



Collegio Ghislieri

Centro per la Comunicazione e la Ricerca

Progetto "Progressi in Biologia e Medicina"

6° Corso di formazione avanzata

Predizione molecolare e bioinformatica in ematologia/oncologia

16 - 20 aprile 2007, Collegio Ghislieri, Pavia

A cura di Carlo Bernasconi

6° Corso di formazione avanzata

**Predizione molecolare e bioinformatica
in ematologia/oncologia**



Collegio Ghislieri

Centro per la Comunicazione e la Ricerca

Progetto “Progressi in Biologia e Medicina”

6° Corso di formazione avanzata

**Predizione molecolare e bioinformatica
in ematologia/oncologia**

16-20 aprile 2007, Collegio Ghislieri, Pavia

A cura di Carlo Bernasconi

© Copyright 2007 

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

Citogenetica molecolare

1. Metodiche innovative di citogenetica molecolare » 3
*C. Matteucci, R. La Starza, G. Barba, C. Crescenzi, S. Romoli,
P. Gorello, C. Mecucci*
2. L'importanza clinica della FISH nelle leucemie » 9
Maria Christina Cox
3. Nuove anomalie citogenetiche nelle malattie
mieloproliferative croniche » 15
*C. Matteucci, R. La Starza, G. Barba, C. Crescenzi, S. Romoli,
P. Gorello, C. Mecucci*
4. Significato delle delezioni genomiche associate a traslocazioni
cromosomiche nelle leucemie » 22
Giorgina Specchia, Francesco Albano
5. Inquadramento prognostico delle SMD » 28
*Paolo Bernasconi, Marina Boni, Paola Maria Cavigliano,
Silvia Calatroni, Ilaria Giardini, Barbara Rocca, Rita Zappatore,
Irene Dambruoso, Marilena Caresana*
6. L'apporto della FISH alla prognosi delle LLC » 40
Antonio Cuneo, Francesco Cavazzini, Gian Matteo Rigolin
7. Definizione delle anomalie cromosomiche nel mieloma multiplo » 47
*Nicoletta Testoni, Simona Luatti, Giulia Marzocchi, Chiara Nicci,
Carmen Baldazzi, Monica Stacchini*
8. Necessità di un percorso informatico alla conoscenza biologica .. » 53
Sergio Ferrari

Espressione genica e mutazioni

9. Analisi di geni singoli con la PCR quantitativa:
 espressione e monitoraggio di BCR-ABL nelle leucemie
 Ph1 positive » 61
Fabrizio Pane
10. Analisi in parallelo dell'espressione genica:
 apporto dei DNA microarray in oncologia » 71
Sergio Ferrari
11. Gene expression profiling and mutations in acute myeloid
 leukemia prognostication » 76
*Pier Paolo Piccaluga, Michela Rondoni, Stefania Paolini,
 Claudio Laterza, Cristina Papayannidis, Lea Verlicchi,
 Francesco De Rosa, Giovanni Martinelli*
12. Rilevanza dell'espressione di WT1 nelle sindromi
 mielodisplastiche e nelle LAM » 82
Daniela Cilloni
13. Ricombinazione mitotica e fenotipo: la mutazione JAK2
 nelle malattie mieloproliferative croniche Ph-negative » 85
*Davide Rossi, Daniela Capello, Paola Riccomagno,
 Monia Lunghi, Gianluca Gaidano*
14. Espressione genica e classificazione dei linfomi » 94
*Davide Rossi, Silvia Franceschetti, Michaela Cerri, Silvia Rasi,
 Valeria Spina, Paola Riccomagno, Annarita Conconi,
 Daniela Capello, Gianluca Gaidano*
15. Classificazione molecolare del mieloma multiplo » 99
Carolina Terragna, Michele Cavo
16. Profilo dell'espressione genica nel tumore della mammella: » 104
 rilevanza diagnostica e prognostica
Giovanna Chiorino
17. La cellula di origine della leucemia linfatica cronica » 107
Manlio Ferrarini

Sequenziamento genico, polimorfismi, microRNA

18. Dal sequenziamento del DNA alla post-genomica » 114
Diego di Bernardo

19. Raggruppamenti di famiglie geniche in base alle caratteristiche funzionali » 119
Carlo Sala
20. Significato e classificazione dei polimorfismi genici » 125
Roberto Giorda
21. SNP (*Single Nucleotide Polimorfism*): scoperta e applicazioni » 132
Manuela Sironi
22. Il polimorfismo genico come fattore predisponente allo sviluppo di leucemie acute e sindromi mielodisplastiche » 138
Maria Teresa Voso, Alessandra Scardocci, Francesco D'Alò, Giuseppe Leone
23. I microRNA: una nuova classe di regolatori dell'espressione genica con un ruolo nel cancro » 144
Giuseppe Biamonti
24. I microRNA nella diagnosi e progressione della LLC » 150
Massimo Negrini

Farmacogenomica

25. La conoscenza delle vie di trasduzione del segnale come guida per lo sviluppo di nuove terapie molecolari » 159
Giuseppe Saglio, Daniela Cilloni
26. Leucemia mieloide cronica: un modello di terapia molecolare » 164
Carlo Gambacorti-Passerini
27. New targeted therapy for acute myeloid leukemia » 175
Giovanni Martinelli, Stefania Paolini, Michela Rondoni, Claudio Laterza, Cristina Papayannidis, Lea Verlicchi, Francesco De Rosa, Simona Soverini, Ilaria Iacobucci, Pier Paolo Piccaluga
28. DNA methylation as therapeutic target in myelodysplastic syndromes » 176
Valeria Santini
29. Inibitori del proteosoma nella terapia del mieloma multiplo » 183
Giampaolo Merlini

30. Il sistema telomero/telomerasi: un nuovo bersaglio in ematologia/oncologia	»	191
<i>Paolo Bernasconi, Marina Boni, Paola Maria Cavigliano, Silvia Calatroni, Ilaria Giardini, Barbara Rocca, Rita Zappatore, Irene Dambroso, Marilena Caresana</i>		

31. Ripristino dell'apoptosi mediata da p53: ruolo degli antagonisti di MDM2	»	202
<i>Gian Matteo Rigolin</i>		

Elementi di bioinformatica

32. Informatica biomedica: lo scenario internazionale	»	211
<i>Riccardo Bellazzi</i>		

33. Banche dati e recupero delle informazioni	»	223
<i>Marco Masseroli</i>		

34. Modelli per la predizione della struttura genica e scoperta di siti di splicing	»	232
<i>Paolo Magni</i>		

35. Analisi di dati di microarrays di espressione: predizione di funzione, studio di pathways e reti di regolazione	»	237
<i>Riccardo Bellazzi, Lucia Sacchi</i>		

36. Analisi di dati di spettrometria di massa finalizzata alla scoperta di biomarcatori	»	249
<i>Paolo Magni</i>		

Prefazione

Le nuove conoscenze di genomica e proteomica hanno reso la bioinformatica uno strumento indispensabile per la predizione molecolare in ematologia e in oncologia. Si tratta di uno strumento completamente nuovo, al quale il medico ed il biologo si stanno avvicinando ora, comprendendone le grandi potenzialità.

Questo avvicinamento deve essere graduale, partendo da argomenti che sono più familiari all'ambito biomedico, per giungere poi a nozioni di ordine più propriamente matematico e informatico.

Nello svolgimento del presente corso verranno quindi dapprima trattate le più recenti acquisizioni di citogenetica molecolare, di analisi dell'espressione e delle mutazioni geniche, dello studio del sequenziamento genico, del significato e della classificazione dei polimorfismi, dei microRNA come nuova classe di regolatori dell'espressione genica. Per l'importanza che sempre più sta assumendo, grande spazio verrà poi riservato alla farmacogenomica, intesa come premessa per lo sviluppo di nuovi farmaci diretti verso precisi bersagli molecolari e come guida per la possibile attuazione di terapie personalizzate.

Verranno infine forniti gli elementi fondamentali dell'informatica biomedica: le banche dati e il recupero delle informazioni, i modelli per la predizione della struttura genica, l'analisi dei dati di microarrays di espressione, la ricerca finalizzata alla scoperta di biomarcatori.

Quindi, i principali obiettivi del corso (e di questo volume) sono che i partecipanti (e i lettori) acquisiscano la capacità di valorizzare al meglio l'enorme quantità di informazioni oggi disponibili su un determinato fenomeno, e che conoscano le molteplici possibilità applicative della bioinformatica alla biologia molecolare, alla farmacologia, alla clinica. Questi obiettivi sono certamente di primaria importanza nell'attuale scenario della ricerca biomedica.

Ringrazio tutti i Docenti che, particolarmente esperti degli argomenti che è stato loro richiesto di trattare, hanno generosamente corrisposto all'invito che ho loro rivolto a nome del Collegio Ghislieri per la realizzazione di questo corso e di questo volume.

Carlo Bernasconi

Pavia, 16 aprile 2007

CITOGENETICA MOLECOLARE

Metodiche innovative di citogenetica molecolare

C. Matteucci, R. La Starza, G. Barba, C. Crescenzi, S. Romoli, P. Gorello, C. Mecucci

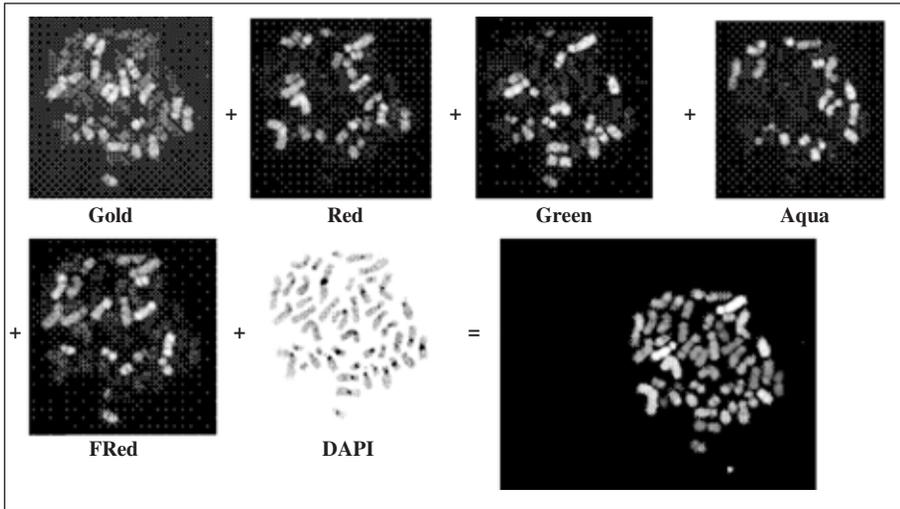
Ematologia, Fondazione IBIT, Università di Perugia

Negli ultimi decenni è stato chiaramente stabilito come le alterazioni genomiche acquisite delle cellule neoplastiche abbiano un ruolo fondamentale nella genesi delle leucemie e dei linfomi, e come trattamenti specificamente diretti verso lesioni genetiche specifiche consentano di raggiungere importanti successi clinici. Durante gli anni di sviluppo del Progetto Genoma Umano sono state sviluppate nuove tecnologie di base e nuovi strumenti per l'analisi del DNA e dell'RNA, a complemento degli studi di citogenetica convenzionale, che costituisce, ancora oggi, l'approccio iniziale nello studio del genotipo delle cellule neoplastiche. Qui di seguito si offre una visione d'insieme delle varie metodiche di citogenetica molecolare e si sottolinea l'importanza dell'integrazione fra queste e la citogenetica convenzionale ai fini diagnostici e di ricerca nelle emopatie maligne.

FISH e MultiFISH

La metodica di FISH e le sue estensioni, è senz'altro una delle metodiche più importanti nello studio del genoma; essa sfrutta la capacità di una sonda di DNA marcata con un fluorocromo o con un aptene, di legarsi in modo specifico con una sequenza bersaglio di DNA complementare, costituita da cromosomi metafasici o cellule in interfase. Il target può inoltre essere costituito anche da cellule intere su vetrino (citospins, apposizioni di tessuto) o sezioni di tessuto.

Alcune sonde di DNA corrispondono ad un intero cromosoma (sonde painting), altre sono specifiche per le regioni centromeriche, altre ancora per piccole porzioni genomiche (sonde locus-specifiche). Con la metodica di FISH in interfase è possibile individuare anomalie numeriche (quali monosomie/delezioni o trisomie/amplificazioni di regioni selezionate) o strutturali (con sonde locus- o gene-specifiche). Le applicazioni degli studi di FISH sono numerose: studio di cariotipi normali o di casi con cariotipo non disponibile per mancanza di mitosi, identificazione di un subclone anormale, valutazione della malattia residua minima o



di eventuali chimerismi post trapianto “sex-mismatch”. La FISH in metafase è invece uno strumento importante per lo studio di traslocazioni o geni di fusione noti, per individuare riarrangiamenti criptici non evidenziabili al cariotipo, per lo studio dei punti di rottura di nuove traslocazioni non conosciute a livello molecolare, e per l’identificazione di nuovi geni coinvolti nella patogenesi di malattie onco-ematologiche o di sindromi congenite.

La Multicolor-FISH utilizza simultaneamente, come sonde di DNA, 24 painting cromosomici in grado di colorare diversamente ciascuna coppia di cromosomi. In tal modo è in grado di valutare, in un unico esperimento, ciascuna coppia di omologhi e i cromosomi sessuali.

I segnali fluorescenti vengono catturati separatamente da un microscopio, dotato di filtri specifici per ogni singola lunghezza d’onda, e collegato ad una telecamera. Le immagini acquisite vengono poi elaborate da un software che le sovrappone dando origine ad un’immagine a 24 colori, ciascuno corrispondente ad una coppia di cromosomi.

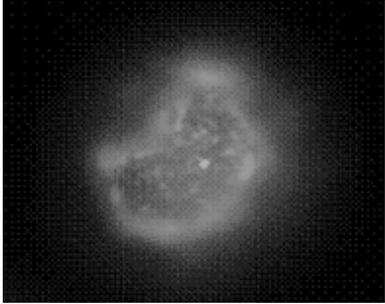
La MultiFISH è uno strumento prezioso per la definizione dei cariotipi complessi e per l’identificazione di materiale cromosomico coinvolto in traslocazioni non bilanciate, nel corso di un singolo esperimento. I limiti di applicazione di questa metodica possono essere così riassunti:

- Piccole delezioni/duplicazioni, inversioni intracromosomiche.
- False inserzioni all’interfaccia dei segmenti traslocati.
- Riarrangiamenti coinvolgenti regioni pericentriche.
- Amplificazione di materiale proveniente da cromosomi diversi in un unico marker.

Inoltre, il potere di risoluzione equivale a quello di un painting cromosomico, pertanto la MultiFISH non fornisce informazioni riguardo la definizione fine dei punti di rottura coinvolti nei riarrangiamenti strutturali.

FICTION (Fluorescence Immunophenotype and Interphase Cytogenetics as a Tool for Investigation Of Neoplasms)

Anti-CD13+probe for20q11



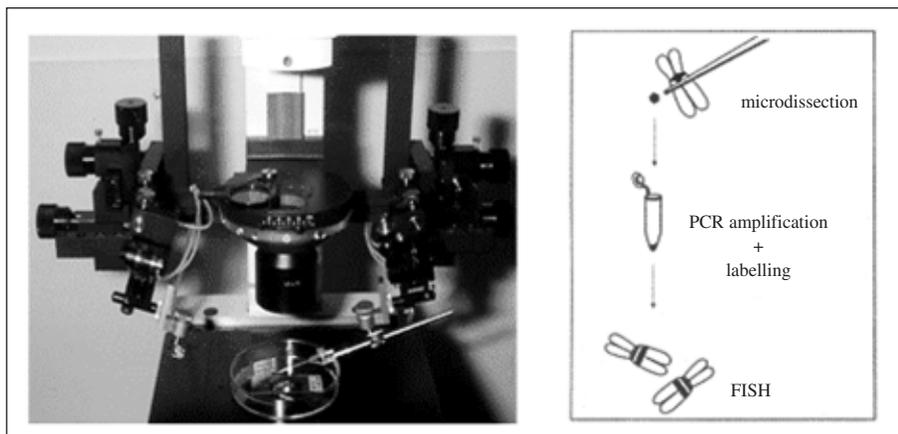
La metodica di FICTION combina le metodiche di FISH in interfase e l'immunofenotipo con anticorpi monoclonali direttamente marcati sulla stessa cellula, permettendo così l'affiliazione di linea di ciascuna anomalia genetica.

Il target cellulare è costituito da cellule intere in sospensione apposte su vetrino (citospin) o da strisci di midollo/sangue periferico, o da sezioni di tessuto incluse in paraffina. Le cellule subiscono un primo passaggio di fissazione in acetone;

successivamente viene eseguita la reazione di immunofenotipizzazione con anticorpi specifici per l'antigene di superficie che si intende valutare; una cascata di anticorpi policlonali marcati con fluorocromi di colore rosso (Cy3), giallo (fluoresceina) o blu (AMCA) rivela le cellule positive per quel determinato antigene, permettendone l'identificazione fenotipica. Successivamente il preparato viene nuovamente fissato in paraformaldeide, denaturato su piastra, e ibridato con una o più sonde di DNA selezionate e direttamente marcate con fluorocromi. Dopo l'ibridazione il preparato viene osservato con un microscopio in fluorescenza, in modo tale che sia possibile valutare simultaneamente immunofenotipo e ibridazione in situ, attribuendo specifici riarrangiamenti cromosomici (numerici o strutturali) alla filiera cellulare oggetto di studio.

Micro-FISH

In questo particolare esperimento di FISH, le sonde di DNA sono ottenute mediante microdissezione cromosomica. L'utilizzo di un micromanipolatore col-

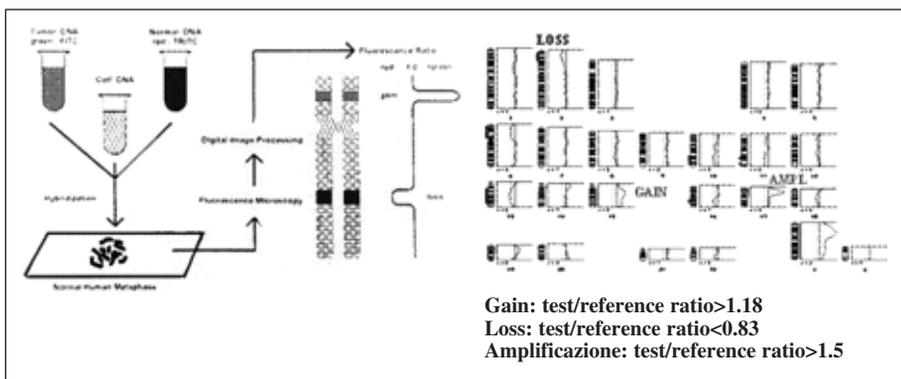


legato ad un microscopio consente, infatti, di “sezionare” interi cromosomi, o parte di essi, per ottenere molecole di DNA che verranno amplificate tramite PCR, marcate con fluorocromi o apteni e utilizzate come normali sonde di FISH. L'utilizzo principale della micro-FISH consiste nel microdisettare cromosomi anormali o marcatori non identificati al cariotipo e di ibridare questi frammenti su una metafase normale, al fine di identificare, con precisione, la provenienza del materiale cromosomico selezionato.

Comparative Genomic Hybridization. Metaphase and array-CGH

La Comparative Genomic Hybridization è uno sviluppo della FISH in metafase, in cui però la miscela di ibridazione è costituita da DNA estratto dalle cellule tumorali e DNA di riferimento (normale) in pari quantità, marcati con due diversi fluorocromi. Il pattern di ibridazione viene rielaborato con un software d'analisi d'immagine e tradotto in un profilo che riproduce, lungo l'asse di ciascun cromosoma, il rapporto fra le intensità di fluorescenza dei due fluorocromi. Se questo rapporto è maggiore di un valore “soglia” fissato a 1,18 o inferiore a 0,83, il cromosoma o la regione cromosomica interessata risulterà rispettivamente duplicata o deleta. I valori “soglia” sono stati definiti in seguito a una serie di ibridazioni competitive fra due DNA normali. Le regioni centromeriche e telomeriche sono escluse dall'analisi in quanto ricche di sequenze altamente ripetute. Per questo motivo potrebbero dare origine a pattern di ibridazione alterato, con maggiore possibilità di falsi positivi. La metodica di CGH non richiede cellule in fase mitotica; inoltre il DNA necessario per l'esperimento può essere estratto anche da campioni d'archivio inclusi in paraffina o da altro materiale biologico fissato, permettendo così studi di tipo retrospettivo.

Nella pratica, consente di evidenziare sbilanciamenti genomici quali duplicazioni o delezioni di interi cromosomi o di parte di essi e di individuare eventi criptici, quali le amplificazioni geniche, non evidenziabili all'analisi cariotipica. D'altro canto non fornisce informazioni su riarrangiamenti genomici bilanciati (traslocazioni o inversioni) in cui non si abbia, cioè, perdita o guadagno netto di materiale cromosomico. L'altro limite nel potere di risoluzione della CGH è



dovuto alle dimensioni del clone neoplastico: un'ampiezza del clone inferiore al 40% della popolazione cellulare totale non produrrà risultati apprezzabili. Anche le delezioni cromosomiche inferiori alle 10 Mb sfuggiranno al potere di risoluzione della metodica. Per ovviare a questi ultimi due inconvenienti è opportuno l'utilizzo di un software per la cosiddetta CGH ad alta risoluzione (HR-CGH). Quest'ultimo ha un potere di risoluzione più elevato, ed è in grado di identificare delezioni cromosomiche dell'ordine di circa 3 Mb. Non utilizza valori soglia fissi, ma si basa su una standardizzazione dei parametri di riferimento, valori cosiddetti "dinamici", formulati sulla base di una serie di controlli normali e rielaborati applicando un intervallo di confidenza variabile fra il 95% e il 99,9%. L'altra "estensione" della metodica di CGH è costituita dalla array-CGH. I principi di questa metodica sono essenzialmente gli stessi della CGH in metafase, ma il potere di risoluzione è maggiore poiché l'ibridazione avviene su un array dove sono "spottate" regioni selezionate di DNA (BAC, PAC o oligonucleotidi) corrispondenti a oncogeni, geni oncosoppressori, o regioni genomiche potenzialmente coinvolte nel processo di trasformazione neoplastica delle cellule. Anche nella tecnologia con array è possibile valutare la perdita o il guadagno di materiale genomico per il pool di cloni a disposizione, tramite l'utilizzo di un sistema composto da uno scanner a raggi laser (per la acquisizione dei segnali fluorescenti) e da un software per la rielaborazione e l'analisi dei dati. L'alta specificità di ibridazione rende la tecnica estremamente sensibile, ma un limite importante della metodica è quello della presenza di falsi positivi legati a siti polimorfici presenti negli array "commerciali" che utilizzano BAC e PAC. L'utilizzo degli array con oligonucleotidi è particolarmente interessante per la soluzione di questi problemi interpretativi.

Bibliografia

1. Cottier M, Tchirkov A, Perissel B, Giollant M, Campos L, Vago P. Cytogenetic characterization of seven human cancer cell lines by combinig G- and R-banding, M-FISH, CGH and chromosome- and locus-specific FISH. *Int J Mol Med* 14(4): 483-95, 2004.
2. El-Rifai W, Larramendy ML, Bjorkqvist AM, Hemmer S, Knuutila S. Optimization of Comparative Genomic Hybridization using fluorochrome conjugated to dCTP and dUTP nucleotides. *Lab Invest* 77(6): 699-700, 1997.
3. Gebhart E. Genomic imbalances in human leukemia and lymphoma detected by comparative genomic hybridization (Review). *Int J Oncol* 27(3): 593-606, 2005.
4. Kallioniemi A, Kallioniemi O-P, Sudar D, Rutovitz D, Gray JW, Waldman F, Pinkel D. Comparative Genomic Hybridization for molecular cytogenetics analysis of solid tumors. *Science* 258: 818-821, 1992.
5. Lichter P, Boyle AL, Cremer T, et al. Analysis of genes and chromosomes by non isotopic in situ hybridization. *GATA* 8: 24-35, 1991.
6. Liebisch P, Viardot A, Bassermann N, Wendl C, Roth K, Goldschmidt H,

- Einsele H, Straka C, Stilgenbauer S, Dohner H, Bentz M. Value of comparative genomic hybridization and fluorescence in situ hybridization for molecular diagnostics in multiple myeloma. *Br J Haematol* 122(2): 193-201, 2003.
7. Mathew S, Raimondi SC. FISH, CGH, and SKY in the diagnosis of childhood acute lymphoblastic leukemia. *Methods Mol Biol* 220: 213-33, 2003.
 8. Speicher MR, Gwyn Ballard S, Ward DC. Karyotyping human chromosomes by combinatorial multi-fluor FISH. *Nat Genet* 12: 368-375, 1996.
 9. Veldman T, Vignon C, Schrock E, Rowley JD, Ried T. Hidden chromosome abnormalities in haematological malignancies detected by multicolour spectral karyotyping. *Nat Genet* 15(4): 406-10, 1997.
 10. Wessendorf S, Fritz B, Wrobel G, Nessling M, Lampel S, Goettel D, Kuepper M, Joos S, Hopman T, Kokocinski F, Dohner H, Bentz M, Schwaenen C, Lichter P. Automated screening for genomic imbalances using Matrix-Based Comparative Genomic Hybridization. *Laboratory Investigation* 82: 47-60, 2002.

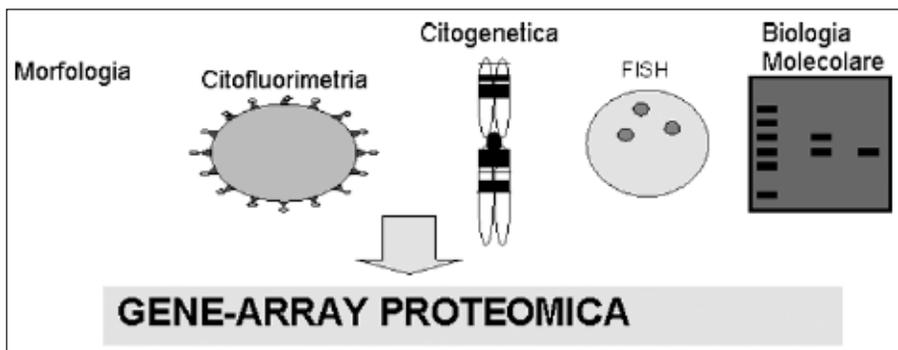
L'importanza clinica della FISH nelle leucemie

Maria Christina Cox

Dipartimento di Ematologia, Azienda Ospedaliera S. Andrea, Università di Roma "La Sapienza"

Le leucemie acute costituiscono un gruppo eterogeneo di emopatie, la cui diagnosi e caratterizzazione si basa sull'integrazione di morfologia, fenotipo immunologico e aberrazioni genetiche. La diagnosi delle aberrazioni genetiche nelle leucemie acute comporta l'integrazione delle informazioni derivanti da Citogenetica, biologia molecolare e FISH. Nuove tecnologie come gli Array e la Proteomica, stanno dando un ulteriore fondamentale contributo alla comprensione della loro eterogeneità clinica e biologica.

Le leucemie acute si suddividono in 2 grandi gruppi mieloidi (LAM) e linfoidi (LAL), entrambe si caratterizzano per l'acquisizione di mutazioni genetiche e citogenetiche che determinano lo sviluppo e l'espansione di un clone leucemico. Circa il 20-25% dei casi di LAM presenta al cariotipo una aberrazione cromosomica bilanciata (traslocazioni, inversioni), nel 35-40% dei casi presenta aberrazioni che comportano perdita e/o acquisto di ampie regioni genomiche (monosomie, trisomie, delezioni, duplicazioni...) senza aberrazioni bilanciate, nel 40-50% dei casi di LAM la citogenetica non evidenzia aberrazioni cromosomiche. Nelle LAL, che costituiscono la seconda neoplasia per frequenza durante l'età evolutiva, l'incidenza delle aberrazioni cromosomiche è dell'80-90%. Le aberrazioni citogenetiche costituiscono attualmente il più importante fattore prognosti-



Favorevole*	Sfavorevole	Molto sfavorevole
T(15:17)	Normale	Complesso ≥ 3 abn
T(8:21)	Altro (# da FAV e SFAV)	Abn 3q
Inv (16)		-t(11q23)
		-del 5q; -5
		-7

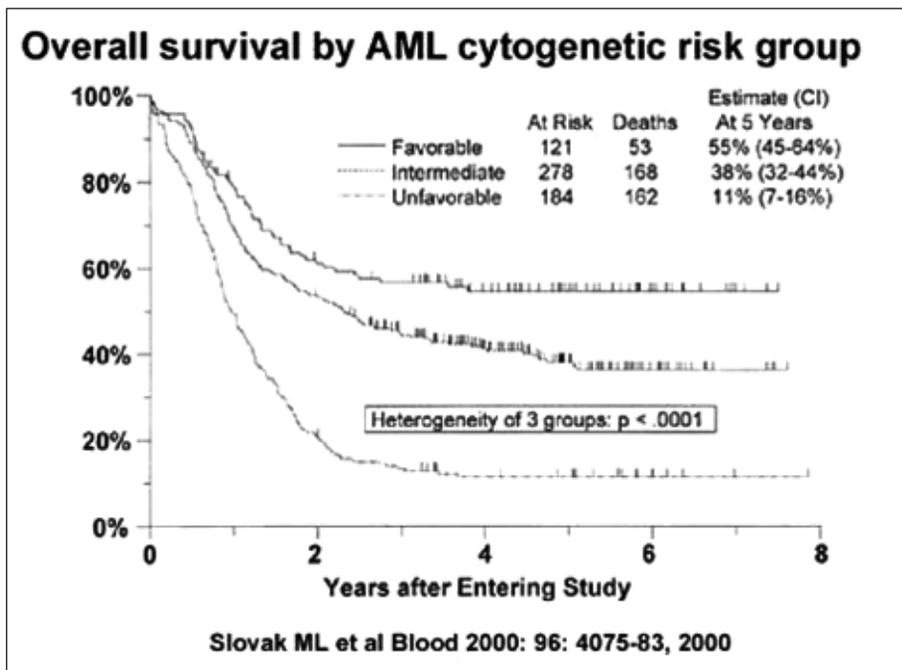
*Classificazione in base ai criteri GIMEMA/EORTC

co biologico delle leucemie acute e la loro determinazione in fase diagnostica è di fondamentale importanza ai fini.

- 1) classificativi,
- 2) prognostici,
- 3) di indirizzo terapeutico,
- 4) per il monitoraggio della malattia minima residua (MMR).

L'impatto clinico delle aberrazioni citogenetiche è tale che i pazienti affetti da LAM vengono suddivisi in base alle aberrazioni cromosomiche in: gruppo a prognosi favorevole, intermedia e sfavorevole.

Nelle LAM i pazienti con cariotipo normale vengono assegnati al gruppo di prognosi intermedia, che rappresenta in realtà un gruppo molto eterogeneo in cui nel corso degli ultimi anni, sono state identificate delle mutazioni genetiche relative ad alcuni geni come FLT3, MLL, NPM, C/EBP α con significativo impatto prognostico. Queste recenti acquisizioni sottolineano ancora di più quanto sia neces-



sario integrare diverse tecnologie per stratificare pazienti che vengono raggruppati in classi solo apparentemente omogenee.

Nelle LAL non esiste una classificazione così stratificata ma alcune traslocazioni sono associate ad una prognosi più severa: t(9;22); abn 11q23, t(8; 14) e varianti. Le LAL dell'adulto e del bambino si distinguono per tipo e frequenza di aberrazioni cromosomiche, nell'adulto circa il 30% dei casi ha la t(9;22).

Nessuna aberrazione è da sola sufficiente ad instaurare e a promuovere il fenotipo leucemico, devono infatti verificarsi un discreto accumulo di mutazioni (teoria multistep della tumorigenesi), prima che la leucemia si manifesti. Le aberrazioni citogenetiche ricorrenti nelle leucemie acute sono diverse decine, tuttavia sono state individuati dei meccanismi molecolari comuni a tutte le leucemie:

- 1) Alterata proliferazione,
- 2) Blocco differenziativo,
- 3) Deficit dei meccanismi apoptotici,
- 4) Auto-rinnovamento (leukemic stem cell),
- 5) Alterazioni del ciclo cellulare,
- 6) Instabilità genomica,
- 7) Disseminazione leucemica.

All'origine della scoperta delle aberrazioni cromosomiche ci sono gli studi di citogenetica classica. L'analisi citogenetica, comporta una serie complessa di procedure: coltura, blocco delle metafasi, ipotonizzazione, spreading dei vetrini, bandeggio, colorazione, osservazione microscopica e analisi dei cromosomi bandeggiati.

Il cariotipo delle cellule leucemiche costituisce tutt'ora uno strumento insostituibile nella fase di diagnosi in quanto consente una analisi ad ampio spettro, in cui tutte le aberrazioni genetiche che comportano un rimodellamento sufficientemente ampio della struttura cromosomica (nell'ordine del milione di basi) vengono individuate. Tuttavia sussistono una serie di limiti tecnici e biologici:

- 1) l'esame è eseguito su cellule in metafase,
- 2) essendo un esame morfologico è strettamente dipendente dalla qualità dei cromosomi,
- 3) insufficiente risoluzione ottica per definire riarrangiamenti che coinvolgono le regioni telomeriche,
- 4) difficoltà o impossibilità nel definire riarrangiamenti complessi,
- 5) lunghezza e complessità del procedimento che richiede una grande manualità e una grande esperienza interpretativa.

Nelle leucemie acute queste difficoltà sono particolarmente acute dal fatto che c'è pochissimo tempo per fare la diagnosi: la chemioterapia deve essere iniziata il prima possibile e inoltre il materiale biologico che viene prelevato (sangue periferico e/o midollare) deve essere suddiviso tra più laboratori.

Molteplici metodi possono essere utilizzati per evidenziare specifiche aberrazioni cromosomiche: PCR-based, Southern-blotting, FISH. Poiché le mutazioni genetiche da identificare sono in continuo aumento i metodi PCR-based che com-

portano l'amplificazione multipla di diversi cDNA, non sono pratici. Fortunatamente lo sviluppo di sonde molecolari da usare in FISH costituiscono una valida alternativa. I metodi FISH non hanno la stessa sensibilità della PCR, la loro maggiore utilità è infatti in fase diagnostica e di recidiva quando la percentuale di cellule leucemiche è elevata. La FISH è una metodica estremamente valida nella rilevazione di riarrangiamenti cromosomici, monosomie, delezioni, trisomie, amplificazioni e può essere utilizzata su sangue periferico, midollare, apposizioni a contatto, sezioni di paraffina. La FISH non è utile per la valutazione della malattia minima residua in quanto la sensibilità dei metodi di routine è di $10e2$. Con una ibridazione di molteplici sonde over-night è possibile nel giro di poche ore fare una diagnosi molecolare estremamente sensibile e specifica.

Nel caso in cui l'esame Citogenetico sia fallito (assenza di metafasi, morfologia inadeguata, poche metafasi), mediante un ampio pannello di sonde per la FISH si possono individuare e/o escludere una serie di aberrazioni ricorrenti.

Più frequentemente invece la FISH è utilizzata per integrare l'informazione derivante dall'esame citogenetico: in particolare nelle LAL anche in presenza di un cariotipo normale l'esame FISH è utilissimo per verificare la presenza del riarrangiamento BCR/ABL correlato alla $t(9;22)$, o ETV6/AML1 correlato alla $t(12;21)$; la FISH è molto utile nel caso in cui si evidenziano riarrangiamenti di regioni cromosomiche terminali come il locus 11q23, infatti spesso non è evidente quale sia il partner traslocativo: grazie all'ampio utilizzo di questa metodica è stato possibile identificare oltre 50 loci cromosomici che possono fungere da partner traslocativi del locus 11q23. La FISH è poi fondamentale per dirimere cariotipi complessi e l'origine di cromosomi derivati da riarrangiamenti non evidenti all'esame del kariogramma. Inoltre una volta identificato un marcatore, in fase di sospetta recidiva la FISH consente rapidamente di verificare la presenza del clone leucemico. In conclusione possiamo dire che la FISH ha consentito enormi progressi nella identificazione di nuove traslocazioni cromosomiche ed è attualmente la metodica più sensibile semplice e robusta, per la diagnosi delle aberrazioni cromosomiche.

Bibliografia

1. Haferlach C, Rieder H, Lillington DM, Dastugue N, Hagemeijer A, Harbott J, Stilgenbauer S, Knuutila S, Johansson B, Fonatsch C. Proposals for standardized protocols for cytogenetic analyses of acuteleukemias, chronic lymphocytic leukemia, chronic myeloid leukemia, chronicmyeloproliferative disorders, and myelodysplastic syndromes. *Genes Chromosomes Cancer*. 2007 May; 46(5): 494-9.
2. Forestier E, Andersen MK, Autio K, Blennow E, Borgstrom G, Golovleva I, Heim S, Heinonen K, Hovland R, Johannsson JH, Kerndrup G, Nordgren A, Rosenquist R, Swolin B, Johansson B. Cytogenetic patterns in ETV6/RUNX1-positive pediatric B-cell precursor acutelymphoblastic leukemia: A Nordic series of 245 cases and review of the literature. *Genes Chromosomes Cancer*. 2007 May; 46(5): 440-50.

3. Mrozek K, Dohner H, Bloomfield CD. Influence of new molecular prognostic markers in patients with karyotypically normal acute myeloid leukemia: recent advances. *Curr Opin Hematol*. 2007 Mar; 14(2): 106-14.
4. Hasle H, Alonzo TA, Auvrignon A, Behar C, Chang M, Creutzig U, Fischer A, Forestier E, Fynn A, Haas OA, Harbott J, Harrison CJ, Heerema NA, van den Heuvel-Eibrink MM, Kaspers GJ, Locatelli F, Noelle P, Polychronopoulou S, Ravindranath Y, Razzouk B, Reinhardt D, Savva NN, Stark B, Suciu S, Tsukimoto I, Webb DK, Wojcik D, Woods WG, Zimmermann M, Niemeyer CM, Raimondi SC. Monosomy 7 and deletion 7q in children and adolescents with acute myeloid leukemia: an international retrospective study. *Blood*. 2007 Feb 13; [Epub ahead of print].
5. Satwani P, Sather H, Ozkaynak F, Heerema NA, Schultz KR, Sanders J, Kersey J, Davenport V, Trigg M, Cairo MS. Allogeneic bone marrow transplantation in first remission for children with ultra-high-risk features of acute lymphoblastic leukemia: A children's oncology group study report. *Biol Blood Marrow Transplant*. 2007 Feb; 13(2): 218-27.
6. Falini B, Nicoletti I, Martelli MF, Mecucci C. Acute myeloid leukemia carrying cytoplasmic/mutated nucleophosmin (NPMc+ AML): biologic and clinical features. *Blood*. 2007 Feb 1; 109(3): 874-85. Epub 2006 Sep 28. Review.
7. Schultz KR, Pullen DJ, Sather HN, Shuster JJ, Devidas M, Borowitz MJ, Carroll AJ, Heerema NA, Rubnitz JE, Loh ML, Raetz EA, Winick NJ, Hunger SP, Carroll WL, Gaynon PS, Camitta BM. Risk- and response-based classification of childhood B-precursor acute lymphoblastic leukemia: a combined analysis of prognostic markers from the Pediatric Oncology Group (POG) and Children's Cancer Group (CCG). *Blood*. 2007 Feb 1; 109(3): 926-35. Epub 2006 Sep 26.
8. Brezinova J, Zemanova Z, Ransdorfova S, Pavlistova L, Babicka L, Houskova L, Melichercikova J, Siskova M, Cermak J, Michalova K. Structural aberrations of chromosome 7 revealed by a combination of molecular cytogenetic techniques in myeloid malignancies. *Cancer Genet Cytogenet*. 2007 Feb; 173(1): 10-6.
9. Rosati R, La Starza R, Barba G, Gorello P, Pierini V, Matteucci C, Roti G, Crescenzi B, Aloisi T, Aversa F, Martelli MF, Mecucci C. Cryptic chromosome 9q34 deletion generates TAF-Ialpha/CAN and TAF-Ibeta/CAN fusion transcripts in acute myeloid leukemia. *Haematologica*. 2007 Feb; 92(2): 232-5.
10. Strefford JC, Worley H, Barber K, Wright S, Stewart AR, Robinson HM, Bettney G, van Delft FW, Atherton MG, Davies T, Griffiths M, Hing S, Ross FM, Talley P, Saha V, Moorman AV, Harrison CJ. Genome complexity in acute lymphoblastic leukemia is revealed by array-based comparative genomic hybridization. *Oncogene*. 2007 Jan 22; [Epub ahead of print].
11. Mrozek K, Marcucci G, Paschka P, Whitman SP, Bloomfield CD. Clinical relevance of mutations and gene-expression changes in adult acute myeloid leukemia with normal cytogenetics: are we ready for a prognostically priori-

- tized molecular classification? *Blood*. 2007 Jan 15; 109(2): 431-48. Epub 2006 Sep 7.
12. MC Cox, Panetta P, Venditti A, Del Poeta G, Franchi A, Buccisano F, Tamburini A, Maurillo L and Amadori S. Comparison between conventional banding analysis and FISH screening with an AML-specific set of probes in 260 patients. *The Hematology Journal* 2003.
 13. Pileri SA, Ascani S, Cox MC, Campidelli C, Bacci F, Piccioli M, Piccaluga PP, Agostinelli C, Asioli S, Novero D, Bisceglia M, Ponzoni M, Gentile A, Rinaldi P, Franco V, Vincelli D, Pileri A Jr, Gasbarra R, Falini B, Zinzani PL, Baccharani M. Myeloid sarcoma: clinico-pathologic, phenotypic and cytogenetic analysis of 92 adult patients. *Leukemia*. 2007 Feb; 21(2): 340-50.

Nuove anomalie citogenetiche nelle malattie mieloproliferative croniche

*C. Matteucci, R. La Starza, G. Barba, C. Crescenzi, S. Romoli,
P. Gorello, C. Mecucci*

Ematologia, Fondazione IBIT, Università di Perugia

Le malattie mieloproliferative croniche sono caratterizzate da proliferazione di uno o più tipi cellulari della linea mieloide (eritroide, granulocitica e megacariocitica). Tale “amplificazione” della linea mieloide, derivante dalla deregolazione della cellula staminale emopoietica, non incide significativamente sulle caratteristiche di maturazione della cellula mieloide. L’Organizzazione mondiale della Sanità (WHO) ha suggerito la classificazione dei disordini mieloproliferativi cronici (DMPC) in due gruppi principali:

- 1) DMPC classici, comprendenti la leucemia mieloide cronica, la policitemia vera (PV), la trombocitemia essenziale (TE) e la mielofibrosi idiopatica (MFI);
- 2) DMPC meno frequenti quali la leucemia neutrofilica cronica, la sindrome ipereosinofila/leucemia eosinofila cronica, e altri disordini mieloproliferativi non ben classificati.

La mastocitosi sistemica è considerata una entità intermedia fra disordini mieloproliferativi e mielodisplastici, così come la leucemia mielomonocitica cronica e la leucemia mieloide cronica atipica. Dal punto di vista della caratterizzazione genetico-molecolare la caratteristica comune più frequente in queste patologie è il coinvolgimento di molecole tirosin kinasiche, la cui attivazione costitutiva può essere dovuta alla formazione di una proteina di fusione con caratteristiche oncogeniche, o ad una mutazione puntiforme che ne modifica la funzione. Il primo modello di studio nei DMPC è stata la leucemia mieloide cronica, in cui la scoperta del riarrangiamento cromosomico (traslocazione 9;22) e, in un secondo tempo, dell’evento molecolare (fusione dei geni *BCR* e *ABL*) ha portato alla precisa caratterizzazione funzionale dell’oncoproteina *BCR/ABL*, e all’utilizzo di farmaci antineoplastici specificamente diretti verso la lesione molecolare, con funzione inibitoria dell’attività tirosin-kinasica costitutiva della proteina di fusione. Da qui l’inizio di una serie di studi volti all’identificazione di altre anomalie citogenetiche ricorrenti per determinare il ruolo di altre tirosin kinasi nei DMPC *BCR/ABL*-negativi.

Riarrangiamenti strutturali di PDGFRB, FGFR1 e JAK2

Fra i geni più frequentemente coinvolti nella patogenesi dei DMPC BCR/ABL-negativi, ritroviamo *PDGFRB*, localizzato sul braccio lungo del cromosoma 5 (5q33). Esso codifica per il recettore per le tirosin kinasi “platelet derived growth factor receptor beta” (PDGFR β), normalmente espresso nelle cellule del sistema emopoietico, nei fibroblasti, nelle cellule endoteliali e negli osteoblasti e che svolge un ruolo fondamentale nella differenziazione monocitica, nei processi infiammatori, di cicatrizzazione delle ferite e di angiogenesi.

Il gene *PDGFRB* è spesso riarrangiato nei DMPC con differenti geni partners, elencati in tabella. In generale i breakpoints sono tali che la risultante proteina di fusione possiede il dominio catalitico intracellulare di PDGFR β ma il dominio extracellulare di questi recettori viene sostituito dalla porzione N-terminale della proteina partner. Le proteine chimeriche non rispondono più allo stimolo del PDGF ma possiedono un’attività tirosin kinasica costitutiva e inviano continuamente segnali proliferativi e anti-apoptotici alla cellula nella quale risiedono.

	Gene Partner	Localizzazione cromosomica
PDGFRB (5q33)	H4	10q22
	ETV6	12p13
	NIN	14q24
	CEV14	14q32
	KIA A 1509	14q32
	TP53BP1	15q22
	RABAPTIN	17p13
	HCMOGT-1	17p11.2
	PDE4DIP	1q23
	HIP	7q11
	NDE1	16p13

Il primo riarrangiamento cromosomico coinvolgente il gene *PDGFRB* descritto è stata la traslocazione t(5;12)(q31;p12). In questo riarrangiamento il gene *ETV6*, localizzato sul braccio corto del cromosoma 12, si fonde con il gene *PDGFRB* e genera una proteina di fusione che, come *BCR/ABL*, possiede attività tirosin-kinasica costitutiva. Proprio per l’analogia di struttura e meccanismo d’azione, anche in questa patologia il trattamento mirato con molecole inibitorie dell’attività tirosin kinasica costituiscono il più efficace trattamento terapeutico.

Riarrangiamenti del gene *FGFR1*, localizzato sul braccio corto del cromosoma 8 (8p11), sono stati osservati in DMPC associati a proliferazione linfoide, le cosiddette sindromi “stem cell leukemia-lymphoma”. Il primo gene partner identificato, e il più frequentemente coinvolto, è lo zinc finger *ZNF198*, localizzato alla banda 13q12. Anche in questo caso la traslocazione porta all’attivazione costitutiva di una tirosin kinasi nella quale il gene partner fornisce la regione codificante per un dominio di oligomerizzazione. Come per *PDGFRB* la lista dei partner di fusione è in continua espansione. I più frequenti sono elencati in tabella.

	Gene Partner	Localizzazione cromosomica
FGFR1 (8p11)	ZNF198	13q12
	TRIM24	7q32
	FGFR1OP	6q27
	FGFR1OP2	12p11
	MYO18A	17q11
	BCR	22q11
	LOC113386	19q13.3
	CEP1	9q33

Più recentemente è stata descritta la specifica associazione fra traslocazione t(8;9)(p22;p24) e DMPC o leucemie acute. La traslocazione porta alla fusione fra i geni *PCMI* e *JAK2*. Analogamente alle fusioni dei geni *ABL1*, *PDGFRB* e *FGFR1*, anche il trascritto *PCMI/JAK2* codifica per una tirosin kinasi costitutivamente attivata che promuove la iperproliferazione cellulare. Inoltre, sono stati descritti rari casi di riarrangiamento *BCR/JAK2* dovuto alla traslocazione reciproca t(9;22)(p24;q11), con caratteristiche fenotipiche sovrapponibili alla LMC *BCR/ABL*-positiva.

Alterazioni genetiche nei DMPC a cariotipo apparentemente normale

Geni di fusione

Accanto alla leucemia mieloide cronica e ad alcuni casi di leucemia eosinofila cronica caratterizzati dai riarrangiamenti strutturali sopra descritti, gran parte dei disordini mieloproliferativi cronici non presentano anomalie cromosomiche evidenziabili all'analisi cariotipica.

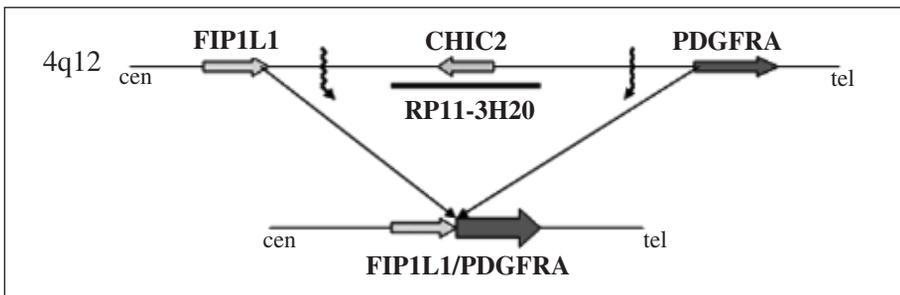
Ciò rende più difficile lo studio genetico di queste malattie, poiché l'anomalia cromosomica fornisce precise informazioni sui possibili geni candidati localizzati nella regione cromosomica riarrangiata. Sulla base dei meccanismi oncogenici di altri DMPC l'attenzione dei ricercatori si è concentrata nella ricerca di altri geni, codificanti per tirosin kinasi, potenzialmente coinvolti nei casi non ancora definiti sul piano genetico.

Questa ipotesi è stata validata dalla individuazione del nuovo gene di fusione *FIP1L1/PDGFR* in un sottogruppo di leucemie eosinofile croniche che spesso, per mancanza di dati di clonalità, venivano classificate come forme "idiopatiche". In questi casi, una delezione criptica dell'ampiezza di circa 800 kb a livello della banda 4q12, porta alla giustapposizione dei geni *FIP1L1* e *PDGFR*. Il trascritto di fusione così formato genera, anche questa volta, una proteina con funzione tirosin-kinasica costitutiva, estremamente sensibile al trattamento con farmaci inibitori anche a basso dosaggio.

Questa alterazione genetica, che riguarda il 20-50% dei casi di leucemia eosinofila cronica e sindrome ipereosinofila, è oggi facilmente diagnosticabile tramite la metodica di FISH con una sonda specifica per il gene *CHIC2* che mappa nella regione di delezione, o tramite RT-PCR per individuare il trascritto di fusione.

Mutazioni geniche

Il difetto molecolare alla base di DMPC quali la trombocitemia essenziale (TE), la policitemia vera (PV) e la mielofibrosi idiopatica (MFI), è stato individuato solo recentemente. Questi disordini mieloproliferativi sono considerate entità “promiscue” in quanto hanno in comune numerose caratteristiche biologiche e cliniche. I progenitori ematopoietici, in queste malattie, mostrano un’elevata sensibilità a varie citochine, compresa la eritropoietina (Epo), la IL-3, la GM-CSF, la trombopoietina (Tpo). Frequentemente, si osserva l’evoluzione clinica da TE a PV o MFI, e da PV a MFI. Per questa serie di motivi si è ipotizzato che il difetto molecolare alla base di queste manifestazioni fenotipiche potesse essere comune. Nel 2005 gruppi diversi, utilizzando differenti approcci metodologici, sono arrivati all’individuazione di una mutazione puntiforme del gene JAK2, facente parte della famiglia JAK con attività tirosin kinasica. Questo gene, in condizioni fisiologiche, codifica per delle proteine che si legano al dominio intracitoplasmatico di vari recettori per le citochine, promuovendo la trasmissione del segnale, tramite la via JAK-STAT. La mutazione di JAK2 descritta nei disordini mieloproliferativi consiste nella sostituzione G>T in posizione 1849 (corrispondente all’esone 14), provocando la sostituzione di una valina con una fenilalanina al codone 617 della proteina corrispondente. Tale mutazione (denominata V617F) è localizzata nel dominio JH2, ad attività pseudokinastica, che svolge un ruolo inibitorio sul dominio kinasico JH1, situato all’estremità C-terminale del gene. La valina 617 e la cisteina 618 sono due aminoacidi importanti per il mantenimento del dominio JH1 in una conformazione inattiva. La sostituzione della valina 617 con una fenilalanina destabilizza questa reazione inibitoria e provoca un aumento dell’attività kinasica, indipendente dalla presenza di citochine. Studi recenti hanno evidenziato come la mutazione JAK2 V617F non sia presente a livello germinale in casi di DMPC di tipo familiare; ciò suggerisce che altri eventi molecolari alternativi concorrano, insieme alla mutazione di JAK2, all’insorgenza della malattia. D’altro canto, studi funzionali della mutazione V617F nel topo hanno chiaramente indicato come questa svolga un ruolo determinante nell’origine della malattia, in quanto topi trapiantati con midollo osseo recante la mutazione hanno sviluppato una condizione di mielofibrosi midollare sovrapponibile alla naturale progressione della PV nell’uomo. Inoltre questi studi hanno evidenziato come bassi livelli di espressione di JAK2 V617F siano associati ad una condizione di trombocitosi, mentre livelli maggiori di espressione produrrebbero una evoluzio-



ne verso una policitemia o mielofibrosi. L'ipotesi "dosaggio" è quindi una possibile spiegazione per l'eterogeneità dei DMPC JAK2 V617F-positivi. In una percentuale rilevante di pazienti con mutazione V617F di JAK2, questa è presente in omozigosi. Studi mirati hanno evidenziato che ciò è dovuto non tanto a delezione dell'allele wild-type, quanto ad un evento di ricombinazione mitotica, che porta allo scambio di materiale genomico fra la regione cromosomica 9p24 contenente l'allele normale con la regione 9p24 recante l'allele mutato, generando, in tal modo, una disomia uniparentale a livello di questa regione. Di conseguenza, i pazienti omozigoti per V617F presentano due copie dell'allele mutato, in assenza della copia wild-type. In accordo con l'ipotesi "dosaggio", la condizione di eterozigosi (=attività kinasica più bassa) favorirebbe una condizione "benigna", mentre lo stato di omozigosi determinerebbe un fenotipo più aggressivo e una progressione di malattia verso la condizione di mielofibrosi. Molto recentemente è stato identificato un sottogruppo di DMPC JAK2 V617F-negativi che presentano altre mutazioni, localizzate all'esone 12 del gene. Le 4 nuove mutazioni descritte, denominate F537-K539delinsL, H538QK539L, K539L e N542-E543del, sono state riscontrate soltanto in pazienti con policitemia vera o con eritrocitosi idiopatica, ma non in pazienti V617F-positivi, né in casi di trombocitemia essenziale o mielofibrosi idiopatica V617F-negativi. Queste mutazioni, pertanto, definiscono un nuovo sottogruppo di DMPC, con caratteristiche cito-morfologiche peculiari, sottolineando, ancora una volta, l'importanza della classificazione genetico-molecolare nella diagnosi e nella terapia di ciascuna entità.

Un altro meccanismo di attivazione costitutiva della pathway JAK-STAT avviene in conseguenza di alterazioni a livello di molecole con funzione recettoriale per le citochine, come il recettore per l'eritropoietina (EPOR), per il "granulocyte-colony stimulating factor" (GCSFR), o per la trombopoietina (MPL). Nel 2006 Pikman et al. hanno studiato, tramite sequenziamento diretto, i geni codificanti per questi recettori, in una serie di pazienti con DMPC JAK2 V617F-negativi. Non sono state individuate mutazioni di EPOR o GCSFR, ma 4/45 pazienti (9%) con mielofibrosi hanno evidenziato una mutazione somatica del gene MPL, e sostituzione di una molecola di triptofano con una di leucina a livello del codone 515 della proteina corrispondente (mutazione MPLW515K). La mutazione descritta non è presente in nessun caso analizzato di TE o di PV, né in 270 individui sani utilizzati come controllo. In saggi di proliferazione in vitro, la mutazione W515L ha conferito ad alcune linee cellulari, una capacità di crescita indipendente dallo stimolo con citochine, la fosforilazione costitutiva della via JAK/STAT, e la ipersensibilità alla TPO. Studi in vivo su topi trapiantati con midollo osseo recante la mutazione hanno evidenziato lo sviluppo di una condizione di splenomegalia, trombocitosi e leucocitosi, un midollo marcatamente ipercellulare e un aumento di megacariociti con morfologia atipica e displastica. Di nuovo una mutazione genica è stata individuata come causa di un altro disordine mieloproliferativo cronico, la mastocitosi sistemica. Questa patologia è un evento abbastanza infrequente, con un quadro biologico e clinico complesso. È caratterizzata da una proliferazione anomala e incontrollata di mastociti in numerosi tessuti come la pelle, il midollo osseo, il tubo digerente. La fisiopatologia

delle mastocitosi è diventata più chiara in seguito alla individuazione delle mutazioni del gene c-KIT, un recettore di membrana dello “stem cell factor”, il principale fattore di crescita del mastocita. In condizioni normali il legame dello “stem cell factor” al suo recettore è necessario per la attività kinasica della proteina KIT, mentre la mutazione del gene rende la proteina costitutivamente attivata, contribuendo alla sopravvivenza e alla proliferazione incontrollata della cellula neoplastica. Sono state descritte diverse mutazioni del gene KIT. La più frequente, tipica della mastocitosi dell’adulto e di forme atipiche della mastocitosi del bambino, consiste nella sostituzione valina → acido aspartico al codone 816 (D816V).

Questa mutazione non è stata riscontrata nella leucemia mieloide cronica, nella trombocitemia essenziale, nella policitemia vera, o nella mielofibrosi. In un numero limitato di altri disordini mieloproliferativi è stata descritta la mutazione D52N dell’esone 2 di KIT. Gli effetti funzionali di questa seconda mutazione non sono stati determinati in vitro, ma la sua associazione con un numero ridotto di pazienti non suggerisce un ruolo fondamentale nella patogenesi di queste malattie. Altre, e diverse, mutazioni di KIT (delezioni o inserzioni a livello dell’esone 8) sono state frequentemente descritte nelle leucemie acute mieloidi, particolarmente nei sottotipi M0 e M1 (69% dei casi) e in associazione con anomalie cariotipiche come la t(8;21) o la inv(16). Il codone 419 è coinvolto in tutti i casi, ma il significato funzionale delle diverse mutazioni non è stato determinato sperimentalmente. Le diverse mutazioni di KIT hanno dimostrato conferire una diversa sensibilità ai farmaci inibitori; la determinazione delle varie mutazioni e dei loro effetti in vitro è quindi di fondamentale importanza per la classificazione, la stratificazione prognostica e il trattamento mirato della malattia in questi pazienti.

Bibliografia

1. De Keersmaecker K and Cools J. Chronic myeloproliferative disorders: a tyrosin kinase tale. *Leukemia* 2006; 20: 200-205.
2. Delhommeau F, Pisani DF, James C, et al. Oncogenic mechanisms in myeloproliferative disorders. *Cell Mol Life Sci* 2006; 63: 2939-2953.
3. Kralovics R, Passamonti F, Buser AS, et al. A gain-of-function mutation of JAK2 in myeloproliferative disorders. *New Engl J Med* 2005; 352: 1779-1790.
4. Levine RL and Gilliland G. JAK-2 mutations and their relevance to myeloproliferative disease. *Curr Opin in Hematol* 2007; 14: 43-47.
5. Longley BJ, Reguera MJ, Ma Y. Classes of c-kit activating mutations: proposed mechanisms of action and implications for disease classification and therapy. *Leukemia Research* 2001; 25: 571-576.
6. Pikman Y, Lee BH, Mercher T, et al. MPLW515L is a novel somating activating mutation in myelofibrosis with myeloid metaplasia. *PLoS Medicine* 2006; 3(7): 1140-1151.
7. Reiter A, Sohal J, Kulkarni S, et al. Consistent fusion of ZNF198 to the

- fibroblast growth factor receptor-1 in the t(8;13)(p11;q12) myeloproliferative syndrome. *Blood* 1998; 92: 1735-1742.
8. Reiter A, Walz C, Watmore A, et al. The t(8;9)(p22;p24) is a recurrent abnormality in chronic and acute leukaemia that fuses PCM1 to JAK2. *Cancer Res* 2005; 65: 2662-2667.
 9. Scott LM, Tong W, Levine RL, et al. JAK2 Exon 12 mutations in polycythemia vera and idiopathic erythrocytosis. *New Engl J Med* 2007; 365: 459-468.

Significato delle delezioni genomiche associate a traslocazioni cromosomiche nelle leucemie

Giorgina Specchia, Francesco Albano

Dipartimento di Ematologia, Università di Bari

La caratterizzazione citogenetico-molecolare dei punti di rottura dei riarrangiamenti ricorrenti nelle leucemie ha consentito l'identificazione di microdelezioni genomiche associate ai riarrangiamenti cromosomici ricorrenti. Questo dato è stato dimostrato per la prima volta nella traslocazione t(9;22) dei pazienti affetti da leucemia mieloide cronica (LMC); in seguito le microdelezioni sono state rilevate anche in altre anomalie ricorrenti delle leucemie acute come la inv(16) ed i riarrangiamenti che coinvolgono il gene MLL. Al contrario, la t(8;21) raramente è associata alle microdelezioni ed inoltre la t(15;17) della leucemia acuta promielocitica non presenta mai perdita di sequenze genomiche.

Alcuni studi di caratterizzazione citogenetico-molecolare dei riarrangiamenti cromosomici hanno rivelato che anche le traslocazioni sporadiche presenti nelle leucemie acute e croniche possono presentare la perdita di sequenze genomiche. Il significato della presenza di microdelezioni associate ad un riarrangiamento cromosomico apparentemente bilanciato rimane ancora insoluto, sia in termini del meccanismo molecolare responsabile di tali anomalie sia per quanto riguarda il significato biologico e prognostico.

La LMC rappresenta la prima malattia neoplastica umana nella quale una specifica anomalia del cariotipo è stata associata alla presenza di un tumore. Nel 90-95% dei pazienti affetti da LMC è possibile osservare la presenza del cromosoma Philadelphia (Ph) derivato dalla traslocazione reciproca t(9;22)(q34;q11). La traslocazione porta alla formazione di due geni ibridi reciproci derivanti dalla fusione tra il proto-oncogene *c-ABL* (9q34) ed il gene *BCR* (22q11); il gene ibrido 5'*BCR/3'ABL* si forma sul cromosoma Ph (derivativo 22), mentre il gene reciproco 5'*ABL/3'BCR* si forma sul derivativo del cromosoma 9. Nel 5% circa dei casi di LMC il cariotipo è apparentemente normale (pazienti LMC "Ph negativi") ma le tecniche molecolari di Southern blot, RT-PCR e di ibridazione in situ fluorescente (FISH) permettono di evidenziare la presenza del gene di fusione 5'*BCR/3'ABL*; generalmente il gene è localizzato in 22q11, ma non sono rari i casi in cui è stato osservato in 9q34.

La proteina di fusione *BCR-ABL* induce la trasformazione cellulare attraverso tre principali meccanismi:

- 1) alterata adesione delle cellule mieloidi allo stroma e alla matrice extracellulare;
- 2) attivazione costitutiva della divisione cellulare;
- 3) riduzione dei meccanismi apoptotici.

Recentemente, in circa il 9-16% dei pazienti affetti da LMC, è stata descritta sul cromosoma der(9) la perdita di sequenze genomiche appartenenti ai cromosomi 9 e/o 22 in corrispondenza del breakpoint della t(9;22). Le delezioni sono state evidenziate sia a monte del gene *ABL* che a valle del gene *BCR* ed hanno un'ampiezza variabile che va da 260 Kb a 41 Mb per le sequenze del cromosoma 9 e da 230 Kb a 12 Mb per le sequenze del cromosoma 22. I punti di rottura hanno una localizzazione eterogenea; maggiore clusterizzazione è stata osservata relativamente ai breakpoints delle delezioni di sequenze del cromosoma 22. Le delezioni sembrano insorgere contemporaneamente al riarrangiamento *BCR/ABL*, dal momento che sono state osservate alla diagnosi in tutte le cellule leucemiche con la t(9;22).

È stata riscontrata un'associazione statisticamente significativa tra la presenza delle delezioni sul der(9) e i casi t(9;22) varianti, così definite per il fatto che nella traslocazione può essere coinvolto un terzo o più cromosomi oltre il 9 e il 22. È stato quindi ipotizzato che i riarrangiamenti molecolari complessi che accompagnano le traslocazioni cromosomiche varianti siano associati ad un aumento della probabilità di insorgenza delle delezioni, dato confermato dal fatto che alcuni casi di LMC con riarrangiamenti varianti della t(9;22) mostrano la perdita di sequenze genomiche appartenenti al terzo cromosoma coinvolto nella traslocazione.

È noto che la presenza delle delezioni sul cromosoma der(9) nei pazienti con LMC in trattamento terapeutico con α -IFN è determinante nello stabilire una prognosi sfavorevole in termini di sopravvivenza, di progressione della malattia e della risposta al trattamento; infatti i pazienti portatori di delezioni mostrano una maggiore resistenza al trattamento terapeutico con α -IFN, una durata più breve della fase cronica ed una riduzione della mediana di sopravvivenza.

Non è chiaro il modo in cui la presenza delle delezioni sul der(9) possa condizionare la progressione della patologia, ma si pensa che un ruolo importante possa essere svolto dalla perdita di una copia allelica di importanti geni che funzionano come oncosoppressori (tumor suppressor gene, TSG); l'inattivazione del secondo allele, secondo il modello di Knudson, potrebbe essere determinata da eventi secondari come mutazioni, delezioni o silenziamento mediante metilazione o imprinting.

Una seconda ipotesi prevede invece che la perdita di un solo allele di un TSG possa già da sola contribuire alla progressione di un tumore, come è stato recentemente osservato per il gene p27kip1. Infatti è stato dimostrato che la monosomia di particolari regioni cromosomiche e/o la perdita di una copia di un gene specifico, sono coinvolti in alcuni difetti dello sviluppo nell'uomo e nel topo per effetto del meccanismo dell'aploinsufficienza (effetto di dose genica insufficiente). Se l'aploinsufficienza per specifici geni può alterare lo sviluppo embrionale

è allora probabile che meccanismi simili possano essere coinvolti nell'induzione della trasformazione clonale alla base dello sviluppo dei tumori.

L'introduzione dell'imatinib (Glivec), una molecola farmacologica capace di inibire in modo specifico l'attività tirosin-chinasica costitutiva dell'oncoproteina BCR-ABL, ha cambiato radicalmente la strategia terapeutica messa in atto per il trattamento della LMC. In recenti studi è stato valutato l'effetto che le delezioni genomiche sul der(9) hanno sulla prognosi dei pazienti affetti da LMC e trattati con l'imatinib; sebbene dati preliminari documentino l'assenza di associazione tra la prognosi e le delezioni sul der(9), al momento questi risultati non possono essere ritenuti definitivi per il fatto che le casistiche dei pazienti studiati non sempre risultavano tra loro omogenee e per il follow-up breve dei casi studiati. Recentemente è stato dimostrato che il sistema di sonde Vysis "dual-color/dual-fusion" è meno efficace nell'identificare le microdelezioni sul der(9) rispetto alle sonde BAC e PAC utilizzate dal nostro gruppo. Infatti la presenza di un "gap" di circa 300 Kb contenuto nella sonda commerciale specifica per il gene *BCR* (Figura 1) non consente il corretto inquadramento dei pazienti portatori di delezioni a carico del cromosoma 22 quando queste sono < 700 kb.

Grazie alla disponibilità di contigui di cloni BAC/PAC che coprono l'intero genoma umano è stato messo a punto un sistema di sonde BAC e PAC in grado di rilevare con efficienza la presenza di delezioni genomiche sul der(9) e di identificare con precisione la presenza di riarrangiamenti cromosomici complessi della t(9;22) (Figura 2).

La caratterizzazione precisa dei breakpoint in ciascuno dei pazienti con traslocazione variante ha permesso di evidenziare, non solo la perdita di sequenze dei cromosomi 9 e/o 22, ma anche delezioni di sequenze appartenenti al terzo cromosoma coinvolto nel riarrangiamento complesso. In generale, le delezioni a carico delle sequenze del cromosoma 9 sono risultate più ampie rispetto a quelle del 22. La massima estensione descritta per le delezioni a carico del 9 e 22 è stata

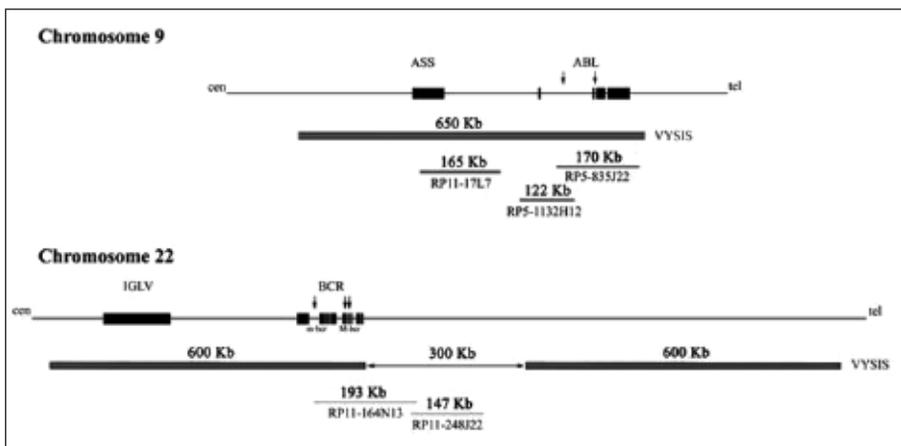


Fig. 1 - Rappresentazione schematica dei cloni specifici per ABL e BCR e della sonda dual color/dual fusion della Vysis.

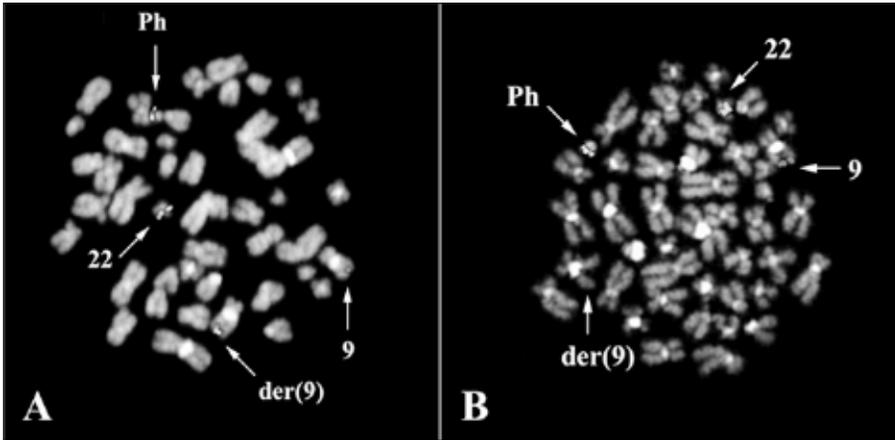


Fig. 2 - Esempi di FISH su metafasi di pazienti con LMC Ph+ impiegando cloni specifici per ABL (RP5-1132H12 e RP5-835J22, in rosso) in coibridazione con quelli specifici per BCR (RP11-164N13, in verde). A) Nei casi di LMC senza delezioni sul der(9) il pattern di FISH risulterà essere in un segnale rosso sul cromosoma 9, un segnale verde sul cromosoma 22 normale e nello splitting dei segnali di ABL e BCR sui due cromosomi derivativi. La simultanea localizzazione del segnale rosso e verde sia sul der(9) che sul cromosoma Ph identifica le 2 regioni di fusione su entrambi i cromosomi derivativi. B) Il pattern di FISH sul der(9) diverso da quello atteso (cioè assenza della colocalizzazione dei due segnali sul derivativo 9) identificherà il caso con la delezione sul der(9).

pari a 41.9 Mb e 12.9 Mb, rispettivamente. Anche le delezioni osservate sul terzo cromosoma derivativo hanno un'estensione variabile da 0,5 Mb a 5,3 Mb.

La consultazione del database genomico dell'Università di Santa Cruz (UCSC) (<http://genome.ucsc.edu>) ha consentito la valutazione dei geni localizzati all'interno delle regioni delete. Infatti, il meccanismo molecolare più probabilmente responsabile dell'associazione tra le delezioni genomiche sul der(9) e la prognosi sfavorevole è costituito dalla perdita di uno o più geni importanti per la progressione della malattia. Nelle regioni delete sono stati identificati quattro TSGs, *PTGES*, *SMARCB1*, *GSTT1* e *PRDM12*.

