono spesso poco attendibili i risultati (8-11).

La proteomica nella gestione clinica dei pazienti con amiloidosi sistemica

Il termine "proteomica" indica lo studio globale delle specie proteiche che costituiscono un campione biologico, basata sull'analisi mediante spettrometria di massa. Una caratteristica peculiare della proteomica consiste nel fatto che le proteine possono essere analizzate senza la necessità di anticorpi specifici, e senza ipotesi a priori sulla composizione del campione. A partire dai dati di spettrometria di massa, infatti, le specie presenti nel campione vengono identificate tramite analisi bioinformatica e ricerca in banche dati. La proteomica è stata applicata allo studio delle amiloidosi sistemiche in tempi relativamente recenti, ma le enormi potenzialità di questo approccio nella descrizione delle specie che compongono i depositi di amiloide e nello studio delle caratteristiche molecolari dei precursori amiloidogenici è apparso immediatamente chiaro. L'introduzione della proteomica nella pratica clinica ha rivoluzionato l'approccio alla tipizzazione delle amiloidosi (12).

Varie metodologie differenti sono state sviluppate, sia per l'analisi proteomica di campioni di tessuto fresco, non fissato, che per tessuti fissati in formalina ed inclusi in paraffina (13-16). Questi metodi sono stati ottimizzati per valutare quantità molto piccole di materiale, dell'ordine di quanto ottenibile mediante agoaspirazione, o di quanto ricavabile da sezioni montate su vetrini per istopatologia. Tali approcci possono essere divisi in tre tipologie:

- 1) gel-free, che non richiedono la separazione delle specie proteiche su gel (14-16);
- 2) gel-based, che prevedono la separazione delle proteine in elettroforesi bidimensionale (2D-PAGE) (13);
- 3) analisi proteomica diretta della sezione tissutale ("*imaging mass spectrometry*") (17).

La maggior parte delle procedure per tipizzare l'amiloide in sezioni tissutali sono

basate sulla selezione, preliminare all'analisi in spettrometria di massa, delle aree contenenti i depositi, tramite micro-dissezione laser (*"laser capture microdissection"*) (15). Le proteine presenti nelle aree selezionate vengono quindi estratte, processate, analizzate ed identificate con spettrometria di massa e bioinformatica. Questo approccio si è dimostrato utile sia per identificare forme comuni di amiloidosi, che forme rare (3), ed è oggi considerato uno dei nuovi metodi gold standard per la tipizzazione di queste malattie. Un approccio alternativo, recentemente sviluppato dal nostro gruppo (16), permette invece la caratterizzazione mediante analisi dell'intero frammento tissutale, senza effettuare preliminarmente la fase di micro dissezione. Questa metodica è stata messa a punto, in particolare, per la valutazione di campioni di grasso periumbelicale sottocutaneo, tessuto di scelta per la valutazione istologica dei depositi di amiloide nei pazienti in cui se ne sospetti la presenza. Questo approccio è basato sulla cosiddetta tecnologia MudPIT (Multidimensional Protein Identification Technology), che prevede, come fase preliminare alla spettrometria di massa, la digestione del campione in peptidi e la separazione degli stessi mediante cromatografia bidimensionale. L'analisi permette l'identificazione di centinaia di proteine differenti in ogni campione; l'assegnazione del tipo di amiloidosi richiede la comparazione con il proteoma del corrispondente tessuto non affetto (ottenuto da donatori), ed è basa-

ene	Diagnos	LC κ	LC λ	TTR	S				
P6		13	212	6	6				
P11		6	165	4	0				
P2		0	130	0	0				
P1		0	88	6	0				
P9		13	61	9	0				
P12	ALλ	8	50	0	0				
P13		14	34	0	0				
P8		7	33	0	0				
P4		0	22	0	0				
P3		0	21	0	0				
P5		0	13	0	0				
P10		0	6	0	0				
P15		372	0	0	0				
P14	ALκ	176	0	0	0				
P25		44	0	0	0				
P26		14	0	0	0				
P19		4	0	1158	0				
P20		0	0	105	0				
P16	ATTR	4	0	145	0				
P18		0	0	89	0				
P7		4	0	16	0				
P22		0	0	0	6				
P23		0	0	0	2				
P17	SAA	10	0	0	1				
P24		0	0	0	9				
P21		0	0	0	7				
Cm		4	2	1	1				
lege		0	1-2	3-5	6-9	10-	20-	30-49	>50

Fig. 1 - Tipizzazione, mediante analisi MudPIT, delle proteine presenti nei depositi di amiloide a livello del tessuto adiposo sottocutaneo addominale, in pazienti con varie forme di amiloidosi. Le cifre nella tabella indicano il numero di spettri di massa (*"spectral count"*) attribuiti a ciascuna delle proteine più frequentemente responsabili di amiloidosi sistemica, indicate in alto (LC, immunoglobulin light chains, TTR, transthyretin, SAA, Serum Amyloid A). I diversi colori sono proporzionali per intensità allo spectral count. I campioni indicati con p provengono da pazienti; quelli indicati con c si riferiscono ai controlli non affetti.

sull'assunto che le proteine anomale (depositate) debbano essere sovra rappresentate nei malati rispetto ai controlli (Fig. 1).

Oltre alle proteine costituenti le fibrille, questi approcci permettono anche di identificare le altre specie proteiche associate ai depositi di amiloide, in primo luogo la siero amiloide P (SAP) e le apolipoproteine (ApoE ed ApoA-IV).

Anche la procedura basata su elettroforesi bidimensionale è centrata sull'analisi differenziale tra il proteoma dell'intero tessuto affetto e quello della controparte non affetta. In questa tecnica, le proteine vengono separate sulla base di peso molecolare e punto isoelettrico e vengono visualizzate su gel. Le proteine anomale (non visibili nel controllo e quindi ipotizzate derivare dai depositi di amiloide) possono essere quindi escisse, processate e identificate mediante spettrometria di massa. Rispetto agli altri approcci, quello basato su elettroforesi bidimensionale permette l'immediata visualizzazione delle specie proteiche ed è il più efficiente nel fornire un quadro globale di tutte le specie ed isoforme che compongono i depositi (Fig. 2).

Un nuovo, promettente approccio, applicato anche all'analisi di tessuti affetti da amiloidosi, è la cosiddetta "imaging mass spectrometry" (17). Questa tecnica per-

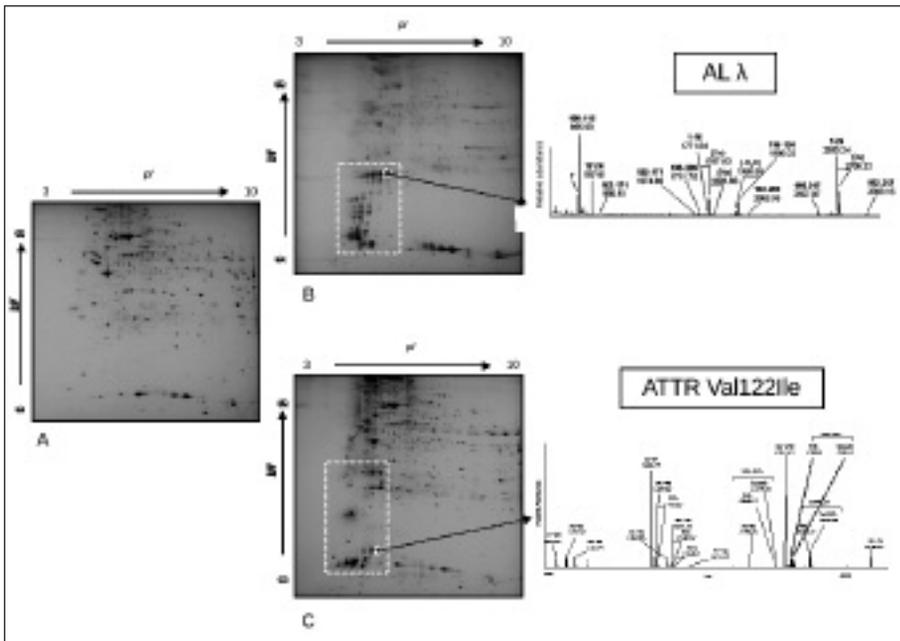


Fig. 2 - Tipizzazione dei depositi di amiloide nel tessuto adiposo sottocutaneo addominale, tramite elettroforesi bidimensionale e spettrometria di massa. A) mappa proteomica del tessuto adiposo di controllo; B) amiloidosi AL λ ; C) amiloidosi ATTR (variante Val122Ile). La comparazione fra la mappa dei pazienti e quella dei controlli permette di localizzare prominenti specie anomale nei campioni affetti (riquadri), che migrano in regioni del gel consistenti, per peso molecolare e punto isoelettrico, con quelle attese per i rispettivi precursori amiloidogenici. In aggiunta, specie corrispondenti a frammenti o polimeri della proteina amiloidogenica sono osservabili. L'identificazione delle specie anomale, e quindi la tipizzazione dell'amiloidosi, è ottenuta mediante spettrometria di massa ed analisi bioinformatica.

mette di mantenere l'integrità del tessuto durante l'analisi proteomica, preservando di conseguenza la distribuzione spaziale delle proteine. In questo approccio, le specie proteiche vengono processate in situ, e la fettina di tessuto è direttamente inserita all'interno dello spettrometro. Gli spettri di massa sono ottenuti da punti ravvicinati del campione; ogni spettro contiene le informazioni riguardanti le proteine presenti nella specifica posizione in cui è stato acquisito. Combinando le informazioni ricavate da punti distinti, si possono ottenere specifiche immagini virtuali, che mostrino la distribuzione spaziale di ciascuna specie proteica. Il prodotto finale di questa tecnica di analisi sono immagini analoghe a quelle ottenibili al microscopio in istologia tradizionale.

In quanto malattie da deposizione proteica, le amiloidosi sono un campo d'applicazione ideale per la proteomica come strumento diagnostico.

Questo si è tradotto nel fatto che gli approcci sopra descritti siano oggi utilizzati correntemente nella pratica clinica, collocando le amiloidosi fra i pochi esempi in cui la proteomica è stata introdotta nella routine diagnostica.

La genomica nella gestione clinica dei pazienti con amiloidosi sistemica

Anche l'analisi del precursore amiloide basato sulla genomica (*i.e.* valutando l'assetto dei geni che ne influenzano la produzione), anziché sulla proteomica (che valuta il prodotto dell'espressione di tali geni), ha un ruolo di primo piano nella gestione clinica delle amiloidosi sistemiche.

Il ruolo determinante del corredo genetico nel determinare l'insorgenza di specifiche forme di amiloidosi è noto da tempo ed è alla base della definizione e della diagnosi delle forme di amiloidosi ereditarie. In particolare, la valutazione di singoli, specifici geni (ad esempio transtiretina, apolipoproteina A-I ed A-II, lisozima, fibrinogeno), per rilevare mutazioni associate con lo sviluppo di malattia è uno strumento consolidato ed indispensabile nella gestione clinica dei pazienti.

Tuttavia, in un'accezione più ampia, l'analisi genomica ha permesso recentemente di formulare nuovi paradigmi sulla patogenesi di alcune forme di amiloidosi.

Un esempio è rappresentato dai polimorfismi in geni associati ad amiloidosi reattiva, quali SAA e MEFV (18, 19), che sono stati dimostrati influenzare in modo sostanziale il rischio di sviluppo e la severità della malattia.

Anche nella amiloidosi AL, lo studio della genetica delle catene leggere amiloidogeniche ha un impatto profondo sulla definizione delle proprietà patogenetiche. Diversi studi, infatti, hanno dimostrato una correlazione fra specifici geni *germline* codificanti la regione variabile delle catene leggere, ed il particolare tipo di tropismo d'organo. Ad esempio è stata confermata da almeno tre gruppi di ricerca l'associazione fra l'uso del gene *germline* *IGVL6-57* (noto in precedenza come 6a) da parte della catena leggera amiloidogenica e la localizzazione renale (20-22). Similmente, la localizzazione ai tessuti molli e articolazioni è significativamente più frequente nei pazienti con catene leggere amiloidogeniche che appartengono alla famiglia KI (23). Più recentemente il nostro gruppo ha riportato una correlazione altamente significativa fra l'uso del gene *germline* *IGLV1-44* e preponderante coinvolgimento cardiaco dei depositi di amiloide (24).

L'approfondimento di tale aspetto potrebbe condurre alla identificazione di specifici aspetti molecolari che spieghino la drammatica eterogeneità clinica di questa forma di amiloidosi, e potrebbero permettere lo studio di terapie specifiche per contrastarne gli effetti patogeni.

Bibliografia

1. Merlini G, Bellotti V. Molecular mechanisms of amyloidosis. *N Engl J Med*. 2003; 349 (6): 583-96.
2. Merlini G, Seldin DC, Gertz MA. Amyloidosis: pathogenesis and new therapeutic options. *J Clin Oncol*. 2011; 29 (14): 1924-33.
3. Sipe JD, Benson MD, Buxbaum JN, et al. Amyloid fibril protein nomenclature: 2010 recommendations from the nomenclature committee of the International Society of Amyloidosis. *Amyloid*. 2010; 17 (3-4): 101-4.
4. Bergstrom J, Murphy C, Eulitz M, et al. Codeposition of apolipoprotein A-IV and transthyretin in senile systemic (ATTR) amyloidosis. *Biochem Biophys Res Commun*. 2001; 285 (4): 903-8.
5. Elimova E, Kisilevsky R, Szarek WA, et al. Amyloidogenesis recapitulated in cell culture: a peptide inhibitor provides direct evidence for the role of heparan sulfate and suggests a new treatment strategy. *Faseb J*. 2004; 18 (14): 1749-51.
6. Pepys MB, Rademacher TW, Amatayakul-Chantler S, et al. Human serum amyloid P component is an invariant constituent of amyloid deposits and has a uniquely homogeneous glycostructure. *Proc Natl Acad Sci USA*. 1994; 91 (12): 5602-6.
7. Kisilevsky R. The relation of proteoglycans, serum amyloid P and apo E to amyloidosis current status, 2000. *Amyloid*. 2000; 130 (2-3): 99-108.
8. Lachmann HJ, Booth DR, Booth SE, et al. Misdiagnosis of hereditary amyloidosis as AL (primary) amyloidosis. *N Engl J Med*. 2002; 346 (23): 1786-91.
9. Comenzo RL, Zhou P, Fleisher M, et al. Seeking confidence in the diagnosis of systemic AL (Ig light-chain) amyloidosis: patients can have both monoclonal gammopathies and hereditary amyloid proteins. *Blood*. 2006; 107 (9): 3489-91.
10. Picken MM. Amyloidosis-where are we now and where are we heading? *Arch Pathol Lab Med*. 2010; 134 (4): 545-51.
11. Solomon A, Murphy CL, Westermark P. Unreliability of immunohistochemistry for typing amyloid deposits. *Arch Pathol Lab Med*. 2008; 132 (1): 14.
12. Lavatelli F, Vrana JA. Proteomic typing of amyloid deposits in systemic amyloidoses. *Amyloid*. 2011; 18 (4): 177-82.
13. Lavatelli F, Perlman DH, Spencer B, et al. Amyloidogenic and associated proteins in systemic amyloidosis proteome of adipose tissue. *Mol Cell Proteomics*. 2008; 7 (8): 1570-83.
14. Murphy CL, Wang S, Williams T, et al. Characterization of systemic amyloid deposits by mass spectrometry. *Methods Enzymol*. 2006; 412: 48-62.

15. Vrana JA, Gamez JD, Madden BJ, et al. Classification of amyloidosis by laser microdissection and mass spectrometry-based proteomic analysis in clinical biopsy specimens. *Blood*. 2009; 114 (24): 4957-9.
16. Brambilla F, Lavatelli F, Di Silvestre D, et al. Reliable typing of systemic amyloidoses through proteomic analysis of subcutaneous adipose tissue. *Blood*. 2012; 119 (8): 1844-7.
17. Groseclose MR, Andersson M, Hardesty WM, et al. Identification of proteins directly from tissue: in situ tryptic digestions coupled with imaging mass spectrometry. *J Mass Spectrom*. 2007; 42 (2): 254-62.
18. Bakkaloglu A, Duzova A, Ozen S, et al. Influence of Serum Amyloid A (SAA1) and SAA2 gene polymorphisms on renal amyloidosis, and on SAA/C-reactive protein values in patients with familial Mediterranean fever in the Turkish population. *J Rheumatol*. 2004; 31 (6): 1139-42.
19. Gershoni-Baruch R, Brik R, Zacks N, et al. The contribution of genotypes at the MEFV and SAA1 loci to amyloidosis and disease severity in patients with familial Mediterranean fever. *Arthritis Rheum*. 2003; 48 (4): 1149-55.
20. Comenzo RL, Zhang Y, Martinez C, Osman K, Herrera GA. The tropism of organ involvement in primary systemic amyloidosis: contributions of Ig V-L germ line gene use and clonal plasma cell burden. *Blood*. 2001; 98 (3): 714-720.
21. Perfetti V, Casarini S, Palladini G, et al. Analysis of V lambda-J lambda expression in plasma cells from primary (AL) amyloidosis and normal bone marrow identifies 3r (lambda III) as a new amyloid-associated germline gene segment. *Blood*. 2002; 100 (3): 948-953.
22. Abraham RS, Geyer SM, Price-Troska TL, et al. Immunoglobulin light chain variable (V) region genes influence clinical presentation and outcome in light chain-associated amyloidosis (AL). *Blood*. 2003; 101 (10): 3801-3808.
23. Prokhaeva T, Spencer B, Kaut M, et al. Soft tissue, joint, and bone manifestations of AL amyloidosis: clinical presentation, molecular features, and survival. *Arthritis Rheum*. 2007; 56 (11): 3858-3868.
24. Perfetti V, Palladini G, Casarini S, et al. The repertoire of lambda light chains causing predominant amyloid heart involvement and identification of a preferentially involved germline gene, IGLV1-44. *Blood*. 2012; 119 (1): 144-150.

Profilo di espressione genica nelle SMD e LAM: attualità e prospettive

Sergio Ferrari

Dipartimento di Scienze della Vita, Università di Modena e Reggio Emilia, Modena

Le leucemie mieloidi acute (LMA) e le sindromi mielodisplastiche (SMD) rappresentano dei disordini molto eterogenei sia dal punto di vista clinico, morfologico, immunologico e genetico (1-3). Le alterazioni genetiche qualitative più frequenti riguardano traslocazioni, inversioni, delezioni amplificazioni, inserzioni e mutazioni puntiformi a carico di numerosi geni di regolazione come fattori di crescita, recettori per fattori di crescita, trasduttori del segnale, efattori trascrizionali, e geni oncosoppressori. Anche le alterazioni quantitative del cariotipo sono molteplici, monosomie, trisomie, cariotipi complessi, perdita o acquisizione di intere regioni cromosomiche (guadagno o perdita di funzioni) ecc. Negli anni recenti si sono aggiunte alle tecniche di studio classiche come la morfologia, l'immunofenotipo, la citogenetica, tecniche come la citogenetica e la genetica molecolare che hanno portato ad una più accurata caratterizzazione sia delle LMA che delle MDS e all'identificazione di numerosi marcatori 4-6). Dal 2000 in avanti si sono sviluppate nuove tecnologie definite omiche (genomiche, trascrittomiche, proteomiche) che applicate a queste patologie hanno permesso di acquisire nuove informazioni sia diagnostiche che prognostiche. In particolare, i profili di espressione genica (PEG), basata sui microarray, hanno contribuito a ridefinire nelle MDS e LM:

- a) alla tassonomia molecolare;
- b) alla diagnosi;
- c) alla prognosi;
- d) alla scoperta di nuove sottoclassi;
- d) alla identificazione di geni bersagli per la terapia e quindi lo sviluppo di nuovi farmaci.

Inoltre le tecnologie genomiche certamente contribuiranno a migliorare la nostra comprensione dei meccanismi molecolari della trasformazione e progressione neoplastica nelle sindromi mieloproliferative in generale ed in particolare nelle LMA e nelle SMD (7, 8). L'applicazione su migliaia di casi di SMD/LMA della tecnologia dei DNA micrarray, studio MILE (Microarray Innovation in Leukemia), ha certamente contribuito alla:

1. *predizione di classe (class prediction)*, cioè l'identificazione di caratteristici patterns di espressione genica nelle sottoclassi delle SMD/LMA associate a specifiche alterazioni sia citogenetiche che molecolari (9-11).
2. *Identificazioni di nuove classi (class discovery)*, cioè la scoperta di nuove sottoclassi di leucemie (9).
3. *Predizione prognostica (outcome prediction)*, cioè predire la risposta alla terapia standard sia nelle SMD che nelle LMA dell'adulto e pediatriche (12-14).
4. *Scoperta di nuovi farmaci (drug discovery)*, cioè l'identificazione di farmaci in grado di interferire con geni bersaglio, cioè geni coinvolti nei meccanismi patogenetici delle SMD e delle LMA, ed in particolare geni coinvolti nelle vie di traduzione del segnale deregolati. In questi casi il PEG orienta anche sulla possibile risposta al farmaco del paziente (15, 16).

Il profilo di espressione genica nelle leucemie mieloidi acute dell'adulto

Il programma internazionale multicentrico (studio MILE), centrato sull'ELN (European Leukemia Network), ha riguardato uno studio retrospettivo di fase I per la scoperta di nuovi marcatori ed ha utilizzato la tecnologia dei DNA microarray per lo studio dei profili di espressione genica. Tale studio è stato fatto su 2.143 pazienti con diversi tipi di leucemie acute e croniche e SMD. A questo primo studio è seguito uno studio prospettico di fase II con 1.191 pazienti reclutati. Il primo studio retrospettivo di fase I ha permesso di classificare 18 distinti tipi di leucemie con il 92,2% di accuratezza e una specificità di media del 99,7% su 2.096 pazienti. Fra i pazienti studiati erano inclusi 542 casi di LMA con cariotipi complessi, aberranti, cariotipi normali e diverse traslocazioni. Nel secondo studio prospettico di fase II, l'analisi di 1.152 pazienti sia con LMA che LLA ha raggiunto una sensibilità media del 95,6% e una specificità del 99,8% (11). Altri studi multicentrici hanno evidenziato PEG specifici in pazienti con mutazione di NPM1, di CEBPa, over-espressione di FLT3 e nei casi di endo-duplicazione FLT3-ITD. Recentemente un insieme di 133 geni (signature) ha assunto un significato prognostico indipendente in tutte le sottoclassi di AML con alterazioni citogenetiche o cariotipo normale (12). Identificare quindi l'impatto prognostico della combinazione di tutti questi marcatori rappresenta un'alternativa al sequenziamento multiplex della pleora dei marcatori attualmente conosciuti. Inoltre, mutazioni geniche casuali hanno un impatto prognostico nettamente inferiore rispetto alla mutazione di geni "driver," cioè geni rilevanti funzionalmente e coinvolti nei meccanismi patogenetici. Lo dimostra l'evidenza che solo mutazioni di geni "driver" e in grado di modificare il GEP.

Il profilo di espressione genica nelle leucemie mieloidi acute pediatriche

Anche le AML pediatriche sono eterogenee come quelle dell'adulto, ma con la fondamentale differenza che nei bambini la sopravvivenza è superiore al 60%. In un primo studio di 130 pazienti pediatriche reclutati de novo (17) e in un secondo studio di 237 bambini con AML, il PEG solamente in alcuni tipi di AML con tras-

locazioni specifiche, inversioni e nei casi di leucemia megacarioblastica acuta, ha evidenziato una accuratezza di sottoclassificazione del 92-93% e una sensibilità del 99%. A differenza dell'adulto, non vi è un PEG predittivo per le alterazioni genetiche che riguardano NPM1, CEBPa; FLT3-ITD o KIT. Una larga casistica di pazienti pediatrici con LLA o LMA avevano un comune pattern di espressione genica quando l'alterazione genetica riguardava fusioni geniche con MLL (17). Alcuni geni come RUNX3, ATRX e PRAME sono rilevanti per le "signature prognostiche". Solo lo studio PEG di casistiche pediatriche più consistenti potrà portare a nuove prospettive anche nelle LMA pediatriche, come si propone il network di eccellenza dell'azione COST "EuGESMA".

Il profilo di espressione genica nelle MDS (14, 18-27)

Le SMD sono molto più eterogenee delle LMA. Questo aspetto è sicuramente uno svantaggio per la classificazione basata sui PEG. Molti studi hanno comparato i PEG di cellule staminali emopoietiche CD34+ o CD133+ ottenute da individui normali con staminali ottenute da pazienti con MDS. Altri studi hanno utilizzato semplicemente cellule mononucleate ottenute dal midollo osseo. In tutti e due i casi si sono tentate "signature" molecolari diagnostiche o prognostiche con scarsi risultati. Alterazioni dell'espressione genica nelle SMD riguardano la deregolazione di geni mitocondriali, trasportatori (Canali ABCB7), geni omeotici, geni ribosomali. Specifiche "signature" sono state ottenute in pazienti con MDS con 5q-, -7/del (7q), e trisomia dell'8. Alcune vie di segnalazione sembrano deregolate nelle staminali delle SMD e fra queste la via dell'interferone, della trombopoietina, di Wnt. L'accuratezza di predizione di classe non supera il 50%. Alcune "signature" sono anche prognostiche. Casi a basso rischio sono arricchite in geni ribosomali, Myc e Wnt.

Il PEG come predittore di risposta o resistenza ai farmaci o per predire nuovi farmaci potenzialmente terapeutici

"Signature" specifiche sono state descritte per la sensibilità all'ATRA dei blasti leucemici mieloidi. (28), la sensibilità agli inibitori delle farnesil-transferasi (FTI) (15), della risposta alla lenalidomide nelle SMD con 5q- (16). Lo studio di GEP predittivi della sensibilità al trattamento farmacologico richiederà una validazione clinica fatta su grandi casistiche. Altra prospettiva dei GEP è quella di facilitare, mediante l'individuazione di vie di trasduzione del segnale deregolate, composti in grado di interagire specificamente con i geni target (terapia bersaglio), deregolati in distinte sottoclassi di leucemie o SMD o, nella migliore delle ipotesi, in singoli pazienti (terapia personalizzata). La deregolazione di vie di segnalazione che coinvolgono Ras, E2F, EGFR, AML1/ETO o NF-Kb in LMA o SMD offrono la possibilità di una terapia bersaglio mediante la identificazione di composti interferenti. Diversi studi stanno affrontando il confronto fra i blasti leucemici e la loro controparte normale. Il confronto fra promileociti normali e leucemici ha portato all'identificazione di farmaci inibenti la funzione di alcuni

geni. Ne sono esempi la Tricostatina A, inibente l'istone deacetilasi; LY294002, inibitore della PI-3chinasi; Quinacrina, inibitore della fosfolipasi A2 (29, 30).

Prospettive dei GEP in associazione con le altre tecnologie omiche

Numerosi studi recenti hanno dimostrato che la de-regolazione di micro-RNA contribuisce ai meccanismi di leucemogenesi (31). Inoltre i profili di espressione di micro-RNA su scala genomica correlano con le sottoclassi citogenetiche (32, 33) e le sottoclassi genetico- molecolari quali le LMA con mutazioni di CEBPa e NPM1 (32). Inoltre la recente osservazione che il meccanismo d'azione dei micro-RNA è primariamente a livello degli RNA messaggeri, mediante la destabilizzazione del RNA bersaglio più che la inibizione della traduzione, ha permesso di sviluppare strumenti bio-informatici per l'analisi integrata dei GEP e dei micro-RNA. Fra le tecnologie "omiche," applicate alle LMA e alle SMD, vi sono le analisi dell'intero genoma mediante array CGH o SNP che hanno rivelato un grande numero di variazioni del numero di copie (Copy number alterations-CNAs) come pure una acquisita omozigotità nella forma di disomia uniparentale segmentale, UPD (34). L'analisi comparativa di queste diverse tecnologie ha permesso di definire non solo le regioni genomiche coinvolte, ma anche i geni deregolati (35). La caratterizzazione di queste regioni è attualmente facilitata dalla possibilità di isolare il DNA di queste regioni e sequenziarle mediante NGS. I risultati ottenuti sono significativi nelle LMA dell'adulto, ma non nelle LLA o nelle LMA pediatriche. Questi risultati dimostrano l'estrema eterogeneità e variabilità del genoma nei meccanismi di leucemogenesi. Inoltre l'intero genoma di cellule leucemiche è stato sequenziato con risultati difficili da interpretare. Anche i "patterns" di metilazione dei promotori genici sono stati studiati mediante un approccio "genome-wide". Lo sforzo futuro sarà quello di combinare l'analisi del pattern di metilazione con i GEP. Infine un altro approccio combinato sarà quello di comparare i GEP con gli studi di proteomica. Questi studi sono rilevanti in quanto si comparano diversi livelli della regolazione o de-regolazione dell'espressione genica, quello trascrizionale, post trascrizionale e traduzionale. Le varianti di "splicing" e le diverse isoforme proteiche sono rilevanti per l'interpretazione funzionale delle patologie in esame. Il profilo proteomico permette l'identificazione di pattern associati al fenotipo neoplastico, alle alterazioni genetiche e citogenetiche e potrebbe rivelarsi importante dal punto di vista prognostico. Esistono già numerosi esempi in clinica di terapie bersaglio che impiegano inibitori di specifiche proteine, come già sinteticamente riportato. Da segnalare infine la rilevanza degli studi del "chinesoma nelle LMA (36).

Bibliografia

1. Corey SJ, Minden MD, Barber DL, Kantarjian H, Wang JC, Schimmer AD. Myelodysplastic syndromes: the complexity of stem-cell diseases. *Nat Rev Cancer*. 2007; 7 (2): 118-29.
2. Nimer SD. Myelodysplastic syndromes. *Blood*. 2008; 111 (10): 4841-51.

3. Dohner K, Dohner H. Molecular characterization of acute myeloid leukemia. *Haematologica* 2008; 93 (7): 976-82.
4. Ebert BL. Genetic deletions in AML and MDS. *Best Pract Res Clin Haematol* 2010; 23: 457-61.
5. Jadersten M, Hellstrom-Lindberg E. New clues to the molecular pathogenesis of myelodysplastic syndromes. *Exp Cell Res* 2010; 316: 1390-6.
6. Haferlach T, Bacher U, haferlach C, Kern W, Schnitter S. Inside the molecular pathogenesis of myeloid malignancies. *Curr Opin Haematol* 2007; 14: 90-7.
7. Wouters BJ, Lowenberg B, Delwel R. A decade of genome-wide gene expression profiling in acute myeloid leukemia: flashback and prospects. *Blood*. 2009; 113 (2): 291-8.
8. Theilgaard-Monch K, Boulwood J, Ferrari S, Giannopoulos K, Hernandez-Riva JM, Kohlmann A et al. gene expression profiling in MDS and AML: pitfalls and potential. *Leukemia*. 2011.
9. Bullinger L, Dohner K, Bair E, Frohling S, Schlenk RF, Tibshirani R, et al. Use of gene-expression profiling to identify prognostic subclasses in adult acute myeloid leukemia. *N Engl J Med*. 2004; 350 (16): 1605-16.
10. Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, et al. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science*. 1999; 286 (5439): 531-7.
11. Haferlach T, Kohlmann A, Wiczorek L, Basso G, Te Kronnie G, Bene MC, et al. Clinical utility of microarray-based gene expression profiling in the diagnosis and subclassification of leukemia: report from the international microarray innovations in leukemia Study Group. *J Clin Oncol*. 2010; 20.
12. Bullinger L, Dohner K, Kranz R, Stirner C, Frohling S, Scholl C, et al. An FLT3 gene-expression signature predicts clinical outcome in normal karyotype AML. *Blood*. 2008; 111 (9): 4490-5.
13. Valk PJ, Verhaak RG, Beijen MA, Erpelinck CA, Barjesteh van Waalwijk van Doorn-Khosrovani S, Boer JM, et al. Prognostically useful gene-expression profiles in acute myeloid leukemia. *N Engl J Med*. 2004; 350 (16): 1617-28.
14. Mills KI, Kohlmann A, Williams PM, Wiczorek L, Liu WM, Li R, et al. Microarray-based classifiers and prognosis models identify subgroups with distinct clinical outcomes and high risk of AML transformation of myelodysplastic syndrome. *Blood*. 2009; 114 (5): 1063-72.
15. Raponi M, Lancet JE, Fan H, Dossey L, Lee G, Gojo I, et al. A 2-gene classifier for predicting response to the farnesyltransferase inhibitor tipifarnib in acute myeloid leukemia. *Blood*. 2008; 111 (5): 2589-96.
16. Ebert BL, Galili N, Tamayo P, Bosco J, Mak R, Pretz J, et al. An erythroid differentiation signature predicts response to lenalidomide in myelodysplastic syndrome. *PLoS Med*. 2008; 5 (2): e35. 30.
17. Ross ME, Mahfouz R, Onciu M, Liu HC, Zhou X, Song G, et al. Gene expression profiling of pediatric acute myelogenous leukemia. *Blood*. 2004; 104 (12): 3679-87.
18. Mano H. DNA micro-array analysis of myelodysplastic syndrome. *Leuk Lymphoma*. 2006; 47 (1): 9-14.

19. Pellagatti A, Fidler C, Wainscoat JS, Boultonwood J. Gene expression profiling in the myelodysplastic syndromes. *Hematology*. 2005; 10 (4): 281-7.
20. Pellagatti A, Cazzola M, Giagounidis AA, Malcovati L, Porta MG, Killick S, et al. Gene expression profiles of CD34+ cells in myelodysplastic syndromes: involvement of interferon-stimulated genes and correlation to FAB subtype and karyotype. *Blood*. 2006; 108 (1): 337-45.
21. Boultonwood J, Pellagatti A, Nikpour M, Pushkaran B, Fidler C, Cattani H, et al. The role of the iron transporter ABCB7 in refractory anemia with ring sideroblasts. *PLoS One*. 2008; 3 (4): e1970.
22. Nikpour M, Pellagatti A, Liu A, Karimi M, Malcovati L, Gogvadze V, et al. Gene expression profiling of erythroblasts from refractory anaemia with ring sideroblasts (RARS) and effects of G-CSF. *Br J Haematol*. 2010; 12.
23. Nilsson L, Eden P, Olsson E, Mansson R, Astrand-Grundstrom I, Strombeck B, et al. The molecular signature of MDS stem cells supports a stem-cell origin of 5q myelodysplastic syndromes. *Blood*. 2007; 110 (8): 3005-14.
24. Pellagatti A, Hellstrom-Lindberg E, Giagounidis A, Perry J, Malcovati L, Della Porta MG, et al. Haploinsufficiency of RPS14 in 5q- syndrome is associated with deregulation of ribosomal- and translation-related genes. *Br J Haematol*. 2008; 142 (1): 57-64.
25. Boultonwood J, Pellagatti A, Cattani H, Lawrie CH, Giagounidis A, Malcovati L, et al. Gene expression profiling of CD34+ cells in patients with the 5q- syndrome. *Br J Haematol*. 2007; 139 (4): 578-89.
26. Sridhar K, Ross DT, Tibshirani R, Butte AJ, Greenberg PL. Relationship of differential gene expression profiles in CD34+ myelodysplastic syndrome marrow cells to disease subtype and progression. *Blood*. 2009; 114 (23): 4847-58.
27. Pellagatti A, Cazzola M, Giagounidis A, Perry J, Malcovati L, Della Porta MG, et al. Deregulated gene expression pathways in myelodysplastic syndrome hematopoietic stem cells. *Leukemia*. 2010; 24 (4): 756-64.
28. Tagliafico E, Tenedini E, Manfredini R, Grande A, Ferrari F, Roncaglia E, et al. Identification of a molecular signature predictive of sensitivity to differentiation induction in acute myeloid leukemia. *Leukemia*. 2006; 20 (10): 1751-8.
29. Hassane DC, Guzman ML, Corbett C, Li X, Abboud R, Young F, et al. Discovery of agents that eradicate leukemia stem cells using an in silico screen of public gene expression data. *Blood*. 2008; 111 (12): 5654-62.
30. Marstrand TT, Borup R, Willer A, Borregaard N, Sandelin A, Porse BT, et al. A conceptual framework for the identification of candidate drugs and drug targets in acute promyelocytic leukemia. *Leukemia* 2010; 27.
31. Chen J, Odenike O, Rowley JD. Leukaemogenesis: more than mutant genes. *Nat Rev Cancer*. 2010; 10 (1): 23-36.
32. Jongen-Lavrencic M, Sun SM, Dijkstra MK, Valk PJ, Lowenberg B. MicroRNA expression profiling in relation to the genetic heterogeneity of acute myeloid leukemia. *Blood* 2008; 111 (10): 5078-85.
33. Li Z, Lu J, Sun M, Mi S, Zhang H, Luo RT, et al. Distinct microRNA expres-

- sion profiles in acute myeloid leukemia with common translocations. *Proc Natl Acad Sci USA*. 2008; 105 (40): 15535-40.
34. Maciejewski JP, Mufti GJ. Whole genome scanning as a cytogenetic tool in hematologic malignancies. *Blood*. 2008; 112 (4): 965-74.
35. Rucker FG, Bullinger L, Schwaenen C, Lipka DB, Wessendorf S, Frohling S, et al. Disclosure of candidate genes in acute myeloid leukemia with complex karyotypes using microarray-based molecular characterization. *J Clin Oncol*. 2006; 24 (24): 3887-94.
36. Kornblau SM, Tibes R, Qiu YH, Chen W, Kantarjian HM, Andreeff M, et al. Functional proteomic profiling of AML predicts response and survival. *Blood*. 2009; 113 (1): 154-64.

**GENOMICA E ATTUALITÀ DI MEDICINA
MOLECOLARE IN TUMORI SOLIDI**

Basi molecolari del rischio genetico di tumori

Paolo Radice

Dipartimento di Medicina Predittiva e per la Prevenzione, Istituto Nazionale Tumori, Campus IFOM-IEO, Milano

L'esistenza di aggregazioni famigliari caratterizzate da casi multipli di tumore, spesso ad insorgenza precoce, è una chiara dimostrazione che fattori ereditari possono contribuire in maniera determinante allo sviluppo delle patologie oncologiche. A partire dalla fine degli anni '80 del secolo scorso, gli studi di genetica molecolare hanno permesso di individuare numerosi geni e varianti alleliche associati alla predisposizione ereditaria al cancro ed è ormai evidente che tali elementi giocano un ruolo non meno importante delle mutazioni somatiche nella cancerogenesi umana (Foulkes, 2008).

I fattori ereditari che predispongono allo sviluppo tumorale costituiscono una classe eterogenea di elementi genetici che possono essere distinti in base alla loro frequenza nella popolazione ed ai rischi associati (Tab. 1). Il carcinoma mammario famigliare costituisce un buon modello per illustrare tale complessità (Mavaddat et al., 2010).

Agli inizi degli anni 90 una serie di studi di *linkage* condotti su famiglie numerose in cui era possibile osservare una ereditarietà della patologia di tipo mendeliano, sebbene a penetranza incompleta, hanno portato all'isolamento dei due principali elementi genetici responsabili della predisposizione al carcinoma della mammella ad oggi noti: *BRCA1* e *BRCA2* (Turnbull and Rahman, 2008). Tali geni sono responsabili di circa il 20% di tutte le aggregazioni famigliari di cancro della mammella e sono anche associati ad un significativo aumento di rischio di carcinoma ovarico.

Lo spettro mutazionale dei geni *BRCA* è molto eterogeneo ed a tutt'oggi sono state identificate diverse centinaia di mutazioni patogeniche, la maggior parte delle quali sono di tipo *nonsense* o *frameshift*. Tali mutazioni portano alla sintesi di prodotti proteici difettivi od instabili, causando la perdita di funzione del gene

Tab. 1 - Fattori genetici associati a predisposizione ereditaria al cancro.

Mutazioni ad alta penetranza	0,001>5	Analisi di <i>linkage</i>
Mutazioni a penetranza moderata	>0,05	1.05-1.3S
Varianti alleliche a bassa penetranza	0,001-0,0042-4	Studi di associazione <i>genome wide</i> Sequenziamento di geni candidati

(Breast cancer Information Core; <http://research.nhgri.nih.gov/bic/index.shtml>). A livello somatico, la perdita o inattivazione dell'allele costitutivamente normale porta allo sviluppo tumorale. Pertanto, i geni BRCA rientrano nella categoria dei geni "onco-soppressori". È interessante rilevare che mutazioni bi-alleliche di *BRCA2* sono state osservate in soggetti affetti da anemia di Fanconi, una sindrome ereditaria caratterizzata di malformazioni congenite, insufficienza del midollo osseo, instabilità cromosomica e predisposizione al cancro (Howlett et al., 2002).

La frequenza delle mutazioni dei geni BRCA nella popolazione generale è stimata nell'ordine dello 0,06-0,1%. Esistono però delle popolazioni in cui tale frequenza può essere fino a 10 volte superiore, come nel caso degli Ebrei Ashkenaziti, in cui è stata osservata la presenza di tre specifiche mutazioni fondatrici (due in *BRCA1* ed una in *BRCA2*) (King et al., 2003).

Il rischio di cancro della mammella a 70 anni per una donna portatrice di una mutazione *BRCA1* è di circa il 60-70% e del 40-50% per una donna con mutazione *BRCA2* (Antoniou et al., 2003). Questi rischi risultano da 5 a 10 superiori a quelli che si osservano nella popolazione generale e pertanto tali **mutazioni** vengono definite ad **alta penetranza**. Vi sono forti indicazioni che i rischi associati alle mutazioni ad alta penetranza possono essere modulati da una serie di fattori non genetici, quali la parità e l'ovariectomia (Narod, 2006). Inoltre, sono stati identificati alcuni fattori genetici in grado di modificare la penetranza della mutazione nei geni BRCA (Milne & Antoniou, 2011).

Oltre a *BRCA1* e *BRCA2* sono noti altri geni che conferiscono elevati rischi di cancro mammario. Tali geni sono mutati assai raramente nella popolazione generale e sono generalmente responsabili di vere e proprie forme sindromiche. Tra questi *TP53*, responsabile della sindrome di Li-Fraumeni (Birch et al., 2001), *PTEN*, responsabile della sindrome di Cowden (Nelen et al., 1996), e *STK11/LKB1*, responsabile delle sindrome di Peutz-Jegher (Jenne et al., 1998).

L'individuazione dei geni sopra descritti ha permesso lo sviluppo di test genetici che sono entrati ormai da diversi anni nella pratica clinica. Tali test vengono utilizzati per identificare, tra i componenti delle famiglie con evidenza di suscettibilità ereditaria allo sviluppo della malattia, coloro i quali, essendo portatori della mutazione responsabile di tale suscettibilità, è opportuno vengano indirizzati a specifici programmi di prevenzione e/o sorveglianza per la riduzione del loro rischio. Va, comunque, tenuto presente che l'utilizzo di tali test presenta alcune criticità, quali la già citata complessità degli spettri mutazionali dei geni in analisi e la incompleta penetranza delle mutazioni. Un altro elemento critico è rappresentato dalla interpretazione del risultato del test. Infatti, oltre alle mutazioni che possono essere immediatamente inferibili come associate a predisposizione ereditaria, quali quelle "troncanti" o di *splicing*, ve ne possono essere altre, ad esempio le *missense* o quelle introniche, la cui classificazione clinica, in rapporto al rischio di cancro, richiede l'esecuzione di ulteriori analisi (Radice et al., 2011).

Oltre alle mutazioni ad alta penetranza, sono state identificate alcune altre varianti alleliche che conferiscono un aumentato rischio di cancro della mammella, sebbene in misura assai più modesta. Queste varianti sono relativamente frequenti

nella popolazione generale (tra il 10 ed il 40% circa) e sono associate a rischi relativi (RR) di sviluppo della malattia variabili tra 1.1 e 1.3. La loro identificazione è stata resa possibile dagli studi di associazione caso-controllo, in particolare quelli condotti sull'intero genoma (*genome wide association studies*, GWAS) (Varghese & Easton 2010). Questi studi analizzano alcune centinaia di migliaia di varianti genetiche, generalmente *single nucleotide polymorphisms* (SNPs), rappresentative delle regioni di *linkage disequilibrium* presenti nel genoma umano e identificano quelle presenti con una frequenza significativamente superiore nei soggetti affetti rispetto a gruppi di individui sani della popolazione generale. Il successo di tali studi dipende in gran parte dalla possibilità di disporre di elevate numerosità di casi e controlli, che vengono solitamente raccolti all'interno di consorzi collaborativi internazionali.

Attualmente, sono stati identificati **25 alleli a bassa penetranza** associati alla predisposizione al cancro mammario e si stima che, complessivamente, tali alleli siano responsabili di circa il 9% di tutte le aggregazioni famigliari di questo tumore (Ghoussaini et al., 2012). È interessante osservare che diverse di queste varianti sono le stesse, precedentemente menzionate, che sono state osservate in grado di modulare la penetranza delle mutazioni dei geni BRCA.

Al momento attuale non vi sono indicazioni a favore della utilità clinica degli alleli a bassa penetranza, in considerazione dei modesti aumenti di rischio conferiti (Pharoah et al., 2008). Inoltre, va considerato che tali aumenti di rischio potrebbero non essere necessariamente dovuti alle varianti identificate negli studi di associazione, ma ad altre più rare nella popolazione generale. In *linkage* con le varianti comuni e possibilmente associate a rischi più elevati. Studi di *fine mapping* e *re-sequencing* delle regioni genomiche di interesse sono attualmente in corso.

Un terzo gruppo di fattori genetici associati alla suscettività ereditaria al cancro mammario è stato identificato negli ultimi anni in seguito al sequenziamento di geni candidati in soggetti a rischio. Questi geni sono stati selezionati in base alla loro "plausibilità" biologica. Ad esempio, è stato osservato che mutazioni eterozigoti in diversi geni del cosiddetto *pathway* della anemia di Fanconi (oltre a BRCA2) conferiscono un aumentato rischio di cancro della mammella. Tra questi *BRIP1/FANCF* (Seal et al., 2006), *PALB2/FANCF* (Rahaman et al., 2007) e *RAD51C/FANCF* (Meindl et al., 2010). I rischi di cancro associati a queste mutazioni sono intermedi tra quelli osservati per le mutazioni dei geni BRCA e quelli conferiti dalle varianti a bassa penetranza (RR: 2-4). Conseguentemente, tali **mutazioni** vengono definite **a penetranza moderata**. Analoghi livelli di rischio di cancro mammario sono stati osservati per mutazioni in altri geni, tra cui *ATM* (Renwick et al., 2006) e *CHEK2* (Meijers-Heijboer et al., 2002), i quali, analogamente ai geni dell'anemia di Fanconi, sono coinvolti nei meccanismi di riparazione delle rotture del DNA. Mutazioni nei geni sopradetti sono riscontrate assai raramente nei casi di carcinoma mammario famigliare e complessivamente sono responsabili di non più del 5% di tutte le aggregazioni famigliari.

Va sottolineato che nelle famiglie in cui sono presenti le mutazioni in uno dei geni sopra detti, non infrequentemente si osserva una incompleta segregazione di tali

mutazioni. Ciò è dovuto, non solo alla loro ridotta penetranza, ma anche alla probabile presenza nelle stesse famiglie di altri fattori ereditari che contribuiscono al rischio di cancro. Pertanto, una classificazione per classi di rischio dei componenti di queste famiglie basata unicamente sulla presenza/assenza di una specifica mutazione in un singolo gene, potrebbe portare ad escludere erroneamente dai programmi di sorveglianza/prevenzione soggetti a rischio aumentato. Per questa ragione, l'utilizzo clinico delle analisi mutazionali in questo particolare gruppo di geni è al momento ancora oggetto di discussione.

Negli ultimi anni notevoli progressi sono stati fatti per quanto riguarda le conoscenze delle basi molecolari della predisposizione ereditaria al cancro. Una parte di queste conoscenze ha già trovato un'applicazione clinica ed altre potranno essere sviluppate in futuro. Tuttavia, i fattori genetici responsabili di una frazione consistente di aggregazioni famigliari di tumori rimangono tuttora sconosciuti. È prevedibile che le nuove tecnologie che sono state sviluppate nel corso degli ultimi anni, ed in particolare le cosiddette tecniche di *next generation sequencing* (Ng et al., 2010), consentiranno la identificazione di tali elementi in un futuro non troppo lontano. Determinante a tale fine sarà la capacità di discriminare, all'interno dell'enorme numero di varianti genetiche che tali approcci permetteranno di individuare, quelle effettivamente associate allo sviluppo tumorale, ed in particolare quelle a maggiore rilevanza clinica. Ciò potrebbe avere un notevole impatto sulla gestione delle patologie oncologiche, con importanti conseguenze dal punto di vista socio-sanitario.

Bibliografia

1. Antoniou A, Pharoah PD, Narod S, et al. Average risks of breast and ovarian cancer associated with BRCA1 or BRCA2 mutations detected in case series unselected for family history: a combined analysis of 22 studies. *Am J Hum Genet* 2003; 72: 1117e1130.
2. Birch JM, Alston RD, McNally RJ, et al. Relative frequency and morphology of cancers in carriers of germline TP53 mutations. *Oncogene* 2001; 20: 4621e4628.
3. Foulkes WD. Inherited susceptibility to common cancers. *N Engl J Med* 2008; 359: 2143-53.
4. Ghoussaini M, Fletcher O, Michailidou et al. Genome-wide association analysis identifies three new breast cancer susceptibility loci. *Nat Genet* 2012; 44: 312-8.
5. Howlett NG, Taniguchi T, Olson S, et al. Biallelic inactivation of BRCA2 in Fanconi anemia. *Science* 2002; 297: 606-9.
6. Jenne DE, Reimann H, Nezu J, et al. Peutz-Jeghers syndrome is caused by mutations in a novel serine threonine kinase. *Nat Genet* 1998; 18: 38e43.
7. King MC, Marks JH, Mandell, JB. Breast and ovarian cancer risks due to inherited mutations in BRCA1 and BRCA2. *Science* 2003; 302: 643e646.
8. Mavaddat N, Antoniou AC, Easton DF, Garcia-Closas M. Genetic susceptibility to breast cancer. *Mol Oncol* 2010; 4: 174-91.

9. Meindl A, Hellebrand H, Wiek C, et al. Germline mutations in breast and ovarian cancer pedigrees establish RAD51C as a human cancer susceptibility gene. *Nat Genet* 2010; 42: 410-4.
10. Meijers-Heijboer H, van den Ouweland A, Klijn J, et al. Low-penetrance susceptibility to breast cancer due to CHEK2(*)1100delC in noncarriers of BRCA1 or BRCA2 mutations. *Nat Genet* 2002; 31: 55e59.
11. Milne RL, Antoniou AC. Genetic modifiers of cancer risk for BRCA1 and BRCA2 mutation carriers. *Ann Oncol* 2011; (Suppl. 1): i11-7.
12. Narod SA. Modifiers of risk of hereditary breast cancer. *Oncogene* 2006; 25: 5832-6.
13. Nelen MR, Padberg GW, Peeters EA, et al. Localization of the gene for Cowden disease to chromosome 10q22-23. *Nat. Genet.* 1996; 13: 114e116.
14. Ng SB, Nickerson DA, Bamshad MJ, et al. Massively parallel sequencing and rare disease. *Hum Mol Genet* 2010; 19: R119-24.
15. Pharoah PD, Antoniou AC, Easton DF, Ponder BA. Polygenes, risk prediction, and targeted prevention of breast cancer. *N Engl J Med* 2008; 358: 2796-803.
16. Radice P, De Summa S, Caleca L, Tommasi S. Unclassified variants in BRCA genes: guidelines for interpretation. *Ann Oncol* 2011; (Suppl. 1): i18-23.
17. Rahman N, Seal S, Thompson D, et al. PALB2, which encodes a BRCA2-interacting protein, is a breast cancer susceptibility gene. *Nat Genet* 2007; 39: 165-7.
18. Renwick A, Thompson D, Seal S, et al. ATM mutations that cause ataxiatelangiectasia are breast cancer susceptibility alleles. *Nat. Genet* 2006; 38: 873e875.
19. Seal S, Thompson D, Renwick A, et al. Truncating mutations in the Fanconi anemia J gene BRIP1 are low-penetrance breast cancer susceptibility alleles. *Nat Genet* 2006; 38: 1239-41.
20. Turnbull C, Rahman N. Genetic predisposition to breast cancer: past present, and future. *Annu Rev Genomics Hum Genet* 2008; 9: 321-45.
21. Varghese JS, Easton DF. Genome-wide association studies in common cancers-what have we learnt? *Curr Opin Genet Dev* 2010; 20: 201-9.

Caratterizzazione molecolare del tumore della mammella

Umberto Magrini

Già Professore Ordinario di Anatomia Patologica, Università di Pavia

Il carcinoma mammario è una malattia eterogenea, che comprende numerose entità con quadri istopatologici, storia naturale e risposte alle terapie estremamente diverse.

Attraverso i “microarray” di espressione genica si è venuta configurando, negli ultimi dieci anni, una nuova classificazione su base molecolare del carcinoma della mammella, che è al centro dell’interesse dei biologi molecolari, dei patologi e degli oncologi.