La perdita di almeno uno di questi TSGs è stata osservata in 55 dei 57 (96,5%) pazienti portatori di delezioni sul der(9). Anche le delezioni sul terzo cromosoma derivativo determinano la perdita di importanti TSGs o geni coinvolti nella progressione del ciclo cellulare. In alcuni casi l'evento cruciale di delezione genica potrebbe interessare le sequenze del terzo cromosoma coinvolto nelle varianti della t(9;22). La possibilità che nella LMC esista la delezione di sequenze genomiche che codificano per microRNA (miRNA) apre un nuovo scenario circa la comprensione dei meccanismi che sono alla base della patogenesi di questa malattia. Queste piccole sequenze di RNA non codificante ma dotate di attività regolatoria della traduzione delle proteine, potrebbero essere la chiave dei meccanismi responsabili della deregolazione proliferativa e differenziativa della cellula emopoietica. Delezioni emi- o omozigoti a livello di 13q14 sono frequenti nella Leucemia Linfatica Cronica (LLC) e ciò suggerisce che in questa regione vi siano uno o più geni oncosoppressori. Recentemente è stato dimostrato che un

microcluster costituito da miR-15a e miR-16-1 mappa all'interno della regione 13q14. Un interessante gene target di questi microcluster di miRNA è il gene antiapoptotico BCL-2. Ad oggi sono ancora in corso gli studi che avranno lo scopo di verificare la presenza di sequenze di miRNA nelle regioni delete sul der(9) e soprattutto di studiare eventuali variazioni quantitative della espressione di questi piccoli RNA. L'integrazione delle informazioni prodotte dalla caratterizzazione citogenetico-molecolare con i dati clinici dei pazienti affetti da LMC consentirà di verificare se la perdita dei miRNA associata alle delezioni sul der(9) ha un significato biologico e prognostico.

Bibliografia

1. Albano F, Specchia G, Anelli L, Zagaria A, Storlazzi CT, Buquicchio C, Roberti MG, Liso V, Rocchi M. Genomic deletions on other chromosomes involved in variant t(9;22) chronic myeloid leukemia cases. *Genes Chromosomes Cancer*. 2003; 36:353-60.
2. Albano F, Specchia G, Anelli L, Zagaria A, Liso A, Liso V, Rocchi M. A novel translocation t(14;15)(q32;q24) bearing deletion on der(14) in Philadelphia-positive chronic myeloid leukemia. *Haematologica*. 2003; 88:1076-7.
3. Albano F, Specchia G, Anelli L, Zagaria A, Archidiacono N, Liso V, Rocchi M. Molecular cytogenetic findings supporting the evidence of a biclonal origin in acute myeloid leukemia. *Annals of Hematology*. 2006; 85: 129-131.
4. Albano F, Anelli L, Zagaria A, Archidiacono N, Liso V, Specchia G, Rocchi M. "Home brew" FISH assay shows higher efficiency than "BCR-ABL dual color, dual fusion" probe in detecting genomic microdeletions and complex rearrangements associated with t(9;22) in chronic myeloid leukemia. *Cancer Genetics & Cytogenetics*, in press. 2007.
5. Calin GA, Ferracin M, Cimmino A, Di Leva G, et al. A microRNA signature associated with prognosis and progression in chronic lymphocytic leukemia. *New England Journal of Medicine* 2005; 353: 1793-1801.
6. Huntly BJ et al. Deletions of the derivative chromosome 9 occur at the time of the Philadelphia translocation and provide a powerful and independent prognostic indicator in chronic myeloid leukemia. *Blood* 2001; 98: 1732-1738.
7. Huntly BJ, Bench A, Green AR. Double jeopardy from a single translocation: deletions of the derivative chromosome 9 in chronic myeloid leukemia. *Blood*. 2003; 102: 1160-8.
8. Huntly BJ, Guillot F, Reid AG, Vassiliou G et al. Imatinib improves but may not fully reverse the poor prognosis of patients with CML with derivative chromosome 9 deletions. *Blood*. 2003; 102: 2205-12.
9. Kolomietz E, Al-Maghrabi J, Brennan S, Karaskova J, et al. Primary chromosomal rearrangements of leukemia are frequently accompanied by extensive submicroscopic deletions and may lead to altered prognosis. *Blood*. 2001; 97: 3581-3588.

10. Quintas-Cardama A, Kantarjian H, Talpaz M, O'Brien S, et al. Imatinib mesylate therapy may overcome the poor prognostic significance of deletions of derivative chromosome 9 in patients with chronic myelogenous leukemia. *Blood*. 2005; 105: 2281-6.
11. Quon KC, Berns A. Haplo-insufficiency? Let me count the ways. *Gene Dev* 2001; 15: 2917-2921.
12. Saglio G, Storlazzi CT, Giugliano E, Surace C, Anelli L, Rege-Cambrin G, Zagaria A, et al. A 76-kb duplication maps close to the *BCR* gene on chromosome 22 and the *ABL* gene on chromosome 9: Possible involvement in the genesis of the Philadelphia chromosome translocation. *PNAS* 2002; 99: 9882-9887.
13. Sinclair PB, Nacheva EP, Laversha M, et al. Large deletions at the t(9;22) breakpoint are common and may identify a poor prognosis subgroup of patients with chronic myeloid leukemia. *Blood* 2000; 95: 738-743.
14. Specchia G, Albano F, Anelli L, Storlazzi CT, Zagaria A, Liso A, Pannunzio A, Pastore D, Mestice A, Greco G, Liso V, Rocchi M.. Derivative chromosome 9 deletions in chronic myeloid leukemia are associated with loss of tumor suppressor genes. *Leuk Lymphoma*. 2004; 45: 689-94.
15. Specchia G, Albano F, Anelli L, Storlazzi CT, Zagaria A, Mancini M, Cuneo A, Pane F, Foa R, Manolelli F, Liso V, Rocchi M. Deletions on der(9) chromosome in adult Ph-positive acute lymphoblastic leukemia occur with a frequency similar to that observed in chronic myeloid leukemia. *Leukemia*. 2003; 17: 528-531.
16. Storlazzi CT, Specchia G, Anelli L, Albano F, Pastore D, Zagaria A, Rocchi M, Liso V. Breakpoint Characterization of der(9) deletions in CML patients. *Genes Chromosomes Cancer* 2002; 35: 271-276.

Inquadramento prognostico delle SMD

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

Divisione di Ematologia, Fondazione IRCCS Policlinico San Matteo, Università di Pavia

Le Sindromi Mielodisplastiche (SMD) sono disordini clonali dell'emopoiesi caratterizzate da un midollo osseo riccamente cellulato che mostra una difettosa maturazione di tutte le linee cellulari con conseguente emopoiesi inefficace e sviluppo di citopenia nel sangue periferico (Heany & Golde, 1999). Il decorso clinico delle SMD è quanto mai eterogeneo e la mediana di sopravvivenza è di circa sei anni se la diagnosi viene fatta precocemente. La maggioranza dei pazienti presenta un progressivo peggioramento della citopenia periferica con facile sviluppo di infezioni e di emorragie, mentre il 30-40% dei pazienti presenta una progressione in LAM di solito resistente alla chemioterapia. Pertanto, sin dalla metà degli anni '80 numerosi studi hanno tentato di individuare i parametri biologici e clinici che potessero predire il decorso clinico delle SMD (Tricot et al., 1985; Tricot et al., 1986; Teerenhovi et al., 1986; Kerkhofs et al., 1987). Tutti questi studi avevano riportato che età, citopenie periferiche e percentuale di blasti midollari erano i soli parametri prognosticamente rilevanti nelle SMD (Aul et al., 2004). Nel 1985 Mufti et al. avevano sviluppato un primo sistema prognostico, chiamato "Bournemouth score", basato sull'esame emocitometrico e sulla morfologia midollare. Il sistema di Bournemouth dava un punteggio ("score") ai seguenti quattro parametri: emoglobina <10g/dl, valore assoluto dei neutrofilici <2.5x10⁹/L, conta piastrinica <100x10⁹/L, percentuale di blasti midollari >5%. La combinazione dei quattro parametri consentiva di individuare tre gruppi di pazienti che presentavano una diversa mediana di sopravvivenza. Il sistema di Bournemouth venne utilizzato da vari Autori sebbene la sua validità prognostica non fosse stata confermata in analisi multivariata. Un secondo sistema prognostico venne elaborato dal gruppo Spagnolo nel 1989. In questa casistica età, percentuale di blasti midollari e conta piastrinica erano i più importanti fattori prognostici. Successivamente, Aul et al. (1992) dimostrarono che i parametri biologici capaci di predire il decorso clinico delle SMD erano una percentuale di blasti midollari >5%, un valore di lattico deidrogenasi >200 Unità Internazionali, un valore di emoglobina <9g/dl ed un valore di piastrine ≤100x10⁹/l. Ad ognuno dei parametri venne dato un punteggio ed in tal modo fu possibile individuare tre

categorie di pazienti a decorso clinico significativamente diverso. Il sistema prognostico elaborato dal gruppo tedesco, chiamato "Dusseldorf score", presentava due vantaggi rispetto allo score di Bournemouth: il valore della lattico deidrogenasi permetteva di identificare con maggior facilità i pazienti con leucemia mielomonocitica cronica (LMMC) e la combinazione dei quattro parametri permetteva di individuare con maggior precisione i pazienti con Anemia Refrattaria con Sideroblasti ad anello/ Anemia Refrattaria (ARS/AR) a breve sopravvivenza e ad alto rischio di trasformazione in LAM. Per questo motivo il sistema prognostico di Dusseldorf identificava con maggior accuratezza i pazienti con SMD ad alto rischio rispetto a tutti gli altri sistemi prognostici. Uno studio eseguito su un campione di 180 pazienti aveva dimostrato che il 32,2% dei pazienti veniva considerato ad alto rischio dal sistema prognostico di Dusseldorf, mentre solo il 13-20% dei pazienti rientrava in tale categoria quando venivano applicati altri sistemi prognostici (Sperr et al., 2001).

Negli stessi anni numerosi studi avevano dimostrato che il 40-70% dei pazienti con SMD *de novo* ed il 95% dei pazienti con SMD secondarie presentano anomalie cromosomiche clonali che identificano particolari entità biologiche e cliniche e che hanno un rilevante impatto sul decorso clinico della malattia (Pierre et al., 1989; Bernasconi et al., 1994; Haase et al., 1995). Tuttavia, solo nel 1993 il gruppo di Lilla sviluppò un nuovo sistema che per la prima volta includeva l'assetto citogenetico tra i parametri che potevano influire sul decorso clinico delle SMD. In base al cariotipo i 235 pazienti entrati nello studio erano stati suddivisi nei seguenti gruppi: cariotipo normale, 5q- come singola anomalia, trisomia 8 (+8) come singola anomalia, -7/7q- come singola anomalia, 20q- come singola anomalia, -Y come singola anomalia, difetti vari, -7/7q- associato ad ulteriori difetti del cariotipo e cariotipo complesso (con ≥ 3 anomalie cromosomiche). In analisi multivariata la sopravvivenza complessiva era significativamente influenzata dalla percentuale di blasti midollari (i pazienti erano stati suddivisi in tre gruppi con percentuali $< 5\%$, tra 5% e 10% , $> 10\%$), dal cariotipo (pazienti a cariotipo normale e a complesso), e dal valore delle piastrine (pazienti con piastrine $> 0 \leq 75 \times 10^9/L$). La combinazione dei tre fattori prognostici permetteva di identificare tre gruppi di pazienti con mediane di sopravvivenza significativamente diverse: 56, 24 e 6 mesi. La probabilità di progressione in LAM era significativamente influenzata dalla percentuale di blasti midollari (pazienti con percentuale $\leq 10\%$ o $> 10\%$) e dal cariotipo (pazienti a cariotipo normale o complesso). La combinazione dei due fattori consentiva d'individuare tre gruppi di pazienti a diverso rischio di progressione in LAM. Dopo due anni di osservazione progressiva in LAM l'8% dei pazienti a basso rischio, il 40% di quelli a rischio intermedio ed il 68% di quelli ad alto rischio. Il maggior limite del sistema prognostico di Lilla consisteva nel fatto che tale casistica non era sufficientemente ampia da consentire di valutare appieno il significativo prognostico del cariotipo dato che alcune anomalie erano state osservate in un numero di pazienti insufficiente per una corretta analisi statistica.

Uno studio successivo sottolineò ulteriormente il valore prognostico dell'assetto cromosomico (Toyama et al., 1993). In questa casistica di 401 pazienti la mono-

somia 7, la delezione del(7q) ed il cariotipo complesso erano le anomalie che identificavano tra i pazienti con ARS/AR quelli a decorso clinico più sfavorevole. Inoltre il cariotipo complesso aveva un impatto sfavorevole anche nei pazienti con anemia refrattaria con eccesso di blasti (AREB) ma non in quelli con AREB in trasformazione (AREB-t). Infatti, un cariotipo con tre o più anomalie comportava nei pazienti con AREB non solo una riduzione dell'intervallo tra diagnosi ed evoluzione in LAM ma anche tra quest'ultima e la morte del paziente. Considerando i dati riportati dalle casistiche sopra elencate, nel 1994 fu creato un gruppo internazionale di esperti che doveva identificare i fattori biologici e clinici capaci di predire il più efficacemente possibile la prognosi dei pazienti con SMD. Inoltre, doveva essere accuratamente valutata l'effettiva rilevanza prognostica del cariotipo. La casistica comprendeva un gruppo di 800 pazienti poco o per nulla trattati con assetto citogenetico definito mediante analisi cromosomica convenzionale. Il gruppo internazionale suddivise i pazienti in tre categorie citogenetiche: basso rischio che includeva i pazienti con cariotipo normale, con 5q- come singola anomalia, con 20q- come singola anomalia e con perdita del cromosoma Y come singola anomalia; rischio intermedio che includeva i pazienti con trisomia 8 e con vari difetti sia singoli che doppi del cariotipo; alto rischio che includeva pazienti con monosomia o delezione delle braccia lunghe del cromosoma 7 (-7/7q-) e pazienti con cariotipo complesso (con un numero di anomalie ≥ 3). I pazienti del primo gruppo erano circa il 70% dell'intera casistica, quelli del secondo gruppo circa il 14% e quelli della categoria ad alto rischio citogenetico il 16% circa. I pazienti del primo gruppo presentavano una sopravvivenza mediana di 3,8 anni, quelli del secondo gruppo di 2,4 anni e quelli del terzo gruppo di 0,8 anni. Il 25% dei pazienti di ciascun gruppo evolveva in LAM a distanza di 5,6, 1,6 e 0,9 anni dalla diagnosi. In analisi multivariata i parametri biologici e clinici più predittivi del decorso clinico erano tre: cariotipo, numero e tipo di citopenie periferiche ($Hb < 10g/dl$, conta assoluta dei neutrofili $< 1,8 \times 10^9/l$ e valore delle piastrine $< 100 \times 10^9/l$) e percentuale di blasti midollari. Successivamente, il gruppo di esperti elaborò un nuovo sistema prognostico, l'"International Prognostic Score" (IPSS), basato sui parametri sopra riportati (Greenberg et al., 1997). Ad ognuno dei tre parametri fu dato un punteggio: da 0 a 0,5 per il numero di citopenie, da 0 a 1,0 per la categoria di rischio citogenetico e da 0 a 2,0 per la percentuale di blasti midollari. Combinando tra loro i tre diversi fattori prognostici si individuarono quattro categorie di pazienti a decorso clinico significativamente diverso: basso rischio (punteggio =0), rischio intermedio 1 (punteggio =0,5-1,0), rischio intermedio 2 (punteggio =1,5-2,0) ed alto rischio (punteggio $\geq 2,5$). Le mediane di sopravvivenza erano significativamente diverse nelle quattro categorie di pazienti: 5,7 anni per la categoria a basso rischio, 3,5 anni per quella a rischio intermedio 1, 1,2 anni per quella a rischio intermedio 2 e 0,4 anni per quella ad alto rischio. Inoltre il 25% dei pazienti di ciascuna categoria aveva sviluppato una LAM a distanza di 9,4, 3,3, 1,1 e 0,2 anni dalla diagnosi. Il punteggio IPSS stressava l'importanza dell'assetto cromosomico del paziente mielodisplastico. Infatti, l'inclusione del dato citogenetico migliorava la capacità dell'IPSS di predire la sopravvivenza ed il rischio di evo-

luzione in LAM rispetto ad altri sistemi prognostici, mentre la mancanza del dato citogenetico impediva una corretta stratificazione prognostica dei pazienti, specialmente di quelli a rischio intermedio-1 ed intermedio-2.

Queste osservazioni sono state ribadite da studi successivi che hanno dimostrato la rilevanza prognostica delle categorie citogenetiche definite dall'IPSS anche nei pazienti sottoposti a chemioterapia intensiva, trapianto autologo e trapianto allogenico. Solo uno studio aveva riportato che il sesso del paziente riduceva la rilevanza prognostica del cariotipo (Mauritzon et al., 2001). In questa casistica comprendente 389 SMD vi era una buona corrispondenza tra categoria citogenetica proposta dall'IPSS e sopravvivenza, ma età, sesso, sottotipo morfologico ed abitudine al fumo riducevano significativamente l'impatto della citogenetica. Infatti, i soggetti di sesso maschile che appartenevano alla categoria a rischio basso ed intermedio avevano mostrato una sopravvivenza statisticamente più breve rispetto a quella dei soggetti di sesso femminile appartenenti alle stesse categorie di rischio. Per i pazienti a basso rischio tale differenza era totalmente dovuta alla maggiore incidenza di sindrome del 5q- e ad un minor consumo di sigarette nei soggetti di sesso femminile, ma per i pazienti a rischio intermedio il diverso decorso clinico non sembrava essere causato da alcun particolare fattore sia clinico che biologico.

Gli stessi studi che confermavano la validità prognostica dell'IPSS, hanno però anche individuato i principali limiti di tale sistema prognostico che consistevano principalmente nell'assoluta necessità del dato citogenetico e nell'inclusione di tutta una serie di anomalie ad impatto ancora sconosciuto sulla sopravvivenza e sul rischio di progressione in LAM nella categoria citogenetica a prognosi intermedia. Un primo studio che cercò di superare quest'ultimo limite fu quello di Pfeilstöcker e Collaboratori (1999) che suddivise i pazienti in cinque categorie: cariotipo normale; cariotipo anormale ma senza anomalie complesse e senza anomalie dei cromosomi 5,7 e 8; qualsiasi anomalia del cromosoma 5; qualsiasi anomalia del cromosoma 7; qualsiasi anomalia del cromosoma 8, cariotipo complesso (≥ 3 anomalie). Gli Autori riportarono che la loro nuova suddivisione delle alterazioni citogenetiche, denominata "Prognostic Index Cytogenetics", migliorava significativamente la stratificazione prognostica dei pazienti con SMD. Un ulteriore tentativo di meglio definire l'impatto del cariotipo fu eseguito da Solé et al. (2000) che comunque ribadirono la validità prognostica dell'IPSS. Il principale obiettivo di questo studio era stato quello di analizzare l'effettiva rilevanza clinica delle alterazioni cromosomiche incluse nella categoria citogenetica che l'IPSS aveva considerato a rischio intermedio. Nella loro casistica che comprendeva 640 pazienti la delezione delle braccia corte di un cromosoma 12, [del(12p)], si associava ad un decorso clinico del tutto sovrapponibile a quello dei pazienti appartenenti alla categoria a basso rischio citogenetico, mentre la trisomia per le braccia lunghe del cromosoma 1, [+1q], e la trisomia 8 comportavano una sopravvivenza peggiore rispetto a quella dei pazienti considerati a rischio citogenetico intermedio. In particolare la trisomia 8 si associava ad una mediana di sopravvivenza di circa un anno ed ad un rischio di trasformazione in LAM ad un anno dalla diagnosi del 34%.

Un altro studio aveva cercato di definire più esattamente il significato prognostico della delezione 5q (Giagounidis et al., 2004). In questa casistica i pazienti con 5q- come singola anomalia presentavano una mediana di sopravvivenza significativamente migliore di quella dei pazienti con 5q- associato ad altre alterazioni del cariotipo (146 mesi versus 45 mesi). Pertanto, sul piano statistico la sopravvivenza dei pazienti con 5q- isolato era del tutto sovrapponibile a quella dei pazienti con punteggio IPSS=0. La sopravvivenza dei pazienti con 5q- isolato non era influenzata dal sesso, dai punti di rottura lungo il cromosoma 5 e dalla gravità degli aspetti displastici, ma invece era esclusivamente determinata dalla percentuale di blasti presenti nel midollo osseo. Infatti, i pazienti con 5q- e percentuale di blasti midollari >5% mostravano una sopravvivenza mediana di 24 mesi, significativamente peggiore rispetto a quella dei pazienti con 5q- e percentuale di blasti midollari ≤5%. Il rischio di progressione in LAM era dell'80% per i pazienti del primo gruppo e del solo 9% per quelli del secondo gruppo.

Queste osservazioni erano state riprese da una casistica più recente che comprendeva 331 pazienti (Bernasconi et al., 2005). Questo studio aveva analizzato anche altre anomalie del cariotipo. Era stato così osservato che i riarrangiamenti delle braccia lunghe del cromosoma 3 (anomalie 3q), incluse dall'IPSS nella categoria a rischio citogenetico intermedio, si associavano invece ad un decorso clinico del tutto sovrapponibile a quello dei pazienti appartenenti alla categoria ad alto rischio citogenetico. Inoltre, in analisi uni- e multivariata i pazienti con delezione del(7)(q31q35) presentavano una sopravvivenza migliore rispetto a quella dei pazienti appartenenti alla categoria citogenetica ad alto rischio, ma quando i pazienti con tale delezione venivano considerati come una categoria citogenetica a se stante la stratificazione prognostica dei pazienti non migliorava. Inoltre, in questa casistica la rilevanza del cariotipo complesso (≥3 anomalie) era indipendente dall'entità della blastosi midollare.

Più recentemente Solè et al. (2005) hanno tentato di ridefinire le categorie citogenetiche dell'IPSS nel tentativo di migliorare ulteriormente la capacità predittiva della citogenetica. I 968 pazienti di questa casistica sono stati suddivisi in quattro categorie citogenetiche. La categoria a prognosi favorevole comprendeva oltre al cariotipo normale, al 5q- isolato, al 20q- isolato e alla perdita del cromosoma Y, il 12p- isolato e l'11q- isolato; la categoria a prognosi intermedia comprendeva i riarrangiamenti 3q21q26, la trisomia 8, la trisomia 9, le traslocazioni 11q e le delezioni 17p; la categoria a prognosi sfavorevole comprendeva anomalie cromosomiche rare considerate a rischio intermedio dall'IPSS. Veniva quindi introdotta una categoria di pazienti a prognosi ancor più sfavorevole che comprendeva le SMD con -7/7q-, con isocromosoma per le braccia lunghe di un numero 17, e con cariotipo complesso (≥3 anomalie). In questa casistica i parametri che maggiormente influenzavano la sopravvivenza erano età, emoglobina <10g/dl, numero di citopenie, sottotipo FAB e nuova classificazione citogenetica elaborata dagli Autori, mentre la progressione in LAM era significativamente influenzata da emoglobina <10g/dl, sottotipo FAB e nuove categorie citogenetiche. La combinazione di questi fattori prognostici, a cui era stato dato un punteggio, permetteva di individuare quattro classi di pazienti a decorso clinico statisticamente diverso. Gli

Autori concludevano che la migliore suddivisione prognostica dei pazienti era stata raggiunta proprio grazie ad una più accurata definizione della rilevanza clinica delle anomalie cromosomiche incluse dall'IPSS nella categoria citogenetica a rischio intermedio. Una casistica ancora più recente (Bernasconi et al., 2007), comprendente 491 pazienti, ha confermato in parte le osservazioni di Solé et al. (2005). Anche in questa casistica la delezione 11q e 12p si associavano ad un decorso clinico del tutto sovrapponibile a quello della categoria a basso rischio citogenetico secondo l'IPSS, le anomalie 3q si associavano ad una sopravvivenza statisticamente più breve rispetto a quella della categoria a rischio citogenetico intermedio e simile a quella della categoria citogenetica ad alto rischio, la delezione del(7)(q31q35) ad una sopravvivenza statisticamente migliore rispetto a quella della categoria ad alto rischio citogenetico e simile a quella della categoria a rischio citogenetico intermedio, la delezione 20q ad un rischio di evoluzione in LAM significativamente più alto rispetto a quello della categoria a basso rischio citogenetico e simile a quello della categoria a rischio intermedio, la trisomia 8 ad un rischio di progressione del tutto sovrapponibile a quello della categoria ad alto rischio citogenetico. Tenendo conto di tutte queste osservazioni gli Autori avevano ridefinito le categorie citogenetiche dell'IPSS ed avevano dimostrato che le nuove categorie citogenetiche avevano lo stesso potere predittivo delle categorie citogenetiche dell'IPSS per quanto riguarda la sopravvivenza ma erano maggiormente predittive del rischio di evoluzione in LAM.

Il gruppo tedesco ha tentato di superare l'altro importante limite dell'IPSS e cioè l'assoluta disponibilità del dato citogenetico. Per raggiungere tale obiettivo Germing et al. (2005) hanno introdotto una nuova variabile costituita dai valori della lattico deidrogenasi. Alti valori di lattico deidrogenasi individuavano pazienti a decorso clinico estremamente sfavorevole anche nei gruppi definiti a rischio basso o intermedio dall'IPSS. Altri Autori hanno tentato di superare i limiti dell'IPSS combinando tale sistema prognostico con la classificazione WHO.

Evoluzione del cariotipo

Consiste nello sviluppo di una nuova popolazione clonale marcata da una particolare alterazione citogenetica se il cariotipo iniziale era normale o nello sviluppo di nuove popolazioni clonali tutte originate da quella iniziale di solito marcata da un'anomalia isolata del cariotipo. Dal punto di vista clinico la maggior parte dei pazienti con SMD sviluppa un quadro di insufficienza midollare che porta a morte il paziente, la metà dei pazienti progredisce in LAM e pochi muoiono per patologie intercorrenti. La malattia può quindi evolvere secondo tre diverse modalità: progressivo peggioramento della citopenia e graduale aumento della blastosi midollare; decorso clinico stabile con improvviso sviluppo di LAM; decorso clinico stabile senza variazioni nella percentuale di blasti midollari. Il primo gruppo di pazienti mostra di solito un assetto citogenetico stabile e sviluppa più frequentemente anomalie molecolari (Es.: sovra-espressione di Flt3); il secondo gruppo di pazienti presenta popolazioni clonali marcate da un cariotipo complesso o da nuove anomalie cromosomiche numeriche (più spesso trisomie);

il terzo gruppo di pazienti mostra un cariotipo decisamente stabile. Comunque sia, pochissimi sono gli studi che hanno valutato quale sia l'incidenza di evoluzione cromosomica nei pazienti con SMD. Anche se il periodo di osservazione di queste casistiche è estremamente breve il 60% dei pazienti con evoluzione clonale progredisce in LAM e la prognosi è particolarmente sfavorevole per coloro che sviluppano nuove popolazioni clonali in meno di 100 giorni.

Utilità della citogenetica molecolare (FISH)

La FISH non necessita che la cellula in studio sia in mitosi e proprio per questa sua caratteristica potrebbe essere impiegata per valutare se i pazienti con SMD a cariotipo normale possano in realtà presentare anomalie citogenetiche non dimostrabili dall'analisi cromosomica convenzionale. Si tratta di un gruppo di pazienti considerati a citogenetica favorevole dall'IPSS ma a decorso clinico quanto mai eterogeneo e non sempre favorevole, visto che alcuni possono rapidamente progredire in LAM. La capacità della FISH di svelare anomalie cromosomiche criptiche in questi pazienti era già stata dimostrata da alcuni studi, ma altri avevano sottolineato che la metodica identificava alterazioni del cariotipo con la stessa frequenza della citogenetica convenzionale che quindi doveva essere considerata un'ottima metodica per l'analisi dei pazienti con SMD, mentre la FISH aveva solo un ruolo ancillare. Altre casistiche avevano osservato che la FISH è comunque in grado di dimostrare popolazioni clonali minori marcate da monosomia 7 in una discreta percentuale di pazienti con SMD a cariotipo normale. Basandosi su queste osservazioni due studi avevano analizzato 108 e 57 pazienti ad assetto citogenetico normale con sonde per le anomalie specificamente presenti nei pazienti con SMD (Rigolin et al., 2001; Bernasconi et al., 2003). L'incidenza di anomalie criptiche era stata del 15-18% e queste erano state più frequentemente osservate nei pazienti con SMD a rischio IPSS intermedio-alto ed in quelli con elevata blastosi midollare. Alla diagnosi la popolazione clonale dimostrata dalla FISH era costituita da un piccolo numero di cellule, ma durante il follow-up tale popolazione cellulare si espandeva progressivamente. Entrambi gli studi avevano sottolineato che i difetti criptici del cariotipo erano svelati dalla sola FISH sia perché di dimensioni al di sotto del potere di risoluzione dell'analisi cromosomica convenzionale sia perché il clone displastico era prevalentemente costituito da cellule quiescenti. Sul piano prognostico la presenza di anomalie svelate dalla FISH si associava ad un più breve sopravvivenza ed ad un più elevato rischio di evoluzione in LAM. Una casistica sottolineava che nei pazienti con FISH anormale il tasso di mortalità ed il tasso di progressione in LAM erano 5 e 18 volte quelli che si osservavano nei pazienti con FISH normale. Il rilevante impatto clinico della FISH era ulteriormente sottolineato dall'osservazione che l'anomalia criptica peggiorava la categoria di rischio IPSS in cinque pazienti. Altre casistiche avevano però riscontrato una più bassa incidenza di difetti cromosomici criptici. Uno studio aveva riportato che un solo paziente dei trentadue analizzati presentava un'anomalia clonale alla FISH ed aveva sottolineato che le lesioni geniche responsabili delle SMD a cariotipo normale potrebbero essere individuate

solo impiegando tecniche molecolari molto più sofisticate della FISH (Ketterling et al., 2002). Altre casistiche avevano sottolineato che la diversa incidenza di anomalie tra analisi cromosomica convenzionale e FISH dipendeva dal fatto che la prima metodica individuava anomalie non rivelate dalla FISH proprio perché quest'ultima impiega sonde che non analizzano tutte le regioni cromosomiche coinvolte nelle SMD (Beyer et al., 2004). L'effettiva potenzialità della FISH potrà essere appieno dimostrata solo in futuro quando sarà disponibile un maggior numero di sonde. Per il momento la FISH dovrebbe essere impiegata insieme alla citogenetica convenzionale per definire con più precisione possibile l'assetto cromosomico del paziente. Più recentemente uno studio comprendente 529 pazienti ha dimostrato che la probabilità che il paziente con SMD abbia effettivamente un cariotipo normale è più alta se vengono compiutamente analizzate non meno di diciannove metafasi (Steidl et al., 2005). Infatti i pazienti con più di venti metafasi normali mostrano un'incidenza di cariotipi anormali statisticamente più bassa rispetto a quella dei pazienti con meno di diciannove metafasi analizzate (47% versus 56%). Questa casistica ha perciò confermato le linee guida per l'esame citogenetico applicate da molti studi internazionali e ha raccomandato che nei pazienti con SMD il cariotipo venga considerato normale solo dopo aver esaminato almeno venti metafasi.

Assetto citogenetico e terapia

È stato osservato che l'assetto citogenetico identifica sottogruppi di pazienti particolarmente sensibili ad una data terapia. Un cariotipo normale e soprattutto un'alta percentuale di cellule eritroidi a cariotipo normale sono predittivi di risposta all'eritropoietina, forse dovuta ad una riduzione dello sbilanciamento tra proteine pro- ed anti-apoptotiche. Questa osservazione è stata raggiunta con l'analisi di tredici pazienti a cariotipo misto, cioè con metafasi normali ed anormali. Alla diagnosi l'esame cromosomico convenzionale aveva dimostrato che i pazienti responsivi all'eritropoietina presentavano una percentuale di mitosi normali significativamente superiore rispetto a quella dei pazienti non responsivi e la FISH dimostrava che questa diversità era presente anche a livello delle cellule eritroidi e mieloidi. Il trattamento con eritropoietina riduceva ulteriormente il numero delle mitosi anormali alla citogenetica convenzionale ed la percentuale di cellule eritroidi e mieloidi anormali alla FISH (Rigolin et al., 2002). Uno studio successivo aveva analizzato le cellule CD34 positive di pazienti parzialmente responsivi o non più responsivi all'eritropoietina, somministrata insieme al fattore di crescita granulocitario (Rigolin et al., 2004). La percentuale di cellule CD34 positive anormali era significativamente inferiore nei pazienti responsivi rispetto a quelli non responsivi e durante il trattamento si riduceva ulteriormente. La risposta al trattamento si correlava con una riduzione delle cellule CD34 positive apoptotiche, ma non si osservava alcuna differenza nella lunghezza dei telomeri dei neutrofili tra pazienti responsivi e non responsivi. Questa osservazione indicava che nei pazienti responsivi l'eritropoietina ed il fattore di crescita inducevano la proliferazione di cellule CD34 positive a cario-

tipo normale ma con importanti deficit funzionali e con telomeri che avevano già subito un processo di erosione.

Nel 2005 List et al. avevano riportato che la delezione 5q- identifica SMD particolarmente sensibili al trattamento con talidomide e con il suo derivato 4-amino-glutamidico, più potente ma meno tossico e teratogeno. Lo studio aveva riportato che il trattamento con una dose di 25 mg o 10 mg al giorno per ventun giorni consecutivi e con intervallo libero di ventotto giorni determinava un'indipendenza dalla terapia di supporto trasfusionale nel 56% dei pazienti. La percentuale di risposta era più alta nei pazienti con 5q- isolato e nei pazienti a basso rischio IPSS. In particolare, una risposta veniva raggiunta nell'83% dei pazienti con 5q-, nel 56% dei pazienti a cariotipo normale e nel 12% dei pazienti con altre anomalie del cariotipo. Il tempo mediano per raggiungere una risposta era significativamente più breve nei pazienti con 5q- rispetto a quelli con cariotipo normale o con altre anomalie del cariotipo (8 verso 11 settimane). Undici dei venti pazienti a cariotipo anormale avevano presentato un dimezzamento della percentuale di metafasi anormali, dieci (nove con 5q-) avevano raggiunto la remissione citogenetica dopo un tempo mediano di 8 settimane. La durata mediana di risposta alla terapia non era stata ancora raggiunta dopo un tempo mediano di osservazione di 81 settimane ed il livello mediano di emoglobina era di 13,2g/dl. Inoltre una risposta veniva raggiunta nel 68% dei pazienti a basso rischio IPSS, nel 50% di quelli a rischio intermedio 1 e nel 20% di quelli a rischio intermedio 2 e ad alto rischio. La terapia determinava inoltre una completa risoluzione della dismegacariocitopoiesi midollare e nei pazienti con ARS una significativa riduzione del numero di sideroblasti ad anello. Quest'ultima osservazione suggeriva che il derivato della talidomide potesse determinare una soppressione clonale selettiva della popolazione marcata dall'anomalia cromosomica con riespansione della popolazione eritroide normale residua. In questo primo studio le più frequenti complicanze del trattamento erano state la neutropenia e la piastrinopenia che avevano richiesto la sospensione del trattamento o una riduzione del dosaggio del farmaco nel 58% dei pazienti. Uno studio più recente che comprendeva 148 pazienti, aveva riportato una risposta eritroide definita maggiore o completa secondo i criteri del "Working Group" internazionale nel 66% dei pazienti, la maggior parte con 5q- (List, 2005). Inoltre, dopo un periodo osservazionale di 9.3 mesi la durata mediana della risposta non era stata ancora raggiunta.

Nel 2001 Lübbert et al. avevano riportato che la decitabina induceva una risposta citogenetica completa o maggiore secondo i criteri già applicati ai pazienti con leucemia mieloide cronica in diciannove dei 69 pazienti valutabili. La decitabina e la 5-azacitidina sono due antimetaboliti fase S specifici che una volta incorporati nel DNA agiscono da potenti inibitori della mutilazione formando composti covalenti con la DNA metil-transferasi 1. Nella casistica sopra riportata le risposte venivano ottenute dopo un numero medio di tre cicli di terapia ed avevano una durata mediana di 7,5 mesi. Considerando il rischio citogenetico, le percentuali di risposta erano state 60% nei pazienti a basso rischio, 20% in quelli a rischio intermedio e 38% in quelli ad alto rischio. Le rispettive mediane di sopravvivenza erano state di 30, 8 e 13 mesi. I pazienti responsivi presentavano un rischio di

morte significativamente più basso rispetto a quelli resistenti alla terapia. I pazienti che avevano raggiunto la risposta citogenetica mostravano una mediana di sopravvivenza di 24 mesi mentre quelli non responsivi di 11 mesi. Uno studio successivo che aveva utilizzato l'azacitidina aveva riportato che il farmaco determinava un maggior numero di risposte cliniche ed un allungamento del tempo di progressione in LAM (Silverman et al., 2002). Tuttavia la sopravvivenza globale era del tutto sovrapponibile a quella raggiunta con la sola terapia di supporto. Un'altra casistica ha confermato questi risultati e ha indicato che l'efficacia terapeutica dell'azacitidina è maggiore nei pazienti ad alto rischio con cariotipo complesso (Wijermans et al., 2000). Infatti, la percentuale di risposta in questo gruppo di pazienti era stata del 64%, mentre per quelli a rischio intermedio 1 ed a rischio intermedio 2 era stata del 48% e 25% rispettivamente.

Bibliografia

1. Heany ML, Golde DW. Myelodysplasia. *New Engl. J Med* 1999; 340: 1649-1653.
2. Tricot G, Vlietinck R, Boogaerts MA, et al. Prognostic factors in the myelodysplastic syndromes: importance of initial data on peripheral blood counts, bone marrow cytology, trephine biopsy and chromosomal analysis *Brit J Haematol* 1985; 60: 19-32.
3. Tricot G, Vlietinck R, Verwilghen RL. Prognostic factors in the myelodysplastic syndromes: a review. *Scand J Haematol* 1986; 36 (Suppl 45): 107-113.
4. Teerenhovi L, Lintula R. Natural course of myelodysplastic syndromes - Helsinki experience. *Scand J Haematol* 1986; 36(Suppl 45): 102-106.
5. Kerkhofs H, Hermans J, Haak HL, et al. Utility of the FAB classification for myelodysplastic syndromes: investigation of prognostic factors in 237 cases. *Brit J Haematol* 1987; 65: 73-81.
6. Aul C, Giagounidis A, Germing U, Ganser A. Evaluating the prognosis of patients with myelodysplastic syndromes. *Ann Hematol.* 2004; 81: 485-497.
7. Mufti GJ, Stevens JR, Oscier DG, Hamblin TJ, Machin D. Myelodysplastic Syndromes: a scoring system with prognostic significance. *Brit. J Haemat.* 1985; 59: 425-433.
8. Sanz GF, Sanz MA, Vallespi T, et al. Two regression models and a scoring system for predicting survival and planning treatment in myelodysplastic syndromes: a multivariate analysis of prognostic factors in 370 patients. *Blood* 1989; 74: 395-408.
9. Aul C, Gattermann N, Heyll A, et al Primary myelodysplastic syndromes: Analysis of prognostic factors in 235 patients and proposal for an improved scoring system. *Leukemia* 1992; 6: 52-59.
10. Sperr WR, Wimazal F, Kundi M, et al. Survival analysis and AML development in patients with de novo myelodysplastic syndromes: comparison of six different prognostic scoring systems. *Ann Hematol* 2001; 80: 272-277.
11. Pierre RV, Catovsky D, Mufti GJ, et al. Clinical-cytogenetic correlations in myelodysplasia (preleukemia). *Cancer Gen. Cytogen.* 1989; 40: 149-161.

12. Bernasconi P, Alessandrino EP, Boni M, et al. Karyotype in myelodysplastic syndromes: Relations to morphology, clinical evolution, and survival. *Am. J Hematol.* 1994; 46: 270-277.
13. Haase D, Fonatsch C, Freund M, Wörmann R, Bodenstern H, Bartels H, Stollmann-Gibbels B, Lengfelder E. Cytogenetic findings in 179 patients with myelodysplastic syndromes. *Ann. Hematol.* 1995; 70: 171-187.
14. Toyama K, Ohyashiki K, Yoshida Y, et al. Clinical implications of chromosomal abnormalities in 401 patients with myelodysplastic syndromes: a multicenter study in Japan. *Leukemia* 1993; 7: 499-508.
15. Greenberg P, Cox C, Le Beau MM, et al. International scoring system for evaluating prognosis in myelodysplastic syndromes. *Blood* 1997; 89: 2079-2088.
16. Estey E, Thal P, Beran M, et al. Effect of diagnosis (refractory anemia with excess of blasts, refractory anemia with excess of blasts in transformation or acute myeloid leukaemia [AML]) on outcome of AML-type chemotherapy. *Blood* 1997; 90: 2969-2997.
17. Oosterveld M, Wittebol SH, Lemmens WAJG, et al. the impact of intensive antileukemic treatment strategies on prognosis of myelodysplastic syndrome patients aged less than 61 years according to International Prognostic Scoring System risk group. *Brit J Haematol* 2003; 123: 81-89.
18. Nevill TJ, Fung HC, Sheperd JD, et al. Cytogenetic abnormalities in primary myelodysplastic syndromes are highly predictive of outcome after allogeneic bone marrow transplantation. *Blood* 1998; 92: 1910-1917.
19. Mauritzson N, Johansson B, Rylander L, et al. The prognostic impact of karyotypic subgroups in myelodysplastic syndromes is strongly modified by sex. *Brit J Haematol.* 2001; 113: 347-356.
20. Pfeilstöcker M, Reisner E, Nösslinger T, et al. Cross-validation of prognostic scores in myelodysplastic syndromes on 386 patients from a single institution confirms importance of cytogenetics. *Brit. J Haematol.* 1999; 106: 455-463.
21. Solè F, Espinet B, Sanz GF, et al. Incidence, characterization and prognostic significance of chromosomal abnormalities in 640 patients with primary myelodysplastic syndromes. *Brit. J Haematol.* 2000; 108: 346-356.
22. Giagounidis AAN, Germing U, Haase S, et al. Clinical, morphological, cytogenetic, and prognostic features of patients with myelodysplastic syndromes and del(5q) including band q31. *Leukemia* 2004; 18: 113-119.
23. Bernasconi P, Klersy C, Boni M. et al. Incidence and prognostic significance of karyotype abnormalities in de novo primary myelodysplastic syndromes: a study on 331 patients from a single institution. *Leukemia* 2005; 19: 1424-1431.
24. Solè F, Luno E, Sanzo C, et al. Identification of novel cytogenetic markers with prognostic significance in a series of 968 patients with primary myelodysplastic syndromes. *Haematologica* 2005; 90: 1168-1178.
25. Bernasconi P, Klersy C, Boni M. et al. WHO classification in combination with cytogenetic markers improves the prognostic stratification of patients with de novo primary myelodysplastic syndromes (MDS). *Brit J Haematol*, 2007 (in press).

26. Germing U, Hildebrandt B, Pfeilstocker M, et al. Refinement of the international prognostic system (IPSS) by including LDH as an additional prognostic variable to improve risk assessment in patients with primary myelodysplastic syndromes (MDS). *Leukemia* 2005; 19: 2223-2231.
27. Rigolin GM, Bigoni R, et al. Clinical importance of interphase cytogenetics detecting occult chromosome lesions in myelodysplastic syndromes with normal karyotype. *Leukemia* 2001; 15: 1841-1847.
28. Bernasconi P, Cavigliano PM, Boni M, et al. Is FISH a relevant prognostic tool in myelodysplastic syndromes with a normal chromosome pattern on conventional cytogenetics? A study on 57 patients. *Leukemia* 2003; 17: 2107-2112.
29. Ketterling RP, Wyatt WA, VanWier SA, et al. Primary myelodysplastic syndrome with normal cytogenetics: utility of "FISH panel testing" and M-FISH. *Leuk Res.* 2002; 26: 235-240.
30. Beyer V, Castagnè C, Muhlematter D, et al. Systematic screening at diagnosis of -5/del(5)(q31),-7, or chromosome 8 aneuploidy by interphase fluorescence in situ hybridisation in 110 acute myelocytic leukaemia and high-risk myelodysplastic syndrome patients: concordance and discrepancies with conventional cytogenetics. *Cancer Gen Cytogen* 2004; 152: 29-41.
31. Steidl C, Steffens R, Gassmann W, et al. Adequate cytogenetic examinations in myelodysplastic syndromes: analysis of 529 patients. *Leuk res* 2005; 29: 987-993.
32. Rigolin GM, Della Porta M, Bigoni R, et al. rHuEpo administration in patients with low-risk myelodysplastic syndromes: evaluation of erythroid precursors' response by fluorescence in situ hybridisation on May-Grunwald-Giemsa-stained bone marrow samples. *Brit j Haematol.* 2002; 119: 652-659.
33. Rigolin GM, Della Porta M, Ciccone M, et al. In patients with myelodysplastic syndromes response to rHuEpo and G-CSF treatment is related to an increase of cytogenetically normal CD34+ cells. *Brit J Haematol.* 2004; 126:501-507.
34. List A, Kurtin S, Rose DJ, et al. Efficacy of lenalidomide in myelodysplastic syndromes. *New Engl J Med* 2005; 352: 549-557.
35. List A. Emerging data on ImiDs in the treatment of myelodysplastic syndromes. *Sem Oncol* 2005; 32 (4 suppl 5: S31-35).
36. Lubbert M, Wijermans P, Kunzmann R, et al. Cytogenetic response in high-risk myelodysplastic syndrome following low-dose treatment with the DNA methylation inhibitor 5-aza-2'-deoxycytidine. *Brit J Haematol.* 2001; 114: 349-357.
37. Silverman LR, Demakos EP, Peterson BL et al. Randomized controlled trial of azacitidine in patients with the myelodysplastic syndrome: a study of the cancer and leukemia study group B. *J Clin Oncol.* 2002; 20: 2429-2440.
38. Wijermans P, Lubbert M, Verhoef G, et al. Low-dose 5-aza-2'-deoxycytidine, a DNA hypomethylating agent, for the treatment of high-risk myelodysplastic syndrome a multicenter phase II study in elderly patients. *J Clin Oncol.* 2000; 18: 956-962.

L'apporto della FISH alla prognosi della LLC

Antonio Cuneo, Francesco Cavazzini, Gian Matteo Rigolin

Dipartimento di Scienze Biomediche e Terapie Avanzate, Sezione di Ematologia, Arcispedale Sant'Anna, Università di Ferrara

La biologia e la clinica della LLC presentano aspetti ampiamente eterogenei che riflettono una patogenesi complessa, che riconosce diversità in alcune caratteristiche fondamentali della cellula d'origine e che si sviluppa attraverso una serie di lesioni citogenetiche e molecolari.

In termini moderni l'analisi della LLC si incentra sulla applicazione di una serie di tecniche di citogenetica classica e molecolare che, assieme alle altre indagini diagnostiche, consentono di meglio definire il profilo clinicobiologico di ciascun paziente (1, 2).

Principi metodologici

Analisi citogenetica convenzionale

Lo studio del cariotipo rappresenta l'unica metodica che potenzialmente permette la visualizzazione di qualunque anomalia cromosomica in un singolo esperimento. Le cellule devono essere coltivate a fresco e sottoposte a stimolazione in vitro con mitogeni B-linfocitari (esteri del forbolo, lipopolisaccaride di E. Coli). L'introduzione della metodica di stimolazione che impiega gli oligodeossinucleotidi consente di rinvenire metafasi anomale nell'80% dei casi (2).

Ibridazione in situ fluorescente (FISH)

Questa metodica, fornisce risultati riproducibili e ha il grande pregio della sensibilità, della relativa rapidità e semplicità (3). Si esegue su campioni preparati e fissati per l'analisi citogenetica convenzionale; con opportuni accorgimenti può essere eseguita su cellule morfologicamente intatte o marcate con metodica immunologica.

È nota come FISH, abbreviazione dall'inglese Fluorescence In Situ Hybridization. Mediante l'impiego di sonde fluorescenti specifiche, permette la visualizzazione di sequenze di DNA bersaglio e rileva quindi la presenza di delezioni, trisomie o traslocazioni di specifiche regioni cromosomiche. Il vantaggio rispetto all'analisi citogenetica convenzionale è duplice in quanto:

- a) visualizza le lesioni anche nei nuclei in interfase, risolvendo il problema dell'induzione della mitosi in vitro nelle cellule della LLC;
- b) rileva delezioni molto sottili, dell'ordine di grandezza delle centinaia di kilobasi.

Il suo limite è che, normalmente, permette la visualizzazione di una o due anomalie per esperimento e non può fornire un quadro completo delle lesioni presenti nel paziente.

Comparative genomic hybridization (CGH) e CGH microarrays

Questa metodica si fonda sul principio della ibridazione su cromosomi normali di eguali quantità di DNA tumorale e DNA normale, marcati con fluorocromi diversi. La prevalenza di uno dei due fluorocromi in caso di delezioni o trisomie nel campione patologico consente di individuare qualunque guadagno o perdita di materiale cromosomico in un singolo esperimento. Il vantaggio deriva dal fatto che questa metodica non dipende dall'entrata in mitosi delle cellule patologiche; il limite è rappresentato dall'impossibilità di visualizzare le anomalie "bilanciate" (traslocazioni, inversioni, inserzioni) e dalla sensibilità che non si spinge oltre le 10 megabasi per delezioni o trisomie parziali di un segmento cromosomico (4). La sensibilità di questa metodica viene portata al livello di centinaia di kilobasi qualora invece di ibridare il DNA tumorale e controllo sui cromosomi normali, si utilizzino "array" precostituiti contenenti sequenze di DNA normale clonate in vettori (BAC, PAC) rappresentative di regioni o geni di interesse nella patologia in esame (5).

Lesioni citogenetiche e molecolari

Profilo citogenetico

Oggi si preferisce l'utilizzo della FISH in campo clinico, per la sua sensibilità e relativa semplicità; in realtà entrambe le tecniche hanno un rilevante impatto prognostico (vide infra). Il profilo di anomalie cromosomiche della LLC è molto ricco e le aberrazioni del cariotipo riportate in letteratura come anomalia singola in almeno due casi sono le seguenti: +12; 13q-, 11q-, 6q-, t(14;18), t(14;19), +3, +18, 17p-, +21, +X, 8q+ (6). In molti casi queste anomalie possono essere presenti contemporaneamente nello stesso paziente: infatti è noto che circa il 40-50% dei pazienti con cariotipo anomalo presenta una singola anomalia, mentre un 20-30% dei pazienti presenta 2 anomalie e 20-30% presenta 3 o più anomalie (cariotipo complesso). In diversi casi, sono stati riportati nello stesso caso cloni con anomalie diverse e non correlate tra loro: ad esempio si possono rinvenire nello stesso paziente cellule con delezione 13, altre cellule con trisomia 12 e cellule con entrambe le anomalie. Questi dati riflettono il fatto che queste anomalie non sempre si manifestano nelle fasi precoci della leucemogenesi, mentre giocano un ruolo importante nelle fasi di espansione clonale. Oggi si utilizza in campo clinico una classificazione "gerarchica" che si fonda sull'idea, supportata dai dati sperimentali, che alcune anomalie hanno

Tab. 1 - Frequenza della principali anomalia cromosomiche nella LLC

	Citogenetica convenzionale	FISH
% pazienti con cariotipo normale	40-60%	20%
% pazienti con cariotipo non valutabile	10%	0%
% pazienti con cariotipo alterato	40-50%	80%
% con anomalie specifiche		
+12	20-40% (*)	15-25% (**)
13q-	10-20%	50%
11q-	5-15%	15-30%
6q-	5-10%	2-5%
t(14q32)(#)	3-5%	2-3%
17p-	1-3%	1-5%

(*) % riferite al totale dei casi anomali; (**) % riferite al totale dei casi analizzati;

(#)= esclusa la t(11;14); nel 50% dei casi si tratta di una t(14;18)(q32;q21)

un ruolo biologico determinante e sono in grado di conferire alla cellula leucemica caratteristiche peculiari che si riflettono a loro volta nel comportamento clinico (7).

Così ciascun paziente viene collocato in una classe citogenetica sulla base delle seguenti priorità: delezione 17p > delezione 11q > trisomia 12 > delezione 6q > delezione 13q > normale. Ad esempio, un paziente che presenta 6q- e 11q- ricade nella categoria 11q-, uno che presenta 13q- e 17p- si colloca nella categoria 17p- e così via.

Così facendo si riesce a collocare in una singola categoria citogenetica circa il 70% dei pazienti, mentre i restanti presentano altre anomalie o assenza di aber-

Tab. 2 - Correlazione tra le principali lesioni citogenetico-molecolari e caratteristiche clinicobiologiche

Anomalia	Geni o regioni coinvolte	Citologia	Immunofenotipo Geni Ig	Caratteristiche clinicobiologiche
17p-	p53	LLC/PL	CD38+++/- ZAP-70+++/- Ig non mutate +++/-	Prognosi altamente sfavorevole (mediana sopravvivenza <5 anni). Resistenza agli analoghi delle purine. Sensibilità all'anti CD52
11q-	ATM	LLC tipica	CD38+++/- ZAP-70+/- Ig non mutate +++/-	Prognosi sfavorevole (mediana sopravvivenza 5-10 anni). Adenopatie profonde e massive. Discreta sensibilità all'antiCD20 (?)
+12	12q13-15	LLC atipica	CD38+/- ZAP-70+++/- Ig non mutate ++/-	Prognosi intermedia (mediana sopravvivenza 10-15 anni) Cattiva sensibilità all'antiCD20 (?)
6q-	6q21	LLC atipica	CD38+++/- Ig non mutate ++/-	Prognosi intermedia (mediana sopravvivenza 10-15 anni) Iperleucocitosi; terapia precoce
13q-	13q14	LLC tipica	CD38- ZAP-70+/- Ig non mutate +/-	Prognosi favorevole (mediana sopravvivenza > 15 anni) Buona sensibilità all'anti CD20 (?)

Tab. 3 - Altre anomalie ricorrenti nella LLC

Lesione	Commenti
Trisomia 8q	Coinvolge la banda 8q24, rinvenuta in prevalenza mediante CGH o FISH nel 5% dei casi, spesso in associazione ad altre anomalie.
Trisomia 3q	Coinvolge le bande 3q26-28, rinvenuta mediante CGH o FISH nel 3% dei casi
Trisomia 19	Rinvenuta spesso in associazione a trisomia 12 e geni IgV mutati
Trisomia 2p24	Presente nel 5% dei casi, rinvenuta mediante CGH arrays, coinvolge il gene MYCN che viene sovraespresso
Delezioni 14q24	Associate ad anomalie addizionali - prognosi sfavorevole (?)
4q35	Cellule linfoplasmocitoidi (LLC atipica); immunofenotipo classico; Componente paraproteinemica; malattia evolutiva; risposta di breve durata alla terapia
4p16	LLC atipica; stadio iniziale, malattia evolutiva
6p24-25	LLC atipica; stadio iniziale, malattia evolutiva
4q21	LLC/PL; fabbisogno precoce di terapia
1p34	LLC tipica; stadio iniziale, malattia indolente
12p	LLC tipica; stadio iniziale, malattia indolente

razioni rilevabili. Come riportato in tabella 1, l'anomalia predominante in ordine di frequenza, rinvenuta nel 40% circa dei casi di LLC è rappresentata dalla delezione 13q, seguita dalla trisomia 12 e dalla delezione 11q, quindi dalla delezione 17p, dal cromosoma 6q-, dalle traslocazioni 14q. Come di seguito descritto e riassunto in tabella 2 ciascuna di queste anomalie si associa ad alcune importanti caratteristiche cliniche e biologiche.

Esiste poi una serie di più rare anomalie, in prevalenza non bilanciate (i.e. che determinano perdita o guadagno di materiale cromosomico), più raramente bilanciate (traslocazioni), globalmente rinvenibili in un 10-20% circa dei pazienti che non ricadono in alcuna delle classi citogenetiche sopradescritte. Queste anomalie sono descritte in tabella 3.

La citogenetica classica può ottenere informazioni aggiuntive mediante l'utilizzo dell'IL2 e degli oligonucleotidi come stimolatori delle metafasi. È stato documentato infatti che in un 30-40% di pazienti apparentemente negativi all'indagine FISH si possono rinvenire anomalie cromosomiche, in genere strutturali (traslocazioni) (2). La presenza di anomalie cromosomiche di tipo non-bilanciato all'indagine citogenetica si associa ad un profilo biologico sfavorevole (geni Ig non mutati). Analogamente, anomalie strutturali bilanciate (traslocazioni), si associano alla positività per il CD38 (8).

Evoluzione del cariotipo

Il cariotipo nella LLC è considerato fondamentalmente stabile, con <20% dei pazienti mostranti aberrazioni aggiuntive durante il decorso della malattia (9). È stato dimostrato che i pazienti con instabilità del cariotipo hanno spesso delezione 11q alla diagnosi (10). Tra le anomalie che compaiono nelle fasi evolutive sono state segnalate la delezione 6q e, l'11q- e il 17p- (11). La comparsa tardiva della delezione 11q si associa alle forme di LLC/PL e ad un'attivazione della malattia (12).

Significato prognostico delle lesioni citogenetico-molecolari

Analisi citogenetica convenzionale

Le informazioni rilevanti e dotate di potere predittivo indipendente, ricavabili dall'indagine citogenetica convenzionale (6) possono essere così riassunte:

- esiste una correlazione significativa tra percentuale di metafasi anomale e sopravvivenza: i pazienti con cariotipo normale alla diagnosi e durante il decorso della malattia (NN) hanno una mediana di sopravvivenza non raggiunta ad 11 anni; i pazienti con una commistione di metafasi normali ed anomale (AN) sopravvivono in media 8 anni circa; quelli con solo metafasi anomale (AA) non raggiungono i 5 anni. Questa osservazione riflette con tutta probabilità la correlazione tra capacità di crescita "in vitro" del clone neoplastico e capacità di crescita "in vivo" e dimostra che i pazienti con un clone anomalo che ha soppiantato le residue cellule normali hanno una malattia più aggressiva.
- Esiste un "gradiente" in decremento della curva di sopravvivenza a partire dai pazienti a cariotipo normale, attraverso quelli con 1 singola anomalia clonale e con 2 anomalie, per terminare con quelli aventi 3 o più anomalie (cariotipo complesso), che mostrano la sopravvivenza peggiore. La correlazione tra instabilità genetica-citogenetica e prognosi è altamente riproducibile in molte neoplasie ematologiche.
- Anomalie cromosomiche specifiche hanno significato prognostico indipendente: la delezione 13q è un fattore "favorevole" se presente come anomalia singola, la trisomia 12 e la delezione 11q hanno un significato "sfavorevole", al pari delle alterazioni 14q. È noto che molti dei pazienti con traslocazioni 14q32, hanno la t(11;14)(q13;q32). Questa anomalia continua ad essere diagnosticata anche oggi in pazienti arruolati in trial clinici multicentrica, ma si ritiene che identifica casi particolari di linfomi del mantello in fase leucemica o, più raramente, come LLC/PL o LLC atipica (13).
- L'impiego di oligodeossinucleotidi ed IL2 come stimolatori delle metafasi consente la identificazione di traslocazioni cromosomiche, spesso non rilevabili con stimolatori convenzionali (TPA, LPS) che identificano un gruppo di pazienti a cattiva prognosi (2, 8).

FISH

Il numero di pazienti con una specifica anomalia cromosomica sale dal 50% al 70-75% impiegando un pannello di 4 sonde che rilevano le anomalie 13q-, 17p-, 11q- e trisomia 12 ed all'82% aggiungendo sonde che rilevano il 6q-, le traslocazioni 14q32, le trisomie parziali 3q ed 8q. Le principali correlazioni tra specifiche anomalie e prognosi emerse in un ampio studio di riferimento (7) sono qui riassunte.

costruendo una classificazione gerarchica che riconosce importanza, in ordine crescente, alla delezione 13q, alla trisomia 12, all'11q- ed al 17p- si rileva che la rispettiva frequenza di queste anomalie è la seguente: 36%; 14%; 17%; 7%. Il 18% dei casi ha un cariotipo normale; l'8% ha anomalie varie. La frequenza di

queste anomalie cromosomiche varia a seconda della tipologia di pazienti considerata: ad esempio nei trias clinici che arruolano pazienti in fase avanzata o che richiedono terapia per malattia attiva la percentuale di 11q- può essere >20% ed il 6q- può essere rinvenuto in oltre il 5% dei casi mentre nei casi o resistenti alla fludarabina 17p- e 11q- assieme possono rappresentare oltre il 50% dei casi (14). La mediana di sopravvivenza in questi gruppi è significativamente diversa (6), attestandosi a 133 mesi per il 13q-, a 111 mesi per l'assenza di anomalie rilevabili, a 114 mesi per la trisomia 12, a 79 mesi per l'11q- e a 32 mesi per il 17p-. I pazienti con 17p- mostrano il più breve intervallo tra diagnosi ed inizio del trattamento; quelli con 13q- il più lungo. Il 17p- e l'11q- sono fattori prognostici sfavorevoli indipendenti. I pazienti con 11q- presentano malattia difficilmente eradicabile, anche con procedure di terapia intensiva seguita da trapianto autologo (14). In una recente esperienza l'impatto prognostico avverso di queste variabili citogenetiche poteva essere parzialmente abbattuto da procedure di trapianto allogenico con regime di induzione ad intensità ridotta (15).

La delezione 6q identifica un gruppo di pazienti a prognosi intermedia (16), come si evince anche dalla più elevata frequenza di questa anomalia in pazienti arruolati in schemi di trattamento per stadi avanzati (14).

L'approccio diagnostico impiegante la FISH riconosce dunque la presenza di almeno due anomalie cromosomiche (17p- ed 11q-), che hanno importanza nella valutazione del rischio. Ha il pregio di essere relativamente semplice ed effettuabile su larga scala, presenta tuttavia il limite di non riconoscere altri indici citogenetici importanti quali il cariotipo complesso, la percentuale di cellule in mitosi anomale. Non consente inoltre di riconoscere il significato della trisomia 12 che risulta omologata ad un ampio gruppo non pienamente informativo includente il cariotipo normale e la delezione 13q. La moderna diagnostica citogenetica della LLC deve oggi includere, oltre all'analisi FISH, anche il cariotipo con efficienti stimolatori delle metafasi. Tenuto conto della comparsa di anomalie aggiuntive durante il decorso della malattia, si profila la utilità di una valutazione citogenetico-molecolare sequenziale nel paziente di età <60 anni, anche alla luce delle nuove opzioni di trattamento che possono includere svariati regimi chemioimmunoterapici e, in casi attentamente selezionati, il trapianto di midollo (17).

Bibliografia

1. Cuneo A, Bigoni R, Castoldi G. Towards a clinically relevant cytogenetic classification of chronic lymphocytic leukemia and related disorders. *Haematologica*. 1998; 83: 577-579.
2. Dicker F, Schnittger S, Haferlach T, Kern W, Schoch C. Immunostimulatory oligonucleotide-induced metaphase cytogenetics detect chromosomal aberrations in 80% of CLL patients: A study of 132 CLL cases with correlation to FISH, IgVH status, and CD38 expression. *Blood*. 2006; 108: 3152-60.
3. Mecucci C. FISH (fluorescent in situ hybridization): the second youth of cytogenetics. *Haematologica*. 1995; 80: 95-97.
4. Bentz M, Huck K, du Manoir S et L. Comparative genomic hybridization in

- chronic B-cell leukemias shows a high incidence of chromosomal gains and losses. *Blood*. 1995; 85: 3610-3618.
5. Schwaenen C, Nessling M, Wessendorf S et al. Automated array-based genomic profiling in chronic lymphocytic leukemia: development of a clinical tool and discovery of recurrent genomic alterations. *Proc Natl Acad Sci U S A*. 2004; 101: 1039-1044.
 6. Juliusson G, Oscier DG, Fitchett M et al. Prognostic subgroups in B-cell chronic lymphocytic leukemia defined by specific chromosomal abnormalities. *N Engl J Med*. 1990; 323: 720-4.
 7. Dohner H, Stilgenbauer S, Benner A et al. Genomic aberrations and survival in chronic lymphocytic leukemia. *N Engl J Med*. 2000; 343: 1910-6.
 8. Mayr C, Speicher MR, Kofler DM, Buhmann R, Strehl J, Busch R, Hallek M, Wendtner CM. Chromosomal translocations are associated with poor prognosis in chronic lymphocytic leukemia. *Blood*. 2006; 107: 742-51.
 9. Juliusson G, Friberg K, Gahrton G. Consistency of chromosomal aberrations in chronic B-lymphocytic leukemia. A longitudinal cytogenetic study of 41 patients. *Cancer*. 1988; 62: 500-506.
 10. Fegan C, Robinson H, Thompson P, Whittaker JA, White D. Karyotypic evolution in CLL: identification of a new sub-group of patients with deletions of 11q and advanced or progressive disease. *Leukemia*. 1995; 9: 2003-2008.
 11. Rai KR, Dohner H, Keating MJ, Montserrat E. Chronic lymphocytic leukemia: case-based session. *Hematology (Am Soc Hematol Educ Program)*. 2001; 140-56.
 12. Cuneo A, Bigoni R, Rigolin GM, et al. Late appearance of the 11q22.3-23.1 deletion involving the ATM locus in B-cell chronic lymphocytic leukemia and related disorders. Clinico-biological significance. *Haematologica*. 2002; 87: 44-51.
 13. Cuneo A, Bigoni R, Negrini M, et al. Cytogenetic and interphase cytogenetic characterization of atypical chronic lymphocytic leukemia carrying BCL1 translocation. *Cancer Res*. 1997; 57: 1144-50.
 14. Byrd JC, Stilgenbauer S, Flinn IW. Chronic lymphocytic leukemia. *Hematology Am Soc Hematol Educ Program*. 2004; 163-83.
 15. Caballero D, Garcia-Marco JA, Martino R, Mateos V, Ribera JM, Sarra J, Leon A, Sanz G, de la Serna J, Cabrera R, Gonzalez M, Sierra J, San Miguel J. Allogeneic transplant with reduced intensity conditioning regimens may overcome the poor prognosis of B-cell chronic lymphocytic leukemia with unmutated immunoglobulin variable heavy-chain gene and chromosomal abnormalities (11q- and 17p-). *Clin Cancer Res*. 2005; 11: 7757-63.
 16. Cuneo A, Rigolin GM, Bigoni R, De Angeli C, Veronese A, Cavazzini F, Bardi A, Roberti MG, Tammiso E, Agostini P, Ciccone M, Della Porta M, Tieghi A, Cavazzini L, Negrini M, Castoldi G. Chronic lymphocytic leukemia with 6q- shows distinct hematological features and intermediate prognosis. *Leukemia*. 2004; 18: 476-83.
 17. Wierda WG. Current and Investigational Therapies for Patients with CLL. *Hematology Am Soc Hematol Educ Program*. 2006; 285-94.

Definizione delle anomalie cromosomiche nel mieloma multiplo

Nicoletta Testoni, Simona Luatti, Giulia Marzocchi, Chiara Nicci, Carmen Baldazzi, Monica Stacchini

Istituto di Ematologia e Oncologia Medica, Policlinico Sant'Orsola, Università di Bologna

Il mieloma multiplo è caratterizzato, almeno nelle prime fasi della malattia, da una scarsa proliferazione delle plasmacellule neoplastiche e, spesso, anche da una loro bassa infiltrazione midollare. L'analisi citogenetica risulta per questo essere limitata e difficoltosa, in quanto le metafasi ottenute ed analizzate rappresentano in molti casi la mielopoiesi normale, per cui spesso è indispensabile l'analisi citogenetica di un elevato numero di metafasi per l'identificazione di un clone anormale (1, 2).

La frequenza delle alterazioni cromosomiche, analizzate con metodiche di citogenetica convenzionale, varia dal 30-50% nei pazienti alla diagnosi, al 40-60% nei pazienti in progressione, ad oltre l'85% nelle leucemie plasmacellulari. Tale aumento deriva verosimilmente da un'incrementata attività proliferativa delle plasmacellule e dall'eventuale insorgenza di alterazioni citogenetiche secondarie, che possono mascherare l'evento primario della malattia.

Negli ultimi anni, lo sviluppo della citogenetica molecolare, che va dalla FISH (fluorescent *in situ* hybridization) alla SKY/M-FISH (spectral karyotype imaging/multicolor-FISH) e alla CGH (comparative genomic hybridization), ha notevolmente aumentato la possibilità di individuare anomalie citogenetiche. L'individuazione di specifiche alterazioni genetiche ha permesso una migliore comprensione della patogenesi di questa malattia (3).

Sicuramente la metodica più utilizzata, e quella che ha permesso di caratterizzare meglio questa patologia da un punto di vista citogenetico, è la FISH. Tuttavia, come detto precedentemente, nei mielomi alla diagnosi la percentuale di plasmacellule neoplastiche presenti nel midollo non è alta. Di conseguenza, per riuscire ad analizzare il clone mielomatoso, è importante, prima di applicare la FISH, effettuare una selezione della popolazione plasmacellulare. A questo scopo possono essere applicate diverse metodiche e sicuramente una delle più pratiche ed utilizzate è la selezione positiva tramite sistema immunomagnetico di separazione su colonna.

Il mieloma multiplo è caratterizzato spesso da un cariotipo complesso con numerose alterazioni strutturali e numeriche. Ciò rende ancora più problematica l'in-

dagine citogenetica e perciò, in molti casi, le alterazioni possono essere non ben caratterizzate. In particolare il MM è caratterizzato da frequente presenza di aneuploidia. Le alterazioni numeriche più frequenti riguardano monosomie dei cromosomi 13, 14, 16, 22 e trisomie dei cromosomi 3, 5, 7, 9, 11, 15, 19 e 21 (1). In base al tipo di aneuploidia presente i pazienti possono essere suddivisi nelle seguenti categorie: ipodiploidi/ipotetraploidi, pseudodiploidi e iperdiploidi.

I primi due gruppi di pazienti sono associati prevalentemente a traslocazioni che coinvolgono la banda 14q32 (locus del gene delle catene pesanti delle immunoglobuline-IgH) e a monosomia/delezione del cromosoma 13. In particolare è evidente l'associazione con le tre traslocazioni 14q32 più ricorrenti: t(4;14)(p16;q32), t(11;14)(q13;q32) e t(14;16)(q32;q23). Mentre nell'iperdiploidia, sono meno comuni queste traslocazioni ma prevalgono duplicazioni di segmenti cromosomici (1, 4, 5). Per quanto riguarda il valore prognostico di questi sottogruppi, diversi studi hanno dimostrato l'associazione, dell'ipodiploidia ad una sopravvivenza più breve (7, 8). Tra le alterazioni strutturali, le più frequenti e maggiormente caratterizzate sono quelle che coinvolgono il locus del gene delle catene pesanti delle immunoglobuline (IgH) sulla banda cromosomica 14q32. Le traslocazioni che coinvolgono i geni delle catene pesanti (IgH) e delle catene leggere (IgL- λ e IgL- κ) delle immunoglobuline sembrano costituire degli eventi primari nella patogenesi del MM (9,10). Studi di FISH indicano che traslocazioni che coinvolgono il locus IgH sono presenti nel 55-70% dei casi di MM e nell'80% dei casi di leucemia plasmacellulare (10, 11). La banda 14q32 è coinvolta in traslocazioni con più di 20 partners cromosomici differenti. Le traslocazioni più frequenti sono quelle con i loci: 11q13, che coinvolge il gene della *ciclina D1* (*CCND1*), presente nel 15-20% dei casi (10-12); 4p16, che coinvolge i geni *MMSET* e *FGFR3*, presente nel 15-20% dei casi (13, 14); 16q23, che coinvolge il gene *c-maf*, presente in circa il 2-10% dei casi (15); 6p21, che coinvolge il gene della *ciclina D3* (*CCND3*), presente in circa il 4% dei casi (16).

t(11;14)(q13;q32)

È la traslocazione più frequente infatti è stata identificata, tramite FISH, in circa il 15-20% dei pazienti con MM. La presenza della t(11;14) è stata correlata con una morfologia linfoplasmacitoide, cioè ad una morfologia matura delle plasmacellule, espressione dell'anticorpo CD20, e con i sottotipi di MM oligo-/non secernente. I breakpoints di traslocazione sul cromosoma 11 sono dispersi fra 330 kb che costituiscono il locus del gene della Ciclina D1 (*CCND1*) (17). Il risultato della traslocazione è l'espressione aumentata della ciclina. Per quanto riguarda le implicazioni cliniche, sono riportati pareri discordanti. I primi studi mostravano l'associazione della traslocazione ad una prognosi sfavorevole (18, 19). In realtà gli studi erano stati condotti su un piccolo numero di pazienti, inoltre erano studi citogenetici retrospettivi condotti su pazienti trattati con diverse strategie terapeutiche e analizzati solo in citogenetica convenzionale. Quindi, probabilmente, la prognosi severa poteva essere correlata ad un alto indice proliferativo delle plasmacellule di questi pazienti, tale da permettere di individuare l'alterazione con la metodica convenzionale.