Esaminando simultaneamente l’espressione di migliaia di geni viene rilevato il profilo molecolare dei tumori che si definisce “intrinseco” in quanto ottenuto da una analisi “unsupervised”, che prescinde cioè dal comportamento clinico della neoplasia e che si basa su differenze di ordine biologico.

Vi sono alcuni gruppi di geni che caratterizzano gruppi definiti di carcinomi. In sintesi: geni correlati con i recettori ormonali, geni correlati con HER2 (Human epidermal growth factor 2), geni identificati come “basal-like”, geni legati ad attività proliferativa.

Sono stati identificati, utilizzando la variabile espressione di questi “clusters” di geni, almeno cinque sottotipi intrinseci: due sottotipi di tipo luminale (A e B) che rappresentano la maggior parte dei tumori dotati di recettori estrogeni, un sottotipo HER2 positivo, un sottotipo basale e infine un sottotipo recentemente proposto come “claudin-low”.

Un ulteriore sottotipo, inizialmente indicato come “normal breast like”, è probabilmente la risultante di un artefatto di campionamento costituito da una quota preponderante di cellule stromali e di epitelii della mammella normale.

I cinque sottotipi sono stati confermati con diverse metodologie e il profilo biomolecolare è risultato uniforme sia nelle lesioni iniziali (carcinoma “in situ”) che nelle metastasi, avvalorando il concetto che i sottotipi costituiscano proprietà intrinseche della cellula neoplastica.

Inoltre, ed è il dato pratico più importante, la correlazione con i dati clinici pone in evidenza differenze significative per quanto attiene l’incidenza, la sopravvivenza e la risposta alle terapie.

La terminologia che definisce i sottotipi è entrata nel linguaggio clinico, non ultimo per il fatto che appaiono connessi a geni e vie metaboliche strettamente legate alla storia naturale dei carcinomi mammari, quali i recettori ormonali e i geni che li regolano, il recettore HER2 e i geni che controllano la proliferazione.

Dato il tumultuoso sviluppo della biologia molecolare, l'avvento di nuove tecnologie, quali ad esempio la possibilità di catalogare tutti i tipi di trascritti (RNA-Seq), è verisimile che questa classificazione vada entro breve tempo rivista con l'introduzione di categorie più omogenee e, soprattutto, con l'identificazione di nuovi bersagli molecolari aggredibili con terapie specifiche.

Lo studio del profilo di espressione genica dei carcinomi mammari mediante "microarray" riconosce, ovviamente, numerose limitazioni: complessità tecnica, esigenze di validazione e standardizzazione, costi.

L'immunoistochimica (ICH) di facile esecuzione, affidabile, ampiamente standardizzata, economica, può fornire un surrogato ai "microarray". Infatti, la dimostrazione di alcuni biomarcatori "chiave" (ER, PgR, HER2, Ki 67), consente di identificare gruppi di tumori corrispondenti a quanto ottenuto con i "microarray". Pertanto, nella pratica clinica, si utilizzano, ai fini diagnostici, i dati ottenuti con i marcatori immunoistochimici. Un Panel di esperti (The St. Gallen International Expert Consensus for Early Breast Cancer 2011) ha convalidato la possibilità di definire i sottotipi su base immunoistochimica con le possibili implicazioni terapeutiche che ne derivano.

Vengono qui di seguito esposte le caratteristiche principali, biologiche e cliniche, dei cinque sottotipi intrinseci a tutt'oggi riconosciuti.

Sottotipi intrinseci

Luminale A

Il termine "luminale" indica il tipo di cellula epiteliale prospiciente e delimitante il lume della struttura funzionale della mammella, detta "unità dotto terminale-lobulo" (UDTL). Lo strato cellulare sottostante alle cellule luminali, con caratteristiche mio-epiteliali, è contraddistinto dal termine "basale".

La UDTL rappresenta l'unità funzionale della mammella, sensibile, sia nella componente epiteliale, sia nella componente stromale circostante, alle stimolazioni ormonali della vita feconda della donna e sede di origine della quasi totalità dei carcinomi mammari.

Il sottotipo luminale A è il più comune tra i carcinomi mammari, rappresentando il 50-60% del totale. La peculiarità di questo tipo di neoplasia è l'espressione di geni attivati dal fattore di trascrizione ER, così come nelle cellule luminali della mammella normale. Altre caratteristiche sono la bassa espressione dei geni di proliferazione e il basso grado istologico.

Dal punto di vista immunoistochimico sono dimostrabili, negli elementi tumorali, oltre ai recettori per estrogeni e progesterone (ER, PgR), citocheratine di tipo luminale (CK 8/18) e la proteina anti-apoptotica bcl2. È invece negativo il recettore HER2 e appare bassa l'attività proliferativa.

Ha, tra tutti i sottotipi, la prognosi migliore, con frequenza di recidive inferiore al

30% e con le sopravvivenze più lunghe. La sede più frequente di recidive è rappresentata dall'apparato scheletrico.

Le caratteristiche biologiche della neoplasia giustificano l'uso elettivo di terapie ormonali (inibitori delle aromatasi, modulatori selettivi dei recettori estrogeni-SERM, regolatori selettivi puri di ER).

Luminale B

Questo sottotipo rappresenta il 10-20% dei carcinomi mammari. La differenza biologica con il Luminale A consiste in una più elevata espressione di geni della proliferazione e dalla frequente espressione di HER2 e EGFR (Epidermal Growth Factor Receptor 1/HER1). Il "cut-off" per separare il Luminale B dal Luminale A è costituito da un indice di proliferazione (Ki 67) superiore al 14%.

Il grado istologico è più elevato ed il comportamento clinico più aggressivo rispetto al Luminale A, con diffusioni metastatiche, oltre che allo scheletro, anche in sede epatica.

Il trattamento con terapia ormonale è meno efficace che per il Luminale A, mentre migliore è la risposta alla chemioterapia neoadiuvante. L'opzione terapeutica ottimale rimane comunque tuttora "sub-judice".

HER2 positivo

Il sottotipo contraddistinto da alta espressione del gene HER2 e di altri geni associati con la via HER2 e/o con l'amplicone di HER2, si riscontra nel 15-20% dei carcinomi mammari.

Oltre ai geni correlati con HER2, questo tipo di tumori esprime in alto grado i geni correlati con la proliferazione, frequentemente denota mutazioni di Tp53 ed ha un alto grado istologico.

Da sottolineare che non vi è perfetta corrispondenza tra sottotipo intrinseco definito dai "microarray" e profilo immunoistochimico, poiché solo il 70% dei casi identificati in base ai "microarray" denota anche alta espressione istochimica della proteina recettoriale.

Clinicamente si tratta di tumori a prognosi grave, radicalmente migliorata negli ultimi dieci anni con l'avvento dei trattamenti anti-HER2 e con la dimostrata alta chemiosensibilità.

L'utilizzo di un predittore di prognosi (HER2-derived prognostic predictor-HDPP) composto da 158 geni, consente di stratificare i tumori HER2 positivi tra forme a buona e cattiva prognosi.

È interessante notare che il gruppo di geni predittivi di prognosi non riguarda l'attività proliferativa né la via di segnale di HER2, quanto piuttosto i geni connessi con la risposta immunitaria e con la capacità invasiva e metastatica.

"Basal-like"

Questo sottotipo intrinseco costituisce il 10-20% di tutti i carcinomi mammari ed esprime geni normalmente presenti nelle cellule basali/mio-epiteliali (CK5/6,

CK17, P-caderina, caveolina 1 e 2, nestina, CD 44, EGFR) dell'unità dotto terminale-lobulo.

Peculiarità di questo sottotipo è l'assenza di tre recettori "chiave" (ER, PgR, HER2), per cui rientra tra i cosiddetti tumori "tripli negativi". Come si vedrà in seguito non vi è però perfetta sovrapposibilità dei termini "basale" e "triplo negativo".

L'identificazione immunostochimica si basa sulla tripla negatività e sulla possibile positività per marcatori basali (CK5/6) e per il recettore EGFR. Nell'85% dei casi sono presenti mutazioni di TP53, con espressione istochimica di p53.

In alcuni casi sono espresse anche citocheratine proprie delle cellule luminali (CK8 e/o 18), ponendo in questione la iniziale attribuzione tassonomica che assegna l'istogenesi a cellule basali mio-epiteliali. In base a tali reperti, è stato suggerito che almeno un sottogruppo di carcinomi "basal-like" possa originare da progenitori luminali. A questo proposito va ricordato che istogenesi e differenziazione sono processi diversi, benché spesso erroneamente usati come sinonimi. I carcinomi "basal-like" si presentano in giovane età, più frequentemente in donne di origine Africana, con dimensioni elevate, alto grado istologico (Carcinomi Duttali Infiltranti non altrimenti specificabili-CDI n.a.s., G3), alto indice proliferativo, necrosi/fibrosi centrale, margini espansivi, infiltrati linfocitari, quadri di tipo midollare tipico/atipico.

Da notare che anche in un gruppo di carcinomi lobulari infiltranti, usualmente compresi tra le forme "luminali", sono state dimostrate citocheratine ad alto peso molecolare proprie delle cellule basali.

L'interesse che i tumori "basal-like" suscitano è connesso con l'aggressività di queste neoplasie, superiore a quella degli altri sottotipi intrinseci, e con l'impossibilità di porre in atto terapie mirate data l'assenza di bersagli specifici. L'aggressività è dimostrata in maniera significativa dal fatto che il picco di rischio di recidive si colloca tra il primo e il terzo anno dalla diagnosi e che la maggior parte dei decessi si riscontra entro cinque anni dalla terapia.

La diffusione metastatica è sia linfonodale che sistemica, con localizzazioni preferenziali al polmone e al sistema nervoso centrale. La chemioterapia neoadiuvante con antracicline o antracicline+taxani induce una completa risposta in percentuali variabili (17-58%). I casi che non presentano completa risposta hanno prognosi infausta.

Recentemente è stato proposto un "set" di geni, indicato come "Basal 14 signature", in grado di stratificare i pazienti in gruppi a buona e cattiva prognosi. Da sottolineare che la presenza, tra i geni espressi, del gene per la proteina citosolica linfocitaria1 (LCPI1), rappresenta una possibile spiegazione dell'infiltrato linfocitario intratumorale che, quando presente, correla con una buona prognosi. La possibilità di distinguere questi due sottogruppi alla diagnosi iniziale, consente di applicare regimi aggressivi per i pazienti con prognosi grave e di evitarli nei pazienti con prognosi migliore.

Un punto che merita approfondimento è costituito dai **rapporti tra carcinomi basali e mutazioni BRCA1 di tipo germinale**.

La maggior parte dei tumori insorti in portatori della mutazione BRCA1 di tipo

germinale, particolarmente quelli di età inferiore ai 50 anni, presenta istotipi simili a quelli riscontrabili nel gruppo di carcinomi “basal-like” e fenotipo coerente in base ai “microarray” di espressione genica e ai marcatori immunoistochimici. Inoltre, sia i carcinomi basali che i carcinomi insorti in portatori della mutazione germinale BRCA1 esprimono un pattern particolare per quanto attiene le proteine correlate con il ciclo cellulare: bassi livelli di p27, alti livelli di Skp2, ciclina E, caspasi-3.

Il fatto che i carcinomi basali, benché non abbiano mutazioni somatiche di BRCA1, mostrino profili molecolari simili ai carcinomi insorti in portatori della mutazione, è dovuto a regolazione alterata della via di BRCA1 (metilazione del gene promoter, espressione di regolatori negativi di BRCA1). Il silenziamento di BRCA1 induce ipoespressione di ER e iperespressione di geni markers del carcinoma basale, come CK5/6, CK17, P-caderina.

In sintesi, la disfunzione di BRCA1 rappresenta uno dei meccanismi patogenetici dei carcinomi basali.

L'alterata funzione di BRCA1, correlata con la riparazione del DNA, porta a instabilità genetica e a un accumulo di errori. In questo contesto si inserisce la strategia terapeutica con inibitori di un enzima chiave nella riparazione del DNA, la poli-ADP riboso-polimerasi-1 (PARP-1).

L'inibizione di PARP-1 porta a un accumulo di rotture ed alla morte cellulare. I risultati ottenuti associando un inibitore di PARP-1 alla chemioterapia con gemcitabina/carboplatino appaiono incoraggianti.

Richiede un breve approfondimento il rapporto tra i termini “triplici negativi” e “basal-like” frequentemente usati come sinonimi in quanto la maggior parte dei carcinomi “basal-like” sono triplici negativi, cioè non esprimono ER, PgR, HER2. Nonostante le numerose somiglianze cliniche e di profilo molecolare tra il sottotipo “basal-like” e i carcinomi triplici negativi, i termini non sono sinonimi, anche se nel linguaggio clinico sono usati come tali. Infatti non tutti i carcinomi definiti in base al profilo di espressione genica con “basal-like” mancano di ER, PgR, HER2 e, per converso, non tutti i carcinomi triplici negativi denotano un fenotipo “basal-like”.

In sintesi, la tripla negatività non può essere utilizzata come un surrogato per la diagnosi di carcinomi “basal-like”. Resta il fatto che i carcinomi “basal-like” rappresentano una categoria molto eterogenea legata a prognosi grave e che rimane problematica un'identificazione di “routine” di questo sottotipo di tumori.

“Claudin-low”

È un sottotipo relativamente raro (12-14%), contraddistinto da bassa espressione di geni coinvolti nelle giunzioni cellulari, quali claudina-3, -4, -7 e caderina-E. Si tratta di tumori compresi tra i triplici negativi, a prognosi grave, con frequente differenziazione di tipo metaplastico o midollare.

Dal punto di vista biologico presenta caratteristiche coerenti con cellule staminali e con possibilità di transizione epitelio-mesenchimale (EMT).

La contestualizzazione di questo sottotipo con gli altri sottotipi delinea una diffe-

rennazione gerarchica assimilabile al normale sviluppo dell'epitelio mammario umano, nell'ambito della quale il sottotipo "claudin-low" denota il fenotipo meno differenziato, ricco in cellule staminali cancerose, responsabili di crescita, progressione, metastatizzazione e resistenza alle terapie.

In attesa di nuovi marcatori che consentano di identificare sottotipi responsivi a trattamenti specifici, il compito del patologo rimane ancora quello di fornire una valutazione il più possibile precisa per quanto riguarda il tipo istologico, il grado di differenziazione, l'assetto recettoriale, gli indici di proliferazione, vale a dire dei parametri pratici che costituiscono tuttora la base per una terapia sistemica.

Bibliografia

1. Badve S, Dabb DJ, Schmitt SJ, et al. Basal-like and triple negative breast cancers: a critical review with emphasis on the implications for pathologists and oncologists, *Modern Pathology*. 2011; 24: 157-67.
2. Eroles P, Bosch A, Perez-Fidalgo JA, et al. Molecular biology in breast cancer: intrinsic subtypes and signaling pathways. *Cancer Treat Rev* 2011; doi:10.1016/j.ctrv.2011.11.005.
3. Sandhu R, Parker JS, Jones WD, et al. Microarray based gene expression profiling for molecular classification of breast cancer and identification of new targets for therapy. *Lab Medicine*. 2010; 41: 364-72.
4. Prat A, Parker JS, Karginova O, et al. Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. *Breast Cancer Research* 2010; doi: 10.1186/bcr2635.
5. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews/ Genetics*. 2009; 10: 57-63.
6. Huber KE, Carey LA, Wazer DE. Breast cancer molecular subtypes in patients with locally advanced disease: impact on prognosis, patterns of recurrence and response to therapy. *Semin Radiat Oncol*. 2009; 19: 204-10.
7. Hallett RM, Dvorkin-Gheva A, Bane A et al. A gene signature for predicting outcome in patients with basal-like breast cancer. *Scientific Reports* 2012; doi: 10.1038/srep00227.
8. Park S, Koo JS, KimMS, et al. Characteristics and outcomes according to molecular subtypes of breast cancer as classified by panel of four biomarkers using immunohistochemistry. *The Breast* 2011; doi: 10.1016/j.breast.2011.07.008.

Terapie personalizzate nel carcinoma mammario

PierFranco Conte, Federico Piacentini

Dipartimento di Oncologia, Ematologia e Malattie dell' Apparato Respiratorio,
Università di Modena e Reggio Emilia, Azienda Ospedaliera Policlinico di Modena

I tassi di sopravvivenza a cinque anni dei pazienti con carcinoma mammario metastatico sono oggi approssimativamente del 27%, mentre quelli del carcinoma mammario diagnosticato in stadio iniziale risultano superiori al 90% (1). Tali risultati si sono ottenuti anche grazie all'impiego di trattamenti antitumorali sempre più efficaci nell'ambito della terapia endocrina (inibitori dell'aromatasi, fulvestrant), della chemioterapia citotossica (regimi a base di antracicline, taxani, anti-metaboliti), dei farmaci a bersaglio molecolare e della combinazioni di questi. Quanto alle cosiddette *target therapies*, si tratta di farmaci biologici in grado di inattivare una o più proteine chiave nelle vie metaboliche intracellulari e, quindi, di spegnere l'interruttore da cui dipende la proliferazione e la crescita tumorale. Interagendo con specifiche molecole presenti per lo più nelle cellule tumorali questi farmaci risultano mediamente più efficaci e meglio tollerati rispetto ai tradizionali trattamenti chemioterapici poiché colpiscono in modo selettivo le cellule neoplastiche e lasciano la maggior parte delle cellule sane mediamente inalterate. È comunque importante ricordare che questi farmaci quando usati in monoterapia consentono risultati in genere modesti mentre sono in grado di potenziare enormemente i benefici della chemioterapia somministrata in associazione. Ed è basandosi sulla dimostrazione della presenza/assenza di biomarcatori o di loro specifiche alterazioni genetiche che lo specialista oncologo cerca di selezionare il trattamento farmacologico più mirato per ogni singolo tumore, evitando così di esporre agli effetti indesiderati di un farmaco i pazienti che non ne trarrebbero sicuro beneficio (2).

Attualmente, solo alcuni biomarcatori hanno dimostrato la capacità di predire la risposta ad uno specifico trattamento, ma numerosi altri sono in fase di sviluppo. Tra tutti, la scoperta di HER2 che ha drasticamente cambiato la storia biologica del carcinoma mammario migliorando la prognosi delle donne affette, rappresenta sicuramente l'esempio più emblematico di questa innovativa area di ricerca oncologica. HER2 è amplificato o overespresso in circa il 20-30% dei tumori della mammella, e correla con maggior aggressività tumorale e prognosi peggiore rispetto alla controparte di tumori cosiddetti HER2-negativi (3). Trastuzumab è un anticorpo monoclonale ricombinante umanizzato con elevata affinità di lega-

me per HER2 raccomandato per il trattamento adiuvante in associazione alla chemioterapia in tutti i casi di malattia HER2 positiva o come prima linea di trattamento in associazione a chemio- o endocrinoterapia nella malattia avanzata (4). Sebbene il meccanismo d'azione di trastuzumab non sia del tutto definito, sembra che esso agisca tramite una complessa interazione tra ADCC (antibody-dependent cellular cytotoxicity), effetti anti-angiogenetici, down-regolazione del recettore HER2, ed inibizione della progressione del ciclo cellulare (Fig. 1).

Uno studio pilota di fase 3 dimostrò che l'aggiunta di trastuzumab alla chemioterapia (doxorubicina-ciclofosfamida o paclitaxel) in 469 donne con tumore mammario metastatico HER2 positivo prolungava significativamente il tempo medio alla progressione di malattia (7,4 vs 4.6 mesi; $p > 0,001$) e la durata mediana di risposta (9,1 vs 6.1; $p < 0,001$), il tasso di risposte obiettive (50% vs 32%; $p < 0,001$) e la sopravvivenza mediana (25,1 vs 20,3; $p = 0,01$) (5). Studi clinici successivi hanno convalidato l'efficacia di trastuzumab in associazione ad altri agen-

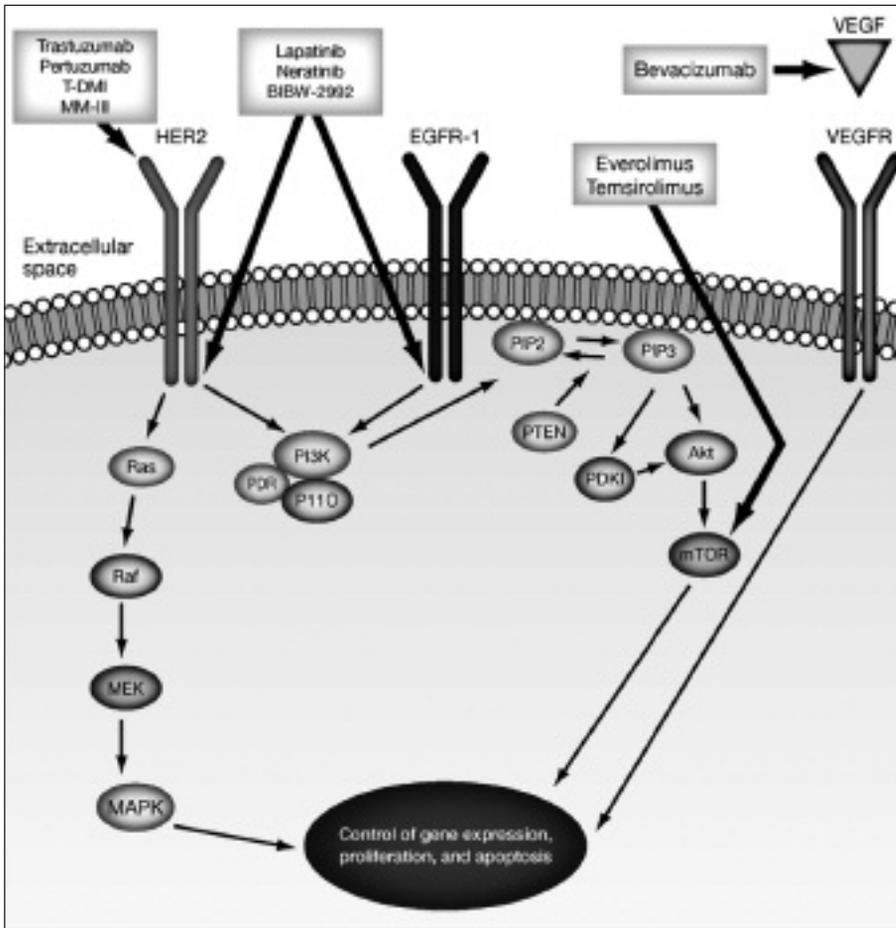


Fig. 1

ti chemioterapici con tassi di ORR variabili dal 50 al 73%, accelerando l'introduzione di questi trattamenti anche nelle fasi più precoci di malattia. Nell'esperienza nel MDACC ad esempio, le pazienti con carcinoma mammario operabile HER2 positivo trattate in neoadiuvante con 4 cicli di paclitaxel seguiti da 4 cicli con FEC in associazione a trastuzumab conseguirono tassi di risposta patologica completa significativamente migliori rispetto alle pazienti che ricevevano la sola chemioterapia senza trastuzumab (66,7% vs 26,3%). Un rivoluzionario vantaggio in termini di sopravvivenza globale e libera da malattia è stato evidenziato in cinque studi randomizzati con trastuzumab adiuvante, includenti oltre 10.000 donne con carcinoma mammario HER2 positivo. In questi trial, un anno di trastuzumab ha ridotto del 50% circa il rischio di recidiva (HR 0.52 nell'analisi combinata degli studi NSABP-B31 e NCCTG N983 (6); 0.76 nello studio HERA (7), 0.64 nel braccio AC-TH e 0.75 nel braccio TCH dello studio BCIRG006 (8). Inoltre, in un'analisi ad interim dello studio N9831, la combinazione di chemioterapia e trastuzumab rispetto alla sequenza si è dimostrata più vantaggiosa (HR 0.77; 99.9% CI, 0.53 a 1.1) (9). In definitiva, dati biologici e clinici supportano gli effetti sinergici citotossici di trastuzumab + chemioterapia, mentre la somministrazione sequenziale di trastuzumab dopo chemioterapia sembra indurre per lo più un effetto citostatico che potrebbe richiedere più tempo per ottimizzare il beneficio clinico. Purtroppo, l'unico studio prospettico progettato per testare durate differenti di trastuzumab adiuvante è lo studio HERA, i cui risultati di confronto di uno rispetto a due anni di trattamento sono ancora attesi. Tuttavia, in un piccolo studio adiuvante finlandese, trastuzumab somministrato per 3 mesi in combinazione con docetaxel o vinorelbina, seguito da tre FEC, ha dimostrato di ridurre il rischio di recidiva in proporzione simile a quanto osservato negli studi precedentemente citati con un anno di trastuzumab, suggerendo che periodi di trattamento più brevi possono produrre efficacia comparabile ma con minor tossicità e costi più bassi. Ad oggi un anno di trattamento con trastuzumab rimane però il gold standard, ma la durata ottimale di trastuzumab adiuvante rimane da stabilire con certezza. E nell'ipotesi che una minor durata di trastuzumab adiuvante concomitante alla chemioterapia possa essere altrettanto efficace, ma meno tossico, abbiamo progettato lo studio ShortHer (10) (Fig. 2) di fase III multicentrico e randomizzato, al fine di valutare se 9 somministrazioni settimanali di trastuzumab rispetto alle 18 somministrazioni trisettimanali standard in combinazione con la chemioterapia adiuvante siano almeno altrettanto efficaci in termini di DFS, e meno tossici dal punto di vista cardiaco.

Altri studi europei stanno affrontando le stesse domande (trial PHARE, SOLD, PERSEPHONE), ma in aggiunta il nostro studio analizzerà l'entità del beneficio anche nei piccoli tumori pT1a,b N0 HER2 positivi così scarsamente rappresentati nella maggior parte degli studi clinici. Per le pazienti a rischio di recidiva nonostante trastuzumab e chemioterapia, sono in corso studi finalizzati a migliorare l'outcome clinico: il trial clinico ALTTO con lapatinib adiuvante è finalizzato a comprendere quale agente anti-HER2 sia più efficace e quale sia il miglior programma di somministrazione (in sequenza o in concomitanza a trastuzumab). Il trial TEACH ha infine recentemente dimostrato che lapatinib adiuvante per 1

anno è in grado di prolungare la DFS solo in caso di ER/PgR negativo o meno di un anno dalla diagnosi (11).