In seguito sono stati condotti degli studi su coorti più numerose di pazienti trattati omogeneamente. Questi studi fanno supporre che la t(11;14) individui un sottogruppo di mielomi che è relativamente “geneticamente stabile”, con conseguente decorso più indolente della malattia ed una maggiore suscettibilità del clone plasmacellulare alla chemioterapia, in particolare a quella ad alte dosi (17).

t(4;14)(p16.3;q32)

La t(4;14)(p16.3;q32) è citogeneticamente criptica e vede coinvolti due oncogeni, MMSET (Multiple Myeloma SET domain, fattore di trascrizione) che permane sul der(4) e FGFR3 (Fibroblast Growth Factor Receptor 3, recettore tirosin chinasi) che trasloca sul der(14). Da notare che in circa il 20% dei casi con t(4;14), FGFR3 sul der(14) è perso o non espresso. Questa traslocazione è stata trovata per la prima volta in tumori che presentavano un isotipo IgA ed è associata a pazienti con caratteristiche cliniche aggressive. Indipendentemente dalla modalità di trattamento, la t(4;14) è associata ad un decorso clinico non favorevole. Alcuni studi hanno osservato una stretta relazione fra t(4;14) e delezione del cromosoma 13, alti livelli sierici di b2-microglobulina, un alto PCLI, alti livelli di componente M sierica e una morfologia immatura delle PC, quindi con caratteristiche associate ad una prognosi sfavorevole (4,20). Recenti studi hanno dimostrato che la t(4;14) è associata ad una OS ed ad una EFS significativamente più brevi, sia nei pazienti trattati con chemioterapia convenzionale che in quelli trattati con terapia ad alte dosi.

t(14;16)(q32;q23)

La traslocazione (14;16)(q32;q23) è citogeneticamente silente e può essere rilevata tramite FISH. Tale traslocazione è rara; infatti presenta una incidenza del 2-10% in pazienti con MM, ed ha come risultato una espressione aumentata del fattore di trascrizione c-maf che regola l'espressione di geni come CCND2. La deregolazione di questo oncogene, quindi, può portare ad un controllo alterato del ciclo cellulare, dell'apoptosi o dell'interazione cellula-cellula. La sua presenza sembra essere associata ad una OS e ad una EFS significativamente più brevi (1, 21).

Monosomia 13/ delezione 13q (-13/13q-)

In studi di FISH queste alterazioni sono presenti in circa il 40-50% dei casi, mentre utilizzando solo la citogenetica convenzionale l'incidenza della delezione del cromosoma 13, è intorno al 10-20%. È stato osservato che l'alterazione sul cromosoma 13 è una monosomia nella maggior parte dei casi (80-90% dei casi). Benchè sia noto che la parziale delezione sia concentrata nella banda cromosomica 13q14, ancora non è stato definito il singolo segmento genomico che viene deletato. La regione critica di rottura (13q14-q21) è costituita da 350kb, zona dove è localizzato il gene RB1 (gene oncosoppressore del retinoblastoma), che gioca un ruolo importante in numerose neoplasie, ma non sembra avere un ruolo nella patogenesi del MM (4).

È per questo che la delezione del cromosoma 13 viene considerata, da molti, sem-

plicemente un marker surrogato di altre alterazioni, a prognosi sfavorevole, a cui risulta essere associata, come l'ipodiploidia o la t(4;14) (6, 8, 22, 23). In realtà, nel mieloma multiplo, la delezione del 13 sembra essere associata ad importanti implicazioni cliniche e molte sue caratteristiche fanno supporre un suo specifico ruolo biologico. Comunque il suo ruolo nella patogenesi del MM non è ancora stato chiarito. Per quanto riguarda le implicazioni cliniche di questa alterazione, diversi studi hanno riportato che la delezione 13 è associata ad una OS e ad una EFS significativamente più brevi, sia in pazienti trattati con chemioterapia convenzionale che in pazienti trattati con terapia ad alte dosi (22, 23). Invece alcuni studi mostrano che se associata ad altri fattori a prognosi sfavorevole, come alti livelli sierici di β_2 -microglobulina o la presenza di un cariotipo alterato, l'alterazione porta ad una prognosi ancora più sfavorevole. Mentre la presenza di alterazioni a prognosi favorevole, come la t(11;14), in pazienti trattati con terapia ad alte dosi, annulla l'effetto negativo sulla prognosi della delezione stessa. Infine diversi studi riportano l'associazione della delezione del cromosoma 13 alla t(4;14)(p16;q32).

Recenti studi hanno evidenziato la significativa associazione di una precoce ricaduta della malattia nei pazienti con alterazioni citogenetiche del cromosoma 13 ma non in quelli con delezioni del cromosoma 13 osservate solo in FISH. Quindi l'analisi citogenetica convenzionale sembrerebbe identificare una popolazione clonale iperproliferante con un'alta frequenza di fallimento terapeutico.

Delezione 17p13

Nella banda 17p13 è presente il locus del gene oncosoppressore p53, che viene perso in caso di delezione. p53 è un fattore trascrizionale che induce la trascrizione dei geni coinvolti nella regolazione del ciclo cellulare, nei meccanismi di riparazione del DNA e di morte cellulare programmata. L'alterazione di p53 può essere il risultato di una mutazione puntiforme, una perdita allelica, un riarrangiamento o una delezione. L'incidenza della del(17p) varia in un range compreso fra il 5% e il 10% dei pazienti alla diagnosi. Si è riscontrato, invece, che la perdita funzionale del gene ha un'incidenza del 40% nei pazienti in stadio avanzato. Tutto ciò sta ad indicare che tale delezione possa essere un marker della progressione della malattia. Per quanto riguarda l'associazione con terapia convenzionale ed ad alte dosi, gli studi finora effettuati riportano che la delezione 17p13 è associata ad una prognosi sfavorevole sia in pazienti trattati con CHT convenzionale che in pazienti con terapie ad alte dosi (4).

La FISH rappresenta comunque lo studio più appropriato per questo tipo di alterazione, rispetto alla citogenetica convenzionale.

Anomalie del cromosoma 1

Anomalie del cromosoma 1 sono state osservate in più del 48% dei casi. Recenti studi hanno descritto il coinvolgimento della banda 1q21 in fenomeni di amplificazione ed overespressione del locus del gene CKS1B (regolatore del ciclo cellulare). Tali implicazioni sembrano essere associate ad una prognosi sfavorevole (4).

Bibliografia

1. Fonseca R, Barlogie B, Bataille R, Bastard C, et al. Plasma cell karyotypes in multiple myeloma. *Blood* 1988; 71: 453.
2. Ho PJ, Campbell LJ, Gibson J, et al. The biology and cytogenetics of multiple myeloma. *Rev Clin Exp Hematol* 2002; 6.3: 276-300.
3. Liebisch P, Döhner H. Cytogenetics and molecular cytogenetics in multiple myeloma. *European Journal of Cancer* 2006, 42: 1520-1529.
4. Chen Z, Issa B, Huang S, et al. A practical approach to the detection of prognostically significant genomic aberrations in multiple myeloma. *Journal of Molecular Diagnostics* 2005, 7: 560-565.
5. Debes-Marun C, Dewald G, Bryant S, et al. Chromosome abnormalities clustering and its implications for pathogenesis and prognosis in myeloma. *Leukemia (Baltimore)* 2003; 17: 427-436.
6. Smadja NV, Leroux D, Soulier J, et al. Further cytogenetic characterization of multiple myeloma confirms that 14q32 translocations are a very rare event in hyperdiploid patients. *Genes Chromosomes Cancer* 2003; 38: 234-239.
7. Fassas AB, Spencer T, Sawyer J, et al. Both hypodiploidy and deletion of chromosome 13 independently confer poor prognosis in multiple myeloma. *Br J Haematol* 2002; 118: 1041-1047.
8. Bergsagel PL, Chesi M, Brents LA, Kuehl WM. Translocations into IgH switch regions - the genetic hallmark of multiple myeloma. *Blood* 1995; 86:223-223.
9. Bergsagel PL, Chesi M, Nardini E, et al. Promiscuous translocations into immunoglobulin heavy chain switch regions in multiple myeloma. *Proc Natl Acad Sci USA* 1996; 93: 13931-13936.
10. Avet-Loiseau H, Facon T, Grosbois B, et al. Oncogenesis of multiple myeloma: 14q32 and 13q chromosomal abnormalities are not randomly distributed, but correlate with natural history, immunological features and clinical presentation. *Blood* 2002; 99: 2185-2191.
11. Fonseca R, Harrington D, Oken M, et al. Myeloma and the t(11;14)(q13;q32); evidence for a biologically defined unique subset of patients. *Blood* 2002; 99: 3735-3741.
12. Chesi M, Nardini E, Brents LA, et al. Frequent translocation t(4;14)(p16.3;q32) in multiple myeloma is associated with increased expression and activating mutations of fibroblast growth factor receptor 3. *Nat Genet* 1997; 16: 260-264.
13. Chesi M, Nardini E, Lim R, et al. The t(4;14) translocation in myeloma dysregulates both FGFR3 and a novel gene, MMSET, resulting in IgH/MMSET hybrid transcripts. *Blood* 1998; 92: 3025-3034.
14. Chesi M, Bergsagel PL, Shonukan OO, et al. Frequent dysregulation of the *c-maf* proto-oncogene at 16q23 by translocation to an Ig locus in multiple myeloma. *Blood* 1998; 91: 4457-4463.
15. Shaughnessy J, Gabrea A, Qi Y, et al. Cyclin D3 at 6p21 is dysregulated by

- recurrent chromosomal translocations to immunoglobulin loci in multiple myeloma. *Blood* 2001; 98: 217-223.
16. Soverini S, Cavo M, Cellini C, et al. *Cyclin D1* overexpression is a favorable prognostic variable for newly diagnosed multiple myeloma patients treated with high-dose chemotherapy and single or double autologous transplantation. *Blood* 2003; 102: 1588-1594.
 17. Lai JL, Michaux L, Dastugue N, et al. Cytogenetics in multiple myeloma: a multicenter study of 24 patients with t(11;14)(q13;q32) or its variant. *Cancer Genet Cytogenet.* 1998; 104(2): 133-138.
 18. Fonseca R, Witzig TE, Gertz MA et al. Multiple myeloma and the translocation t(11;14)(q13;q32): a report on 13 cases. *Br J Haematol* 1998; 101: 296-301.
 19. Garand R, Avet-Loiseau H, Accard F, et al. t(11;14) and t(4;14) translocations correlated with mature lymphoplasmacytoid and immature morphology, respectively, in multiple myeloma. *Leukemia* 2003; 17: 2032-2035.
 20. Fonseca R, Debes-Marun C, Picken E, et al. The recurrent IgH translocations are highly associated with nonhyperdiploid variant multiple myeloma. *Blood* 2003; 102: 2562-2567.
 21. Zojer N, Konigsberg R, Ackermann J, et al. Deletion of 13q remains an independent adverse prognostic variable in multiple myeloma despite its frequent detection by interphase fluorescent in situ hybridization. *Blood* 2000; 95: 1925-1930.
 22. Shaughnessy J, Jacobson J, Sawyer J, et al. Continuous absence of metaphase-defined cytogenetic abnormalities, especially of chromosome 13 and hypodiploidy, ensures long-term survival in multiple myeloma treated with Total Therapy 1: interpretation in the context of global gene expression. *Blood* 2003; 101: 3849-3856.
 23. Facon T, Avet-Loiseau H, Guillermin G, et al. Chromosome 13 abnormalities identified by FISH analysis and serum β_2 -microglobulin produce a powerful myeloma staging system for patients receiving high-dose therapy. *Blood* 2001; 97: 1566-1571.

Necessità di un percorso informatico alla conoscenza biologica

Sergio Ferrari

Dipartimento di Scienze Biomediche, Sezione di Chimica Biologica,
Università di Modena e Reggio Emilia, Modena

Il completamento del progetto genoma ha enormemente aumentato le conoscenze sull'organizzazione del genoma umano ed in particolare dei geni che codificano per proteine. Va subito precisato che il genoma non è semplicemente l'insieme dei geni codificanti in quanto questi occupano meno del 10% dell'intera sequenza del DNA. La larga prevalenza dei trascritti e quindi del trascrittoma è dovuta a RNA non codificanti.

Le tecniche di studio dell'espressione genica ed in particolare i DNA microarrays offrono la possibilità di monitorare contemporaneamente le decine di migliaia di diversi trascritti che una cellula è in grado di produrre.

Il profilo di espressione genica permette di avere una più completa comprensione della funzione dei geni, della loro regolazione e delle loro possibili interazioni. Inoltre, l'identificazione di "patterns" di espressione genica e il loro raggruppamento in classi di espressione permette una maggiore comprensione dei processi biologici. Larga parte di questa breve trattazione riguarderà l'analisi dei dati e la loro interpretazione.

Basi tecnologiche dei DNA microarrays

La tecnologia dei DNA microarrays si è sviluppata a partire dalla fine degli anni novanta e ha consentito lo studio dell'espressione di migliaia di geni contemporaneamente. La comunità scientifica ha poi cominciato ad applicarla ad un ampio spettro di processi biologici sia normali che patologici. Tale metodologia permette infatti la fenotipizzazione molecolare di un contesto cellulare.

Due tipi di DNA microarrays sono attualmente in uso, gli oligonucleotidi arrays e gli spotted arrays. Numerose sono inoltre le piattaforme tecnologiche che permettono di eseguire questo tipo di indagine con procedura semiautomatizzata.

Una delle applicazioni più importanti degli arrays è, come già accennato, il monitoraggio dell'espressione genica globale di una cellula a livello degli RNA messaggeri.

Il profilo di espressione genica globale o trascrittoma

L'insieme delle migliaia di geni espressi o trascritti dal DNA genomico in un determinato contesto cellulare prende il nome di profilo di espressione genica (PEG) o trascrittoma, e permette di determinare il fenotipo molecolare della cellula. Come primo passo nel processo di sintesi proteica, il DNA genomico viene trascritto in RNA messaggero, e si assume che ad un RNA messaggero corrisponda una proteina, anche se in realtà ciò non sempre è vero, in quanto esiste un controllo traduzionale dell'espressione genica. Possiamo comunque assumere che differenze nell'espressione genica sono responsabili delle differenze morfologiche e fenotipiche della cellula in risposta a stimoli microambientali o a perturbazioni di vario tipo (nutrizionali, condizionamento genico, temperatura, interazioni patogeni-ospite, trattamento con composti biologicamente attivi, ecc.). A differenza del genoma, il trascrittoma è altamente dinamico e si modifica rapidamente e drammaticamente in risposta a perturbazioni o anche durante i normali eventi cellulari quali proliferazione, differenziamento, apoptosi, attivazione, fagocitosi ecc. Per la comprensione della funzione dei geni, sapere quando, dove e come un gene è espresso (in termini di abbondanza di RNA messaggero) è fondamentale per capire l'attività e il ruolo biologico della proteina da esso codificata. Inoltre, e più rilevante, le variazioni dei pattern di espressione permettono di valutare i meccanismi di regolazione a livello trascrizionale, di sviluppare mappe trascrizionali e quindi di capire un ampio spettro di funzioni biologiche, quali vie metaboliche sono attivate o quali vie di trasduzione del segnale sono coinvolte nel processo biologico oggetto di studio. Lo studio del trascrittoma è stato negli ultimi 6 anni ampiamente utilizzato non solo per lo studio dei processi biologici normali ma anche per studiare le cause e le conseguenze di malattie, le interazioni tra farmaci e funzioni cellulari, lo sviluppo di nuovi farmaci mirati su bersagli genici. L'uso di questa tecnologia ha richiesto la risoluzione di problemi legati al disegno sperimentale, all'uso di piccole quantità di RNA messaggero ottenibile da campioni biologici (ad es. biopsie), allo sviluppo di software per la corretta analisi dei dati e alla valutazione della significatività statistica dei risultati ottenuti, al fine di estrarre dati biologicamente rilevanti utilizzabili nella pratica clinica. Sulle riviste internazionali sono stati pubblicati migliaia di articoli scientifici nei quali è stato studiato il trascrittoma. I campi applicativi di tale tecnologia in biologia e medicina sono molteplici, e fra questi troviamo lo studio dei programmi genetici di processi biologici attivabili dalle cellule (quali quello proliferativo, differenziativo, apoptotico), lo studio dei meccanismi molecolari alla base della trasformazione e progressione delle cellule tumorali, la classificazione molecolare dei diversi tipi di tumori, lo studio della interferenza dei farmaci con la trascrizione genica, lo studio della farmacoresistenza, lo studio dei meccanismi patogenetici di malattie complesse e l'individuazione di nuovi geni bersaglio per lo sviluppo di strategie terapeutiche mirate. L'insieme dei geni che sono coespressi in un processo biologico normale o patologico viene definito "signature" e può avere una valenza diagnostica, prognostica o patogenetica se individuato in una malattia. Segnaliamo inoltre la possibilità di consultare in rete

banche dati relative agli esperimenti di DNA microarrays, in particolare Microarrays and gene expression (<http://www.mged.org/mage>) e Stanford microarrays Database (<http://genome-www5.stanford.edu>).

Analisi dei dati di microarrays

Cinque sono le fasi fondamentali dell'analisi dei dati di microarrays:

- 1) disegno sperimentale;
- 2) processamento;
- 3) inferenza;
- 4) classificazione;
- 5) validazione.

1) Disegno sperimentale

Il disegno sperimentale interferisce con l'efficienza e la validità degli esperimenti. C'è un consenso generale della comunità scientifica sul fatto che la replica biologica degli esperimenti è essenziale. Possiamo programmare una replica tecnica, cioè usare l'RNA estratto da un singolo campione biologico per l'ibridazione a diversi DNA arrays. Questa replica permette di valutare la variabilità tecnica dei diversi DNA chip e quindi costituisce un controllo di qualità. La replica biologica viene fatta partendo invece da diversi esperimenti biologici e quindi diversi campioni ibridati e permette di valutare la variabilità e quindi la riproducibilità del sistema biologico oggetto dello studio. La domanda che sorge spontanea è: quanti replicati biologici sono necessari per poter avere una significatività statistica dei risultati? Se quello che si deve valutare è l'espressione differenziale fra due contesti cellulari sono necessarie dalle 3 alle 5 repliche. Riunire in pool gli RNA derivati da esperimenti o campioni omogenei è possibile e in alcuni casi necessario quando l'RNA che si può estrarre dai singoli campioni è piccolo e non sufficiente per programmare un esperimento di DNA microarray. Questa risoluzione elimina le possibili varianti biologiche individuali e dà risultati in parte diversi dalla ibridazione dei singoli campioni. Infine gli esperimenti biologici vanno rigorosamente programmati usando le stesse condizioni sperimentali e possibilmente gli stessi reagenti.

2) Processamento

La scansione delle immagini dei dati di ibridazione su DNA microarrays è una procedura di acquisizione necessaria sia per la normalizzazione, cioè il processo per cui l'intensità dei vari spot sono aggiustati tenendo conto della variabilità fra i diversi esperimenti e piattaforme, che per rimuovere le variazioni sistematiche. Inoltre permette la trasformazione di dati, cioè l'applicazione di specifiche funzioni matematiche cosicché i dati sono cambiati in una forma differente, ad esempio \log_2 (fold change: comparazione fra i livelli di espressione fra due distinti campioni). Il filtraggio dei dati, cioè la selezione dei dati importanti per l'interpretazione dell'esperimento. Ad esempio se sono interessato ai geni differenzialmente espressi è evidente che filtro, cioè escludo dall'analisi, i geni espressi

costituttivamente o non espressi. Un esempio di filtraggio dei dati che riguarda gli arrays a due colori è la sottrazione del fondo.

3) *Inferenza*

Si tratta di trarre conclusioni sulla veridicità dell'ipotesi formulata. Ad es. l'ipotesi da verificare è: esistono differenze nell'espressione genica quando coltivo le stesse cellule nella condizione A o nella condizione B? Nel trarre le conclusioni si possono commettere molti errori quali la presenza di falsi positivi o falsi negativi. Esistono molti metodi di analisi dei dati che permettono di evitare questi errori che non abbiamo la possibilità di trattare dettagliatamente. Fra questi possiamo annoverare il t-test, ANOVA, la tecnica di regressione logistica ecc. Nella comunità scientifica vi è unanime accordo che non è più sufficiente prender in considerazione il solo "fold change" come parametro dell'espressione differenziale.

4) *Classificazione*

Il processo di classificazione consiste nell'ordinare i geni in categorie preesistenti, in questo caso si parla di "supervised classification" anche denominata assegnazione o predizione di classe o discriminazione, o di sviluppare un set di categorie in cui i geni vengono raggruppati "unsupervised classification", detta anche scoperta di classe. Diversi algoritmi sono stati sviluppati per questo tipo di analisi di "clustering".

5) *Validazione*

Nella maggior parte dei casi si preferisce validare i dati di espressione mediante altre tecniche ed in particolare mediante RT-PCR quantitativa. È evidente che se limito il mio studio all'espressione genica valuto di base l'attività trascrizionale della cellula in quanto pochi sono gli esempi di stabilizzazione e compartimentalizzazione degli RNA messenger. Posso validare i dati microarrays anche studiando le proteine codificate dagli RNA messenger mediante le diverse metodologie della proteomica, ma in realtà sono numerosi ormai gli esempi in letteratura che non tutti gli RNA messenger vengono decodificati a proteine.

Rappresentazione grafica dei dati derivati dall'analisi del trascrittoma

Il percorso informatico per analizzare i dati di espressione ottenuti dagli esperimenti di DNA microarrays può essere lungo e laborioso ma fondamentalmente viene adattato alla domanda che il ricercatore si pone in relazione al sistema biologico oggetto di studio. Una visualizzazione relativamente semplice che mi permette di valutare fra due o più contesti cellulari la statistica dei geni espressi nella sua globalità è il *diagramma di Venn*. Questa rappresentazione grafica è utile nelle analisi esplorative basilari sui profili di espressione genica, per confrontare geni espressi o liste di geni differenzialmente espressi.

Anche lo *scatter plot* consente di visualizzare graficamente differenze nel profilo di espressione genica tra due contesti cellulari. Il *clustering gerarchico* per-

mette di visualizzare graficamente il grado di similitudine tra i profili di espressione di più campioni o di più geni. Il grado di similitudine reciproca tra i differenti profili di espressione è indicato dalla lunghezza dei rami del dendrogramma. Con lo stesso data set utilizzato per il clustering gerarchico posso, mediante *l'analisi delle componenti principali (PCA)* stabilire la distanza Euclidea fra i diversi insiemi di geni. Se l'esperimento che ho programmato prevede diversi campioni che fanno parte ad esempio di una cinetica di induzione del differenziamento cellulare posso valutare mediante l'uso di programmi analitici specifici, come Gene Spring, l'andamento coerente di insiemi di geni che vengono riuniti per tipologia di espressione (indotti, silenziati o costitutivamente espressi).

Annotazioni funzionali e classificazione dei geni in famiglie

L'annotazione funzionale dei geni è un passaggio obbligato verso la comprensione delle interazioni tra i geni e i prodotti genici, come realizzato attraverso gli studi di espressione genica e di proteomica, che è fondamentale per la comprensione dell'attivazione o inattivazione di funzioni cellulari specifiche ed in particolare di moduli funzionali (insieme di geni che permettono l'attivazione di un processo biologico). Attualmente il consorzio Gene Ontology (GO) ha in atto un progetto per annotare tutte le caratteristiche funzionali del gruppo completo di geni identificati dai diversi progetti genoma e quindi anche quello umano (HGP); il sito di GO è accessibile all'indirizzo <http://www.gene-ontology.org>. Tale sito classifica le funzioni dei geni in tre categorie principali, che sono:

- 1) processo biologico, cioè il processo in cui rientra la funzione della proteina, ad esempio proliferazione, differenziamento, apoptosi, immunità naturale ecc.;
- 2) funzione molecolare, che definisce l'attività della proteina, ad esempio fattore trascrizionale, molecola d'adesione, recettore, enzima, proteina legante il calcio ecc.;
- 3) componente cellulare, che si riferisce alla localizzazione della proteina nei diversi compartimenti cellulari, ad esempio nucleo, citoplasma, mitocondri, nucleolo, lisosomi, reticolo, membrane ecc.

Ciascuna di queste categorie è organizzata in modo gerarchico dalla funzione generale a quella più specifica. L'organizzazione del sito è dinamica e prevede un continuo aggiornamento delle categorie, che va di pari passo con la disponibilità di nuove acquisizioni. Le annotazioni funzionali dei geni sono riportate anche in altri siti quali ad esempio Genecards <http://www.genecards.org/index.shtml> e OMIM (On line Mendelian Inheritance in Men) <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM>.

Sono disponibili programmi di analisi dei dati, disponibili in rete, utili allo scopo di classificare funzionalmente i geni come GenMAPP o ricostruire vie di trasduzione del segnale, MAPPFinder (<http://www.GenMAPP.org>), o NetAffx, software di analisi funzionale dei geni a disposizione degli utenti della piattaforma Affymetrix, che sono fondamentale per l'interpretazione dei dati di espressione genica.

Le mappe trascrizionali

Un ulteriore sviluppo dell'analisi dei dati ottenuti dai profili di espressione genica, reso possibile dallo sviluppo di nuovi softwares, è l'integrazione dei dati relativi al trascrittoma con le informazioni della localizzazione dei geni sul DNA genomico. Tale tipo di analisi permette di stabilire interessanti correlazioni tra posizione genomica, funzione biologica e pattern di espressione genica. Con questa metodica di analisi applicata a diversi sistemi biologici si sono individuati domini cromatinici iperespressi o silenti.

Bibliografia

1. International Human genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. *Nature* 431: 931-945, 2004.
2. Lockhart DJ and Winzeler EA. Genomics, gene expression and DNA arrays. *Nature* 405: 827-836, 2000.
3. Quackenbush J. Computational analysis of microarray data. *Nature Reviews in Genetics* 2: 418-427, 2001.
4. Mark Schena *Microarray Analysis*, Wiley-Liss 2003.
5. Allison DB, Cui X, Page GP, Sabripour M. Microarray data analysis : disarray to consolidation and consensus. *Nature Reviews in Genetics* 7: 55-65, 2006.
6. Doniger SW, Salomonis N, Dahlquist KD, Vranizan K, Lawlor SC, Conklin BR. MAPPFinder: using Gene Ontology and GenMAPP to create a global gene-expression profile from microarray data. *Genome Biology* 4: R7, 2003.
7. Ashurst JL, Collins JE. Gene annotation: prediction and testing. *Annu. Rev. Genomics Hum. Genet.* 4:69-88, 2003.
8. Ashburner M, Ball CA, Blake JA. et al Gene Ontology: tool for unification of biology. *Nature Genetics* 25: 25-29, 2000.
9. Yamishita T, Honda M, Takatori H, Nishino R, Hoshino N, Kaneko S. Genome wide transcriptome mapping analysis identifies organ specific gene expression pattern along human chromosomes. *Genomics* 84: 867-875, 2004.

**ESPRESSIONE GENICA
E MUTAZIONI**

Analisi di geni singoli con la PCR quantitativa: espressione e monitoraggio di BCR-ABL nelle leucemie Ph1 positive

Fabrizio Pane

Dipartimento di Biochimica e Biotecnologie Mediche, Università di Napoli Federico II

Il cromosoma Philadelphia

Il cromosoma Philadelphia (Ph) è un cromosoma anomalo che si forma in seguito ad una traslocazione cromosomica bilanciata tra i cromosomi 9 ed il 22, e determina la fusione, sul cromosoma 22 derivativo, tra i geni BCR ad ABL. Il cromosoma Ph si può trovare nelle cellule neoplastiche di almeno tre distinti tipi di patologie leucemiche: la leucemia mieloide cronica (LMC), la leucemia linfoblastica acuta (LLA), e la leucemia mieloide cronica a neutrofili (LMC-N) (1). Numerose evidenze sperimentali e cliniche indicano che il cromosoma Ph ed in particolare, il gene di fusione BCR/ABL che si forma su di esso, è patogeneticamente correlato con l'insorgenza della malattia e costituisce pertanto anche un marcatore diagnostico del clone leucemico. La caratterizzazione molecolare dei geni BCR ed ABL ha mostrato che, nelle cellule normali, essi regolano la funzione di una cascata di altri geni o proteine appartenenti a pathways intracellulari molteplici e a volte divergenti coinvolti nei meccanismi di regolazione della proliferazione, differenziazione ed apoptosi delle cellule emopoietiche; questi studi hanno da una parte consentito di identificare almeno in parte i meccanismi della trasformazione neoplastica, e dall'altra fornito il razionale per una più moderna e mirata terapia di questi tipi di leucemie. Il gene di fusione BCR/ABL che si forma sul cromosoma Ph codifica per una proteina chimerica ad attività oncogenica che, a differenza della normale proteina ABL la cui attività tirosin-chinasi è strettamente regolata, ha un'attività chinasi costitutivamente attivata, e determina la fosforilazione un grande numero di substrati intracitoplasmatici, tra cui molecole adattatrici come Grb-2 e CRKL, CBL e SHC, proteine associate con l'organizzazione del citoscheletro e della membrana cellulare, come la paxillina e la talina ed un terzo gruppo di proteine che hanno funzione catalitica come la PI-3K, STAT-1 e -5. Inoltre, il dominio SH2 di Grb2 collega la proteina BCR/ABL al pathway di Ras (1). La rete della trasduzione del segnale controllata da BCR/ABL è molto complessa e queste proteine oncogeniche sembrano alte-

rare pathway molteplici e divergenti che controllano la proliferazione (come ad esempio le MAP e le JUN chinasi e c-Myc), la differenziazione (PI3 chinasi e paxillina) e apoptosi (1). L'attività tirosino-chinasi costitutivamente attivata gioca pertanto un ruolo centrale nella trasformazione leucemica, e ciò ha portato allo sviluppo di una serie di farmaci che agiscono direttamente sul difetto molecolare e cioè mediante inibizione selettiva dell'attività tirosino-chinasi della proteina oncogenica BCR/ABL. In tal senso le leucemie Ph positive costituiscono un esempio paradigmatico di come i dati scientifici ottenuti nei laboratori di ricerca sono rapidamente stati applicati alla pratica clinica.

Il gene BCR/ABL come marcatore di cellule leucemiche

È da sottolineare che, anche prima della sua definizione come bersaglio terapeutico, il gene BCR/ABL è stato utilizzato come marcatore delle cellule leucemiche in tecniche diagnostiche avanzate basate sulla reazione di amplificazione genica. A partire dalla fine degli anni 80, è stato infatti chiaramente dimostrato che nei pazienti con LMC si poteva dimostrare la presenza in modo riproducibile, oltre che del cromosoma Ph, anche del trascritto del gene BCR/ABL, utilizzando la reazione di amplificazione PCR delle regioni di giunzione tra le sequenze BCR e ABL del cDNA ottenuto dalle cellule leucemiche (2). Ben presto, fu evidente che la tecnica PCR poteva essere utilmente affiancata alla analisi citogenetica per la diagnosi della malattia, e che, anzi, le due tecniche avevano aspetti complementari. L'analisi citogenetica conserva ancora la tutta la sua validità, perché consente di apprezzare oltre che la presenza del cromosoma Ph, anche quella di eventuali anomalie cromosomiche aggiuntive, la cui presenza ha un valore prognostico rilevante. Tuttavia l'analisi citogenetica consente di dimostrare la presenza del cromosoma Ph in circa il 90-95% dei pazienti non trattati. Nei restanti casi, il cariotipo dimostra la presenza di traslocazioni cromosomiche complesse, difficili da interpretare senza ricorrere alle complesse e costose tecniche di FISH, o più spesso un cariotipo apparentemente normale (microtraslocazioni). In tutti questi casi l'analisi molecolare consente sempre di dimostrare la presenza del trascritto BCR/ABL causativo di malattia e confermare pertanto la diagnosi di LMC (3).

La malattia minima residua

Accanto alla possibilità di dimostrare la presenza di traslocazioni Ph criptiche (traslocazioni complesse o microtraslocazioni), fu anche evidente che la ricerca del trascritto ibrido BCR/ABL mediante tecnica di PCR aveva anche una elevata sensibilità analitica, e consentiva perciò di dimostrare la presenza di un ridotto numero di cellule leucemiche Ph+ (fino ad una cellula Ph+ in 10^5) in pazienti in remissione citogenetica completa, come ad esempio quelli trattati con trapianto di cellule staminali (4). Questo dato si rivelò rapidamente di grande interesse nel follow-up clinico dei pazienti sottoposti a trapianto allogenico di cellule staminali (Allo-SCT) e cioè trattati intensivamente. Pertanto, accanto ad un indubbio valore diagnostico, la tecnica di PCR con la sua possibilità di evidenziare la presenza

di ridottissimi quantitativi di cellule leucemiche residue ha assunto rapidamente un importante valore prognostico. I dati clinici dimostravano infatti che nel 20-30% dei pazienti con LMC in remissione clinica e citogenetica dopo Allo-SCT, si poteva dimostrare un residuo di cellule leucemiche mediante RT-PCR, e che la associazione tra la presenza del residuo di cellule e la ricaduta clinica era correlato a due fattori: il tempo dal trapianto ed il tipo di trapianto (5).

In uno studio condotto su 364 pazienti con LMC e sottoposti a Allo-SCT, fu dimostrato che, se a 3 e 6 mesi dal trapianto la presenza del trascritto BCR/ABL non aveva alcun impatto sulla probabilità di ricaduta e quindi sulla prognosi, nella finestra temporale compresa tra 6 e 12 mesi dal trapianto, la persistente positività alla reazione PCR si associava ad un più elevato rischio di ricaduta (42% vs 3% nei PCR negativi; $p < 0,0001$) ed a una più breve sopravvivenza (74% vs 83%; $p = 0,002$). All'opposto la persistente positività a più di 12 mesi dal trapianto si associa ad una probabilità di recidiva molto bassa (6). Questi risultati, confermati successivamente anche da altri gruppi, indicavano quindi che i pazienti con LMC hanno frequente positività alla PCR per un residuo di malattia nei primi 6 mesi dopo il trapianto, ma che successivamente almeno i 2/3 di questi convertono ad una negatività all'esame PCR probabilmente perché le cellule Ph+ residue sono state eradicade dal sistema immunitario del donatore (Graft Vs Leukemia reaction - GVL) (7, 8). L'importanza della GVL in questi casi è stata poi ulteriormente supportata dai dati della malattia residua nei pazienti che ricevevano un trapianto allogenico T-depleto. In seguito a questo tipo di trapianto, la positività PCR per la malattia residua dopo 6 mesi dalla procedura arriva al 80% (9-11). Nell'insieme tutti questi dati indicavano chiaramente la importanza della PCR per BCR/ABL per la definizione prognostica della LMC.

Molto ridotto era però il significato clinico del monitoraggio mediante PCR qualitativa del residuo di cellule Ph+ nei pazienti non eleggibili ad Allo-SCT e trattati con chemoterapia tradizionale o Interferone (IFN). La ridotta efficacia della terapia e la elevata sensibilità della tecnica PCR, fanno sì che in tutti i pazienti trattati con idrossiurea ed nella stragrande maggioranza di quelli trattati con IFN, il residuo di cellule Ph+ sia sempre evidente alla reazione di PCR (12). È tuttavia importante ricordare che una piccola frazione di casi trattati con IFN mantengono una remissione citogenetica completa per un lungo periodo (superiore a 6-7 anni); tra questi una frazione significativa di casi convertono ad una negatività della analisi PCR per il residuo di cellule Ph+ e possono essere considerati curati, poiché mantengono la remissione anche alla sospensione del trattamento (13).

Monitoraggio quantitativo della malattia residua

Numerosi sforzi furono compiuti durante gli anni 90 per studiare in modo quantitativo la riduzione delle cellule Ph+ mediante PCR, allo scopo di valutare il significato prognostico della cinetica delle variazioni della massa Ph+ durante il trattamento. Tuttavia la tecnologia di PCR tradizionale con la rilevazione dei prodotti di amplificazione alla fine della reazione (tecniche end-point) mal si adattava allo sviluppo di metodologie che uniscono una elevatissima sensibilità alla

capacità di misurare il numero di copie del trascritto ibrido BCR/ABL, in modo da fornire una stima quantitativa del residuo di cellule Ph+. Pertanto, nonostante l'utilizzo di una serie di artifici tecnici (ad es. PCR competitiva), le tecniche disponibili per la valutazione quantitativa delle cellule Ph+ erano piuttosto imprecise e molto indaginose e lo studio cinetico della malattia residua è stato limitato ed applicato a ristretti gruppi di pazienti.

Tuttavia, a cavallo con l'inizio degli anni 2000, e quindi in contemporanea con l'introduzione nella sperimentazione prima, e nella pratica clinica poi, della terapia con imatinib, divenne disponibile una nuova tecnologia di amplificazione genica, la "Real Time PCR". Questa è una metodologia innovativa e sensibile basata sulla reazione 5' nucleasica della Taq polimerasi e sull'impiego di un probe fluorogenico (Taqman probe) (14). Ad ogni ciclo di PCR, l'attività 5'-esonucleasica della Taq polimerasi, degrada il probe determinando l'aumento della fluorescenza emessa dai fluorocromi legati al probe. L'accumulo dei prodotti di amplificazione produce un incremento dell'emissione di fluorescenza che è monitorata in tempo reale durante la PCR da uno strumento in grado di generare un grafico corrispondente all'intensità di emissione di fluorescenza durante i vari cicli di PCR. Il numero dei cicli al quale l'emissione di fluorescenza raggiunge un dato valore soglia (CT) viene utilizzata come unità di misura dell'accumulo dei prodotti di amplificazione ed è inversamente proporzionale al numero iniziale delle molecole target presenti nel campione. La reazione di "Real Time PCR", nella sua preparazione, se non per i reagenti, non si discosta molto da una reazione di PCR tradizionale e si è rivelata estremamente riproducibile ed accurata. È stata quindi rapidamente adattata per lo sviluppo di una metodologia per la stima quantitativa del numero di copie di trascritto ibrido BCR/ABL nei campioni di sangue midollare o periferico dei pazienti con LMC. Agli inizi degli anni 2000, quindi, come risultato di uno sforzo cooperativo di 20 laboratori europei interessati allo sviluppo di metodologie innovative in oncematologia, è stata sviluppata e validificata una metodologia per la misura della malattia residua basata sulla quantizzazione dei geni di fusione (15).

La malattia residua minima nei pazienti con LMC trattati con imatinib

La terapia con imatinib ha radicalmente cambiato le prospettive prognostiche dei pazienti affetti da LMC; la sperimentazione clinica ha molto rapidamente dimostrato sia nei pazienti trattati tardivamente dopo la perdita della risposta a IFN (fase cronica tardiva) ed a maggior ragione in quelli trattati front-line (fase cronica precoce), la impressionante superiorità di questo farmaco su tutti gli altri precedentemente impiegati (16, 17).

Gli studi multicentrici internazionali di fase 2 in pazienti in fase cronica tardiva, resistenti o intolleranti ad IFN, hanno dimostrato che la risposta al farmaco è molto rapida ed efficace; praticamente tutti i pazienti raggiungono la remissione ematologica completa e quasi i 2/3 dei pazienti trattati mostrano una conversione cariotipica maggiore (<33% di cellule Ph+ - MCR) dopo un solo anno di trattamento. È inoltre importante notare che quasi la metà dei pazienti trattati in fase

cronica tardiva mostrano una conversione cariotipica completa (CCR) dopo un solo anno di trattamento (17). Il gruppo cooperatore italiano ha condotto una analogo sperimentazione di fase 2 in 200 pazienti con LMC in fase cronica tardiva (resistenti o refrattari ad IFN) per studiare la cinetica della risposta mediante PCR quantitativa in Real Time (RQ-PCR) (18). Nei 191 pazienti valutabili, i risultati confermavano l'elevata efficacia del trattamento: il 97% dei pazienti otteneva una risposta ematologica completa, nel 64% dei casi il farmaco induceva in 12 mesi di trattamento una MCR ed nel 44% dei pazienti una CCR. La risposta citogenetica si dimostrava stabile nell'89% dei pazienti CCR durante il primo anno di trattamento, e il 79% conservava la risposta alla fine del secondo anno di follow-up (18). Lo studio della cinetica di risposta al trattamento, valutata mediante RQ-PCR, ha consentito di ottenere una serie di informazioni di elevato significato speculativo e clinico. Innanzitutto, è stato dimostrato che sebbene non sia stata osservata alcuna differenza nella quantità di trascritto ibrido BCR/ABL, misurata prima del trattamento nei pazienti che raggiungono la CCR e nei pazienti che raggiungono la MCR, i 31 pazienti, che non hanno avuto alcuna risposta al trattamento con Imatinib, presentavano livelli di trascritto ibrido BCR/ABL significativamente più alti rispetto alle altre categorie di pazienti (19). I dati del monitoraggio mediante RQ-PCR indicavano inoltre, che i pazienti che raggiungevano la risposta citogenetica parziale o instabile, il livello della massa leucemica residua valutata mediante quantizzazione del trascritto ibrido BCR/ABL rimaneva elevato durante tutto il periodo del trattamento (18). Nei pazienti che raggiungevano una risposta citogenetica parziale si osservava un decremento del residuo leucemico inferiore a due logaritmi alla fine dello studio. I pazienti che ottenevano la CCR stabile, mostravano livelli di malattia residua che si riducevano molto rapidamente: più di un logaritmo dopo tre mesi e più di tre dopo 6 mesi. Inoltre il monitoraggio della risposta durante il secondo anno di trattamento, indicava che nei casi in cui si era osservata una rapida risposta citogenetica completa si poteva osservare il rapido decremento del residuo di malattia che si riduceva rapidamente di oltre 4 logaritmi. Il monitoraggio quantitativo dimostrava inoltre che anche nei pazienti con lenta conversione alla CCR si poteva osservare dopo 2 anni di trattamento la riduzione di quasi 4 logaritmi della massa leucemica residua (18). Ciò dimostra che il farmaco ha un'efficacia progressiva sul clone neoplastico. Infine si dimostrava una correlazione inversa tra il livello di malattia residua valutato mediante RQ-PCR e tendenza alla recidiva. Infatti nei pazienti con la più elevata riduzione della massa neoplastica, la perdita della risposta al farmaco era osservata molto sporadicamente (18).

Come atteso, il trattamento con Imatinib in fase cronica precoce (front-line) si è rivelato ancora più efficace; molte delle attuali conoscenze sulla cinetica di risposta al trattamento front-line derivano dallo studio multicentrico internazionale "International Randomized Study of Interferon and STI571 (IRIS)", che ha randomizzato 1106 pazienti con CML di nuova diagnosi in due bracci di trattamento, 553 nel braccio imatinib (400 mg/die) ed i rimanenti 553 nel braccio IFN/Ara-C (il trattamento standard dei pazienti negli anni 90) (20). Dopo soli 18 mesi di trattamento la superiorità dei risultati del braccio imatinib era già molto eviden-

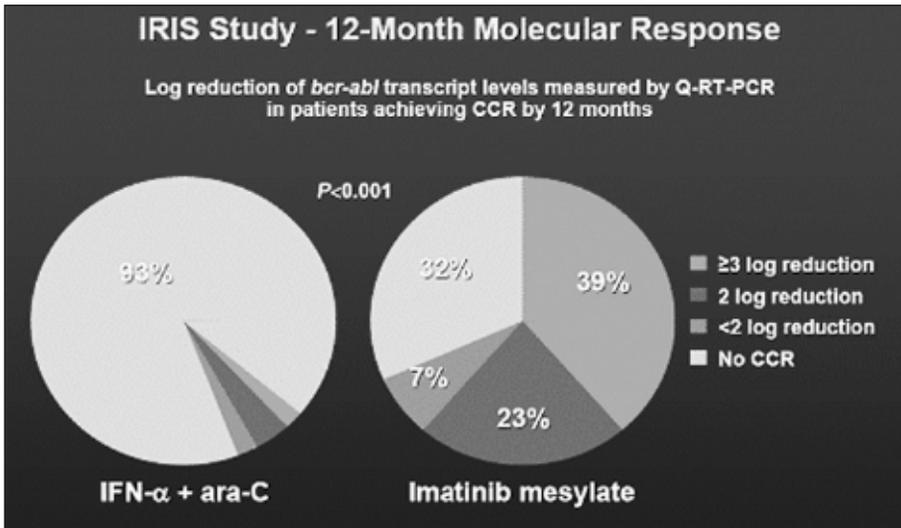


Fig. 1

CCR = complete cytogenetic response Hughes et al.
N. Engl. J. Med. 2003; 349: 1423.

te; la percentuale di MCR e di CCR nei due bracci di trattamento era rispettivamente 87,1% vs 34,7% ($p < 0,001$) e 76,2% vs 14,5% ($p < 0,001$) (20). Lo studio della malattia residua in questi pazienti mediante RQ-PCR mostrava alcuni importanti aspetti; la riduzione superiore a tre logaritmi della massa neoplastica residua, definita risposta molecolare maggiore (MMR), era osservata a 12 mesi

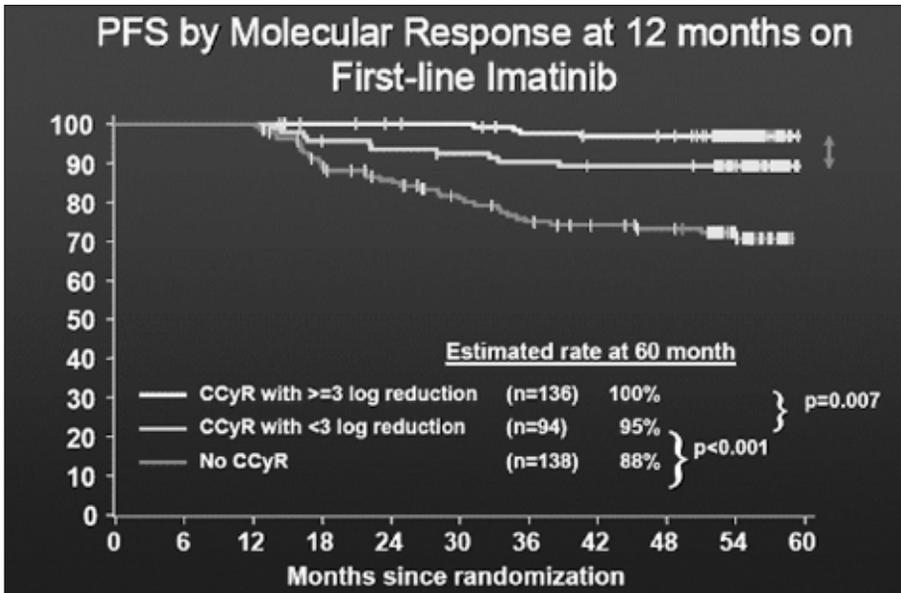


Fig. 2

nel 39% dei casi trattati con imatinib ma solo nel 2% di quelli del braccio IFN/Ara-C (21) (Figura 1).

Il follow-up successivo, necessariamente ristretto ad un gruppo più limitato di pazienti del braccio imatinib, dimostrava che la riduzione mediana della malattia residua nei pazienti in CCR era progressiva nel tempo senza però raggiungere un plateau fino ad un follow-up di 48 mesi (22), così come già dimostrato dal gruppo cooperatore italiano nei pazienti trattati in fase cronica tardiva (18). Tuttavia, la percentuale di casi in cui il livello di malattia residua era al di sotto della sensibilità della tecnica RQ-PCR (-4,5 log), definito risposta molecolare completa (CMR), era molto limitato, indicando che la malattia sembra essere molto difficilmente eradicabile. Il livello raggiunto dalla malattia residua ha un elevato impatto sulla "progression free survival" (PFS); la MMR entro i primi 12 mesi di trattamento è predittiva di PFS a 60 mesi, e la maggior parte degli eventi di progressione (perdita di risposta clinica o progressione a fasi avanzate di malattia) riguardano i pazienti che non ottengono una riduzione superiore a 2 log della malattia residua nei 6 mesi di trattamento, e che raramente raggiungono la CCR (23) (Figura 2)

Gli studi di cinetica mediante Q-PCR hanno anche mostrato una buona correlazione tra i livelli di malattia residua misurata nel sangue periferico con quelli misurati nel sangue midollare, aprendo quindi la possibilità di un più agevole monitoraggio di questi pazienti, che possono quindi evitare ulteriori mielobiopsie dopo l'ottenimento della CCR, se il livello di malattia residua misurato nel sangue periferico si mantiene stabile (25, 25).

La malattia residua minima nei pazienti con LLA Ph+

La leucemia linfoblastica acuta Ph+ (ALL) appartiene ad un sottogruppo delle ALL con prognosi infausta ed alto rischio di fallimento del trattamento. Nonostante le alte dosi di chemioterapia inducano una risposta ematologica completa (CHR) nel 70-80% dei pazienti ALL Ph+, molti di questi pazienti ricadono precocemente. Recenti studi hanno mostrato che le ALL Ph+ mostrano una risposta eterogenea al trattamento e che la risposta al prednisone e la risposta molecolare consentono di prevedere la durata della remissione. Per questo motivo, il gruppo GIMEMA ha iniziato un'attenta valutazione della cinetica di riduzione del livello di malattia residua in un gruppo di 45 pazienti affetti da ALL Ph+ per verificare se la valutazione precoce della chemiosensibilità attraverso la RQ-PCR può essere utilizzata per prevedere l'outcome e per identificare il rischio di una precoce ricaduta dei pazienti (26). I livelli di malattia residua sono stati valutati nel midollo dei pazienti mediante la tecnica di RQ-PCR nelle prime fasi di trattamento (alla diagnosi, dopo induzione e dopo consolidamento). Sono stati esclusi dalla valutazione finale quattro pazienti resistenti all'induzione e due con un follow-up clinico inferiore a sei mesi. I 39 pazienti valutabili per l'analisi dell'outcome (mediana del follow-up di 14 mesi, range 6-48), presentavano alla diagnosi un livello medio di trascritto BCR/ABL di 120 (range 23-1030). I pazienti sono poi stati stratificati in due gruppi sulla base della cinetica di riduzione della massa

leucemica residua: i 27 pazienti con un livello inferiore a 0,01 (>2 log di riduzione) dopo l'induzione, e inferiore a 0,001 (>3 log di riduzione) dopo il consolidamento sono stati considerati "good molecular responders" (GMRs), mentre i restanti 12 casi che, nonostante l'ottenimento della risposta ematologica completa, avevano un minore decremento della malattia minima residua ad entrambi i time point, sono stati considerati "poor molecular responders" (PMRs). Il follow up ha mostrato che la precoce riduzione della massa residua leucemica costituisce un parametro prognostico indipendente e statisticamente significativo in questi pazienti.

Nei GMRs la probabilità attuariale della sopravvivenza libera da ricaduta (RFS), sopravvivenza libera da malattia (DFS) e sopravvivenza totale (OS) a due anni era del 31%, 44%, e 44%, paragonata allo 0%, 0%, and 0% nei PMRs ($p=0.0029$, 0,001 e 0,03, rispettivamente) (26). Alla fine dello studio 13 pazienti appartenenti al gruppo dei GMR e solo uno dei PMR erano vivi. Inoltre, solo in quattro casi, tutti appartenenti al gruppo dei GMR, la terapia di salvataggio ha indotto una seconda remissione completa (26). I dati indicano che, anche le ALL Ph+ dell'adulto comprendono pazienti con un sensibilità eterogenea al trattamento e che la valutazione quantitativa precoce della malattia minima residua è un eccellente parametro prognostico in questa malattia.

Considerazioni conclusive

Il monitoraggio della malattia residua mediante RQ-PCR è diventata uno strumento indispensabile per la corretta e moderna gestione della terapia nei pazienti con leucemie Ph+. È comunque da notare che, nonostante la tecnologia di RQ-PCR sia stata notevolmente rifinita e migliorata nel tempo rispetto all'era dei primi pionieristici studi della fine degli anni 80, la tecnologia è ancora da ritenere dedicata a centri specializzati. Inoltre è ancora in corso uno sforzo collaborativo tendente alla standardizzazione interlaboratorio dei risultati mediante l'uso di appropriati standard di riferimento internazionali.

Bibliografia

1. Pane F, Intrieri M, Quintarelli C, et al. BCR/ABL genes and leukemic phenotype: from molecular mechanisms to clinical correlations. *Oncogene* 2002; 21(56): 8652-67.
2. Potter MN, Cross NCP, van Dongen JJ, et al. Molecular evidence of minimal residual disease after treatment for leukaemia and lymphoma: an updated meeting report and review. *Leukemia* 1993; 7: 1302-14.
3. Faderl S, Talpaz M, Estrov Z, et al. The biology of chronic myeloid leukemia. *N Engl J Med* 1999; 341: 164-72.
4. Lin F, Goldman JM, Cross NCP: A comparison of the sensitivity of blood and bone marrow for the detection of minimal residual disease in chronic myeloid leukaemia. *Br J Haematol* 1994; 86: 683.
5. Oehler VG, Radich JP. Monitoring bcr-abl by polymerase chain reaction in

- the treatment of chronic myeloid leukemia. *Curr Oncol Rep* 2003; 5: 426 - 35.
6. Radich JP, Gooley T, Bryant E, et al. The significance of bcr-abl molecular detection in chronic myeloid leukemia patients "late," 18 months or more after transplantation. *Blood* 2001; 98: 1701-7.
 7. Cross NCP, Lin F, Chase A, et al. Competitive polymerase chain reaction to estimate the number of BCR-ABL transcripts in chronic myeloid leukaemia after bone marrow transplantation. *Blood* 1993; 82: 1929-36.
 8. Reiter E, Greinix HT, Keil F, et al. Long-term follow up of patients after related- and unrelated donor bone marrow transplantation for chronic myelogenous leukemia. *Ann Hematol* 1999; 78: 507-13.
 9. Pichert G, Roy D-C, Gonin R, et al. Distinct patterns of minimal residual disease associated with graft-versus-host disease after allogeneic bone marrow transplantation for chronic myelogenous leukemia. *J Clin Oncol* 1995; 13: 1704-13.
 10. Gratwohl A, Brand R, Apperley J, et al. Graft-versus-host disease and outcome in HLA-identical sibling transplantations for chronic myeloid leukemia. *Blood* 2002; 100: 3877-86.
 11. Mackinnon S, Barnett L, Heller G. Polymerase chain reaction is highly predictive of relapse in patients following T cell-depleted allogeneic bone marrow transplantation for chronic myeloid leukemia. *Bone Marrow Transplant* 1996; 17: 643-7.
 12. Chomel JC, Brizard F, Veinstein A, et al. Persistence of BCR-ABL genomic rearrangement in chronic myeloid leukemia patients in complete and sustained cytogenetic remission after interferon-alpha therapy or allogeneic bone marrow transplantation. *Blood* 2000; 95: 404-8.
 13. Bonifazi F, De Vivo A, Rosti G, et al. European Study Group on Interferon in Chronic Myeloid Leukemia. Chronic myeloid leukemia and ainterferon: a study of complete cytogenetic responders. *Blood* 2001; 98: 3074-81.
 14. Heid CA, Stevens J, Livak KJ, et al. Real time quantitative PCR. *Genome Res* 1996; 6: 986-94.
 15. Gabert J, Beillard E, van der Velden VH, et al. Standardization and quality control studies of 'real-time' quantitative reverse transcriptase polymerase chain reaction of fusion gene transcripts for residual disease detection in leukemia - a Europe Against Cancer program. *Leukemia* 2003; 17(12): 2318-57.
 16. Druker BJ, Talpaz M, Resta DJ, Peng B, Buchdunger E, Ford JM, et al. Efficacy and safety of a specific inhibitor of the BCR-ABL tyrosine kinase in chronic myeloid leukemia. *N Engl J Med* 2001; 344: 1031-7.
 17. Kantarjian H, Sawyers C, Hochhaus A, Guilhot F, Schiffer C, Gambacorti-Passerini C, et al. Hematologic and cytogenetic responses to imatinib mesylate in chronic myelogenous leukemia. *N Engl J Med* 2002; 346: 645-52.
 18. Rosti G, Martinelli G, Bassi S, Amabile M, Trabacchi E, Giannini B, Cilloni D, Izzo B, De Vivo A, Testoni N, Cambrin GR, Bonifazi F, Soverini S, Luatti S, Gottardi E, Alberti D, Pane F, Salvatore F, Saglio G, Baccarani M.

- Molecular response to imatinib in late chronic-phase chronic myeloid leukemia. *Blood* 2004; 103(6): 2284-90.
19. Iacobucci I, Rosti G, Amabile M, et al. Comparison between patients with Philadelphia-positive chronic phase chronic myeloid leukemia who obtained a complete cytogenetic response within 1 year of imatinib therapy and those who achieved such a response after 12 months of treatment. *J Clin Oncol*. 2006; 24: 454-459.
 20. O'Brien SG, Guilhot F, Larson RA, Gathmann I, Baccarani M, Cervantes F, et al. Imatinib compared with interferon and low-dose cytarabine for newly diagnosed chronic-phase chronic myeloid leukemia. *N Engl J Med* 2003; 348: 994-1004.
 21. Hughes T, Kaeda J, Branford S, Rudzki Z, Hochhaus A, Hensley M, et al. Frequency of major molecular responses to imatinib or interferon alpha plus cytarabine in newly-diagnosed chronic myeloid leukemia. *N Engl J Med* 2003; 349: 142-32.
 22. Goldman J et al, Continuing Reduction in Level of Residual Disease after 4 Years in Patients with CML in Chronic Phase Responding to First-Line Imatinib (IM) in the IRIS Study. *ASH 2005*; Abs#2345.
 23. Druker BJ, Guilhot F, O'Brien SG, Gathmann I, Kantarjian H, et al. Five-year follow-up of patients receiving imatinib for chronic myeloid leukemia. *N Engl J Med* 2006; 355(23): 2408-17.
 24. Branford S, Hughes TP, Rudzki Z. Monitoring chronic myeloid leukaemia therapy by real-time quantitative PCR in blood is a reliable alternative to bone marrow cytogenetics. *Br J Haematol* 1999; 107: 587-99.
 25. Kantarjian HM, Talpaz M, Cortes J, et al. Quantitative polymerase chain reaction monitoring of BCR-ABL during therapy with imatinib mesylate (STI571; Gleevec) in chronic-phase chronic myelogenous leukemia. *Clin Cancer Res* 2003; 9: 160-6.
 26. Pane F, Cimino G, Izzo B, et al. Significant reduction of the hybrid BCR/ABL transcripts after induction and consolidation therapy is a powerful predictor of treatment response in adult Philadelphia-positive acute lymphoblastic leukemia. *Leukemia*. 2005; 19: 628-635.

Analisi in parallelo dell'espressione genica: apporto dei DNA microarray in oncologia

Sergio Ferrari

Dipartimento di Scienze Biomediche, Sezione di Chimica Biologica,
Università di Modena e Reggio Emilia, Modena

La ricerca biologica e biomedica è in un momento di transizione significativo, determinato da due principali fattori: la conoscenza della sequenza del DNA genomico umano e lo sviluppo di nuove tecnologie per utilizzarla.

Sfortunatamente, la conoscenza della sequenza del DNA non fornisce informazioni su come i singoli geni agiscano, su come la cellula operi, su come le cellule si organizzino a formare organismi, su cosa non “funzioni” in caso di malattia, o come poter dirigere lo sviluppo di un farmaco. Entrano quindi in gioco gli studi funzionali, il cui scopo non è semplicemente quello di fornire un elenco di tutti i geni e le informazioni sulle loro funzioni, ma di capire come le componenti lavorino insieme a formare cellule ed organismi funzionanti. Tra gli strumenti più efficaci e versatili per questo genere di studi vi sono gli arrays di oligonucleotidi o di DNA complementare, chiamati DNA microarrays, microarrays di oligonucleotidi o chips. L'insieme dei geni trascritti dal DNA genomico è indicato come profilo d'espressione o “trascrittoma”, e caratterizza il fenotipo e la funzione della cellula. Al contrario del genoma, il “trascrittoma” è molto dinamico e cambia rapidamente e drammaticamente in risposta a qualsiasi cambiamento od anche durante eventi cellulari, come la replicazione del DNA, la divisione cellulare, l'apoptosi o la trasformazione neoplastica. Inoltre, cambiamenti nel quadro multi genico dell'espressione possono fornire indizi circa i meccanismi regolatori, le funzioni e le vie biochimiche cellulari. Ancora, studi sul profilo di espressione genica sono utili ad individuare cause e conseguenze delle malattie, come agiscono i farmaci sulle cellule e quali prodotti genici possono essere bersagli adatti per la terapia. Va notato tuttavia che l'RNA messaggero è soltanto un prodotto intermedio della via di produzione di una proteina, ma gli studi delle proteine si mostrano a tutt'oggi molto difficili e meno sensibili. I livelli di RNA messaggero offrono importanti informazioni sullo stato della cellula e, per la maggior parte dei geni, un aumento del livello trascrizionale si traduce in un aumento della quantità di RNA messaggero e di proteina tradotta. Il sequenziamento e l'analisi del genoma umano indicano che circa 25.000-30.000 geni vengono attivamente trascritti, potendo però

generare con lo splicing alternativo un numero maggiore di prodotti proteici. Analisi comparative in diversi contesti cellulari, utilizzando la tecnologia degli arrays, approfondiranno le conoscenze su processi biologici fondamentali, come la proliferazione, il differenziamento, il processo apoptotico e la trasformazione neoplastica. Storicamente, studi tesi in questa direzione hanno sempre proceduto un gene per volta, ma l'avvento di questa nuova tecnologia rende possibile raccogliere informazioni da migliaia di geni, disegnando un quadro globale dell'espressione genica e suggerendo nuovi meccanismi di regolazione genica. Ad esempio, è stato possibile indagare interessanti contesti biologici:

- a) espressione genica differenziale nelle fasi successive del ciclo cellulare o nel differenziamento cellulare;
- b) meccanismi molecolari alla base della trasformazione e della progressione neoplastica;
- c) identificazione dei geni espressi durante il programma genetico pro-apoptotico;
- d) modificazioni dell'espressione genica durante le infezioni virali; e individuazione dei geni interessati nel metabolismo di farmaci.

Le analisi di espressione e di clustering nei diversi contesti cellulari forniranno utili informazioni sulla complessità di molti programmi genetici ed aiuteranno l'individuazione delle loro alterazioni in caso di malattia. Esistono più di 200 tipi diversi di cancro e ciascuno può essere considerato un contesto cellulare unico, ciò che rende la classificazione tramite analisi con DNA microarrays un valido strumento nella lotta contro il cancro. Un recente studio ha preso in considerazione i profili di espressione mediante cDNA microarrays in 60 linee cellulari neoplastiche di diversa origine, anche emopoietica, con la prospettiva di comparare tali profili con i tumori originari.

Analisi di DNA microarray e classificazione dei tumori

Il potenziale di indagine dei DNA microarrays nello studio globale dell'espressione genica è stato recepito fin da subito ed è stato largamente usato per definire nuove sottoclassi di malattie, per identificare nuovi marcatori biologici associati a malattie o per distinguere sottoclassi all'interno della stessa patologia. In particolare, lo studio del PEG è stato utilizzato per classificare diversi tipi di tumore ("signature diagnostica") o per predire la risposta alla chemioterapia e quindi la sopravvivenza dei pazienti trattati ("signature prognostica"). Un esempio è lo studio di DNA microarrays condotto in Olanda su pazienti affette da carcinoma mammario. Tale studio ha permesso l'individuazione di una signature prognostica di 70 geni predittivi dello sviluppo di metastasi e quindi di un aggravamento del tumore. Tale evidenza ha portato a modificare gli schemi terapeutici ed ha una valenza predittiva molto superiore rispetto alla diagnostica convenzionale basata sull'esame istologico del tumore e sull'esame clinico. Un altro esempio, fra i molti che si potrebbero riportare, è la sottoclassificazione molecolare delle leucemie mieloidi acute. Tali leucemie sono tradizionalmente divise in 7 sottoclassi diverse in base ai criteri diagnostici convenzionali, che riguardano

lo studio della morfologia, l'espressione di enzimi cellulari evidenziata con coloranti cromogenici (citochimica), l'espressione di proteine di superficie riconosciute da anticorpi specifici (immunofenotipo), la citogenetica per lo studio delle alterazioni cromosomiche, e le metodiche molecolari per lo studio di alterazioni di geni specifici. Uno studio cooperativo fra diversi gruppi di ricerca, pubblicato recentemente, ha sottoclassificato tali leucemie in base al PEG in 16 sottogruppi che hanno una diversa risposta alla terapia, che va da una buona responsività alla refrattarietà al trattamento. Uno studio analogo, sempre condotto su pazienti affetti da leucemie mieloidi acute, ha invece individuato una "signature" prognostica di 133 geni predittivi che correla con la sensibilità dei pazienti alla terapia. Tuttavia deve essere ricordato che questo tipo di studi richiede una stretta collaborazione fra clinici, laboratoristi e bioinformatici, e che spesso più istituzioni devono collaborare a distanza per la raccolta di un numero sufficiente di pazienti da studiare. Lo schema riportato in illustra una "signature" mediante DNA microarrays. Un'altra applicazione importante dei DNA arrays riguarda la farmacogenomica, che tratteremo nel paragrafo successivo.

DNA microarrays e farmacogenomica

La farmacogenomica è lo studio dell'impatto della variabilità genomica individuale sulla risposta ai farmaci, sulla loro efficacia terapeutica e sul loro metabolismo, ed è forse da considerare una delle prime applicazioni del HGP nella pratica clinica in medicina. Negli anni cinquanta questo campo di studio era chiamato farmacogenetica, ma con l'aumento delle conoscenze sul genoma umano è stato adottato il termine di farmacogenomica. Con i DNA microarrays è virtualmente possibile esaminare la risposta dell'intero genoma della cellula quando esposta a farmaci e quindi stabilire come il farmaco è in grado di modificare l'espressione genica. Nel paragrafo precedente sono stati riportati alcuni esempi di come si possa arrivare ad una "signature" prognostica, cioè a definire un insieme di geni associati alla farmacoresistenza. Il relativamente nuovo settore della farmacogenomica contribuirà notevolmente a capire il meccanismo d'azione dei farmaci e il meccanismo di risposta o di adattamento delle cellule in risposta all'esposizione a farmaci. Infatti, le differenze nel profilo di espressione genica indotte da farmaci o da composti biologicamente attivi possono essere analizzate per individuare nuove attività di questi prodotti. Un simile approccio sperimentale, chiamato "forward pharmacology", può portare all'individuazione di nuovi effetti biologici di un farmaco già conosciuto e possibilmente a nuove applicazioni terapeutiche in considerazione dei geni coinvolti nell'attività del farmaco stesso. Inoltre possono essere individuati e studiati nuovi geni bersaglio. È altrettanto evidente che lo sviluppo di nuovi farmaci prevede necessariamente lo studio della loro interferenza con il genoma, e quindi lo studio del PEG diventerà indispensabile per individuare possibili effetti tossici e stabilire l'indice terapeutico. Quando questa metodologia è applicata allo studio dei tumori, ed in particolare allo sviluppo di nuove strategie terapeutiche, il termine più comunemente usato è Oncogenomica (Strausberg R.L. et al., Nature 2004).

Bibliografia

1. International Human genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. *Nature* 431: 931-945, 2004.
2. Lockhart DJ, and Winzeler EA. Genomics, gene expression and DNA arrays. *Nature* 405: 827, 2000.
3. Lipshutz RJ, Fodor SPA, Gingeras TR, Lockhart DJ. High density synthetic oligonucleotide arrays. *Nature Genetics (Suppl.)* 21: 20, 1999.
4. Epstein CB, Butow RA. Microarray technology - enhanced versatility, persistent challenge. *Curr. Opinion Biotech.* 11: 36, 2000.
5. Pandey A, Mann M. Proteomics to study genes and genomes. *Nature* 405: 837, 2000.
6. Young R.A. Biomedical discovery with DNA arrays. *Cell* 102: 9, 2000.
7. Cho RJ, Campbell MJ, Winzeler EA., Steinmetz L., Conway A., Wodicka L., Wolfsberg T.G., Gabriele A.E., Landsman D., Lockhart D.J., & Davis R.W. A genome-wide transcriptional analysis of the mitotic cell cycle. *Molecular Cell* 2: 65, 1998.
8. Zhu H, Cong J.P., Mamtora G., Gingeras T., Shenk T. Cellular gene expression altered by human cytomegalovirus: global monitoring with oligonucleotide arrays. *PNAS* 95: 14470, 1998.
9. Der SD, Zhou A, Bryan R, Williams G, Silverman RH. Identification of genes differentially regulated by interferon a, b, or g using oligonucleotide arrays. *PNAS* 95: 15623, 1998.
10. Tagliafico E, Tenedini E, Bergamaschi A, Manfredini R, Percudani R, Siena M, Zanocco-Marani T, Grande A, Montanari M, Gemelli C, Torelli U, Ferrari S. Gene expression profile of Vitamin D3 treated HL-60 cells shows a phenotypic but not a complete functional conversion to monocytes. Submitted.
11. Ross DT, Scherf U, Eisen MB, Perou CM, Rees C, Spellman P. Systematic variation in gene expression patterns in human cancer cell lines. *Nature Genetics* 24: 227, 2000.
12. Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, Coller H, Loh ML, Downing JR, Caligiuri MA, Bloomfield CD, & Lander ES. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* 286: 531, 1999.
13. Quackenbush J. Microarray analysis and tumor classification. *New England Journal of Medicine* 354: 2463-2472, 2006.
14. Liang P, Pardee AB. Analysing differential gene expression in cancer. *Nature Rev. Cancer* 3: 869-876, 2003.
15. Van de Vijver MJ, He YD, van't Veer LJ, et al. A gene expression signature as a predictor of survival in breast cancer. *New England Journal of Medicine* 347: 1999-2009, 2002.
16. Valk PJ, Verhaak RGW, Beijen MA, et al. Prognostically useful gene-expression profiles in acute myeloid leukemia. *New England Journal of Medicine* 350: 1617-1628, 2004.

17. Bullinger L, Dohner K, Bair E, et al. Use of Gene-expression profiling to identify prognostic subclasses in adult acute myeloid leukemia. *New England Journal of Medicine* 350: 1605-1616, 2004.
18. Tagliafico E, Tenedini E, Manfredini R, Grande A, Ferrari F, Roncaglia E, Biciato S, Zini R, Salati S, Bianchi E, Gemelli C, Montanari M, Vignudelli T, Zanocco-Marani T, Parenti S., Paolucci P, Martinelli G, Piccaluga PP, Baccarani M, Specchia G, Torelli U, Ferrari Ser. "Identification of a molecular signature predictive of sensitivity to differentiation induction in Acute Myeloid Leukemia. *Leukemia* 20: 1751-1758, 2006.
19. Ginsburg GS, Konstance RP., Allsbrook JS, Schulman KA. Implications of pharmacogenomics for drug development and clinical practice. *Archives of Internal Medicine* 165: 2331-2336, 2005.
20. Strausberg RL, Simpson AJG, Old LJ, Riggins GJ. Oncogenomics and the development of new cancer therapies. *Nature* 429: 469-474, 2004.

Gene expression profiling and mutations in acute myeloid leukemia prognostication

Pier Paolo Piccaluga, Michela Rondoni, Stefania Paolini, Claudio Laterza, Cristina Papayannidis, Lea Verlicchi, Francesco De Rosa, Giovanni Martinelli

Istituto di Ematologia e Oncologia Medica "L. e A. Seragnoli", Policlinico Sant'Orsola-Malpighi, Università di Bologna

The development of acute myeloid leukemia (AML) is associated with accumulation of acquired genetic alterations and epigenetic changes in hematopoietic progenitor cells that alter normal mechanisms of cell growth, proliferation, and differentiation.

At diagnosis, most patients with AML harbor at least 1 chromosome aberration in their marrow blasts. Numerous recurrent structural and numeric cytogenetic aberrations have been identified and many of them not only are diagnostic markers for specific AML subtypes but also constitute independent prognostic factors for attainment of complete remission (CR), relapse risk, and overall survival (OS).

However, in a sizable group of AML patients, 40% to 49% of adults and 25% of children, no microscopically detectable chromosome abnormality can be found on standard cytogenetic analysis (normal karyotype, NK, AML). Of note, however, in the last years, several gene mutations and changes in gene expression have been discovered that strongly affect clinical outcome of NK AML patients.

FMS-related tyrosine kinase 3 (FLT3) gene abnormalities

The *FLT3* gene encodes a member of the class III receptor tyrosine kinase family that is normally expressed on the surface of bone marrow (BM) hematopoietic progenitor cells and plays an important role in the survival and/or differentiation of multipotent stem cells. Since 1996, *FLT3*-ITD has been shown to be among the most prevalent genetic abnormalities in AML, being especially frequent in t(15;17) - positive and NK patients. Among the latter, *FLT3*-ITD has been detected in 28% to 33% of patients. In addition, 2 types of point mutations in *FLT3* have been reported: those localized within the activation loop of the tyrosine kinase domain (TKD), found in 5% to 14% of NK patients; and those localized in the JMD, whose incidence is 2% among all AML patients but has not yet been estab-

lished in NK AML. The *FLT3*-ITD encodes an abnormal protein that undergoes ligand-independent receptor dimerization, autophosphorylation, and constitutive activation, which turns on downstream signaling pathways involved in cell proliferation, differentiation, and survival. Similarly, both *FLT3*-TKD and JMD mutations bestow constitutive phosphorylation and activation of the FLT3 receptor, although the autophosphorylation of FLT3 and consequently transforming potential conferred by JMD mutations are weaker than those of *FLT3*-ITD and *FLT3*-TKD mutations.

Clinically, *FLT3*-ITD-positive patients present with increased white blood cell counts (WBCs) and percentages of blood and BM blasts and are more often diagnosed with de novo than secondary AML. Notably, an inferior prognosis for both remission duration and survival is conferred by *FLT3*-ITD in NK AML as well as in cases carrying “intermediate-risk” chromosome abnormalities. Conversely, the *FLT3*-TKD has not been reported to have an adverse prognosis among either NK patients or those with intermediate-risk cytogenetics. In addition to being a prognostic marker, *FLT3*-ITD is a potential therapeutic target. Several small-molecule inhibitors of FLT3 tyrosine kinase activity are currently in clinical trials.