Anche nel contesto della malattia metastatica, le opzioni di trattamento sono in rapido aumento (ixabepilone, eribulin e nab-paclitaxel), con significativo miglioramento dei tassi di risposta e sopravvivenza. Ad esempio, nello studio di fase 3 EMBRACE, eribulin mesilato ha migliorato la sopravvivenza globale (mediana di 13.1 mesi, 95% CI 11,8-14,3) rispetto ad altro trattamento a scelta del medico (mediana 10.6 mesi, CI 9,3-12,5), in pazienti che avevano ricevuto da due a cinque regimi chemioterapici precedenti (comprendenti antracicline e taxani) per la malattia avanzata (12). Nab-paclitaxel a dosaggio trisettimanale si è dimostrato più efficace e meglio tollerato rispetto a paclitaxel a dosaggio trisettimanale (tempo mediano alla progressione di 23 vs 16,9 settimane, HR 0.75 con $p=0,006$) (13); la schedula settimanale ha poi prodotto risultati significativi anche in pazienti già pretrattate con altri taxani (14). Al fine di offrire opzioni terapeutiche sempre più personalizzate, la ricerca oncologica punta ad una sempre miglior caratterizzazione del sottotipo molecolare di tumore: il carcinoma mammario è tradizionalmente classificabile in carcinoma ormono-sensibile (generalmente caratterizzato da un decorso più indolente, con un lungo intervallo libero da malattia e una tendenza a recidivare all'osso o ai tessuti molli), HER2 positivo (più aggressivo e con maggiore propensione per le recidive viscerali) o triplo negativo (TNBC - con tendenza alla recidiva viscerale e al sistema nervoso centrale). Ogni sottotipo mole-

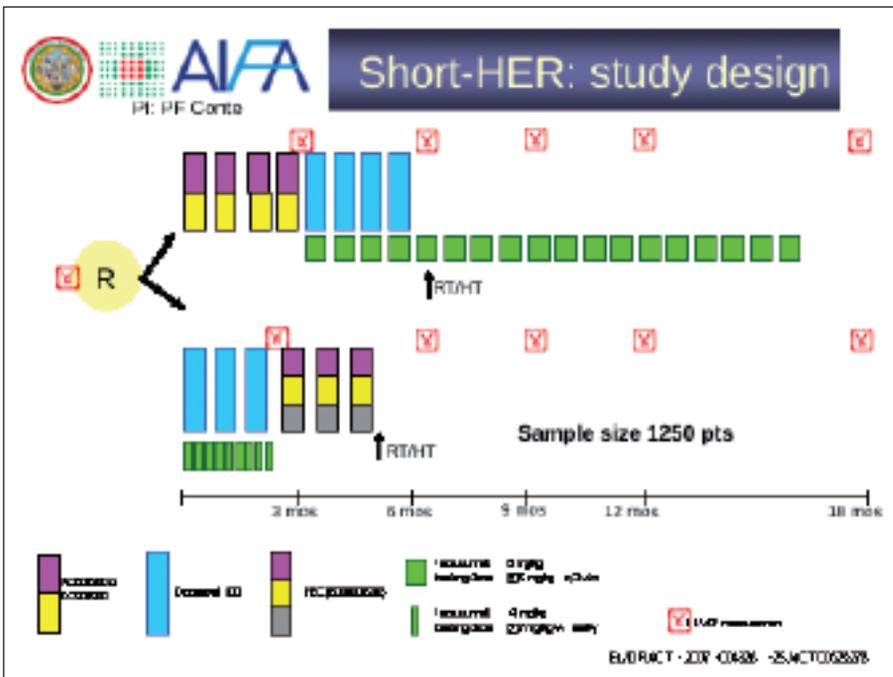


Fig. 2

colare richiede specifici approcci terapeutici. In caso di tumore a recettori ormonali positivi, la manipolazione endocrina è la pietra miliare del trattamento, e deve tenere in considerazione altre determinanti come lo stato menopausale e l'intervallo libero da malattia: inibitori dell'aromatasi non steroidei e steroidei, tamoxifene e fulvestrant rappresentano le possibili opzioni, ma in presenza di malattia a rapida crescita o life-threatening, il ricorso alla chemioterapia diviene inevitabile. Studi recenti hanno dimostrato che tumori a recettori ormonali positivi possono trarre beneficio anche da agenti a target molecolare: lo studio BOLERO-2 ha mostrato un notevole miglioramento in termini sopravvivenza libera da progressione con everolimus + exemestane vs exemestane come prima o seconda linea di trattamento, dopo fallimento di un antiaromatase non-steroidico (mediana di PFS 10.6 vs 4.1 mesi, HR 0.36, IC 95% 0.27-0.47, $p < 0,001$) (15). Anche l'associazione di trastuzumab o lapatinib agli agenti endocrini è un'opzione interessante nei casi di tumori a recettori ormonali ed HER2 positivi, per migliorare la reattività del sistema endocrino e superare l'ormono-resistenza primaria o acquisita. Lo studio randomizzato di fase 2 TAnDEM ha incluso 207 pazienti con carcinoma mammario metastatico triplo positivo e ha riportato un raddoppio della sopravvivenza libera da progressione con l'aggiunta di trastuzumab ad anastrozolo (HR 0,63, 95% CI, 0,47-0,84; PFS mediana, 4,8 vs 2,4 mesi, $p=0,016$) (16). Infine, i risultati di uno studio di fase 3 su 1.286 pazienti con carcinoma mammario metastatico a recettori ormonali positivi randomizzate a ricevere letrozolo da solo o in combinazione con lapatinib, concludono che l'aggiunta di lapatinib riduce significativamente il rischio di progressione rispetto al trattamento con letrozolo da solo nel sottogruppo di 219 pazienti con concomitante positività per HER2 (PFS 8,2 vs 3,0 mesi, HR 0,71, 95% CI, 0,53-0,96, $p=0,019$).

Molti altri farmaci anti-HER2 sono in fase di sviluppo, come T-DM1 (trastuzumab coniugato con un farmaco antimicrotubulo che viene internalizzato nella cellula neoplastica dopo il legame di HER2), neratinib e pertuzumab. In uno studio di fase 2, a singolo braccio con T-DM1 ($n=112$ pazienti con malattia metastatica HER2 positiva progrediti dopo trastuzumab), ad un follow-up di ≥ 12 mesi, la PFS mediana è risultata di 4.6 mesi (95% CI, 3,9-8,6) con tasso di risposta globale del 26% (17). T-DM1 è attualmente in fase di ulteriore sviluppo nel contesto di altri studi clinici: lo studio randomizzato di fase III EMILIA prevede il confronto tra T-DM1 agente singolo con la combinazione di lapatinib e capecitabina in pazienti la cui malattia HER2-positiva è progredita con trastuzumab; nel trial di fase III MARIANNE, T-DM1 in monoterapia è confrontato con trastuzumab + taxano. Neratinib/HKI-272 è un inibitore irreversibile dell'attività tirosin-chinasica di EGFR/HER1, HER2 e HER4: in uno di fase II, i pazienti in con malattia HER2 positiva in progressione (con o senza precedente trattamento con trastuzumab) che hanno ricevuto neratinib, hanno perseguito una sopravvivenza libera da progressione a 16 settimane del 59% in caso di precedente trastuzumab ($n=63$) e del 78% in sua assenza ($n=64$), con PFS mediana di 22,3 e 39,6 settimane rispettivamente (18). In corso la valutazione dell'efficacia di paclitaxel + neratinib vs paclitaxel + trastuzumab (trial Nefertiti), o del solo neratinib verso la combinazione di capecitabina e lapatinib. Infine, pertuzumab è un anticorpo monoclonale umaniz-

zato che si lega al dominio II del recettore HER2, inibendone l'eterodimerizzazione con HER1, HER2, HER3 o HER4. Recenti dati di uno studio randomizzato di fase 3 mostrano come la combinazione di trastuzumab, pertuzumab e docetaxel in prima linea migliora significativamente la sopravvivenza libera da progressione, rispetto alla combinazione di trastuzumab e docetaxel (PFS 18,5 vs 12,4 mesi, HR 0,62, 95% CI 0,51-0,75, $p < 0,001$) (19).

Per contro, la chemioterapia rimane l'unica opzione disponibile per il sottotipo triplo negativo di carcinoma mammario; al momento non esistono infatti agenti mirati specificamente approvati per il trattamento del TNBC. Bevacizumab sembra prolungare la sopravvivenza libera da progressione, quando aggiunto alla chemioterapia, ma non migliora la sopravvivenza globale; gli inibitori di PARP non hanno sostenuto i dati preliminari di efficacia; una grande varietà di altre terapie mirate, compresi gli inibitori di PI3K o agenti che inibiscono la riparazione del DNA sono in corso di indagine attiva.

Per concludere, un interessante campo di ricerca è la ri-caratterizzazione molecolare della malattia metastatica. Routinariamente, la biopsia della sede metastatica di malattia non viene eseguita e la decisione terapeutica si basa principalmente sullo stato recettoriale del tumore primitivo. Tuttavia, i tassi di discordanza (motivati da diverse procedure tecniche, eterogeneità del tumore, deriva genica del tumore durante la progressione, pressione selettiva delle terapie) nell'espressione dei recettori ormonali e/o di HER2 tra tumori primari e recidivi sono variabili tra 10% e 35% (20), potendo rivestire anche un importante ruolo prognostico (21).

È pertanto indispensabile poter integrare tutte le informazioni relative alla malattia della paziente per riuscire ad assicurare la migliore e più personalizzata possibile strategia di trattamento.

Bibliografia

1. Jemal A, et al. Cancer statistics, 2009. *CA Cancer J Clin* 2009; 59 (4): 225-49.
2. Awada A, et al. New therapies in HER2-positive breast cancer: A major step towards a cure of the disease? *Cancer Treat Rev* (2012), doi:10.1016/j.ctrv.2012.01.001.
3. Slamon DJ, et al. Human breast cancer: correlation of relapse and survival with amplification of the HER2/neu oncogene. *Science*. 1987; 235 (4785): 177-82.
4. National comprehensive cancer network. NCCN clinical practice guidelines in oncology™. Breast Cancer V.2.2010. <http://www.nccn.org/professionals/physician_gls/PDF/breast.pdf>
5. Slamon DJ, et al. Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2. *N Engl J Med*. 2001; 344 (11): 783-92.
6. Perez EA, et al. Four-year follow-up of trastuzumab plus adjuvant chemotherapy for operable human epidermal growth factor receptor 2-positive breast cancer: joint analysis of data from NCCTG N9831 and NSABP B-31. *J Clin Oncol*. 2011; 29 (25): 3366-73

7. Gianni L, et al. Herceptin Adjuvant (HERA) Trial Study Team. Treatment with trastuzumab for 1 year after adjuvant chemotherapy in patients with HER2-positive early breast cancer: a 4-year follow-up of a randomised controlled trial. *Lancet Oncol.* 2011; 12 (3): 236-44.
8. Slamon D, Eiermann W, Robert N, et al. Adjuvant trastuzumab in HER2-positive breast cancer. *N Engl J Med.* 2011; 365 (14): 1273-83.
9. Perez EA, et al. Sequential versus concurrent trastuzumab in adjuvant chemotherapy for breast cancer. *J Clin Oncol.* 2011; 29 (34): 4491-7.
10. Guarneri V, et al. Multicentric, randomized phase III trial of two different adjuvant chemotherapy regimens plus three versus twelve months of trastuzumab in patients with HER2-positive breast cancer (Short-HER Trial; NCT00629278). *Clin Breast Cancer.* 2008; 8 (5): 453-6.
11. Goss P, et al. Results of a randomized, double-blind, multicenter, placebo-controlled study of adjuvant lapatinib in women with early-stage ErbB2-overexpressing breast cancer. *Cancer Res.* 2011; 71 (24 Suppl.) December 15.
12. Cortes J, et al. Eribulin monotherapy versus treatment of physician's choice in patients with metastatic breast cancer (EMBRACE): a phase 3 open-label randomized study. *Lancet.* 2011; 377: 914-23.
13. Gradishar WJ, et al. Superior efficacy of albumin-bound paclitaxel, ABI-007, compared with polyethylated castor oil-based paclitaxel in women with metastatic breast cancer: results of a Phase III trial. *J Clin Oncol.* 2005; 23: 1-10.
14. Blum JL, et al. Phase II study of weekly albumin-bound paclitaxel for patients with metastatic breast cancer heavily pretreated with taxanes. *Clin Breast Cancer* 2007; 7: 850-6.
15. Baselga J, et al. Everolimus in Postmenopausal Hormone-Receptor-Positive Advanced Breast Cancer. *N Engl J Med.* 2011; 7.
16. Kaufman B, et al. Trastuzumab plus anastrozole versus anastrozole alone for the treatment of postmenopausal women with human epidermal growth factor receptor 2-positive, hormone receptor-positive metastatic breast cancer: results from the randomized TAnDEM study. *J Clin Oncol.* 2009; 27: 5529937.
17. Burris HA, et al. Phase II study of the antibody drug conjugate trastuzumab-DM1 for the treatment of human epidermal growth factor receptor 2 (HER2)-positive breast cancer after prior HER2-directed therapy. *J Clin Oncol.* 2011; 29: 398-405.
18. Burstein HJ, et al. Neratinib, an irreversible ErbB receptor tyrosine kinase inhibitor, in patients with advanced ErbB2-positive breast cancer. *J Clin Oncol.* 2010; 28: 1301-7.
19. Baselga J, et al. Pertuzumab plus Trastuzumab plus Docetaxel for Metastatic Breast Cancer. *N Engl J Med.* 2011; 7.
20. Guarneri V, et al. Comparison of HER-2 and hormone receptor expression in primary breast cancers and asynchronous paired metastases: impact on patient management. *Oncologist.* 2008; 13: 838-44.
21. Dieci MV, et al. Change in triple-receptor status between primary and recurrent breast cancer: prognostic impact. ABC1 Lisbon; *The Breast.* 2011; abstract BP32.

Caratterizzazione molecolare del tumore del colon-retto

Alberto Bardelli

Dipartimento di Scienze Oncologiche, Istituto per la Ricerca e Cura del Cancro, Candiolo (TO)

I pazienti affetti da carcinoma del colon-retto metastatico (mCRC) hanno un tasso di sopravvivenza a 5 anni inferiore al 5%. Nuovi farmaci diretti contro il recettore per il fattore di crescita epidermico (EGFR) sono di recente stati approvati per uso clinico, come gli anticorpi monoclonali (mAb) cetuximab e panitumumab. Tali agenti, spesso in combinazione con farmaci chemioterapici di ultima generazione, sono in grado di indurre la regressione del tumore e prolungare la sopravvivenza di pazienti affetti da neoplasie maligne, dimostrando l'importante ruolo delle vie di trasduzione dipendenti dall'EGFR nella crescita tumorale. Da recenti studi è emerso che solo il 10-20% di questi pazienti ha tratto un beneficio clinico dal trattamento con anticorpi anti-EGFR.

Un grande interesse si è quindi sviluppato attorno a studi volti a comprendere i meccanismi molecolari responsabili della sensibilità e resistenza a questi farmaci. Il nostro gruppo di ricerca ha contribuito alla scoperta che mutazioni somatiche in KRAS, presenti nel 35-45% dei pazienti mCRC, costituiscono un importante predittore negativo per i pazienti trattati con cetuximab o panitumumab (3). Inoltre, tra i tumori con KRAS wild type, mutazioni in BRAF o PIK3CA o perdita di espressione di PTEN, possono anche predire resistenza al trattamento con anticorpi anti-EGFR (2, 4, 5, 6). Molto recentemente abbiamo inoltre dimostrato come la presenza della amplificazione del gene HER2 sia correlata alla mancata risposta alle terapie anti EGFR nei carcinomi coloretali. In conclusione lo sviluppo clinico degli anticorpi monoclonali anti-EGFR cetuximab e panitumumab costituisce il primo esempio di trattamento individualizzato del cancro colonrettale metastatico.

Bibliografia

1. Bertotti A, GM, Salimi F, Sassi F, Torti D, et al. A molecularly annotated platform of patient-derived xenografts ('xenopatients') identifies HER2 as an effective therapeutic target in cetuximab-resistant colorectal cancer. *Cancer Discovery*. 2011.

2. Bardelli, Siena. Molecular mechanisms of resistance to cetuximab and panitumumab in colorectal cancer. *J Clin Oncol.* 2011.
3. Benvenuti S, et al. Oncogenic activation of the RAS/RAF signaling pathway impairs the response of metastatic colorectal cancers to anti-epidermal growth factor receptor antibody therapies. *Cancer Res.* 2007.
4. De Roock W, et al. Association of KRAS p.G13D mutation with outcome in patients with chemotherapy-refractory metastatic colorectal cancer treated with cetuximab. *JAMA.* 2010.
5. Di Nicolantonio F, et al. Wild-type BRAF is required for response to panitumumab or cetuximab in metastatic colorectal cancer. *J Clin Oncol.* 2008.
6. Sartore-Bianchi A, et al. PIK3CA mutations in colorectal cancer are associated with clinical resistance to EGFR-targeted monoclonal antibodies. *Cancer Res.* 2009; 69: 1851-7.

Diagnosi molecolare del tumore del polmone

Federico Cappuzzo

Direttore Dipartimento di Oncologia Medica, Ospedale Civile di Livorno

Il tumore polmonare rappresenta la prima causa di morte tumore-correlata nei Paesi industrializzati. Rappresenta il 20% dei tumori diagnosticati nel sesso maschile, mentre nel sesso femminile l'incidenza è storicamente più bassa, sebbene in progressivo incremento come risultato dell'abitudine al fumo da parte delle donne. In Italia muoiono per tumore del polmone circa 27.500 persone all'anno (circa 22.000 uomini e 5.500 donne), rappresentando la prima causa di morte oncologica negli uomini e la seconda nelle donne. Fino a pochi anni fa, i tumori polmonari venivano distinti in due grosse categorie istologiche, il tumore polmonare non a piccole cellule (*Non-Small-Cell Lung Cancer, NSCLC*) ed il microcitoma (*Small Cell Lung Cancer, SCLC*).

L'approccio terapeutico variava unicamente in base a tale distinzione. Nell'ultimo decennio, però, notevoli progressi sono stati fatti nella caratterizzazione istopatologica e nel trattamento soprattutto del NSCLC e, dopo anni di un approccio terapeutico del tipo "*one size fits all*", si è reso evidente come le sottocategorie istologiche del NSCLC rappresentino forme tumorali distinte, che possono beneficiare di trattamenti specifici, come ad esempio il bevacizumab ed il pemetrexed nelle forme non squamose.

La chemioterapia con l'associazione di platino e un farmaco di terza generazione ha rappresentato lo standard di trattamento per il NSCLC in stadio avanzato fino a poco tempo fa, quando importanti novità nella conoscenza della biologia del cancro hanno permesso di identificare alterazioni molecolari di rilievo, dette mutazioni "*driver*". Queste alterazioni genetiche interessano geni che codificano per proteine di segnale di cruciale importanza per la proliferazione e la sopravvivenza cellulare e possono essere responsabili della formazione e del sostentamento del tumore.

Tali conoscenze hanno permesso di creare e studiare una serie di trattamenti "target", ossia rivolti al blocco della funzione di una proteina o di un pathway specifico. Durante gli ultimi tre anni, pertanto, si è reso evidente come la selezione del paziente sia d'importanza cruciale per la scelta di un trattamento appropriato ed oggi una stratificazione dei pazienti sulla base dell'istologia e delle alterazioni molecolari è fondamentale per un'appropriata indicazione ad un trattamento di prima linea.

L'Epidermal Growth Factor Receptor

La scoperta nel 2004 delle mutazioni attivanti del gene del recettore del fattore di crescita epidermico (EGFR) ha rappresentato una pietra miliare nella storia del trattamento del NSCLC. La maggior parte delle mutazioni del gene EGFR osservate nel NSCLC interessano gli esoni codificanti il dominio tirosin-chinasico (esoni 18-21) e si ritrovano in circa il 10-15% dei pazienti non asiatici, mentre nelle popolazioni asiatiche tale percentuale è più elevata e si assesta intorno a valori del 30%. La presenza di mutazioni di EGFR è più comune, oltre che nelle popolazioni asiatiche, anche nel sesso femminile, nell'adenocarcinoma e nei pazienti non fumatori.

La scoperta dell'EGFR ha portato i ricercatori a condurre numerosi studi sia retrospettivi sia prospettici, che hanno messo in evidenza come la presenza di mutazioni attivanti di tale gene sia un importante fattore prognostico. I pazienti mutati, infatti, hanno una sopravvivenza più lunga rispetto ai pazienti che non presentino mutazioni di EGFR. Oltre ad essere un fattore prognostico, le mutazioni attivanti di EGFR sono anche un fattore predittivo di risposta a una specifica classe di farmaci, gli inibitori tirosin-chinasici di EGFR (*EGFR-TKIs*). Studi preclinici hanno dimostrato come linee cellulari di tumore polmonare con mutazione di EGFR siano inibite dagli EGFR-TKIs a concentrazioni molto inferiori rispetto a linee cellulari che non presentino tale mutazione. Il dato preclinico è stato ampiamente riportato anche in studi clinici controllati. Dopo i deludenti risultati ottenuti in pazienti non selezionati, si è reso evidente come il beneficio dato da erlotinib e gefitinib fosse ristretto a particolari sottocategorie cliniche di pazienti, ossia gli asiatici, le donne, l'adenocarcinoma e i non fumatori (studi IPASS e First-SIGNAL). Dalle analisi per sottogruppi di questi studi è emerso come i pazienti che beneficiavano maggiormente dal trattamento con EGFR-TKIs fossero solo i pazienti con mutazioni attivanti di EGFR. L'identificazione del valore predittivo delle mutazioni di EGFR, ha portato alla conduzione di diversi studi clinici che hanno valutato l'utilizzo degli EGFR-TKIs in pazienti selezionati dal punto di vista molecolare. Almeno quattro studi clinici di fase III (WJTOG 3405, NEJ002, EURTAC, OPTIMAL) hanno dimostrato come i pazienti con mutazione di EGFR abbiano un beneficio maggiore in termini di risposte obiettive (fino all'80-90%), tempo alla progressione (PFS, fino a 24 mesi) e qualità di vita da un trattamento con EGFR-TKIs rispetto ad una chemioterapia standard a base di platino. Questi studi hanno portato alla registrazione in tutto il mondo di erlotinib e gefitinib come trattamento di prima linea in pazienti con NSCLC con mutazioni attivanti di EGFR in fase avanzata di malattia. Sulla base dei dati riportati da questi studi, la conoscenza dello stato mutazionale di EGFR prima dell'inizio di un trattamento di prima linea è necessaria in tutti i pazienti con NSCLC in fase metastatica o localmente avanzata. Altri studi clinici che hanno valutato l'utilizzo degli EGFR-TKIs in seconda-terza linea (BR.21, INTEREST, TITAN) e come trattamento di mantenimento (SATURN, ATLAS) hanno messo in evidenza come il beneficio di questa classe di farmaci nei pazienti mutati sia notevole anche in queste fasi di trattamento. Oggi, pertanto, un trattamento con

EGFR-TKIs non può essere negato a nessun paziente non NSCLC in fase avanzata con mutazione attivante di EGFR.

Inevitabilmente tutti i pazienti con NSCLC, anche in presenza di mutazioni attivanti di EGFR, sviluppano resistenza al trattamento con TKIs. Sono stati identificati due meccanismi principali di resistenza secondaria ai TKIs, ossia la presenza della mutazione T790M del gene EGFR (50% dei casi) e l'amplificazione del gene MET (20% dei casi). Sono attualmente in studio farmaci che almeno teoricamente siano in grado di superare questi meccanismi di resistenza, come l'afatinib (un inibitore duplice ed irreversibile di EGFR ed HER2), l'ARQ197 (una piccola molecola inibitore di MET) e il ficlatuzumab (un anticorpo anti-MET).

Il gene Rat Sarcoma (RAS)

La famiglia dei gene RAS include H-RAS, K-RAS e N-RAS e codifica per proteine di membrana che legano la guanosina trifosfato. Tale proteina regola la crescita cellulare, la differenziazione e l'apoptosi mediante l'interazione con multipli effettori, come le MAP chinasi, PI3K e STAT. Le proteine RAS acquisiscono capacità trasformante quando una mutazione puntiforme nel gene sostituisce un aminoacido alla posizione 12, 13 o 61, portando alla formazione di forme di RAS con un'attività GTP-asica alterata e causando una attivazione costitutiva del pathway di segnale di RAS. Le mutazioni di RAS sono molto frequenti nel NSCLC, soprattutto nell'adenocarcinoma (20-30% dei casi), meno nel carcinoma squamoso (circa 7%). Nel NSCLC la maggior parte delle mutazioni di KRAS interessa i codoni 12 e 13 del gene e sono solitamente associate con l'abitudine al fumo di sigaretta. La frequenza di tali mutazioni varia fra i diversi gruppi etnici: sono osservate con minor frequenza negli asiatici, mentre sono molto più comuni tra afro-americani e caucasici. Recentemente mutazioni di KRAS sono state identificate in una percentuale notevole di pazienti non fumatori (15%), indicando che la presenza di tali mutazioni non può essere predetta sulla base della storia di abitudine al fumo del paziente, come invece accade per l'EGFR.

Diversi studi retrospettivi hanno valutato l'importanza di KRAS come fattore prognostico nel NSCLC, ma i risultati di questi studi sono contrastanti e il valore prognostico di KRAS rimane ancora controverso.

Ad oggi non abbiamo a disposizione farmaci inibitori di KRAS, ma è stato studiato se la presenza di mutazione di KRAS influenza la sensibilità alla chemioterapia o agli EGFR-TKIs. I dati attualmente disponibili non sono univoci e ad oggi l'uso dell'analisi mutazionale di KRAS non è consigliata per decidere quali pazienti devono essere avviati o meno a ricevere un trattamento con chemioterapici o con agenti biologici.

L'Anaplastic Lymphoma Kinase (ALK)

Le mutazioni attivanti o le traslocazioni del gene ALK sono state identificate in diversi tipi di neoplasie, compreso il NSCLC. EML4-ALK è un gene aberrante di fusione che codifica per un proteina chimerica citoplasmatica che presenta attivi-

tà chinasica intrinseca e potere trasformante in vitro. Tale alterazione può essere identificata con svariate tecniche includenti l'immunoistochimica e l'ibridazione in situ (FISH). La traslocazione EML4-ALK non è un'alterazione genica comune, essendo presente nel 2-7% dei NSCLC, soprattutto in giovani pazienti di sesso maschile con adenocarcinoma e mai fumatori o lievi fumatori. Dal punto di vista prognostico, la presenza della traslocazione EML4-ALK sembra avere un valore prognostico negativo. I pazienti con tale alterazione genica sembrano, infatti, avere una sopravvivenza minore ed una più alta incidenza di metastasi cerebrali ed epatiche rispetto ai pazienti che non presentino tale alterazione.