Nucleophosmin member 1 (NPM1) gene abnormalities

Mutations of *NPM1*, are the most frequent gene mutations in NK AML patients, being reported in 46% to 62% of cases. These mutations are always heterozygous and result in both loss of tryptophan residues 290 and 288 (or 290 only), critical for nucleolar protein localization, and acquisition of an additional nuclear export signal at the C-terminus of nucleophosmin, causing its aberrant cytoplasmic localization. Altered distribution in cell compartments likely interferes with the normal functions of nucleophosmin, which include preventing nucleolar protein aggregation, regulation of ribosomal protein assembly and the nucleocytoplasmic transport of these proteins, as well as the regulation of the ARF and p53 tumor-suppressor pathways.

The prognostic relevance of *NPM1* mutations has been shown by most, but not all, recent studies that analyzed NK patients or those classified in the intermediate-risk cytogenetic category. Falini et al first reported that cytoplasmic localization of nucleophosmin, which is highly correlated with the presence of *NPM1* mutation, was an independent favorable prognostic factor for achievement of CR. In subsequent studies, comparisons of the clinical outcome involving patients with and without *NPM1* mutations regardless of the presence of other genetic rearrangements have yielded somewhat inconsistent results. However, when *FLT3*-ITDs, which are detected in approximately 40% of patients with *NPM1* mutations, have also been considered, all studies have found that *NPM1* mutations are associated with a significantly improved outcome in the absence of *FLT3*-ITD. Patients with *NPM1* mutations lacking *FLT3*-ITD had significantly better CR rates, EFS, RFS, DFS, and OS, whereas *NPM1* mutations did not impact the poor outcome of patients harboring *FLT3*-ITD. In multivariable analy-

ses, the combined status of mutated *NPM1* and the absence of *FLT3*-ITD constituted an independent favorable prognostic factor for the achievement of CR, DFS, RFS, EFS, and OS.

Other prognostic factors in karyotypically normal AML

Additional molecular abnormalities associated with peculiar prognosis in NK AML are:

- 1) partial tandem duplication (PTD) of the myeloid/lymphoid or mixed-lineage leukemia (*MLL*) gene;
 - 2) over-expression of *BAALC* gene;
 - 3) mutations of the CCAAT/enhancer-binding protein (*CEBPA*) gene.
- *MLL*-PTD occurs in approximately 8% (range, 5%-11%). Pretreatment characteristics do not differ significantly between NK patients with and without *MLL*-PTD. However, different studies showed that CR duration, RFS and EFS were significantly shorter for *MLL*-PTD-positive patients. Interestingly, Whitman et al have recently shown that the combination of decitabine, a DNA methyltransferase inhibitor, and depsipeptide, a histone deacetylase inhibitor, can reactivate the transcription of the *MLL* wild-type allele in *MLL*-PTD-positive cells leading to enhanced apoptosis.
 - *BAALC* over-expression at diagnosis seems to negatively affect EFS, DFS, and OS and appears to be particularly useful as a prognostic marker in NK AML patients lacking *FLT3*-ITD and *CEBPA* mutations.
 - The initial clinical studies revealed that *CEBPA* mutations were associated with a significantly better EFS, DFS, and OS among patients included in the cytogenetic intermediate-risk category. This favorable prognostic significance of *CEBPA* mutations was subsequently confirmed by investigations restricted to NK patients, 15% to 19% of whom carry *CEBPA* mutations.

Gene expression profiling (GEP)

GEP was recently used to ascertain patterns of gene activation and silencing (ie, gene-expression signatures) in NK AML with the goal of identifying patient subsets with diverse prognoses. Bullinger et al found that the majority of NK patients segregated into 2 gene-expression clusters with significantly different survival. Patients in cluster I, who had worse survival, more frequently harbored *FLT3* aberrations and were more commonly diagnosed with AML FAB M1 and M2. More recently, Radmacher et al validated this signature and created a class-prediction algorithm that allows for the prediction of outcome for subsequent individual patients. However, a strong association of the outcome classifier with the *FLT3*-ITD was observed that might explain the prognostic significance of the signature. However, these required confirmation, especially because in other studies NK patients segregated into several distinct clusters. Moreover, the prediction accuracy of the classifier for dichotomized outcome classes was modest, with OS and DFS of approximately 60% of the patients being correctly predicted.

In conclusion, it is now clear that NK AML is very heterogeneous at the molec-

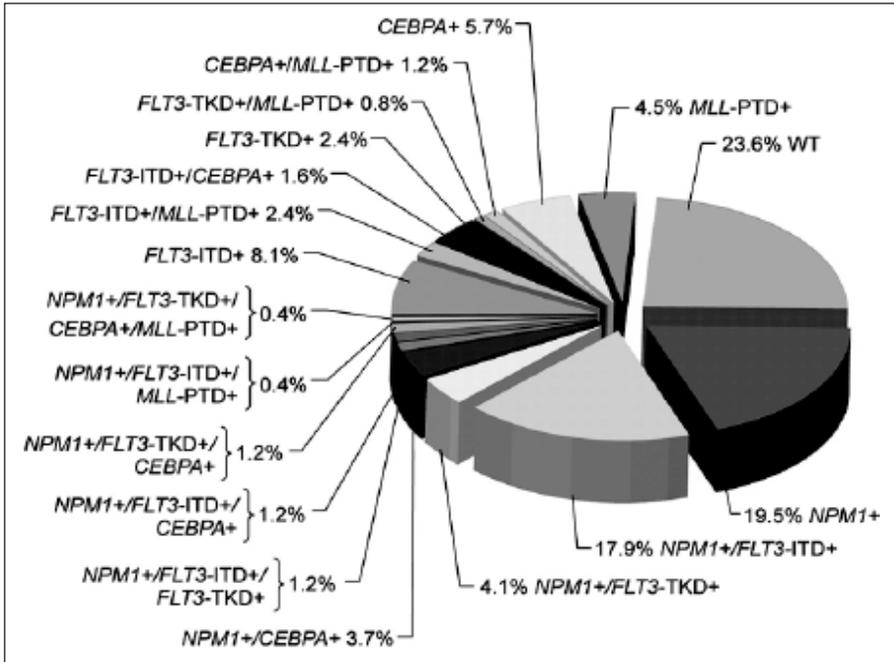


Fig. 1 - AML molecular subtypes. Adapted from Mrózek et al. Blood 2007, Vol. 109, No. 2, pp. 431-448.

ular level. We believe that upcoming modification to the World Health Organization (WHO) classification of AML should consider the above discussed recurrent molecular genetic rearrangements. Studies that would test all currently known molecular alterations in order to determine the relative impact

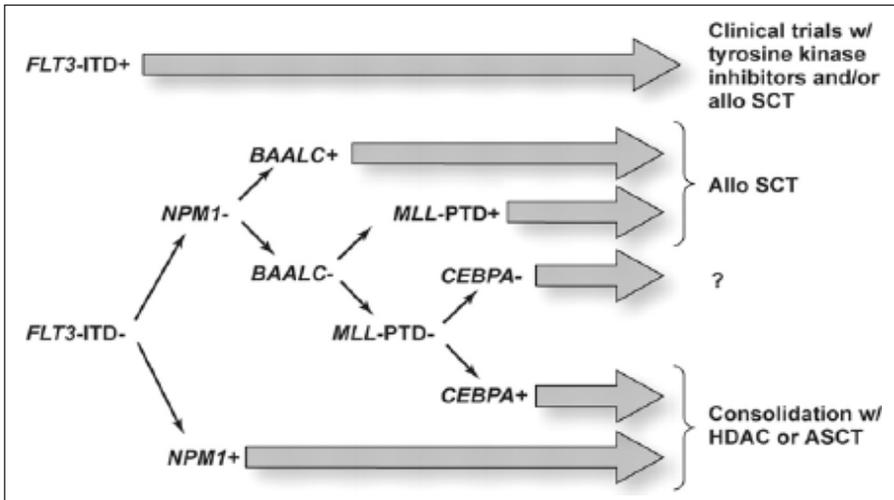


Fig. 2 - Proposal for AML post-remission therapy assignment according to molecular features. Adapted from Mrózek et al. Blood 2007, Vol. 109, No. 2, pp. 431-448.

of each of them are underway. Finally, while the recent results of gene-expression profiling are promising, it remains to be determined whether microarray technology will become more useful clinically and thus be able to replace testing for several individual genetic alterations with prognostic significance in NK AML patients.

Acknowledgment:

This work was supported in part by Centro Interdipartimentale per la Ricerca sul Cancro "G. Prodi", AIRC, BolognAIL, PRIN and RFO (Proff. M. Bacarani and SA. Pileri), Fondazione del Monte di Bologna e Ravenna grants, and LeukemiaNet.

References

1. Mrózek K, Marcucci G, Paschka P et al. Clinical relevance of mutations and gene-expression changes in adult acute myeloid leukemia with normal cytogenetics: are we ready for a prognostically prioritized molecular classification? *Blood* 2007, Vol. 109, No. 2: 431-448.
2. Nakao M, Yokota S, Iwai T, et al. Internal tandem duplication of the *flt3* gene found in acute myeloid leukemia. *Leukemia* 1996; 10: 1911-1918.
3. Fröhling S, Schlenk RF, Breitnick J, et al. Prognostic significance of activating *FLT3* mutations in younger adults (16 to 60 years) with acute myeloid leukemia and normal cytogenetics: a study of the AML Study Group Ulm. *Blood* 2002; 100: 4372-4380.
4. Falini B, Mecucci C, Tiacci E, et al. Cytoplasmic nucleophosmin in acute myelogenous leukemia with a normal karyotype [erratum appears in *N Engl J Med*. 2005; 352: 740]. *N Engl J Med* 2005; 352: 254-266.
5. Falini B, Nicoletti I, Martelli MF, Mecucci C. Acute myeloid leukemia carrying cytoplasmic/mutated nucleophosmin (NPMc+ AML): biologic and clinical features. *Blood*. 2007; 1, 109(3): 874-85.
6. Döhner K, Schlenk RF, Habdank M, et al. Mutant nucleophosmin (*NPM1*) predicts favorable prognosis in younger adults with acute myeloid leukemia and normal cytogenetics: interaction with other gene mutations. *Blood* 2005; 106: 3740-3746.
7. Caligiuri MA, Schichman SA, Strout MP, et al. Molecular rearrangement of the *ALL-1* gene in acute myeloid leukemia without cytogenetic evidence of 11q23 chromosomal translocations. *Cancer Res* 1994; 54: 370-373.
8. Baldus CD, Thiede C, Soucek S, Bloomfield CD, Thiel E, Ehninger G. *BAALC* expression and *FLT3* internal tandem duplication mutations in acute myeloid leukemia patients with normal cytogenetics: prognostic implications. *J Clin Oncol* 2006; 24: 790-797.
9. Fröhling S, Schlenk RF, Stolze I, et al. *CEBPA* mutations in younger adults with acute myeloid leukemia and normal cytogenetics: prognostic relevance and analysis of cooperating mutations. *J Clin Oncol* 2004; 22: 624-633.
10. Bullinger L, Döhner K, Bair E, et al. Use of gene-expression profiling to

- identify prognostic subclasses in adult acute myeloid leukemia. *N Engl J Med* 2004; 350: 1605-1616.
11. Radmacher MD, Marcucci G, Ruppert AS, et al. Independent confirmation of a prognostic gene-expression signature in adult acute myeloid leukemia with a normal karyotype: a Cancer and Leukemia Group B study. *Blood* 2006; 108: 1677-1683.
 12. Valk PJ, Verhaak RG, Beijen MA, Erpelinck CA, Barjesteh van Waalwijk van Doorn-Khosrovani S, Boer JM, Beverloo HB, Moorhouse MJ, van der Spek PJ, Lowenberg B, Delwel R. Prognostically useful gene-expression profiles in acute myeloid leukemia. *N Engl J Med*. 2004; 15, 350(16): 1617-28.

Rilevanza dell'espressione di WT1 nelle sindromi mielodisplastiche e nelle LAM

Daniela Cilloni

Dipartimento di Scienze Cliniche e Biologiche, Università di Torino

Il gene del tumore di Wilms (WT1), isolato la prima volta nell'omonima neoplasia renale pediatrica, codifica per un fattore di trascrizione zinc finger coinvolto nella regolazione della proliferazione, differenziazione e apoptosi cellulare. Sebbene l'esatto ruolo di WT1 non sia completamente chiaro, è noto che inibisce la trascrizione di diversi fattori di crescita tra cui il PDGF, M-CSF, IGF-II, di recettori di fattori di crescita come IGF-IR e EGFR e di altri geni tra cui RAR- α , c-myc, c-myb, N-myc e bcl-2, ma può anche attivare la trascrizione dello stesso bcl-2 e di altri geni come Bax-1. WT1 è espresso durante il periodo dell'organogenesi in cui svolge un ruolo cruciale. Topi knock-out per il gene WT1 muoiono a livello embrionale, probabilmente a causa di uno scompenso cardiaco e presentano alterazioni a livello dell'apparato urogenitale.

Inizialmente a questo gene è stato dato un ruolo oncosoppressore in quanto sono state trovate delezioni o mutazioni nel tumore di Wilms e mutazioni nelle linee germinali in pazienti che presentavano una predisposizione allo sviluppo di leucemie. Al contrario di altri oncosoppressori come Rb e p53 che sono espressi in tutti i tessuti, l'espressione del gene WT1, oltre ad avere valori molto bassi, è limitata solo alle cellule delle gonadi, dell'utero, dei reni, della milza, dei mesoteli e dei progenitori ematopoietici. Recentemente è stato però proposto per la forma wild-type del gene WT1 un ruolo pro-oncogenico in quanto si sono accumulate evidenze che vi sono elevati livelli di espressione in varie neoplasie maligne ematologiche e tumori solidi. Il gene WT1 è iperespresso nella leucemia mieloide e linfoblastica acuta, nelle malattie mieloproliferative croniche e nelle sindromi mielodisplastiche. Infine l'abnorme espressione di WT1 è caratteristica anche di numerosi tumori solidi tra cui il tumore del polmone, mammella, tiroide, stomaco e molti altri. Inoltre l'espressione di WT1 nelle leucemie è altamente correlato con la quantità di malattia. Si è infatti dimostrato un ottimo indicatore per monitorizzare la malattia residua dopo trattamento chemioterapico nei pazienti con leucemia acuta. In particolare, nei pazienti affetti da LAM, la persistenza di elevati livelli di espressione post chemioterapia di induzione è indice di

elevato rischio di recidiva precoce. Al contrario, i pazienti che ottengono la normalizzazione di WT1 dopo chemioterapia necessitano di frequenti valutazioni poiché, un aumento di WT1 precede sempre la ricaduta ematologia, in alcuni casi anche di diversi mesi. Essendo WT1 iperespresso anche nel sangue periferico, il monitoraggio della malattia risulta egualmente sensibile su PB facilitando pertanto il monitoraggio molecolare.

Nei pazienti affetti da sindromi mielodisplastiche i livelli di espressione di WT1 alla diagnosi correlano con il rischio di progressione in leucemia, ovvero con l'IPSS. WT1 possiede pertanto un significato fortemente prognostico e può essere considerato una valida alternativa all'IPSS soprattutto nei casi in cui non risulta possibile ottenere tutti i parametri necessari alla valutazione di tale indice. Inoltre, i molti casi di mielodisplasia con bassa blastosi, WT1 risulta di grande utilità anche in fase di diagnosi, soprattutto quando la diagnosi differenziale con forme secondarie risultasse difficoltosa. WT1 infatti è altamente specifico e non si innalza oltre i livelli soglia nei casi di anemie carenziali, emolitiche ecc. La presenza di livelli anomali di WT1 si associa sempre a un'emoipoiesi clonale.

Nonostante il significato di WT1 in termini di marcatore molecolare di emopoiesi clonale e di malattia minima residua, il significato dell'overespressione di questo gene rimane pressoché oscuro. Quello che ormai è chiaro che WT1 nel contesto dell'emoipoiesi leucemica svolge un ruolo di oncogene. In particolare esistono 4 isoforme che svolgono un ruolo pressoché diverso. In particolare due isoforme derivate da uno splicing alternativo dell'esone 9 dette KTS+ e KTS-, posseggono, proprio grazie alla presenza o assenza di 3 aminoacidi che si inseriscono tra il terzo e quarto zinc finger, proprietà funzionali diverse. Infatti, solo l'isoforma KTS- è in grado di entrare nel nucleo e di attivare la trascrizione genica. Al contrario l'isoforma KTS+ risulta prevalentemente citoplasmatica. Nelle cellule emopoietiche normali queste due isoforme distribuite in modo pressoché uguale con un rapporto fisiologico variabile tra 1 e 1,5. Dallo studio sull'espressione delle due isoforme di WT1 in campioni di sangue midollare prelevato da pazienti leucemici all'esordio di malattia (AML e CML), risulta evidente come ci sia uno sbilanciamento delle due isoforme a favore della forma KTS+. Il rapporto tra le isoforme ritorna a valori fisiologici dopo terapia ovvero al raggiungimento della remissione di malattia. Pertanto l'assenza di funzione trascrizionale di WT1 potrebbe rappresentare un meccanismo di leucemogenesi.

Infine, poiché WT1 sembrerebbe quindi avere un ruolo essenziale nella tumorigenesi e nel mantenimento del fenotipo neoplastico si è cominciato a valutare il possibile ruolo di WT1 come antigene contro cui l'organismo potrebbe indirizzare il proprio sistema immunitario nel tentativo di eliminare le cellule tumorali. Attualmente vi sono chiare evidenze cliniche dell'efficacia della risposta immunitaria antitumorale indotta da vaccini peptidici. In studi preclinici l'inoculo di questi peptidi nel topo determina la formazione di linfociti T citotossici CD8+ specifici verso l'antigene WT1 che in vitro sono in grado di uccidere selettivamente cellule tumorali che iperesprimono WT1. In vivo i topi immunizzati con questi peptidi rigettano l'inoculazione di cellule tumorali iperesprimenti WT1 e

sopravvivono per lungo tempo. Inoltre nessun topo vaccinato con questi peptidi mostra segni di tossicità d'organo da parte dei linfociti T citotossici CD8+ specifici verso l'antigene WT1. Nell'uomo è stata osservata la formazione spontanea di linfociti T citotossici CD8+ in pazienti affetti da leucemia mieloide acuta e sindrome mielodisplastica, ma anche in pazienti con tumore del polmone e della mammella. WT1 sembrerebbe quindi essere una proteina altamente immunogena e un promettente antigene tumore specifico. Sono stati quindi individuati vari peptidi in grado di evocare una risposta immunitaria citotossica nell'uomo verso l'antigene WT1. Sulla base dei risultati preclinici sono iniziati i primi studi clinici di fase I e II su pazienti affetti da leucemia acuta, sindrome mielodisplastiche, tumore del polmone e della mammella. I risultati preliminari mostrano che la vaccinazione è molto sicura e priva di tossicità e in grado di indurre risposte cliniche in termine di riduzione della quota di blasti nei pazienti leucemici.

Ricombinazione mitotica e fenotipo: la mutazione JAK2 nelle malattie mieloproliferative croniche Ph-negative

*Davide Rossi, Daniela Capello, Paola Riccomagno, Monia Lunghi
e Gianluca Gaidano*

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

Le malattie mieloproliferative croniche Philadelphia negative (Ph- CMPD) comprendono la policitemia vera (PV), la trombocitemia essenziale (ET) e la mielofibrosi idiopatica (IMF). La patogenesi delle Ph- CMPD ha come minimo comune denominatore la attivazione della via di traduzione del segnale JAK-STAT. A fronte di questo meccanismo patogenetico unificante, le Ph- CMPD sono caratterizzate da un fenotipo eterogeneo.

I meccanismi che sottendono la eterogeneità fenotipica delle Ph- CMPD sono noti solo in parte e verosimilmente riflettono differenze nei percorsi genetici e biologici coinvolti nella attivazione della via JAK-STAT.

Struttura e funzione del gene JAK2

JAK2 è una tirosin-chinasi citoplasmatica composta da tre domini funzionali (1). La regione carbossi-terminale di JAK2 contiene il dominio catalitico tirosin-chinasico (JH1). Adiacente al dominio JH1 è il dominio pseudo-chinasico (JH2), che possiede una funzione regolatoria (1).

Poiché la distruzione del dominio JH2 di JAK2 causa un incremento *in vitro* della attività tirosin-chinasica di JAK2, è verosimile che il dominio JH2 inibisca l'attivazione del dominio catalitico JH1 quando JAK2 è privo di ligando (1).

La regione amino-terminale di JAK2 contiene il dominio che media il legame di JAK2 alla regione intracitoplasmatica dei recettori per le citochine (1). JAK2 è coinvolto nella via di trasduzione del segnale dai recettori per eritropoietina (EPOR), trombopoietina (TPOR), G-CSF (G-CSFR) e GM-CSF (GM-CSFR) (1). L'interazione ligando-recettore induce oligomerizzazione del recettore e attivazione tramite transfosforilazione di JAK2.

La forma attiva di JAK2 fosforila una cascata di trasduttori del segnale tra cui STAT5, ERK e AKT. Ne consegue l'attivazione delle vie STAT, Ras/RAF/MEK/MAPK e PIK3/Akt/mTOR (1).

JAK2 è indispensabile per una eritropoiesi quantitativamente e qualitativamente normale.

Infatti, nei modelli murini JAK2^{-/-}, l'assenza di JAK2 risulta letale per l'embrione in quanto determina un difetto severo della eritropoiesi simile a quello indotto nei topi EPO^{-/-} o EPOR^{-/-} (1).

La mutazione JAK2V617F

JAK2 è bersaglio della mutazione JAK2V617F nel 70-90% dei casi di PV, nel 25-70% dei casi di ET e nel 35-60% dei casi di IMF (2-7). La mutazione JAK2V617F inoltre ricorre con frequenza variabile nelle patologie con fenotipo simile alle Ph- CMPD. Infatti, la mutazione JAK2V617F è presente nel 50% dei casi di mielodisplasia con piastrinosi e nel 50% dei casi di trombosi mesenterica, portale e delle vene sovraepatiche (8, 9). Al contrario, la mutazione JAK2V617F è assente nella eritrocitosi idiomatica (10).

La mutazione JAK2V617F causa la sostituzione del residuo di valina in posizione 617 con un residuo di fenilalanina.

La valina 617 si trova nel dominio pseudo-chinasico JH2. La sostituzione della valina con fenilalanina determina la perdita della attività auto-inibitoria esercitata dal dominio pseudo-chinasico JH2 sul dominio chinasico JH1. Pertanto, JAK2V617F si comporta come una tirosin-chinasi costitutivamente attivata, la cui funzione è indipendente dalla interazione ligando-recettore. JAK2V617F agisce da attivatore endogeno della via di JAK/STAT.

Evidenze derivate da modelli *in vitro* e *in vivo* hanno definito la capacità trasformante di JAK2V617F. JAK2V617F induce fosforilazione di STAT5 e, *in vitro*, causa proliferazione cellulare indipendente dalla stimolazione del recettore da parte di EPO, TPO e G-CSF (2-7).

Modelli murini transgenici per JAK2V617F sviluppano un fenotipo che mima la policitemia vera e la sua evoluzione in mielofibrosi post-policitemica. Infatti, JAK2V617F induce un fenotipo policitemico caratterizzato da eritrocitosi, reticolocitosi, leucocitosi, soppressione di EPO sierica, proliferazione di colonie eritroidi EPO-indipendenti, splenomegalia, panmielosi midollare e mielo-eritropoiesi extramidollare. Nel tempo, il fenotipo policitemico regredisce, le dimensioni della milza aumentano e compaiono fibrosi midollare, anemia, piastrinopenia e neutrofilia (11-13).

Correlazioni clinico-molecolari della mutazione JAK2V617F

I meccanismi che sottendono alla eterogeneità fenotipica delle Ph- CMPD JAK2V617F- positive comprendono:

- 1) il *mutation burden* di JAK2V617F;
- 2) il livello e il tipo di differenziazione del progenitore colpito dalla mutazione JAK2V617F;
- 3) la preesistenza, coesistenza o acquisizione di ulteriori lesioni genetiche/epigenetiche.

Mutation burden di JAK2V617F

La omozigosi per JAK2V617F risulta dalla ricombinazione mitotica della banda cromosomica 9p, sito in cui mappa il locus JAK2. La ricombinazione mitotica di 9p causa la perdita dell'allele wild type di JAK2 e la duplicazione dell'allele JAK2V617F mutato (2). L'impatto della omozigosi per JAK2V617F sulla biologia della cellula di Ph- CMPD è dimostrato da studi di espressione genica. Infatti, la omozigosi per JAK2V617F determina una alterazione del profilo di espressione genica più marcato rispetto alla eterozigosi per JAK2V617F, sia per un effetto di gene-dosage, sia per l'effetto dominante negativo esercitato dall'allele JAK2 wild type sull'allele JAK2V617F (14).

Numerose osservazioni suggeriscono che la omozigosi per JAK2V617F, e quindi il *gene-dosage*, possa promuovere il fenotipo delle Ph- CMPD. La omozigosi per JAK2V617F è un evento frequente nella PV, mentre è virtualmente assente nella ET. I granulociti del sangue periferico risultano omozigoti per JAK2V617F nel 25-33% dei casi di PV e nel 0-3% dei casi di ET (2-7). Inoltre, virtualmente tutti i casi di PV eterozigoti per JAK2V617F su granulociti del sangue periferico hanno progenitori emopoietici omozigoti per JAK2V617F fin dalla diagnosi, e la proporzione di granulociti omozigoti per JAK2V617F nel sangue periferico aumenta con l'aumentare del tempo di osservazione (15-17). Al contrario, virtualmente tutti i casi di ET eterozigoti per JAK2V617F su granulociti del sangue periferico hanno progenitori emopoietici eterozigoti per JAK2V617F e la acquisizione di omozigosi per JAK2V617F durante il tempo di osservazione è un evento raro (15-17).

Il fenotipo PV, quindi, si associa alla omozigosi per JAK2V617F, mentre il fenotipo ET si associa alla eterozigosi per JAK2V617F. Poiché la predisposizione alla omozigosi per JAK2V617F è conseguenza di una predisposizione al fenomeno di ricombinazione mitotica del cromosoma 9p, i meccanismi genetici o ambientali che predispongono alla ricombinazione mitotica del cromosoma 9p possono influenzare lo sviluppo di un fenotipo PV o di un fenotipo ET in un determinato clone di Ph- CMPD JAK2V617F-positiva.

L'effetto del *gene-dosage* sul fenotipo è suggerito da studi di correlazione clinico-molecolare nella PV e nella ET. Rispetto ai pazienti con PV eterozigote per JAK2V617F, i pazienti con PV omozigote per JAK2V617F hanno un fenotipo clinico più aggressivo, caratterizzato da conta leucocitaria ed ematocrito più elevati, conta piastrinica più bassa, maggiore frequenza di splenomegalia, maggiore rischio di eventi vascolari, e maggiore necessità di chemioterapia per il controllo della malattia (18, 19). Al contrario, rispetto ai pazienti con PV omozigote per JAK2V617F, i pazienti con PV eterozigote per JAK2V617F hanno un fenotipo clinico più simile a quello della ET, caratterizzato da leucociti ed ematocrito meno elevati, conta piastrinica più alta, minore prevalenza di splenomegalia e minore rischio di eventi vascolari (18, 19).

I pazienti con ET eterozigote per JAK2V617F, rispetto ai pazienti con ET wild type per JAK2, hanno un fenotipo clinico più simile a quello della PV, caratterizzato da valori più elevati di emoglobina e leucociti, una conta piastrinica più bassa, maggior prevalenza di pannellosi midollare, valori di EPO più bassi, mag-

giore rischio di eventi vascolari e maggiore rischio di trasformazione in PV e in mielofibrosi (20-24). Inoltre, nei pazienti affetti da ET, il passaggio da eterozigosi a omozigosi per JAK2V617F si associa alla acquisizione di un fenotipo policitemico (21).

Il ruolo della mutazione JAK2V617F nel determinare il fenotipo della IMF è meno rilevante. La mutazione JAK2V617F ricorre nel 30-60% dei casi di IMF, con una prevalenza di omozigosi pari a 10-30% (2-7). I casi di IMF con mutazione JAK2V617F hanno una minore necessità trasfusionale, a sottolineare l'impatto della spinta eritropoietica indotta dalla mutazione JAK2V617F. Tuttavia, poiché nessuna delle altre caratteristiche cliniche della IMF correla con la presenza della mutazione JAK2V617F, è verosimile che altre lesioni genetiche oggi non note contribuiscano in maniera più sostanziale al fenotipo clinico (25, 26). Tale ipotesi è sostenuta anche del frequente riscontro di anomalie citogenetiche nei casi di IMF con e senza mutazione JAK2V617F (27).

JAK2V617F e progenitori emopoietici

Il grado di differenziazione del progenitore emopoietico al momento della acquisizione della mutazione JAK2V617F potrebbe influenzare il fenotipo delle Ph-CMPD. Le evidenze sperimentali suggeriscono che:

- 1) nella PV, tutte le linee emopoietiche (cellule staminali emopoietiche, progenitori mieloidi comuni, progenitori granulocitario/macrofagici e progenitori eritroidi/megacariocitari) portano la mutazione JAK2V617F (28, 29);
- 2) nella PV e nella IMF, la mutazione JAK2V617F colpisce una cellula staminale emopoietica con potenzialità differenziative mieloidi, eritroidi, megacariocitarie e linfoidi (30).

Lesioni molecolari addizionali

Molteplici evidenze suggeriscono un ruolo di lesioni molecolari addizionali nella patogenesi delle Ph-CMPD:

- 1) lesioni citogenetiche addizionali ricorrono nel 5-10% dei casi di Ph-CMPD;
- 2) una frazione consistente di ET e IMF presenta una emopoiesi clonale in assenza della mutazione JAK2V617F (31, 32);
- 3) in pazienti affetti da ET, e in misura minore PV, la percentuale di granulociti mutati per JAK2V617F è inferiore alla percentuale di granulociti clonali (31).

Tale osservazione, anche se controversa, sembra suggerire che il clone originale della Ph-CMPD pre-esista all'acquisizione della mutazione JAK2V617F. Inoltre, nei casi di Ph-CMPD familiari, la predisposizione alla malattia è indipendente dalla mutazione JAK2V617F (33). Infine, alcuni casi di leucemia acuta secondaria a Ph-CMPD perdono la mutazione JAK2V617F e possono acquisire un cariotipo differente e meno complesso rispetto a quello della fase cronica (35).

Tali osservazione suggeriscono che le leucemie acute secondarie a Ph-CMPD, almeno in una frazione di casi, originano da un sub-clone JAK2V617F negativo. Alcune delle lesioni molecolari addizionali sono state identificate e comprendo-

no le mutazioni dell'esone 12 di JAK2, la mutazione MPLW515L, e la inattivazione tramite metilazione dei geni SHP1, SOCS1 e SOCS3 (36-39).

Le mutazioni nell'esone 12 di JAK2 sono state identificate in 6 casi di PV JAK2V617F-negativa e in 4 casi di eritrocitosi idiopatica JAK2V617F-negativa, ma non nei casi di ET o IMF JAK2V617F-negativi (36). In modelli sperimentali *in vitro* e *in vivo*, le mutazioni dell'esone 12 di JAK2 dimostrano una capacità trasformante simile a quella di JAK2V617F (36). Infatti, le mutazioni dell'esone 12 di JAK2 inducono fosforilazione di STAT5 e causano proliferazione cellulare indipendente dalla stimolazione del recettore da parte di EPO, TPO e G-CSF (36). Inoltre, in modelli murini transgenici, le mutazioni dell'esone 12 di JAK2 inducono un fenotipo caratterizzato da eritrocitosi, reticolocitosi, leucocitosi e trombocitosi (36). Il fenotipo clinico dei pazienti affetti da mutazioni dell'esone 12 di JAK2 è peculiare e si caratterizza per eritrocitosi isolata e iperplasia eritroide midollare, in assenza di organomegalia, leucocitosi o trombocitosi (36). Rispetto ai pazienti affetti da PV con mutazione JAK2V617F, i pazienti con PV/eritrocitosi idiopatica e mutazione dell'esone 12 di JAK2 sono più giovani, hanno valori di ematocrito più elevati e una conta piastrinica e leucocitaria più bassa (36).

Il gene MPL codifica per il recettore per la trombopoietina. La mutazione MPLW515L colpisce il 5-10% dei casi di IMF JAK2V617F-negativa e l'1% dei casi di ET JAK2V617F-negativa (37, 38). In modelli sperimentali *in vitro* e *in vivo*, la mutazione MPLW515L dimostra capacità trasformante (38).³⁸ Infatti, la mutazione MPLW515L induce fosforilazione di STAT5 e causa proliferazione cellulare indipendente dalla stimolazione del recettore da parte di EPO, TPO e G-CSF. Inoltre, in modelli murini transgenici, la mutazione MPLW515L induce un fenotipo mielofibrotico caratterizzato da leucocitosi, trombocitosi, splenomegalia, fibrosi midollare, iperplasia megacariocitaria atipica, eritroblastosi ed emopoiesi extramidollare (38).

SHP1, SOCS1 e SOCS3 sono proteine citoplasmatiche coinvolte nel controllo e nella inibizione del segnale promosso da JAK2. SHP1 è indispensabile per il controllo della emopoiesi. Infatti, i topi SHP1^{-/-} sviluppano un fenotipo caratterizzato da iperplasia mieloide, splenomegalia ed emopoiesi extramidollare (1). Nelle Ph- CMPD, SHP1, SOCS1 e SOCS3 sono frequentemente inattivati tramite metilazione del DNA (39). Poiché la metilazione di SHP1, SOCS1 e SOCS3 ricorre con frequenza simile nella Ph- CMPD JAK2V617F-negative e nelle Ph- CMPD JAK2V617F-positive, è verosimile che la metilazione di SHP1, SOCS1 e SOCS3 sia un meccanismo complementare a JAK2V617F nella attivazione della via JAK-STAT (39).

Bibliografia

1. Khwaja A. The role of Janus kinases in haemopoiesis and haematological malignancy. *Br J Haematol.* 2006; 134: 366-84.
2. Kralovics R, Passamonti F, Buser AS, Teo SS, Tiedt R, Passweg JR, Tichelli A, Cazzola M, Skoda RC. A gain-of-function mutation of JAK2 in myeloproliferative disorders. *N Engl J Med.* 2005; 352: 1779-90.

3. Baxter EJ, Scott LM, Campbell PJ, East C, Fourouclas N, Swanton S, Vassiliou GS, Bench AJ, Boyd EM, Curtin N, Scott MA, Erber WN, Green AR; Cancer Genome Project. Acquired mutation of the tyrosine kinase JAK2 in human myeloproliferative disorders. *Lancet*. 2005; 365: 1054-61.
4. Jones AV, Kreil S, Zoi K, Waghorn K, Curtis C, Zhang L, Score J, Seear R, Chase AJ, Grand FH, White H, Zoi C, Loukopoulos D, Terpos E, Vervessou EC, Schultheis B, Emig M, Ernst T, Lengfelder E, Hehlmann R, Hochhaus A, Oscier D, Silver RT, Reiter A, Cross NC. Widespread occurrence of the JAK2 V617F mutation in chronic myeloproliferative disorders. *Blood*. 2005; 106: 2162-8.
5. Levine RL, Wadleigh M, Cools J, Ebert BL, Wernig G, Huntly BJ, Boggon TJ, Wlodarska I, Clark JJ, Moore S, Adelsperger J, Koo S, Lee JC, Gabriel S, Mercher T, D'Andrea A, Frohling S, Dohner K, Marynen P, Vandenberghe P, Mesa RA, Tefferi A, Griffin JD, Eck MJ, Sellers WR, Meyerson M, Golub TR, Lee SJ, Gilliland DG. Activating mutation in the tyrosine kinase JAK2 in polycythemia vera, essential thrombocythemia, and myeloid metaplasia with myelofibrosis. *Cancer Cell*. 2005; 7: 387-97.
6. Zhao R, Xing S, Li Z, Fu X, Li Q, Krantz SB, Zhao ZJ. Identification of an acquired JAK2 mutation in polycythemia vera. *J Biol Chem*. 2005; 280: 22788-92.
7. James C, Ugo V, Le Couedic JP, Staerk J, Delhommeau F, Lacout C, Garcon L, Raslova H, Berger R, Bennaceur-Griscelli A, Villeval JL, Constantinescu SN, Casadevall N, Vainchenker W. A unique clonal JAK2 mutation leading to constitutive signalling causes polycythaemia vera. *Nature*. 2005; 434: 1144-8.
8. Patel RK, Lea NC, Heneghan MA, Westwood NB, Milojkovic D, Thanigaikumar M, Yallop D, Arya R, Pagliuca A, Gaken J, Wendon J, Heaton ND, Mufti GJ. Prevalence of the activating JAK2 tyrosine kinase mutation V617F in the Budd-Chiari syndrome. *Gastroenterology*. 2006; 130: 2031-8.
9. Szpurka H, Tiu R, Murugesan G, Aboudola S, Hsi ED, Theil KS, Sekeres MA, Maciejewski JP. Refractory anemia with ringed sideroblasts associated with marked thrombocytosis (RARS-T), another myeloproliferative condition characterized by JAK2 V617F mutation. *Blood*. 2006; 108: 2173-81.
10. Rossi D, Cortini F, Deambrogi C, Barbieri C, Cerri M, Franceschetti S, Conconi A, Capello D, Gaidano G. Usefulness of JAK2V617F mutation in distinguishing idiopathic erythrocytosis from polycythemia vera. *Leuk Res*. 2007; 31: 97-101.
11. Wernig G, Mercher T, Okabe R, Levine RL, Lee BH, Gilliland DG. Expression of Jak2V617F causes a polycythemia vera-like disease with associated myelofibrosis in a murine bone marrow transplant model. *Blood*. 2006; 107: 4274-81.
12. Zaleskas VM, Krause DS, Lazarides K, Patel N, Hu Y, Li S, Van Etten RA.. Molecular Pathogenesis and Therapy of Polycythemia Induced in Mice by JAK2 V617F. *PLoS ONE*. 2006; 1: e18.

13. Lacout C, Pisani DF, Tulliez M, Gachelin FM, Vainchenker W, Villeval JL. JAK2V617F expression in murine hematopoietic cells leads to MPD mimicking human PV with secondary myelofibrosis. *Blood*. 2006; 108: 1652-60.
14. Kralovics R, Teo SS, Buser AS, Brutsche M, Tiedt R, Tichelli A, Passamonti F, Pietra D, Cazzola M, Skoda RC. Altered gene expression in myeloproliferative disorders correlates with activation of signaling by the V617F mutation of Jak2. *Blood*. 2005; 106: 3374-6.
15. Scott LM, Scott MA, Campbell PJ, Green AR. Progenitors homozygous for the V617F mutation occur in most patients with polycythemia vera, but not essential thrombocythemia. *Blood*. 2006; 108: 2435-7.
16. Lippert E, Boissinot M, Kralovics R, Girodon F, Dobo I, Praloran V, Boiret-Dupre N, Skoda RC, Hermouet S. The JAK2-V617F mutation is frequently present at diagnosis in patients with essential thrombocythemia and polycythemia vera. *Blood*. 2006; 108: 1865-7.
17. Gale RE, Allen AJ, Nash MJ, Linch DC. Long-term serial analysis of X-chromosome inactivation patterns and JAK2 V617F mutant levels in patients with essential thrombocythemia show that minor mutant-positive clones can remain stable for many years. *Blood*. 2007; 109: 1241-3.
18. Vannucchi AM, Antonioli E, Guglielmelli P, Pancrazzi A, Bogani C, Pieri L, Bosi A. Influence of the JAK2V617F mutational load at diagnosis on major clinical aspects in patients with polycythemia vera. *Blood*. 2006; 108: 6a.
19. Tefferi A, Lasho TL, Schwager SM, Strand JS, Elliott M, Mesa R, Li CY, Wadleigh M, Lee SJ, Gilliland DG. The clinical phenotype of wild-type, heterozygous, and homozygous JAK2V617F in polycythemia vera. *Cancer*. 2006; 106: 631-5.
20. Finazzi G, Rambaldi A, Guerini V, Carobbo A, Barbui T. Risk of thrombosis in patients with essential thrombocythemia and polycythemia vera according to JAK2 V617F mutation status. *Haematologica*. 2007; 92: 135-6.
21. Wolanskyj AP, Lasho TL, Schwager SM, McClure RF, Wadleigh M, Lee SJ, Gilliland DG, Tefferi A.. JAK2 mutation in essential thrombocythaemia: clinical associations and long-term prognostic relevance. *Br J Haematol*. 2005; 131: 208-13.
22. Carobbio A, Finazzi G, Guerini V, Spinelli O, Delaini F, Marchioli R, Borrelli G, Rambaldi A, Barbui T. Leukocytosis is a risk factor for thrombosis in essential thrombocythemia: interaction with treatment, standard risk factors and Jak2 mutation status. *Blood*. 2006; DOI 10.1182/blood-2006-09-046342.
23. Cheung B, Radia D, Pantelidis P, Yadegarfar G, Harrison C. The presence of the JAK2 V617F mutation is associated with a higher haemoglobin and increased risk of thrombosis in essential thrombocythaemia. *Br J Haematol*. 2006; 132: 244-5.
24. Campbell PJ, Scott LM, Buck G, Wheatley K, East CL, Marsden JT, Duffy A, Boyd EM, Bench AJ, Scott MA, Vassiliou GS, Milligan DW, Smith SR, Erber WN, Bareford D, Wilkins BS, Reilly JT, Harrison CN, Green AR;

- United Kingdom Myeloproliferative Disorders Study Group; Medical Research Council Adult Leukaemia Working Party; Australasian Leukaemia and Lymphoma Group. Definition of subtypes of essential thrombocythaemia and relation to polycythaemia vera based on JAK2 V617F mutation status: a prospective study. *Lancet*. 2005; 366: 1945-53.
25. Tefferi A, Lasho TL, Schwager SM, Steensma DP, Mesa RA, Li CY, Wadleigh M, Gary Gilliland D. The JAK2(V617F) tyrosine kinase mutation in myelofibrosis with myeloid metaplasia: lineage specificity and clinical correlates. *Br J Haematol*. 2005; 131: 320-8.
 26. Campbell PJ, Griesshammer M, Dohner K, Dohner H, Kusec R, Hasselbalch HC, Larsen TS, Pallisgaard N, Giraudier S, Le Bousse-Kerdiles MC, Desterke C, Guerton B, Dupriez B, Bordessoule D, Fenaux P, Kiladjian JJ, Viallard JF, Briere J, Harrison CN, Green AR, Reilly JT. V617F mutation in JAK2 is associated with poorer survival in idiopathic myelofibrosis. *Blood*. 2006; 107: 2098-100.
 27. Tefferi A, Strand JJ, Lasho TL, Elliott MA, Li CY, Mesa RA, Dewald GW. Respective clustering of unfavorable and favorable cytogenetic clones in myelofibrosis with myeloid metaplasia with homozygosity for JAK2(V617F) and response to erythropoietin therapy. *Cancer*. 2006; 106: 1739-43.
 28. Ishii T, Bruno E, Hoffman R, Xu M. Involvement of various hematopoietic-cell lineages by the JAK2V617F mutation in polycythemia vera. *Blood*. 2006; 108: 3128-34.
 29. Jamieson CH, Gotlib J, Durocher JA, Chao MP, Mariappan MR, Lay M, Jones C, Zehnder JL, Lilleberg SL, Weissman IL. The JAK2 V617F mutation occurs in hematopoietic stem cells in polycythemia vera and predisposes toward erythroid differentiation. *Proc Natl Acad Sci USA*. 2006; 103: 6224-9.
 30. Delhommeau F, Dupont S, Tonetti C, Masse A, Godin I, Le Couedic JP, Debili N, Saulnier P, Casadevall N, Vainchenker W, Giraudier S. Evidence that the JAK2 G1849T (V617F) mutation occurs in a lymphomyeloid progenitor in polycythemia vera and idiopathic myelofibrosis. *Blood*. 2007; 109: 71-7.
 31. Levine RL, Belisle C, Wadleigh M, Zahrieh D, Lee S, Chagnon P, Gilliland DG, Busque L. X-inactivation-based clonality analysis and quantitative JAK2V617F assessment reveal a strong association between clonality and JAK2V617F in PV but not ET/MMM, and identifies a subset of JAK2V617F-negative ET and MMM patients with clonal hematopoiesis. *Blood*. 2006; 107: 4139-41.
 32. Kralovics R, Teo SS, Li S, Theodorides A, Buser AS, Tichelli A, Skoda RC. Acquisition of the V617F mutation of JAK2 is a late genetic event in a subset of patients with myeloproliferative disorders. *Blood*. 2006; 108: 1377-80.
 33. Bellanne-Chantelot C, Chaumarel I, Labopin M, Bellanger F, Barbu V, De Toma C, Delhommeau F, Casadevall N, Vainchenker W, Thomas G,

- Najman A.. Genetic and clinical implications of the Val617Phe JAK2 mutation in 72 families with myeloproliferative disorders. *Blood*. 2006; 108: 346-52.
34. Campbell PJ, Baxter EJ, Beer PA, Scott LM, Bench AJ, Huntly BJ, Erber WN, Kusec R, Larsen TS, Giraudier S, Le Bousse-Kerdiles MC, Griesshammer M, Reilly JT, Cheung BY, Harrison CN, Green AR. Mutation of JAK2 in the myeloproliferative disorders: timing, clonality studies, cytogenetic associations, and role in leukemic transformation. *Blood*. 2006; 108: 3548-55.
 35. Rossi D, Deambrogi C, Capello D, Cerri M, Lunghi M, Parvis G, Saglio G, Gaidano G, Cilloni D. JAK2 V617F mutation in leukaemic transformation of philadelphia-negative chronic myeloproliferative disorders. *Br J Haematol*. 2006; 135: 267-8.
 36. Scott LM, Tong W, Levine RL, Scott MA, Beer PA, Stratton MR, Futreal PA, Erber WN, McMullin MF, Harrison CN, Warren AJ, Gilliland DG, Lodish HF, Green AR. JAK2 exon 12 mutations in polycythemia vera and idiopathic erythrocytosis. *N Engl J Med*. 2007; 356: 459-68.
 37. Pardanani AD, Levine RL, Lasho T, Pikman Y, Mesa RA, Wadleigh M, Steensma DP, Elliott MA, Wolanskyj AP, Hogan WJ, McClure RF, Litzow MR, Gilliland DG, Tefferi A.. MPL515 mutations in myeloproliferative and other myeloid disorders: a study of 1182 patients. *Blood*. 2006; 108: 3472-6.
 38. Pikman Y, Lee BH, Mercher T, McDowell E, Ebert BL, Gozo M, Cuker A, Wernig G, Moore S, Galinsky I, DeAngelo DJ, Clark JJ, Lee SJ, Golub TR, Wadleigh M, Gilliland DG, Levine RL. MPLW515L is a novel somatic activating mutation in myelofibrosis with myeloid metaplasia. *PLoS Med*. 2006; 3: e270.
 39. Rossi D, Deambrogi C, Lischetti T, Cerri M, Spunton M, Rasi S, Franceschetti S, Conconi A, Riccomagno P, Lunghi M, Capello D, Gaidano G. Hypermethylation of the SH²-containing phosphatase-1 (SHP-1) and Suppressor of cytokine signaling-1 and -3 (SOCS-1 and -3) in Philadelphia negative chronic myeloproliferative disorders (Ph- CMPD). *Blood*. 2006; 108: 39a.

Espressione genica e classificazione dei linfomi

Davide Rossi, Silvia Franceschetti, Michaela Cerri, Silvia Rasi, Valeria Spina, Paola Riccomagno, Annarita Conconi, Daniela Capello e Gianluca Gaidano

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

Concetti generali

Il gene expression profiling (GEP; profilo di espressione genica) con la tecnica dei microarrays è emerso rapidamente come approccio di caratterizzazione molecolare delle neoplasie ematologiche, tra cui i linfomi. Le “signatures” (letteralmente, firme) molecolari derivate dall’analisi GEP e associate a specifiche categorie clinico-patologiche di linfoma rappresentano una opzione estremamente raffinata e selettiva per interpretare biologicamente l’alto grado di complessità clinica e morfologica dei linfomi. L’analisi GEP ha quindi offerto uno strumento nuovo per definire la variabilità nell’outcome clinico di pazienti con la stessa diagnosi, ed ha fornito elementi innovativi per la diagnosi, la diagnostica differenziale e la classificazione dei linfomi. Inoltre, la rivelazione di specifici percorsi molecolari associati a determinati linfomi può contribuire a identificare bersagli nuovi per strategie di “target therapy”. Malgrado queste importanti prospettive, l’analisi GEP applicata ai linfomi non è scevra di problematiche e richiede un alto livello tecnologico e bioinformatico non immediatamente applicabile alla pratica clinica quotidiana. Pertanto, attualmente, l’analisi GEP nel contesto dei linfomi rappresenta uno strumento di “discovery” di specifici marcatori biologici, la cui traslazione alla pratica clinica richiede il disegno di modalità operative più semplici e più diffuse a livello territoriale.

I dati oggi disponibili riguardo al GEP dei linfomi riguardano sia neoplasie delle cellule B periferiche sia, in misura minore, delle cellule T periferiche. Della grande mole di dati disponibili, solo una parte è di interesse clinico ai fini della classificazione, della diagnostica e della stratificazione prognostica dei linfomi, e una parte ancora minore ha trovato oggi applicazione clinica.

Neoplasie delle cellule B periferiche

Nell’ambito dei linfomi derivati da cellule B periferiche, l’analisi GEP è stata particolarmente informativa nel caso del linfoma diffuso a grandi cellule B.

Inoltre, dati potenzialmente utili ai fini clinici sono stati ottenuti anche per il linfoma follicolare, il linfoma mantellare, e il primary effusion lymphoma.

Linfoma diffuso a grandi cellule B

Traslocazioni cromosomiche coinvolgenti la banda 3q27 e svariati siti cromosomici, inclusi i siti delle immunoglobuline, rappresentano una alterazione frequente nei linfomi diffusi a grandi cellule B (diffuse large B-cell lymphoma, DLBCL). Il proto-oncogene coinvolto, definito *BCL-6*, appartiene alla famiglia dei fattori trascrizionali contenenti i domini “zinc finger” ed è espresso selettivamente dalle cellule B del centro germinativo. L’analisi molecolare di *BCL-6* ha definito che questo gene è coinvolto in circa il 40% dei linfomi diffusi a grandi cellule B, e rappresenta un marcatore specifico di questa patologia. Oltre al riarrangiamento cromosomico, le alterazioni molecolari di *BCL-6* includono anche le mutazioni in prossimità delle regioni regolatorie del gene, in corrispondenza della regione che va incontro a riarrangiamento.

Studi GEP su DLBCL de novo hanno identificato due gruppi prognostici distinti. Il primo gruppo, definito “germinal center B-cell-like” (GCB-DLBCL) riflette il GEP di cellule B del centro germinativo e presenta prognosi favorevole. Il secondo gruppo, definito “activated B-cell like” (ABC-DLBCL) riflette il GEP di cellule B attivate “in vitro” e di cellule post-centro germinativo e presenta prognosi sfavorevole. Un terzo gruppo, con prognosi simile ai ABC-DLBCL ma con GEP distinto, è stato identificato da alcuni studi. I GCB-DLBCL e i ABC-DLBCL presentano patogenesi molecolare differente: ad esempio, i riarrangiamenti di *BCL-2* sono praticamente esclusivi dei casi di GCB-DLBCL, mentre i casi di ABC-DLBCL sono associati ad attivazione di NF- κ B ed a mutazioni di *BLIMP-1*. Le differenze fra GCB-DLBCL e ABC-DLBCL sono anche rivelate da studi di array-comparative genomic hybridization (CGH), che identificano specifiche delezioni/amplificazioni nei due diversi sottogruppi della malattia.

Poiché l’analisi GEP dei DLBCL è verosimilmente importante ai fini prognostici, ma non è praticabile nella gestione clinica di routine, studi immunoistochimici hanno tentato di riprodurre l’impatto prognostico del GEP utilizzando marcatori “surrogati” derivati dall’analisi GEP. Questi studi hanno prevalentemente utilizzato *CD10* e *bcl-6* come marcatori dei GCB-DLBCL, e *MUM-1/IRF4* come marcatore di ABC-DLBCL. L’utilità della immunoistochimica ai fini della stratificazione prognostica dei DLBCL è comunque ancora dibattuta. Si deve notare che gli studi di stratificazione prognostica dei DLBCL sulla base del GEP o dei marcatori “surrogati” immunoistochimici si basano su casistiche retrospettive trattate in epoca pre-Rituximab.

Infine, l’analisi GEP ha documentato che, come già ipotizzato in base alla classificazione WHO, il linfoma a cellule B primitivo del mediastino è realmente una entità clinico-patologica distinta dal DLBCL.

Linfoma follicolare

I linfomi follicolari sono caratterizzati da attivazione di *BCL-2* nel 90% dei casi. A differenza della maggior parte dei proto-oncogeni, *BCL-2* non regola diretta-

mente la proliferazione cellulare, bensì È preposto al controllo della morte cellulare programmata (apoptosi).

L'analisi GEP dei linfomi follicolari ha condotto alla identificazione di due profili di espressione genica distinti, associati a diversa prognosi dei pazienti. Le due "signatures" GEP identificate nei linfomi follicolari sono definite "risposta immune 1" e "risposta immune 2", perché i geni coinvolti includono loci noti per essere espressi differenzialmente nelle cellule del sistema immunitario normale. Si deve notare che entrambi le "signatures" sono espresse in maniera altamente differenziale nella frazione CD19-negativa delle biopsie di linfoma follicolare, ovvero nella frazione composta da cellule T, monocito/macrofagi ed altre cellule del microambiente. La "signature" risposta immune 1 si associa a prognosi favorevole, mentre la "signature" risposta immune 2 si associa a prognosi sfavorevole. La "signature" risposta immune 1 comprende geni ristretti alle cellule T, tra cui CD7, CD8B, ITK e LEF1. Al contrario, la "signature" risposta immune 2 comprende geni iperespressi da macrofagi e cellule dendritiche, tra cui CD64, TLR5 e varie componenti del complemento. La forte sinergia statistica fra le due "signatures" probabilmente riflette l'abbondanza relativa di popolazioni cellulari immunologiche non-neoplastiche che infiltrano il tumore. È questo un esempio di come il GEP possa rivelare categorie prognostiche non solo sulla base delle caratteristiche del clone neoplastico, ma anche del microambiente tumorale.

Linfoma mantellare

La traslocazione t(11;14) associata al linfoma mantellare attiva il proto-oncogene ciclina D1 (anche noto come *BCL-1*), un fattore di controllo del ciclo cellulare. Poiché la traslocazione si associa selettivamente al linfoma mantellare nello spettro dei linfomi non-Hodgkin, essa rappresenta un utile marcatore diagnostico per la diagnosi differenziale di questo linfoma. L'analisi GEP ha rivelato più di 1000 geni espressi differenzialmente fra i linfomi mantellari ed altri tipi di linfoma. Questi studi hanno quindi permesso di identificare una "signature" propria del linfoma mantellare, indipendentemente dalla iperespressione di ciclina D1. Ne deriva che alcuni linfomi, pur ciclina D1 negativi, esprimono la "signature" propria dei linfomi mantellari; in alcuni di questi casi, sono coinvolte la ciclina D2 o la ciclina D3. Inoltre, l'analisi GEP ha permesso di identificare categorie prognostiche distinte di linfoma mantellare. Infine, l'analisi GEP permette il riconoscimento molecolare della variante blastoide del linfoma mantellare, caratterizzata da iperespressione di cyclin-dependent kinase (CDK) 4 e di CDC28 protein kinase 1. CDK4 si associa con ciclina D1 e favorisce la progressione del ciclo cellulare attraverso il checkpoint G1/S. L'iperespressione di CDC28 protein kinase 1 abroga l'inibizione del complesso ciclina D1/CDK4 da parte dell'inibitore CDK p27/Kip1. In base ai dati GEP, l'espressione di vari altri geni coinvolti nella regolazione del ciclo cellulare è alterata nel linfoma mantellare variante blastoide.

Primary effusion lymphoma

Il primary effusion lymphoma (PEL) rappresenta una rara neoplasia linfoide associata a infezione del clone neoplastico da parte di HHV-8. Il PEL tende a

manifestarsi con interessamento esclusivo o predominante delle cavità sierose in fase liquida, e predilige l'ospite immunodeficiente. L'analisi GEP di casi di PEL ha rivelato una stretta somiglianza al profilo delle plasmacellule. In particolare, il GEP del PEL è ben distinto dal GEP di tutti gli altri linfomi aggressivi, sia della popolazione immunocompetente sia immunodeficiente. Il GEP del PEL è tipicamente plasmablastico, con aspetti immunoblastici (CD30, CD38, MUM1/IRF-4) e plasmacellulari (BLIMP1, CD138).

Altri marcatori GEP associati a PEL e potenzialmente utili ai fini di diagnosi differenziale o terapia mirata includono: aquaporina-1, IL-2R β , recettore per la vitamina D3 (VDR), mucina-1. La mucina-1 rappresenta un possibile bersaglio per strategie di immunoterapia, mentre l'espressione di VDR offre la possibilità di terapie differenzianti con agonisti recettoriali.

Neoplasie delle cellule T periferiche

L'analisi GEP ha permesso di distinguere, da un punto di vista molecolare, i linfomi di cellule T periferiche (PTCL) da linfomi linfoblastici T. Studi successivi hanno quindi evidenziato le peculiarità GEP dei PTCL "not otherwise specified" (NOS) rispetto a cellule T normali o altre neoplasie T. Uno studio recente ha evidenziato l'iperespressione di platelet derived growth factor receptor alpha (PDGFRA) come un marcatore biologico dei PTCL NOS. PDGFRA rappresenta un potenziale bersaglio di terapia molecolare con piccoli inibitori, tra cui imatinib.

Prospettive

Le conoscenze derivate dall'analisi GEP devono poter essere applicate alla pratica clinica convenzionale dei linfomi, come già è avvenuto, ad esempio, nel caso di ZAP-70 nel contesto della leucemia linfatica cronica. A tale fine, la tecnologia GEP deve essere tradotta in sistemi più semplici e di basso costo, quali ad esempio le tecniche di immunoistochimica o di Q-PCR.

Inoltre, l'analisi GEP ci permetterà di esplorare settori della patologia dei linfomi ancora oscuri, con particolare riguardo alla diagnostica differenziale dei linfomi indolenti non-follicolari, ai linfomi post-trapianto, ed ai linfomi delle cellule T periferiche.

Bibliografia

1. Alizadeh AA, Eisen MB, Davis RE, Ma C, Lossos IS, Rosenwald A, et al. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* 2000; 403: 503-511.
2. Colomo L, Lopez-Guillermo A, Perales M, Rives S, Martinez A, Bosch F, et al. Clinical impact of the differentiation profile assessed by immunophenotyping in patients with diffuse large B-cell lymphoma. *Blood* 2003; 101: 78-84.
3. Dave SS, Wright G, Tan B, et al. Prediction of survival in follicular lympho-

- ma based on molecular features of tumor-infiltrating immune cells. *N Engl J Med* 2004; 351: 2159-2169.
4. De Paepe P, De Wolf-Peeters C. Diffuse large B-cell lymphoma: a heterogeneous group of non-Hodgkin lymphomas comprising several distinct clinico-pathologic entities. *Leukemia* 2007; 21: 37-43.
 5. Deeve S. Gene expression profiling and outcome prediction in non-Hodgkin lymphoma. *Biol Blood Marrow Transpl* 2006; 12: 50-52.
 6. Dunphy CH. Gene expression profiling data in lymphoma and leukemia. *Arch Pathol Lab Med* 2006; 130: 483-520.
 7. Gaidano G, Dalla-Favera R. Pathobiology of non-Hodgkin's lymphomas. In *Hematology. Basic Principles and Practice*, 4th edition. Hoffman R, Benz EJ, jr., Shattil SJ, Furie B, Cohen HJ, Silberstein L, McGlave P (eds.). Philadelphia: Elsevier, 2005; 1307-1324.
 8. Hans CP, Weisenburger DD, Greiner TC, Gascoyne RD, Delabie J, Ott G, et al. Confirmation of the molecular classification of diffuse large B-cell lymphoma by immunohistochemistry using a tissue microarray. *Blood* 2004; 103: 275-282.
 9. Klein U, Ghoghini A, Gaidano G, Chadburn A, Cesarman E, Dalla-Favera R, Carbone A. Gene expression profile analysis of AIDS-related primary effusion lymphoma (PEL) suggests a plasmablastic derivation and identifies PEL-specific transcripts. *Blood* 2003; 101: 4115-4121.
 10. Klein U, Ghoghini A, Tirelli U, Gaidano G, Cesarman E, Carbone A, Dalla-Favera R. Gene expression profile of AIDS-related non-Hodgkin lymphomas. *Blood* 100: 316a, 2002.
 11. Lossos IS, Morgensztern D. Prognostic biomarkers in diffuse large B-cell lymphoma. *J Clin Oncol* 2006; 24: 995-1007.
 12. Martinez-Delgado B, Melendez B, Cuadros M, et al. Expression profiling of T-cell lymphomas differentiates peripheral and lymphoblastic lymphomas and defines survival related genes. *Clin Cancer Res* 2004; 10: 4971-4982.
 13. Martinez-Delgado B. Peripheral T-cell lymphoma gene expression profiles. *Hematol Oncol* 2006; 24: 113-119.
 14. Piccaluga PP, Agostinelli C, Zinzani PL, Baccarani M, Dalla-Favera R, Pileri SA. Expression of platelet-derived growth factor receptor alpha in peripheral T-cell lymphoma not otherwise specified. *Lancet Oncol* 2005; 6: 440.
 15. Rosenwald A, Wright G, Chan WC, Connors JM, Campo E, Fisher RI, et al. The use of molecular profiling to predict survival after chemotherapy for diffuse large B-cell lymphoma. *N Engl J Med* 2002; 346: 1937-1947.
 16. Rosenwald A, Wright G, Wiestner A, et al. The proliferation gene expression signature is a quantitative integrator of oncogenic events that predict survival in mantle cell lymphoma. *Cancer Cell* 2003; 3: 185-197.
 17. Shipp MA, Ross KN, Tamayo P, Weng AP, Kutok JL, Aguiar RCT, et al. Diffuse large B-cell lymphoma outcome prediction by gene-expression profiling and supervised machine learning. *Nat Med* 2002; 8: 68-74.

Classificazione molecolare del mieloma multiplo

Carolina Terragna, Michele Cavo

Istituto di Ematologia ed Oncologia Medica "L. e A. Seràgnoli", Università di Bologna

Nel mieloma multiplo (MM) le plasmacellule midollari presentano molteplici e complesse alterazioni citogenetico-molecolari, la cui presenza, un tempo ritenuta casuale, è associata a peculiari caratteristiche di presentazione e ad un differente decorso clinico. Questa stretta correlazione ha reso possibile la suddivisione della malattia in differenti sottogruppi sulla base delle caratteristiche citogenetico-molecolari, che di ciascun sottogruppo rappresentano veri e propri marcatori biologici, clinici e prognostici.

Insorgenza e progressione del mieloma multiplo

Il MM è una neoplasia della linfocitopoiesi B, la cui insorgenza può avvenire *de novo* oppure rappresentare la trasformazione di una pregressa gammopatia monoclonale di significato non determinato (MGUS). Nel corso della storia naturale della malattia, sono fondamentalmente riconoscibili due differenti fasi. La prima fase, intramidollare e stroma-dipendente, è caratterizzata da un decorso clinico relativamente indolente e responsivo alla terapia; ad essa fa invariabilmente seguito una seconda fase, di trasformazione, extramidollare e stroma-indipendente, che è caratterizzata dall'acquisizione di una elevata resistenza alla terapia. È stato riportato che, alla progressione del MM attraverso queste fasi contribuisce in misura determinante la progressiva acquisizione di alterazioni genético-molecolari. Nell'ambito della marcata instabilità genomica che caratterizza la storia naturale del MM, è stato inoltre possibile identificare un certo numero di alterazioni cromosomiche ricorrenti che rappresentano, al momento attuale, uno dei principali criteri che sono alla base della scelta di opzioni terapeutiche differenzialmente aggressive o di strategie terapeutiche comprensive dell'impiego di nuovi farmaci.

Le alterazioni genético-molecolari

Da un punto vista molecolare, il clone neoplastico mielomatoso è definito da un insieme di entità biologiche geneticamente diverse fra loro. Attraverso l'applica-

zione di tecniche di citogenetica molecolare (FISH) e biologia molecolare (profilo di espressione genica) è stato possibile mettere in evidenza che un certo numero di alterazioni cromosomiche, numeriche e strutturali sono ricorrenti, e inoltre caratterizzano sottogruppi di pazienti con differenti espressioni cliniche, diverso decorso, prognosi e risposta alla terapia. Ad esempio, sulla base della presenza o assenza di iperdiploidia, i pazienti con MM possono essere divisi in due grandi gruppi (1, 2): la condizione di non-iperdiploidia è tipicamente accompagnata da una elevata frequenza della delezione del cromosoma 13 e dalla presenza di traslocazioni che coinvolgono il cromosoma 14, a livello del locus della catena pesante delle immunoglobuline (IgH), e alcuni cromosomi ricorrenti. Questo tipo di traslocazione determina la de-regolazione trascrizionale di proto-oncogeni normalmente silenti, quali la ciclina D1 (*CCND1*), iper-espressa per effetto della $t(11;14)$; *FGFR3* e *MMSET*, in pazienti con la $t(4;14)$; *MAF* e *MAFB*, in associazione, rispettivamente, alla $t(14;16)$ e $t(14;20)$; ed infine la ciclina D3 (*CCND3*) in pazienti con $t(6;14)$. I pazienti con iperdiploidia sono invece caratterizzati da trisomie, più frequentemente dei cromosomi 3, 5, 7, 9, 11, 15, 19 e 21.

Come già segnalato, studi eseguiti su ampie serie di pazienti con malattia di nuova diagnosi, trattati sia con chemioterapia convenzionale, che sottoposti a chemioterapia ad alte dosi, hanno consentito di attribuire ad alcune di queste alterazioni genico-molecolari un differente significato prognostico. In particolare, la presenza della $t(4;14)$, della $del(13)$ e della amplificazione del cromosoma 1, sono state associate a cattiva prognosi, sia dopo l'impiego di chemioterapie convenzionali, che con chemioterapie ad alte dosi e successivo trapianto di cellule staminali emopoietiche (3-7). Al contrario, la $t(11;14)$ è stata associata a prognosi favorevole dopo singolo o doppio trapianto autologo (8, 9). L'espressione di un diverso rischio prognostico del MM in associazione a particolari alterazioni genico-molecolari supporta l'ipotesi che la presenza di queste ultime possa tradursi nell'attivazione di particolari *pathways*, non ancora identificati, responsabili di differenti risposte alla terapia.

Recentemente, anche lo studio del profilo di espressione genica di plasma cellulare ottenute da pazienti con MM di nuova diagnosi, ha messo in evidenza che pazienti portatori all'esordio della stessa alterazione cromosomica presentano un profilo di espressione genica significativamente simile tra loro, supportando ulteriormente l'ipotesi che esistano sottotipi di MM ben definiti anche da un punto di vista molecolare (10).

La de-regolazione trascrizionale delle cicline D

Nonostante le plasmacellule mielomatose siano caratterizzate da un basso indice proliferativo, in tutti i pazienti all'esordio esse presentano la de-regolazione trascrizionale di almeno una delle tre cicline D (11, 12). Le cicline D1, D2 e D3 hanno un ruolo molto importante nella regolazione del ciclo cellulare; esse infatti, interagendo con le chinasi ciclino-dipendenti *cdk4* o *cdk6*, regolano la fosforilazione del retinoblastoma (RB1), consentendo la transizione G1/S del ciclo cellulare. La de-regolazione trascrizionale delle cicline nel clone mielomatoso può

avvenire direttamente o indirettamente: circa il 25% dei MM presenta una traslocazione coinvolgente il locus IgH che ha come conseguenza diretta la de-regolazione trascrizionale della ciclina D1 (11q23), oppure della ciclina D3 (6p21), oppure di un gene MAF (*MAF*, 16q23 o *MAFB*, 20q11), che a sua volta codifica per un fattore di trascrizione che ha come target la ciclina D2. Circa il 40% dei MM, pur non presentando la t(11;14), è iperdiploide e presenta una polisomia del cromosoma 11, e di conseguenza una espressione della ciclina D1 significativamente superiore a quella delle plasmacellule di un midollo normale. Infine i restanti MM, inclusi quelli che presentano la t(4,14), presentano una over-espressione della ciclina D2, rispetto alle plasmacellule di midolli sani.

La classificazione TC

Sulla base di queste osservazioni è stata proposta una nuova classificazione dei pazienti con MM basata essenzialmente sulle caratteristiche molecolari del clone neoplastico, che tiene conto non solo della de-regolazione trascrizionale di alcuni oncogeni come conseguenza delle traslocazioni che coinvolgono il locus IgH, ma anche della de-regolazione trascrizionale di una delle tre cicline D (11, 12). Da questo nuovo punto di vista, i pazienti con MM possono essere suddivisi all'esordio in un certo numero di "sottogruppi TC" (laddove T indica *translocation* e C indica *cyclin D*), a seconda che le rispettive plasmacellule midollari presentino o meno almeno una delle alterazioni cromosomiche ricorrenti e contemporaneamente esprimano, o meno, in maniera de-regolata una delle cicline D. Il valore di questa suddivisione è stato ulteriormente confermato dal fatto che in studi più recenti, l'analisi supervisionata del profilo di espressione genica di un consistente numero di pazienti con MM all'esordio, ha mostrato come la classificazione TC identifichi gruppi omogenei di tumore, con un pattern distinto di espressione genica e quindi di fenotipo (10), suggerendo inoltre che i diversi sottotipi TC possano condividere anche importanti caratteristiche biologiche e cliniche. A conferma di ciò, un recente studio ha integrato i risultati dell'analisi del profilo di espressione genica nel contesto di un determinato trattamento terapeutico (13), dimostrando significative correlazioni tra i vari sottogruppi di MM e la prognosi della malattia.

La patogenesi del MM

La de-regolazione trascrizionale della ciclina D, come evento biologico comune a tutti i pazienti con MM, potrebbe rappresentare l'elemento universale alla base della patogenesi di questa neoplasia. Infatti l'espressione de-regolata di una delle tre cicline D potrebbe essere determinante nell'aumentare la suscettibilità della cellula mielomatosa nei confronti degli stimoli proliferativi. Questo potrebbe determinare l'espansione selettiva del clone neoplastico, grazie anche alla stretta interazione con le cellule stromali del midollo osseo che, a loro volta, producono citochine e fattori di crescita in maniera de-regolata. Questi eventi, insieme alla perdita del cromosoma 13 e alla presenza delle traslocazioni primarie che coin-

volgono il locus IgH potrebbero rappresentare gli eventi più precoci durante la patogenesi del MM, cui si sovrappongono in un secondo tempo alterazioni cromosomiche secondarie, altre alterazioni cariotipiche e infine mutazioni puntiformi attivanti nei geni *K-RAS*, *N-RAS* o *FGFR3*. Quindi, nonostante l'enorme varietà di sequenziali eventi genetico-molecolari che avverrebbero durante la storia naturale del MM, il fenotipo tumorale finale può essere efficacemente ricondotto all'esistenza di sottogruppi definiti dalla classificazione TC; questi sottogruppi potrebbero rappresentare delle entità patologiche diverse, se pur simili, che potrebbero richiedere diversi approcci terapeutici.

Conclusioni

La proposta di questa nuova classificazione su base molecolare dei pazienti con MM, similmente a quanto già osservato anche nei linfomi e in altre patologie leucemiche, oltre ad un interesse formale, potrebbe anche avere un importante impatto sulla clinica del MM. Infatti, sapendo che il contesto biomolecolare di un paziente con MM di nuova diagnosi può avere un'influenza sul suo decorso clinico, diventa possibile impostare precocemente un approccio terapeutico mirato, che possa quindi risultare maggiormente efficace per il paziente stesso.