Recenti sono i dati di uno studio di fase I/II che ha utilizzato un farmaco, il crizotinib, che era nato come un inibitore di MET ma che si è dimostrato, in realtà, un potente inibitore di ALK. In questo studio, sebbene i pazienti fossero pretrattati con una o più linee di chemioterapia, è stato osservato un tasso di controllo di malattia molto elevato (circa l'80%). Questi promettenti risultati hanno portato subito alla progettazione di studi di fase III che sono attualmente in corso e di cui si aspettano i risultati a breve.

Conclusioni

Il miglioramento delle conoscenze nell'ambito della biologia dei tumori sta consentendo di definire strategie terapeutiche sempre più efficaci grazie all'impiego mirato di farmaci a bersaglio molecolare. Alcuni di questi farmaci, come l'erlotinib, il gefitinib o il crizotinib, sono già entrati nella nostra pratica clinica e offrono speranze più concrete ai pazienti affetti da NSCLC. Purtroppo ad oggi nessun paziente viene guarito in modo definitivo quando presenta una malattia in fase metastatica, in quanto inevitabilmente insorgono fenomeni di resistenza. Occorre, quindi, definire in modo sempre più preciso i meccanismi biologici che sostengono la crescita neoplastica con l'obiettivo di prolungare significativamente la vita dei pazienti.

Cancro: malattia genetica di cellule somatiche

Lucio Luzzatto

Direttore Scientifico dell'Istituto Toscano dei Tumori, Firenze

Per molto tempo il termine malattia genetica è stato considerato essenzialmente sinonimo di malattia ereditaria: quando uno od entrambi i genitori trasmettono un cromosoma o un gene anormale si può avere una malattia nei figli. Con lo sviluppo delle culture cellulari e con l'introduzione di tecniche per ottenere fusione tra cellule si è sviluppata la genetica delle cellule somatiche); e al tempo stesso con l'analisi citogenetica si scoprivano numerose anomalie cromosomiche non trasmesse dai genitori, ma evidentemente insorte *ex novo* in una cellula somatica. Per lo più queste anomalie venivano osservate in cellule tumorali: ed il loro significato patogenico divenne chiaro con la scoperta che una stessa traslocazione si trovava regolarmente in un certo tipo di tumore. I prototipi furono nella leucemia mieloide cronica il cromosoma Philadelphia), risultato di una traslocazione t(9; 22); e nel linfoma di Burkitt la traslocazione t(8; 14) (3).

Da questi dati apparve evidente, e dagli studi con virus oncogeni apparve sempre più chiaro che il tumore (o cancro, o neoplasia) insorge in seguito alla trasformazione di una cellula normale in una cellula tumorale (o cancerosa, o neoplastica). Credo che si debba a John Cairns (4) la prima formulazione esplicita del modello di tumore come malattia clonale in cui mutazioni successive conferiscono alle cellule mutanti un sempre maggiore vantaggio selettivo nel senso Darwiniano. Da allora si è accumulata una messe di dati che sostengono in pieno questo modello; e nel 2004 si potevano annoverare con certezza più di 400 geni ben identificati le mutazioni somatiche dei quali hanno un ruolo causale nell'oncogenesi (5). La maggioranza di questi geni ha ruoli pure identificabili in processi che hanno a che fare con il controllo della crescita cellulare, l'apoptosi, la risposta a stress, e il riparo del DNA. Inoltre, mutazioni in geni coinvolti nei processi di adesione e di migrazione *hanno* probabilmente un ruolo cruciale nel conferire a una cellula tumorale il potenziale metastatico (6).

L'elemento centrale della selezione Darwiniana di popolazioni è la eredità fedele dei geni, normali o mutati (fino quando non subentra una nuova mutazione). Lo stesso vale per le popolazioni cellulari selezionate nel corso della tumorigenesi. La dovizia di dati generati dalle tecnologie contemporanee (7) non deve confon-

derci: ognuna delle innumerevoli alterazioni del trascrittoma, del proteoma, del metaboloma, del miRNoma (8) che si trovano in una cellula tumorale può avere grande importanza funzionale, ma contribuirà alla tumorigenesi solo se è ereditabile. Pertanto, è sempre più cruciale identificare, per ogni tipo di tumore, quali siano le mutazioni somatiche che vi sottendono. Storicamente ciò è avvenuto in alcuni casi la scoperta di eventi genetici rivelabili mediante l'analisi citogenetica (il primo esempio è stato il famoso cromosoma Filadelfia, ed oggi sappiamo che a livello molecolare ciò che conta è il gene di fusione *BCR-ABL*); ma oggi è fattibile, almeno sulla carta, per qualsiasi tumore, attraverso il sequenziamento completo del suo genoma, usando come controllo interno il genoma di cellule non tumorali dello stesso paziente (9).

È evidente che per i tumori, come del resto per tutte le altre malattie, esistono cause ereditarie e cause ambientali: ma non è facile ottenere stime quantitative sui contributi relativi di questi due ordini di cause (10); e la questione viene resa assai complessa non solo dalla varietà dei possibili geni che predispongono al cancro, non solo dalla varietà degli agenti potenzialmente carcinogeni, ma anche dal grande numero di tipi del cancro stesso. Al momento attuale le stime per le cause ambientali vanno dal 40 all'80%; e per le cause ereditarie dal 5 al 20%, ma considerando il numero sempre crescente di geni a bassa penetranza che si stanno individuando, questa percentuale potrebbe aumentare.

In ogni caso, e qualunque sia la causa o siano le cause di un certo tumore, perché esso si sviluppi occorre che abbia luogo un certo numero di eventi genetici discreti, cioè mutazioni somatiche: queste avvengono essenzialmente a caso e pos-

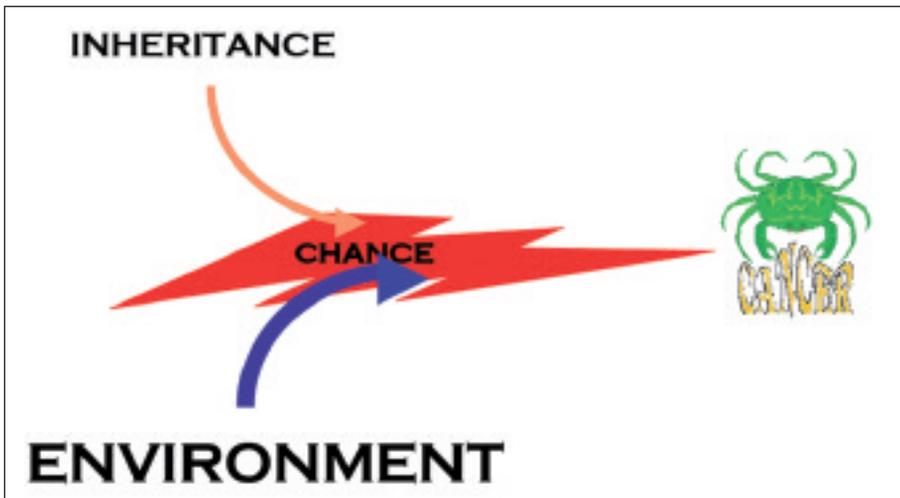


Fig. 1 - Sia fattori ambientali (esposizione) sia fattori ereditari (predisposizione) influiscono sul rischio di sviluppare il cancro. D'altro canto nessun cancro, o tumore, può svilupparsi senza che siano avvenute mutazioni somatiche. Queste sono eventi stocastici, o casuali: pertanto, anche un soggetto non predisposto e poco esposto può sfortunatamente ammalarsi di cancro; ed anche un soggetto predisposto od esposto può fortunatamente non ammalarsi. L'ereditarietà e l'ambiente agiscono in buona parte attraverso il loro effetto sul tasso di mutazioni somatiche: il caso perciò ha sempre un ruolo. (Da (10)).

sono perciò considerarsi eventi stocastici. Mutazioni somatiche hanno luogo in ogni fase dello sviluppo, anche prima della nascita, e anche indipendentemente dall'intervento di agenti esterni che chiamiamo mutageni: ma la loro frequenza è assai bassa, poiché il meccanismo di replicazione del DNA è assai fedele, protetto, ed in larga misura capace di auto-verificarsi (*proof-reading*) e di auto-correggersi (*repair*). La maggior parte delle mutazioni somatiche sono innocue: è solo quando certi geni sono mutati in un certo modo che una cellula normale diventa cancerosa. Ne possiamo dedurre che il *tasso di mutazioni somatiche* (μ) è probabilmente un parametro importante nel determinare il rischio di cancro. In linea di massima fattori ereditari possono influenzare μ , ad esempio perchè implicati nella efficienza del riparo; e i mutageni ambientali, per definizione, fanno aumentare μ . Recentemente abbiamo sviluppato una metodologia (11) che permette di misurare μ direttamente in linee cellulari linfoblastoidi (ottenibili da qualsiasi soggetto); e la frequenza di mutanti f , un parametro correlato con μ , si può misurare addirittura su un piccolo campione di sangue periferico.

Bibliografia

1. Pontecorvo G. Symposium No. 19: somatic cell genetics. Introduction by the chairman. *Genetics*. 1975; 79 (Suppl.): 339-341.
2. Chandra HS, et al. Philadelphia Chromosome Symposium: commemoration of the 50th anniversary of the discovery of the Ph chromosome. *Cancer Genetics*. 204 (4): 171-9.
3. George K. Specific chromosomal translocations and the genesis of B-cell-derived tumors in mice and men. *Cell*. 1983; 32 (2): 311-5.
4. Cairns J. Mutation, selection, and the natural history of cancer *Nature*. 1975; 255 (5505): 197-200.
5. Futreal PA, et al. A census of human cancer genes. *Nat Rev Cancer*. 2004. 4 (3): 177-83.
6. Nguyen DX, Bos PD, Massague J. Metastasis: from dissemination to organ-specific colonization. *Nat Rev Cancer*. 2009; 9 (4): 274-84.
7. Hanash SM, Baik CS, Kallioniemi O. Emerging molecular biomarkers (mdash) blood-based strategies to detect and monitor cancer. *Nat Rev Clin Oncol*. 8 (3): 142-50.
8. Croce CM, Calin GA. miRNAs, cancer and stem cell division. *Cell*. 2005; 122: 6-7.
9. Stratton MR, Campbell PJ, Futreal PA. The cancer genome. *Nature*. 2009; 458 (7239): 719-24.
10. Luzzatto L. Somatic mutations in cancer development. *Environmental Health*. 10 (Suppl. 1): S12.
11. Peruzzi B, et al. The use of PIG-A as a sentinel gene for the study of the somatic mutation rate and of mutagenic agents in vivo. *Mutation Research/Reviews in Mutation Research*. 705 (1): 3-10.

**GENOMICA E ATTUALITÀ
DI MEDICINA MOLECOLARE
IN EMOPATIE NON-NEOPLASTICHE**

Hemoglobinopathies and Thalassemia: origin of molecular medicine and new insights

Maria Domenica Cappellini, Alessia Marcon, Giovanna Graziadei

Fondazione "Ca Granda" Policlinico, Dipartimento di Medicina Interna, Università di Milano

The haemoglobinopathies, inherited disorders of the structure or synthesis of haemoglobin, are the commonest monogenic diseases. Approximately 80% of the annual births of babies with these conditions occur in low-or middle-income countries, many of which have extremely limited facilities for their control and management. They are divided into two main groups, the structural haemoglobin variants and the thalassaemias, which result from defective synthesis of the globin chains. The β -thalassaemias pose by far the most important global public health problem and they also lead the way to the concepts of 'molecular disease' and 'molecular medicine'.

It was becoming apparent that thalassaemia might not be a single disorder, but a complex syndrome characterized by wide phenotypic diversity that ranges from profound anaemia in early infancy, to moderate anaemia of the kind described by Rietti and colleagues, and symptomless carrier states identified only by morphological or osmotic abnormalities of the red blood cells. The problem of how it is possible for a single defective gene to produce the diverse clinical manifestations of the disease was the central issue in many of the reviews of the field in the late. In 1965, a method was developed for studying the *in vitro* synthesis of the α and β -globin chains of haemoglobin in a quantitative fashion. This work provided clear evidence that the thalassaemias are diseases of imbalanced globin-chain synthesis; in β -thalassaemia, for example, there are many phenotypes that range from an absence of β -chain production to a mild reduction in its output. In every variety there is excess α -chain production. Towards the end of the 1970s, techniques for the direct analysis of the globin genes were becoming available. Southern blotting was first applied to the study of the α -thalassaemias in 1978. It was confirmed that the severe forms result from deletions of both α -globin genes and that the milder forms result from the deletion or loss of function of one of the linked pairs. Southern blotting was less productive for the study of β -thalassaemia. However, by the early 1980s, it was possible to clone and sequence the globin genes, leading to the first successful identification of a single-base mutation as the cause of β -thalassaemia in 1981. It soon became apparent that sequencing β -globin genes from

random individuals would lead to the repeated identification of the same mutation. Both the α and β -globin gene clusters were found to have polymorphic restriction-enzyme sites that were in a series of patterns, or haplotypes, that varied in structure and frequency between different populations. It was reasoned that different types of β -thalassaemia mutation would be found in association with particular haplotypes. Using this approach, the discovery of β -thalassaemia mutations moved extremely quickly, and by the end of the 1980s, almost 100 different mutations had been identified. Further progress in speeding up the identification of mutations followed the development of PCR, oligonucleotide probes and more rapid DNA sequencing methods; by 2001, more than 200 β -thalassaemia mutations had been discovered, and close to 100 deletion and nondeletion varieties of β -thalassaemia had been identified. A complete updated list is available at the globin Gene Server Web Site (<http://globin.cse.psu.edu>).

It turned out that homozygotes or compound heterozygotes for β -thalassaemia mutations usually result in the phenotype of transfusion dependent thalassaemia major. However, in a consistent proportion of the cases, there are genotypes associated with milder forms, which are referred to as thalassaemia intermedia. These forms of β -thalassemias show a marked phenotypic heterogeneity which ranges from the severe thalassaemia major to the asymptomatic β -thalassaemia carrier state (1-4). Well-recognized mechanisms for the production of a milder phenotype are:

- a) the inheritance of a mild or silent β -thalassaemia mutation;
- b) the co-inheritance with homozygous β -thalassaemia of mutations of the β -globin gene associated with α -thalassaemia which by reducing the output of α -chains decrease the α /non α chain imbalance;
- c) the coinheritance of genetic determinants able to sustain a persistent production of γ -chain in the adult life, which again act by reducing the α /non α globin chain imbalance (Tab. 1).

Most commonly the persistence of HbF production depends on the co-transmission of a specific determinant of hereditary persistence of fetal hemoglobin (HPFH). The inherited persistent production of HbF in adult life belongs to two different categories according to the distribution of HbF within the red blood cells, which can be pancellular (homogeneous distribution) or heterocellular (het-

Tab. 1 - Determinants of disease severity.

-
- Molecular factors
 - inheritance of a mild or silent β -chain mutation
 - presence of a polymorphism for the enzyme Xmn-1 in the $G\gamma$ -promoter region, associated with increased HbF
 - co-inheritance of α -thalassaemia
 - increased production of α -globin chains by triplicated or quadruplicated α -genotype associated to β -heterozygosity; also from interaction of β - and $\delta\beta$ -thalassaemia
 - other genetic non-globin linked polymorphisms
 - Environmental factors may influence severity of symptoms, e.g.
 - social conditions
 - nutrition
 - availability of medical care
-

HbF = fetal hemoglobin. Taher A, et al. Blood Cells Mol Dis. 2006; 37: 12-20.

erogeneous distribution). Pancellular HPFH are caused either by deletion at variable extent of the β -globin cluster or by point mutation in the $G\gamma$ or $A\gamma$ promoter. Heterocellular HPFH are associated with quantitative trait loci linked or unlinked to the β -globin cluster.⁷ These heterocellular HPFH determinants were recently shown to play a major role in determining the phenotype of β -thalassemia. The major advances at this regards have being made in the discovery of critical modifier genes, such as Myb and especially BCL11A (B cell lymphoma 11A), a master regulator of HbF (fetal hemoglobin) and hemoglobin switching. Polimorphysms of BCL11A, Myb and γ -globin genes account for most of the variability in the clinical phenotypes in β -thalassemia and sickle cell anemia patients (Fig. 1).

So why did the haemoglobin field lead the way to the concepts of ‘molecular disease’ and ‘molecular medicine’? “There might be a crucial time in every field at which the drawing together of scientists of disparate interests provides the impetus for important progress. Equally importantly, studies of unusual phenotypes also led to the concept of multiple levels of genetic modifiers. The history of

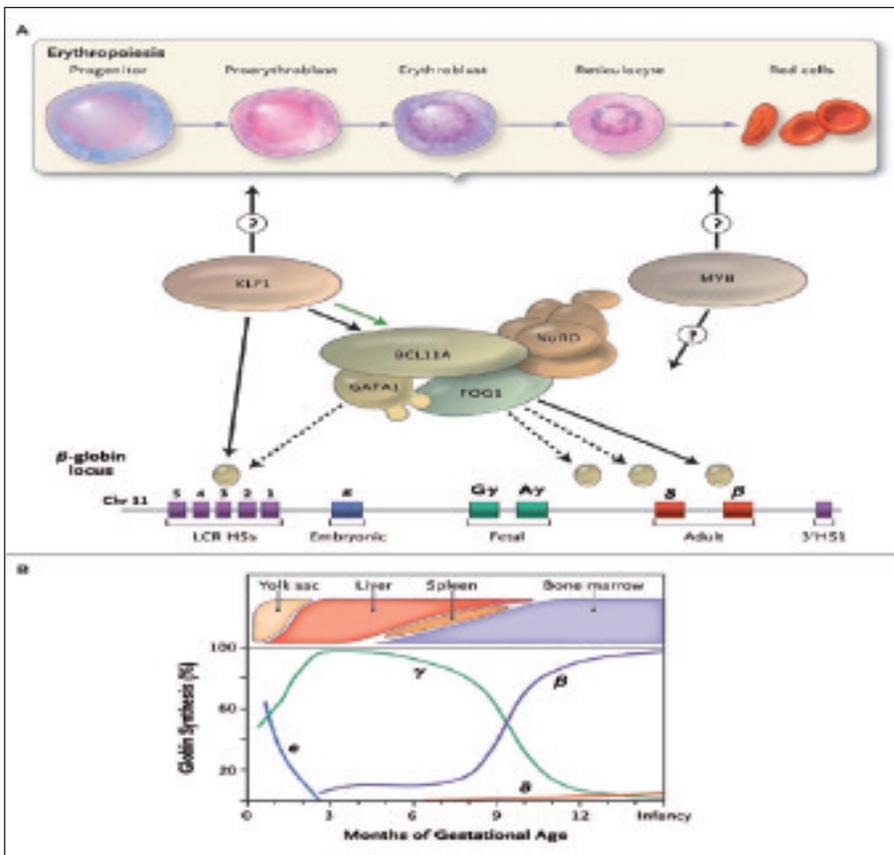


Fig. 1 - Molecular regulation of the fetal to the adult hemoglobin switch.

haemoglobin research also emphasizes how the clinical and basic sciences can feed off each other to their mutual benefit, and, incidentally, also highlights the importance of identifying and studying phenotypes that fall outside the usual run of particular genetic diseases” (Sir David Weatherall)

Bibliografia

1. Weatherall DJ. Phenotype-genotype relationships in monogenic disease: lessons from the thalassaemias. *Nature Genetics*: 2001; 2: 245.
2. Weatherall DJ. Thalassaemia: the long road from bedside to genome. *Nature Genetics*. 2004; 5: 1.
3. Cao A, Moi P, Galanello R. Recent advances in β -thalassemias. *Pediatric Reports* 2011; 3:e17.
4. Taher AT, Musallam KM, Karimi M, El-Beshlawy A, Belhoul K, Daar S, Saned MS, El-Chafic AH, Fasulo MR, Cappellini MD. Overview on practices in thalassemia intermedia management aiming for lowering complication rates across a region of endemicity: the OPTIMAL CARE study. *Blood*. 2010; 115 (10): 1886-92.

Disordini ereditari del metabolismo del ferro

Clara Camaschella

Divisione di Genetica e Biologia Cellulare, Istituto Scientifico San Raffaele,
Università Vita-Salute, Milano

Il ferro è essenziale per molte funzioni cellulari vitali quali il trasporto di ossigeno o la proliferazione, ma è tossico se presente in eccesso. Poiché non esiste un sistema di escrezione di ferro nei mammiferi, la sua disponibilità è strettamente controllata sia a livello cellulare che a livello sistemico. A livello cellulare la regolazione è operata dalle Iron Regulatory Proteins (IRPs), che controllano a livello posttrascrizionale numerosi geni del ferro tra cui il recettore della transferrina e la ferritina.

La regolazione dell'assorbimento intestinale e il rilascio dai macrofagi sono controllati dall'ormone del ferro epcidina, che regola l'espressione in membrana dell'esportatore di ferro ferroportina (1) in relazione alle necessità dell'eritropoiesi e dei depositi di ferro (2). Il ferro introdotto con la dieta è ridotto da ferro ferrico a ferroso dal citocromo b duodenale (Dcytb), una reduttasi dell'orletto a spazzola, e trasportato da DMT1, trasportatore di metalli divalenti, all'interno dell'enterocita.

Qui il ferro è utilizzato per le funzioni cellulari o stivato nella ferritina o ceduto dalla ferroportina alla transferrina circolante (circa 1-2 mg/die). Ferroportina è l'unico esportatore del ferro anche nei macrofagi splenici che riciclano il ferro (25-30 mg/die) proveniente dalla distruzione dei globuli rossi senescenti (2).

L'epcidina è un peptide di 25 aminoacidi secreto dal fegato la cui sintesi è stimolata dalle Bone Morphogenic Proteins (BMPs) in presenza di ferro, e dalle citochine infiammatorie, soprattutto interleuchina 6 (IL-6) (3). La produzione di epcidina è ridotta quando il fabbisogno di ferro è elevato, nella sideropenia, ipossia ed espansione eritropoietina, mentre è soppressa quando i depositi di ferro nel fegato e nel sistema macrofagico sono aumentati (3).

Il sistema epcidina-ferroportina e la sua regolazione sono stati chiariti grazie agli studi dell'emocromatosi ereditaria ed ai modelli murini sviluppati per chiarire i meccanismi patogenetici della malattia.

L'identificazione di questo peptide ha rivoluzionato la fisiopatologia dei disordini del metabolismo del ferro, ha permesso di riconoscere nuove entità nosologiche ed ha aperto prospettive terapeutiche innovative.

Emocromatosi ereditaria: malattia da deficit di epcidina

L'emocromatosi è una malattia genetica, caratterizzata da accumulo di ferro nel fegato e in altri organi. Il ferro in eccesso favorisce la formazione di radicali liberi dell'ossigeno, che danneggiano le cellule, determinando necrosi e fibrosi. Le complicanze cliniche comprendono cirrosi epatica, diabete ed altre endocrinopatie, scompenso cardiaco ed aritmie, danni articolari multipli e pigmentazione cutanea. L'epatocarcinoma è la causa principale di morte nei pazienti diagnosticati in fase cirrotica (4).

Le forme genetiche note di emocromatosi e le loro principali caratteristiche sono riassunte nella Tabella 1.

Il deficit di epcidina correla con la gravità del sovraccarico di ferro. Tutte le proteine dell'emocromatosi modulano la sintesi di epcidina: emogiuvellina è corecettore delle BMP, essenziale per la trasmissione del segnale attraverso le proteine SMAD, HFE e TFR2 sono verosimilmente modulatori del *pathway* principale di attivazione (Fig. 1), mentre ferroportina ne rappresenta il recettore (5).

L'emocromatosi dovuta a mutazioni del gene HFE è la forma più comune di emocromatosi, dovuta ad una mutazione prevalente (C282Y) in condizioni di omozigosi (4). L'emocromatosi giovanile, dove la produzione di epcidina è fortemente soppressa, è la forma più grave che può portare a sovraccarico cardiaco e ipofisario e, se non diagnosticata, anche a morte in giovane età. Il tipo 3 è molto raro ed ha un fenotipo intermedio: esordio giovanile ma minor gravità rispetto al tipo 2 (5). Il tipo 4 ha ereditarietà dominante e dipende da mutazioni di ferroportina: a seconda delle mutazioni il ferro si può accumulare nei macrofagi determinando ridotto rilascio e eritropoiesi ferro carente (tipo 4A) o, qualora le mutazioni riducano il legame epcidina-ferroportina, un sovraccarico di ferro da "resistenza all'epcidina" (tipo 4B).

La diagnosi di emocromatosi viene sospettata per l'alterazione della saturazione della transferrina (>45%) e i livelli elevati di ferritina. La biopsia epatica, un tempo indispensabile alla diagnosi, per determinare la concentrazione di ferro tissutale, è attualmente sostituita dalla Risonanza Magnetica Nucleare (RMN), che è un metodo non invasivo per valutare l'accumulo di ferro, che tuttavia non può evidenziare l'eventuale danno istologico. Il test genetico è necessario per la conferma diagnostica. Nei casi positivi è utile lo screening familiare.

Tab. 1 - Classificazione genetica dell'emocromatosi e differenze fenotipiche.

Patologia	Gene	N. OMIM	Meccanismo	Sovraccarico di ferro
Tipo 1	HFE	+235200	riduz parziale di HAMP	adulto, moderato
Tipo 2 A	HJV	#602390	riduz totale di HAMP	giovanile, grave
Tipo 2 B	HAMP	#602390	riduz totale di HAMP	giovanile, grave
Tipo 3	TFR2	#604250	riduz parziale di HAMP	intermedio
Tipo 4 A	FPN	#606069	ridotta FPN	macrofagico
Tipo 4 B	FPN	#606069	epcidina-resistenza	grave

HJV = hemojuvelin, HAMP = epcidina, TFR2 = recettore 2 della transferrina, FPN = ferroportina. OMIM = on line mendelian inheritance in man.

La terapia mira alla rimozione del ferro tramite salassi periodici (1 ml di emazie contiene circa 1 mg di ferro) e in rari casi terapia chelante del ferro.