Bibliografia

1. Smadja NV, Bastard C., Brigaudeau C., Leroux D. et al. (2001) Hypodiploidy is a major prognostic factor in multiple myeloma. *Blood*, 2001; 98: 2229-2238.
2. Cremer FW, Bila J, Buck I Kartal M et al. Delineation of distinct sub groups of multiple myeloma and a model for clonal evolution based on interphase cytogenetics. *Genes Chromosomes cancer* 2005; 44,
3. Keats JJ, Reiman T, Maxwell CA, Taylor BJ et al. In multiple myeloma, t(4;14)(p16;q32) is an adverse prognostic factor irrespective of FGFR3 expression. *Blood*, 2002; 101: 1520-1529.
4. Stewart AK, Fonseca R. Prognostic and therapeutic significance of myeloma genetics and gene expression profiling. *J.Clin. Oncol.* 2005; 23: 6339-6344.
5. Magrangeas F, Lode L, Wuilleme S, Minvielle S, Avet-Loiseau H. Genetic heterogeneity in multiple myeloma. *Leukemia*, 2005; 19: 191-194.
6. Chng WJ, Santana-Davila R, Van Wier SA, Ahmann GJ et al. Prognostic factors for hyperdiploid-myeloma: effects of chromosome 13 deletions and IgH translocations. *Leukemia*, 2006; 20: 807-813.
7. Cavo M., Terragna C., Renzulli M., Zamagni E. et al. Poor outcome with front-line autologous transplantation in t(4;14) multiple myeloma: low complete remission rate and short duration of remission. *J Clin Oncol*, 2006 24 (3): e4-5.
8. Soverini S, Cavo M, Cellini C, Terragna C et al. Cyclin D1 overexpression is a favorable prognostic variable for newly diagnosed multiple myeloma

- patients treated with high-dose chemotherapy and single or double autologous transplantation. *Blood*, 2003; 102: 1588-1594.
9. Chang H, Li D, Zhuang L, Nie E et al. Detection of chromosome 13q deletions and IgH translocations in patients with multiple myeloma by FISH: comparison with karyotype analysis. *Leuk Lymphoma*, 2004; 45: 965-969.
 10. Agnelli, L, Bricciato S, Mattioli M, Fabris S et al. Molecular classification of multiple myeloma: a distinct transcriptional profile characterizes patients expressing CCND1 and negative for 14q32. *J Clin Oncol*, 2005; 23: 7296-7306.
 11. Hideshima T, Bergsagel PL, Kuehl WM, Anderson K Advances in biology of multiple myeloma: clinical applications. *Blood*, 2004; 104: 607-618.
 12. Bergsagel PL, Kuehl WM, Zhan F et al. Cyclin D dysregulation: an early and unifying pathogenetic event in multiple Myeloma. *Blood*, 2005; 106: 296-303.
 13. Zhan F, Huang Y, Colla S, Stewart JP et al. The molecular classification of multiple myeloma. *Blood*, 2006; 108: 2020-2028.

Profilo dell'espressione genica nel tumore della mammella: rilevanza diagnostica e prognostica

Giovanna Chiorino

Laboratorio di Farmacogenomica Oncologica, Fondo Edo Tempia, Biella

La tecnologia dei microarrays consente l'analisi dei livelli di espressione di diverse migliaia di geni in un singolo esperimento e offre un'enorme potenzialità non solo nel miglioramento della nostra conoscenza sulla biologia molecolare dei tumori, ma anche nella scoperta di nuovi marker molecolari e bersagli terapeutici. Nel caso del tumore della mammella studi di grande rilevanza, seppur criticati dal punto di vista strettamente statistico, hanno ad esempio consentito l'estrazione di un numero limitato di geni utili per la scelta terapeutica post-intervento in pazienti linfonodo-negative. Nessun altro parametro clinico-patologico è infatti in grado di predire altrettanto bene l'aggressività del tumore in tale classe di pazienti. Generalmente, nei progetti di ricerca che prevedono l'utilizzo dei microarrays, si parte effettuando uno screening su vasta scala con supporti "whole genome", per poi successivamente focalizzare l'attenzione su un sottoinsieme ristretto di sequenze differenzialmente espresse in un sottogruppo specifico di campioni, o in seguito ad un particolare trattamento degli stessi.

Il primo step fondamentale per l'analisi dei profili di espressione genica in campo oncologico, che ne costituisce tuttavia un grosso limite, è l'estrazione di RNA integro e puro da tessuto tumorale (in particolare per i tumori solidi). Recentemente stanno uscendo nuovi protocolli che permettono l'analisi di un numero limitato di geni (fino a circa 1.500) utilizzando RNA anche parzialmente degradato estratto da materiale d'archivio fissato ed incluso in paraffina, ma per un'analisi su vasta scala di tutto il genoma è fondamentale disporre di RNA di ottima qualità. È pertanto necessario che il tessuto sia o immediatamente congelato in azoto liquido oppure immerso in soluzioni che preservano l'RNA subito dopo il prelievo (sia biotico sia proveniente da pezzo chirurgico).

Un altro grosso punto di dibattito, per alcuni tumori solidi quali il carcinoma mammario, riguarda il grado di omogeneità del tipo cellulare, sia per quanto concerne la distinzione tra cellule neoplastiche e non, sia per quanto riguarda la separazione non solo tra cellule epiteliali, stromali e infiltrati di varia origine, ma anche tra cellule epiteliali luminali e cellule mioepiteliali basali. Per poter rende-

re uno studio il più generalizzabile possibile, teoricamente non dovrebbe essere necessario effettuare le analisi su tipi cellulari distinti (attraverso l'uso di microdissezione laser, ad esempio), perché la natura stessa dei prelievi bioptici è costituita da una mescolanza di tipi diversi. Bisogna però tenere presente che la componente non tumorale presente nei campioni tumorali bioptici ha un effetto significativo sull'identificazione dei profili di espressione genica degli stessi. Risulta pertanto fondamentale l'avere a disposizione casistiche molto numerose per poter veramente apprezzare differenze nel profilo di espressione associabili a sottotipi tumorali distinti o correlabili con la prognosi, e quindi applicabili nella pratica clinica. La necessità di casistiche numerose è di natura molteplice: oltre che per le ragioni sopra descritte, anche - e soprattutto - perché con la tecnica dei microarrays si effettuano analisi di diverse migliaia di geni in simultanea, pertanto c'è un grosso sbilanciamento tra il numero dei campioni e il numero di test fatti per estrarre i geni differenzialmente espressi. Dovendo applicare criteri di correzione, proprio per tenere conto dei molteplici test, si rischia di non considerare significativo alcun cambiamento di espressione tra le classi esaminate, anche se in effetti la differenza c'è. Aumentando il numero di casi, si può ridurre questo tipo di sbilanciamento.

Un altro elemento da tenere in considerazione è l'omogeneità dei casi. Bisogna infatti stare attenti a lavorare con casistiche non affette da eterogeneità riguardo a stadio, trattamento terapeutico e/o altri parametri clinico-patologici, per non incrementare ulteriormente la variabilità dei dati introducendo elementi di bias. La tecnologia dei microarrays è stata molto applicata nello studio del tumore della mammella, sia dal punto di vista della classificazione tumorale sia per quanto riguarda l'identificazione di specifiche firme molecolari associate alla prognosi. Gli articoli scientifici e le rassegne più rilevanti pubblicati negli ultimi anni possono essere così suddivisi:

Classificazione molecolare del tumore alla mammella

- Sorlie T. Molecular classification of breast tumors: toward improved diagnostics and treatments. *Methods Mol Biol.* 2007; 360: 91-114.
- Grigoriadis A et al., "Establishment of the epithelial-specific transcriptome of normal and malignant human breast cells based on MPSS and array expression data." *Breast Cancer Res.* 2006; 8(5): R56.
- Sorlie T et al., Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci USA* 2001; 98: 10869-74.
- Perou CM et al., Molecular portraits of human breast tumours. *Nature* 2000; 406: 747-52.

Predizione di outcome e di comparsa di metastasi nel tumore della mammella

- Reis-Filho JS et al., The impact of expression profiling on prognostic and predictive testing in breast cancer. *J Clin Pathol.* 2006 Mar; 59(3): 225-31.

- Cleator SJ et al., The effect of the stromal component of breast tumours on prediction of clinical outcome using gene expression microarray analysis. *Breast Cancer Res.* 2006; 8(3): R32.
- Weigelt B et al. Molecular portraits and 70-gene prognosis signature are preserved throughout the metastatic process of breast cancer. *Cancer Res* 2005; 65(20): 9155-58).
- Hennessy BT et al. Individualization of neoadjuvant therapy for breast cancer according to molecular tumor characteristics. *Nature clinical practice oncology* 2005; 2(12): 598-99.
- van 't Veer LJ et al. Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 2002; 415: 530-6.
- van de Vijver MJ et al. A gene expression signature as a predictor of survival in breast cancer. *N Engl J Med* 2002; 347: 1999-2009.

Durante il seminario verrà illustrato in modo dettagliato lo studio condotto da Grigoriadis et al., nel quale è stata applicata una metodica di purificazione del campione attraverso separazione immunomagnetica di tipi cellulari diversi. Inoltre, un'analisi di "arricchimento" (gene set enrichment analysis) ha permesso di evidenziare le classi funzionali maggiormente rappresentate in insiemi di geni differenzialmente espressi, ottenuti peraltro utilizzando 4 diverse piattaforme per microarrays e un metodo di sequenziamento. In particolare, è stato possibile definire delle firme molecolari associate non solo alla transizione tra epitelio normale luminale ad epitelio tumorale, ma anche allo specifico fenotipo epiteliale (luminale o mioepiteliale basale). In particolare, si è visto che anche tumori positivi per il recettore per gli estrogeni (ER+), che erano stati precedentemente classificati come "luminal like", possono overesprimere geni di tipo "mioepiteliale" ed avere un fenotipo associato a prognosi negativa. Alcuni di questi geni sono stati inoltre validati con tecniche di immunohistochimica su tissue-arrays di tumore mammario

La cellula di origine della leucemia linfatica cronica

Manlio Ferrarini

Istituto Nazionale per la Ricerca sul Cancro e DOBIG, Università di Genova

La LLC, nella sua definizione classica, accettata da generazioni di Ematologi, è stata vista come una malattia da accumulo cellulare. Tale accumulo, si pensava, fosse dovuto alla lenta espansione di un clone, le cui cellule erano incapaci di iniziare un normale processo apoptotico. Questi concetti hanno subito nei tempi più recenti una radicale rivisitazione sulla base di nuove osservazioni. Ad esempio, con il miglioramento delle tecniche di cultura cellulare in vitro, si è visto che la cellula di LLC ha una forte tendenza alla apoptosi, la quale può essere modulata da citochine o interazioni con cellule del microambiente. Inoltre, con la conoscenza dei meccanismi enzimatici di apoptosi, si è potuto affrontare in maniera rigorosa il problema della capacità apoptotica della cellula neoplastica e constatare l'assenza di evidenti difetti intrinseci. Infine, lo studio del repertorio dei geni delle immunoglobuline (geni VH/VL, che codificano per il sito combinatorio dell'anticorpo) ha rivelato che la cellula leucemica è una cellula memoria, cioè con esperienza di incontro con l'antigene, prima, durante e magari anche dopo il processo di trasformazione. Queste osservazioni hanno condotto alla rivisitazione di un altro dei classici concetti che definivano la LLC, cioè quello di una malattia generata da cellule immuno-incompetenti incapaci di rispondere ad una stimolazione antigenica. Il risultato degli studi più recenti ha portato alla definizione della LLC come una entità dinamica, le cui cellule sono capaci di proliferare attivamente e di essere anche eliminate per apoptosi. La proliferazione, la sopravvivenza e anche le dimensioni del clone neoplastico sono condizionate non solo dalle caratteristiche intrinseche delle cellule, ma anche dal microambiente, inclusa la stimolazione da parte di antigeni e più specificamente di auto-antigeni.

I nuovi concetti riassunti sopra comportano una ridefinizione della cellula di origine della LLC, che passa da cellula vergine-immuno-incompetente, a cellula memoria-immuno-competente, e che in una visione patogenetica attuale deve avere particolari requisiti di risposta agli antigeni e di capacità interattive con il micro-ambiente.

La cellula di LLC è una cellula memoria

Nell'ambito del sistema immunitario, una cellula memoria viene definita come quella cellula che nella sua vita è già stata stimolata dall'antigene e, probabilmen-

te, è andata incontro ad un processo espansivo clonale. Questo secondo requisito, tuttavia, non è necessario, in quanto, una cellula può incontrare l'antigene ed essere resa "tollerante", cioè incapace di rispondere a ulteriori stimoli.

Non sempre è possibile distinguere tra cellule memoria e cellule "tolleranti", pertanto spesso si allude alle cellule che hanno incontrato l'antigene come cellule "antigen-experienced".

La cellula di LLC possiede queste caratteristiche di cellula antigen-experienced? Per la metà circa dei casi di LLC, la risposta è facile ed è positiva. In queste LLC, infatti, le cellule neoplastiche presentano mutazioni somatiche nei geni VH/VL che codificano per la parte variabile delle Ig espresse in membrana. Dato che l'acquisizione di una quantità sostanziale di mutazioni (cioè una deviazione maggiore del 2% della sequenza nucleotidica rispetto al gene "germ-line") avviene solamente durante una risposta immunitaria, è necessario ammettere che in questi casi le cellule sono "antigen-experienced".

L'incontro con l'antigene è probabilmente avvenuto prima e/o durante la trasformazione neoplastica. Si discuterà, di seguito, se questa stimolazione continui ad avvenire dopo la trasformazione neoplastica, agendo da fattore promotore della espansione clonale.

Le cellule di LLC che non utilizzano geni VH/VL mutati, sono anch'esse cellule "antigen-experienced"? Una risposta a questo importante quesito proviene dall'ampio studio condotto in questi anni sul repertorio genico utilizzato dai cloni di LLC per codificare per le molecole immunoglobuliniche.

Il razionale alla base di questi studi è il seguente: se, nel corso di una trasformazione neoplastica o antecedentemente ad essa, vi è una stimolazione antigenica, il repertorio dei geni VH/VL della LLC, desunto dallo studio di un vasto numero di casi, dovrebbe essere differente da quello delle cellule B normali (vergini), poiché la stimolazione antigenica deve aver operato una selezione positiva in favore di certi geni o combinazioni geniche VH/VL e negativa a sfavore di altri.

In assenza di un processo selettivo, e quindi di un repertorio uguale a quello delle cellule normali, la stimolazione antigenica deve considerarsi non operativa. I dati sino ad ora in nostro possesso hanno dimostrato che:

- 1) le cellule di LLC usano più frequentemente certi segmenti VH in paragone alle cellule vergini;
- 2) esse usano spesso combinazioni particolari di geni VH/VL che non si ritrovano nelle cellule vergini con altrettanta frequenza;
- 3) vi sono LLC che presentano combinazioni VH/VL in comune; studi approfonditi sulla struttura degli anticorpi prodotti da queste cellule dimostrano come questi potrebbero reagire con lo stesso determinante antigenico. Si calcola che circa il 20 delle LLC utilizzi recettori Ig il cui sito combinatorio rientra in una categoria nota, codificata da combinazioni VH/VL predeterminate.

Nel complesso questi dati suggeriscono che vi sia una forte pressione antigenica a selezionare quei cloni che poi saranno destinati alla trasformazione neoplastica.

La stimolazione antigenica può agire come agente promotore nella espansione del clone neoplastico?

Per rispondere a questa domanda, è necessario rispondere preliminarmente ad altri due quesiti e cioè:

- 1) se nelle cellule di LLC, sia ancora conservata la pathway di traduzione del segnale iniziata da IgM;
- 2) se vi sia una differenza nella prognosi a seconda dei recettori espressi dalle cellule e dello stato della pathway recettoriale.

In realtà, in una certa percentuale, sebbene non in tutte le LLC, le cellule conservano una funzionante via di traduzione del segnale iniziata da IgM. Queste caratteristiche si ritrovano per lo più nei casi con marcatori prognostici sfavorevoli, cioè assenza di mutazioni somatiche nei geni VH, ed espressione di CD38 e di ZAP-70. Queste osservazioni servono in parte a rispondere alla seconda domanda e cioè che vi è correlazione fra prognosi infausta e capacità delle cellule a essere stimulate da un potenziale antigene. A questo proposito, è interessante notare che le cellule di LLC “non mutate” generalmente esprimono recettori di membrana con reattività a tipo anticorpi polispecifici, capaci cioè di reagire con una molteplicità di antigeni, inclusi autoantigeni.

Questa reattività non si ritrova nelle Ig di membrana espresse dalle cellule di LLC “mutate”. Nel loro complesso, queste osservazioni hanno suggerito il seguente modello patogenetico: le cellule di LLC “non mutate” sarebbero continuamente stimulate a proliferare da auto-antigeni presenti in vivo; da qui una maggiore espansione del clone neoplastico, una facilità ad accumulare modificazioni citogenetiche tardive (come mutazioni del gene ATM o di p53) e, in ultima analisi, un decorso clinico più sfavorevole. Questa stimolazione continua porterebbe anche all'espressione da parte delle cellule di svariati marcatori di attivazione, inclusi ZAP-70 e CD38. Questa stimolazione invece non avverrebbe per le cellule di LLC “mutate” che non esprimono Ig di membrana con autoreattività e la cui proliferazione sarebbe dettata solo da alterazioni citogenetiche “intrinseche” alle cellule e ancora non definite. Si potrebbe anche ipotizzare che le cellule di LLC “mutate” che, come si è detto, non rispondono in genere a stimolazioni attraverso le Ig di membrana, potrebbero essere cellule rese tolleranti attraverso pregresse stimolazioni antigeniche, avvenute durante il processo di accumulo delle mutazioni VH/VL.

Da quale cellula origina la LLC?

I dati sin qui descritti confermano che la LLC chiaramente deriva da una cellula matura, e memoria secondo la definizione precedentemente espressa. Questa affermazione non basta a indicare la cellula di origine della LLC, in quanto oggi si conoscono svariate sottopopolazioni di cellule B mature ognuna delle quali potenzialmente capace di originare una cellula memoria. Il problema non è attualmente risolto, anche se si può tentare dapprima di delineare le sottopopolazioni

finora conosciute e la loro fisiologia e poi tentare di definire quelle cellule che potrebbero essere capaci di dare luogo a una LLC.

- a) sottopopolazioni di cellule B. Le cellule B, una volta completato il loro processo maturazionale nel midollo osseo o nel fegato fetale, escono come cellule transizionali, capaci di maturare ulteriormente prima di poter essere stimulate dall'antigene. Una volta completato il processo maturativo, le cellule B possono insediarsi nel follicolo linfatico degli organi linfoidi periferici o nella zona marginale della milza o nei suoi equivalenti (come l'area subepiteliale delle tonsille o le aree subcapsulari del linfonodo). In queste aree, le cellule vanno incontro a processi differenziativi differenti, ma sempre dipendenti dalla stimolazione antigenica. Nel follicolo, le cellule, una volta stimulate, passano dal mantello al centro germinativo; qui iniziano un processo di mutazione somatica dei geni VH/VL che portano all'espressione di Ig di membrana con nuove specificità. L'antigene stimolante provvederà a selezionare quelle cellule dotate del sito combinatorio con maggiore affinità. Questo processo di stimolazione/selezione, che esita nella produzione di anticorpi IgG ad alta affinità da parte di plasma-cellule, che si formano nel centro germinativo prima di uscire e localizzarsi nel midollo osseo, si svolge sotto il controllo e con l'aiuto di cellule T. Ugualmente, nel centro germinativo, si formano i linfociti B memoria, che una volta usciti, vanno a localizzarsi per lo più nella zona marginale dello stesso o di altri organi linfoidi. Anche questo processo richiede il controllo delle cellule T. Nella zona marginale, le cellule B vergini possono essere stimulate da antigeni (generalmente polisaccaridi di membrana batterici) in modo T indipendente. Le cellule così stimulate si espandono e danno origine a plasmacellule, per lo più secernenti IgM. Queste plasmacellule hanno una tendenza residenziale e tendono a rimanere nella zona marginale dello stesso organo, anche se possono emigrare in quella di altri organi o nel midollo. Nello stesso modo, si generano cellule B memoria. Nelle cellule stimulate in maniera timo indipendente si può verificare l'accumulo di mutazioni VH/VL, anche se questa, a differenza di ciò che avviene nel centro germinativo, non è una tappa obbligata. Nello stesso modo, le cellule della zona marginale possono andare incontro a uno switch isotipico passando dalla produzione di IgM a quella di IgG o IgA. Anche in questo caso, e contrariamente di ciò che avviene nel centro germinativo, lo switch isotipico non è un passo differenziativo obbligato.
- b) La LLC deriva da una o da due cellule differenti? Da quanto detto sopra ci si pone il problema se i due principali sottotipi di LLC derivino dalla stessa cellula o da cellule differenti. Vediamo gli scenari delle due ipotesi:
- 1) origine dalla stessa cellula. In questo caso la cellula di origine sarebbe una cellula della zona marginale. Questa potrebbe essere stimolata in maniera timo indipendente nella zona marginale e accumulare mutazioni trasformanti durante la stimolazione. Durante questa stessa stimolazione, si potrebbe avere accumulo di mutazioni di geni V delle Ig che darebbero luogo alle LLC "mutate", mentre, se queste mutazioni non avessero luogo, si avrebbe l'origine delle LLC non mutate. È interessante notare che studi recenti hanno dimo-

strato che le cellule LLC mutate (che non producono anticorpi polispecifici) derivano da cellule producenti anticorpi polispecifici, che hanno perso questa capacità, a seguito dell'accumulo delle mutazioni VH/VL. Questa osservazione è quindi in supporto della origine "unitaria" della CLL, anche se non si ha ancora informazione esauriente sulle specificità e polispecificità degli anticorpi prodotti dalle cellule della zona marginale.

- 2) origine da due cellule differenti. Il modello è simile a quello discusso sopra per l'origine delle LLC "non mutate", ma differisce in quanto postula un passaggio nel centro germinativo per le cellule mutate. Su questa base le LLC sarebbero due, in quanto originate da sottopopolazioni di cellule B differenti (una dalla zona marginale, una dalle cellule uscite dal centro germinativo). La difficoltà ad accettare questo modello risiede nel fatto che le cellule memoria che fuoriescono dal centro germinativo sono di solito IgG, mentre le LLC mutate producenti IgG sono rare.

Bibliografia

1. Stevenson FK, Caligaris-Cappio F. Chronic lymphocytic leukemia: revelations from the B-cell receptor. *Blood*. 2004 Jun 15; 103(12): 4389-95. Epub 2004 Feb 12. Review.
2. Caligaris-Cappio F, Hamblin TJ. B-cell chronic lymphocytic leukemia: a bird of a different feather. *J Clin Oncol*. 1999 Jan; 17(1): 399-408.
3. Chiorazzi N, Ferrarini M. Evolving view of the in-vivo kinetics of chronic lymphocytic leukemia B cells. *Hematology Am Soc Hematol Educ Program*. 2006; 273-8.
4. Chiorazzi N, Rai KR, Ferrarini M. Chronic lymphocytic leukaemia. *New Engl. J. Med*. 2005; 352: 804-15.
5. Chiorazzi N., Ferrarini M. B cell chronic lymphocytic leukemia: lessons learned from studies of the B cell antigen receptor. *Annu. Rev. Immunol*. 2003; 21: 841/894.

**SEQUENZIAMENTO GENICO,
POLIMORFISMI, microRNA**

Dal sequenziamento del DNA alla post-genomica

Diego di Bernardo

Fondazione Telethon, TIGEM, Napoli

La Biologia del nostro secolo è entrata in una nuova fase, quella ‘post-genomica’, il cui inizio è stato segnato dal completamento del genoma umano. Grazie allo sviluppo tecnologico degli ultimi venti anni, oggi è possibile osservare e misurare, simultaneamente ed economicamente, le diverse componenti di una cellula in maniera quantitativa, standardizzata ed automatica. Le nuove biotecnologie mettono a disposizione della comunità scientifica enormi quantità di dati sperimentali che, per risultare utili, devono essere adeguatamente archiviati, facilmente accessibili e correttamente interpretati. La necessità di gestire ed analizzare una grossa mole di dati ha pertanto prodotto il rapido sviluppo della *bioinformatica* e della *biologia computazionale*.

Sebbene la distinzione tra queste due discipline non sia molto netta, la bioinformatica può essere definita come l’applicazione delle discipline informatiche all’archiviazione e alla gestione dei dati sperimentali, mentre la biologia computazionale può descriversi come l’applicazione delle discipline ingegneristiche e fisico-matematiche per l’analisi di questi stessi dati sperimentali.

La conoscenza completa delle sequenze genomiche di diversi organismi, le tecnologie per misurare i cambiamenti a livello molecolare della cellula durante una specifica funzione biologica, insieme allo sviluppo della bioinformatica e della biologia computazionale, hanno preparato il campo per la *biologia dei sistemi*.

La biologia dei sistemi è lo studio integrato e quantitativo dei modi in cui i diversi componenti della cellula interagiscono fra loro per dar luogo ad una specifica funzione biologica. Questo studio è eseguito da un gruppo di ricerca interdisciplinare formato non solo da biologi e medici, ma anche da fisici, matematici ed ingegneri. La ricerca biologica indica le nuove tecnologie e gli strumenti computazionali da sviluppare. Una volta implementati, tali strumenti aprono ulteriori frontiere nella ricerca biologica, innescando così un circolo virtuoso che, oltre a favorire l’integrazione tra diverse discipline scientifiche, sta cambiando radicalmente la biologia.

La biologia dei sistemi può essere definita, semplificando, come l’opposto dell’approccio riduzionistico in cui ogni elemento della cellula, ad esempio una proteina, è studiato come una singola unità e non come parte di un ingranaggio più

complesso. L'approccio riduzionistico, che ha guidato la scoperta scientifica in biologia nell'ultimo secolo, ha permesso un'approfondita conoscenza dei singoli geni, delle proteine e delle loro rispettive funzioni. Nonostante la validità e il successo di tale approccio, è però sempre più evidente che una particolare funzione biologica può soltanto raramente essere attribuita ad una singola molecola. Al contrario, la maggior parte delle funzioni biologiche nasce dalla complessa interazione tra i numerosi componenti della cellula, quali proteine, geni, RNA e metaboliti.

Le nuove parole del genoma

Il completamento del sequenziamento del genoma umano e, più recentemente, anche di quello murino e di altri organismi evolutivamente più distanti, rappresenta una delle più straordinarie conquiste della biologia moderna. Tuttavia, questo è soltanto il primo passo di un lungo cammino da compiere per arrivare a capire come l'informazione contenuta nella sequenza del DNA è elaborata dalla cellula.

Il DNA contiene le istruzioni per la costruzione di tutti i 'pezzi' necessari al funzionamento della cellula e allo sviluppo dell'organismo. Tale informazione però non è facile da leggere, tanto che può paragonarsi ad un antico papiro scritto in un linguaggio sconosciuto e rinvenuto tra le rovine di una città scomparsa: il testo è in nostro possesso ma, non conoscendo l'idioma in cui è redatto, il messaggio in esso contenuto resta inaccessibile. In realtà il codice non è del tutto ignoto, allo stato attuale della ricerca, gli studiosi conoscono alcune delle parole del genoma, come ad esempio i geni. Fino ad oggi, però, si è ritenuto che oltre il 95% del genoma fosse privo di significato e dovuto a piccoli errori accumulati nel nostro DNA durante l'evoluzione.

Con gran sorpresa è stato di recente scoperto che un gran parte del genoma ha invece un significato: circa il 60% del DNA contiene parole necessarie alla cellula per costituire la propria struttura e a svolgere ciascuna funzione. Si è inoltre scoperto che, al contrario di quanto si pensava, la maggioranza di queste parole non è costituita solo da geni, ma anche da nuove parole chiamate *RNA non-codificanti*, la cui funzione è tutt'ora ignota.

Un'altra scoperta ha, di fatto, sconvolto il cosiddetto *dogma centrale della biologia molecolare*, secondo il quale ad ogni parola corrisponderebbe un gene che è trascritto in RNA ed infine tradotto in una proteina. Si è, infatti, scoperto che la cellula ha la possibilità di leggere ogni parola in molti modi diversi, ciò significa che ad ogni parola corrispondono moltissimi componenti. Proprio questa corrispondenza non biunivoca tra parole e componenti rende i mammiferi diversi da un verme o da un moscerino della frutta.

Reti genetiche

La nuova sfida che la biologia molecolare si trova ad affrontare consiste nel capire come le varie molecole, già individuate attraverso la ricerca fin qui svolta, interagiscano tra loro: è, appunto, come avere tutti i pezzi di un motore su un tavolo,

ma senza le istruzioni per ricostruirlo, non sapremmo come procedere al montaggio. Al fine di svolgere le proprie funzioni vitali, la cellula coordina i geni, le proteine e gli RNA non-codificanti in una complessa rete di regolazione, che è chiamata *rete genetica*. Uno degli obiettivi della biologia dei sistemi è dunque l'identificazione delle diverse reti genetiche che danno luogo alle varie funzioni cellulari. Un esempio di tali funzioni è il cosiddetto *ciclo cellulare* che permette alla cellula di duplicarsi. Il *ciclo cellulare* è un processo molto complesso che coinvolge centinaia di geni (e non solo) la cui accensione e il cui spegnimento devono essere finemente regolati e coordinati nel tempo. Un piccolo errore può provocare gravi conseguenze nel funzionamento della cellula e dell'organismo di cui essa fa parte, ad esempio si ritiene che un difetto nei meccanismi di controllo del ciclo cellulare sia una delle cause del cancro. L'insieme delle interazioni tra geni, proteine e RNA che regolano il ciclo cellulare costituisce la *rete genetica* del *ciclo cellulare*.

Identificare le reti genetiche coinvolte nei diversi processi biologici è di straordinaria importanza. Difatti, una volta scoperte e rese note le interazioni tra le diverse componenti della rete genetica sotto osservazione, sarà possibile costruire un modello matematico del processo biologico preso in esame. La possibilità di costruire un modello matematico è di assoluta rilevanza perché attraverso di esso un calcolatore è in grado di simulare il comportamento del sistema biologico, in modo simile a quanto avviene oggi con i simulatori di volo. Tali simulatori sono modelli matematici che descrivono i rapporti tra tutte le variabili necessarie a permettere ad un aeroplano di volare (equazione di fluido-dinamica, parametri fisici del velivolo, etc.). I simulatori di volo consentono di predire il comportamento del velivolo nel caso di perturbazioni atmosferiche e di malfunzionamenti dei sistemi di bordo. Analogamente, i modelli matematici delle reti genetiche potrebbero essere impiegati per predire il comportamento della cellula in seguito a perturbazioni del suo stato, come il trattamento con un farmaco, od in seguito alla mutazione di un gene, come avviene per le malattie genetiche. La creazione di modelli matematici siffatti potenzierà enormemente la medicina, in quanto consentirà di conoscere in dettaglio i meccanismi molecolari delle diverse patologie e di sviluppare terapie mirate ed efficaci.

Applicazioni della biologia dei sistemi

Le metodologie proprie della biologia dei sistemi sono oggi applicabili grazie ai progressi degli ultimi anni nella misurazione della concentrazione delle migliaia di molecole presenti in un determinato istante nella cellula. Tra queste nuove tecnologie, quella di maggiore successo è senz'altro il 'microarray' che permette di misurare il profilo trascrizionale (la concentrazione di mRNA dei singoli geni) di una popolazione di cellule. Un microarray è un vetrino con una superficie di circa 1 cm² sul quale sono 'stampati' decine di migliaia di punti. Ogni punto corrisponde ad una breve sequenza di DNA o RNA (oligonucleotidi), ripetuta milioni di volte. Ai diversi punti corrispondono differenti oligonucleotidi, tanti quanti sono i geni di cui si vuole misurare la concentrazione (di solito circa 40'000). Il

funzionamento dei microarray si basa sulla proprietà di complementarità degli acidi nucleici; se l'oligonucleotide è complementare alla sequenza di un gene che si vuole misurare, quest'ultimo riconoscerà l'oligonucleotide come complementare e si formerà una doppia elica (ibridazione). In tal modo, è possibile misurare la concentrazione di un gene, contando, nel punto corrispondente sul microarray, quanti oligonucleotidi sono ibridati.

Malattie genetiche e multifattoriali

La patogenesi di una malattia, ovvero la catena d'eventi che causa una specifica patologia, è alla base della terapia e della prevenzione. Solo analizzando le cause di una malattia a livello molecolare è, infatti, possibile intervenire per ripristinare il corretto funzionamento del sistema biologico compromesso. In termini biomolecolari, la patogenesi di una malattia può essere descritta come il malfunzionamento di una *rete genetica*, ossia di quel complesso processo di interazioni tra geni, proteine, RNA e metaboliti la cui sinergia permette l'espletamento di una determinata funzione biologica. Ad esempio nelle malattie monogeniche dovute al cattivo funzionamento di un singolo gene, la proteina codificata dal gene non interagisce più con gli altri componenti (in questo caso si parla di "loss of function" cioè di "perdita di funzione"), oppure interagisce con componenti diversi da quelli con cui dovrebbe (nella quale evenienza si parla piuttosto di "gain of function" o, in italiano, di "acquisizione di funzione"). Appare chiaro quindi che per trovare una cura ad una patologia è fondamentale capire come la rete genetica cambi quando il gene non funziona in maniera appropriata. La biologia dei sistemi fornisce gli strumenti adatti a fronteggiare questa esigenza. In particolare, sono stati recentemente proposti metodi che permettono di identificare una rete genetica misurando la risposta della cellula (cioè il cambiamento nella concentrazione di tutti gli mRNA, ovvero di tutte le copie dei geni attivi presenti nella cellula) a diverse perturbazioni esterne ed analizzando le variazioni nella concentrazione degli mRNA al calcolatore con algoritmi atti a ricostruire le interazioni tra i diversi geni (ref).

In un prossimo futuro, questi stessi metodi permetteranno di capire come la rete genetica cambi in seguito ad una patologia e di individuare un metodo d'intervento che ne ripristini il corretto funzionamento.

Per le malattie multifattoriali più comuni, come il cancro, il problema è più complesso poiché il malfunzionamento non avviene più al livello del singolo gene, ma è dovuto a piccole imperfezioni di molte elementi della cellula, la cui contemporaneità predispone alla malattia. La biologia dei sistemi si sta sviluppando molto velocemente ed è facile prevedere che nell'arco di un decennio sarà possibile identificare le reti genetiche malfunzionanti anche per le più malattie complesse.

Farmacologia

In genere, le molecole farmacologiche attive interagiscono funzionalmente solo con un piccolo numero di geni, detti *geni-obiettivo*, benché siano migliaia i geni

attivi in una cellula (in realtà il farmaco interagisce con le proteine e non con i geni, che sono semplicemente l'istruzione a livello del DNA per la costruzione della proteina). Uno dei problemi che s'incontrano durante il processo di sviluppo di nuovi farmaci, consiste proprio nell'individuare i geni-obiettivo ('targets'). La conoscenza di tutti i geni-obiettivo di un farmaco è indispensabile per l'ottimizzazione dell'efficacia terapeutica e per la riduzione della tossicità di una molecola farmacologica. Tuttavia, i geni-obiettivo di molte molecole farmacologiche sono tuttora difficili da individuare.

Ancora una volta le ricerche in corso nell'ambito della biologia dei sistemi sembrano avere buone possibilità di risolvere questo problema. Si ricordi il paragone tra modelli di reti genetiche e simulatori di volo: l'azione di una molecola farmacologica può essere simulata mediante modelli al computer di reti genetiche allo scopo di capire quali geni siano alterati dal farmaco.

Conclusioni

La ricerca post-genomica ha portato alla convergenza di discipline tecnico-scientifiche con discipline mediche e biologiche. Questa cooperazione tra ambiti disciplinari diversi permette di affrontare lo studio dei sistemi biologici integrando gli approcci sperimentali con metodi informatici, matematici e computazionali. Lo studio della patogenesi e della terapia delle malattie genetiche e multifattoriali, nonché lo sviluppo di nuovi farmaci, saranno rivoluzionati da questo approccio innovativo che permette di integrare e riassumere l'enorme quantità di dati sperimentali in modelli matematici che descrivono il funzionamento a livello molecolare della cellula. Con tali modelli si potrà simulare il comportamento dei sistemi biologici al fine di comprenderne il funzionamento e di predirne il comportamento nella patologia. In futuro, infine, la conoscenza dettagliata delle reti genetiche permetterà anche di modificarle per prevenire o curare le malattie del secolo e per mitigare gli effetti dell'invecchiamento.

Raggruppamenti di famiglie geniche in base alle caratteristiche funzionali

Carlo Sala

CNR Istituto di Neuroscienze, Dipartimento di Farmacologia, Università di Milano

La classificazione degli essere viventi è stata tradizionalmete fatta comparando l'apparenza esteriore dei vari organismi. Possiamo chiaramente affermare che un pesce ha occhi, ossa, cervello, mentre un verme ne è privo, e possiamo interpretare le similitudini in termini di evoluzione da un comune organismo ancestrale identificabile in qualche fossile. In questo modo è possibile ricostruire un albero genealogico osservando l'aspetto fisico degli organismi di tutti gli esseri viventi individuando le linee di discendenza ed i vari punti di separazione dove un organismo ancestrale di un specie diventa differente da quello di un'altra. Quando le disparità diventano grandi comunque questo metodo comincia a diventare impreciso. Del resto anche le comparazioni basate sulle differenze comportamentali e nutrizionali non ci consentono di ricostruire fedelmente l'albero genealogico delle speci viventi. Attualmente il metodo più affidabile è quello di utilizzare l'analisi delle sequenze genomiche dei vari organismi individuando e comparando le famiglie geniche comuni in ciascun organismo o specie. Il sequenziamento del DNA dei vari organismi ci ha consentito di paragonare e comparare le distanze tra i vari organismi. Un gene originariamente presente in un organismo ancestrale è evoluto creando una famiglia il cui albero genealogico ci aiuta a ricostruire la storia dell'evoluzione della vita. All'interno della famiglia due geni di due speci diverse più si assomigliano nella loro sequenza più le speci sono simili e viceversa. L'evoluzione della famiglia genica avviene tramite divisioni e mutazioni della sequenza del DNA che portano all'acquisizione o alla perdita di alcune proprietà molecolari di ciascun componente durante l'evoluzione. Attualmte tramite sofisticate analisi bioinformatiche i geni sono suddivisi in famiglie geniche consentendoci di ricostruire la nostra storia evolutiva ma anche di classificare i geni per funzione aiutandoci a comprendere il loro funzionamento.

L'importanza dell'analisi bioinformatica in genomica

Il termine "bioinformatica" è stato introdotto solo recentemente, non se ne trova traccia in letteratura prima del 1991. Il concetto corrente di "bioinformatica" deriva dalla convergenza di due rivoluzioni tecnologiche degli ultimi anni dovute alla

crescita esponenziale delle biotecnologie, gemellata a quella delle tecnologie informatiche. Questa sinergia è illustrata in modo evidente dalla constatazione che sia le dimensioni della banca dati di DNA GenBank sia la potenza di calcolo degli elaboratori elettronici sono raddoppiate circa allo stesso ritmo (ogni 18-24 mesi) per molti anni.

Del resto la possibilità di usare potenti calcolatori ha consentito un'analisi sempre più approfondita delle informazioni ricavate dal sequenziamento genomico implementando in modo considerevole le rudimentali tecniche, usate fin dagli anni 70 per studiare le sequenze geniche.

Chiaramente le tecnologie informatiche in genomica hanno assunto un valore determinate per il completamento dell'ambizioso progetto di determinare la sequenza nucleotidica completa del DNA di varie specie viventi. Questa massa di dati fornisce una materia prima privilegiata per la identificazione di sequenze potenzialmente codificanti come prima tappa verso la scoperta e l'identificazione dei geni. Un altro passo fondamentale è stata la costituzione di banche dati contenenti sequenze parziali derivate da RNA messaggeri (e quindi delle sequenze codificanti le proteine) e ricavate dall'analisi automatizzata di un gran numero di cloni ottenuti da genoteche di cDNA per diversi tessuti, in specie diverse. Queste sequenze chiamate EST per Expressed Sequence Tags sono disponibili per l'analisi di sequenza consentendoci di individuare rapidamente sequenze del genoma che siano espresse in mRNA senza dover occuparsi dello "sfondo" delle sequenze extrageniche e non codificanti, come avviene nei "progetti genoma" tradizionali. La disponibilità di queste banche dati, da una parte, e l'evoluzione degli strumenti di programmazione necessari per la loro analisi, dall'altra, hanno rapidamente portato a un nuovo approccio alla identificazione genica che consiste nell'utilizzare il dato informatico come punto di partenza per una serie di esperimenti in silico, in quanto realizzati col solo utilizzo del calcolatori.

L'operazione basilare in questo senso consiste nel confrontare le sequenze tra loro, deducendone e quantificandone la reciproca "somialianza". Questo termine indica in modo puramente descrittivo la relazione tra le due sequenze che è più significativa di quella dovuta al caso, mentre più formalmente con il termine di "omologia" si intende la nozione di una comune origine evolutiva delle sequenze. In pratica, sulla base delle relazioni di somiglianza tra sequenze può essere possibile inferire una omologia, anche se al di fuori di un modello biologico integrato la discendenza da un gene comune ancestrale rimane ipotetica. I programmi che analizzano la somiglianza tra sequenze si basano, in breve, su un punteggio assegnato in base al numero di sostituzioni, inserzioni e delezioni che occorre effettuare per convertire una sequenza nell'altra. I diversi programmi differiscono per i criteri impiegati nell'assegnazione del punteggio. Attualmente, il programma più impiegato per il confronto di sequenza è basato sull'algoritmo BLAST, un processo euristico che identifica molto rapidamente sequenze simili tra loro, avente la specifica caratteristica di assegnare anche un valore di significatività statistica alla corrispondenza trovata (<http://www.ncbi.nlm.nih.gov/BLAST/>).

Questo valore ("expect value", o valore "E") corrisponde al numero di confronti tra due sequenze con un punteggio di somiglianza uguale o superiore che si

potrebbero trovare, in quella particolare banca dati, solo per effetto del caso; quanto più è piccolo, tanto più la corrispondenza è significativa.

Ad esempio, è possibile partire dalla sequenza amminoacidica di una proteina nota e cercare quali sequenze siano simili all'interno dei frammenti nucleotidici delle EST, di cui può essere predetta automaticamente la sequenza codificante basandosi sul codice genetico (versione TBLASTN di BLAST); oppure è possibile confrontare le sequenze di DNA genomico, attualmente disponibili in banca dati, con le sequenze di mRNA (noti o ricavati dall'analisi delle EST), per una determinazione rapida della struttura genomica dei geni d'interesse. La disponibilità di banche dati per molte specie diverse rende anche possibile una ricostruzione della evoluzione molecolare delle sequenze di interesse, permettendo di distinguere tra ortologia (conservazione di un determinato gene tra diverse specie) e paralogia (presenza di un gruppo di geni omologhi all'interno di una singola specie). Infine, vi sono molte collezioni di brevi tratti (motivi o domini proteici) di sequenze amminoacidiche che indicano particolari elementi strutturali o funzionali. Le ricerche su queste raccolte a partire da nuove sequenze identificate consente di effettuare predizioni di funzione particolarmente affidabili.

Identificazione di famiglie geniche tramite analisi bioinformatica

La maggior parte delle cellule del nostro organismo contiene un genoma diploide e ogni gene è presente sotto forma di due copie alleliche situate sul cromosoma paterno e materno. Oltre a questo grado di ripetizione, circa il 40% del genoma nucleare umano è composto, sia nelle cellule aploidi sia in quelle diploidi, di gruppi di sequenze di DNA non allelico ma strettamente correlate (famiglie di sequenze di DNA, o DNA ripetitivo). All'interno della considerevole varietà di sequenze di DNA ripetute, vi sono anche famiglie di sequenze di DNA i cui singoli membri comprendono geni funzionali (famiglie multigeniche). La definizione operativa di una famiglia di sequenze di DNA è il livello relativamente alto di somiglianza di sequenza tra i membri della famiglia, a livello dell'intera sequenza o di sue regioni localizzate.

I membri di una famiglia genica possono essere identificati mediante:

- 1) ibridizzazione del DNA e clonaggio, impiegando un frammento genico come sonda per lo screening in laboratorio di librerie geniche;
- 2) clonaggio mediante amplificazione con la reazione a catena della polimerasi (PCR), utilizzando degli oligonucleotidi primers degenerati che si leghino alle regioni conservate tra i membri della famiglia;
- 3) analisi di sequenza e lo screening elettronico delle banche dati, che permette di effettuare il calcolo diretto del grado di relazione tra i geni.

Il fatto che due membri di una famiglia di sequenze di DNA mostrino un alto grado di somiglianza è indicativo di una comune origine evolutiva, ed è normalmente in relazione con la conservazione di una funzione.

Una grande percentuale di geni umani attivamente espressi sono membri di famiglie di sequenze di DNA. Il catalogo PFam, mantenuto presso il Sanger Center

(Hinxton, Cambridge, UK), classificava 2.478 famiglie geniche nella versione del 2001, attualmente ne cataloga più di 9.000 per una copertura del genoma degli organismi presi in considerazione intorno al 75% (<http://www.sanger.ac.uk/>).

Possiamo distinguere diversi tipi di famiglie geniche. Nelle famiglie geniche più classiche, i membri mostrano un alto grado di omologia di sequenza lungo la maggior parte dell'estensione dei geni o, almeno, della loro sequenza codificante. Questa caratteristica identifica in pratica una correlazione evolutiva, e funzionale, di queste sequenze. Un esempio è costituito dalle famiglie geniche degli istoni o di molte altre proteine con funzione molto simile. In alcune famiglie geniche, invece, l'omologia è particolarmente pronunciata all'interno di regioni fortemente conservate dei geni, mentre la somiglianza tra le porzioni rimanenti della sequenza codificante può essere molto piccola. Spesso queste famiglie codificano per fattori di trascrizione che giocano un ruolo importante nelle prime fasi dello sviluppo, e la sequenza conservata codifica per un dominio proteico richiesto per il legame selettivo al DNA di specifici geni bersaglio (come ad esempio, il dominio Homeobox). Infine, esistono anche famiglie geniche i cui membri non sono correlati in modo ovvio a livello della sequenza di DNA, ma codificano per prodotti caratterizzati da una funzione generale comune e dalla presenza di brevi o relativamente brevi tratti di sequenza conservati; ad esempio, il dominio PDZ (una specifica sequenza amminoacidica di circa 90-100 aminoacidi) si ritrova in geni differenti, i cui prodotti sembrano tutti funzionare come proteine cosiddette scaffold. I membri delle famiglie geniche possono essere occasionalmente localizzati in prossimità l'uno dell'altro in specifiche regioni subcromosomiche, come i geni del complesso maggiore di istocompatibilità di classe I (HLA), ma più spesso si trovano dispersi nel genoma.

Nuovi geni si generano da geni preesistenti attraverso diversi processi tra i quali:

- 1) mutazioni intrageniche, tramite mutazioni nella sequenza di DNA;
- 2) duplicazione genica, cioè la duplicazione di un gene esistente per creare un nuovo gene simile all'interno della stessa cellula;
- 3) scambio di frammenti, quando due geni vengono frammentati e ricongiunti per creare un nuovo gene ibrido;
- 4) trasferimento orizzontale di materiale genetico, quando un gene o parte di esso viene trasferito tra una specie e l'altra.

In tutti questi casi l'analisi bioinformatica ci consente di identificare le tracce lasciate da ciascun processo e ricostruire la storia della famiglia genica.

Come detto precedentemente confrontando il materiale genetico di organismi rappresentativi dei tre regni viventi (flora, fauna e organismi unicellulari) è possibile cercare per la presenza di omologie. Ovviamente l'evoluzione ha fatto sì che alcuni geni si siano estinti in alcune specie mentre altri siano rimasti conservati e trasmessi fino alle specie più evolute. Tra le più di 9.000 famiglie geniche identificate in realtà solo circa 76 sono da considerare veramente ubiquitarie e vengono considerate come famiglie geniche universali.

Nonostante l'analisi delle famiglie geniche viene effettuata usando la sequenza di DNA, recentemente alcune banche dati hanno affrontato il problema di raggrup-

pare le sequenze di proteine in famiglie. I vari approcci differiscono nel loro grado di automazione, nella loro completezza, nella loro focalizzazione sulla sequenza completa delle proteine o sui domini proteici. In effetti, le relazioni tra i geni e tra le famiglie geniche sono così complesse che è difficile usare un semplice schema gerarchico per rendere i dati facilmente comprensibili, a causa della frequente composizione modulare delle proteine.

Tra i vari strumenti specificamente pensati per la ricostruzione delle famiglie geniche attraverso l'analisi della sequenza amminoacidica, di particolare rilievo sono anche dati che si possono trovare on-line, tra cui la più completa è PROSITE (<http://expasy.org/prosite/>). È possibile utilizzare queste banche dati per identificare domini proteici presenti nella sequenza e quindi risalire alla funzione della proteina e alla eventuale famiglia di appartenenza.

L'analisi delle proteine viene eseguita con programmi come PSI-BLAST la cui ricerca iterativa si basa sull'analisi del profilo. Innanzitutto si effettua una ricerca di somiglianza su una banca dati a partire da una singola sequenza, usando BLAST. Le sequenze significativamente simili sono allineate alla sequenza di interrogazione, e viene costruito un profilo, un sistema di punteggio posizione-specifico derivato dalla frequenza con cui un dato residuo amminoacidico si osserva in una colonna dell'allineamento. Poiché le famiglie di sequenze conservano preferenzialmente determinati residui e motivi critici, questa informazione può permettere di svolgere ricerche più sensibili, in successioni ripetute (iterazioni). Altri programmi basati su HMM impiegano invece un particolare metodo statistico per il riconoscimento della configurazione di una serie di valori (la sequenza) che può essere usato per rappresentare l'allineamento di sequenze multiple o di segmenti di sequenza, con l'obiettivo di identificare la conservazione di ripetizioni conservate o di singoli residui. L'interesse principale dello studio delle famiglie geniche umane consiste operativamente nell'ottenimento di indicazioni sulle probabili funzioni di un gene che sia simile ad un gene già caratterizzato funzionalmente, eventualmente permettendo il recupero di informazioni ottenute in organismi modello di specie diverse. Del resto nonostante l'avanzamento dei grandi progetti di sequenziamento del DNA totale di specie diverse, a tutt'oggi alcuni nuovi geni identificati non sono stati assegnati a famiglie geniche.

Il confronto tra uomo e scimpanzè

Quanto è simile l'uomo allo scimpanzè? In un recente studio (Demuth et al. PLoS ONE. 2006 Dec 20;1:e85) sono stati analizzati i genomi di uomini e scimpanzè, stabilendo che la differenza genetica tra le due specie sarebbe di circa il 6%. Quindi uomini e scimpanzè hanno un patrimonio genetico comune di circa il 94% e presentano geni esclusivi di ciascuna specie per solo il 6% del genoma.

I ricercatori hanno focalizzato la loro attenzione sulle "famiglie di geni" dell'uomo (*Homo sapiens sapiens*), dello scimpanzè (*Pan troglodytes*), del topo (*Mus musculus*) e del cane (*Canis lupus familiaris*). Sono stati esaminati 110.000 geni appartenenti a 9.990 famiglie geniche delle quattro specie in esame. Lo studio si è poi concentrato sulle famiglie geniche comuni a uomo e scimpanzè, nel tenta-

tivo di ricostruire le modificazioni geniche che hanno portato all'evoluzione dell'uomo a partire dal progenitore comune alle altre scimmie antropomorfe.

L'analisi sembrerebbe dimostrare che molte delle differenze genetiche tra le due specie sono dovute alla duplicazione o alla perdita di alcuni geni, soprattutto quelli che influenzano le funzioni cerebrali, piuttosto che alla mutazione di essi. Di particolare interesse è l'osservazione, tramite un metodo statistico, che la nostra specie abbia guadagnato 689 nuovi geni, attraverso la duplicazione di geni già esistenti, e perso altri 86 dal momento della divergenza tra noi e gli scimpanzè dal nostro progenitore comune. Includendo nell'analisi i 729 geni che i nostri "cugini" sembrano aver perduto, la differenza totale tra le due specie ammonta a circa il 6% del patrimonio genetico.

Questi risultati non confutano la comune convinzione che il DNA delle due specie differiscano solo per 1,5% della sequenza nucleotidica, ma evidenziano che esistono vari modi per stabilire la distanza filogenetica tra specie diverse. Non si esclude che in futuro vengano creati metodi comparativi in grado di rendere conto sia delle differenze esistenti tra le sequenze di basi, sia, a livello più macroscopico, delle differenze strutturali dei genomi.

Significato e classificazione dei polimorfismi genici

Roberto Giorda

Laboratorio di Bioinformatica, IRCCS Eugenio Medea, Bosisio Parini (Lecco)

Cenni storici

Le basi della genetica

Il fatto che le caratteristiche fisiche e comportamentali umane, come quelle di animali e piante domestiche, avessero una componente ereditabile era noto fin dall'antichità, ma solo duecento anni fa Maupertuis descrisse il passaggio della polidattilia da una generazione alla generazione seguente. Se un genitore esprimeva l'anomalia, circa la metà dei figli sarebbe stato affetto. Oggi definiremmo questa modalità di ereditarietà "*autosomica dominante*".

Allo stesso modo nel primo '800 fu descritto come l'emofilia e la cecità ai colori (acromatopsia) erano trasmesse dal ramo materno, in cui metà degli zii risultavano affetti, e le sorelle sane del figlio affetto a volte trasmettevano la malattia ai loro figli maschi, mostrando quindi una ereditarietà "*legata al sesso*".

John Adams a sua volta osservò che certi tratti o malattie potevano emergere in famiglie con genitori apparentemente normali. Sorprendentemente, il fenomeno avveniva più frequentemente in famiglie in cui i genitori erano imparentati (consanguineità). Questa descrizione non era altro che l'ereditarietà "*recessiva*" decritta quasi un secolo più tardi da Mendel. Il matematico e monaco Gregor Mendel trasformò la genetica in una disciplina quantitativa postulando l'esistenza di fattori ereditabili che determinano le caratteristiche fisiche delle generazioni successive. Questi fattori erano presenti in due copie in ogni genitore, e in una copia nei gameti. Mendel descrisse anche come ogni coppia di fattori fosse ereditata indipendentemente, e che certi fattori erano "*dominanti*" su altri. Il primo esempio di variazione genetica ben caratterizzata, il gruppo sanguigno ABO (Landsteiner), è stato descritto all'inizio del ventesimo secolo.

Sutton, nel 1904, attribuì la localizzazione dei fattori ereditabili ai "*cromosomi*", organelli colorati nucleari. Garrod predisce che certe malattie umane erano dovute alla perdita di "*fattori mendeliani*", che furono chiamati "*geni*" da Johannsen nel 1909. Dall'anno seguente Morgan e i suoi allievi iniziarono ad usare la *Drosophila melanogaster* come modello per studiare l'ereditarietà dei caratteri genetici.

Linkage

Hogben negli anni '30 suggerì che geni situati a breve distanza sullo stesso cromosoma potevano essere usati per localizzare la posizione (“locus”) di altri geni sul cromosoma, ciò che chiamiamo “mappatura” mediante “linkage genetico”. Dal momento che le variazioni di un gene (gli “alleli”) erano frequentemente ereditate insieme a una specifica variazione in un altro gene, i due geni dovevano essere fisicamente vicini, o “linked”. I topi per esempio rivelarono 20 gruppi di linkage, corrispondenti al numero dei loro cromosomi. Poco tempo dopo Haldane calcolò la distanza genetica tra emofilia e acromatopsia sul cromosoma X. Al 1949 (Pauling) risale l'identificazione molecolare della prima mutazione, una sostituzione aminoacidica nell'emoglobina che causa anemia falciforme. L'assegnazione di geni a vari cromosomi umani rimaneva difficile, tanto che solo nel 1951 Mohr notava l'associazione del fattore sanguigno “Lutheran” con “secretor”.

Il DNA

Nel 1953 Watson e Crick descrissero la struttura molecolare del materiale genetico, la “doppia elica”. L'accoppiamento delle basi del DNA è la base della biologia molecolare di oggi, perché permette di riconoscere qualsiasi porzione della molecola utilizzando la sequenza complementare marcata con una molecola fluorescente o radioattiva (“sonda”) per riconoscerla e legarsi specificamente a lei (“ibridare”), anche in un cromosoma intero.

Negli anni seguenti Tijo e Levan dimostrarono che l'uomo ha 46 (e non 48) cromosomi; Lejeune notò un cromosoma addizionale nei soggetti affetti da sindrome di Down; Mary Lyon descrisse l'inattivazione di uno dei cromosomi X nelle femmine; McKusick pubblicò la prima edizione di “Mendelian Inheritance in Man (MIM)”, che conteneva 1500 caratteristiche ereditate, o “fenotipi mendeliani”. Solo nel 1968 Donahue osservò l'associazione tra una variazione osservabile al microscopio del cromosoma 1 e un allele del gruppo sanguigno “Duffy”, che fu così assegnato al cromosoma 1.

Una serie di miglioramenti tecnologici permisero dalla metà degli anni '70 l'accelerazione esponenziale della ricerca genetica. Weiss e Green fusero cellule umane con una linea di Hamster, ottenendo “ibridi somatici” stabili in cui rimaneva a volte solo 1 cromosoma umano, che poteva così venire esaminato singolarmente; Caspersson disegnò un metodo per ottenere un migliore “bandeggio cromosomico” e riconoscere più facilmente i diversi cromosomi e le loro anomalie; Sanger e Maxam e Gilbert nel 1975 descrissero metodi diversi per leggere la sequenza di frammenti di DNA (“sequenziamento”); Kan e colleghi nel 1976 usarono l'ibridazione molecolare per identificare la perdita di geni della globina nella anemia Mediterranea. Due anni dopo Kan e Dozy descrissero una variazione della sequenza del DNA vicino al gene dell'anemia falciforme che poteva essere usata in una data famiglia come marcatore di linkage con il gene. Questa variazione, denominata “polimorfismo” fu il primo esempio dell'uso di differenze molecolari comuni tra persone diverse per la diagnosi e in seguito la mappatura di un carattere nell'intero genoma.

Polimorfismi e mappatura su larga scala

Nel 1979 Solomon e Bodmer suggerirono che altre variazioni nella sequenza del genoma potevano essere usate come marcatori per altre malattie e caratteri genetici. Infatti lo stesso anno White e Wyman isolarono la loro prima (ma non ultima) sonda anonima, che mostrava molte variazioni tra individui diversi e fu perciò chiamata “*altamente polimorfica*”. Negli anni seguenti un numero di laboratori iniziò la ricerca di altri marcatori polimorfici, che per la metodica usata per evidenziarli furono chiamati polimorfismi di lunghezza di frammenti di restrizione (Restriction Fragment Length Polymorphisms, RFLP). Nel frattempo a Parigi Jean Dausset fondava il Centre d’Etude du Polymorphisme Humain (CEPH) che inizialmente preparò campioni di DNA da 40 grandi famiglie multigenerazionali e li distribuì a chiunque volesse mappare loci genetici umani.

Così negli anni seguenti vennero mappati i loci per la malattia di Huntington (Gusella 1983), la malattia del rene policistico adulto (Reeders 1985), la distrofia muscolare (Kunkel 1987) e la fibrosi cistica (Collins; Tsui 1989). Intanto nel 1983 Kary Mullis inventava la reazione polimerasica a catena (Polymerase Chain Reaction, PCR) che permette di ottenere facilmente e velocemente miliardi di copie di qualsiasi frammento di DNA.

Categorie di variazioni strutturali del genoma

Eteromorfismi citogenetici

Gli eteromorfismi (sinonimo di polimorfismi o varianti normali) di regioni cromosomiche sono stati osservati fin dai primi studi sulla cromatina e caratterizzati con varie tecniche di bandeggio cromosomico. Sono considerati clinicamente non significativi, ma ancora oggi non è facile definire con certezza se una nuova variazione cromosomica è veramente un eteromorfismo.

Single nucleotide polymorphisms (SNPs)

La prima categoria di polimorfismi umani ad essere riconosciuta è stata la sostituzione di singoli nucleotidi (Single Nucleotide Polymorphism, SNP), che analogamente alle mutazioni avvengono in maniera essenzialmente casuale durante la replicazione del DNA. Nella popolazione umana ci sono circa 10 milioni di SNP in cui l’allele meno frequente ha una frequenza superiore all’1%. Gli SNP sono meno variabili dei microsatelliti ma anche meno soggetti a mutazioni ripetute, e per questo preferibili in studi di mappatura.

Da molto tempo è noto che gli SNP sono associati con la variazione fenotipica umana sia per un loro effetto diretto che per la loro associazione con altre variazioni causali con cui sono in linkage disequilibrium. L’analisi di SNP, fino a pochi anni fa limitata a uno o pochi polimorfismi alla volta, può oggi essere condotta su centinaia di migliaia in un singolo esperimento, anche grazie alla creazione di mappe contenenti la localizzazione di milioni di possibili polimorfismi nelle regioni codificanti (cSNP) e non codificanti del genoma umano. Le caratteristiche degli SNP li rendono adatti a studi di associazione per la ricerca delle basi genetiche di malattie complesse.

Tab. 1 - Variazioni del genoma umano.

Tipo di variazione	Definizione	Frequenza nel genoma umano
Eteromorfismi citogenetici	Variazioni cromosomiche visibili al microscopio	Rari
SNP	Variazione di una singola coppia di basi trovati in >1% dei cromosomi in una popolazione	Circa 107 SNP nella popolazione umana (0,3% del genoma)
Microsatelliti	Sequenze contenenti ripetizioni variabili di sequenze di 1-6 basi	>106 microsatelliti nel genoma umano (3% del genoma)
Minisatelliti e VNTR	Sequenze polimorfiche contenenti 20-50 ripetizioni di sequenze di 6-100 basi	Circa 150,000 minisatelliti, di cui 20% polimorfici
Inserzioni/Delezioni (InDel)	Delezioni o inserzioni di un segmento di DNA di dimensioni 1-1000 basi	Circa 106 indel >1 base nel genoma umano
CNV	Delezioni o Inserzioni di segmenti di DNA più grandi di 1000 basi	Circa 1.500 CNV >50 Kb nel genoma (12% del genoma). La frequenza dei CNV tra 1-50 Kb è ignota
Inversioni	Riarrangiamento che causa la presenza di un segmento di DNA in orientazione opposta	La frequenza delle inversioni microscopiche è 0,12-0,7% (pericentriche e 0,1-0,5% (paracentriche). La frequenza delle inversioni submicroscopiche non è nota

Microsatelliti e minisatelliti

Il termine “DNA satellite” ha origine da osservazioni negli anni '60 di una frazione di DNA che mostrava una densità diversa, visibile come un “picco satellite” nelle centrifugazioni per gradienti di densità. Questo DNA venne poi identificato come grandi sequenze ripetute centromeriche. In seguito, quando vennero identificate sequenze ripetute più corte, esse furono designate micro- e minisatelliti (STR e VNTR), regioni contenenti un numero variabile di ripetizioni di sequenze di 1-4 o 15-1.000 nucleotidi. I microsatelliti sono estremamente sovrarappresentati nel genoma umano (oltre 1 milione di copie) giungendo al 3% della sequenza totale. Sono tra le sequenze di DNA più variabili nel genoma, con tassi di mutazione tra 1:100 e 1:10.000. In contrasto con gli SNP, questi polimorfismi derivano in gran parte da variazioni di lunghezza piuttosto che nella sequenza. Possono presentare un numero di forme diverse (alleli) anche piuttosto elevato, e dopo l'invenzione della PCR sono stati usati con successo per localizzare la posizione di mutazioni di geni ignoti in studi familiari di patologie a eredità mendeliana. La maggioranza dei microsatelliti risiede nelle regioni intergeniche e negli introni e si assume che la loro evoluzione sia neutrale. Una classe speciale di sequenze ripetute trinucleotidiche nel DNA codificante è associata a una serie di patologie umane. Questi loci sono soggetti ad espansione della sequenza ripetuta con un meccanismo che si pensa sia diverso da quello della maggior parte dei microsatelliti nel genoma.

Alcune classi di microsattelliti, in particolare quelli mononucleotidici (A)_n e (T)_n, e altri tipi di microsattelliti ricchi in A sono associati a elementi ripetuti come SINE e LINE; Per esempio, le sequenze Alu umane spesso hanno microsattelliti nella loro estremità 3', che derivano dall'inserzione di code di poli (A) durante la loro retroscrittura.

Polimorfismi Alu e LINE

Il termine "elemento ripetuto" descrive varie sequenze di DNA presenti in molteplici copie in un genoma. Tra essi, gli elementi "interspersi" sono dispersi in modo casuale e sono classificati sulla base delle loro dimensioni in elementi più corti di 500 basi (Short Interspersed elements, SINE) ed elementi più lunghi (Long Interspersed elements, LINE). Gli elementi Alu (così chiamati perché possiedono un sito di restrizione per l'omonimo enzima) sono stati inizialmente identificati come una componente nelle curve di rinaturazione del DNA. Sono lunghi circa 300 basi, presenti in oltre 1,000,000 copie nel genoma umano, di cui sono i SINE più abbondanti, e comprendono più del 10% della massa del genoma. La loro origine ed amplificazione è un fenomeno evolutivamente recente che ha coinciso con la radiazione dei primati negli ultimi 65 milioni di anni. Si pensa che l'amplificazione degli Alu avvenga mediante la trascrizione inversa di un trascritto derivato da un Alu. Dal momento che gli Alu non codificano per i fattori necessari alla propria mobilizzazione, si pensa che sfruttino quelli codificati dagli elementi LINE. La maggioranza delle sequenze Alu non sembra oggi in grado di retrotrasporre, e il tasso attuale di amplificazione è di circa una inserzione ogni 200 nascite. Gli elementi Alu sono classificabili in famiglie di età genetica diversa. La maggioranza degli Alu si è integrata nel genoma prima della divergenza tra uomo e scimmie africane, ma alcune famiglie comprendenti circa 5.000 elementi sono presenti solo nella linea umana.

Di questi, circa 1.200 (il 25%) si sono integrati dopo la radiazione umana fuori dall'Africa, perciò sono presenti solo in alcune popolazioni e sono perciò polimorfici rispetto alla loro presenza o assenza. I polimorfismi Alu, come quelli LINE che però sono molto più rari, sono stati utilizzati per rispondere a varie domande sull'origine della specie e delle popolazioni umane grazie alla loro stabilità. La maggioranza delle nuove inserzioni Alu non ha effetti negativi sul genoma e rappresenta essenzialmente una mutazione neutra; esistono però la possibilità di vari effetti negativi dovuti a un'inserzione e ci sono esempi di malattie genetiche causate dall'inserzione di un elemento Alu.

Variazione nel numero di copie

Come abbiamo visto, le prime differenze osservate nella nostra composizione genetica sono state rare differenze nella quantità e struttura dei cromosomi, tra cui riarrangiamenti, aneuploidie, eteromorfismi abbastanza grandi da essere visti con il microscopio. Queste varianti, che sono grandi più di 3 Megabasi (Mb), sono definite "variazioni strutturali microscopiche".

Il sequenziamento del genoma umano ha rivelato una grande quantità di piccole variazioni, tutte più corte di 1 Kilobase (Kb), riguardanti il numero di copie di un segmento di DNA, quali inserzioni o delezioni, che vengono genericamente chiamate “indels”.

Le nuove tecnologie per l'analisi del genoma, tra cui la ibridazione comparativa su array (Array Comparative Genome Hybridisation, Array-CGH) e l'analisi comparativa delle sequenze di DNA, hanno permesso di rivelare variazioni del DNA che coinvolgono segmenti più piccoli di quelli visibili microscopicamente ma più grandi di quelli osservabili mediante l'analisi di sequenza convenzionale. Queste variazioni, che hanno dimensioni tra 1 Kb e 3 Mb, sono definite “variazioni strutturali submicroscopiche”. Di queste variazioni in numero di copie (Copy-Number Variations, CNV) ne sono state descritte finora varie centinaia, oltre a una serie di polimorfismi di inversione. Quando la loro frequenza è >1% vengono chiamati Copy-Number Polymorphisms, CNP. Si stima che nel genoma umano ci siano circa un milione di indels maggiori di 1 base, mentre finora sono stati rilevati circa 1500 CNV più grandi di 50 Kb, che coprono il 12% del genoma umano. La differenza media tra due genomi, che a livello di SNP e STR era circa 2 Mb (corrispondente allo 0,1% del genoma) diventa così 20 Mb (1%). Inoltre alcune di queste varianti contengono interi geni con le loro regioni regolatrici, potrebbero avere un ruolo nelle variazioni fenotipiche ed essere associate a fenomeni di rilevanza clinica. L'impatto relativo di SNP e CNV sull'espressione genica è stato valutato recentemente verificando l'associazione tra i livelli di espressione di circa 15000 geni e SNP/CNV in una serie di soggetti di tre diverse etnie. L'effetto degli SNP sulla variazione nell'espressione genica era circa quattro volte maggiore di quello dei CNV, ma se teniamo presente che solo una piccola parte dei CNV, quelli più grandi di 100 Kb, è stata testata possiamo ipotizzare che l'effetto complessivo dei CNV sia almeno uguale a quello degli SNP

Inversioni e duplicazioni segmentali

Un'ultima, almeno per ora, categoria di polimorfismi riguarda l'orientazione di segmenti del genoma di varie dimensioni, da poche migliaia a vari milioni di basi, che possono esistere in due “alleli” con orientazione opposta. Questa categoria di polimorfismi, mentre non modifica la sequenza del genoma, può influenzare profondamente il processo di appaiamento e ricombinazione meiotica e di conseguenza l'evoluzione del nostro genoma. Le inversioni sono spesso circondate da “duplicazioni segmentali”, segmenti di DNA che sono presenti in due o più copie nel genoma aploide, in cui le copie hanno una identità di sequenza maggiore di 90%; esse sono spesso variabili in numero di copie e perciò possono essere considerate anche CNV. Inversioni e duplicazioni segmentali sono coinvolte in una serie di patologie con base genetica.

Bibliografia essenziale

1. Watson, JD, Crick, FH. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. Nature. 1953 Apr 25; 171(4356): 737-8.

2. Weiss MC, Green H. Human-mouse hybrid cell lines containing partial complements of human chromosomes and functioning human genes. *Proc Natl Acad Sci USA*. 1967 Sep; 58(3): 1104-11.
3. Gilbert W, Maxam A. The nucleotide sequence of the lac operator. *Proc Natl Acad Sci U S A*. 1973 Dec; 70(12): 3581-4.
4. Sanger F, Coulson AR. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *J Mol Biol*. 1975 May 25; 94(3): 441-8.
5. Kan YW, Holland JP, Dozy AM, Charache S, Kazazian HH. Deletion of the beta-globin structure gene in hereditary persistence of foetal haemoglobin. *Nature*. 1975 Nov 13; 258(5531): 162-3.
6. Kan YW, Golbus MS, Dozy AM. Prenatal diagnosis of alpha-thalassemia. Clinical application of molecular hybridization. *N Engl J Med*. 1976 Nov 18; 295(21): 1165-7.
7. Solomon E, Bodmer WF. Evolution of sickle variant gene. *Lancet*. 1979 Apr 28; 1(8122): 923.
8. Reeders ST, Breuning MH, Davies KE, Nicholls RD, Jarman AP, Higgs DR, Pearson PL, Weatherall DJ. A highly polymorphic DNA marker linked to adult polycystic kidney disease on chromosome 16. *Nature*. 1985 Oct 10-16; 317(6037): 542-4.
9. Monaco AP, Bertelson CJ, Colletti-Feener C, Kunkel LM. Localization and cloning of Xp21 deletion breakpoints involved in muscular dystrophy. *Hum Genet*. 1987 Mar; 75(3): 221-7.
10. Kerem B, Rommens JM, Buchanan JA, Markiewicz D, Cox TK, Chakravarti A, Buchwald M, Tsui LC. Identification of the cystic fibrosis gene: genetic analysis. *Science*. 1989 Sep 8; 245(4922): 1073-80.
11. Saiki RK, Scharf S, Faloona F, Mullis KB, Horn GT, Erlich HA, Arnheim N. Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science*. 1985 Dec 20; 230(4732): 1350-4.

SNP (Single Nucleotide Polimorfism): scoperta e applicazioni

Manuela Sironi

Laboratorio di Bioinformatica, IRCCS Eugenio Medea, Bosisio Parini (Lecco)

Gli SNP sono polimorfismi (ovvero presenza di più alleli all'interno di una specie o popolazione) che coinvolgono un solo nucleotide; essi costituiscono una delle maggiori fonti di variazione genetica nelle popolazioni umane. È possibile assegnare a ciascuno SNP una frequenza dell'allele meno rappresentato (Minor allele frequency, MAF) e, generalmente, variazioni nucleotidiche con MAF >1% sono classificate come SNPs.

I dati sulla frequenza di SNPs in diverse popolazioni (<http://www.hapmap.org>) indicano che la maggior parte degli SNPs è rara (il 46% ha MAF <0,05), ma, allo stesso tempo, la maggior parte degli SNP eterozigoti in un individuo è rappresentata da varianti comuni (the International HapMap Consortium, 2005).

Il termine *aplotipo* indica la combinazione di alleli polimorfici lungo un cromosoma (o una porzione cromosomica, Figura 1); gli *aplotipi* che si osservano sono molti meno di quelli che ci si potrebbe aspettare dati il numero e le frequenze alleliche dei singoli SNP. Questo è dovuto all'esistenza del *linkage disequilibrium* (LD). Con LD si intende la associazione non casuale degli alleli di due o più loci polimorfici (Figura 1) e la presenza di LD dipende dalla discendenza dei cromosomi umani attuali da un'unica popolazione ancestrale (africana). L'interruzione dei blocchi di LD è dovuta alla presenza degli *hotspot* di ricombinazione, per cui la genotipizzazione estensiva degli SNPs consente di costruire mappe fini delle frequenze di ricombinazione nel genoma umano. Solitamente l'LD viene misurato attraverso il calcolo di D' o r^2 (Figura 1). Nel caso di due SNP biallelici con alleli a/A e b/B e nel caso di non LD, avremo $p(AB)=p(A)p(B)$; se esiste LD, $p(AB)=p(A)p(B)+D$, ovvero $D=p(AB)-p(A)p(B)$. Per semplicità si utilizza in realtà $D'=D/D_{max}$ che varia tra 0 e 1; $D'=1$ indica assenza di ricombinazione. r^2 è invece calcolato come $D/p(A)p(a)p(B)p(b)$; questo parametro varia quindi sia in funzione della ricombinazione sia in ragione dell'ordine di comparsa dei 2 SNPs nelle popolazioni ancestrali.