Iron-refractory iron deficiency anemia (IRIDA): malattia da eccesso di epidina

Studi recenti hanno identificato le basi molecolari di forme rare di anemie ereditarie dovute ad alterazione del trasporto, uptake e utilizzo del ferro (6). La principale di queste anemie è l'**Iron-refractory iron deficiency anemia (IRIDA)**, una condizione recessiva rara causata da mutazioni del gene *TMPRSS6*, che codifica per matriptasi-2 (7), una proteasi che inibisce epidina. L'eccesso di epidina promuove la degradazione di ferroportina, limitando il rilascio di ferro con il risultato di un'anemia ipocromico-microcitica. La condizione è speculare all'emocromatosi. L'anemia non risponde al ferro orale, dato lo scarso assorbimento

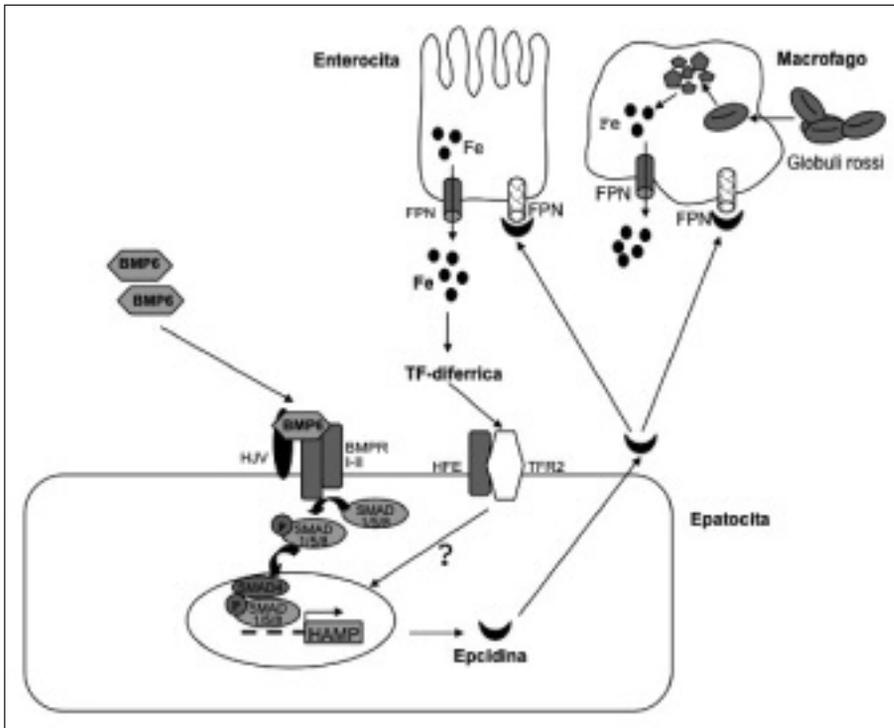


Fig. 1 - Meccanismo di azione di epidina e sua regolazione ferro-dipendente. La trascrizione di epidina nell'epatocita è regolata dal ferro attraverso i recettori (BMPRI e II) di Bone Morphogenetic Protein (BMP6) che, in presenza del corecettore emogiuelina (HJV), fosforilano (P) le SMAD, che nel nucleo attivano la trascrizione di HAMP a produrre epidina. Il gene dell'emocromatosi tipo 1 (HFE) ed il recettore 2 della Transferrina (TFR2) sono i verosimili sensori del ferro (Fe) legato alla transferrina (TF) e partecipano all'attivazione di epidina con meccanismo ancora poco chiaro indicato da?. Il legame epidina-ferroportina (FPN) promuove l'internalizzazione e la degradazione della ferroportina nell'enterocita e nel macrofago con conseguente blocco dell'assorbimento intestinale del ferro e del suo rilascio (modificata da Camaschella Nat Genet. 2009; 41: 386-8).

e risponde solo parzialmente al ferro parenterale, che viene sequestrato nei macrofagi (7). Epcidina, dosata su siero o urine è elevata in assoluto o relativamente, considerando che in genere è indosabile nel corso di sideropenia. La maggioranza dei casi descritti sono bambini, che nonostante la sideropenia hanno una crescita e sviluppo psico-fisico normali.

Esistono modelli murini sia per le diverse forme di emocromatosi (3) che per l'IRIDA. Nel topo "Mask" matriptasi-2 manca del dominio catalitico, nel "knock out" la proteasi è inattiva (7). Il fenotipo è sovrapponibile a quello umano per quanto riguarda il sovraccarico di ferro per l'anemia sideropenica ferro-resistente con eccesso di epcidina rispettivamente.

L'avanzamento delle conoscenze relative al metabolismo del ferro degli ultimi anni è stato spettacolare. Gli avanzamenti nel settore hanno prodotto nuovi bersagli terapeutici e potranno condizionare in futuro nuovi approcci a queste patologie (8). I modelli murini dei disordini del ferro saranno un buon modello per studi preclinici sull'utilizzo di agonisti/antagonisti di epcidina in vivo. Questi studi hanno implicazioni anche per il sovraccarico di ferro secondario che si presenta in soggetti anemici con espansione eritropoietica e bassi livelli di epcidina (2) o in disordini infiammatori caratterizzati da eccesso di epcidina che comporta sequestro di ferro macrofagico, eritropoiesi ferro-carente e anemia dei disordini cronici (10) per cui anche questi settori potranno beneficiare dell'avanzamento delle conoscenze nel settore.

Bibliografia

1. Nemeth E, Tuttle MS, Powelson J, et al. Heparin regulates cellular iron efflux by binding to ferroportin and inducing its internalization. *Science*. 2004; 306: 2090-3.
2. Andrews NC. Forging a field: the golden age of iron biology. *Blood*. 2008; 112: 219-30.
3. Hentze MW, Hentze MW, Muckenthaler MU, Galy B, et al. Two to tango: regulation of Mammalian iron metabolism. *Cell*. 2010; 142: 24-38.
4. Pietrangelo A. Hereditary hemochromatosis: pathogenesis, diagnosis and treatment. *Gastroenterology*. 2010; 139: 393-408.
5. Camaschella C. Understanding iron homeostasis through genetic analysis of hemochromatosis and related disorders. *Blood*. 2005; 106: 3710-7.
6. Iolascon A, De Falco L, Beaumont C. Molecular basis of inherited microcytic anemia due to defects in iron acquisition or heme synthesis. *Haematologica*. 2009; 94: 395-408.
7. Finberg KE. Iron-refractory iron deficiency anemia. *Semin Hematol*. 2009; 46: 378-86.
8. Nemeth E. Targeting the hepcidin-ferroportin axis in the diagnosis and treatment of anemias. *Adv Hematol*. 2010; 2010: 750643.
9. Weiss G, Goodnough LT. Anemia of chronic disease. *N Engl J Med*. 2005; 352: 1011-23.

The evolving spectrum of inherited thrombocytopenias

Carlo L. Balduini

Direttore Clinica Medica III, IRCCS Policlinico San Matteo, Pavia

Knowledge of inherited thrombocytopenias greatly improved during the last ten years, in that several new disorders have been discovered and some of them were unexpectedly found to be among the most frequent forms of genetic thrombocytopenia. This is the case of *MYH9*-related disease (*MYH9*-RD) (1) and *ANKRD26*-related thrombocytopenia (*ANKRD26*-RT) (2, 3). Moreover, a better understanding of the pathogenetic mechanisms of many forms allowed to realize that diseases previously considered separate entities derive from mutations of the same gene and are indeed different clinical expressions of a single disorder. For instance, identification of *MYH9* as the gene whose mutations cause May-Hegglin anomaly (MHA, MIM 155100) (4) allowed to recognize that Sebastian platelet syndrome (SBS, MIM 605249), Epstein syndrome (EPTS, MIM 153650) and Fechtner syndrome (FTNS, MIM 153640), derive from mutations of the same gene and describe overlapping disorders. Similarly, Montreal platelet syndrome has been cancelled from the list of inherited thrombocytopenias because it has been shown that subjects with this diagnosis were actually affected by von Willebrand Diseases 2B (5). The same fate occurred to Mediterranean macrothrombocytopenia after demonstration that some patients carried the genetic defects of sitosterolaemia (6) or had a mild form of Bernard-Soulier syndrome (BSS) (7).

Notwithstanding considerable progress, nearly half of patients with inherited thrombocytopenias remain without a definite diagnosis because their illnesses have not yet been described (8).

In this review we will discuss briefly the general aspects of inherited thrombocytopenias and will describe in more detail *MYH9*-RD and *ANKRD26*-RT, which have been identified recently and probably represent the less rare forms of genetic thrombocytopenia.

General aspects of inherited thrombocytopenias

Classification of inherited thrombocytopenias

Different parameters have been used to classify inherited thrombocytopenias, as the inheritance pattern or the presence of symptoms other than thrombocytopenia.

nia. However, the inheritance pattern is not always easy to identify because sporadic cases deriving from *de novo* mutations occur frequently (see below *MYH9-RD*). Moreover, transmission of some disorders, as BSS, may be both dominant and recessive. Likewise, a classification based on clinical symptoms is not always reliable, as syndromic and non-syndromic forms may result from mutations of the same gene, as in *MYH9-RD*. Moreover, patients with isolated thrombocytopenia at birth may develop the typical picture of a syndromic form in adult life (see *MYH9-RD*). Among other types of classification, one of the most satisfactory is based on platelet size. It is very effective for diagnostic purposes, in that platelet size is easy to determine by microscopic observation of peripheral blood smears and represents a constant feature of each illness (9). Table 1 describes the essential features of inherited thrombocytopenias classified according to this parameter. It also reports some references to relevant papers that describe these disorders.

Clinical picture

The bleeding diathesis of inherited thrombocytopenias is not different from that of acquired forms and consists of muco-cutaneous hemorrhages, such as spontaneous cutaneous bruising, nose bleeds, menorrhagia, and gastrointestinal bleeding. Few patients with platelet counts lower than $20\text{-}30 \times 10^9/\text{L}$ suffer from spontaneous hemorrhages since birth, but in most cases the degree of thrombocytopenia is mild and bleeding is occasional and/or related to trauma, surgery or invasive procedures. However, the bleeding tendency is sometime disproportionate to the degree of thrombocytopenia because of associated defects of platelet function. This is the case of patients with biallelic BSS, who may suffer from severe bleeding episodes also when thrombocytopenia is relatively mild (10). Some inherited thrombocytopenias differ from acquired forms also for the presence of additional defects to low platelet count (syndromic forms). Table 1 reports syndromic forms and their clinical picture.

Epidemiology

Inherited thrombocytopenias are considered exceedingly rare, but their prevalence is unknown because no population-based study has been performed. The relative frequency of different forms in our database, which currently includes nearly 500 patients from 160 families, is reported in Figure 1. In our experience, the most frequent disorder is a mild, dominant form of BSS due to monoallelic Bolzano mutation in GPIb alpha (11). However, this is an Italian disease, since analysis of the geographic origin of affected pedigrees and haplotypes indicated that this mutation originated in southern Italy (11). Moreover, no case has been described so far in other countries. The second, most frequent form is *MYH9-RD*, which has been diagnosed in 12% of our patients. According to our database, the prevalence of *ANKRD26-RT* is a little lower than that of *MYH9-RD*.

The Italian Registry for *MYH9-RD* (www.registromyh9.org) now includes 200 affected individuals, indicating that the prevalence of the disorder in Italy is at

Tab. 1 - Essential features of inherited thrombocytopenias classified according to platelet size. Same relevant papers describing these disorders in detail are indicated.

Disease (abbreviation, MIM #), (ref.)	Inheritance	Gene (localization)	Other features
Large platelets			
Biallelic (BSS, 231200) (10) and monoallelic Bernard-Soulier syndrome (nd, nd), (11)	AR-AD	GP1BA (17p13), GPIBB (22q11), GP9 (3q21)	None
<i>MYH9</i> -related disease (<i>MYH9</i> -RD, nd), (1)	AD	<i>MYH9</i> (22q12-13)	Leukocyte inclusions, cataracts, nephropathy and/or deafness
Dyserythropoietic anaemia with thrombocytopenia (nd, 300367) (13)	XL	<i>GATA1</i> (Xp11)	Haemolytic anaemia
X-linked thrombocytopenia with thalassemia (XLTT, 314050) (14)			Haemolytic anaemia, unbalanced globin chain synthesis
Paris-Trousseau thrombocytopenia (TCPT, 188025/600588), Jacobsen syndrome (JBS, 147791) (64)	AD	Large deletion (11q23)	Cardiac and facial defects, developmental
Gray platelet syndrome (GPS, 139090), (65-68)	AD-AR	<i>NBEAL2</i> (3p21.1)	“Pale” platelets, myelofibrosis, splenomegaly, high serum vitamin B12
Sitosterolaemia (STSL, 21025) (6)	AR	<i>ABCG5</i> , <i>ABCG8</i> (2p21)	Stomatocytosis, possible anaemia, tendon xanthomas, atherosclerosis
Platelet-type von Willebrand disease (PTvWD, 177820) (69)	AD	<i>GPIBA</i> (17pter-p12)	The platelet count goes down under stress
<i>ITGA2/ITGB3</i> -related thrombocytopenia (nd, nd) (70)	AD	<i>ITGB3</i> (17q21.32)	None
<i>TUBB1</i> -related macrothrombocytopenia (nd, nd) (71)	AD	<i>TUBB1</i> (6p21.3)	None
<i>FLNA</i> -related macrothrombocytopenia (nd, nd) (72)	X-L	<i>FLNA</i> (Xq28)	Possible periventricular nodular heterotopia, bone dysplasia, mental retardation, and other malformations
Normal-sized platelets			
<i>ANKRD26</i> -related thrombocytopenia (THC2, 313900) (2, 3)	AD	<i>ANKRD26</i> (10p2)	Risk of leukemia?
Congenital amegakaryocytic thrombocytopenia (CAMT, 604498), (18)	AR	<i>c-MPL</i> (1p34)	Reduced megakaryocytes. Evolution into bone marrow aplasia
Congenital amegakaryocytic thrombocytopenia with radio-ulnar synostosis (CTRUS, 605432), (19)	AD	<i>HOXA11</i> (7p15-14)	Reduced megakaryocytes. Possible evolution into aplastic anaemia. Radio-ulnar synostosis +/- other defects
Familial platelet disorder and predisposition to acute myelogenous leukemia (FPD/AML, 601399), (20)	AD	CBFA2 (21q22)	Development of leukemia or MDS in 40% of patients
Thrombocytopenia with absent radii (TAR, 274000), (73)	AR	Unknown (unknown)	Platelet count tends to rise and often normalizes in adult life. Reduced megakaryocytes. Bilateral radial aplasia +/- other malformations
<i>CYCS</i> -related thrombocytopenia (THC4, 612004), (74)	AD	<i>CYCS</i> (7p15.3)	None
Small platelets			
Wiskott-Aldrich syndrome (WAS, 301000), (75)	X-L	<i>WAS</i> (Xp11)	Severe immunodeficiency
X-linked thrombocytopenia (XLT, 313900), (75)			No or mild immunodeficiency
Abbreviations: ref. = reference.			

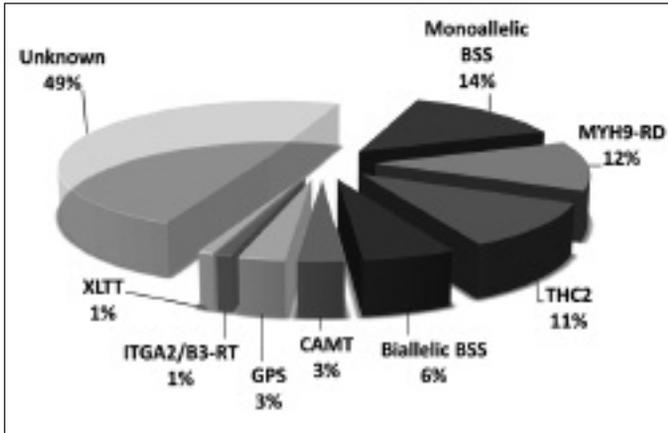


Fig. 1 - Relative frequency of different forms of inherited thrombocytopenia in 160 families (500 patients) observed in Pavia in the last 10 years. Despite recent progress, nearly half of patients remain without a diagnosis because their diseases have not yet been described.

least 3.2 in 1,000,000. Based on this figure and the relative frequency of *MYH9*-RD in our case series, we calculate that the prevalence of inherited thrombocytopenias in Italy is at least 2.7 in 100,000. Because mild forms are discovered incidentally and severe forms are often misdiagnosed with other disorders, the actual prevalence of these disorders is expected to be higher.

Pathogenesis

Megakaryocytes (Mks) are formed from hematopoietic stem cells in the bone-proximal proliferative osteoblastic niche. During maturation, they migrate from the osteoblastic to the vascular niche, which is enriched in vascular sinusoids. Here, mature Mks bind to vessels and extend proplatelets, which protrude into the vascular lumen and release nascent platelets directly into the bloodstream (12). Although the genes whose mutations result in thrombocytopenia have been identified for the vast majority of known disorders (Table 1), the exact molecular mechanisms that translate mutations into low platelet counts are still a matter of debate. Figure 2 describes the current model of megakaryocytopoiesis and the steps that have been hypothesized to be affected by some genetic defects. Table 2 makes an attempt to classify inherited thrombocytopenias according to their pathogenesis. In most cases, defects in the complex mechanisms that regulate Mk maturation (for instance, in thrombocytopenias deriving from *GATA1* mutations) (13, 14) and proplatelet formation (BSS and *MYH9*-RD) (15-17) are responsible for reduced platelet production. More rarely, a defect of Mk commitment and/or differentiation reduces the number of bone marrow Mks. These forms are often severe and put the patients at risk of developing bone marrow aplasia (Congenital amegakaryocytic thrombocytopenia, CAMT, and Congenital amegakaryocytic thrombocytopenia with radio-ulnar synostosis, CTRUS) or leukemia (Familial platelet disorder and predisposition to acute myelogenous leukemia, FDP/AML) (18-20). In a few cases, reduced platelet survival has been demonstrated or hypothesized.

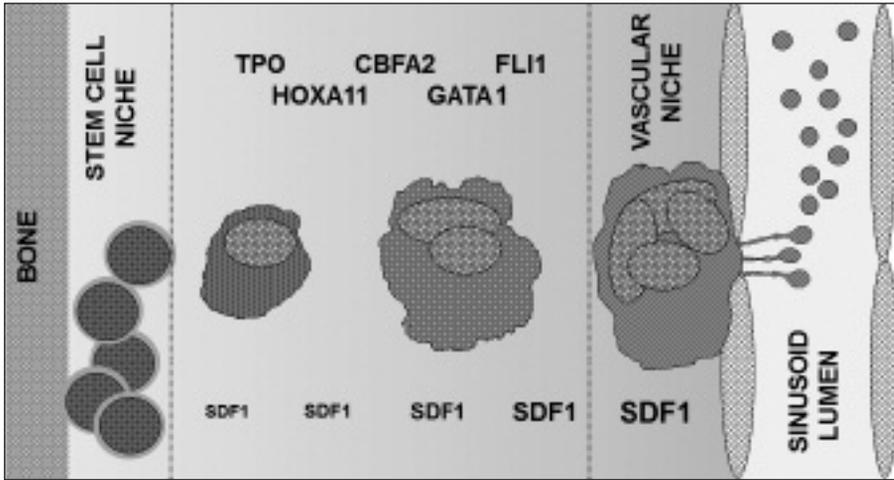


Fig. 2 - Schematic representation of megakaryocytopoiesis. Thrombopoietin (TPO) stimulation induces megakaryocytic differentiation of the hematopoietic stem cells located in the bone-proximal osteoblastic niche. Then, different transcription factors and TPO regulate maturation of Mk's while they migrate from the osteoblastic to the vascular niche guided by a stromal derived factor-1 (SDF-1) gradient. Here, mature Mk's bind to vessels and extend proplatelets, which protrude into the vascular lumen and release nascent platelets directly into the bloodstream (12). A defect of cellular response to TPO is responsible for both congenital amegakaryocytic thrombocytopenia and thrombocytopenia with absent radii. Defective HOXA-11 and CBFA2 impair megakaryocytic commitment-differentiation and result in amegakaryocytic thrombocytopenia with radio-ulnar synostosis and familial platelet disorder with predisposition to acute myelogenous leukemia, respectively. Deletion of the gene Fli-1 induces Paris-Trousseau type thrombocytopenia and Jacobsen's syndrome, characterized from defective megakaryocytic maturation. GATA-1 mutations cause dyserythropoietic anemia with thrombocytopenia and X-linked thrombocytopenia with thalassemia. GATA-1 and its cofactor FOG-1 regulate Mk commitment, differentiation, and maturation. However, GATA-2 can substitute GATA-1 for the interaction with FOG-1 in early but not late phases of megakaryocytopoiesis, and therefore subjects with GATA-1 mutations have recognizable Mk's that fail to mature properly. Finally, mutations of the genes MYH9 and CYCS, as well as those that induce BSS, severely affect proplatelet formation and platelet release in flowing blood.

Diagnosis

Differentiation between inherited and acquired thrombocytopenias

The recognition of the genetic origin of a thrombocytopenia is often hampered by several difficulties. Especially in the mild forms, thrombocytopenia is frequently identified incidentally late in infancy or in adult life, and, therefore, suspecting a congenital form is not obvious, particularly when patient have *de novo* mutations and no other family members are affected. Moreover, in inherited macrothrombocytopenias both platelet count and volume are regularly underestimated by the automated cell counters that measure platelet parameters by evaluating particles within a specified volume window (e.g., 2-20 fL). As a consequence, subjects with very large platelets (as in BSS and MYH9-RD) are at risk of misdiagnosis with severe immune thrombocytopenia and of undue medical therapy or splenectomy, as has been reported to have occurred in several cases (21). To avoid this error, careful investigation of each patient with isolated thrombocytopenia is

essential, especially in those cases with no record of previously normal platelet counts. Morphologic examination of peripheral blood smears is the most informative single examination because it can reveal qualitative platelet abnormalities, such as platelet macro- or microcytosis, that strongly support an inherited disorder. Cut off values of platelet size for distinguishing between inherited macrothrombocytopenias and autoimmune forms have been recently proposed (22). Also medical history and physical examination have an important role in the diagnostic process since they can allow to identify one of the many defects that are associated with thrombocytopenia in inherited syndromic forms.

Identification of specific forms of inherited thrombocytopenia

Although requiring an update to include some new forms, the algorithm proposed in 2003 by the Italian Platelet Study Group (9) and modified in 2004 by Noris et al. (8) may be useful for diagnostic purposes. It is based on a first phase of clinical investigations (differentiation between syndromic and non-syndromic forms) and simple laboratory tests (evaluation of platelet size, *in vitro* platelet aggregation and flow cytometry for platelet glycoproteins) to put forward diagnostic hypotheses. Then, more complex studies are required to confirm the diagnostic suspicion. Since the genes responsible for nearly all known inherited forms of thrombocytopenia have been identified, mutation screening can further confirm diagnosis and, in some cases, it can also define prognosis (see below *MYH9*-RD). Unfortunately, not all inherited forms of thrombocytopenia have been identified and about 50% of patients with (usually mild) genetic thrombocytopenias remain without a definite diagnosis because they do not fit the criteria of any known disorder.

Therapy

The most important aspect of management of inherited thrombocytopenias is to anticipate risks and to prevent bleeding. A major measure is to instruct patients to avoid drugs that impair platelet function (above all, aspirin). Regular dental care is essential to prevent gingival bleeding. Oral contraceptives should be used to prevent menorrhagia.

Transfusions of platelet concentrates may be used to transiently increase platelet count. However, they expose to the risks of alloimmunization, subsequent refractoriness to platelet infusions, infectious diseases, and febrile reactions. Thus, platelet transfusions should only be used in case of hemorrhages that cannot be otherwise managed and as prophylaxis prior to surgery or other major haemostatic stresses. When available, platelets from HLA-matched donors should be used to prevent or overcome alloimmunization. However, recent demonstration that an oral TPO mimetic increased platelet count in patients with *MYH9*-RD (see below *MYH9*-RD) opened new therapeutic perspectives also for other forms of inherited thrombocytopenia (23).

Allogeneic hematopoietic stem cell transplantation is an appealing therapy for restoring normal megakaryocytopoiesis. However, in most cases the risk of this

procedure is higher than that induced by the bleeding tendency, and transplantation has been used so far only to cure patients with very severe forms (CAMT, BSS, and Wiskott-Aldrich syndrome, WAS) (24, 25).

MYH9-related disease

MYH9-RD is the autosomal dominant syndromic thrombocytopenia deriving from mutations in the gene for the heavy chain of the non-muscle myosin IIA (NMMHC-IIA) (4, 26, 27). The term *MYH9*-RD encompasses four thrombocytopenias that were previously described as distinct diseases, namely MHA, SBS, FTNS and EPTS. The identification of *MYH9* as the gene responsible for all these conditions and the subsequent analysis of large case series of patients demonstrated that MHA, SBS, FTNS and EPTS represent different clinical presentations of the same disease, which is characterized by congenital macrothrombocytopenia often associated with “Döhle-like” leukocyte inclusions, and a variable risk of developing proteinuric nephropathy, sensorineural deafness, and/or cataract during infancy or adult life (28, 29).

Etiology

MYH9 (MIM 160775) is a large gene with 40 coding exons localized on chromosome 22q12.3. Non-muscle myosin IIA is a conventional, non-sarcomeric myosin expressed in most cell types and tissues (Fig. 3). As the other conventional myosins, it consists of two heavy chains (NMMHC-IIAs) and two couples of light chains with regulatory functions (30). The NMMHC-IIA molecule includes the N-terminal domain and the tail domain: the former consists of the catalytic globular head (or motor domain), which interacts with actin and binds and hydrolyzes ATP, and the neck domain, which acts as lever arm for transducing force generated by the motor domain and binds the light chains. The tail domain includes a long, α -helical coiled-coil portion, which is required for dimerization between the heavy chains and self-associations of dimers to form bipolar functional myosin filaments, and a short C-terminal non-helical tailpiece, which is a phosphorylation site with regulatory function (31, 32). Non-muscle myosin IIA participates in functions requiring the production of intracellular chemomechanical force by the cytoskeleton and the sliding along actin filaments, including cytokinesis, cell motility and polarization, cell-cell and cell-matrix interaction, and maintenance of cell shape (30).

To date, at least 44 different *MYH9* mutations have been reported in 218 *MYH9*-RD unrelated families (1). The spectrum of mutations is limited, suggesting that only specific alterations of the NMMHC-IIA molecule are compatible with the *MYH9*-RD phenotype. Most mutations are single aminoacid substitutions hitting only 19 of the 1960 residues of the protein. All the nonsense and frame-shift alterations affect the 34 residues of the C-terminal non-helical tailpiece. In rare cases, the disease derived from in frame deletions or duplications of a few nucleotides, occurring prevalently in the exon 24. A relevant aspect is the high frequency of

sporadic cases (about 35%), most of them carrying mutations in the head domain (1).