La struttura *aplotipica* a blocchi (presenza di LD) del genoma umano deve essere tenuta in considerazione quando si intraprendano studi di associazione. Gli studi di associazione rappresentano il tentativo di associare un allele marcatore ad una malattia. Essi sono particolarmente utili nel caso di malattie complesse, in cui il fenotipo patologico è spesso dovuto a più alleli di suscettibilità (oltre che a pos-

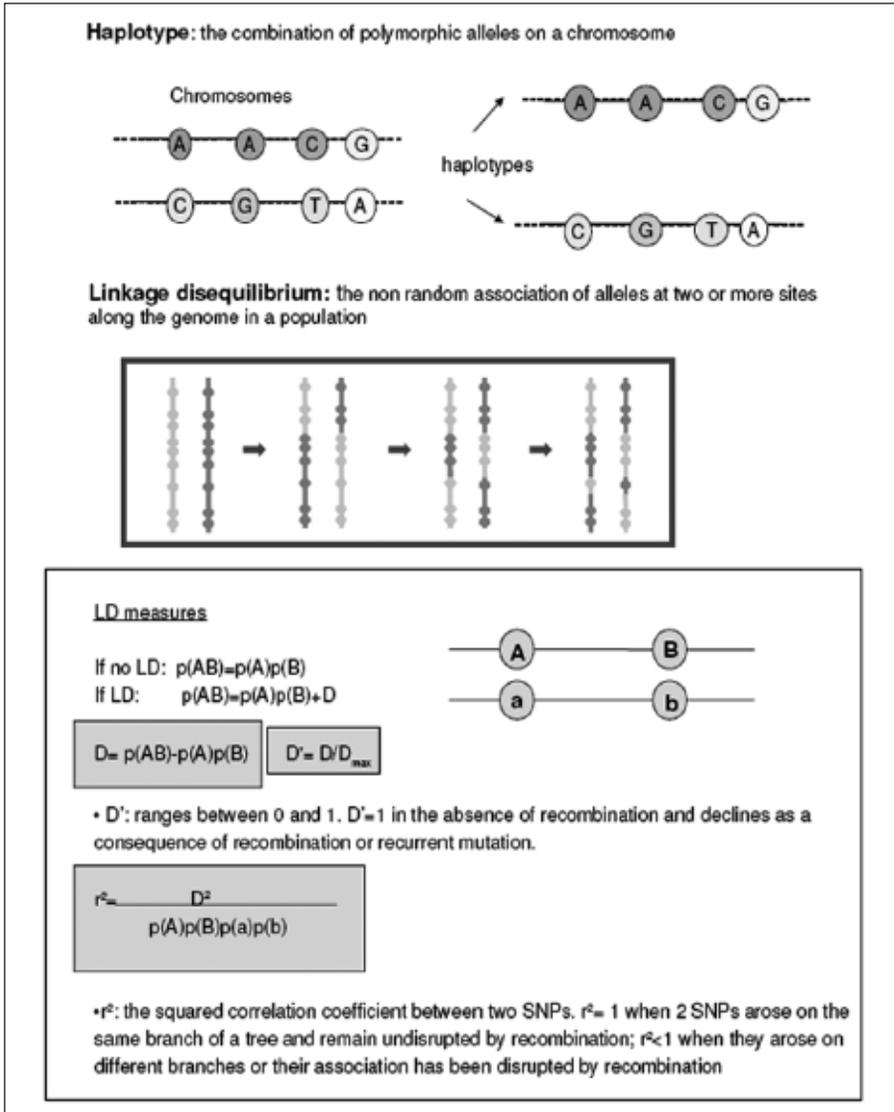


Fig. 1

sibili fattori ambientali). Gli SNP sono estremamente utili negli studi di associazione in quanto, nonostante siano meno informativi (la maggior parte è costituita da SNP biallelici) dei microsatelliti, hanno elevata densità e possono essere genotipizzati su ampia scala con apparecchiature automatiche sfruttando diversi approcci (Figura 2).

In uno studio di associazione sull'intero genoma, la presenza di LD rende ridondante la tipizzazione di tutti gli SNP: si può identificare una sotto-popolazione di SNP (detti Tag SNP) da tipizzare perdendo nessuna o poca informazione (in

dipendenza del valore LD) riguardo alla struttura aploipica dell'intero genoma. Sebbene l'uso in studi di associazione sia una delle applicazioni più diffuse e rilevanti, la genotipizzazione degli SNP è applicata in molti campi delle scienze biomediche. Ad esempio, la genotipizzazione genome-wide di SNP su supporto solido consente di identificare delezioni, amplificazioni e tratti estesi di omozigotità in regioni genomiche con una risoluzione approssimativamente nell'ordine delle 50 kb. Le applicazioni di una tale tecnica sono evidenti; vale la pena ricordare, ad

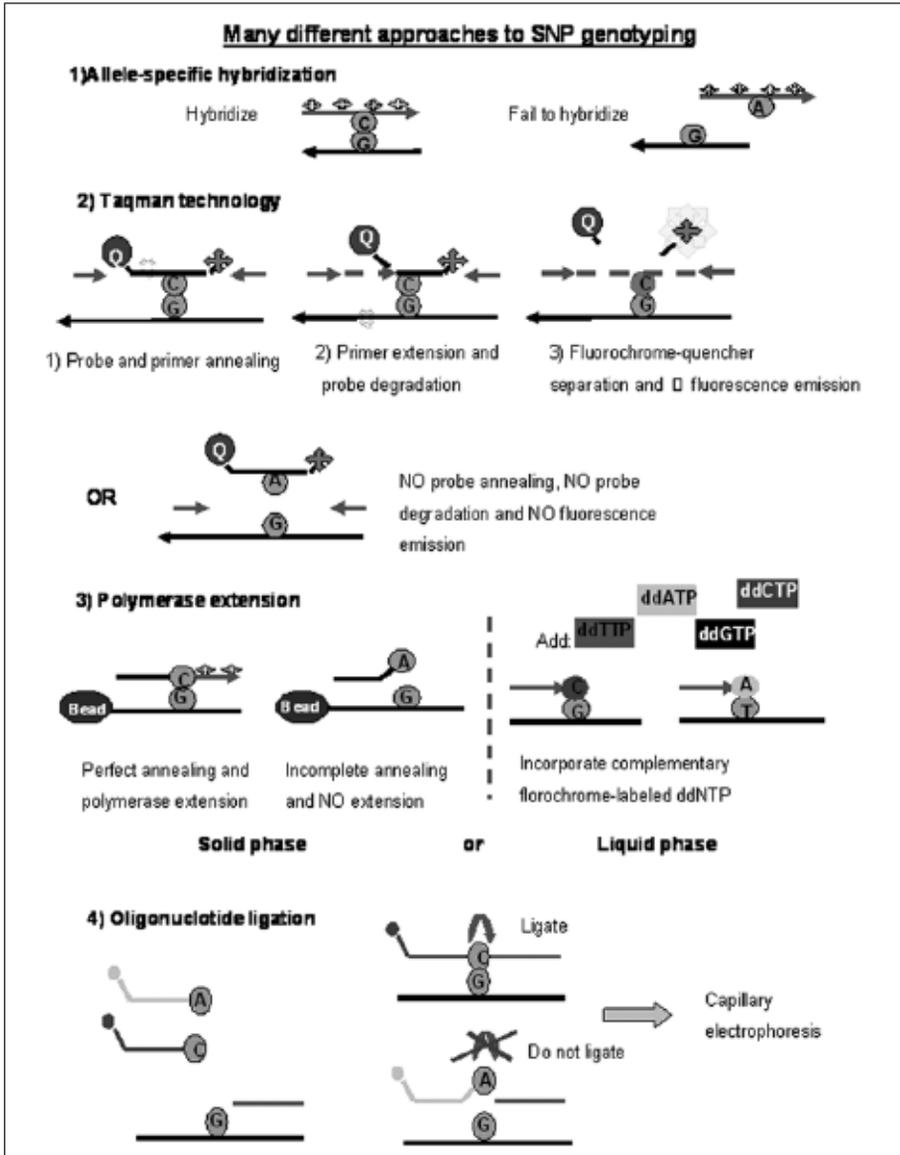


Fig. 2

esempio, che la localizzazione di tratti estesi di omozigosità in famiglie con consanguineità parentale (o in popolazioni isolate) può consentire la mappatura di mutazioni recessive. Inoltre, utilizzando un approccio di questo tipo è possibile valutare la presenza e la frequenza di anomalie strutturali in tessuti tumorali o linee cellulari.

Un altro campo in cui gli SNP hanno avuto una grande importanza è lo studio dell'evoluzione a livello molecolare (per una review, si veda Nielsen, 2005). Infatti, l'analisi della frequenza di occorrenza di SNP, così come lo studio delle loro frequenze alleliche e dell'estensione dei blocchi aploipici, consente di identificare (o verificare) la presenza di regioni genomiche (o siti funzionali) sottoposte a selezione negativa o positiva. Alcuni esempi in questi campo riguardano l'uso di SNP per verificare la selezione attiva su enhancer di splicing, sequenze conservate non codificanti e siti di legame per microRNA. In tali studi ha particolare rilevanza l'analisi delle frequenze alleliche (come raffigurato in figura 3).

Ovviamente gli SNP non rappresentano solo degli utili marcatori, ma una parte di essi ha effetto sulla variabilità fenotipica della nostra specie. Considerando un gene e le sue regioni funzionali, può non essere difficile immaginare come SNP non sinonimi nelle regioni codificanti (SNP che determinano la variazione di un residuo aminoacidico) possano influire sull'attività della proteina codificata e, quindi sul fenotipo. Tuttavia, studi recenti dimostrano come SNP localizzati nelle regioni non codificanti o SNP codificanti sinonimi possano indurre alterazioni funzionali agendo sui diversi processi che a partire dalla sequenza genomica sono responsabili di un prodotto proteico maturo. Alcuni esempi in questo senso comprendono alterazioni di enhancer di splicing esonici o di siti di splicing, introduzione di codoni rari che alterino il folding del polipeptide nascente e creazione di nuovi promotori.

In particolare, è stato recentemente identificato (Kimchi-Sarfaty et al. 2007) un aplotipo del gene MDR1 in cui la presenza di SNP sinonimi induce un'alterazione nell'affinità del prodotto genico (una pompa ATPasi coinvolta nel metabolismo dei farmaci e nella resistenza ai chemoterapici delle cellule tumorali) verso alcuni inibitori. Gli autori hanno dimostrato come gli SNP non alterino la sequenza aminoacidica, ma il folding della proteina. Essi introducono infatti alcuni codoni rari; la presenza di codoni rari in cluster può modificare la velocità di traduzione e, quest'ultima, a sua volta è in grado di influenzare il folding. Infatti l'introduzione di una variazione sinonima che introduca un codone ancora più raro modifica in modo ancora più vistoso le proprietà biochimiche della pompa. Un altro esempio recente è quello dell'identificazione (De Gobbi et al. 2006) di uno SNP (diffuso in una popolazione della Melanesia) localizzato nel cluster dei geni globinici e responsabile di alfa-talassemia. Lo SNP è localizzato in una regione non codificante e non conservata; gli autori dimostrano come l'effetto patologico sia dovuto alla creazione di un promotore che interferisce con la normale attivazione dei geni alfa-globinici. Tale SNP rappresenta quindi una variazione che determina acquisizione di funzione e il messaggio è quello di non classificare come sicuramente non funzionali SNP localizzati in regioni apparentemente "inutili".

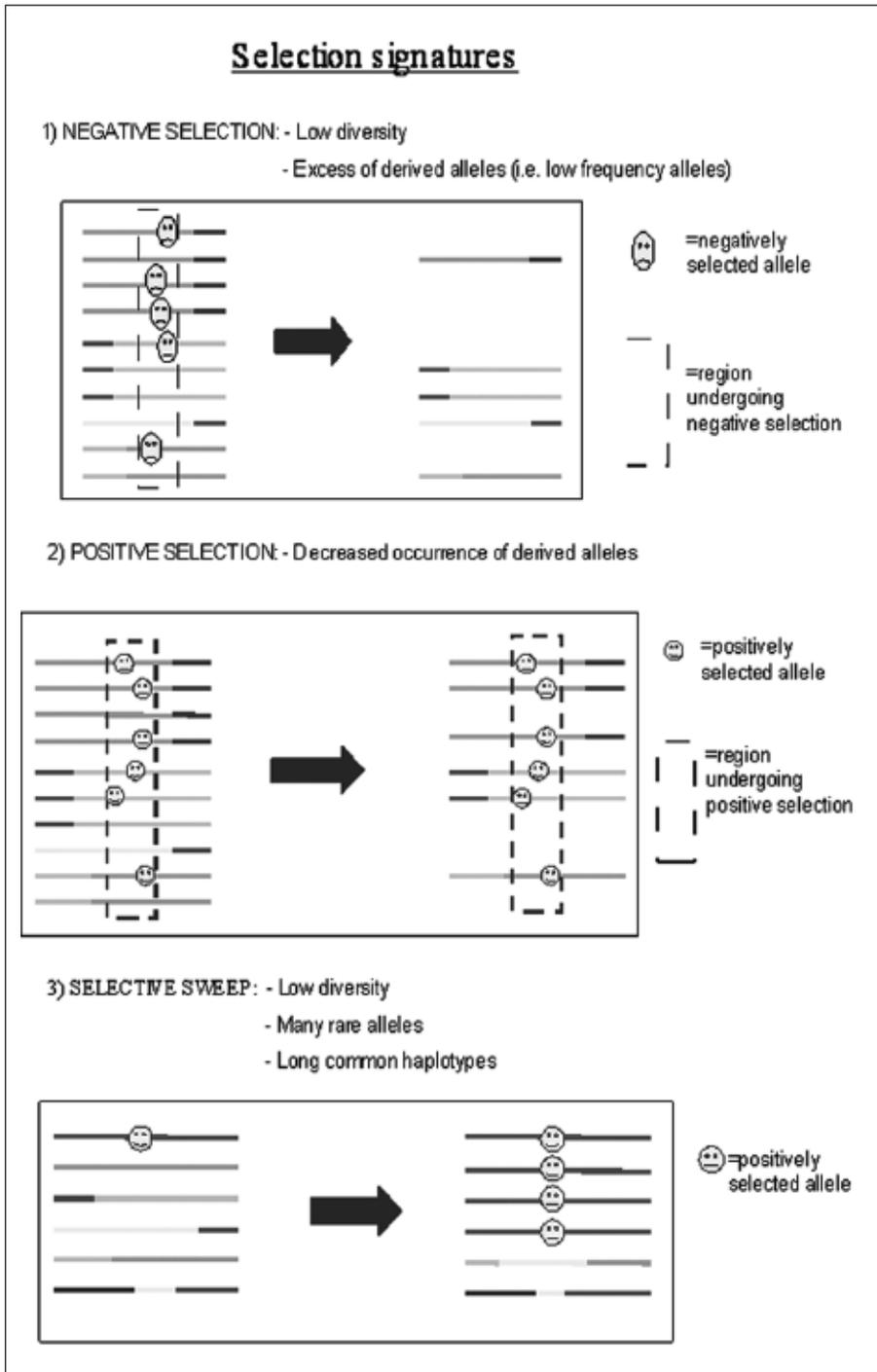


Fig. 3

Un terzo esempio riguarda l'identificazione (Valentonyte et al. 2005) di uno SNP come fattore di rischio per la sarcoidosi. Lo SNP è localizzato nel gene *BTNL2* (che codifica per una proteina della super-famiglia delle immunoglobuline coinvolta nell'attivazione dei linfociti T) e si localizza in posizione -1 rispetto allo splice site donatore dell'esone 5. Gli autori dimostrano come l'allele A porti all'utilizzo di un sito di slicing localizzato 5 bp upstream rispetto a quello canonico con la conseguente delezione di 4 bp e, quindi, causando un frameshift. Questo dimostra come specifici SNP possano rappresentare in realtà variazioni patologiche.

Infine, è utile ricordare l'importanza degli SNP per la farmacogenetica (per una review, Goldstein et al. 2003). Un'importante aspetto della variazione fenotipica tra gli individui è infatti rappresentato dalla risposta ai farmaci. Nel corso degli ultimi anni sono aumentati i singoli SNP (o aplotipi) per cui sia stata riportata un'associazione con la risposta ad un farmaco (farmacogenetica). La maggior parte di tali SNP localizza in geni coinvolti nel metabolismo o nel trasporto dei farmaci. A questo riguardo uno degli esempi più incoraggianti in cui la farmacogenetica ha avuto effetto sulla pratica clinica è quello degli SNP nel gene *TPMT* (thiopurine S-methyltransferase). Alcuni genotipi determinano attività ridotta o assente (1/300 caucasici) di *TPMT*: questo risulta in un'elevata tossicità (anche fatale) in questi soggetti di farmaci tiopurinici utilizzati per la terapia delle leucemie e di alcune malattie autoimmuni. Di conseguenza in molti Paesi la genotipizzazione di *TPMT* è pratica comune prima della somministrazione di farmaci tiopurinici.

Bibliografia

1. De Gobbi M, Viprakasit V, Hughes JR, Fisher C, Buckle VJ, Ayyub H, Gibbons RJ, et al. A regulatory SNP causes a human genetic disease by creating a new transcriptional promoter. *Science*. 2006 May 26; 312(5777): 1215-7.
2. Goldstein DB, Tate SK, Sisodiya SM. Pharmacogenetics goes genomic. *Nat Rev Genet*. 2003 Dec; 4(12): 937-47.
3. International HapMap Consortium. A haplotype map of the human genome. *Nature*. 2005 Oct 27; 437(7063): 1299-320.
4. Kimchi-Sarfaty C, Oh JM, Kim IW, Sauna ZE, Calcagno AM, Ambudkar SV, Gottesman MM. A "silent" polymorphism in the *MDR1* gene changes substrate specificity. *Science*. 2007 Jan 26; 315(5811): 525-8.
5. Myers S, Bottolo L, Freeman C, McVean G, Donnelly P. A fine-scale map of recombination rates and hotspots across the human genome. *Science*. 2005 Oct 14; 310(5746): 321-4.
7. Nielsen R. Molecular signatures of natural selection. *Annu Rev Genet*. 2005; 39: 197-218.
8. Valentonyte R, Hampe J, Huse K, Rosenstiel P, Albrecht M, Stenzel A, Nagy M, Gaede KI, et al. Sarcoidosis is associated with a truncating splice site mutation in *BTNL2*. *Nat Genet*. 2005 Apr; 37(4): 357-64.

Il polimorfismo genico come fattore predisponente allo sviluppo di leucemie acute e sindromi mielodisplastiche

Maria Teresa Voso, Alessandra Scardocci, Francesco D'Alò, Giuseppe Leone

Istituto di Ematologia, Università Cattolica S. Cuore, Roma

Lo sviluppo delle leucemie mieloidi acute (LMA) e delle sindromi mielodisplastiche (SMD) è influenzato oltre che da fattori relativi alla cellula tumorale, anche dalle caratteristiche dell'ospite. Il ruolo dell'ospite è dimostrato in particolare dal fatto che tali malattie si possano presentare in soggetti predisposti, dopo esposizione a carcinogeni ambientali o come sequela di un trattamento chemio- o radio-terapico, per altro tipo di neoplasia. L'attività dei carcinogeni è modificata dal loro metabolismo, che coinvolge enzimi che provvedono alla ossidazione e quindi alla attivazione degli xenobiotici o alla loro coniugazione e quindi alla detossificazione. Si è visto che varianti funzionali di questi enzimi, che risultano dal polimorfismo del DNA dei geni interessati, influenzano non solo il rischio di una varietà di tumori, ma anche la risposta al trattamento citotossico. Una ridotta detossificazione negli individui con genotipi che conferiscono una attività enzimatica più bassa può essere vantaggiosa nell'aumentare l'efficacia dei farmaci. D'altra parte, la ridotta detossificazione potrebbe prolungare l'effetto tossico e indurre lo sviluppo di leucemie secondarie.

Vari studi hanno analizzato diversi polimorfismi in geni che codificano per enzimi implicati nel metabolismo di farmaci, tossici ambientali e xenobiotici, tra cui i geni della famiglia del citocromo P450 (CYP), coinvolti nel metabolismo di fase I (ossidazione/attivazione), i geni della famiglia della glutatione S-transferasi (GST), implicati nel metabolismo di fase II (coniugazione/detossificazione) e il gene dell'enzima chinone ossidoreduttasi (NQO1), che interviene nei processi di metabolismo dei radicali liberi-stress ossidativo.

I geni della famiglia del citocromo P450 intervengono nella bioattivazione di diversi carcinogeni chimici, quali gli idrocarburi policiclici aromatici, con la produzione di metaboliti intermedi altamente reattivi, che possono interagire con il DNA, portando alla formazione di addotti. Una più alta prevalenza dell'allele CYP1A1*4 è stata osservata nei pazienti con LMA. Krajinovic, invece, ha riportato un'aumentata incidenza dell'allele CYP1A1*2A nei bambini con leucemia linfoblastica acuta, con un più alto rischio di recidive. CYP3A4 è la più abbondante isoforma dell'enzima nel fegato umano; esso interviene anche nella con-

versione di diversi agenti antineoplastici in metaboliti intermedi reattivi, alcuni dei quali possono danneggiare il DNA. Felix ha mostrato che solo il 3% delle LMA therapy-related presenta il polimorfismo CYP3A4-V, suggerendo che i portatori dell'allele CYP3A4-wild type potrebbero avere un' aumentata produzione di metaboliti intermedi potenzialmente dannosi per il DNA, con un conseguente aumentato rischio di LMA/SMD therapy-related.

Sono stati riportati dei polimorfismi funzionali in almeno tre dei geni che codificano per GST, cioè GSTM1, GSTT1, GSTP1. Il polimorfismo che riguarda i geni di GSTM1 e GSTT1 è molto diffuso, presenta differenze nell'incidenza tra i diversi gruppi etnici e consiste nella completa delezione del gene, che causa la perdita dell'attività enzimatica. Il polimorfismo di GSTP1 I105V, chiamato anche GSTP1*B, causa la sostituzione aminoacidica isoleucina-valina nel sito attivo di legame elettrofilico del peptide, determinando variazioni nell'attività catalitica e nella stabilità termica dell'enzima. È stato dimostrato che pazienti con un genotipo GST-null presentano una ridotta detossificazione degli agenti genotossici ambientali e dei farmaci chemioterapici, con un aumentato rischio di neoplasie primitive e secondarie e di complicanze legate al trattamento. In particolare, sembra che GSTP1 protegga dalla tossicità da alcuni agenti chemioterapici, come gli alchilanti, le antracicline ed i metaboliti della ciclofosfamida. I polimorfismi di GST possono, così, contribuire alla suscettibilità individuale allo sviluppo di LMA/SMD therapy-related. Inoltre, il coinvolgimento di questi enzimi nella detossificazione di carcinogeni ambientali, come il benzopirene ed altri idrocarburi policiclici aromatici, potrebbe giustificare l'impatto dei polimorfismi GST sullo sviluppo di leucemie de novo.

Sono stati descritti due polimorfismi nel gene NQO1, NQO1*2 (C609T) e NQO1*3 (C465T), che determinano le sostituzioni aminoacidiche P187S e R139W, rispettivamente. Il polimorfismo NQO1*3 determina una diminuita attività enzimatica, mentre NQO1*2 causa una completa perdita dell'attività dell'enzima. NQO1 normalmente protegge dalla tossicità da benzene; uno studio caso-controllo effettuato in Cina su lavoratori esposti al benzene ha mostrato un aumentato rischio di ematotossicità e leucemia negli individui portatori dell'allele NQO1*2. Inoltre, diversi studi hanno mostrato un'associazione tra il polimorfismo NQO1*2 e le LMA/SMD therapy-related; tale polimorfismo, poi, sembra associato alla presenza di alterazioni citogenetiche specifiche, quali le anomalie dei cromosomi 5 e/o 7.

Anche le variazioni nell'attività di geni implicati nel metabolismo dell'acido folico, come la metilene-tetraidrofolato reduttasi (MTHFR), la serina idrossimetiltrasferasi (SHMT1) la timidilato sintasi (TYMS) e la metionina sintasi-reduttasi (MTRR), che svolgono un ruolo indiretto nella corretta sintesi del DNA, possono avere un ruolo importante nella suscettibilità individuale allo sviluppo di leucemie. In particolare, diversi studi hanno mostrato una riduzione del rischio di sviluppo di leucemia linfoblastica acuta sia negli adulti che nei bambini portatori del polimorfismo MTHFR C677T da solo o in combinazione con il polimorfismo MTHFR A1298C, così come nei portatori dei polimorfismi SHMT1 1420CT o 1420TT. Krajcinovic ha riportato che la presenza del genotipo omozigote TYMS

3R si associa ad una cattiva risposta al trattamento nei bambini con leucemia linfoblastica acuta, mentre Rocha ha mostrato un aumento del rischio di recidiva ematologia nei bambini con LLA portatori del genotipo TYMS 3R/3R in associazione con il genotipo GSTM1 non-null.

Fra i fattori relativi al tumore, in particolare nelle leucemie acute e nelle sindromi mielodisplastiche therapy-related è stato dimostrato un ruolo dei polimorfismi anche degli enzimi di riparazione del DNA, quali RAD51, XRCC3, XPD, hMSH2, nell'indurre maggiore instabilità genica, in particolare se combinati ai polimorfismi della detossificazione. È stato riportato che la presenza combinata degli alleli RAD51-135C e XRCC3-241M aumenta il rischio di sviluppo di LMA/SMD therapy-related. Un importante studio di Allan ha valutato la presenza del polimorfismo K751Q del gene XPD in 341 pazienti anziani affetti da LMA, trovando un'associazione tra la presenza dell'omozigosi 751Q e il rischio di LMA therapy-related. Worilow ha riportato un'associazione tra la presenza del polimorfismo IVS12-6C>T del gene hMSH2 e lo sviluppo di leucemie acute secondarie al trattamento con agenti alchilanti.

Nel nostro Istituto di Ematologia, abbiamo studiato la frequenza dei polimorfismi di geni implicati nella detossificazione degli xenobiotici e nella riparazione del DNA nei pazienti affetti da LMA e SMD, paragonati ad una popolazione di controllo costituita da un gruppo di individui "matched" per sesso ed età, e con anamnesi negativa per ogni tipo di neoplasia. In un primo studio, abbiamo osservato una più alta prevalenza dell'allele 4 del citocromo CYP1A1 e delle delezioni omozigoti di GSTT1 nei pazienti affetti da LMA rispetto ai controlli. Inoltre la combinazione del genotipo null per GSTT1 e la presenza degli alleli 2B o 4 di CYP1A1 ulteriormente incrementava il rischio di sviluppare LMA, suggerendo una possibile interazione tra un'iperattività degli enzimi di fase I e un difetto degli enzimi di fase II nella leucemogenesi. In un altro studio abbiamo osservato un rischio aumentato di sviluppare LMA associato la variante allelica A290G del citocromo CYP3A4, mentre lo status omozigote per l'allele *B dell'enzima GSTA1 rappresentava un fattore protettivo. Per quel che riguarda gli enzimi di riparazione, la variante allelica C alla posizione -135 di RAD51 e lo stato omozigote per l'allele 241Met di XRCC3 sono più frequenti nelle LMA rispetto ai controlli. Inoltre il rischio di sviluppare LMA è significativamente aumentato dalla presenza simultanea di RAD51-G135C e CYP3A4-A-290G, suggerendo che anche difetti combinati dei sistemi di detossificazione e riparazione del DNA possano cooperare nel determinare la suscettibilità individuale allo sviluppo di LMA. Allo stesso modo nelle SMD, abbiamo osservato una più elevata frequenza dell'allele RAD51-G135C, e dell'associazione tra i polimorfismi RAD51-G135C e NQO1-Pro187Ser rispetto ai controlli. Inoltre, abbiamo visto come alcuni di questi polimorfismi, come le delezioni di GSTM1 e T1 e la presenza dell'allele 2a del citocromo CYP1A1, siano in grado di influenzare negativamente la prognosi dei pazienti affetti da LMA.

In conclusione, possiamo dire che i polimorfismi genici degli enzimi coinvolti nella detossificazione e nella riparazione del DNA hanno senza dubbio un'influenza molto importante sul rischio di sviluppo di leucemie acute de novo e di

Tab. 1 - Polimorfismi genici (modificata da Bolufer et al., Leukemia Research 2006).

Gene	Polimorfismo	Effetto	Frequenza (%)	Effetto biologico
CYP1A1	CYP1A1*2A	I462V	12-16	Aumento espressione
	CYP1A1*2B	T461N	8-9	Aumento espressione
	CYP1A1*4		10	Sconosciuto
CYP3A4	CYP3A4*1B/V	Promotore	22	Riduzione espressione
GSTM1	Del(GSTM1)	Delezione	50	Abolita espressione
GSTT1	Del(GSTT1)	Delezione	41	Abolita espressione
GSTP1	GSTP1*B	I105V	54	Dim. attività catalitica
NQO1	NQO1*2	P187S	48	Abolisce att. enzimatica
	NQO1*3	R139W	4	Diminuita espressione
MTHFR	A1298C	E429A	49	Dim. attività enzimatica
	C677T	A222V	51	Dim. attività catalitica
SHMT1	C1420T o C1444T	L435F	38	Altera att. enzimatica
TYMS	TYMS2R o 3R		38-54	Aumento espressione
MTRR	A2756G	D919G	17	Altera att. enzimatica
RAD51	G135C		11	Altera ricombinazione omologa
XRCC3	Ex8-53C>T	T241M	35	Altera ricombinazione omologa
XPD	Ex23+61A>C	K751Q	39	Altera "nucleotide excision DNA repair"
hMSH2	IVS12-6C>T		15	Altera il mismatch repair

LMA/SMD therapy-related, e possono rappresentare un elemento predittivo della risposta al trattamento e del rischio di recidiva. Sarebbe, quindi, importante ampliare le famiglie di geni e i polimorfismi studiati, per poter provare a fornire una più accurata definizione prognostica ed un approccio terapeutico più individualizzato.

Bibliografia

1. Rushmore TH, Kong AN. Pharmacogenomics, regulation and signaling pathways of phase I and II drug metabolizing enzymes. *Curr Drug Metab.* 2002 Oct; 3(5): 481-90.
2. Fujita K. Cytochrome P450 and anticancer drugs. *Curr Drug Metab.* 2006 Jan; 7(1): 23-37.
3. D'Alo' F, Voso MT, Guidi F, Massini G, Scardocci A, Sica S, Pagano L, Hohaus S, Leone G. Polymorphisms of CYP1A1 and glutathione S-transferase and susceptibility to adult acute myeloid leukemia. *Haematologica* 2004; Jun 89(6): 664-670.
4. Krajcinovic M, Labuda D, Richer C, Karimi S, Sinnott D. Susceptibility to

- childhood acute lymphoblastic leukemia: influence of CYP1A1, CYP2D6, GSTM1 and GSTT1 genetic polymorphism. *Blood* 1999; 93: 1496-501.
5. Felix CA, Walker AH, Lange BJ, Williams TM, Winick NJ, Cheung NK, Lovett BD, Nowell PC, Blair IA, Rebbeck TR. Association of CYP3A4 genotype with treatment-related leukemia. *Proc Natl Acad Sci U S A*. 1998 Oct 27; 95(22): 13176-81.
 6. Allan JM, Wild CP, Rollinson S, Willett EV, Moorman AV, Dovey GJ, Roddam PL, Roman E, Cartwright RA, Morgan GJ. Polymorphism in glutathione S-transferase P1 is associated with susceptibility to chemotherapy-induced leukaemia. *Proc Natl Acad Sci USA*. 2001 Sep 25; 98(20): 11592-7.
 7. Sasai Y, Horiike S, Misawa S, Kaneko H, Kobayashi M, Fujii H, Kashima K, Taniwaki M Genotype of glutathione S-transferase and other genetic configurations in myelodysplasia. *Leuk Res*. 1999 Nov; 23(11): 975-81.
 8. Smith MT. Benzene, NQO1 and genetic susceptibility to cancer. *Proc Natl Acad Sci USA* 1999; 96: 7624-6.
 9. Lan Q, Zhang L, Li G, Vermeulen R, Weinberg RS, Dosemeci M, Rappaport SM, Shen M, Alter BP, Wu Y, Kopp W, Waidyanatha S, Rabkin C, Guo W, Chanock S, Hayes RB, Linet M, Kim S, Yin S, Rothman N, Smith MT. Hematotoxicity in workers exposed to low levels of benzene. *Science*. 2004 Dec 3; 306(5702): 1774-6
 10. Larson RA, Wang Y, Banerjee M, Wiemels J, Hartford C, Le Beau MM, Smith MT. Prevalence of the inactivating 609C->T polymorphism in the NAD(P)H:quinone oxidoreductase (NQO1) gene in patients with primary and therapy-related myeloid leukaemia *Blood*. 1999 Jul 15; 94(2): 803-7.
 11. Smith MT, Wang Y, Skibola CF, Slater DJ, Lo Nigro L, Nowell PC, Lange BJ, Felix CA Low NAD(P)H:quinone oxidoreductase activity is associated with increased risk of leukemia with MLL translocations in infants and children *Blood*. 2002 Dec 15; 100(13): 4590-3.
 12. Krajcinovic M, Lamothe S, Labuda D, Lemieux-Blanchard E, Theoret Y, Moghrabi A, Sinnett D. Role of MTHFR genetic polymorphisms in the susceptibility to childhood acute lymphoblastic leukemia. *Blood*. 2004 Jan 1; 103(1): 252-7.
 13. Krajcinovic M, Costea I, Chiasson S. Polymorphism of the thymidylate synthase gene and outcome of acute lymphoblastic leukaemia. *Lancet*. 2002 Mar 23; 359(9311): 1033-4.
 14. Rocha JC, Cheng C, Liu W, Kishi S, Das S, Cook EH, Sandlund JT, Rubnitz J, Ribeiro R, Campana D, Pui CH, Evans WE, Relling MV. Pharmacogenetics of outcome in children with acute lymphoblastic leukemia. *Blood*. 2005 Jun 15; 105(12): 4752-8.
 15. Seedhouse C, Faulkner R, Ashraf N, Das-Gupta E, Russell N Polymorphisms in genes involved in homologous recombination repair interact to increase the risk of developing acute myeloid leukemia. *Clin Cancer Res*. 2004 Apr 15; 10(8): 2675-80.
 16. Allan JM, Smith AG, Wheatley K, Hills RK, Travis LB, Hill DA, Swirsky

- DM, Morgan GJ, Wild CP. Genetic variation in XPD predicts treatment outcome and risk of acute myeloid leukemia following chemotherapy. *Blood*. 2004 Dec 15; 104(13): 3872-7.
17. Worrillow LJ, Travis LB, Smith AG, Rollinson S, Smith AJ, Wild CP, Holowaty EJ, Kohler BA, Wiklund T, Pukkala E, Roman E, Morgan GJ, Allan JM. An intron splice acceptor polymorphism in hMSH2 and risk of leukemia after treatment with chemotherapeutic alkylating agents. *Clin Cancer Res*. 2003 Aug 1; 9(8): 3012-20.
 18. Voso MT, D'Alo' F, Putzulu R, Mele L, Scardocci A, Chiusolo P, Latagliata R, Lo-Coco F, Rutella S, Pagano L, Hohaus S, Leone G. Negative prognostic value of glutathione S-transferase (GSTM1 and GSTT1) deletions in adult acute myeloid leukemia. *Blood*. 2002 Oct 15; 100(8): 2703-7.
 19. Voso MT, D'Alo' F, Gumiero D, Guidi F, Hohaus S, Leone G. The CYP1A1*2a allele is an independent prognostic factor for acute myeloid leukemia. *Haematologica*. 2005 Jul; 90(7): 982-4.

I microRNA: una nuova classe di regolatori dell'espressione genica con un ruolo nel cancro

Gisuseppe Biamonti

Istituto di Genetica Molecolare del CNR, Pavia

RNA interference (RNAi) è un meccanismo di silenziamento genico altamente specifico e robusto che dipende da molecole di RNA a doppio filamento (*dsRNA*).

La prima evidenza sperimentale del ruolo dei dsRNA risale a meno di 10 anni fa ed è stata ottenuta nel nematode *Caenorhabditis elegans*. Successivamente, la caratterizzazione di questo complesso meccanismo biochimico ha completamente rivoluzionato tutti gli aspetti della biologia molecolare ampliando la nostra comprensione della regolazione dell'espressione genica e offrendo nuovi strumenti per approcci terapeutici.

RNAi è un fenomeno naturale conservato nell'evoluzione che si verifica nei funghi, nelle piante e negli animali. In tutti questi organismi molecole di dsRNA possono indurre RNAi. Inizialmente si è ritenuto che questo non fosse estendibile alle cellule di mammifero dove si sa che lunghe molecole di dsRNA inducono sia una risposta antivirale mediata dall'interferone che l'attivazione di meccanismi aspecifici di degradazione dei dsRNA.

La prima evidenza che il macchinario biochimico di RNAi è conservato anche nei mammiferi è stata l'osservazione che la degradazione di specifiche molecole di RNA messaggero (*mRNA*) può venir ottenuta introducendo nelle cellule piccole molecole di dsRNA lunghe 21-23 nucleotidi (nt). Questi *siRNA - small interfering RNA* - inducono RNAi senza attivare gli altri meccanismi indotti dalle molecole di dsRNA lunghe.

Gli siRNA iniziano gene silencing legandosi alla sequenza bersaglio di mRNA in maniera sequenza specifica (secondo le regole di appiamento delle basi A:T e G:C) e inducendone la degradazione.

Nelle cellule esistono anche piccoli RNA non-codificanti (19-25 nt) noti come *miRNA - micro RNA* - che hanno un ruolo importante ed evolutivamente conservato nel controllo dello sviluppo, della proliferazione cellulare, nell'ematopoiesi e nell'apoptosi.

Diversamente dai siRNA, i miRNA solitamente non inducono la degradazione del mRNA bersaglio ma ne impediscono la traduzione in proteine.

Un terzo meccanismo, meno compreso, tramite il quale i piccoli RNA controllano l'espressione genica è la capacità di indurre silenziamento trascrizionale modulando l'organizzazione epigenetica ed inducendo la formazione di eterocromatina. Riassumendo: piccole molecole di dsRNA non codificanti controllano l'espressione genica:

- 1) inducendo la degradazione di mRNA bersaglio;
- 2) impedendo la traduzione di mRNA bersaglio;
- 3) modulando lo stato epigenetico di specifiche sequenze bersaglio di DNA.

Gli siRNA

RNAi può venir indotto da molecole di RNA con origine differente come virus a RNA, transposoni, RNA introdotti nelle cellule con esperimenti di biologia molecolare. In tutti questi casi non si parla di RNA endogeni cellulari.

La capacità di lunghe molecole di dsRNA di indurre RNAi richiede la loro conversione in siRNA di 21-23 nt da parte di *Dicer* – una *RNasi di tipo III*. *Dicer* uti-

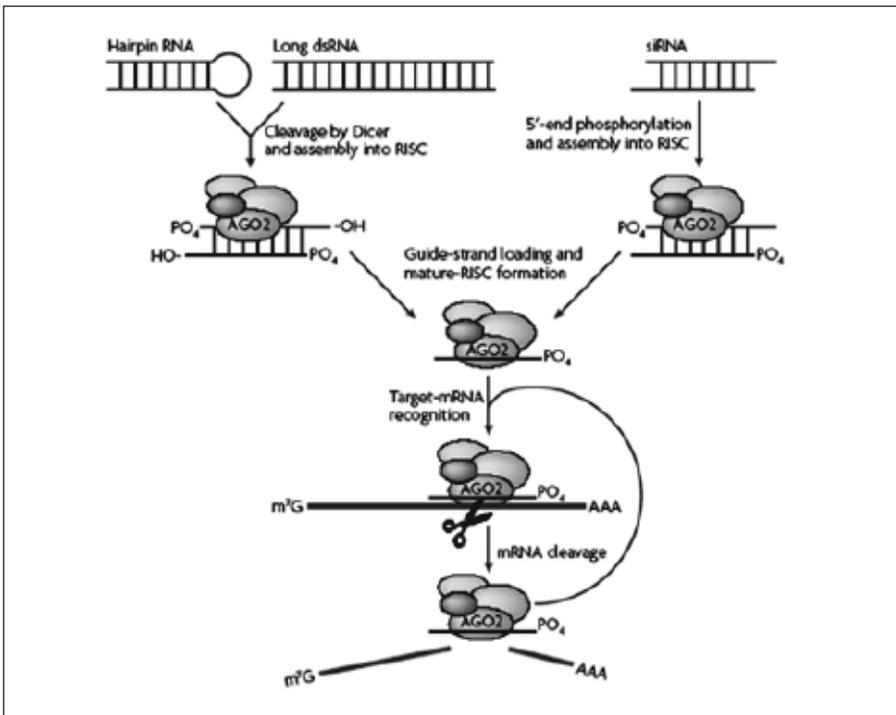


Fig. 1 - *Dicer* utilizza come substrato molecole di dsRNA o hairpin di RNA e li processa in piccoli (21 nt) small interfering RNA (siRNA). Questi presentano 2 nt protrudenti all'estremità 3'-OH e l'estremità 5'-fosforilata. Gli siRNA possono venir anche introdotti nelle cellule dove vengono fosforilati da chinasi cellulari. Gli siRNA vengono assemblati nel complesso RISC che contiene AGO2, *Dicer*, e altri fattori cellulari (per semplicità viene indicato solo AGO2). Il legame con siRNA attiva RISC. Nel complesso è presente solo il filamento anti-senso o guida. Grazie alla complementarità di sequenza tra le molecole di RNA, il complesso attivato riconosce il bersaglio e lo digerisce, il complesso viene riciclato.

lizza le molecole lunghe di dsRNA come substrato per produrre siRNA con estremità 3'-OH che protrudono di 2 nt e estremità 5' fosforilate. I due filamenti di RNA presenti nei siRNA sono rispettivamente "senso" e "anti-senso" rispetto alla molecola di mRNA bersaglio. Il filamento anti-senso è chiamato "guida" perché riconosce la molecola di mRNA bersaglio e ne induce la degradazione. Nelle applicazioni di biologia molecolare si può ottenere RNAi sia introducendo nelle cellule siRNA sintetizzati *in vitro* che utilizzando vettori in grado di esprimere RNA substrati di Dicer. Gli siRNA vengono reclutati in un complesso multiproteico detto RISC - *RNA induced silencing complex*. Solo il filamento guida viene incorporato nel complesso.

L'RNA guida si appaia alla molecola di mRNA bersaglio dirigendo su questa le proteine del complesso RISC. L'estremità 5' della molecola guida determina il

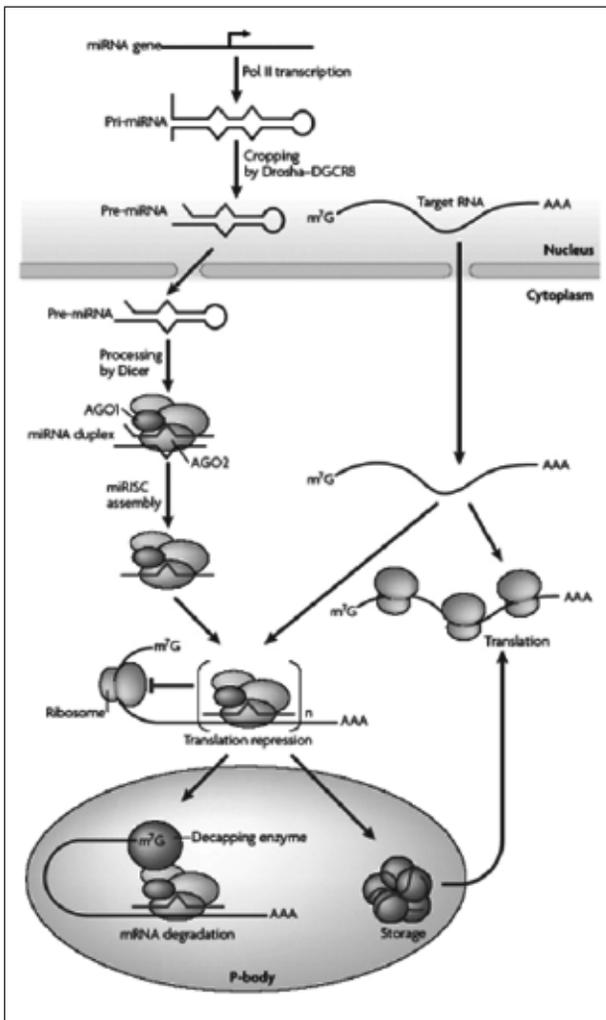


Fig. 2 - I geni per i microRNA (miRNA) sono trascritti dalla RNA polimerasi II (Pol II) in lunghi (60-100 nt) pri-miRNA (primary miRNA) che assumono strutture stem-loop.

I pri-miRNA vengono digeriti dal complesso Drosha-DGCR8 a formare molecole di 70 nt dette pre-miRNA che presentano 2 nt protrudenti all'estremità 3'-OH. I pre-miRNA vengono esportati nel citoplasma dalla esportina 5. Nel citoplasma vengono digeriti da Dicer in piccoli RNA (22 nt) detti miRNA che si associano al complesso RISC. Il complesso RISC attivato si lega al mRNA bersaglio ma a causa della non completa complementarietà l'ibrido mRNA:miRNA non può venir digerito da RISC e si accumula nei P-bodies. In questo modo viene inibita la traduzione del mRNA.

La molecola di mRNA viene conservata nei P-bodies per utilizzi futuri o distrutta con il meccanismo deputato alla degradazione dei mRNA.

sito di taglio del mRNA bersaglio che è 10-11 nt a monte del 5'-fosfato della molecola *guida*. Il processo richiede la perfetta complementarità tra la molecola guida ed il bersaglio ed è conseguentemente altamente specifico.

I miRNA

Il macchinario di RNAi può anche venir programmato dai miRNA, piccoli RNA codificati dal genoma cellulare. I miRNA hanno un ruolo importante nella regolazione post-trascrizionale dell'espressione genica e controllano un elevato numero di geni in tutti gli organismi superiori, incluso l'uomo. Anche i miRNA vengono reclutati nel complesso RISC. La biogenesi dei miRNA inizia nel nucleo. I geni che codificano miRNA sono nelle regioni intergeniche e in unità trascrizionali definite in cui sono presenti sia le sequenze senso che quelle anti-senso. I geni sono trascritti dalla RNA polimerasi II sotto forma di *miRNA primari - pri-miRNA*. Un tipico *pri-miRNA* consiste di una struttura "stem and loop" e da regioni fiancheggianti a singolo filamento. Queste molecole vengono riconosciute da *Drosha*, una RNasi tipo III come Dicer ma a localizzazione nucleare. *Drosha* agisce in un complesso con *DGCR8* (*DiGeorge syndrome critical region gene 8*) noto anche come *Pasha* (*partner of Drosha*). Il prodotto di *Drosha* è una molecola di 70 nt detta *pre-miRNA - precursore di miRNA*.

I *pre-miRNA* vengono esportati nel citoplasma dalla *esportina 5* dove vengono maturati in miRNA da Dicer e successivamente assemblati nel complesso RISC. La molecola di miRNA non è perfettamente complementare alla molecola di mRNA bersaglio. Generalmente il sito di appaiamento si trova nella regione non tradotta al 3' del mRNA. In assenza di una complementarità completa il complesso RISC non riesce a tagliare il bersaglio ma, attraverso un meccanismo ancora poco compreso, ne impedisce la traduzione.

Studi recenti indicano che il complesso RISC si localizza nei *P-bodies*, particolari strutture citoplasmatiche che contengono gli mRNA legati ai complessi che ne impediscono la traduzione. Qui gli mRNA possono venir conservati per un uso futuro o venir degradati dai normali meccanismi di degradazione dei trascritti.

Il complesso RISC

Tra le proteine del complesso RISC ci sono:

- 1) *Dicer*;
- 2) le proteine *Argonauta*;
- 3) *TRBP* (HIV-1 trans-activation responsive element (TAR) RNA-binding protein);
- 4) la dsRNA-binding protein *PACT*.

Dicer è conservato da *Schizosaccharomyces pombe* (un lievito) all'uomo. Le proteine Argonauta (AGO) sono anch'esse conservate e hanno un ruolo sia nel taglio della molecola bersaglio che nella produzione dei piccoli RNA.

Ruolo dei miRNA nella regolazione dell'espressione genica

È stato stimato che il genoma umano codifica per almeno 300 (forse 1000) miRNA. Questo rende i miRNA una delle famiglie più grandi di regolatori dell'espressione genica.

La maggior parte dei miRNA si trovano negli introni di geni che codificano per proteine o in trascritti che non codificano per proteine. Altri miRNA sono associati in gruppi (cluster) all'interno di unità trascrizionali dedicate. Uno di questi cluster si trova sul cromosoma 19 e contiene ben 54 miRNA. I miRNA possono venir raggruppati in famiglie in base alle omologie di sequenza. Molti sono evolutivamente conservati suggerendo un ruolo in processi biologici essenziali.

Uno dei problemi principali nello studio dei miRNA è quello di identificare i bersagli della loro azione. Questa difficoltà è dovuta al fatto che:

1) non esiste una perfetta complementarità tra bersaglio e miRNA. Inoltre 2) la presenza di una regione di complementarità non vuol dire necessariamente che la molecola di mRNA sia un bersaglio.

Il sito di complementarità infatti potrebbe essere mascherato e non accessibile. È stato calcolato che ogni miRNA ha almeno 200 mRNA bersaglio. Questo indica che almeno 1/3 dei geni umani vengono controllati da miRNA.

Tab. 1 - MicroRNAs that are associated with human cancers.

miRNA	Gene Loci	Cancer association	Function
miR-15a miR-16-1	Chromosome 13q14	Frequently deleted or downregulated in B-cell chronic lymphocytic leukemia; negatively regulates the anti-apoptotic gene BCL2	TS
miR-143	Chromosome	Decreased abundance in colorectal cancer; downregulated in breast, prostate, cervical and lymphoid cancer cell lines; miR-145 is decreased in breast cancer	TS
miR-21	Chromosome 17q23.2	Anti-apoptotic factor; upregulated in glioblastomas and breast cancer	OG
let-7 family members	Multiple loci	Negatively regulated the Ras oncogenes; direct cell proliferation and differentiation; decreased abundance in lung cancer	TS
miR-142	Chromosome 17q22	A t(8;17) translocation that places the MYC oncogene downstream of the mir-142 hairpin, resulting in an aggressive B-cell leukemia that is due to MYC overexpression	N/A
BIC/ miR-155	Chromosome 21q21	Upregulated in paediatric Burkitt, Hodgkin, primary mediastinal and diffuse large-B-cell lymphomas; upregulated in human breast cancer	OG
miR-17-19b cluster	Chromosome 13q31-32	upregulated by MYC; negatively modulates the E2F1 oncogene; loss of heterozygosity of this cluster is found in hepatocellular carcinoma; overexpressed in B-cell lymphomas	TS/OG

N/A, not applicable; OG, oncogene; TS, tumour suppressor.

Associazione con il cancro

Ovviamente l'importanza che i piccoli RNA rivestono nei processi di regolazione dell'espressione genica suggerisce che una alterazione nei livelli o nel profilo di espressione di queste molecole possa avere delle ricadute sull'identità cellulare con conseguenti effetti patologici.

È stato osservato che in alcuni tumori si ha una deregolazione di geni che codificano per il macchinario di RNAi. Ad esempio una ridotta espressione di Dicer correla con una bassa sopravvivenza post-operatoria di individui con tumori polmonari da non-small-cells. Le proteine AGO sono frequentemente delete nel tumore di Wilms del rene e nei tumori neuroectodermici.

L'espressione di particolari miRNA correla con differenti tipi di cancro. Questi miRNA possono avere sia funzione di soppressori del tumore che di oncogeni. Verranno presentati modelli esemplificativi del ruolo di miRNA nella progressione tumorale.

Bibliografia

1. Rana. (2007) Illuminating the silence: understanding the structure and function of small RNAs. *Nature Reviews. Molecular Cell Biology* 8: 23-36.
2. Dalmay and Edwards (2006) MicroRNAs and the hallmarks of cancer. *Oncogene* 25: 6170-6175.
3. Esquela-Kerscher and Slack (2006) Oncomirs – microRNAs with a role in cancer. *Nature Reviews. Cancer* 6: 259-269.

I microRNA nella diagnosi e progressione della LLC

Massimo Negrini

Dipartimento di Medicina Sperimentale e Diagnostica,
Sezione di Microbiologia, Università di Ferrara

La LLC è costituita da una popolazione monoclonale di linfociti B dal fenotipo maturo caratterizzate dal punto di vista immunofenotipico dall'espressione di CD5, CD19 e CD23 e da bassi livelli di IgM e IgD di membrana.

Normale sviluppo linfocitario

I linfociti B maturano nel midollo osseo dove vanno incontro ad un processo di riarrangiamento di segmenti immunoglobulinici per produrre migliaia di immunoglobuline diverse nella porzione variabile, segmento responsabile del riconoscimento di antigeni estranei. Quando un antigene incontra una immunoglobulina con sufficiente affinità, il linfocita B entra nel centro germinativo del follicolo linfonodale dove si divide ripetutamente (centroblasto), dando origine ad una popolazione di linfociti discendenti (centrociti). In questa fase, la regione variabile delle Ig va incontro ad un processo di ipermutazione somatica che cambia significativamente, incrementandola o riducendola, l'affinità delle Ig per l'antigene: i centrociti con maggiore affinità per l'antigene si attivano e proliferano in presenza dell'antigene. In questo processo, generalmente ma non esclusivamente T-dipendente, si sviluppano plasmacellule mature che secernono Ig e cellule B con memoria. I centrociti non in grado di legare l'antigene o autoreattivi sono normalmente eliminati. Nel processo di attivazione, i segnali ricevuti attraverso le immunoglobuline recettoriali sono trasferiti al nucleo attraverso l'attivazione di membri della famiglia Src di tirosin-chinasi, come Syc e Lyn. Nei linfociti T, un ruolo analogo viene svolto dalla proteina *zeta-chain associated protein 70* (ZAP-70). Inoltre, le proteine di superficie si modificano nei linfociti attivati: tra queste, CD38, che in certe condizioni è in grado di aumentare il segnale proveniente dalle Ig e regolare l'apoptosi dei linfociti B.

Caratteristiche della LLC

Nell'ultima decade, la LLC si è dimostrata una leucemia eterogenea sia dal punto di vista clinico che molecolare. Alcuni pazienti con LLC sopravvivono per molti

Tab. 1 - Incidence of Chromosomal Abnormalities in 325 Patients with Chronic Lymphocytic Leukemia (da Dohner et al., NEJM, 343: 1910, 2000).

Aberration	No. of Patients (%)*
13q deletion	178 (55)
11q deletion	58 (18)
12q trisomy	53 (16)
17q deletion	23 (7)
6q deletion	21 (6)
8q trisomy	16 (5)
t(14q32)	12 (4)
3q trisomy	9 (3)
Clonal abnormalities	268 (82)
Normal karyotype	57 (18)

*one hundred seventy-five patients had one aberration, 67 had two aberrations, and 26 had more than two aberrations.

anni senza terapia ed il decesso può essere non relato alla leucemia. Altri invece hanno un decurso rapido e fatale della malattia. Gli attuali metodi di stadiazione (Rai e Binet) non sono in grado di predire il decorso clinico.

Tuttavia, è stato ripetutamente osservato che lo stato mutazionale delle regioni variabili delle Ig e l'espressione di ZAP-70 correla con il decorso clinico. Pazienti privi o con poche (<2%) mutazioni nelle IgV e con numerose cellule esprimenti ZAP-70 hanno generalmente un decorso aggressivo della malattia; viceversa, pazienti con >2% mutazioni somatiche nelle regioni IgV e bassi livelli di ZAP-70 hanno un decorso clinico indolente. Zap-70 è una proteina intracellulare che trasferisce il segnale di attivazione che deriva da linfociti T e NK attraverso i loro recettori per gli antigeni; sebbene raramente espressa in linfociti B normali, è stata trovata espressa in numerosi casi di LLC., dove può facilitare la trasmissione del segnale in seguito ad interazione Ag-Ig.

Dal punto di vista della patogenesi molecolare, la LLC rappresenta un caso unico tra le neoplasie del sistema ematopoietico.

Infatti, le traslocazioni cromosomiche sono rare e, in generale, anomalie citogenetiche sono rare nelle fasi iniziali della malattia, mentre compaiono durante la progressione.

È stato ipotizzato un ruolo della stimolazione antigenica persistente (antigeni microbici o autoantigeni) nella espansione clonale iniziale di linfociti B precursori della LLC. Durante la progressione, tuttavia, la presenza di aberrazioni cromosomiche si evidenzia in circa l'80% dei casi (Tabella 1).

La delezione del cromosoma 13q14 è l'aberrazione più caratteristica della LLC ed è evidenziabile in oltre il 50% dei casi. Altre alterazioni coinvolgono delezioni ai cromosomi 11q23, 17p13 e 6q, e trisomie del cromosoma 12q e 8q; traslocazioni che coinvolgono il locus IgH sono presenti in meno del 5% dei casi. Queste alterazioni genetiche sono in larga parte responsabili del decorso della malattia (Figura. 1).

Tra le alterazioni di espressione frequentemente osservate nelle cellule leucemiche, importanti appaiono la sovraespressione dei geni anti-apoptotici BCL-2, survivin e MCL1.

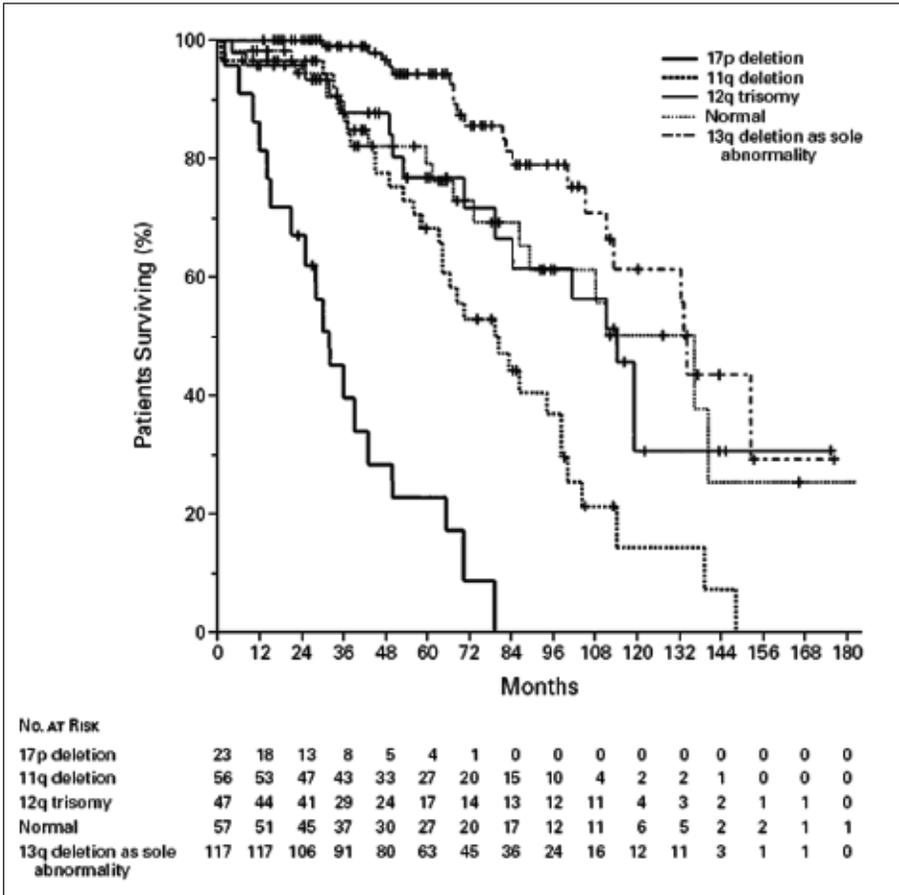


Fig. 1 - (da Dohner et al. NEJM, 343: 1910, 2000) Curve di sopravvivenza correlate a specifiche alterazioni cromosomiche nella LLC.

Basi genetiche del coinvolgimento di microRNA nella LLC

La delezione del cromosoma 13q14 è la più frequente e caratteristica aberrazione cromosomica osservabile nella LLC. Lo studio di questa delezione ha fornito la prima evidenza del coinvolgimento di microRNA in neoplasie umane.

La presenza di delezioni cromosomiche ricorrenti suggerisce la presenza di un gene oncosoppressore nella regione coinvolta. Pertanto, con lo scopo di identificare il gene oncosoppressore coinvolto nella patogenesi della LLC, indagini genetiche dettagliate sui geni presenti nella regione di delezione non sono state in grado di provare la presenza di mutazioni inattivanti o alterazioni di metilazione del DNA in nessuno dei numerosi geni presenti nella regione. Attraverso l'analisi di microdelezioni in 13q14 presenti in due casi di LLC è stato possibile definire una regione di circa 30 kb. In questa regione, sono presenti gli esoni 3 e 4 del gene LEU2, che tuttavia è stato scartato come candidato oncosoppressore, ma

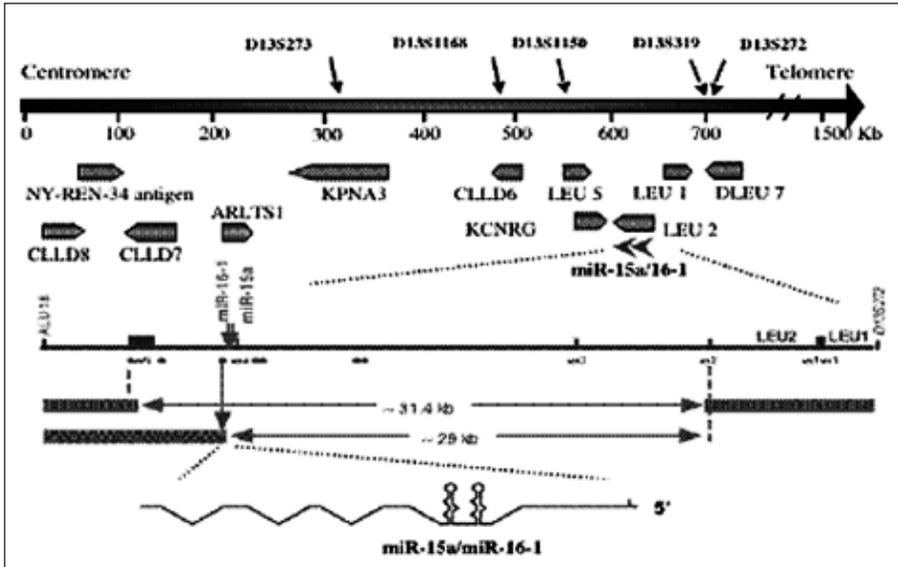


Fig. 2 -Mappa della regione 13q14 colpita da delezioni nella LLC. Nella parte inferiore sono schematizzate le microdelezioni caratterizzate in due casi di LLC, dove si può notare la localizzazione dei microRNA mir-15a e mir-16-1.

sono anche stati riconosciuti due miRNA, mir-15a e mir-16-1, che si sono dimostrati essere sottoespressi in circa il 70% dei casi di LLC ed in tutti i casi associati a delezione 13q (Figura 2).

Mir-15a e mir-16-1 costituiscono un piccolo cluster. Essi sono infatti distanti meno di 200 bp e presentano un unico trascritto primario. L'espressione di questi microRNA è ubiquitaria, con i livelli più alti osservati nei linfociti CD5+, la normale controparte dei linfociti caratteristici della LLC. Dallo sviluppo ed applicazione di un microarray contenente sonde per tutti i miRNA umani noti, abbiamo confermato che la perdita di espressione di mir-15 e mir-16 è una caratteristica distintiva delle LLC con delezione in 13q14. Nel loro complesso, questi dati suggeriscono che il cluster mir-15/16 possa svolgere una funzione oncosoppressoria nei linfociti CD5+, eliminata dalla delezione in 13q14. A sostegno di questa ipotesi, la scoperta di una mutazione (sostituzione C a T) associata a delezione dell'allele normale in due casi di LLC. Tale mutazione è assente in oltre 160 individui non affetti da neoplasia (P<0.001). L'effetto di questa sostituzione, presente nella sequenza del trascritto primario ma non nel precursore, è quello di ridurre i livelli di miRNA maturo, svolgendo perciò un ruolo inibitorio sulla funzione del miRNA.

Funzioni bio-molecolari di mir-15a/mir-16-1

Numerosi studi hanno evidenziato che il gene anti-apoptotico BCL2 è sovraespresso nella LLC con un meccanismo sconosciuto. La regione 3'UTR del gene BCL2 umano (ed anche murino) presenta siti potenzialmente riconosciuti dai

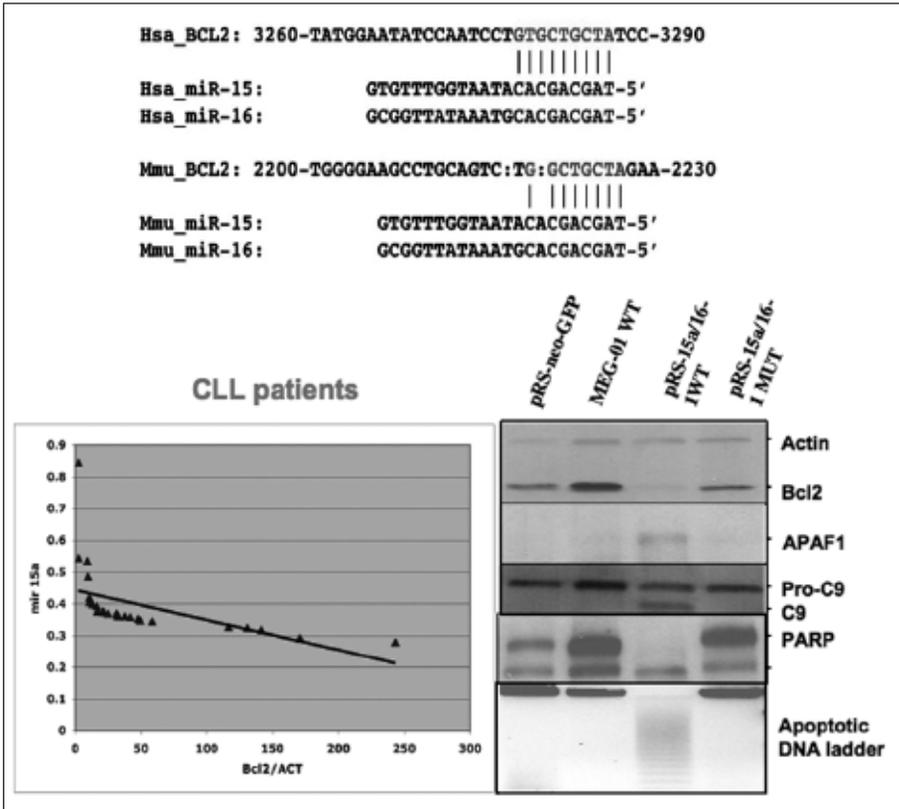


Fig. 3 -BCL2 è un bersaglio di mir-15a e mir-16-1. In alto la sequenza presente nella regione 3'UTR del trascritto di BCL2 umano e murino. In basso a sinistra, la correlazione inversa tra espressione di BCL2 ed espressione di mir15a. In basso a destra, dimostrazione della perdita di espressione di BCL2 in seguito a transfezione del vettore pRS-15a/16-1WT e concomitante attivazione della via apoptotica intrinseca.

microRNA mir-15 e mir-16 (Figura 3). Dalla analisi di dati di espressione di BCL2 e di mir15/16 in 30 diverse LLC, abbiamo osservato una correlazione inversa tra i rispettivi livelli di espressione, suggerendo che i due miRNAs potessero regolare negativamente l'espressione di BCL2 (Figura 3). Abbiamo perciò studiato l'effetto di mir-15/16 sull'espressione di BCL2 impiegando una linea cellulare, MEG-01, in cui l'espressione dei miRNA è assente e BCL2 è espresso ad alti livelli. I risultati hanno dimostrato che i livelli di BCL2 si riducono in seguito al ripristino dell'espressione dei miRNA (Figura 3).

Inoltre, la costruzione di vettori esperimenti la luciferasi legata alla regione 3'UTR di BCL2 presentano una significativa riduzione dei livelli di attività in seguito a transfezione di mir-15/16, indicando che l'effetto di mir-15/16 su BCL2 sia direttamente imputabile alla interazione dei miRNA con sequenze presenti nella regione 3'UTR di BCL2. Infatti, mutazioni nella putativa sequenza bersaglio del 3'UTR abrogano l'effetto dei miRNA. Infine, l'effetto inibitorio non è da imputare a cambiamenti nei livelli di RNA messaggero, ma piuttosto ad inibizione

della traduzione. La inibizione di BCL2 nelle cellule MEG-01 in seguito a ripristino di mir-15/16 ha anche l'effetto di incrementare il numero di cellule in apoptosi, attraverso l'attivazione della via intrinseca (Figura 3). Come atteso, l'inoculazione in topi atimici di cellule transfettate con miR-15/miR-16 produce tumori significativamente più piccoli rispetto ai controlli. Questi risultati indicano che l'incremento dei livelli di BCL2 nella LLC è almeno in parte da imputare alla perdita di espressione di mir-15/16 e suggeriscono un modello di patogenesi riguardante la frazione maggiore di LLC, secondo cui la delezione del cromosoma 13q14, causando una ridotta espressione di mir-15/16, favorirebbe la sopravvivenza cellulare attraverso l'incremento di espressione di BCL2.

Poichè ciascun miRNA è in grado di regolare multipli trascritti, analisi trascrittomiche e proteomiche combinate con analisi "in silico" hanno permesso di identificare numerosi altri geni potenziali bersagli dell'azione di mir-15 e mir-16. Tra questi, gli oncosoppressori WT1 e PDCD4 e l'oncogene RAB9B. Questi risultati contribuiscono a decifrare i meccanismi molecolari coinvolti nella patogenesi della LLC ed hanno implicazioni importanti per lo sviluppo di strategie basate sull'impiego di mir-15 e mir-16 nella strategia terapeutica.

miRNA e prognosi nella LLC

Come precedentemente riportato, la LLC è una malattia eterogenea in cui si possono riconoscere almeno due gruppi distinti per la diversa prognosi. Marcatori predittivi di prognosi comprendono la presenza/assenza di mutazioni somatiche

Tab. 2 - Firma molecolare basata sull'espressione di microRNA associata a fattori prognostici (ZAP-70 e mutazioni IgVH) e progressione di malattia.

microRNA	Map	P value	Espressione in CLL aggressive
miR-15a	13q14.3	0,018	High
miR-195	17p13	0,017	High
miR-221	Xp11.3	0,010	High
miR-23b	9q22.1	0,009	High
miR-155	21q21	0,009	High
miR-223	Xq12-13.3	0,007	Low
miR-29a-2	7q32	0,004	Low
miR-24-1	9q22.1	0,003	High
miR-29b-2	1q32.2-32.3	0,0007	low
miR-146	5q34	0,0007	high
miR-16-1	13q14.3	0,0004	high
miR-16-2	3q26.1	0,0003	high
miR-29c	1q32.2-32.3	0,0002	low

nella regione variabile delle Ig e l'espressione differenziale di ZAP-70. Per meglio comprendere le basi molecolari di tale differenza abbiamo prima sviluppato e poi applicato ad una casistica di 94 LLC un microarray contenente le sonde per tutti i miRNA noti. Da questa analisi abbiamo dimostrato l'esistenza di una firma molecolare costituita dall'espressione di 13 miRNA in grado di discriminare i due gruppi di pazienti ZAP-70 + / IgVH nonmutate (malattia aggressiva) e ZAP-70-/IgVH mutate (malattia indolente) (Tabella 2).

Anche mir-15a e mir-16-1 sono presenti nella firma molecolare e presentano espressione più bassa nel gruppo con malattia indolente, a conferma del fatto che la delezione 13q14 è associata alla forma meno aggressive di LLC. La terapia della LLC è spesso posposta a quando la malattia diviene sintomatica. Dalla applicazione della firma molecolare in relazione al tempo di inizio della terapia, è stato possibile rilevare che il profilo di espressione associato a malattia aggressiva si correla anche a tempi brevi tra diagnosi ed inizio terapia.

Questi risultati indicano che una specifica firma molecolare basata sull'espressione di un subset di miRNA si associa alla prognosi ed alla rapidità nella progressione della malattia. Il riconoscimento dei diversi bersagli genici riconosciuti dai miRNA presenti nella firma potrebbero offrire indicazioni utili a comprendere meccanismi patogenetici differenzialmente espressi nei due diversi gruppi di LLC.