Pathogenesis

NMMHC-IIA mutations cause thrombocytopenia by altering the late processes of platelet release by mature Mks, while differentiation and maturation of Mks are unaffected (33). Recent studies showed that myosin IIA is a negative regulator of platelet production through the inhibition of proplatelet extension by Mks (34, 35). In particular, a functional NMMHC-IIA is essential for the suppression of proplatelet formation initiated by Mks interaction with type I collagen through the $\alpha 2\beta 1$ integrin. This binding activates the small GTPase Rho and the Rho-kinase ROCK, which in turn phosphorylates the regulatory light chains of NMMHC-IIA (34-37). Being an abundant extracellular protein of the osteoblastic niche of bone marrow, type I collagen prevents Mks from a premature release of platelets in this space (38). Consistent with this model, Mks cultured from blood progenitors of patients with two different *MYH9* mutations extended proplatelets even in adhesion to type I collagen, suggesting that *MYH9*-RD thrombocytopenia derives from an ectopic platelet release that results in ineffective platelet production (17). Moreover, proplatelets formed by *MYH9*-RD Mks showed evident morphological abnormalities, such as a reduced number and an increased size of proplatelet tips (17). *Myh9* knock-out mice show early embryonic death, mainly due to defects in cell-cell adhesion and consequent failure to form a visceral endoderm (39). Recently, a knock-in mouse model of *MYH9*-RD recapitulating the phenotype of the human disorder has been developed (40). Proplatelets extended by Mks of these mice presented morphological alterations that were remarkably similar to those observed in patients' Mks (23, 40). However, these Mks presented a reduced propensity to form proplatelets, in contrast to what would be expected from the previous achievements on the role of NMMHC-IIA in platelet production (40). Another recent *in vitro* study demonstrated that NMMHC-IIA is essential for chemotaxis of a megakaryocytic cell line in response to stromal derived factor-1 (SDF-1), and that three mutations responsible for *MYH9*-RD result in a reduced efficiency of this process (41). These observations suggest that an impaired SDF-1-driven migration of Mks from the osteoblastic niche to the perivascular space contributes to thrombocytopenia of *MYH9*-RD.

Clinical features

Congenital haematological defects

Macrothrombocytopenia and bleeding diathesis. Platelet macrocytosis is present in all *MYH9*-RD patients and is more marked than in immune thrombocytopenia and other forms of inherited macrothrombocytopenia (22). Giant platelets (larger than erythrocytes) are always present at examination of peripheral blood smears (42). Platelet count varies greatly among different patients, and usually remains

stable in each individual throughout life. Because of the extreme degree of platelet macrocytosis, automated cell counters usually underestimate platelet count and mean platelet volume (see above) (22). In a case series of 108 consecutive patients, platelet count ranged from 3 to 178 x 10⁹/L (median: 68 x 10⁹/L) by microscopic counting, and from 3 to 130 (median: 31 x 10⁹/L) by automated counters. By microscopy evaluation, the proportion of patients with less than 50 x 10⁹/L platelets was 30% (43). Bleeding tendency is globally proportionate to platelet count and therefore varies widely among different *MYH9*-RD patients. Spontaneous bleeding, sometimes requiring transfusions, is observed in subjects with platelet counts less than 50x10⁹/L (23); patients with higher platelet concentrations are usually asymptomatic and thrombocytopenia is often recognized only on the occasion of a blood cell count performed for unrelated clinical problems or for an anomalous bleeding after minor surgery, such as dental procedures. *Döhle-like leukocyte inclusions* are detected at evaluation of peripheral blood smears upon conventional staining in 42-84% of *MYH9*-RD patients (1). They appear as faint, light blue areas in the cytoplasm of neutrophils with diameter ranging from 1 to 7 µm and different shape (round, oval, spindle-shaped). Their recognition is pathognomonic for the disease but it is often difficult because of their faint appearance.

Extra-haematological manifestations

Kidney damage occurs in about 30% of patients and in 37-48% of pedigrees; it is often present in only some of the affected family members (43-45). Proteinuria, with or without microscopic haematuria, usually appears before the age of 30, and about two-third of patients develop renal failure and then end-stage renal disease within a few years. In the remaining cases, progression of the kidney impairment is slower with proteinuria and mild alteration of renal function stable for years (45-48).

Sensorineural hearing loss is the most frequent extra-haematological manifestation, being reported in 60% of patients and in 36-71% of pedigrees (43, 44). At the onset or in mild forms, the defect is evident only for high tones, but in more severe cases it extends toward the low frequencies. The age at onset is homogeneously distributed along the decades from first to sixth. Forms with onset in the childhood have severe progression and lead to deafness by the age of 30 (47).

Presenile cataract occurs in 16% of patients and 14-17% of families. The mean age at onset is 23 years, but congenital forms have been reported (49).

Diagnosis

The diagnostic suspicion of *MYH9*-RD is based on the recognition that thrombocytopenia is congenital and that it is characterized by giant platelets. Difficulties in identifying these features have been already discussed in the “general aspects” section. The most challenging points in *MYH9*-RD are the high frequency of sporadic cases, who therefore present with a negative familiar history, and the fact that automated cell counters underestimate platelet macrocytosis in almost all

patients. Once suspected, diagnosis can be easily confirmed by the immunofluorescence test for NMMHC-IIA distribution within neutrophils. This assay allows to identify typical aggregates of the MYH9 protein in the cytoplasm of neutrophils, which is a specific feature of *MYH9*-RD (27, 50). NMMHC-IIA aggregates are easily recognizable in all the neutrophils (28), with the only exception of the rare patients with mosaicisms, who have also neutrophils without aggregates (51). In a prospective analysis of 118 consecutive unrelated subjects with suspected *MYH9*-RD, the assay demonstrated 100% sensitivity and 95% specificity with respect to the presence of *MYH9* mutations (42). This assay can be easily centralized, as it can be effectively carried out on shipped blood films (www.registromyh9.org). Even if the immunofluorescence test has very high diagnostic power, identification of the causative mutations has great value in terms of prognostic assessment (see below).

Genotype-phenotype correlations

In 2005, a review of 85 previously published *MYH9*-RD pedigrees suggested that mutations involving the Arg702, in the motor domain of NMMHC-IIA, were associated with a syndromic disease, while alterations in the C-terminal portion of the tail domain of the molecule were more frequently responsible for isolated thrombocytopenia (44). In 2008, detailed characterization of 108 consecutive patients from 50 families confirmed and extended these findings, demonstrating that mutations affecting the head domain were associated with a significantly higher incidence of kidney damage and deafness and more severe thrombocytopenia with respect to mutations involving the tail domain. In particular, all the patients with aminoacid substitutions in the head domain of NMMHC-IIA were expected to develop nephropathy and deafness before the age of 40. Among the most frequent mutations of the tail domain, the p.R1933X alteration correlated with a negligible risk of extra-haematological complications, while the substitutions of the Asp1424 or Glu1841 were associated with an intermediate risk (43). Interestingly, pedigrees with these latter mutations were those characterized by a variability of phenotypic expression within affected family members. The smaller case series published later on were consistent with this findings and, in particular, pointed out the rapid deterioration of kidney and hearing function correlated with the Arg702 mutations (47, 48, 52).

Therapy

The recent demonstration that eltrombopag, an orally available TPO receptor agonist, is effective in patients with *MYH9*-RD opened new prospects in the treatment of this disorder (23). In 12 patients with platelet counts below $50 \times 10^9/L$, thrombocytopenia significantly improved at the end of 3-6 weeks treatment, and 7 subjects achieved platelet concentration over $100 \times 10^9/L$. Importantly, remission of bleeding tendency was obtained also in patients with minor improvements in platelet counts. Since no important side effects have been recorded, short-term

eltrombopag is a possible alternative to platelet transfusion in preparation of *MYH9*-RD patients to elective surgery, while further testing is required before using this drug for long-term treatment of subjects with frequent and severe episodes of spontaneous bleeding (23). There are no available data about the platelet concentration required for safe surgery in *MYH9*-RD. However, since *in vitro* platelet function of platelets of these patients is normal or only slightly reduced (23, 33, 53), physicians can refer to general recommendations for thrombocytopenias (54). Recent observations suggested that treatment with angiotensin-converting enzyme inhibitors and/or angiotensin receptor blockers reduce proteinuria of patients with early stage kidney involvement (46, 47). Subjects with end-stage kidney failure require dialysis and kidney transplantation. Standard cataract surgery should be carried out when indicated.

ANKRD26-related thrombocytopenia

Thrombocytopenia 2 (THC2) is an autosomal dominant form of thrombocytopenia originally described in two large families, one in Italy (55), the second one in the US (56) whose molecular defects were mapped to chromosome 10p11.1-p12. Subsequently, two missense mutations in the *MASTL* (MIM 608221) and *ACBD5* genes were identified as the causative defects in the American family (57) and Italian family (58), respectively.

Etiology

The hypothesis that mutations in genes other than *MASTL* and *ACBD5* may be responsible for this disorder derived from the observation that neither *MASTL* nor *ACBD5* were mutated in 4 additional Italian pedigrees with an autosomal dominant thrombocytopenia mapping to the THC2 locus. Then, mutational screening of all genes, other than *MASTL* and *ACBD5*, located in this region identified nucleotide changes within the 5'-untranslated region (5'UTR) of *ankyrin repeat domain 26* (*ANKRD26*) in three of the investigated families (3). Subsequently, analysis of a large database recognized 12 different mutations in 5'UTR of *ANKRD26* in 21 of 210 pedigrees (2), making this disorder one of the less rare forms of inherited thrombocytopenia (Fig. 1). Interestingly, *ANKRD26* was mutated also in the Italian family carrying the *ACBD5* change (58), which, therefore, was probably a private polymorphism rather than the causative mutation. No information is presently available on the *ANKRD26* state in the American family with *MASTL* changes. Thus, we propose the name of *ANKRD26*-related thrombocytopenia (*ANKRD26*-RT) to identify THC2 patients with mutations in the 5'UTR of *ANKRD26*.

Pathogenesis

ANKRD26 is the ancestral gene for a primate-specific family of genes named POTE (Prostate-, Ovary-, Testis-, and placenta-Expressed genes), which are

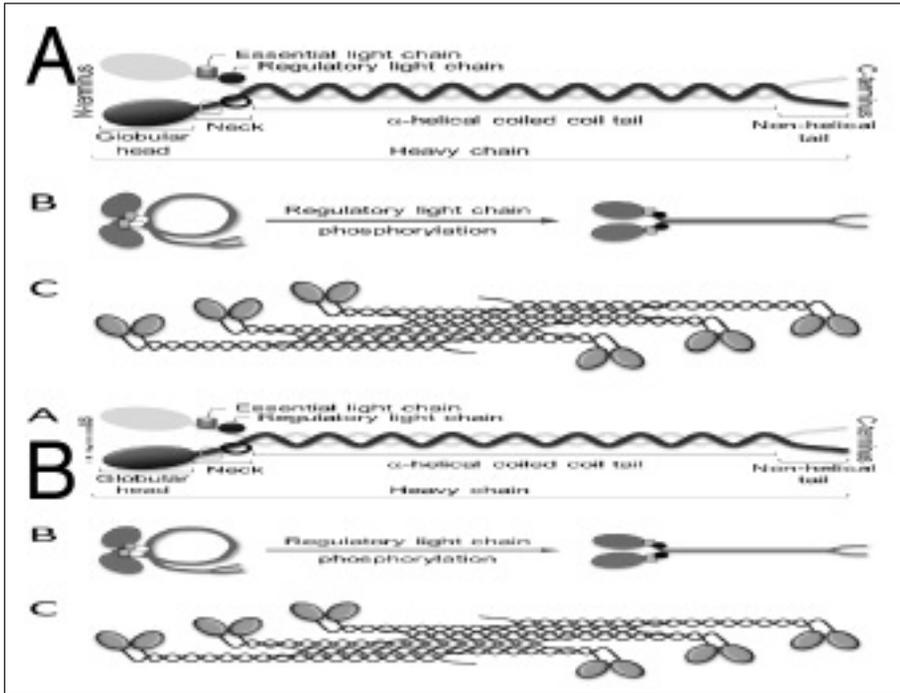


Fig. 3 - General structure of non-muscle myosin IIA. A: Non-muscle myosin II consists of two pairs of light chains and two heavy chains (NMMHC-IIAs) which form a dimer through interactions between the α -helical coiled-coil portions of the tail domains. The heavy chain globular heads contain the actin-binding regions and the enzymatic Mg^{2+} -ATPase catalytic site. The essential light chains and the regulatory light chains bind to the neck domains of the heavy chains. B: Extended non-muscle myosin II dimers assemble into functional bipolar filaments through interaction between their coiled-coil portions of the tail domain. These filaments interact with actin filaments through their head domains. Modified from reference 1.

expressed in a few normal tissues and in a large number cancer cells (59). The functional role of *ANKRD26* is still poorly known. Mice with partial inactivation of *Ankrd26* have marked hyperphagia, which results in extreme obesity, and an increase of body length due to a direct effect of ANKRD26 in the regulation of the food intake; they also develop an obesity-dependent diabetes in presence of insulin resistance in liver, skeletal muscle and brown adipose tissue, but increased insulin sensitivity in white adipose tissue (60, 61). Yet, they are not thrombocytopenic.

Thus, *ANKRD26* haploinsufficiency does not seem to result in low platelet count, at least in mouse. However, the evidence that James D. Watson is not thrombocytopenic although he carries a heterozygous deletion of about 31.5 Kb involving the last 6 exons of *ANKRD26* (Database of Genomic Variants, Variation_39047) suggests that also in humans haploinsufficiency does not lower platelet count (62). Since mutations in 5'UTR can cause either loss or gain of function, it has been therefore suggested that *ANKRD26*-RT derives from overexpression of the gene, and this hypothesis has been supported by the results of luciferase reporter

assay experiments revealing that 5'UTR mutations enhance *ANKRD26* expression (3). Thrombocytopenia of patients with *ANKRD26*-RT probably derives from reduced platelet production, since bone marrow examination in a few patients revealed clear signs of dysmegakaryopoiesis, as revealed by the presence of many small and dystrophic megakaryocytes, often with hypolobulated nuclei (2, 3).

Clinical picture and laboratory features

A thorough analysis of 78 patients from 21 families with *ANKRD26*-RT has been published recently (2).

Bleeding diathesis was usually absent or mild; when present, patients mostly referred epistaxis, gum bleedings, petechiae, ecchymosis and menorrhagia, while only in single cases life-threatening bleedings were reported. Only a few bleeding complications developed after surgical procedures or in women who gave birth both vaginally or by caesarian section.

In some patients previously diagnosed as suffering from immune thrombocytopenia (ITP), neither steroids nor splenectomy ensured a sustained increase in platelet count; however, at least four patients had an increase in their platelet count over $150 \times 10^9/L$ during infectious episodes, probably due to an increase in blood TPO levels.

The most important clinical remark of *ANKRD26*-RT patients is the association with haematological malignancies, in particular myeloid acute leukemias (AML): this observation raised the suspicion that *ANKRD26* mutations favor the appearance of hematopoietic neoplasms. This hypothesis is consistent with the report that *ANKRD26* was mutated in the first AML patient whose entire genome has been sequenced (63), although it needs a confirmation in larger case series of *ANKRD26*-RT patients.

Thrombocytopenia was moderate in the majority of patients, although in few cases the platelet count was even lower than $10 \times 10^9/L$. So far, only one subject carrying an *ANKRD26* mutation had a platelet count at the lower limit of the normal range, suggesting an almost complete penetrance of the trait.

Platelet volumes and *in vitro* aggregation are usually normal. In some patients platelets present reduced α -granule content and defective surface expression of GPIa. However, these two defects are inconstant even in single pedigrees, and, therefore, they are not valuable tools for a diagnosis suggestion. Serum TPO is about 7 fold increased compared to normal subjects and up to 3 times compared to ITP patients with a superimposable platelet count.

In the majority of patients both hemoglobin level and leukocyte count are within the normal range; however, besides a few patients with a mild anemia consistent with a hemorrhagic disorder, increased levels of hemoglobin and leucocytes were observed in a significant number of patients. To note, the last finding was reported also in *THC2* patients carrying the *MASTL* mutation (56).

Since no peculiar clinical or laboratory abnormalities have been identified so far, the diagnosis of *ANKRD26*-RT requires mutation screening.

References

1. Balduini CL, Pecci A, Savoia A. Recent advances in the understanding and management of MYH9-related inherited thrombocytopenias. *Br J Haematol.* 2011; 154: 161-74.
2. Noris P, Perrotta S, Seri M, et al. Mutations in ANKRD26 are responsible for a frequent form of inherited thrombocytopenia: analysis of 78 patients from 21 families. *Blood.* 2011; 117: 6673-80.
3. Pippucci T, Savoia A, Perrotta S, et al. Mutations in the 5' UTR of ANKRD26, the ankirin repeat domain 26 gene, cause an autosomal-dominant form of inherited thrombocytopenia, THC2. *Am J Hum Genet.* 2011; 88: 115-20.
4. The May-Hegglin/Fechtner Syndrome Consortium. Mutations in MYH9 result in the May-Hegglin anomaly, and Fechtner and Sebastian syndromes. *Nat Genet.* 2000; 26: 103-5.
5. Jackson SC, Sinclair GD, Cloutier S, Duan Z, Rand ML, Poon MC. The Montreal platelet syndrome kindred has type 2B von Willebrand disease with the VWF V1316M mutation. *Blood.* 2009; 113: 3348-51.
6. Rees DC, Iolascon A, Carella M, et al. Stomatocytic haemolysis and macrothrombocytopenia (Mediterranean stomatocytosis/macrothrombocytopenia) is the haematological presentation of phytosterolaemia. *Br J Haematol.* 2005; 130: 297-309.
7. Savoia A, Balduini CL, Savino M, et al. Autosomal dominant macrothrombocytopenia in Italy is most frequently a type of heterozygous Bernard-Soulier syndrome. *Blood.* 2001; 97: 1330-5.
8. Noris P, Pecci A, Di Bari F, et al. Application of a diagnostic algorithm for inherited thrombocytopenias to 46 consecutive patients. *Haematologica.* 2004; 89: 1219-25.
9. Balduini CL, Cattaneo M, Fabris F, et al. Inherited thrombocytopenias: proposal of a diagnostic algorithm by the Italian "Gruppo di studio delle piastrine". *Haematologica.* 2003; 88: 582-92.
10. Savoia A, Pastore A, De Rocco D, et al. Clinical and genetic aspects Bernard-Soulier syndrome: searching for genotype/phenotype correlations. *Haematologica.* 2011; 96: 417-23.
11. Noris P, Perrotta S, Bottega R, et al. Clinical and laboratory features of 103 patients from 42 Italian families with inherited thrombocytopenia derived from the monoallelic Ala156Val mutation of GPIIb/IIIa (Bolzano mutation). *Haematologica.* 2011 Sep 20 (Epub ahead of print).
12. Thon JN, Italiano JE. Platelet formation. *Semin Hematol.* 2010; 47: 220-6.
13. Ciovacco WA, Raskind WH, Kacena MA. Human phenotypes associated with GATA-1 mutations. *Gene.* 2008; 427: 1-6.
14. Balduini CL, Pecci A, Loffredo G, et al. Effects of the R216Q mutation of GATA-1 on erythropoiesis and megakaryocytopoiesis. *Thromb Haemost.* 2004; 91: 129-40.
15. Balduini A, Malara A, Balduini CL, Noris P. Megakaryocytes derived from

- patients with the classical form of Bernard-Soulier syndrome show no ability to extend proplatelets in vitro. *Platelets*. 2011; 22: 308-11.
16. Balduini A, Malara A, Pecci A, et al. Proplatelet formation in heterozygous Bernard-Soulier syndrome type Bolzano. *J Thromb Haemost*. 2009; 7: 478-84.
 17. Pecci A, Malara A, Badalucco S, et al. Megakaryocytes of patients with MYH9-related thrombocytopenia present an altered proplatelet formation. *Thromb Haemost*. 2009; 102: 90-6.
 18. Geddis AE. Congenital amegakaryocytic thrombocytopenia. *Pediatr Blood Cancer*. 2011; 57: 199-203.
 19. Castillo-Caro P, Dhanraj S, Haut P, Robertson K, Dror Y, Sharathkumar AA. Proximal radio-ulnar synostosis with bone marrow failure syndrome in an infant without a HOXA11 mutation. *J Pediatr Hematol Oncol*. 2010; 32: 479-85.
 20. Liew E, Owen C. Familial myelodysplastic syndromes: a review of the literature. *Haematologica* 2011; 96: 1536-42.
 21. Geddis AE, Balduini CL. Diagnosis of immune thrombocytopenic purpura in children. *Current Opinion Hematol*. 2007; 14: 520-5.
 22. Noris P, Klersy C, Zecca M, et al. Platelet size distinguishes between inherited macrothrombocytopenias and immune thrombocytopenia. *J Thromb Haemost*. 2009; 7: 2131-6.
 23. Pecci A, Gresele P, Klersy C, et al. Eltrombopag for the treatment of the inherited thrombocytopenia deriving from MYH9 mutations. *Blood*. 2010; 116: 5832-7.
 24. Pai SY, DeMartini D, Forino C, et al. Stem cell transplantation for the Wiskott-Aldrich syndrome: a single-center experience confirms efficacy of matched unrelated donor transplantation. *Bone Marrow Transplant*. 2006; 38: 671-9.
 25. Locatelli F, Rossi G, Balduini C. Hematopoietic stem-cell transplantation for the Bernard-Soulier syndrome. *Ann Intern Med*. 2003; 138: 79.
 26. Kelley MJ, Jawien W, Ortel TL, Korczak JF. Mutation of MYH9, encoding non-muscle myosin heavy chain A, in May-Hegglin anomaly. *Nat Genet*. 2000; 26: 106-8.
 27. Kunishima S, Kojima T, Matsushita T, et al. Mutations in the NMMHC-A gene cause autosomal dominant macrothrombocytopenia with leukocyte inclusions (May-Hegglin anomaly/Sebastian syndrome). *Blood*. 2001; 97: 1147-9.
 28. Seri M, Pecci A, Di Bari F, et al. MYH9-related disease: May-Hegglin anomaly, Sebastian syndrome, Fechtner syndrome, and Epstein syndrome are not distinct entities but represent a variable expression of a single illness. *Medicine*. 2003; 82: 203-15.
 29. Heath KE, Campos-Barros A, Toren A, et al. Non-muscle myosin heavy chain IIA mutations define a spectrum of autosomal dominant macrothrombocytopenias: May-Hegglin anomaly and Fechtner, Sebastian, Epstein, and Alport-like syndromes. *Am J Hum Genet*. 2001; 69: 1033-45.
 30. Sellers JR. Myosins: a diverse superfamily. *Biochim Biophys Acta* 2000; 1496: 3-22.

31. Niederman R, Pollard, TD. Human platelet myosin. II. In vitro assembly and structure of myosin filaments. *J Cell Biol.* 1975; 67: 72-92.
32. Ronen D, Ravid S. Myosin II tailpiece determines its paracrystal structure, filament assembly properties, and cellular localization. *J Biol Chem.* 2009; 284: 24948-57.
33. Pecci A, Canobbio I, Balduini A, et al. Pathogenetic mechanisms of haematological abnormalities of patients with MYH9 mutations. *Hum Mol Genet* 2005; 14: 3169-78.
34. Chang Y, Auradé F, Larbret F, et al. Proplatelet formation is regulated by the Rho/ROCK pathway. *Blood.* 2007; 109: 4229-36.
35. Chen Z, Naveiras O, Balduini A, et al. The May-Hegglin anomaly gene *Myh9* is a negative regulator of platelet biogenesis modulated by the Rho-ROCK pathway. *Blood.* 2007; 100: 171-9.
36. Balduini A, Pallotta I, Malara A, et al. Adhesive receptors, extracellular proteins, and myosin IIA orchestrate proplatelet formation by human megakaryocytes. *J Thromb Haemost.* 2008; 6: 1900-7.
37. Eckly A, Strassel C, Freund M, et al. Abnormal megakaryocyte morphology and proplatelet formation in mice with megakaryocyte-restricted MYH9 inactivation. *Blood.* 2009; 113: 3182-9.
38. Larson MK, Watson SP. A product of their environment: do megakaryocytes rely on extracellular cues for proplatelet formation? *Platelets.* 2006; 17: 435-40.
39. Conti MA, Even-Ram S, Liu C, Yamada KM, Adelstein RS. Defects in cell adhesion and the visceral endoderm following ablation of non-muscle myosin heavy chain II-A in mice. *J Biol Chem.* 2004; 279: 41263-6.
40. Zhang Y, Conti MA, Malide D, et al. Mouse models of MYH9-related disease: mutations in non-muscle myosin II-A. *Blood.* 2011 Sep 8 (Epub ahead of print).
41. Pecci A, Bozzi V, Panza E, et al. Mutations responsible for MYH9-related thrombocytopenia impair SDF-1-driven migration of megakaryoblastic cells. *Thromb Haemost.* 2011; 106: 693-704.
42. Savoia A, De Rocco D, Panza E, et al. Heavy chain myosin 9-related disease (MYH9 -RD): neutrophil inclusions of myosin-9 as a pathognomonic sign of the disorder. *Thromb Haemost.* 2010; 103: 826-32.
43. Pecci A, Panza E, Pujol-Moix N, et al. Position of non-muscle myosin heavy chain IIA (NMMHC-IIA) mutations predicts the natural history of MYH9-related disease. *Hum Mutat.* 2008; 29: 409-17.
44. Dong F, Li S, Pujol-Moix N, et al. Genotype-phenotype correlation in MYH9-related thrombocytopenia. *Br J Haematol.* 2005; 130: 620-7.
45. Ghiggeri GM, Caridi G, Magrini U, et al. Genetics, clinical and pathological features of glomerulonephritis associated with mutations of non-muscle myosin IIA (Fechtner syndrome). *Am J Kidney Dis.* 2003; 41: 95-104.
46. Pecci A, Granata A, Fiore CE, Balduini CL. Renin-angiotensin system blockade is effective in reducing proteinuria of patients with progressive nephropathy caused by MYH9 mutations (Fechtner-Epstein syndrome). *Nephrol Dial Transplant.* 2008; 23: 2690-2.