Il ruolo dei miRNA nella patogenesi di neoplasie ematologiche non si limita a mir-15a e mir-16-1 nella LLC, ma è stata provata anche per mir-155 la cui sovraespressione è stata riportata in linfomi di Hodgkin e svariati tipi di linfomi non-Hodgkin. È stato anche dimostrato in un modello animale transgenico che la sovraespressione di mir-155 nel comparto linfocitario B induce un'espansione clonale che sfocia in una leucemia costituita da cellule indifferenziate. Inoltre, il cluster di miRNA mir-17-92 coopera con l'oncogene MYC nello sviluppo di linfomi nel topo. In sintesi, il ruolo della deregolazione dei miRNA nella patogenesi neoplastica sta emergendo come uno dei meccanismi più frequenti. Il riconoscimento di quali siano i miRNA implicati e quali i bersagli genici potrebbe avere implicazioni nello sviluppo di strategie terapeutiche mirate.

FARMACOGENOMICA

La conoscenza delle vie di trasduzione del segnale come guida per lo sviluppo di nuove terapie molecolari

Giuseppe Saglio, Daniela Cilloni

Divisione di Medicina Interna ed Ematologia, Dipartimento di Scienze Cliniche e Biologiche dell'Università di Torino, Ospedale San Luigi, Orbassano (Torino)

Analogamente a quanto è avvenuto per la leucemia mieloide cronica con il Glivec, per la terapia di moltri altri tipi di leucemia e di tumore ci si avvia ad assistere ad una vera e propria rivoluzione determinata dallo sviluppo di nuovi farmaci, in parte alternativi alla tradizionale chemioterapia. Questi farmaci, che potremmo definire “intelligenti”, sono in grado di colpire in maniera molto più selettiva le cellule tumorali rispetto alle cellule sane, perché costruiti per inibire specificamente i difetti che determinano il comportamento tumorale e per inceppare le alterazioni “molecolari” dei geni alterati e dei loro prodotti che sono alla base del comportamento tumorale o neoplastico delle cellule. Per questo motivo questo tipo di terapia viene spesso riferita come “terapia molecolare”.

Questa nuova evoluzione della terapia antitumorale (e non solo) è stata resa possibile da una conoscenza più precisa e dettagliata dei geni e delle proteine coinvolte nei processi di proliferazione e di differenziazione. Questi complessi di processi biochimici e molecolari sono generalmente riferiti e denominati come “vie di traduzione del segnale”. Lo sviluppo di tutti gli organismi richiede infatti la perfetta esecuzione di programmi complessi nei quali diversi geni sono attivati o repressi in specifici tipi cellulari e con una precisa sequenza temporale. È intuitivo come questo processo debba essere regolato in modo molto preciso da un complesso network di segnali finemente regolati. Tuttavia, nonostante le accresciute conoscenze nel campo della biologia cellulare, il meccanismo coinvolto nel coordinamento dei diversi *pathways* non è completamente chiaro se non addirittura in alcuni casi completamente oscuro.

Le diverse modulazioni dell'espressione genica sono generate da segnali extracellulari che agiscono su specifici recettori di membrana. Molti di questi modulatori sono molecole solubili che agiscono in modo paracrina. Tuttavia alcune proteine coinvolte nella traduzione del segnale sono situate esse stesse alla superficie cellulare e interagiscono con recettori di superficie delle cellule adiacenti attivando segnali intracellulari. Anche le cellule che hanno ormai raggiunto lo stadio maturativo terminale cambiano costantemente il loro pattern di espressione genica poiché ricevono costantemente stimoli extracellulari che inducono l'at-

tivazione di specifici fattori trascrizionali che a loro volta stimolano o reprimono specifici geni target.

Esistono evidenze sempre più consistenti del ruolo chiave delle proteine tirosino-chinasi nell'alterazione dei processi di proliferazione e differenziazione che caratterizzano diverse neoplasie ematologiche tra cui le leucemie acute e croniche. Le tirosino-chinasi, insieme ai recettori delle citochine rappresentano le due importanti classi di recettori di membrana. I recettori delle citochine, che posseggono un domain citosolico associato a un membro della famiglia delle tirosino-chinasi citosoliche ovvero le chinasi JAK e i recettori tirosino-chinasi (RTKs) che posseggono attività tirosino chinasi nel domain citosolico posseggono un meccanismo di attivazione molto simile. In entrambi i casi il legame con il ligando favorisce la dimerizzazione del recettore, nella maggior parte dei casi presente in forma monometrica, che in questa forma attiva innesca un processo di autofosforilazione. Infatti nella forma monometrica inattiva l'attività TK del recettore è molto bassa. Al contrario, nella forma dimerica la chinasi di una subunità è in grado di fosforilare uno o più residui vicino al domain catalitico dell'altra subunità. La fosforilazione induce un cambio conformazionale che facilita il legame dell'ATP in alcuni recettori quali il recettore dell'insulina oppure o il legame di proteine substrato in recettori quali *FGF receptor*. Tutto questo induce la fosforilazione di una seconda tirosino-chinasi. Il processo di fosforilazione della tirosina nell'activation loop è un meccanismo comune attraverso il quale vengono attivate la maggior parte delle TK.

Molte proteine TK interagiscono con proteine adattatrici, ovvero delle piccole proteine che contengono dei domains SH2, PTB o SH3 ma che non posseggono intrinseca attività chinasi. Queste proteine inducono il cross-talk tra le TKs e altre proteine responsabili dell'attivazione di pathways che sfociano ad esempio nell'attivazione di Ras. Ras è una sorta di proteina interruttore che lega il GTP. Per essere attiva deve legare il GTP, al contrario nella forma legata al GDP è inattiva. Le proteine Ras assumono un ruolo centrale in campo oncologico poiché è noto che la presenza di mutazioni di Ras si associa allo sviluppo di neoplasie. Infatti le forme mutate non sembrano essere in grado di idrolizzare il GTP e rimangono pertanto costantemente attivate. In seguito all'attivazione delle RTK si forma sulla superficie citosolica della membrana plasmatica un complesso contenente il recettore attivato, e le molecole adattatrici GRB2 e Sos. Questo legame porta alla localizzazione di Sos dal citosol alla membrana vicino al suo substrato Ras-GDP. Il legame tra Sos e Ras-GDP porta all'attivazione di Ras, e al conseguente innesco di una cascata mediata dalla MAP chinasi attraverso l'attivazione di Raf, MEK, La chinasi MAP, una volta entrata nel nucleo è in grado di fosforilare molte proteine tra cui fattori trascrizionali che regolano l'espressione di geni coinvolti nei processi di differenziazione cellulare, crescita proliferazione e apoptosi.

Lo stimolo con alcuni fattori di crescita quali EGF o PDGF a cellule quiescenti in fase G₀ del ciclo cellulare provoca un rapido aumento dell'espressione di oltre 100 geni, i cosiddetti geni della risposta precoce. Tra questi i geni che codificano per i fattori trascrizionali c-jun e c-fos che inducono l'espressione di proteine necessarie alla progressione attraverso le diverse fasi del ciclo cellulare. Oltre alla

chinasi MAP appartengono alla superfamiglia delle MAP chinasi anche JNKs e P38 che, come MAP sono serino-treonino chinasi attivate da RTKs nel citosol in risposta a stimoli extracellulari e in grado di traslocare nel nucleo ad attivare la trascrizione genica. Oltre al segnale innescato dalla fosforilazione delle RTK, che sicuramente gioca un ruolo di primo piano nell'indurre la trascrizione genica, la cellula possiede altri meccanismi di regolazione della dell'attività trascrizionale in risposta a diversi segnali extracellulari. Tra questi il meccanismo di attivazione di NF-kB è sicuramente tra i più rilevanti.

NF-kB è un fattore trascrizionale nucleare che viene rapidamente attivato in risposta a infezioni, infiammazioni tramite il segnale indotto da citochine infiammatorie come TNF α e IL1, e in risposta a numerosi situazioni di stress come le radiazioni ionizzanti. Nella situazione basale NF-kB è sequestrato nel citosol complessato alla chinasi IKB che ne maschera per legame diretto il segnale di localizzazione nucleare (NLS). Il complesso IKB è un punto di convergenza di tutti i segnali extracellulari che attivano NF-kB. L'attivazione di IKB in seguito ai diversi stimoli porta alla fosforilazione di IKB stesso nei residui di serina 32 e 36 per opera della chinasi IKK rendendola pertanto accessibile alle ubiquitinoligasi, che legando le fosfoserine sono in grado di indurre la degradazione operata dal proteasoma di IKB stesso. La degradazione di IKB smaschera le sequenze di localizzazione nucleare di NF-kB che è così libero di entrare nel nucleo e di iniziare il processo di trascrizione genica. NF-kB stimola la trascrizione di oltre 150 geni che codificano per citochine, chemochine e per proteine coinvolte nella regolazione della proliferazione e apoptosi. Recentemente diversi studi hanno dimostrato un'incrementata attività di NF-kB nei blasti di leucemia acuta. Pertanto il pathway di NF-kB rappresenta un ottimo target per composti selettivi in grado di inibire a diversi livelli l'attività sregolata. In particolare sono già in fase di applicazione clinica gli inibitori del proteasoma.

Le numerose conoscenze dei difetti a livello di proteine coinvolte nella trasduzione del segnale che sono alla base di numerose neoplasie offrono la possibilità di realizzare un approccio terapeutico alternativo in grado di bloccare in modo selettivo il difetto molecolare o il pathway anomalo sregolato. In alcune neoplasie di origine epiteliale con coinvolgimento del recettore per EGF, l'inibitore selettivo di questa chinasi offre attualmente nuove opportunità terapeutiche. Allo stesso modo sono in sperimentazione gli inibitori di Ras che può essere bloccato mediante degli inibitori dell'enzima farnesil transferasi. Sicuramente l'esempio di terapia molecolare più riuscita è rappresentata dall'Imatinib in grado di bloccare l'attività TK di BCR-ABL nella leucemia mieloide cronica.

L'Imatinib mesilato è stato introdotto cinque anni fa nella terapia della LMC e si è subito rivelato un farmaco estremamente promettente. In pazienti trattati alla diagnosi, è stata osservata una risposta citogenetica completa in almeno il 70% dei casi, e nel 50% circa, si è ottenuta una buona remissione molecolare. L'attività dell'Imatinib è stata anche riscontrata in pazienti affetti da LMC in fase accelerata e in crisi blastica e da leucemie acute linfoblastiche (LLA) Ph-positive, ma con una più bassa incidenza di tutti i tipi di risposta e con una più elevata percentuale di ricaduta come è stato descritto in questi pazienti.

Le ricadute sono generalmente causate dallo sviluppo di cloni resistenti, e la resistenza all'Imatinib rappresenta ancora un problema significativo per una certa percentuale di pazienti affetti da LMC in fase cronica, in particolare per quelli con un profilo ad alto rischio Sokal. La resistenza si definisce acquisita quando si associa alla perdita della risposta ematologica o citogenetica precedentemente ottenuta, oppure primaria, quando non si è ottenuto alcun tipo di risposta all'imatinib. La resistenza di tipo acquisito si associa generalmente con la ripresa dell'attività tirosin-chinasica di BCR-ABL e può essere dovuta a diversi meccanismi: mutazioni di BCR-ABL che potrebbero interferire con il legame del farmaco al substrato, aumentata espressione di BCR-ABL, diminuzione "in vivo" della disponibilità o riduzione dei livelli intracellulari di Imatinib. Le mutazioni di BCR-ABL sembrano rappresentare il meccanismo di resistenza più comune e si possono dividere in due gruppi a seconda della regione di BCR-ABL in cui cadono; possono essere localizzate nel sito di legame dell'Imatinib (quali la T315I che rappresenta una delle mutazioni più frequenti), nel P-loop oppure nelle regioni regolatorie. È stato riportato che la presenza di mutazioni nel P-loop si associa ad una prognosi peggiore rispetto alle altre e, nella maggior parte dei pazienti in cui tali mutazioni insorgono, ad una progressione verso la crisi blastica che avviene poco tempo dopo l'identificazione della mutazione. Sembra che, almeno in alcuni casi, le mutazioni che conferiscono resistenza all'Imatinib siano già presenti prima del trattamento e selezionate dal farmaco. Altre evidenze in grado di supportare questa ipotesi giungono dall'osservazione che l'insorgenza di mutazioni che conferiscono resistenza sia correlata con la durata della malattia prima della terapia con Imatinib. Per quanto concerne gli altri meccanismi di resistenza, il legame dell'Imatinib all' γ 1 glicoproteina acida potrebbe ridurre la disponibilità del farmaco in vivo e aumentare i livelli di BCR-ABL a causa dell'amplificazione genica.

Le mutazioni di BCR-ABL generalmente non sono presenti nei pazienti con resistenza primaria, suggerendo l'esistenza di un differente meccanismo di resistenza. In questi casi, anomalie oncogeniche secondarie, potrebbero indurre l'attivazione di meccanismi BCR-ABL indipendenti capaci di inviare segnali proliferativi e antiapoptotici, riducendo in questo modo l'efficacia della terapia. È possibile che nuovi e più potenti inibitori di BCR-ABL come il dasatinib o il nilotinib e la terapia combinata di imatinib con composti in grado di alterare altri meccanismi molecolari significativi, possano superare eventualmente la resistenza (attualmente sono in corso diverse strategie in tale direzione). Con queste terapie ci si aspetta una minore incidenza di fenomeni di resistenza, ma ciò necessita di ulteriori conferme sperimentali. Un altro grande problema riguardante l'efficacia della terapia con l'Imatinib è rappresentato dall'assenza di una completa eradicazione delle cellule tumorali nella maggior parte dei casi di LMC in trattamento con l'Imatinib. Infatti, sebbene più del 50% dei pazienti ottenga una buona risposta molecolare e possa considerarsi guarito, solo in un 5-10% dei casi il clone Ph+ diventa non rivelabile in nested RT-PCR. Ciò rappresenta un grosso problema e compromette l'efficacia a lungo termine del trattamento con l'Imatinib e la possibilità di interrompere ad un dato momento la terapia. Le basi molecolari in grado di spiegare la persistenza di questa minima quota di malattia residua non

sono attualmente conosciute, ma esiste una sempre maggior evidenza che un certo numero di cellule staminali di LMC primitive e quiescenti non vengono eliminate in vivo dall'Imatinib. La percentuale di queste cellule CD34+ di LMC non proliferanti è all'incirca lo 0,5% dell'intero compartimento CD34+. Tali cellule sono in grado di attecchire in topi NOD/SCID e produrre una progenie leucemica dopo 6-8 settimane, suggerendo che la loro quiescenza sia in vivo un fenomeno temporaneo e reversibile e sottolineando l'importanza di questa popolazione come bersaglio terapeutico per eradicare in maniera definitiva la malattia.

Le attuali conoscenze sulla resistenza ai farmaci delle cellule staminali primitive di LMC sono limitate, ma sembra che tale fenomeno sia multifattoriale. Una migliore comprensione di questi meccanismi è di fondamentale importanza se si vuole trovare un modo più efficace per curare in maniera definitiva la LMC senza dover ricorrere a terapie come il trapianto di midollo osseo.

Sin dal momento in cui è stata scoperta, la LMC è stata considerata come un modello ideale per studiare la patogenesi dei processi neoplastici nell'uomo e il modo attraverso cui possiamo tentare di curarli. Le conoscenze dedotte dallo studio della LMC e del modo con cui l'imatinib agisce in questa leucemia stanno servendo da base per sviluppare molti concetti utili allo sviluppo e al miglioramento delle terapie molecularmente mirate.

Bibliografia

1. Kimura F. Molecular target drug discovery. *Intern Med.* 2007; 46(2): 87-9. Epub 2007 Jan 15.
2. He YD. Genomic approach to biomarker identification and its recent applications. *Cancer Biomark.* 2006; 2(3-4): 103-33.
3. Mujtaba S, Zeng L, Zhou MM. Modulating molecular functions of p53 with small molecules. *Cell Cycle.* 2006 Nov; 5(22): 2575-8. Epub 2006 Nov 15. Review.
4. Rodriguez-Nieto S, Zhivotovsky B. Role of alterations in the apoptotic machinery in sensitivity of cancer cells to treatment. *Curr Pharm Des.* 2006; 12(34): 4411-25.
5. Dolle L, Depypere HT, Bracke ME. Anti-invasive/anti-metastasis strategies: new roads, new tools and new hopes. *Curr Cancer Drug Targets.* 2006 Dec; 6(8): 729-51.
6. Ritchie E, Nichols G. Mechanisms of resistance to imatinib in CML patients: a paradigm for the advantages and pitfalls of molecularly targeted therapy. *Curr Cancer Drug Targets.* 2006 Dec; 6(8): 645-57.
7. Hwang JJ, Ghobrial IM, Anderson KC. New frontiers in the treatment of multiple myeloma. *Scientific World Journal.* 2006 Dec 6; 6: 1475-503.
8. Iliopoulos O. Molecular biology of renal cell cancer and the identification of therapeutic targets. *J Clin Oncol.* 2006 Dec 10; 24(35): 5593-600.
9. 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 Dec 7; 355(23): 2408-17.

Leucemia mieloide cronica: un modello di terapia molecolare

Carlo Gambacorti-Passerini

Sezione di Ematologia, Ospedale San Gerardo (Monza), Università di Milano Bicocca

Gli ultimi cinque-dieci anni hanno visto un cambiamento radicale nella terapia e prognosi della Leucemia Mieloide Cronica (LMC). Un miglioramento di questa portata, ottenuto grazie all'utilizzo delle conoscenze sulla patogenesi molecolare della malattia, non era mai stato ottenuto in passato, ed è stato paragonato alla scoperta degli antibiotici per il trattamento delle malattie causate da batteri (1). Per questa ragione la maggior parte di questo articolo sarà incentrato sulla terapia della LMC, mentre un capitolo finale descriverà i progressi nel campo di altre malattie neoplastiche di tipo ematologico.

LMC: da malattia mortale a malattia curata

La LMC è un disordine mieloproliferativo derivante dalla trasformazione neoplastica delle cellule ematopoietiche staminali. La malattia è in origine monoclonale e può coinvolgere le linee mieloide, monocitica, eritroide, megacariocitica e talora anche il compartimento linfoide; le cellule stromali del midollo non sono mai coinvolte. La LMC è una delle prime malattie per cui sia stato possibile individuare una specifica anomalia cromosomica quale causa della malattia: il cromosoma Filadelfia (o Ph) così chiamato in onore della città americana (Philadelphia in inglese) in cui venne scoperto. Il cromosoma Ph (Figura 1) è il risultato della traslocazione del proto-oncogene Abelson (ABL) dal cromosoma 9 ad una regione del cromosoma 22 denominata "regione di raggruppamento dei punti di rottura", breakpoint cluster region (BCR) in inglese, con la formazione di un gene chimerica BCR/ABL. Tale difetto cromosomico è presente in circa il 95% di tutti i pazienti con LMC ma anche nel 30-50% dei pazienti adulti con leucemia linfoblastica acuta (ALL). Nella proteina di fusione risultante dalla traslocazione, parte del prodotto del gene ABL (Abl), è fuso alla porzione aminoterminale del prodotto del gene BCR (Bcr). Esistono due differenti forme di BCR/ABL (Figura 2), una del peso molecolare di 210 kDa associata alla LMC e alla leucemia linfatica acuta (LLA) cr. Ph⁺, ed una del peso molecolare di 190 kDa associata principalmente alla LLA Ph⁺. Il differente peso molecolare deriva dal diverso numero di aminoacidi codificati dal gene BCR che entrano nella struttura della molecola.

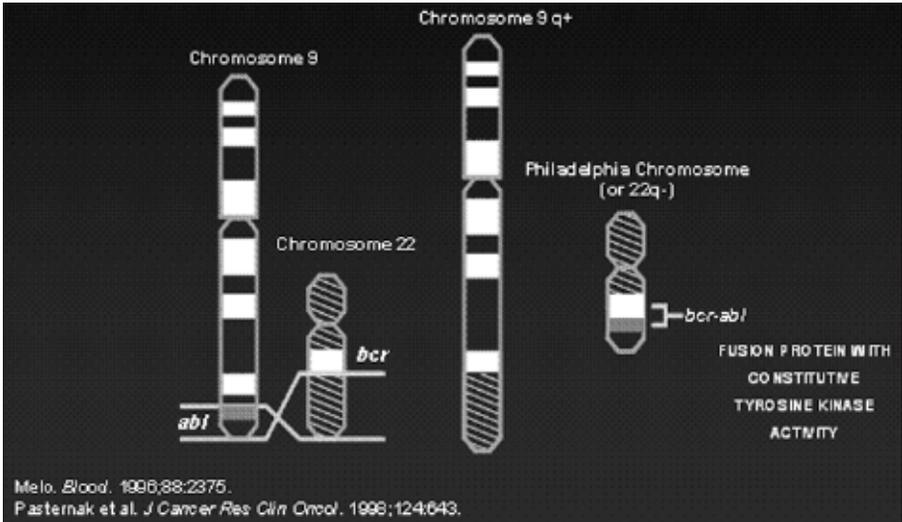


Fig. 1 - The Ph Chromosome and the *bcr-abl* Gene: The t(9;22) Translocation.

Nella parte di Abl che partecipa alla fusione è contenuto un dominio ad attività tirosino chinasi, cioè con capacità di fosforilare (cioè di trasferire gruppi fosfato donati da una molecola di ATP) residui di tirosine appartenenti a proteine in grado di legarsi ad Abl e per questo detti substrati.

Il risultato di questa fusione è quello di rendere l'attività chinasi di Bcr/Abl (ereditata da Abl) costitutiva; in altre parole l'attività tirosinico chinasi di Bcr/Abl non è più regolata, a differenza che nella proteina normale Abl. Inoltre Bcr/Abl viene anche delocalizzata; mentre Abl normale è una proteina nucleare,

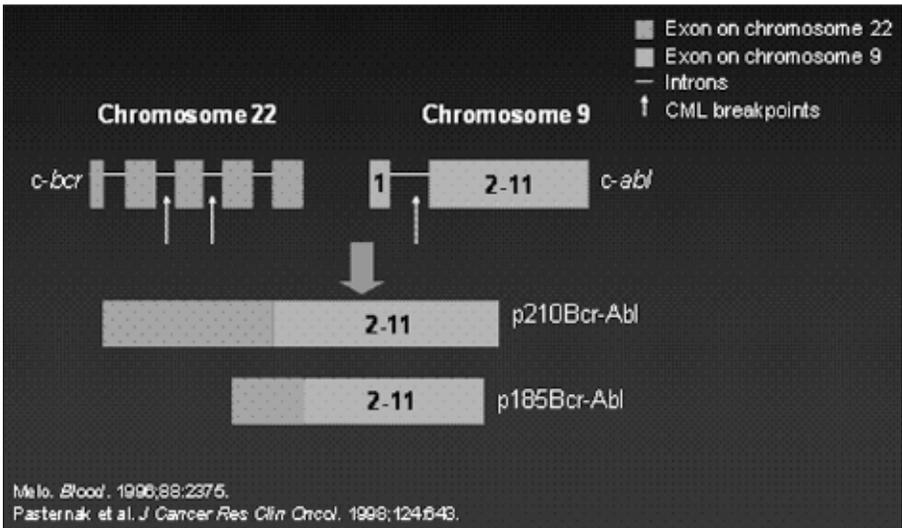


Fig. 2 - The Ph Chromosome and the *bcr-abl* Gene: *bcr-abl* Gene Structure.

Bcr/Abl è esclusivamente localizzata nel citoplasma. La deregolazione e la delocalizzazione dell'attività chinasica di Bcr/Abl portano all'attivazione inappropriata di vie di trasduzione del segnale quali le vie di Ras/Raf, PI3 chinasi/Akt e all'aumento dei livelli di Bcl2; questi fenomeni trasferiscono alla cellula Bcr/Abl+ forti segnali di tipo proliferativo ed antiapoptotico, mentre la differenziazione cellulare non è inizialmente compromessa (2). La proteina Bcr/Abl è stata una delle prime tirosino chinasi ad essere chiaramente associata con una neoplasia umana. L'attività chinasica di Bcr/Abl è indispensabile affinché questa possa svolgere un ruolo patogenetico; mutazioni di Bcr/Abl che ne inattivano l'attività chinasica infatti inibiscono anche la crescita neoplastica.

Da un punto di vista clinico il decorso della malattia si divide (o si divideva) in tre fasi: una fase cronica, della durata di circa 5 anni, che attraverso l'accumulo di ulteriori mutazioni o altre lesioni genetiche evolveva in una fase intermedia di accelerazione e in una di trasformazione in leucemia acuta (crisi blastica), invariabilmente fatale (2).

Sulla base di queste conoscenze ci si può aspettare quindi che una interferenza farmacologica con l'attività tirosino chinasica di Bcr/Abl possa avere un importante effetto sulla regolazione delle attività biologiche delle cellule leucemiche cr. Ph+. Bcr/Abl è una tirosino chinasi intracellulare non recettoriale; dato che non risiede sulla membrana cellulare e non può essere quindi bloccata da anticorpi, sono necessarie piccole molecole in grado di superare la membrana cellulare e di legarsi al sito attivo dell'enzima.

Imatinib (precedentemente noto come CGP57148B e STI571) è stato sviluppato per inibire l'attività enzimatica di Bcr/Abl. Imatinib rappresenta il primo farmaco disegnato per bloccare specificamente una proteina che causa un tumore umano; anche per questo motivo i risultati e le diverse vicende legate alla sua sperimentazione hanno ottenuto un così grande risalto (1).

È difficile riassumere in poche pagine quasi dieci anni di ricerche precliniche prima e cliniche poi, che hanno cambiato il modo di concepire e sviluppare nuovi farmaci in oncologia e in ematologia.

Studi preclinici

Imatinib agisce legando una tasca presente in Bcr/Abl e bloccandone l'attività chinasica (Figura 3) (3). Si era inizialmente ritenuto che imatinib inibisse l'attività enzimatica di Bcr/Abl legando la proteina nel sito ove va ad alloggiarsi la molecola di ATP donatrice del gruppo fosfato, competendo pertanto con l'ATP stesso. Ulteriori ricerche (4) hanno invece mostrato chiaramente che imatinib lega una regione della proteina solo in parte corrispondente a quella dell'ATP e che soprattutto imatinib lega Bcr/Abl in una conformazione "inattiva", incapace di funzionare proprio perché in quella conformazione la chinasi è incapace di legare ATP. Va precisato a questo proposito che tutte le quasi tutte le chinasi ciclano continuamente tra una conformazione attiva o "aperta", in grado cioè di permettere l'alloggiamento dell'ATP nella propria tasca, ed una inattiva o "chiusa" nella quale la tasca per l'ATP non è disponibile (3).

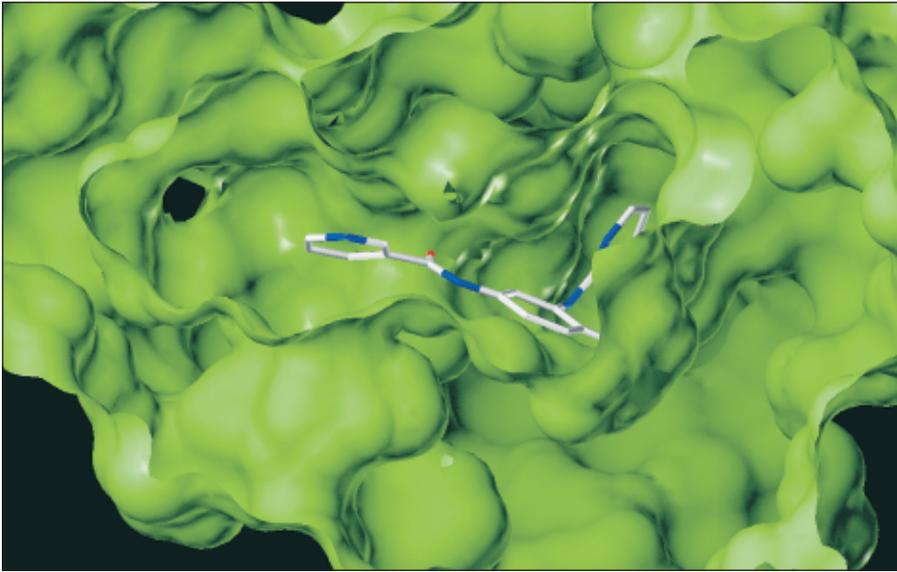


Fig. 3 - Imatinib (bianco e azzurro) inserito nella tasca di bcr/Abl (verde). Da referenza #3.

Tutto ciò non era ovviamente noto nel 1996-7, quando le prime ricerche su questa molecola iniziarono in tre laboratori nel mondo, uno dei quali in Italia. Dati ottenuti tra il 1996 e l'inizio del 1999 permisero di comprendere che imatinib aveva tutte le caratteristiche per poter diventare un farmaco innovativo (5-9). Queste ricerche infatti dimostrarono quattro importanti punti:

- era possibile inibire la proliferazione di cellule leucemiche Ph+, sia ottenute da linee cellulari che direttamente da pazienti, in maniera specifica e a concentrazioni nanomolari, raggiungibili in vivo; l'attività su cellule normali era minima;
- l'attività biologica correlava strettamente con il grado di inibizione dell'attività enzimatica di Bcr/Abl;
- era possibile trasferire "in vivo" (cioè in modelli animali) questi risultati ottenuti "in vitro" cioè in provetta, a patto che la somministrazione di imatinib fosse tale da garantire una inibizione continua dell'attività chinasi di Bcr/Abl; in queste condizioni era possibile eradicare la malattia in percentuali comprese tra 87% e 100% degli animali trattati;
- l'inibizione di bcr/Abl non solo induceva blocco proliferativo ma anche morte delle cellule leucemiche (apoptosi).

Quest'ultimo punto merita una ulteriore discussione. Teoricamente il blocco dell'attività di Bcr/Abl in una cellula staminale dovrebbe riportare la cellula a funzionare come una cellula normale. Lo sviluppo di apoptosi rappresenta un fenomeno non inquadrabile in questo schema e può essere spiegato solamente col fatto che già alla diagnosi le cellule di LMC (che hanno iniziato a moltiplicarsi da 2-5 anni) contengono già ulteriori lesioni geniche, e che esse sopravvivono

all'attività proapoptotica indotta da molte di queste lesioni addizionali solamente grazie all'attività anti-apoptotica di Bcr/Abl, bloccata la quale l'apoptosi diviene inevitabile.

Studi clinici

Per meglio comprendere la portata dei cambiamenti indotti da imatinib nella terapia e prognosi della LMC, è utile riassumere brevemente quali sono stati gli approcci terapeutici utilizzati per la LMC nel passato e con quali risultati.

Nella seconda metà dell'800, contestualmente alla definizione clinica della malattia (10), venne inizialmente proposto l'uso degli arsenicali (soluzione di Flower, 1856), che continuò fino agli anni '50 del XIX secolo. Nei primi decenni del '900 anche la terapia radiante sistemica o splenica si rivelò efficace nel controllare i segni e i sintomi della malattia rappresentando il trattamento di riferimento per i successivi anni. Negli anni '50 fu dimostrato che la chemioterapia con agenti alchilanti orali risultava più efficace della terapia radiante, e farmaci come il busulfan diventarono il caposaldo della terapia. L'idrossiurea fu introdotta negli anni '70 e sostituì il busulfan per la sua maggior maneggevolezza nel controllo della soppressione midollare e per la mancanza di tossicità polmonare. Negli anni '80 fu sperimentato nei pazienti affetti da LMC l'uso di interferone-alfa (INFa) e per tutta la decade passata schemi contenenti INFa hanno costituito la terapia di prima linea per la maggior parte dei pazienti (11). L'uso dell'INFa induce una risposta ematologica completa nell'50-70% dei pazienti ed una risposta citogenetica (definita come una % di cellule midollari Ph+ inferiore al 35%) nel 10-20% dei casi. Questo significa che sia l'INFa, sia tutti i farmaci che l'avevano preceduto non erano in grado di ripristinare una emopoiesi normale nella maggior parte dei pazienti trattati. Inoltre la terapia con INFa è gravata da alcuni effetti collaterali anche importanti: sintomi simil influenzali, calo ponderale e, meno frequentemente, diarrea, stomatite e neurotossicità che si presenta con depressione e deficit mnesici (11). Gli agenti alchilanti e l'INFa non hanno comunque un effetto curativo sulla LMC, riuscendo tutt'al più a ritardare di alcuni mesi l'inevitabile evoluzione verso la crisi blastica. Fino ad oggi le speranze di un approccio eradicante sono state riposte nel trapianto allogenico di cellule midollari. Questa metodica è però gravata da tossicità gravi, anche mortali ed anche a lungo termine; ad esempio il rischio acuto (30-60 giorni) di mortalità legata al trapianto varia dal 10% al 30-40% a seconda di vari fattori quali l'età del paziente, la fase della malattia, il tipo di trapianto (da donatore consanguineo o no). Inoltre questa terapia è proponibile, data la tossicità insita nella terapia, solo a pazienti giovani, di età inferiore a 50-55 anni nel caso di trapianto da familiare e a 40-45 anni nel caso in cui il donatore non sia consanguineo, e nei casi in cui sia presente un donatore compatibile. Tenendo conto anche che l'età media di insorgenza della LMC è di 45-50 anni, risulta evidente come questa terapia sia comunque proponibile solo ad una minoranza di pazienti. Nel complesso il trapianto di midollo nella LMC si è rivelato efficace solo per alcuni pazienti ed il suo utilizzo è rimasto limitato dalla disponibilità di donatori compatibili, dalla

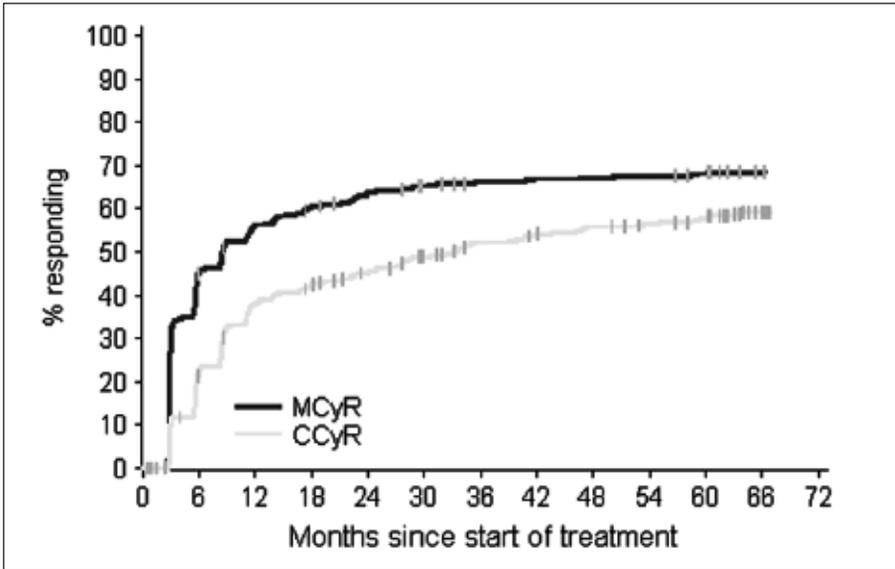


Fig. 4 - Proporzione di pazienti in risposta citogenetica Maggiore (MCyR) o completa (CCyR). Da referenza #13.

tossicità e dagli effetti collaterali importanti anche a lungo termine. I farmaci disponibili all'inizio di questo secolo per la terapia di questa patologia si rivelavano quindi, per motivi connessi con il loro stesso meccanismo d'azione, scarsamente efficaci e gravati da un'alta incidenza di effetti collaterali e di tossicità.

I primi studi clinici con imatinib iniziarono nella seconda metà del 1998 negli USA e nella prima metà del 1999 in Italia e in altri paesi europei.

I primi risultati non tardarono a confermare le aspettative sorte dopo i tests in laboratorio: nella maggior parte dei pazienti trattati i valori dell'emocromo si normalizzavano, se elevati, entro 2-4 settimane. Ma il risultato più importante arrivò dopo alcuni mesi. Infatti già a tre-sei mesi dall'inizio della terapia circa la metà dei pazienti mostrava una risposta citogenetica, generalmente completa (12). Questo risultato non era mai stato raggiunto nei precedenti 150 anni di tentativi. La percentuale di pazienti in risposta citogenetica saliva poi ulteriormente nei successivi mesi fino a raggiungere un valore di circa 70% (Figura 4) (13).

Va inoltre ricordato che questi pazienti, seppure in fase cronica, non erano trattati alla diagnosi, bensì dopo diversi anni (mediamente 3) dalla diagnosi e dopo aver fallito la terapia con interferon. Uno studio effettuato successivamente ha esaminato pazienti trattati con imatinib alla diagnosi e confrontati con un gruppo di controllo trattato con interferon: in questa popolazione la percentuale di risposte citogenetiche arriva fino al 85% (13), all'incirca 10 volte più alto di quanto raggiunto con la miglior terapia precedente. Le risposte ottenute con imatinib hanno anche retto alla prova del tempo: infatti a 5 anni di distanza dall'inizio della terapia nel primo protocollo per cui esistono dati disponibili, oltre 80% dei pazienti [83% per la precisione(13)] che avevano ottenuto risposta cito-

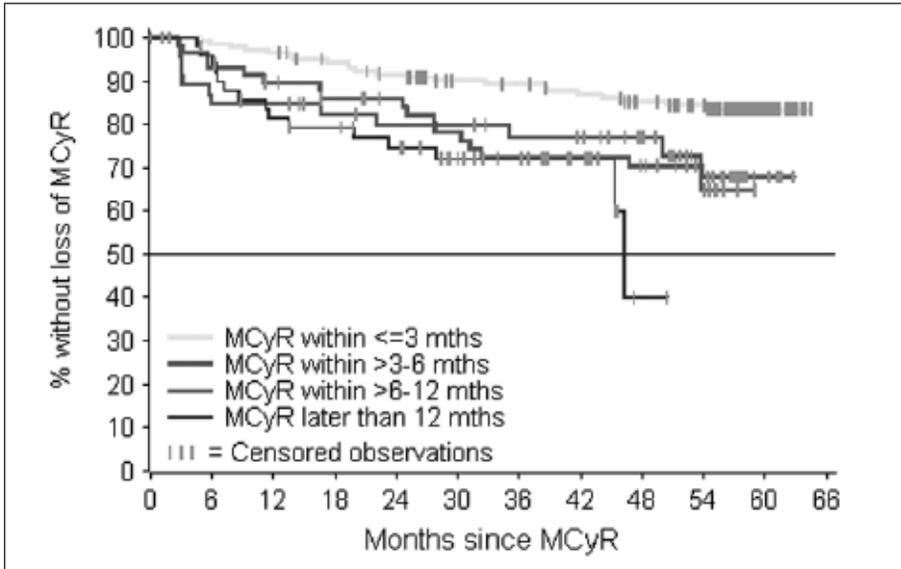


Fig. 5 - Durata della risposta citogenetica in funzione del tempo al quale essa viene ottenuta. Da referenza #13.

genetica nei primi 3-6 mesi di trattamento mantiene tale risposta (Figura 5), e il 94% dei pazienti in risposta citogenetica completa è vivo (13). Questi risultati vanno valutati considerando che questi pazienti avevano, in base al decorso della malattia in era pre-imatinib, una aspettativa di vita non superiore a 2 anni. È anche possibile calcolare il rischio annuo di progressione della malattia, che è compreso tra 1 e 2% (Tabella 1) nei pazienti con risposta citogenetica completa. Studi successivi effettuati in pazienti trattati con imatinib alla diagnosi hanno confermato questi risultati, mostrando addirittura una diminuzione del rischio annuo di recidiva al di sotto dell'uno per cento dopo i primi tre anni di terapia (14) (Tabella 2).

Gli effetti collaterali causati da imatinib sono risultati assai contenuti: essi consi-

Tab. 1 - Rischio di progressione in funzione degli anni di trattamento. Da osservare in particolare la colonna 5 (pazienti con remissione citogenetica a 3 mesi). Da referenza #13.

Year since start of treatment	All (n=454)	MCyR at any time (n=304)	No MCyR at any time (n=150)	MCyR at 3 mths (n=148)	No MCyR at 3 mths (n=295)
1st	7.8%	2.6%	18.6%	2.7%	8.9%
2nd	6.0%	0.3%	21.1%	0.7%	8.9%
3rd	7.2%	3.6%	20.6%	2.3%	10.2%
4th	7.3%	4.0%	24.7%	2.5%	10.6%
5th	7.3%	4.4%	28.1%	1.8%	11.6%

Tab. 2 - Rischio di morte da qualsiasi causa e di progressione della malattia in funzione degli anni di trattamento in pazienti trattati alla diagnosi. Da referenza #14.

Annual Event rates on First-line Imatinib		
Year	All events*	AP/BC
1st	3.3%	1.5%
2nd	7.5%	2.8%
3rd	4.8%	1.6%
4th	1.5%	0.9%
5th	0.9%	0.6%

stono principalmente in edemi dovuti ad alterazioni nella permeabilità dei piccoli vasi, crampi muscolari e rash cutanei. L'entità di questi effetti collaterali è però assai contenuta e meno del 5% dei pazienti mostra tossicità di grado "medio o severo". Altri effetti meno frequenti sono la comparsa di epatotossicità (in genere temporanea), ginecomastia secondaria a diminuita produzione di testosterone (15), riduzione della pigmentazione cutanea.

Risultati meno positivi (16) si ottengono invece in pazienti trattati in fase di leucemia acuta (LMC in crisi blastica e LLA Ph+): mentre i tassi di risposta iniziale sono simili a quelli riportati qui sopra, la durata di queste risposte è in genere limitata ad alcuni mesi, non più di sei in genere, e seguita dallo sviluppo di resistenza a imatinib. Il fenomeno della resistenza ha rappresentato una sfida notevole dal punto di vista scientifico, e nel giro di pochi anni sono stati individuati i principali meccanismi che causano resistenza: amplificazione genica o mutazioni di BCR/ABL (3). In altre parole, la centralità di Bcr/Abl in questa patologia si riconferma anche per ciò che riguarda lo sviluppo di resistenza: si selezionano cioè cellule che producono maggiori quantità di proteina Bcr/Abl o che ne producono di un tipo modificato, contro la quale imatinib non è più attivo.

Va però detto che attualmente il problema della resistenza a imatinib è molto diminuito in frequenza per quanto riguarda la LMC, dato che i pazienti vengono trattati alla diagnosi, in fase cronica. Esso rappresenta invece un problema molto importante nella gestione dei pazienti con LLA Ph+. In questi ultimi l'uso di imatinib alla diagnosi (invece che alla recidiva) sembra consentire risultati migliori (17). Inoltre la disponibilità di nuove molecole (vedi sotto) sarà sicuramente di grande utilità nella gestione di questi pazienti.

Problematiche ancora aperte

- *I pazienti con LMC sono guariti dalla terapia con imatinib?* L'analisi citogenetica del midollo osseo ha una sensibilità tra il 5 e il 10%; ciò significa che per contaminazioni leucemiche inferiori a questi valori, l'esame risulta negativo pur in presenza di cellule leucemiche residue. Lo sviluppo di metodiche quali la PCR, con sensibilità assai maggiore (intorno a 1/100.000) ha permesso di rilevare come la gran parte dei pazienti in remissione citogenetica com-

pleta contenga ancora cellule leucemiche, usualmente in proporzione tra 1/100 e 1/10000. Questi dati da un lato hanno permesso di rifinire ulteriormente la prognosi (un paziente con valori di PCR inferiori a 1/1000 ha un rischio annuo di progressione ancora piú ridotto, attorno a 4-5%), dall'altro rende i pazienti "non guariti" da un punto di vista molecolare e pertanto con necessit  di continuare la terapia per una durata indefinita. Due considerazioni meritano di essere fatte a questo proposito:

- Considerando l'et  media di insorgenza della malattia ed il rischio annuo di progressione (approssimabile a 1%), si pu  concludere che la maggior parte dei pazienti rientrer  nella curva di sopravvivenza della popolazione generale prima di aver avuto una recidiva della malattia; in altre parole la maggior parte dei pazienti morir  per altre cause, ma non per la LMC. In altre parole la LMC non   ancora guarita molecolarmente ma lo   probabilmente da un punto di vista funzionale e pratico per la maggior parte dei pazienti, richiedendo solo una terapia continua ed attenta, come ad esempio avviene per l'ipertensione arteriosa. Queste considerazioni dovrebbero aiutare a togliere alla LMC quella stimate di "malattia probabilmente mortale" che sicuramente non le si addice piú.
- I dati noti a tutt'oggi devono essere considerati ancora provvisori: non   noto per esempio se la continuazione della terapia negli anni porti ad un progressivo abbassamento dei valori di PCR, come risultati recenti sembrano suggerire (18).

Questo potrebbe significare l'ottenimento, anche se in tempi lunghi, di risposte molecolari. Inoltre   importante sottolineare che nuovi inibitori di Bcr/Abl (Dasatinib, AMN107, SKI-606) sono ora in sperimentazione clinica (limitatamente ai pazienti resistenti a imatinib); queste molecole si sono dimostrate 30-300 volte piú potenti di imatinib in laboratorio ed in grado di legare conformazioni di Bcr/Abl diverse da quella a cui imatinib si complessa (19).   possibile ipotizzare che il loro utilizzo possa ottenere una inibizione piú completa di Bcr/Abl e pertanto possa portare ad una maggiore eliminazione delle cellule neoplastiche nei pazienti.

Conclusioni

Lo sviluppo di imatinib per il trattamento della LMC rappresenta sicuramente il maggior progresso ottenuto in oncologia ed ematologia negli ultimi decenni.   possibile ottenere risultati terapeutici importanti e duraturi in assenza di effetti collaterali importanti, cambiando cos  la storia biologica, clinica e sociale della malattia. Le conseguenze di questo successo travalicano comunque il campo specifico della malattia ed abbracciano piú in generale lo sviluppo di nuovi farmaci in oncoematologia. Il messaggio che emana da questi studi, che ormai appartengono alla storia della medicina,   chiaro:   possibile replicare questo successo a patto che due condizioni essenziali siano soddisfatte.   indispensabile ottenere conoscenze dettagliate circa i meccanismi di patogenesi molecolare, ed   altret-

tanto indispensabile, una volta conosciuti i meccanismi patogenetici, sviluppare farmaci diretti contro molecole che *causano* un determinato tumore (meglio se coinvolte nella patogenesi iniziale del tumore) e non sono semplicemente associate alla sua presenza.

Bibliografia

1. M. Lemonick, A Park, D. Cray, C Gorman. New hope for cancer. Time, 28/05/2001.
2. Goldman JM, Melo JV. Chronic myeloid leukemia-advances in biology and new approaches to treatment. N Engl J Med, 2003; 349: 1451-1464.
3. Gambacorti-Passerini CB, Gunby RH, Piazza R, Galiotta A, Rostagno R, Scapozza L. Molecular mechanisms of resistance to imatinib in Philadelphia-chromosome-positive leukaemias. Lancet Oncol, 2003; 4: 75-85.
4. Schindler T, Bornmann W, Pellicena P, Miller WT, Clarkson B, Kuriyan J. Structural mechanism for STI-571 inhibition of abelson tyrosine kinase. Science 2000; 289(5486): 1938-42.
5. Druker BJ, Tamura S, Buchdunger E, et al. Effects of a selective inhibitor of the Abl tyrosine kinase on the growth of Bcr-Abl positive cells. Nat Med 1996; 2(5): 561-6.
6. Gambacorti-Passerini C, le Coutre P, Mologni L, et al. Inhibition of the ABL kinase activity blocks the proliferation of BCR/ABL+ leukemic cells and induces apoptosis. Blood Cells Mol Dis 1997; 23(3): 380-94.
7. Deininger MW, Goldman JM, Lydon N, Melo JV. The tyrosine kinase inhibitor CGP57148B selectively inhibits the growth of BCR-ABL-positive cells. Blood 1997; 90(9): 3691-8.
8. Le Coutre P, Mologni L, Cleris L, et al. In vivo eradication of human BCR/ABL-positive leukemia cells with an ABL kinase inhibitor. J Natl Cancer Inst 1999; 91(2): 163-8.
9. Sausville E: A BCR/ABL kinase antagonist for chronic myelogenous leukemia: a promising path for progress emerges. J. Natl. Cancer Inst., 91: 102-103, 1999.
10. Donne, A. (1844) Cours de microscopie complementaire des etudes medicales. Balliere, Paris, 135, 196. Craigie, D. (1845) Case of disease of the spleen in which death took place from suppuration of the blood. Edinburgh Medical and Surgical Journal, 64, 400±412. Bennett, J.H. (1845) Case of hypertrophy of the spleen and liver in which death took place from suppuration of the blood. Edinburgh Medical and Surgical Journal, 64, 413±423. Virchow, R. (1845) Weisses Blut. Frorieps Notizen, 36, 151±156. Virchow, R. (1856) Die Leukaemie. In: Gesammelte Abhandlungen Zur Wissenschaftlichen Medizin, pp. 190-211. Meidinger, Frankfurt.
11. Jorge E. Cortes, Moshe Talpaz, Hagop Kantarjian. Chronic myelogenous leukemia: a review The American Journal of Medicine 1996; 100: 555-570.
12. HM Kantarjian, CL Sawyers, A Hochhaus, F Guilhot, CA Schiffer, C

- Gambacorti-Passerini, et al. Gleevec™ (imatinib mesylate) Induces Hematologic and Cytogenetic Responses in the Majority of Patients with Chronic Myeloid Leukemia in Chronic Phase: Results of a Phase II Study. *New Engl. J. Med.*, 2002; 346: 645-652.
13. Gambacorti C, Talpaz M, Sawyers C et al. Five Year Follow-Up Results of a Phase II Trial in Patients with Late Chronic Phase (L-CP) Chronic Myeloid Leukemia (CML) Treated with Imatinib Who Are Refractory/Intolerant of Interferon-. *Blood (ASH Annual Meeting Abstracts)*, Nov 2005; 106: 1089.
 14. Druker BJ, Guilhot F, O'Brien SG. et al. Five-Year Follow-up of Patients Receiving Imatinib for Chronic Myeloid Leukemia. *New Engl. J. Med.*, 355, 2408-2417, 2006.
 15. C Gambacorti-Passerini, L Tornaghi, F Cavagnini, P Rossi, F Pecori-Giraldi, L Mariani, N Cambiaghi, E Pogliani, G Corneo, L Gnessi. Gynecomastia in men with chronic myeloid leukemia after imatinib. *Lancet*, 361, 1954-56, 2003.
 16. Sawyers CL, Hochhaus A, Feldman E, Goldman JM, Miller CB, Ottmann OG, Schiffer CA, Talpaz M, Guilhot F, Deininger MW, Fischer T, O'Brien SG, Stone RM, Gambacorti-Passerini CB et al. Imatinib induces hematologic and cytogenetic responses in patients with chronic myelogenous leukemia in myeloid blast crisis: results of a phase II study. *Blood*. 2002 May 15; 99(10): 3530-9.
 17. Yanada M, Takeuchi J, Sugiura I, Akiyama H, Usui N, Yagasaki F, Kobayashi T, Ueda Y, Takeuchi M, Miyawaki S, Maruta A, Emi N, Miyazaki Y, Ohtake S, Jinnai I, Matsuo K, Naoe T, Ohno R; Japan Adult Leukemia Study Group. High complete remission rate and promising outcome by combination of imatinib and chemotherapy for newly diagnosed BCR-ABL-positive acute lymphoblastic leukemia: a phase II study by the Japan Adult Leukemia Study Group. *J Clin Oncol*. 2006 Jan 20; 24(3): 460-6.
 18. Goldman JM, Hughes T, Radich J et al. Continuing Reduction in Level of Residual Disease after 4 Years in Patients with CML in Chronic Phase Responding to First-Line Imatinib (IM) in the IRIS Study. *Blood (ASH Annual Meeting Abstracts)*, Nov 2005; 106: 163.
 19. Gambacorti-Passerini C, Gasser M, Assouline S, Scapozza L. BMS-354825 binding mode in Ablason kinase revealed by molecular docking studies. *Leukemia*, 19, 1267-1269, 2005.

New targeted therapy for acute myeloid leukemia

Giovanni Martinelli, Stefania Paolini, Michela Rondoni, Claudio Laterza, Cristina Papayannidis, Lea Verlicchi, Francesco De Rosa, Simona Soverini, Ilaria Iacobucci, Pier Paolo Piccaluga

Istituto di Ematologia e Oncologia Medica, Policlinico Sant'Orsola, Università di Bologna

In the past 40 years, advances in supportive care and development of chemotherapeutic agents have led to improved outcomes in patients with acute myeloid leukemia. High relapse rates following remission have led to extensive efforts to develop techniques and regimens for detecting and eliminating minimal residual disease. However, the best post-remission therapy has not been identified. Better understanding of the biology and the molecular pathogenesis of AML has led to the development of new, more specific agents and strategies for AML treatment. In particular, in this setting, some categories of drugs have been specially developed, such as tyrosin-kinase inhibitors, farnesyl-transferase inhibitors, and monoclonal antibodies.

FLT3 is a tyrosin-kinase which is mutated in almost 30% of AML patients, thus acquiring constitutive aberrant activity. Specific FLT3 inhibitors have been tested in clinical trials showing promising results.

Farnesyl-transferases are relevant enzymes in cell biology as they are necessary for activating key proteins such as RAS. The inhibition of such enzymes has been demonstrated to be able to induce leukemic cell apoptosis. Phase I-II clinical trials have candidate R115777 (Zarnestra) as promising agent in AML treatment, especially in elderly, frail patients.

Finally, monoclonal antibodies (MoAb) have been recently used as monotherapy or in combination with chemotherapy in AML treatment. In particular, Mylotarg, an anti-CD33 MoAB conjugated to calicheamicin (a potent antibiotic derivate), has been approved for the treatment of relapsed patients as single agent. In addition, it was recently demonstrated to be able to improve disease free survival in younger patients when associated with chemotherapy as first line therapy.

In conclusion, targeted therapy has improved outcomes in some patients. Most of the new agents are less toxic than their predecessors, and they can be used in combination with the more intensive traditional regimens.

DNA methylation as therapeutic target in myelodysplastic syndromes

Valeria Santini

Unità Funzionale Ematologia, Azienda Ospedaliero-Universitaria Careggi,
Università di Firenze

DNA mutations are often found in cancer cells, but only in few cases (i.e. chronic myeloid leukaemia) have they been shown to be the only cause of the neoplastic transformation. More often, several subsequent damages of DNA lead to cancer. It is extremely difficult to target pharmacologically DNA mutations, and therefore it is frequently impossible to modify the DNA altered sequence present in neoplastic cells. Epigenetic events regulate gene expression, and alterations in these mechanisms are frequently present in cancer. Epigenetic modifications are potentially reversible DNA and chromatin modifications transmitted from a cell to its progeny, able to modulate gene expression without changing DNA sequence and without any new genetic information. Epigenetic alterations are frequently involved in cancer and are ideal candidates for therapy as we may modify them without interfering with DNA information of the transformed cell.

Chromatin remodelling by nucleosome re-organization at the site of promoter genes enables transcriptional regulation. Histone acetylation and DNA methylation are the two best characterized epigenetic modifications, to which histone methylation, phosphorylation and ubiquitination must be added. Histone acetylation and DNA methylation are interdependent and their equilibrium contributes to regulation of gene expression (1, 2).

DNA methylation is a covalent chemical modification, adding a CH₃ group at the carbon in position 5 of the cytosines situated in the sequence context 5'CG3'.

In mammalian DNA, normal methylation is restricted to cytosine followed by guanosine (the so named CpG dinucleotide). These CpG sites are rarer than their predicted frequency, but the human genome also contains small regions of DNA called "CpG islands," where CpGs are at higher frequency (CpG Islands 0.5-5kb every 100kb). About half of all human genes (including most housekeeping genes) have CpG islands in their 5'-promoter regions which are usually unmethylated in normal tissues, regardless of the transcriptional status of the gene. DNA methylation is associated with changes in chromatin organization and consequent repression of gene transcription and has a fundamental role during embryogenesis, differentiation, X inactivation, imprinting but also in cancer genesis. Gene

silencing by CpG island methylation is physiologically irreversible during the lifetime of affected cells. Demethylation by pharmacologic agents results in fact in reactivation of gene expression. In normal cells, methylation of CpG islands recruits methyl binding proteins (such as MeCP2), which form complexes including transcription binding repressors, histone deacetylases (HDAC) and DNA methyltransferases (DNMT). This event leads to inaccessibility of DNA to RNA polymerase. In cancer cells, DNMT are often over expressed and hypermethylate CpG islands in ectopic and atypical manner inducing a very stable binding to the co repressor complex, finally leading to permanent repression of gene transcription (2).

The other essential actors of epigenetic mechanisms are the chromatin organizing proteins histones. Histones (H2A, H2B, H3 and H4) associate in duplets to form an octameric complex. This complex is the nucleosome, the spool around which DNA is wrapped (3, 4). Histones can be modified by acetylation of lysine and arginine, phosphorylated and ubiquitinated. The best known modification is acetylation. Histone acetylation occurs on lysine residues at the N-terminal tail of all histones and this process facilitates the access to DNA by destabilizing histone-DNA interaction. Histone acetylation is regulated by the equilibrium of two enzymes histone acetyltransferase (HAT) and HDAC which are recruited locally by sequence specific DNA binding proteins, attracted to the site by CpG methylated islands. Post transcriptional modifications of histones form combinatorial "histone code".

This code is read by transcription factors (and other proteins) and leads to specific gene expression pattern required for particular phenotypic output (4). Euchromatin regions are highly enriched for acetylated histones whereas heterochromatin contains deacetylated histones. Histone acetylation is a transcriptional activatory event and more precisely, acetylation of specific lysine residues and specific patterns of acetylation of different types of histones mark the active chromatin regions and contribute to separate distinct chromatin domains. Lysine bears a positively charged ammonium group at neutral pH. The addition of an acetyl group reduces the affinity for DNA. The loosening of the histone complex from the DNA exposes additional DNA regions to the transcription effectors. In addition, the acetylated lysine residues interact with a specific acetyl lysine-binding domain that is present in many proteins. Thus, acetylation of the histone tails provides a mechanism for recruiting other components of the transcriptional machinery and can activate transcription through a combination of three mechanisms: by reducing the affinity of the histones for DNA, by recruiting other components of the transcriptional machinery, and by initiating the active remodelling of the chromatin structure (2, 4).

Less is known about regulation and role of histone methylation operated by histone methyltransferases (HMT) and specifically associated with euchromatin or heterochromatin.

In transcriptionally active nucleosomes, DNA promoter CpG islands are not methylated. However, if they become methylated due to the activity of DNA methyltransferases (DNMT1, 2, 3a and 3b), methyl binding proteins (MeCP2,

MBP1, 2 and 3) are recruited to the DNA. These proteins in turn recruit corepressor complexes (Sin3), containing histone deacetylases (HDAC) and DNMTs (2, 4, 6). HDACs cause gene silencing by forcing detachment of acetyl groups from lysine tails of histones, resulting in tight wrapping of DNA around the positively charged core histones, which subsequently makes it inaccessible to transcription factors (heterochromatin). The equilibrium between eu- and heterochromatin is essential for gene expression regulation in normal cells, but this balance is disrupted in neoplastic cells. A common shared leukemogenic mechanism is the altered interaction between transcriptional co-regulators (NCoR/SMRT) which are responsible for the recruitment of HDACs and DNMTs. This imbalance is present in human acute promyelocytic leukemia (APL) (PML/RARalpha and PLZF/RARalpha positive) as well as in human acute myeloid leukemia with translocations involving core binding factors (CBF AMLs).

In leukemic cells, DNMTs are often overexpressed. As a result, CpG islands at the promoter site of critical genes involved in cell cycle or cell maturation (e.g. p16ink4a, e-cadherin) are frequently hypermethylated, repressing the expression of these genes. The hypermethylated regions bind to methyl binding protein (MBP), which tightly link co-repressor complexes to DNA and thus render the local chromatin permanently inaccessible to co-activators. In myelodysplastic syndromes, the promoters of several critical genes are hypermethylated and their silencing correlates with disease progression (5-7). Quite recently, evidences have been gathered indicating that in MDS combined hypermethylation of several genes constitutes a strong prognostic factor.

DNA methylation and histone acetylation have been shown to be interdependent, therefore they represent targets for combination therapy using DNA methyltransferase (DNMT) inhibitors (DNMTinhibitors) plus HDAC inhibitors (8).

DNMT inhibitors constitute the first agents approved in the US specifically for treatment of MDS (9). Although in Europe approval by EMEA is still under way, there is no doubt that they constitute the first choice treatment for high risk MDS patients, with no age limitation.

Both 5'-azacytidine and 5'-aza-deoxycytidine (decitabine) have been synthesized more than 40 years ago as cytotoxic drugs, analog of cytosine arabinoside. Although the two molecules have demonstrated similar efficacy in MDS patients, they seem to have different molecular activity due to the fact that decitabine is directly integrated into DNA, whereas azacitidine is first incorporated into RNA and, after reduction, into DNA. Azacitidine has been approved by FDA in 2004 (9) on the basis of studies which demonstrated its efficacy in modifying the clinical history of MDS, by prolonging survival and postponing progression to AML. After early phase I/II studies in which azacitidine was used as intravenous infusion 75 mg/m² for 7 days every month, with 49% hematological improvement (10), the drug has been used in subsequent trials subcutaneously (11), with similar bioavailability (12). Milestone study has been the phase III trial conducted performed by Cancer and Leukemia Group B (CALGB) in 191 MDS patient of all FAB subtypes randomized to receive azacitidine s.c 75 mg/m²/day for 7 days every 4 weeks or best supportive care (13). While therapy related mortality was

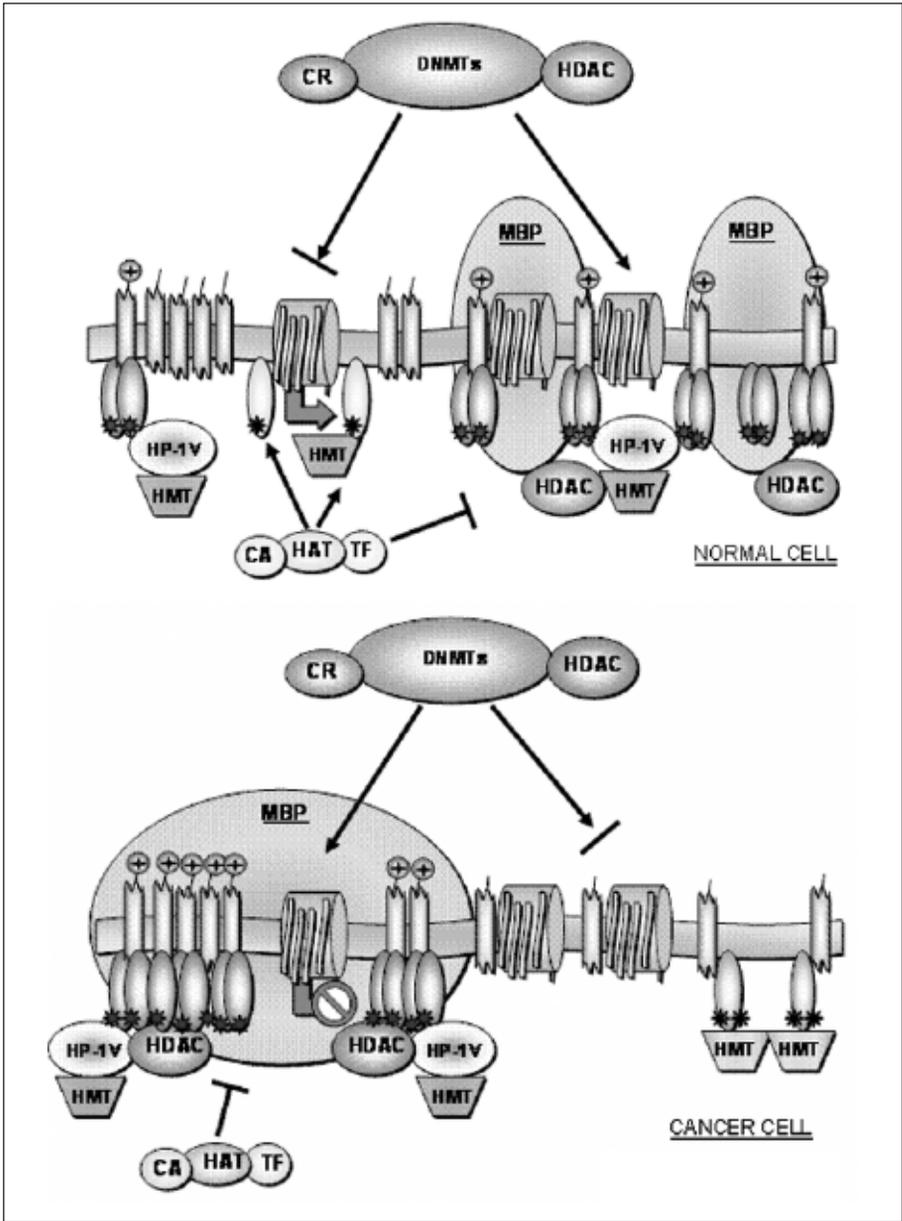


Fig. 1 - Modulation of gene expression by epigenetic modifications.

In normal (a) and in cancer cells (b) chromatin organization and gene expression is modulated by epigenetic modifications, which determine chromatin remodeling. DNA methylation and histone acetylation are interdependent and modulate gene expression. Chromatin status is determined by chemical modifications occurring directly on DNA, on cytosine within CpG islands, and on lysine of histones. Cytidine residues are methylated by DNA methyltransferases (DNMT). Methylation of CpG islands recruits locally methyl binding proteins (MBP) and the latter bind protein complexes containing corepressors (CR), DNA methyltransferases (DNMT) and histone deacetylases (HDAC). HDAC locally deacetylate histone lysines and this modification is accompanied by changes in lysine methylation in the same histone, promoted by histone methyl transferases (HMT). In cancer repression of gene expression is maintained by corepressor complexes of proteins recruited by regionally hypermethylated DNA.

negligible, overall response reached 60% versus 5% in the control arm. Recent re-evaluation of the results of all azacitidine trials, with the application of the revised IWG criteria of response to treatment, has confirmed the efficacy of the drug, indicating a 10-17% CR rate and validating an overall response of 44-47%. In all the studies, achievement of response was obtained after at least 4 cycles of therapy in 75% of MDS responding patients, but in some cases there were very late responders. Mean response duration was 15 months. There are promising results for the use of azacitidine in combination with phenylbutyrate (14) or with high doses of oral valproic acid (50-75 mg/kg/day) for 7 days, plus ATRA 45mg/m²/day for 5 days starting day 3 of azacitidine. Although this cohort of pre-treated, elderly patients had also 83% of abnormal cytogenetics, the responses occurred early (after 1-3 cycles) and accounted for 39% of patients with an overall response of 56% observed in patients >60 years and previously untreated. Dose limiting toxicity in this trial was indeed neurological and due to valproic acid dosage. Because of the unusually rapid onset of response and the high percentage of cytogenetic complete responses, this kind of combination represents at present the most powerful therapy in MDS. At present a GIMEMA trial is ongoing, treating high risk MDS patients with azacitidine s.c. 75 mg/m²/day for 7 days every 4 weeks and plus valproic acid 200-1200 mg/day plus ATRA in case of no response after 4 cycles.

Decitabine has an extremely powerful activity in terms of hypomethylation (15). Decitabine is administered as continuous intravenous infusion 15mg/m² three times daily for 3 days every 6 weeks (16-19). A recent re-evaluation of the European studies analyzed 177 high, int-2 and int-1 risk MDS patients treated with decitabine. Overall response was 49%, with 24% CR, and disease was stabilized in 20% of patients. Age was not influencing negatively the outcome. Mean response duration was 36 weeks (19). Recently, more than a hundred patients with high risk MDS received decitabine 100 mg/m² per course every 4 weeks in 3 different schedules:

- 1) 20 mg/m² intravenously daily x 5;
- 2) 20 mg/m² subcutaneously daily x 5;
- 3) 10 mg/m² intravenously daily x 10.

Decitabine was given for a median of 7 courses and overall response was 70% (19). In multivariate analysis, poor prognostic factors for achieving CR were MDS (vs chronic myelomonocytic leukemia), longer duration of MDS, and prior MDS therapy. For survival, independent adverse prognostic factors were chromosome 5 and/or 7 abnormalities, older age, and prior MDS therapy. The adaptive randomization Bayesian design selected the decitabine schedule of 20 mg/m² intravenously daily x 5 as the schedule associated with the higher CR rate (20, 21). A group of MDS patients has been treated for 10 days with decitabine (15 mg/m² by IV daily) and with escalating doses of oral VPA. This treatment determined 22% of response, with 19% CR. Overall survival was 15.3 months in responders (22). In conclusion, most probably therapies for MDS should combine in the future DNMT inhibitors and HDAC inhibitors (23).

References

1. Turner BM. Reading signals on the nucleosome with a new nomenclature for modified histones. *Nat Struct Mol Biol* 2005; 12 2: 110-2.
2. Santini V, Kantarjian HM and Issa JP. Changes in DNA methylation in neoplasia: pathophysiology and therapeutic implications. *Ann Intern Med* 2001; 134, 7: 573-86.
3. Richmond TJ and Davey CA. The structure of DNA in the nucleosome core. *Nature* 2003; 423, 6936: 145-50.
4. Spotswood HT and Turner BM. An increasingly complex code. *J Clin Invest* 2002; 110 5:577-82.
5. Voso MT, Scardocci A, Guidi F, et al. Aberrant methylation of DAP-kinase in therapy-related acute myeloid leukemia and myelodysplastic syndromes. *Blood* 2004; 103 2: 698-700.
6. Teofili L, Martini M, Luongo M, et al. Hypermethylation of GpG islands in the promoter region of p15(INK4b) in acute promyelocytic leukemia represses p15(INK4b) expression and correlates with poor prognosis. *Leukemia* 2003; 17 5: 919-24.
7. Aggerholm A, Holm MS, Guldborg P, et al. Promoter hypermethylation of p15, HIC1, CDH1, and ER is frequent in myelodysplastic syndrome and predicts poor prognosis in early-stage patients. *Eur J Haematol* 2005.
8. Minucci S and Pelicci PG. Histone deacetylase inhibitors and the promise of epigenetic (and more) treatments for cancer. *Nat Rev Cancer* 2006; 6, 1: 38-51.
9. Kaminskis E, Farrell A, Abraham S, et al. Approval summary: azacitidine for treatment of myelodysplastic syndrome subtypes. *Clin Cancer Res* 2005; 11, 10: 3604-8.
10. Silverman LR, Holland JF, Weinberg RS, et al. Effects of treatment with 5-azacytidine on the in vivo and in vitro hematopoiesis in patients with myelodysplastic syndromes. *Leukemia* 1993; 7, Suppl 1: 21-9.
11. Silverman LR. Targeting hypomethylation of DNA to achieve cellular differentiation in myelodysplastic syndromes (MDS). *Oncologist* 2001; 6, Suppl 5: 8-14.
12. Marcucci G, Silverman L, Eller M, et al. Bioavailability of azacitidine subcutaneous versus intravenous in patients with the myelodysplastic syndromes. *J Clin Pharmacol* 2005; 45, 5: 597-602.
13. Silverman LR, Demakos EP, Peterson BL, et al. Randomized controlled trial of azacitidine in patients with the myelodysplastic syndrome: a study of the cancer and leukemia group B. *J Clin Oncol* 2002; 20, 10: 2429-40.
14. Gore SD. Combination therapy with DNA methyltransferase inhibitors in hematologic malignancies. *Nat Clin Pract Oncol* 2005; 2, Suppl 1: S30-5.
15. Pinto A, Zagonel V, Attadia V, et al. 5-Aza-2'-deoxycytidine as a differentiation inducer in acute myeloid leukaemias and myelodysplastic syndromes of the elderly. *Bone Marrow Transplant* 1989; 4, Suppl 3: 28-32.
16. Wijermans PW, Krulder JW, Huijgens PC and Neve P. Continuous infusion

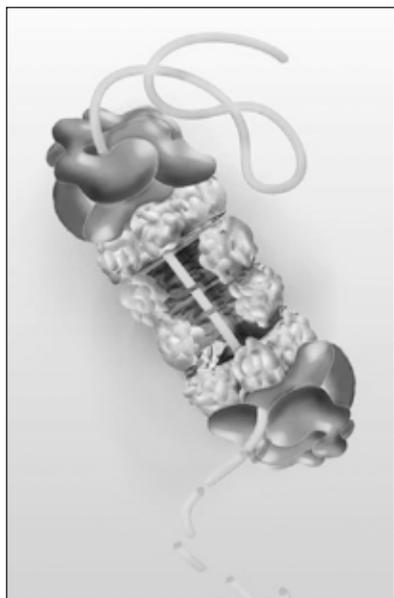
- of low-dose 5-Aza-2'-deoxycytidine in elderly patients with high-risk myelodysplastic syndrome. *Leukemia* 1997; 11 1: 1-5.
17. Lubbert M, Wijermans P, Kunzmann R, et al. Cytogenetic responses in high-risk myelodysplastic syndrome following low-dose treatment with the DNA methylation inhibitor 5-aza-2'-deoxycytidine. *Br J Haematol* 2001; 114, 2: 349-57.
 18. Wijermans P, Lubbert M, Verhoef G, et al. Low-dose 5-aza-2'-deoxycytidine, a DNA hypomethylating agent, for the treatment of high-risk myelodysplastic syndrome: a multicenter phase II study in elderly patients. *J Clin Oncol* 2000; 18 5: 956-62.
 19. Lubbert M and Wijermans PW. Epigenetic therapy in MDS and acute AML: focus on decitabine. *Ann Hematol* 2005; 84 Suppl 13: 1-2.
 20. Kantarjian H, Oki Y, Garcia-Manero G, et al. Results of a randomized study of 3 schedules of low-dose decitabine in higher-risk myelodysplastic syndrome and chronic myelomonocytic leukemia. *Blood* 2007; 109, 1: 52-7.
 21. Kantarjian H, Issa JP, Rosenfeld CS, et al. Decitabine improves patient outcomes in myelodysplastic syndromes: results of a phase III randomized study. *Cancer* 2006; 106, 8: 1794-803.
 22. Garcia-Manero G, Kantarjian HM, Sanchez-Gonzalez B, et al. Phase 1/2 study of the combination of 5-aza-2'-deoxycytidine with valproic acid in patients with leukemia. *Blood* 2006; 108, 10: 3271-9.
 23. Gore SD. Six (or more) drugs in search of a mechanism: DNA methyltransferase and histone deacetylase inhibitors in the treatment of myelodysplastic syndromes. *J Natl Compr Canc Netw* 2006; 4 1: 83-90.