47. Sekine T, Konno M, Sasaki S, et al. Patients with Epstein-Fechtner syndromes owing to MYH9 R702 mutations develop progressive proteinuric renal disease. *Kidney Int.* 2010; 78: 207-14.
48. Han KH, Lee H, Kang HG, et al. Renal manifestations of patients with MYH9-related disorders. *Pediatr Nephrol.* 2011; 26: 549-55.
49. De Rocco D, Pujol-Moix N, Pecci A, et al. Identification of the first duplication in MYH9-related disease: A hot spot for hot unequal crossing-over within exon 24 of the MYH9 gene. *Eur J Med Genet.* 2009; 52: 191-4.
50. Pecci A, Noris P, Invernizzi R, et al. Immunocytochemistry for the heavy chain of the non-muscle myosin IIA as a diagnostic tool for MYH9-related disorders. *Br J Haematol.* 2002; 117: 164-7.
51. Kunishima S, Matsushita T, Yoshihara T, et al. First description of somatic mosaicism in MYH9 disorders. *Br J Haematol.* 2005; 128: 360-5.
52. Pecci A, Panza E, De Rocco D, et al. MYH9 related disease: four novel mutations of the tail domain of myosin-9 correlating with a mild clinical phenotype. *Eur J Haematol.* 2010; 84: 291-7.
53. Canobbio I, Noris P, Pecci A, Balduini A, Balduini CL, Torti M. Altered cytoskeleton organization in platelets from patients with MYH9-related disease. *J Thromb Haemost.* 2005; 3: 1026-35.
54. Liunbruno G, Bennardello F, Lattanzio A, Piccoli P, Rossetti G. Recommendations for the transfusion of plasma and platelets. *Blood Transfus.* 2009; 7: 132-50.
55. Savoia A, Del Vecchio M, Totaro A, et al. An autosomal dominant thrombocytopenia gene maps to chromosomal region 10p. *Am J Hum Genet.* 1999; 65: 1401-5.
56. Drachman JG, Jarvik GP, Mehaffey MG. Autosomal dominant thrombocytopenia: incomplete megakaryocyte differentiation and linkage to human chromosome 10. *Blood.* 2000; 96: 118-25.
57. Gandhi MJ, Cummings CL, Drachman JG. FLJ14813 missense mutation: a candidate for autosomal dominant thrombocytopenia on human chromosome 10. *Hum Hered.* 2003; 55: 66-70.
58. Punzo F, Mientjes EJ, Rohe CF, et al. A mutation in the acyl-coenzyme A binding domain containing protein 5 gene (ACBD5) identified in autosomal dominant thrombocytopenia. *J Thromb Haemost.* 2010; 8: 2085-7.
59. Hahn Y, Bera TK, Pastan IH, Lee B. Duplication and extensive remodeling shaped POTE family genes encoding proteins containing ankyrin repeat and coiled coil domains. *Gene.* 2006; 66: 238-45.
60. Bera TK, Liu XF, Yamada M, et al. A model for obesity and gigantism due to disruption of the Ankrd26 gene. *Proc Natl Acad Sci USA.* 2008; 105: 270-5.
61. Raciti GA, Bera TK, Gavrilova O, Pastan I. Partial inactivation of Ankrd26 causes diabetes with enhanced insulin responsiveness of adipose tissue in mice. *Diabetologia.* 2011; 54: 2911-22.
62. Wheeler DA, Srinivasan M, Egholm M, et al. The complete genome of an individual by massively parallel DNA sequencing. *Nature.* 2008; 452: 872-6.
63. Mardis E, Ding L, Dooling DJ, Larson DE, et al. Recurrent mutations found

- by sequencing an acute myeloid leukemia genome. *N Engl J Med.* 2009; 361: 1058-66.
64. Mattina T, Perrotta CS, Grossfeld P. Jacobsen syndrome. *Orphanet J Rare Dis.* 2009; 4: 9.
65. Gunay-Aygun M, Zivony-Elboum Y, Gumruk F, et al. Gray platelet syndrome: natural history of a large patient cohort and locus assignment to chromosome 3p. *Blood.* 2010; 116: 4990-5001.
66. Kahr WH, Hinckley J, Li L, et al. Mutations in NBEAL2, encoding a BEACH protein, cause gray platelet syndrome. *Nat Genet.* 2011; 43: 738-40.
67. Gunay-Aygun M, Falik-Zaccari TC, Vilboux T, et al. NBEAL2 is mutated in gray platelet syndrome and is required for biogenesis of platelet α -granules. *Nat Genet.* 2011; 43: 732-4.
68. Albers CA, Cvejic A, Favier R, et al. Exome sequencing identifies NBEAL2 as the causative gene for gray platelet syndrome. *Nat Genet.* 2011; 43: 735-7.
69. Hamilton A, Ozelo M, Leggo J, et al. Frequency of platelet type versus type 2B von Willebrand disease. An international registry-based study. *Thromb Haemost.* 2011; 105: 501-8.
70. Gresele P, Falcinelli E, Giannini S, et al. Dominant inheritance of a novel integrin beta3 mutation associated with a hereditary macrothrombocytopenia and platelet dysfunction in two Italian families. *Haematologica.* 2009; 94: 663-9.
71. Kunishima S, Kobayashi R, Itoh TJ, Hamaguchi M, Saito H. Mutation of the beta1-tubulin gene associated with congenital macrothrombocytopenia affecting microtubule assembly. *Blood.* 2009; 113: 458-61.
72. Nurden P, Debili N, Coupry I, et al. Thrombocytopenia resulting from mutations in filamin A can be expressed as an isolated syndrome. *Blood.* 2011; Sep 29 (Epub ahead of print).
73. Houeijeh A, Andrieux J, Saugier-Verber P, et al. Thrombocytopenia-absent radius (TAR) syndrome: a clinical genetic series of 14 further cases. Impact of the associated 1q21.1 deletion on the genetic counselling. *Eur J Med Genet.* 2011; 54: e471-477.
74. Morison IM, Cramer Bordé EM, Cheesman EJ, et al. A mutation of human cytochrome c enhances the intrinsic apoptotic pathway but causes only thrombocytopenia. *Nat Genet.* 2008; 40: 387-9.
75. Notarangelo LD, Miao CH, Ochs HD. Wiskott-Aldrich syndrome. *Curr Opin Hematol.* 2008; 15: 30-6.

Genomica e malattie autoimmuni

Eleonora Gambineri

Dipartimento della Salute della Donna e del Bambino, Università di Firenze,
Ospedale Pediatrico "Anna Meyer"

Il ruolo principale del sistema immunitario è di mantenere l'omeostasi. Questo significa l'abilità a riconoscere agenti patogeni esterni o tumorali e ad instaurare una risposta immune appropriata per difendere l'organismo. Allo stesso tempo il sistema immunitario riconosce anche antigeni *self* verso i quali le cellule non reagiscono, ma instaurano uno stato di tolleranza. Un'alterazione di queste due funzioni fondamentali provoca rispettivamente uno stato di immunodeficienza o di autoimmunità, sempre considerati due condizioni opposte. D'altronde, un'analisi recente più attenta delle caratteristiche cliniche associate alle malattie da immunodeficienza primitiva (IDP) e dei meccanismi molecolari e cellulari coinvolti nelle immunodeficienze e nell'autoimmunità ha mostrato che in realtà queste condizioni sono tra di loro correlate. In particolare, lo studio delle manifestazioni autoimmuni associate alle IPD ha chiarito la fisiopatologia molecolare delle reazioni autoimmuni e delle risposte infiammatorie evidenziando come il mantenimento della tolleranza immunologica sia una funzione importante quanto la difesa da microorganismi patogeni. Inoltre, nonostante che la maggior parte dei disordini autoimmuni sia di origine multifattoriale, lo studio di disordini monogenici della tolleranza centrale e periferica come IPEX (Immunodisregolazione, poliendocrinopatia, enteropatia, X-linked) e APECED (Poliendocrinopatia autoimmune con candidosi e displasia ectodermica), dove immunodeficit e autoimmunità coesistono, ha mostrato la rilevanza clinica dello sviluppo e della funzione di quelle cellule del sistema immune che sono coinvolte nel mantenimento dell'omeostasi immunitaria.

IPEX

La sindrome IPEX è un raro disordine del sistema immunitario associato al cromosoma X, caratterizzato da fenomeni autoimmuni multisistemici e causata da mutazioni del gene *FOXP3*, localizzato sul cromosoma X e codificante l'omonima proteina FOXP3 che appartiene ad una famiglia di regolatori trascrizionali che possono agire da repressori o da attivatori e risultano coinvolti in processi importanti come lo sviluppo del timo, il differenziamento cellulare e la regolazione della

funzione linfocitaria. In particolare, è stato osservato che *FOXP3* è fondamentale per lo sviluppo ed il mantenimento dei linfociti T regolatori CD4+CD25+ (Treg), un sottotipo di linfociti regolatori di derivazione timica. *FOXP3* si localizza nel nucleo, dove controlla l'espressione di diversi geni al fine di mantenere l'omeostasi immunologica. Inoltre *FOXP3* può essere temporaneamente espresso anche dalle cellule T effettrici suggerendone un ruolo nel mantenimento dell'omeostasi immunitaria oltre all'ontogenesi delle Treg. *FOXP3* e le cellule Treg sono quindi un ruolo cruciale nel mantenimento della tolleranza periferica.

La sindrome IPEX si manifesta fin dalla nascita o nella prima infanzia in individui maschi; raramente sono stati descritti casi ad esordio tardivo.

Le principali caratteristiche cliniche identificate sono:

- *enteropatia*: è la manifestazione più comune e può mostrarsi come diarrea acquosa o accompagnata da perdite ematiche e mucose. Si presenta spesso fino dalla nascita e comporta mal assorbimento e ritardo della crescita. All'analisi della biopsia intestinale si osserva, nella maggior parte dei casi, una severa atrofia dei villi con infiltrato linfocitario nella sottomucosa e/o nella lamina propria;
- *endocrinopatie*: il diabete di tipo I ad esordio precoce, spesso neonatale, è la forma più comune, ma possono essere osservate anche anomalie della tiroide di origine autoimmune da ipotiroidismo clinico ad alterazioni dei livelli ormonali senza manifestazioni cliniche evidenti;
- *malattie cutanee*: principalmente eczema, ma in letteratura sono state riportate anche altre manifestazioni come eritrodermia, dermatite psoriasiforme, orticaria, pemfigoide nodulare e alopecia.

A questa triade di sintomi si possono associare anche altre manifestazioni autoimmuni, come citopenie autoimmuni (anemia emolitica, piastrinopenia, neutropenia), nefropatia, epatite colestatica, vasculite, artrite, linfadenopatia e epatosplenomegalia. Sono riportate anche anomalie neurologiche come ritardo di sviluppo psico-motorio e crisi epilettiche. Tali sintomi clinici tuttavia si manifestano con minor frequenza e difficilmente sono presenti all'esordio, ma tendono a comparire nel tempo. È stato osservato anche un aumento del rischio di infezioni ricorrenti probabilmente correlato alla disregolazione del sistema immunitario o conseguenza dei trattamenti terapeutici con immunosoppressori. La gravità delle infezioni è variabile e si possono riscontrare infezioni moderate del sistema respiratorio o del tratto gastrointestinale o anche infezioni gravi come sepsi, peritonite, polmonite e meningite i cui agenti patogeni più comuni sono le specie *Enterococco* e *Stafilococco*, *Citomegalovirus* e *Candida* spp.

Se non trattata la malattia è rapidamente fatale e le cause principali di morte sono emorragie, diarrea intrattabile, sepsi e complicanze del diabete.

Le femmine portatrici eterozigoti risultano in salute, sebbene il processo di inattivazione del cromosoma X avvenga casualmente. Dalle analisi di inattivazione del cromosoma X effettuate in una femmina portatrice è stato dimostrato che gli alleli normale e mutato di *FOXP3* sono espressi ugualmente nei linfociti T circolanti, mascherando forse la patologia tramite un meccanismo di compensazione. I pazienti con IPEX non mostrano caratteristiche laboratoristiche tipiche. In particolare, non sono riportate né grossolane alterazioni quantitative, né alterazioni

funzionali. In genere si osserva un aumento variabile di IgE sieriche e di eosinofili, mentre i livelli delle altre classi di immunoglobuline sono nella norma. La maggior parte dei pazienti produce autoanticorpi verso vari organi, quali tiroide, isole pancreatiche, piastrine, eritrociti, muscolo liscio ed intestino. In alcuni casi è stata descritta la presenza di autoanticorpi contro un antigene di 75 kDa specifico del rene e dell'intestino (AIE-75).

I linfociti totali e dei loro sottotipi non mostrano grossolane anomalie, ma talvolta è riportato un aumento di indici di attivazione linfocitaria (espressione dei marcatori HLA-DR e CD25). La proliferazione linfocitaria in vitro in risposta a mitogeni e ad antigeni specifici è generalmente nella norma. D'altronde è stato dimostrato che i pazienti con mutazione di *FOXP3* presentano un'alterata funzione delle cellule Treg e nella capacità delle cellule T effettrici a produrre IL-2 e IFN- γ .

La diagnosi di IPEX dovrebbe essere considerata in base alle caratteristiche cliniche del paziente, in particolare in presenza di diarrea intrattabile, atrofia dei villi intestinali e ritardo della crescita ed ulteriormente indicata in presenza di manifestazioni cutanee, diabete insulino-dipendente ad esordio precoce e/o anomalie tiroidee.

In ogni caso la diagnosi di IPEX deve essere confermata dall'analisi genetica di mutazione in *FOXP3* o dall'assenza di linfociti T regolatori CD4⁺CD25⁺.

Una terapia aggressiva precoce è di estrema importanza. La scelta primaria ad oggi per il trattamento è solitamente l'uso di agenti immunosoppressivi, come ciclosporina A o tacrolimus (FK 506), spesso in combinazione con farmaci steroidei che agiscono sulle cellule T attivate. Inizialmente la maggior parte dei pazienti così trattati risponde positivamente e in alcuni casi la remissione dei sintomi può durare a lungo, ma l'immunosoppressione cronica a lungo termine può portare tossicità farmacologica ed aumento del rischio di infezioni. Recentemente sembra molto promettente l'impiego della rapamicina, un farmaco immunosoppressore che agisce con meccanismo diverso rispetto agli inibitori della calcineurina e sembra avere un vantaggio selettivo sulle Treg favorendone l'espansione e la sopravvivenza. Attualmente, il trapianto con cellule staminali ematopoietiche è l'unica cura effettiva, soprattutto se effettuato precocemente.

In genere IPEX è conosciuta come una malattia grave a decorso clinico rapidamente fatale se non trattata. Tuttavia in letteratura sono descritti casi clinici affetti da forme incomplete o atipiche di malattia senza la completa espressione della triade di sintomi. Inoltre sono stati osservati pazienti che mostravano un fenotipo caratteristico di IPEX senza però alcuna aberrazione genica di *FOXP3* (IPEX-like).

Sindromi IPEX-like (Deficit di CD25 e STAT5b)

La sindrome da deficit di CD25, la catena α del recettore dell'interleuchina-2 (IL-2), descritta al momento in due pazienti non consanguinei, è considerata una malattia IPEX-like. Dopo attivazione, le cellule T esprimono la catena α -(CD25), che aumenta l'affinità del recettore per IL-2 e permettono l'attivazione dei segnali di trasduzione intracellulare fondamentali per l'attivazione della risposta immunologica. Le cellule Treg esprimono invece costitutivamente CD25 in modo che

siano in grado di rispondere anche basse concentrazioni di IL-2 in ogni momento. Questa capacità di rispondere a basse concentrazioni di IL-2 è un fattore critico per il mantenimento di espressione FOXP3 nelle Treg.

Le caratteristiche cliniche della malattia sono simili sia ad IPEX che ad una immunodeficienza combinata grave (SCID). Come in IPEX, i pazienti sviluppano una grave diarrea cronica con atrofia dei villi nella prima infanzia ed uno dei due anche diabete insulino-dipendente ad esordio precoce ed eczema. In seguito possono comparire altri segni di disregolazione immunologica come auto-anticorpi, epatosplenomegalia, linfadenopatia, e infiltrati linfocitari in vari organi (intestino, fegato, ecc). A differenza dei pazienti con mutazioni di FOXP3, livelli sierici di IgE sono normali o solo lievemente elevati.

In aggiunta alle caratteristiche autoimmuni, i pazienti con deficit di CD25 hanno anche complicanze infettive suggestive di un difetto più ampio dell'immunità cellulare. Fra le infezioni più comuni sono state descritte la polmonite da CMV, il mugugno persistente, l'esofagite da candida, la gastroenterite cronica e l'infezione da EBV.

In entrambi i casi descritti, l'ereditarietà è autosomica recessiva e le mutazioni descritte portano alla mancata espressione di CD25 sui linfociti T. La citometria a flusso è quindi un ottimo metodo di screening iniziale per valutare i pazienti sospettati di avere deficit di CD25. Il sequenziamento del gene IL2RA (CD25) è comunque fondamentale per confermare la diagnosi in tutti i casi.

La terapia di supporto è fondamentale per trattare i sintomi della malattia così come l'immunosoppressione, analogamente ad IPEX. Anche in questo caso, considerare anche le somiglianze cliniche con la SCID, il trapianto di midollo osseo è l'unica cura risolutiva.

Il deficit del fattore di trascrizione STAT5b è stato descritto nel 2003 in pazienti con deficit significativo di crescita staturale-ponderale ed autoimmunità (29). Come altri fattori di trascrizione STAT, STAT5b gioca un ruolo essenziale nella trasmissione del segnale intracellulare di citochine e fattori di crescita, in particolare in risposta a IL-2, IL-15, e all'ormone della crescita (GH). Come risultato, la carenza di STAT5b si traduce in una combinazione di mancata crescita (da -3,0 a -9,9 deviazioni standard sotto la media per l'età per altezza) e immunodeficienza con autoimmunità. Oltre agli effetti sulla crescita, le caratteristiche cliniche più consistenti sono la polmonite linfocitica cronica interstiziale (LIP) nel 90% dei pazienti ed un rash (spesso eczematoso) che è presente nel 80% dei pazienti. La polmonite causa insufficienza respiratoria e morte in alcuni individui affetti. Spesso sono stati identificati anche autoanticorpi organo-specifici (circa il 50% dei pazienti).

Poiché STAT5b è essenziale per la trasmissione intracellulare del segnale di IL-2 e IL-15, fattori di crescita chiave per i linfociti T e le cellule NK rispettivamente, i pazienti hanno tipicamente una moderata linfopenia e spesso soffrono di infezioni virali ricorrenti da agenti patogeni come il citomegalovirus (CMV). Dal momento che l'IL-2 è essenziale per la espressione di FOXP3 nelle cellule Tregs, i pazienti con deficit STAT5b mostrano una diminuita espressione di FOXP3 ed una funzione difettosa delle Treg.

STAT5b è uno dei 7 membri della famiglia di fattori di trascrizione STAT. Queste proteine che legano il DNA giocano un ruolo chiave nella risposta cellulare ad una varietà di citochine e fattori di crescita. In circostanze normali, STAT5b è presente nel citoplasma in forma inattiva, non fosforilata. Viene fosforilato in seguito ad attivazione dei recettori di citochine e sottoforma di dimero si localizza nel nucleo dove, in seguito a legame con il DNA, regola la trascrizione di una varietà di geni bersaglio compreso *FOXP3*. Tutte le mutazioni identificate *STAT5b* fino ad oggi sono in forma autosomica recessiva e distruggono la funzione di entrambi gli alleli di STAT5b.

La diagnosi può essere suggerita dal fenotipo clinico (deficit grave di crescita, deficit immunitario, e disregolazione immunitaria) e da esami di laboratorio di routine (linfopenia, basse cellule NK, ecc). Gli esami endocrinologici mostrano tipicamente normali livelli plasmatici di ormone della crescita, e bassi livelli di IGF-1 e IGFBP-3 con aumento della prolattina. La diagnosi definitiva è comunque fatta mediante sequenziamento del gene *STAT5b*.

La maggior parte dei trattamenti per il deficit di STAT5b si sono concentrati sulla terapia sintomatica e la profilassi contro le infezioni. In alcuni casi sono stati utilizzati anche regimi di immunosoppressione anche se non sono stati riportati schemi di trattamento efficaci per la grave malattia polmonare che colpisce la maggior parte dei pazienti. Ad oggi, non sono stati segnalati casi di trapianto di midollo osseo, anche se in teoria potrebbe essere efficace per correggere l'immunodeficienza e la disregolazione immunitaria in genere associata alla carenza STAT5b. Mentre la correzione di queste funzioni potrebbe ridurre la morbilità e la mortalità, il trapianto non migliorerebbe la grave insufficienza di crescita e questa è un fattore che va considerato prima di intraprendere un eventuale percorso trapiantologico.

APECED

APECED è un raro disordine genetico della regolazione immunitaria ereditato in modo autosomico recessivo e causato da mutazione del gene *AIRE* (*autoimmune regulator*) coinvolto nella tolleranza immune. APECED è caratterizzata dalla variabile combinazione di candidiasi muco-cutanea cronica (CMC), poliendocrinopatia e/o epatite e distrofia dello smalto dentario e delle unghie. In realtà la triade di CMC, ipoparatiroidismo e insufficienza surrenalica costituisce la presentazione più tipica della sindrome e la diagnosi clinica si basa sulla presenza di almeno due dei sintomi riportati. La frequenza di questa sindrome nella popolazione generale non sembra molto elevata, è stata tuttavia notata un'alta incidenza in popolazioni selezionate come i finlandesi (1: 25000), i sardi (1:14400) e gli ebrei iraniani (1:9000).

Il gene *AIRE* è localizzato sul cromosoma 21q22.3 e codifica per un regolatore della trascrizione. L'espressione di *AIRE* è ristretta alle cellule epiteliali della zona midollare del timo (mTECs) e alle cellule dendritiche timiche che svolgono un ruolo cruciale per lo sviluppo la tolleranza immunologica centrale eliminando le cellule T reattive contro antigeni *self* e favorendo l'ontogenesi delle cellule

Treg (CD4+25+FOXP3+). Sono state identificate più di 60 mutazioni in *AIRE*, distribuite per tutta la lunghezza del gene e particolari mutazioni tendono a segregare nelle diverse popolazioni dove APECED è più frequente.

La presentazione clinica di questa malattia è molto eterogenea e può essere caratterizzata dalla presenza di un solo sintomo (in genere CMC) oppure dalla combinazione di diversi tipi di manifestazioni endocrine e disordini del sistema ectodermico. Nella maggior parte dei pazienti il sintomo all'esordio è la CMC che solitamente è resistente alla terapia anti-fungina tradizionale e tende a ricorrere frequentemente soprattutto nell'infanzia, anche se poi l'incidenza di ricadute diminuisce nel tempo.

L'ipoparatiroidismo è il secondo sintomo più frequente che compare in genere nel secondo anno di vita, mentre l'insufficienza surrenalica si manifesta dopo il quarto anno. L'ipoparatiroidismo colpisce circa l'85% degli affetti ed in particolare le femmine. Anche l'ipogonadismo, che si manifesta dopo i 13 anni, colpisce prevalentemente il sesso femminile (60%) rispetto ai maschi (14%) mentre le altre endocrinopatie come l'insufficienza adrenocorticale, presente nel 70% dei pazienti, il diabete insulino-dipendente e l'ipotiroidismo non sembrano avere una prevalenza di sesso.

Inoltre i pazienti APECED possono manifestare una serie di altri sintomi autoimmuni a carico del tratto gastrointestinale come l'anemia perniciosa per deficit di vitamina B12 o malassorbimento intestinale con esordio variabile nell'arco della vita. È importante sottolineare che il malassorbimento e la steatorrea sono sintomi abbastanza frequenti e possono anche caratterizzare l'esordio di malattia.

Anche la cute e altri tessuti di derivazione ectodermica possono essere coinvolti: l'alopecia interessa il 29-36% dei pazienti e spesso esordisce precocemente mentre la vitiligine è descritta nel 13-26% dei casi. Le alterazioni ungueali sono solitamente da riferirsi alla displasia ectodermica anche se a volte l'esame colturale di campioni prelevati dalle unghie ha mostrato la presenza di candida che quindi potrebbe essere in parte responsabile del quadro.

Da una valutazione dell'espressività clinica e del decorso della malattia si è notato che esistono molte differenze fra un paziente e l'altro anche nell'ambito di pazienti di una stessa famiglia, è pertanto difficile stabilire una precisa correlazione fra genotipo e fenotipo. Al contrario, sembra che sia presente una diversa predisposizione genetica, a seconda della popolazione, a sviluppare i vari sintomi caratteristici di APECED, per esempio difficilmente pazienti ebrei iraniani sviluppano CMC che notoriamente è il sintomo cardine della malattia.

L'immunodeficienza di cui questi pazienti sono affetti tende a manifestarsi localmente a carico di cute e mucose dove solitamente si localizza la candidosi, anche se talvolta i pazienti possono sviluppare forme di infezione polmonare.

Una diagnosi clinica di APECED può essere fatta in individui con due delle tre caratteristiche di base della malattia (ipoparatiroidismo, insufficienza surrenalica, e candidosi mucocutanea). La presenza di autoanticorpi organo-specifici può essere anche suggestiva, ma una diagnosi definitiva è tipicamente fatta attraverso il sequenziamento del gene *AIRE*. In quasi tutti i casi, la malattia viene ereditata in modo autosomico recessivo.

La condotta terapeutica più comune per APECED comprende la terapia sintomatica (es. sostituzione ormonale per endocrinopatie e antifungini per il trattamento candidiasi mucocutanea). L'immunosoppressione non viene utilizzata di routine anche se varie combinazioni di ciclosporina, tacrolimus, metotressato, azatioprina e steroidi sembrano efficaci in caso di grave autoimmunità renale, epatica e malattie intestinali. Lo sviluppo di autoanticorpi patogenetici in molti casi suggerisce che la terapia di deplezione delle cellule B (cioè con anti-CD20 anticorpo monoclonale) può essere utile. Ci sono poche informazioni in letteratura sia su modelli animali che nell'uomo per quanto riguarda l'efficacia del trapianto di midollo osseo per APECED. Dato che la carenza di AIRE coinvolge principalmente la funzione di cellule non emopoietiche, non è chiaro se il trapianto fornirebbe un significativo beneficio clinico dal momento che le cellule T del donatore sarebbero soggette agli stessi difetti di selezione timica tipici della malattia.

Conclusioni

L'identificazione di difetti genetici in una varietà di sindromi di disregolazione del sistema immunitario sta fornendo informazioni importanti sui meccanismi per comprendere meglio le basi fisiopatologiche di altre malattie autoimmuni più comuni dal momento che l'alterazione di due punti cruciali della regolazione immunitaria (tolleranza centrale e periferica) è probabilmente alla base di tutte le forme cliniche di disregolazione immunitaria ad eziopatogenesi multifattoriale. È probabile che queste nuove conoscenze possano offrire potenziali bersagli per lo sviluppo di terapie allo scopo di gestire efficacemente le malattie da autoimmunità in generale.

Bibliografia

1. Gambineri E, Perroni L, Passerini L, et al. Clinical and molecular profile of a new series of IPEX patients: inconsistent correlation between FOXP3 protein expression and disease severity. *J Allergy Clin Immunol.* 2008; 122 (6): 1105-12.
2. Gambineri E, Torgerson TR. Genetic disorders with immune dysregulation. *Cell Mol Life Sci.* 2012; 69 (1): 49-58.