Inibitori del proteosoma nella terapia del mieloma multiplo

Giampaolo Merlini

Centro per lo Studio e la Cura delle Amiloidosi Sistemiche, Fondazione IRCCS Policlinico San Matteo, Dipartimento di Biochimica, Università di Pavia

Il sistema ubiquitina-proteosoma degrada più dell'80% delle proteine cellulari che controllano la divisione, crescita, funzione e morte cellulare (1). Queste proteine includono molti fattori di trascrizione, cicline, chinasi ciclina-dipendenti e loro inibitori, oncogeni, e geni oncosoppressori. Il proteosoma è un grande complesso (circa 2,5 mDa) proteasico di forma cilindrica composto da 44 polipetidi (2). Il centro del cilindro è ricoperto da uno o due complessi regolatori di 890 kDa che svolgono le proteine globulari poliubiquitinate e le iniettano nel cilindro previa dissociazione delle molecole di ubiquitina che possono essere riutilizzate. Il cilindro è composto da due anelli alfa esterni e due anelli beta mediali. Gli anelli alfa formano uno stretto canale che richiede la linearizzazione delle proteine globulari. Il riconoscimento del proteosoma nei confronti delle proteine da degradare è di tipo aspecifico, esso le individua per il fatto che sono legate alla catena di poliubiquitina.



Rappresentazione schematica del proteosoma 26S (da www.molbio.su.se)

Successivamente, con consumo energetico, srotola la proteina e la degrada.

Una volta avvenuta la proteolisi le molecole di poliubiquitina vengono depolimerizzate per essere riciclate. Il sito attivo del proteosoma è dato da treonine attive appartenenti ad alcune subunità, degli anelli centrali, queste treonine sono responsabili del taglio proteolitico. Le unità, contengono i siti proteolitici, che idrolizzano la proteina in alcune posizioni: dopo grandi residui idrofobici (chimo tripsina-simili), basici (tripsina-simili), acidi, "branched" e piccoli residui neutri. La proteina è digerita in piccoli peptidi di 6-7 residui (2-4).

Sono noti numerosi inibitori del proteosoma, naturali e di sintesi. Gli inibitori sintetici possono causare una inibizione del sito

chimotripsina-simile reversibile (ad esempio le amidi, acidi e esteri boronici) o irreversibile (eossi-chetoni e vinil-solfoni). Il bortezomib, un derivato dipeptidico dell'acido borbonico, è il principale rappresentante di una classe di farmaci che inibiscono reversibilmente e con elevata specificità l'attività del proteosoma legandosi alla treonina nel sito attivo chimotriptico (5). Più recentemente, è stata sviluppata una nuova generazione di inibitori, come il nuovo inibitore NPI-0052, che si sono dimostrati efficaci in linee cellulari resistenti al bortezomib (6). Studi in vitro hanno mostrato che bortezomib ha una significativa attività contro numerose linee cellulari tumorali (5). Le cellule tumorali sono più sensibili alla inibizione della attività del proteosoma rispetto alla loro controparte benigna, indipendentemente dalla loro capacità proliferativa (7). Gli studi in vivo hanno mostrato marcata attività, oltre che nel mieloma multiplo, nei confronti del tumore della prostata (8), della leucemia T dell'adulto (9) e in numerosi linfomi B (10). In considerazione della sua attività antitumorale, della sua specificità del favorevole profilo farmacocinetico e facilità di sintesi, il bortezomib è stato il primo inibitore del proteosoma ad essere sperimentato in clinica nei pazienti neoplastici.

Inibizione del proteosoma nel mieloma multiplo

Il proteosoma svolge un ruolo fondamentale nella attivazione del fattore di trascrizione "nuclear factor-kB (NF-kB) attraverso la degradazione del suo inibitore endogeno, I-kB (11). La forma attivata di NF-kB trasloca dal citoplasma nel nucleo, dove aumenta la trascrizione di geni bersaglio coinvolti nella sopravvivenza cellulare, nella adesione e nel dialogo citochinico. In particolare NF-kB promuove la sintesi di citochine (IL-6 e TNF α) fattori di sopravvivenza (IAPa, Bcl-XI) e molecole di adesione (ICAM, VCAM e selectina E). Le cellule mielomatose presentano una aumentata attività dell'NF-kB rispetto alle cellule emopoietiche normali e linee cellulari di mieloma chemioresistente mostrano una maggiore attività NF-kB rispetto alle linee cellulari mielomatose chemiosensibili (12). Nel mieloma multiplo l'adesione delle plasmacellule alle proteine della matrice extracellulare conferisce chemioresistenza e l'interazione con le cellule stromali midollari induce la trascrizione e la secrezione di citochine e delle molecole di adesione mediate da NF-kB (13). L'inibizione del proteosoma nelle plasmacellule mielomatose si traduce in importanti effetti sui meccanismi di segnale coinvolti nella sopravvivenza cellulare e sulla chemioresistenza mediata dalla adesione alle proteine della matrice extracellulare e sulla sintesi di citochine NF-kB-dipendente mediata dalla interazione con le cellule dello stroma midollare. Oltre alla inibizione del sistema NF-kB i farmaci anti-proteosoma esercitano numerosi altri effetti che si traducono in una ridotta degradazione di proteine pro-apoptotiche e regolatorie, quali:

- a) la stabilizzazione e l'accumulo di p53;
- b) la deregolazione del turnover delle cicline;
- c) effetti sulla stabilità delle proteine della famiglia cdc25, KIP1 and WAF1;
- d) sovraespressione dei fattori pro-apoptotici e ridotta espressione di proteine anti-apoptotiche come Bcl-2;

- e) stabilizzazione di JNK, aumentata fosforilazione di c-Jun con up-regolazione di Fas;
- f) deregolazione della via di segnale pro-apoptotica TRAIL;
- g) aumento dei radicali liberi e dello stress ossidativo intracellulare;
- h) annullamento dell'effetto anti-apoptotico di IL-6 e di IGF-1;
- i) riduzione della capacità migratoria delle cellule mielomatose in risposta al VEGF;
- l) effetto anti-angiogenico (14-16).

In questo contesto risulta rilevante l'osservazione che le plasmacellule presentano una ridotta capacità proteosomiale (17). Recentemente è stato inoltre proposto un effetto immunomodulatore del bortezomib attraverso l'inibizione della funzione delle cellule dendritiche (18).

In vitro il bortezomib sopprime significativamente la crescita di colture primarie di mieloma e di linee cellulari, e l'effetto inibitorio è significativamente potenziato dal desametasone (19). Inoltre il bortezomib aumenta marcatamente la sensibilità alla doxorubicina ed al melphalan nelle linee di mieloma chemiosensibili e la reintegra nelle linee cellulari e nelle colture primarie di mieloma chemioresistente (14, 20). È importante sottolineare che la combinazione del bortezomib con le antracicline o con il melphalan non esercita citotossicità nei confronti delle cellule staminali CD-34+. È stato riportato che il bortezomib inibisce la trascrizione di numerosi fattori coinvolti nella risposta cellulare protettiva nei confronti dello stress genotossico, reintegrando pertanto la sensibilità agli agenti citotossici che danneggiano il DNA, come è il caso appunto delle antracicline e del melphalan (14, 15).

Riassumendo, l'inibizione del proteosoma si traduce in un effetto anti-mieloma attraverso:

- a) l'inibizione dell'NF- κ B e la conseguente ridotta espressione di proteine anti-apoptotiche;
- b) la soppressione di altre vie correlate alla chemioresistenza, come la via MAP-chinasi;
- c) interferenza con il meccanismo protettivo di interazione con la componente stromale;
- d) l'attenuazione della risposta cellulare protettiva nei confronti dello stress genotossico.

I meccanismi che mediano la resistenza al bortezomib non sono completamente chiariti. È probabile che la sovraespressione di proteine anti-apoptotiche svolga un ruolo importante (15).

Farmacocinetica e farmacodinamica del bortezomib

Dopo l'infusione il bortezomib si distribuisce in circa 15 minuti per più del 90% nei tessuti, compreso il midollo. Causa una inibizione reversibile della attività del proteosoma, che ritorna alla linea di base in 48-72 ore. La somministrazione di due dosi nella settimana consente il recupero tra le dosi e minimizza la tossicità

(21). Il bortezomib è inattivato per mezzo di deboronazione ossidativa da parte del citocromo P450 e di meccanismi non enzimatici ed è poi escreto nella bile e nelle urine. L'esperienza in pazienti con mieloma e insufficienza renale o in dialisi suggerisce che il bortezomib può essere somministrato in questi pazienti ottenendo una attività antimieloma e con un profilo tossicologico simile a quello osservato nei pazienti con normale funzione renale (22, 23). In generale non sembrano esserci importanti limitazioni all'uso prolungato del bortezomib (24).

Esperienze cliniche del bortezomib nel trattamento del mieloma multiplo

Il bortezomib è stato utilizzato in studi di fase II e III come singolo agente o in combinazione per il trattamento del mieloma multiplo avanzato refrattario o

Tab. 1 - Bortezomib alone or in combination for relapsed/refractory MN.

Type of study	Schedule	Assessable patients	Response %	Reference
Phase 2	Bort. 1.3 mg/m ² on days 1, 4, 8, 11 every 21 days; dexa 20 mg daily, if PD or SD after two cycles	193	CR 4, nCR 6, PR 18 (bort. alone)	Richardson et al.
Phase 2	Bort. 1.1 vs 1.3 mg/m ² days 1, 4, 8, 11 every 21 days; dexa 20 mg daily, if PD or SD after two cycles	53	CR (CR+nCR) 11 (bort. 1.0 alone) 19 (bort. 1.0+dexa), 4 (bort. 1.3 alone), 4 (bort. 1.3+dexa); PR 19, 19, 35, 46	Jagannath et al.
Phase 3	Bort. 1.3 mg/m ² on days 1, 4, 8, 11 every 21 days	315	CR 6, nCR 7, PR 25	Richardson et al.
Phase 2	Bort. 1.3 mg/m ² on days 1, 4, 8, 11 every 21 days	18	CR 9, PR 32	Musto et al.
Phase 2	Bort. 0.90-1.50 mg/m ² on days 1, 4, 8, 11 every 21 days; pegylated liposomal doxorubicin on day 4 at 30 mg/m ²	22	CR 36.3, PR 36.3	Orlowsky et al.
Phase 2	Bort. from 0.7 to 1.0 mg/m ² on days 1, 4, 8, and 11 every 28 days. Oral melphalan from 0.025 to 0.25 mg/kg on days 1-4	34	CR 6, nCR 9, PR 32	Berenson et al.
Phase 2	Bort. 1.0 mg/m ² on days 1, 4, 8, 11; oral melphalan (0.15 mg/kg) days 1-4; thalidomide (100 mg/day); dexa (12 mg/m ²) on days 1-4 and 17-20 every 28 days	25	CR 8, PR 48	Terpos et al.
Phase 2	Bort. from 1 to 1.6 mg/m ² on days 1, 4, 15, 22 every 35 days; melphalan 6 mg/m ² on days 1-5; prednisone 60 mg/m ² on days 1-5; thalidomide 100 mg continuously	15	CR 10, nCR 5, PR 35	Palumbo et al.

Abbreviations: bort. = bortezomib; dexa = dexamethasone; CR = complete response; nCR = near complete response; PD = progression of disease; PR = partial response; SD = stable disease

Tab. 2 - Bortezomib alone or in combination for previously untreated MM.

Type of study	Schedule	Assessable patients	Response %	Reference
Phase 2	Bort. 1.3 mg/m ² on days 1, 4, 8, 11 every 21 days; dexa 40 mg daily, if <PR after two cycles or <CR after four cycles	32	CR 3 (bort. alone), 6 (bort.+dexa); nCR 9 (bort. alone), 19 (bort+dexa); PR 28 (bort. alone), 63 (bort.+dexa)	Jagannath et al.
Phase 2	Bort. 1.3 mg/m ² on days 1, 4, 8, 11; dexa 40 mg on days, 1-4, 1-8 and 15-18 during cycle one and days 1-4 during cycles 2-4; doxorubicin 0, 4.5 or 9 mg/m ² on days 1-4	21	CR 24, PR 71	Oakervee et al.
Phase 2	Bort. 1.0-1.3 mg/m ² on days 1, 4, 8, 11, 22, 25 29, 32; oral melphalan, 9 mg/m ² , prednisone 60 mg/m ² on days 1-4	53	CR 28, nCR 11, PR 45	Mateos et al.
Phase 2	Bort. 1.3-1.6 mg/m ² twice weekly for 4 infusions; thalidomide 100-200 mg daily; dexa 20 mg/m ² for 4 days beginning on days 1, 9, 17	36	CR 19, PR 58	Wang et al.
Phase 2	Bort. 1.3 mg/m ² on days 1, 4, 8, 11; dexa 40 mg on days 1-4, 9-12	30	CR 16.7, VGPR 13.3, PR 43.3	Harousseau et al.
Phase 2	Bort. 1.0 mg/m ² on days 1, 4, 8, and 11; adriamycin 9 mg/m ² on days 1-4; dexa 40 mg on days 1-4, 8-11 and 15-18 on cycle 1 and days 1-4 during cycles 2-4	18	CR 16.6, VGPR 22.2, PR 50	Popat et al.

Abbreviations: bort. = bortezomib; dexa = dexamethasone; CR = complete response; nCR = near complete response; PR = partial response; VGPR = very good partial response

recidivato. La tabella 1, tratta da Nencioni et al. (6), riporta gli studi principali. La notevole efficacia e il buon profilo di tossicità osservato negli studi sui pazienti con mieloma multiplo avanzato hanno condotto alla formulazione di numerose sperimentazioni del bortezomib da solo o in associazione con altri agenti in pazienti con mieloma multiplo di nuova diagnosi, non trattati. La tabella 2 riporta alcuni di questi studi.

Una dettagliata descrizione di queste sperimentazioni è riportata in eccellenti revisioni della letteratura (16, 25-28). La combinazione del bortezomib con desametasone aumenta la percentuale di risposta, ma non sembra prolungare la sopravvivenza libera da eventi e la sopravvivenza totale. La combinazione con antracicline, melphalan, ciclofosfamide e talidomide ha mostrato una notevole attività consentendo di ridurre la dose di bortezomib ed i suoi effetti collaterali. Questi sono rappresentati principalmente da trombocitopenia, astenia, sintomi gastrointestinali, neuropatia, vomito e anoressia (29, 30). La neuropatia periferica è l'effetto collaterale più severo, soprattutto in pazienti con sintomi neurologici

ci prima del trattamento. Il precoce rilevamento della neuropatia e l'impiego di un algoritmo per la riduzione del dosaggio del farmaco o per la sua sospensione consentono un rapido miglioramento della neuropatia nella maggior parte dei pazienti. Una importante osservazione preliminare, ottenuta dall'analisi dei risultati di studi di fase 2 e 3, è che il bortezomib sembra efficace anche nei mielomi che presentano delezione del cromosoma 13 annullandone l'effetto prognostico sfavorevole (31). Lo sviluppo a breve termine della ricerca in questo campo sarà dedicata agli inibitori del proteosoma somministrabili per os e con un migliore profilo tossicologico. Inoltre gli studi di farmacogenomica contribuiranno a chiarire i meccanismi di resistenza al bortezomib e guideranno la scelta di più efficaci combinazioni di farmaci.

Bibliografia

1. Adams J. The proteasome: structure, function, and role in the cell. *Cancer Treat Rev.* 2003;29 Suppl 1: 3-9.
2. Coux O, Tanaka K, Goldberg AL. Structure and functions of the 20S and 26S proteasomes. *Annu Rev Biochem.* 1996; 65: 801-47.
3. Groll M, Bajorek M, Kohler A, Moroder L, Rubin DM, Huber R, et al. A gated channel into 3. the proteasome core particle. *Nat Struct Biol.* 2000; 7: 1062-7.
4. Voges D, Zwickl P, Baumeister W. The 26S proteasome: a molecular machine designed for controlled proteolysis. *Annu Rev Biochem.* 1999; 68: 1015-68.
5. Adams J, Palombella VJ, Sausville EA, Johnson J, Destree A, Lazarus DD, et al. Proteasome inhibitors: a novel class of potent and effective antitumor agents. *Cancer Res.* 1999; 59: 2615-22.
6. Chauhan D, Catley L, Li G, Podar K, Hideshima T, Velankar M, et al. A novel orally active proteasome inhibitor induces apoptosis in multiple myeloma cells with mechanisms distinct from Bortezomib. *Cancer Cell.* 2005; 8: 407-19.
7. Voorhees PM, Orlowski RZ. The proteasome and proteasome inhibitors in cancer therapy. *Annu Rev Pharmacol Toxicol.* 2006; 46: 189-213.
8. Yang W, Monroe J, Zhang Y, George D, Bremer E, Li H. Proteasome inhibition induces both pro- and anti-cell death pathways in prostate cancer cells. *Cancer Lett.* 2006; 243: 217-27.
9. Tan C, Waldmann TA. Proteasome inhibitor PS-341, a potential therapeutic agent for adult T-cell leukemia. *Cancer Res.* 2002; 62: 1083-6.
10. O'Connor OA. Marked clinical activity of the proteasome inhibitor bortezomib in patients with follicular and mantle-cell lymphoma. *Clin Lymphoma Myeloma.* 2005; 6: 191-9.
11. Palombella VJ, Rando OJ, Goldberg AL, Maniatis T. The ubiquitin-proteasome pathway is required for processing the NF-kappa B1 precursor protein and the activation of NF-kappa B. *Cell.* 1994; 78: 773-85.
12. Berenson JR, Ma HM, Vescio R. The role of nuclear factor-kappaB in the

- biology and treatment of multiple myeloma. *Semin Oncol.* 2001; 28: 626-33.
13. Adams J. The development of proteasome inhibitors as anticancer drugs. *Cancer Cell.* 2004; 5: 417-21.
 14. Mitsiades N, Mitsiades CS, Richardson PG, Poulaki V, Tai YT, Chauhan D, et al. The proteasome inhibitor PS-341 potentiates sensitivity of multiple myeloma cells to conventional chemotherapeutic agents: therapeutic applications. *Blood.* 2003; 101: 2377-80.
 15. Mitsiades N, Mitsiades CS, Poulaki V, Chauhan D, Fanourakis G, Gu X, et al. Molecular sequelae of proteasome inhibition in human multiple myeloma cells. *Proc Natl Acad Sci U S A.* 2002; 99: 14374-9.
 16. Nencioni A, Grunebach F, Patrone F, Ballestrero A, Brossart P. Proteasome inhibitors: antitumor effects and beyond. *Leukemia.* 2007; 21: 30-6.
 17. Cenci S, Mezghrani A, Cascio P, Bianchi G, Cerruti F, Fra A, et al. Progressively impaired proteasomal capacity during terminal plasma cell differentiation. *Embo J.* 2006 Mar 8; 25(5): 1104-13.
 18. Nencioni A, Grunebach F, Patrone F, Ballestrero A, Brossart P. The proteasome and its inhibitors in immune regulation and immune disorders. *Crit Rev Immunol.* 2006; 26: 487-98.
 19. Hideshima T, Richardson P, Chauhan D, Palombella VJ, Elliott PJ, Adams J, et al. The proteasome inhibitor PS-341 inhibits growth, induces apoptosis, and overcomes drug resistance in human multiple myeloma cells. *Cancer Res.* 2001; 61: 3071-6.
 20. Ma MH, Yang HH, Parker K, Manyak S, Friedman JM, Altamirano C, et al. The proteasome inhibitor PS-341 markedly enhances sensitivity of multiple myeloma tumor cells to chemotherapeutic agents. *Clin Cancer Res.* 2003; 9: 1136-44.
 21. Orłowski RZ, Stinchcombe TE, Mitchell BS, Shea TC, Baldwin AS, Stahl S, et al. Phase I trial of the proteasome inhibitor PS-341 in patients with refractory hematologic malignancies. *J Clin Oncol.* 2002; 20: 4420-7.
 22. Jagannath S, Barlogie B, Berenson JR, Singhal S, Alexanian R, Srkalovic G, et al. Bortezomib in recurrent and/or refractory multiple myeloma. Initial clinical experience in patients with impaired renal function. *Cancer.* 2005; 103: 1195-200.
 23. Chanan-Khan AA, Kaufman JL, Mehta J, Richardson PG, Miller KC, Lonial S, et al. Activity and safety of bortezomib in multiple myeloma patients with advanced renal failure: a multicenter retrospective study. *Blood.* 2007; 109: 2604-6.
 24. Richardson P, Jagannath S, Colson K. Optimizing the efficacy and safety of bortezomib in relapsed multiple myeloma. *Clin Adv Hematol Oncol.* 2006; 4: 1; discussion 8; suppl 13.
 25. Cavo M. Proteasome inhibitor bortezomib for the treatment of multiple myeloma. *Leukemia.* 2006 Aug;20(8):1341-52.
 26. Richardson PG, Schlossman R, Mitsiades C, Hideshima T, Munshi N, Anderson K. Emerging trends in the clinical use of bortezomib in multiple myeloma. *Clin Lymphoma Myeloma.* 2005; 6: 84-8.

27. Voorhees PM, Orlowski RZ. Emerging role of novel combinations for induction therapy in multiple myeloma. *Clin Lymphoma Myeloma*. 2006; 7: 33-41.
28. Kropff M, Bisping G, Wenning D, Berdel WE, Kienast J. Proteasome inhibition in multiple myeloma. *Eur J Cancer*. 2006 Jul; 42(11): 1623-39.
29. Richardson PG, Barlogie B, Berenson J, Singhal S, Jagannath S, Irwin D, et al. A phase 2 study of bortezomib in relapsed, refractory myeloma. *N Engl J Med*. 2003; 348: 2609-17.
30. Richardson PG, Sonneveld P, Schuster MW, Irwin D, Stadtmauer EA, Facon T, et al. Bortezomib or high-dose dexamethasone for relapsed multiple myeloma. *N Engl J Med*. 2005; 352: 2487-98.
31. Jagannath S, Richardson PG, Sonneveld P, Schuster MW, Irwin D, Stadtmauer EA, et al. Bortezomib appears to overcome the poor prognosis conferred by chromosome 13 deletion in phase 2 and 3 trials. *Leukemia*. 2007; 21: 151-7.

Il sistema telomero/telomerasi: un nuovo bersaglio in ematologia/oncologia

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

Divisione di Ematologia, Fondazione IRCCS Policlinico San Matteo, Università di Pavia

Studi recenti indicano che la telomerasi non viene espressa dai tessuti somatici normali, fatta eccezione per quelli ad alta capacità rigenerativa, ma viene invece espressa dall'85% dei tumori. I meccanismi che portano all'attivazione della telomerasi sono tuttora sconosciuti. In particolare, non si sa se nelle cellule tumorali la riattivazione della telomerasi sia una conseguenza delle alterazioni geniche o se invece sia conseguenza di un processo di selezione delle cellule neoplastiche telomerasi positive in corso di progressione neoplastica. Comunque sia, la telomerasi è ormai considerata un nuovo marcatore della cellula neoplastica e quindi si ritiene che molecole ad attività anti-telomerasica potrebbero rappresentare una nuova classe di farmaci antitumorali (Kelland, 2005; Ju & Rudolph, 2006).

Telomero e telomerasi: struttura e funzioni

Telomeri

I telomeri sono le porzioni terminali di ogni cromosoma e vennero identificati per la prima volta negli anni 30 da Hermann Muller e Barbara McClintock. Essi svolgono alcune importanti funzioni: rivestono l'estremità dei cromosomi ("telomere capping") che vengono così stabilizzati, fanno sì che l'estremità del cromosoma sia distinta dalle rotture della doppia catena del DNA, impediscono che si verifichino fenomeni di fusione cromosomica e di degradazione nucleolitica, partecipano al controllo dell'organizzazione della cromatina e della crescita cellulare. Inoltre, i telomeri con la loro posizione regolano la trascrizione di alcuni geni. Questa loro azione, denominata "Telomere Position Effect" (TPE) è tanto più importante quanto più lungo è il telomero e quanto più breve è la distanza tra telomero e gene in questione, dipende dallo stato di acetilazione degli istoni, blocca l'espressione di alcuni geni. I telomeri sono costituiti da DNA ripetitivo altamente conservato durante l'evoluzione e da proteine. La sequenza ripetitiva del DNA telomerico umano è 5'-TTAGGG-3'. I telomeri umani sono lunghi 5-15kb, ma la loro lunghezza varia in base all'età, al tipo di tessuto ed allo stato della cellula. La catena 3' ricca in guanina si estende oltre il DNA a doppia catena a formare un filamento di DNA a singola catena ("3' overhang"), lungo 150-200

nucleotidi (Pendino et al, 2006; Stewart & Weinberg, 2006, Artandi, 2006). Il DNA sub-telomerico abbastanza ripetitivo è poco conservato durante l'evoluzione, non contiene geni, ma è assolutamente necessario per mantenere integro il genoma. La cromatina dei telomeri contiene i nucleosomi e quindi anch'essa subisce le modificazioni dipendenti dallo stato di acetilazione, metilazione e fosforilazione degli istoni. Perciò, modificazioni epigenetiche delle zone di eterocromatina del telomero possono causare un'alterata funzionalità del telomero stesso: la perdita dell'eterocromatina può fare in modo che la cromatina sia più rilasciata favorendo il legame della telomerasi. Dal punto di vista strutturale i telomeri sono considerati lineari. Tuttavia, le sequenze di DNA che contengono tratti di guanina possono formare strutture a quattro catene, chiamate "G-quadruplex". Si tratta di strutture co-planari prodotte dall'appaiamento di quattro guanine del "3' overhang" per azione di legami d'idrogeno. Studi *in vitro* indicano che le strutture "G-quadruplex" rallentano l'allungamento del telomero ad opera della telomerasi. La catena complementare, ricca in citosina può anch'essa avvolgersi in strutture simili, chiamate "I-motif" prodotte dall'appaiamento delle citosine. L'esistenza delle strutture "G-quadruplex" è stata solo recentemente dimostrata. Inoltre, studi di microscopia hanno dimostrato che l'estremità dei telomeri è doppia e formata da due anelli ("loop"). Il "T-loop", prodotto dall'avvolgimento del telomero su se stesso, ed il "D-loop", prodotto dal fatto che il "3' overhang" si intercala nella sequenza ripetitiva a doppio filamento della catena 5' del telomero stesso. La struttura formata dalla porzione del "3' overhang" intercalata viene stabilizzata dalla proteina telomerica TRF2 che ne impedisce la degradazione ad opera di nucleasi, la ricombinazione e l'allungamento ad opera della telomerasi.

Proteine telomeriche

Le proteine telomeriche, che si legano al DNA a doppia catena del telomero, al "3' overhang" o si connettono con altre proteine telomeriche, regolano la lunghezza del telomero e ne mantengono l'integrità e la funzione. Le due proteine TRF1 (Telomeric Repeat Binding Factor 1) e TRF2 si legano al DNA a doppio filamento del telomero ed interagiscono con altre proteine telomeriche. Dal punto di vista funzionale l'azione di TRF1 e TRF2 si esplica seguendo vie diverse dato che esse partecipano a sistemi di segnale diversi. TRF1 controlla la lunghezza dei telomeri, mentre TRF2 ne controlla la lunghezza ed il "capping". Inoltre, l'omeostasi dei telomeri dipende dalle interazioni tra telomero, proteine telomeriche e telomerasi.

TRF1. Si tratta di una proteina telomerica che si lega alla sequenza TTAGGG della doppia catena del "T-loop". Una sovra-espressione di TRF1 induce una progressiva riduzione della lunghezza dei telomeri, mentre la sua inibizione determina un allungamento del telomero. Questa proteina interagisce con tutto un insieme di altre proteine, tra le quali la più importante è sicuramente la tankirasi. Quest'ultima impedisce il legame del TRF1 al DNA telomerico che conseguentemente diviene accessibile alla telomerasi che determina l'allungamento del telomero. Al contrario il legame tra TRF1 e TIN2 determina una stabilizzazione del "T-loop" e riduce quindi l'accessibilità del telomero alla telomerasi.

TRF2. È una proteina che determina una stabilizzazione del “T-loop” poiché va a localizzarsi nella porzione a doppia catena e nella coda del “T-loop” stesso. Una sovra-espressione di TRF2 determina l’accorciamento e la degradazione dei telomeri con un meccanismo indipendente dalla telomerasi, mentre la sovra-espressione di un mutante dominante-negativo provoca il distacco di TRF2 dal telomero, la perdita del “3’ overhang”, l’attivazione dei meccanismi di riparo del DNA mediati da TP53, lo sviluppo di fusioni cromosomiche “end-to-end” e l’arresto della crescita cellulare. Come TRF1, anche TRF2 interagisce con varie altre proteine. Tra queste quelle sicuramente più importanti sono RAD50, MER11 e NSB1, tutte coinvolte nei processi di riparo del DNA e soprattutto nei processi di ricombinazione omologa e non omologa dell’estremità geniche. Inoltre TRF2 interagisce con la kinasi ATM, l’elicasi WRN (Werner) e BLM (Bloom) e con TRF1.

POT1. È una proteina (Protection of Telomere 1, POT1) che si lega al DNA a singolo filamento del telomero mantenendone l’integrità e regolandone la lunghezza. POT1 interagisce con il telomero direttamente, legandosi all’estremità del “3’ overhang”, o indirettamente, attraverso il legame al complesso formato da TRF1. Si ritiene che questo legame sia assolutamente necessario per trasferire le informazioni dalla regione a doppia catena del telomero alla regione a singola catena del “3’ overhang”.

Telomerasi

La telomerasi è una ribonucleoproteina che consente alla cellula di mantenere la lunghezza dei telomeri. È formata da due componenti: un RNA che funziona da stampo, denominato hTR, ed una trascrittasi inversa, denominata hTERT.

“Human Telomerase RNA” (hTR)

Ha una lunghezza di 451 nucleotidi e contiene uno stampo di 11 nucleotidi necessario alla sintesi della sequenza ripetitiva del telomero. Il gene per hTR è mappato in 3q26.3. hTR possiede una struttura ben definita con regioni altamente conservate che ne stabilizzano la struttura secondaria. L’interazione con hTERT è necessaria per l’attività catalitica, per il processamento del complesso, per l’attacco al telomero, per l’acquisizione di una corretta struttura secondaria e per l’assemblaggio di altre ribonucleoproteine. L’assemblaggio della telomerasi avviene nel nucleolo e richiede la partecipazione di Hsp90 e di p23. Mutazioni di alcune porzioni di hTR e di geni che codificano per proteine necessarie all’assemblaggio della telomerasi (diskerina) sono responsabili delle forme autosomiche dominanti e cromosoma-X legate della diskeratosis congenita. Quest’ultima è una malattia che colpisce i tessuti ad intensa attività proliferativa.

“Human Telomerase Reverse Transcriptase” (hTERT)

È una proteina ad attività catalitica. Si tratta di una DNA polimerasi che appartiene alla famiglia delle trascrittasi inverse. Il gene che codifica per questo componente della telomerasi è mappato in 5p15-33, in una regione cromosomica frequentemente amplificata nei tumori del polmone. Il gene si estende per 37 kb, è

formato da 16 esoni e codifica per una proteina di 1132 amminoacidi. La proteina contiene sei regioni caratteristiche delle trascrittasi inverse ed una regione telomerasi specifica. L'estremità ammino-terminale della proteina è essenziale per la localizzazione nucleolare di hTERT, per l'assemblaggio con hTR e per l'allungamento dei telomeri; la regione carbossi-terminale è essenziale per il processamento dell'enzima e per la sua localizzazione nucleare. L'enzima per essere funzionante deve essere in forma dimerica. Mutazioni di hTERT sono state descritte in pazienti con anemia aplastica.

Sul piano funzionale la telomerasi è necessaria per l'allungamento dei telomeri. L'enzima trasferisce le sequenze ripetitive telomeriche in tandem all'estremità 3-OH del telomero. La proteina TRF1, controllando l'accesso della telomerasi al telomero svolge un importante ruolo nell'allungamento di quest'ultimo. Inoltre, la telomerasi impedisce che si verifichino fenomeni apoptotici facendo sì che la cellula assuma un fenotipo immortale, controlla l'espressione di proteine coinvolte nella proliferazione cellulare e partecipa ai meccanismi d'induzione della chemioresistenza.

La telomerasi è di solito espressa da cellule che possiedono un'importante capacità proliferativa, mentre non viene espressa da cellule quiescenti. Durante l'embriogenesi essa viene espressa a partire dalla blastocisti, ma poi la sua espressione gradualmente si riduce man mano che la cellula si differenzia. Nell'adulto le cellule staminali del tessuto emopoietico, della cute e dell'intestino presentano bassi livelli di telomerasi. Nel tessuto emopoietico la telomerasi è poco espressa dalle cellule staminali quiescenti CD34+,CD38-, ma presenta un aumento di espressione con il "commitment" verso progenitori CD34+, CD38+. I bassi livelli di telomerasi potrebbero essere necessari alla cellula staminale emopoietica quiescente per poter sopravvivere senza proliferare. Questa osservazione fa supporre che l'enzima possa svolgere altre importanti funzioni oltre a quella di mantenere l'integrità dei telomeri, come suggerito dal fatto che la cellula CD34+ in fase proliferativa presenta un progressivo accorciamento dei telomeri nonostante gli alti livelli di telomerasi. Entrambe le osservazioni fanno ipotizzare che le cellule staminali sebbene dotate della capacità di automantenimento abbiano in realtà una potenzialità replicativa limitata che non riesce ad essere modificata dalla telomerasi. Quest'ultima non riesce a prevenire la perdita del DNA telomericoprovocata dall'età e dalla storia proliferativa della cellula. Le cellule staminali emopoietiche del fegato fetale e del cordone ombelicale presentano telomeri 4 kilobasi più lunghi rispetto a quelli delle cellule staminali del midollo osseo e del sangue periferico. Anche nel tessuto emopoietico i livelli di telomerasi si riducono man mano che la cellula si differenzia. La ragione di questa repressione della telomerasi è a tutt'oggi sconosciuta (Deville et al., 2006).

Il controllo dell'espressione della telomerasi avviene a livello della trascrizione, dello "splicing" e della maturazione del RNA messaggero, della struttura proteica di hTERT e hTR. Quest'ultimo viene espresso anche da cellule hTERT negative. L'espressione di hTERT è molto controllata durante la differenziazione cellulare ed il gene è poco o per nulla espresso dalle cellule somatiche normali.

Per quanto concerne la regolazione a livello trascrizionale, è stato osservato che

il promotore di hTERT è molto ricco in isole CpG ed in siti di legame per vari fattori trascrizionali. Tra questi ultimi bisogna ricordare c-Myc, Sp1 e gli ormoni steroidei che fanno aumentare l'espressione di hTERT, mentre Mad1, SIP1, p53, Rb1, E2F, WT1 MZF-2, l'interferone γ , il TNF α ed il TGF β la fanno diminuire. La presenza di isole CpG nel promotore di hTERT suggerisce che la metilazione possa essere uno dei principali meccanismi di controllo dell'espressione di hTERT sia nella cellula normale che in quella cellula cancerosa. Tuttavia, anche se è stato riportato che l'ipermetilazione del promotore di hTERT riduce o blocca del tutto l'espressione del gene e che nella leucemia acuta linfoblastica il trattamento con 5-azacitidina può ripristinare l'espressione di hTERT, la metilazione del promotore non sembra essere il più importante meccanismo di controllo dell'espressione del gene. Anche le modificazioni degli istoni contenuti nel promotore di hTERT e l'effetto TPE più sopra riportato non sembrano svolgere un ruolo particolarmente importante nell'espressione del gene.

Un importante controllo potrebbe essere invece svolto dalle varianti di "splicing". Come per tutti i geni anche il pre-mRNA corrispondente a hTERT va incontro a processi di "splicing" alternativo che producono diverse isoforme di hTERT, ma solo il trascritto completo codifica per una proteina funzionalmente attiva. Tra le isoforme quelle codificate dalle varianti α e β , prodotte da una delezione, sono quelle meglio caratterizzate. La variante α perde 32 nucleotidi dell'esone 6, mentre la variante, 182 nucleotidi che si estendono dall'esone 7 all'esone 8. L'isoforma α ha un'azione dominante negativa nei confronti della telomerasi, mentre l'isoforma, dà luogo ad una proteina tronca. Vi sono altre varianti di splicing prodotte da una delezione e da diversi tipi di inserzione nucleotidica. Le varianti di "splicing" sono tessuto specifiche, la loro funzione è del tutto sconosciuta, potrebbero modulare l'espressione di hTERT e regolare la quantità di proteina attiva. Un altro meccanismo che controlla l'espressione della telomerasi è rappresentato dalle proteine che veicolano l'enzima e che partecipano alla costituzione della forma attiva dell'enzima. A tale proposito un ruolo fondamentale viene svolto dalle proteine hsp90, p27 e "Makorin RING finger protein 1" (MKRN1) che rimangono sempre legate alla telomerasi regolandone l'attività. È stato recentemente osservato che MKRN1 favorisce l'ubiquitinazione di hTERT. Un altro meccanismo che favorisce l'attivazione di hTERT è la fosforilazione a livello di residui specifici di serina/treonina e di tirosina per azione della protein chinasi C (PKC), B (PKB/Akt) e della protein fosfatasi A. L'attivazione della telomerasi dipende anche dalla sua localizzazione all'interno della cellula. Infatti una localizzazione citoplasmatica di hTERT fa sì che la componente catalitica dell'enzima sia inattiva. L'attivazione di hTERT avviene con il suo trasferimento al nucleo per opera delle proteine 14.3.3 e di NF-kb.

Successivamente, la regolazione dell'attività telomerasica dipende dalla sua localizzazione all'interno del nucleo. Normalmente la telomerasi si trova nel nucleolo e passa nel nucleoplasma solo in particolari fasi del ciclo cellulare, ad esempio al momento della replicazione del DNA.

I meccanismi che controllano l'espressione di hTR sono molto meno noti, ma anche per questa componente della telomerasi il controllo si verifica a livello trascrizionale e post-trascrizionale.

Telomeri/telomerasi: invecchiamento e neoplasie

Nel 1960 Hayflick dimostrò che dopo un numero limitato di divisioni le cellule coltivate *in vitro* vanno incontro ad un processo di senescenza. In questa condizione, chiamata “Hayflick limit”, la cellula è bloccata in modo irreversibile nella fase G1 del ciclo cellulare, è vitale e metabolicamente attiva. Come vedremo, l’“Hayflick limit” è un processo di soppressione della crescita tumorale. Negli stessi anni Watson dimostrò che la sintesi di nuovo DNA ad opera di una polimerasi DNA-dipendente avviene solo in presenza di una sequenza di RNA lunga 8-12 paia di basi che funge da estremità 3'-OH e che la sintesi di nuovo DNA procede solamente in direzione 5' → 3'. Pertanto, durante questo processo di sintesi l'estremità 5' del cromosoma non è mai completamente replicata ed una piccola zona del telomero non viene copiata. Questo fenomeno, chiamato “the end-replication problem” determina una progressiva riduzione della lunghezza dei telomeri ad ogni divisione cellulare. Nel 1973 Olovnikov ipotizzò che il grado di accorciamento dei telomeri esprima il numero di divisioni compiute dalla cellula prima di giungere all’“Hayflick limit”. Pertanto, la lunghezza del telomero può essere considerata come l'orologio biologico della cellula (“mitotic clock”). Di solito, la perdita di sequenze telomeriche nei vari tessuti somatici è di circa 25-200 paia di basi. Tuttavia, il rapporto tra grado di accorciamento del telomero e senescenza non è sempre così stretto: alcune cellule con lunghi telomeri sono senescenti, altre con corti telomeri sono vitali. Questa discrepanza dipende dal fatto che la senescenza non è innescata dall'accorciamento medio dei telomeri, ma dal telomero che ha perso il maggior numero di sequenze ripetitive ed in alcuni casi anche da una alterata struttura del telomero (Stewart & Weinberg, 2006).

In ogni caso l’“Hayflick limit” limita significativamente la potenzialità replicativa della cellula. Il segnale molecolare che causa le alterazioni morfologiche e biochimiche che inducono il fenotipo senescente rimane a tutt'oggi sconosciuto anche se molti dati indicano che le proteine p53 e Rb1 svolgono un ruolo importante in tale processo. Difatti, la perdita di tali proteine fa sì che la cellula superi il limite di Hayflick, presenti telomeri sempre più corti, che hanno ormai perso la propria funzione protettiva (“capping”), e si avvia verso un nuovo blocco proliferativo chiamato “Crisis”. Quest'ultima è caratterizzata da una rilevante instabilità cromosomica con frequente morte cellulare. Solo poche cellule superano questa fase o perché, come avviene nella maggioranza dei casi (85-90%), esprimono la telomerasi che le rende immortali e permette loro di compiere un numero illimitato di divisioni impedendo lo sviluppo di anomalie cromosomiche letali, o perché, come avviene nella minoranza dei casi (10-15%), attivano il meccanismo alternativo di allungamento dei telomeri (“Alternative Lengthening of Telomeres”, ALT). L'attivazione tardiva della telomerasi non riesce comunque a ripristinare la lunghezza dei telomeri, ma semplicemente ne permette la stabilizzazione. Quindi si può ipotizzare che nelle fasi precoci della trasformazione neoplastica l'accorciamento dei telomeri limiti la proliferazione cellulare e l'attivazione della telomerasi protegga i telomeri bloccando la trasformazione neoplastica.

ca, viceversa nelle fasi tardive della trasformazione neoplastica la perdita delle sequenze telomeriche induce un'importante instabilità genomica responsabile dell'aumentata incidenza di anomalie cromosomiche, di nuove mutazioni e di una progressione della malattia.

Telomero e telomerasi nell'aplasia midollare e nei i disordini onco-ematologici

Come già riportato la discheratosi congenita è stata la prima malattia ad essere correlata alla presenza di corti telomeri. La discheratosi può essere causata da una mutazione del gene *DKC1*, che codifica per la diskerina, un componente del complesso telomeratico, o da una mutazione del gene *hTR*. Nel primo caso la malattia viene trasmessa come carattere legato al cromosoma X, nel secondo come carattere autosomico dominante. Una mutazione del gene *hTR* è stata descritta anche in pazienti con anemia aplastica.

Sebbene nei pazienti con sindrome mielodisplastica (SMD) i livelli di telomerasi siano più alti rispetto a quelli del soggetto normale, una sovra-espressione della telomerasi è evento piuttosto raro nelle SMD. Il 40% dei pazienti mostra invece un accorciamento dei telomeri che può causare instabilità genomica ed ulteriori lesioni geniche (perdita di eterozigosi, amplificazioni geniche e nuove anomalie cromosomiche). Pertanto, la maggior parte dei pazienti con SMD mostra telomeri corti e normale espressione della telomerasi, ma una piccola frazione può invece presentare una sovra-espressione della telomerasi. Quest'ultima si associa ad un elevato rischio di progressione in leucemia acuta mieloide (LAM) e ad una breve sopravvivenza. Inoltre, il 40% dei pazienti che evolvono in LAM dopo una precedente fase di mielodisplasia presentano alti livelli di telomerasi. Perciò, è stato ipotizzato che nelle SMD la telomerasi venga attivata nelle fasi tardive della malattia, quando l'accorciamento dei telomeri ha raggiunto dimensioni critiche tali da richiedere l'attivazione di quei meccanismi che consentono alla cellula di divenire immortale.

Nelle LAM i telomeri della cellula leucemica sono più corti rispetto a quelli della cellula normale probabilmente per il gran numero di divisioni compiute dalla popolazione leucemica. Uno studio ha riportato che esiste una stretta correlazione tra riduzione della lunghezza dei telomeri e anomalie cromosomiche. Infatti, i telomeri delle LAM che mostrano difetti del cariotipo sono più brevi rispetto a quelli delle LAM a cariotipo normale. Inoltre, lo sviluppo di ulteriori anomalie del cariotipo comporta un'ulteriore riduzione della lunghezza dei telomeri e l'attivazione di *hTERT*. Nel 70-80% delle LAM e LAL quest'ultima componente della telomerasi raggiunge un livello di espressione che è dieci volte quello osservato nei soggetti sani.

Al momento della remissione completa si osserva una normalizzazione della lunghezza dei telomeri e dei livelli di espressione di *hTERT* per una riespansione dell'emopoiesi normale. Vari studi hanno dimostrato che alla recidiva i livelli *hTERT* sono superiori a quelli osservati alla diagnosi e che la sovraespressione di *hTERT* si associa ad una minore percentuale di remissioni ed una più breve sopravvivenza. Queste osservazioni suggeriscono che l'attivazione di *hTERT* sia

necessaria alla cellula per acquisire un vantaggio proliferativo ed una chemioresistenza causata dall'azione anti-apoptotica di hTERT.

Durante il decorso della LLC i telomeri riducono progressivamente la loro lunghezza e quando raggiungono dimensioni critiche causano l'attivazione della telomerasi. Nelle fasi iniziali della leucemia linfatica cronica (LLC) l'espressione di hTERT e l'attività della telomerasi sono inferiori a quelle dei soggetti normali, ma aumentano durante il decorso clinico e raggiungono i livelli più alti quando si verifica l'evoluzione in sindrome di Richter. Un recente studio ha indicato che la sovra-espressione di hTERT e un aumento nell'attività della telomerasi si associano ad un decorso clinico sfavorevole.

Anche nella leucemia mieloide cronica (LMC) si assiste ad una progressiva riduzione della lunghezza dei telomeri durante il decorso clinico e vari studi hanno stabilito che esiste una correlazione tra riduzione dei telomeri e progressione della malattia. Pertanto anche in questo disordine onco-ematologico la lunghezza dei telomeri riflette la storia proliferativa della popolazione leucemica e la gravità della malattia. Siccome nella fase iniziale della LMC il grado di accorciamento dei telomeri è strettamente connesso allo stadio della malattia ed alla restante durata della fase cronica, la determinazione della lunghezza dei telomeri già alla diagnosi è molto importante per identificare i pazienti ad alto rischio di trasformazione. Al momento dell'evoluzione in fase accelerata ed in crisi blastica si assiste ad un'ulteriore riduzione della lunghezza dei telomeri e ad un'aumentata espressione di hTERT.

L'ulteriore perdita di sequenze ripetitive telomeriche è responsabile dello sviluppo di quelle anomalie cromosomiche e della mutazione o dell'amplificazione di BCR-ABL che frequentemente precedono o compaiono contemporaneamente alla crisi blastica. Queste nuove lesioni geniche favoriscono l'insorgenza di chemioresistenza. È stato recentemente riportato che nei pazienti con LMC in fase cronica, in fase accelerata ed in crisi blastica il trattamento con imatinib induce una normalizzazione della lunghezza dei telomeri. Questo risultato potrebbe essere causato non solo dalla riespansione della popolazione Ph⁻ negativa, ma anche dal fatto che l'imatinib induce un aumento dell'attività telomerasica con conseguente allungamento dei telomeri. Viceversa, la resistenza all'imatinib che frequentemente si osserva nei pazienti con malattia in fase accelerata o in crisi blastica potrebbe essere provocata dalla fosforilazione di hTERT che inibisce la telomerasi.

Pertanto, nei pazienti con vari disordini onco-ematologici hTERT deve essere considerato un ottimo marcatore molecolare capace di predire la sopravvivenza del singolo paziente e la progressione della malattia. Tuttavia, rimane ancora da stabilire se la riattivazione della telomerasi sia un evento tardivo nella leucemogenesi, indotto dall'instabilità genomica e dalla riduzione della lunghezza dei telomeri o se non sia invece un evento molto precoce che consente alla cellula leucemica di acquisire un vantaggio proliferativo o che semplicemente riflette l'intensa attività proliferativa della cellula. Se si considera valida l'ipotesi dell'attivazione precoce della telomerasi, altri meccanismi oltre a quelli indicati devono partecipare alla stabilizzazione dei telomeri.

Terapia antitelomerastica delle leucemie

Siccome le leucemie sono caratterizzate da telomeri corti e da un'intensa attività telomerastica sono considerate un ottimo bersaglio di terapia anti-telomerastica. Il complesso telomerastico può essere inibito a vari livelli che vengono qui di seguito elencati.

Cellula telomerasi positiva

La cellula che esprime la telomerasi può essere bersaglio di due tipi di terapia: la vaccinoterapia e la terapia genica. Il razionale della vaccinoterapia consiste nel fatto che hTERT è espresso dalle sole cellule tumorali e contiene epitopi che possono indurre una reazione immunologica. Quest'ultima è mediata da linfociti T citotossici ed è stata dimostrata per il tumore della prostata e della mammella. L'altra terapia che ha come bersaglio la cellula tumorale che esprime la telomerasi consiste nella possibilità di utilizzare il promotore di hTERT per indurre adenovirus oncolitici o geni suicidi. Il razionale di questo tipo di terapia consiste nel fatto che il promotore di hTERT sarà attivo esclusivamente nelle cellule neoplastiche, dato che solo queste possiedono un'attività telomerastica. È stato dimostrato che il promotore di hTERT induce l'espressione di TRAIL, di caspase 6 e caspasi 8 solamente nelle cellule telomerasi positive, mentre queste attivazioni non si sviluppano nelle cellule normali.

Componenti della telomerasi ed i telomeri

Le terapie che hanno come bersaglio i componenti della telomerasi ed i telomeri sono: oligonucleotidi antisense e meccanismi che inducono una interferenza a livello di RNA, "hammerhead ribozymes, forme dominanti-negative di hTERT, agenti ad azione stabilizzatrice nei confronti delle strutture "G-quadruplex", altre piccole molecole. Sia hTERT che hTR possono essere bersaglio di oligonucleotidi antisense. Tuttavia, solo gli oligonucleotidi antisense specifici per hTR inibiscono efficacemente la telomerasi, forse perché hTR, essendo un template, è un bersaglio eccezionalmente accessibile. Pertanto, è stata sintetizzata tutta una serie di nuovi antagonisti di hTR nel tentativo di identificare composti di lunghezza ottimale per ottenere la massima inibizione e dotati della migliore biodisponibilità. Tra questi il più noto è GRN 163. Altre ricerche hanno portato allo sviluppo di oligonucleotidi chimerici, composti da una sequenza antisense che ha come bersaglio la regione che funge da copia di hTR e da una porzione che interagisce con la parte proteica.

Gli "Hammerhead ribozymes" sono piccole molecole di RNA ad attività catalitica capaci di tagliare RNA messengeri a livello di siti specifici. L'introduzione nella cellula cancerosa di "Hammerhead ribozymes" che hanno come bersaglio gli RNA messengeri corrispondenti a hTERT e hTR blocca ogni attività telomerastica. La forma catalitica inattiva di hTERT, chiamata dominante-negativa (DN-hTERT), è anch'essa in grado di inibire selettivamente la telomerasi causando il progressivo accorciamento dei telomeri ed inducendo fenomeni apoptotici.

La stabilizzazione delle strutture "G-quadruplex", le cui caratteristiche sono già

state riportate, da parte di alcune piccole molecole (antrachinoni, porfirine e acridine) impedisce alla telomerasi di raggiungere il telomero bloccandone l'allungamento ad opera dell'enzima. Pertanto, la cellula non riesce a mantenere la lunghezza del telomero. Nonostante, questa importante azione gli stabilizzatori delle strutture "G-quadruplex" sono stati poco sviluppati a causa della loro scarsa selettività di legame. Infatti, questi composti possono legarsi anche al duplex DNA. Recentemente però, è stata identificata tutta una serie di nuovi stabilizzatori. Di questi ultimi fa parte la telomestatina, composto in grado di inibire selettivamente l'attività telomerasica delle cellule BCR-ABL positive. Dati recenti indicano tra l'altro che la telomestatina aumenta l'efficacia terapeutica dell'imatinib e della chemioterapia, suggerendo un sinergismo terapeutico.

Tra le piccole molecole bisogna ricordare gli analoghi nucleosidici (ad esempio l'AZT). Questi composti agiscono legandosi al sito di legame nucleosidico della trascrittasi inversa e sono incorporati nel DNA durante la fase di allungamento, interferendo in tal modo con l'incorporazione dei dNTP. Nonostante, questa loro azione possiedono una scarsa attività inibitoria nei confronti della telomerasi. Un inibitore non nucleosidico specifico di hTERT, BiBR 1532, possiede un'importante e selettiva citotossicità non solo nei tumori solidi ma anche nelle leucemie acute. Le cellule leucemiche sembrano essere più sensibili a questo composto dei normali progenitori emopoietici. Per quanto riguarda il meccanismo d'azione, è stato proposto che, oltre all'inibizione della telomerasi, BiBR 1532 possieda un effetto citotossico forse causato dalla sua interazione con le proteine telomeriche.

Meccanismi di regolazione della telomerasi

In questo caso il possibile bersaglio terapeutico è rappresentato dai meccanismi di controllo trascrizionale e post-trascrizionale della telomerasi e di hTERT. Infatti nella cellula immortale che matura l'inibizione della telomerasi è preceduta dall'inibizione di hTERT. Pertanto, il controllo trascrizionale di quest'ultima componente della telomerasi è divenuto il principale bersaglio di una possibile terapia anti-telomerasica. Studi recenti hanno dimostrato che nella leucemia acuta promielocitica (LAP) i retinoidi svolgono la loro azione antitelomerasica agendo secondo due diverse modalità. Utilizzando linee cellulari NB4 (linea cellulare derivata da LAP) sensibili ed insensibili alla maturazione è stato osservato che la prima modalità consiste in una rapida repressione della telomerasi che si accompagna alla maturazione della linea cellulare NB4, mentre la seconda modalità, identificata impiegando linee cellulari NB4 resistenti alla maturazione, consiste in una lenta repressione della telomerasi. Infatti, questa repressione si sviluppa solo dopo trattamento prolungato con retinoidi e comporta l'accorciamento dei telomeri, l'arresto della crescita e la morte cellulare. L'effetto anti-telomerasico dei retinoidi è stato confermato utilizzando anche altri citotipi di LAM. Ancor più recentemente è stato osservato che il successo della terapia con acido trans retinoico (ATRA) e triossido di arsenico (As_2O_3) in combinazione nei pazienti con LAP dipende dal sinergismo terapeutico dei due farmaci. Infatti, entrambi causano una repressione della telomerasi con conseguente riduzione della lunghezza dei telomeri e morte cellulare.

Bibliografia

1. Kelland LR. Overcoming the immortality of tumour cells by telomere and telomerase based cancer therapeutics – current status and future prospects. *Eur J Cancer* 41: 971-979, 2005.
2. Ju Z & Rudolph L. Telomere and telomerase in cancer stem cells. *Eur J Cancer* 42: 1197-1203, 2006.
3. Muller HJ The remaking of chromosomes. *Collect. Net.* 13: 181-195, 1938.
4. McClintock B The behaviour in successive nuclear division of a chromosome broken in meiosis. *Proc Natl Acad Sci USA* 25: 405-416, 1938.
5. Pendino F, Tarkanyi I, Dudognon C, et al Telomeres and telomerase: pharmacological targets for new anticancer strategies? *Curr Cancer Drugs Targets* 6: 147-180, 2006.
6. Stewart SA & Weinberg RA Telomeres: Cancer and human aging. *Cells Dev Biol* 22: 531-557, 2006.
7. Artandi SE Telomeres, telomerase, and human disease. *New Engl J Med* 355: 1195-1197, 2006.
8. Deville L, Hillion J, Lanotte M, et al Diagnostic, prognostic and therapeutic exploitation of telomeres and telomerase in leukemias. *Curr Pharm Biotec* 7: 171-183, 2006.
9. Hayflick L. The limited in vitro lifetime of human diploid cell strains. *Exp. Cell Res.* 37: 614-636, 1965.
10. Watson JD. Origin of concatemeric T7 DNA. *Nat. New Biol.* 239: 197-201, 1972.
11. Olovnikov AM. A theory of marginotomy. The incomplete copying of template margin in enzymic synthesis of polynucleotides and biological significance of the phenomenon. *J Theor. Biol.* 41: 181-190, 1973.

Ripristino dell'apoptosi mediata da p53: ruolo degli antagonisti di MDM2

Gian Matteo Rigolin

Sezione di Ematologia, Azienda Ospedaliero-Universitaria Arcispedale S. Anna, Università di Ferrara

p53 come target terapeutico

La proteina p53 è un fattore di trascrizione che regola il ciclo cellulare e ricopre la funzione di soppressore tumorale. In presenza di segnali di stress cellulare quali un danno a carico del DNA o una iperespressione di oncogeni, p53 attiva un complesso network di segnali ("the p53 pathway") che possono esitare in un arresto del ciclo cellulare, riparazione del DNA, senescenza replicativa, o apoptosi (1). p53 può tuttavia anche indurre risposte cellulari indipendenti dalla sua azione trascrizionale allo scopo di controllare in modo molto preciso ed irreversibile il processo di autodistruzione cellulare ed evitare quindi di perpetuare un eventuale danno a carico del DNA (2). p53 è la proteina che più frequentemente è inattivata nelle neoplasie umane come risultato di delezioni o mutazioni a carico del gene TP53 presenti in circa il 50% dei tumori umani (3) o di alterazioni dei meccanismi che presiedono alla sua regolazione (4).

Uno dei meccanismi di soppressione di p53 è rappresentato dalla iperespressione del suo regolatore negativo MDM2 (murine double minute 2) (5) per effetto di una amplificazione del gene MDM2 o della iperespressione della proteina senza amplificazione genica (6). Poiché nei tumori iperesprimenti MDM2 p53 è quasi esclusivamente in configurazione wild-type, una affascinante strategia terapeutica potrebbe consistere nel tentativo di ripristinare l'apoptosi mediata da p53 antagonizzando la proteina MDM2 (4). L'utilità terapeutica della riattivazione di p53 mediante gli antagonisti di MDM2 deve tuttavia considerare che:

- 1) MDM2 non è il solo regolatore negativo noto di p53 e pertanto la p53 liberata da MDM2 può non essere completamente attivata;
- 2) un difettivo signaling di p53 nelle cellule tumorali con una p53 wild type potrebbe attenuare o disattivare una risposta agli antagonisti di MDM2;
- 3) una possibile attività soppressiva della crescita e/o una attività apoptotica di p53 nei tessuti normali potrebbe ridurre od eliminare l'azione terapeutica (7).

MDM2: regolatore della stabilità ed attività di p53

MDM2 e p53 partecipano ad un meccanismo di feedback autoregulatorio attraverso il quale le 2 proteine controllano mutualmente i loro livelli. Il promoter del gene *Mdm2* contiene un elemento responsivo a p53 che, in presenza di elevati livelli di p53, determina una attivazione della trascrizione di MDM2 che legandosi a p53 la inattiva impedendone l'attività trascrizionale. MDM2 funziona inoltre come E3 ligasi nel processo di ubiquitilazione di p53 inducendone la distruzione per proteolisi nel proteasoma (8). MDM2 e p53 si legano ciascuno all'altro mediante i loro domini N terminali. Come risultato p53 e MDM2 vengono mantenute, nelle cellule non sottoposte a stress, a livelli molto bassi. Vi sono numerose evidenze che suggeriscono che MDM2 non sia l'unica proteina regolatoria di p53. Recentemente sono state scoperte ulteriori 3 proteine, PIRH2 (p53-induced protein with RING-H2 domain) (9), COP1 (constitutively photomorphogenic 1) (10) e ARF-BP1 (ARF-binding protein 1) (11) in grado di legare p53 e di fungere da ligasi di p53 nel processo di ubiquitilazione. La presenza di queste nuove ligasi di p53 suggerisce che più regolatori possano essere coinvolti in condizioni di stress ed in diversi tessuti. Il loro ruolo nella regolazione di p53 è tuttavia ancora poco conosciuto e non ci sono evidenze che queste ligasi possano sostituirsi a MDM2 nella regolazione della stabilità di p53.

MDMX (MDM4 nel topo) è una proteina con un elevato livello di omologia con MDM2 che lega p53. Similmente a MDM2, MDMX si lega a p53 con elevata affinità ed inibisce la sua capacità di transattivazione (12). MDMX non è invece capace di ubiquitilare p53. Non avendo inoltre nella regione promoter elementi responsivi a p53 non è regolato trascrizionalmente da p53. MDM4 sembra tuttavia essere essenziale per lo sviluppo in quanto topi *Mdm4*^{-/-} muoiono allo stato embrionale e similmente a quelli *Mdm2*^{-/-}, possono essere salvati dalla delezione del gene *TP53* (13). Similmente a MDM2, MDMX è iperespresso in alcuni tumori umani e sembra facilitare la tumorigenesi anche se le sue esatte funzioni molecolari e la sua regolazione devono essere ancora definiti in modo chiaro.

Antagonizzare la funzione di MDM2 per attivare p53

MDM2 è il più importante soppressore della funzione di p53 in tumori con aberrante espressione di MDM2 (14). Antagonizzando pertanto la funzione di MDM2, si ottiene la stabilizzazione di p53, il suo rapido accumulo a livello nucleare e l'attivazione di specifici geni target che determinano un arresto del ciclo cellulare ed apoptosi. Diversi approcci sono stati utilizzati allo scopo di incrementare il rilascio di p53 dal controllo di MDM2 (15).

- 1) Un primo meccanismo consiste nell'utilizzo di oligonucleotidi antisense allo scopo di downregolare MDM2 (16). Si è riscontrato che non solo le cellule con p53 wild type ma anche quelle con p53 mutata rispondono a questo tipo di blocco di MDM2. Questa osservazione è di difficile interpretazione. Recentemente è stato suggerito che l'attività antitumorale degli oligonucleo-

- tidi antisenso anti-MDM2 sia dovuta alla stabilizzazione p53-indipendente dell'inibitore delle chinasi ciclino dipendente p21^{WAF1} (17).
- 2) Poiché l'ubiquitilazione e la degradazione di p53 sono i principali meccanismi di controllo da parte di MDM2, una strategia per attivare p53 utilizza come bersaglio l'attività ligasica di MDM2. Recentemente sono state identificate alcune piccole molecole inibitrici che hanno come target principale l'attività E3 ligasica di MDM2 (18). Queste sostanze hanno una tuttavia ridotta potenza e selettività di azione.
 - 3) L'inibizione del legame MDM2-p53 è una ulteriore possibile strategia per ottenere una stabilizzazione di p53. Tre aminoacidi di p53 (Phe19, Trp 23 e Leu26) che si inseriscono in una profonda tasca idrofobica sulla superficie della molecola MDM2, sono essenziali per il legame tra p53 e MDM2 (19). Sfruttando queste caratteristiche recentemente è stata identificata una classe di potenti e selettivi attivatori della p53 wild type, le Nutlins.

Piccole molecole inibitrici della interazione tra p53 e MDM2

Le nutlins sono la prima classe di antagonisti di MDM2 con un potente e selettivo meccanismo di azione (15, 20). Questi inibitori, identificati a partire da una classe di composti cis-imidazolici, sono in grado di spiazzare in vitro p53 da MDM2 a concentrazioni nanomolari. Studi cristallografici hanno evidenziato come le nutlins si leghino a MDM2 nella tasca per la p53 mimando l'interazione molecolare di p53. Le nutlins inibendo con elevata specificità l'interazione tra p53 e MDM2, stabilizzano p53 ed attivano la via di p53, bloccando la proliferazione delle cellule tumorali in fase G1 e G2 ed inducendo l'apoptosi (21). Solo le cellule con p53 wild type sono sensibili alle nutlins. La somministrazione di nutlin-3a a topi nudi portatori di xenotrapianti di tumori umani ha determinato una inibizione della crescita tumorale ed una riduzione del tumore a dosi non tossiche (22). Ulteriori classi di attivatori di p53 sono state recentemente descritte (7, 23, 24). Questi composti tuttavia spesso presentano una bassa potenza ed una ridotta selettività di azione.

Inibitori di MDM2 come agenti antitumorali

La possibilità di usare le nutlins come agenti antitumorali dipende dallo stato mutazionale di TP53 e dalla funzionalità della via di attivazione di p53. La maggior parte delle mutazioni riduce infatti la capacità di p53 di legarsi alla sequenza di riconoscimento del DNA nella regione promoter dei suoi geni bersaglio. p53 mutata è pertanto di solito inattiva come fattore trascrizionale (25). p53 mutata non è nemmeno in grado di attivare il suo regolatore negativo MDM2 che pertanto non può influenzare i livelli e l'attività di p53. Esperimenti cellulari con le nutlins hanno confermato che gli antagonisti di MDM2 richiedono per la loro azione una p53 wild-type. La determinazione dello stato di p53 è pertanto un pre-requisito per qualsiasi approccio terapeutico avente come bersaglio l'interazione MDM2-p53 (26).

Anche in presenza di una p53 wild-type, è tuttavia possibile riscontrare nelle neoplasie difetti nella regolazione o nel signaling di p53 (27). p53 è infatti al centro di un complesso network di trasduzione di segnali cellulari. Segnali a monte nella via di p53 assicurano la sua stabilizzazione ed attivazione, mentre segnali a valle permettono l'esecuzione dell'appropriato programma cellulare. Nelle neoplasie, alterazioni sia nella parte a monte che in quella a valle della via di p53 sono in eguale modo dannose in quanto possono attenuare o completamente disattivare la funzione di p53. Gli inibitori della interazione tra p53-MDM2 possono compensare i difetti nella parte a monte nella via di p53. Al contrario difetti a valle nella via di p53 potrebbero ridurre in modo significativo la loro efficacia.

Gli antagonisti di MDM2 sono maggiormente efficaci nelle neoplasie che presentano come unica alterazione l'aberrante espressione di MDM2 suggerendo pertanto che lo stato di MDM2 possa essere un utile biomarcatore di risposta (22). Tuttavia una risposta è stata anche osservata in tumori con p53 wild type esprimenti livelli normali di MDM2 nei quali i circuiti a valle nella via apoptotica di p53 sono conservati. L'identificazione di difetti a valle nella via di p53 può essere pertanto un utile indicatore per stratificare i pazienti con p53 wild type in un contesto clinico. È stato dimostrato che anche MDMX è un potente inibitore dell'attività trascrizionale di p53 (12). Pertanto una efficace attivazione di p53 potrebbe richiedere non solo la dissociazione di MDM2 ma anche di MDMX dal suo dominio N-terminale. Studi recenti hanno dimostrato che la impossibilità di spiazzare MDMX da p53 può attenuare la risposta apoptotica a nutlin-3 (28). L'attività apoptotica di Nutlin-3 può pertanto essere incrementata riducendo l'espressione di MDMX (28). Pur essendovi un certo grado di omologia tra i domini di MDM2 e MDMX in grado di legare p53, nutlin-3 sembra essere incapace di interferire direttamente nell'interazione tra p53 e MDMX. Nonostante ciò, si può tuttavia ipotizzare che nutlin-3 possa influenzare i livelli di MDMX modificando i livelli cellulari di MDMX e p53 e facilitando la degradazione MDM2-dipendente di MDMX. Entrambi questi meccanismi nutlins-dipendenti potrebbero diminuire la funzione inibitoria di MDMX su p53. Pertanto una diretta inibizione dell'interazione p53-MDMX potrebbe ulteriormente incrementare l'attività antitumorale degli inibitori di MDM2 specialmente in tumori con alterazioni della espressione e della regolazione di MDMX (7).

Tossicità derivante della attivazione di p53 sui tessuti normali

Uno dei principali problemi legati all'uso terapeutico di agenti attivanti la p53 è la possibile tossicità per i tessuti normali (29). Le nutlins rappresentano la prima classe di composti antagonisti di MDM2 in grado di attivare la via di p53 selettivamente senza danni a carico del DNA. Il trattamento con nutlin-3 determina in modelli animali murini una effettiva inibizione della crescita e regressione tumorale senza significativa tossicità, confermando l'ipotesi che una attivazione di tipo non genotossico può essere meno tossica e meglio tollerata di una di tipo genotossico (21, 22). Ulteriori studi sono tuttavia necessari per confermare la sicurezza di questo approccio terapeutico.

Monoterapia e terapia di combinazione

Tra i diversi stress cellulari in grado di attivare la via di p53, il danno a carico del DNA è sicuramente uno dei più potenti ed è implicato nel meccanismo di azione di molti agenti chemioterapici genotossici (1). È tuttavia noto che i farmaci genotossici possono anche indurre vie di attivazione p53 indipendenti oltre che severi effetti collaterali tra cui, per un effetto cumulativo, neoplasie secondarie. Per questi motivi, gli attivatori non genotossici di p53 che abbiano come bersaglio l'interazione p53-MDM2 potrebbero rappresentare una interessante alternativa all'uso di chemioterapici genotossici. In risposta ad uno stress genotossico, p53 va incontro a modificazioni post traduzionali quali fosforilazione, acetilazione, e sumoilazione. Queste modificazioni ed in particolare la fosforilazione nella porzione N-terminale si ritiene abbiano un ruolo cruciale non solo nella dissociazione di p53 da MDM2 ma anche nella attivazione di p53 come fattore trascrizionale (30). In realtà recentemente è stato dimostrato che la fosforilazione non sia indispensabile per l'attivazione di p53 e per la sua attività trascrizionale (31). Una ulteriore importante aspetto è che gli inibitori dell'interazione p53-MDM2 non necessitano di un concomitante danno a carico del DNA per una attività massima e possono essere pertanto utilizzati come singoli agenti nella terapia antitumorale (22). Inoltre poiché sia gli antagonisti di MDM2 che gli agenti chemioterapici genotossici attivano la via di p53, ci si può attendere che questi due tipi di agenti abbiano un effetto additivo o sinergico. Studi recenti hanno valutato la combinazione di nutlin-3 e diversi agenti chemioterapici nel mieloma multiplo (32), nella leucemia mieloide acuta (33) e nella leucemia linfatica cronica (34) ed hanno dimostrato che gli antagonisti di MDM2 possono incrementare in vitro l'attività di un agente chemioterapico convenzionali.

Gli inibitori di MDM2 come agenti chemoprotettori

I farmaci che arrestano la progressione del ciclo cellulare prima che entri in una particolare fase potrebbero proteggere le cellule proliferanti da agenti che richiedono, per la loro azione, il passaggio attraverso questa fase (35). Si potrebbe pertanto pensare di proteggere i normali tessuti proliferanti manipolando il ciclo cellulare durante il trattamento chemioterapico. Al riguardo, considerando che gli antagonisti di MDM2, attivando la via di p53, arrestano la progressione del ciclo cellulare nelle fasi G1 e G2 nelle cellule con p53 wild type, è possibile disegnare strategie terapeutiche che considerino un effetto protettivo degli inibitori di MDM2 sulle cellule proliferanti normali di pazienti trattati con agenti antimitotici per tumori con p53 mutata senza influenzare l'efficacia terapeutica del farmaco antitumorale sulle cellule neoplastiche (36).

Conclusioni

L'attivazione delle vie di soppressione tumorale di p53 nelle neoplasie maligne rappresenta un approccio terapeutico particolarmente attraente anche se il

potenziale clinico è ancora incerto. Il successo degli antagonisti di MDM2 dipenderà da una serie di aspetti che in futuro dovranno essere chiariti quali la loro tossicità nei riguardi dei tessuti normali, la possibilità di sviluppare resistenze acquisite alla attivazione di p53, la possibilità di identificare neoplasie con via della p53 preservata.

Pertanto altri studi sono necessari per valutare l'utilità ed i possibili vantaggi degli antagonisti di MDM2 rispetto alle attuali terapie.

Bibliografia

1. Vogelstein B, et al. Surfing the p53 network. *Nature* 408; 307-10, 2000.
2. Vousden KH, et al. Live ore let die: the cell's response to p53. *Nat Rev Cancer* 2; 594-604, 2002.
3. Hainaut P and Hollstein M. p53 and human cancer: the first ten thousand mutations. *Adv Cancer Res* 77, 81-137, 2000.
4. Bullock AN and Fersht AR. Rescuing the function of mutant p53. *Nat Rev Cancer* 1; 68-76, 2000.
5. Fakharzadeh et al. Structure and organization of amplified double minutes containing the mdm2 oncogene. *Genomics* 15; 283-90, 1993.
6. Oliner JD et al. Amplification of a gene encoding p53-associated protein in human sarcomas. *Nature* 358; 80-3, 1992.
7. Vassilev LT. MDM2 inhibitors for cancer therapy. *Trends Molec Med* 13, 23-31, 2007.
8. Hoeller et al. Ubiquitin and ubiquitin-like proteins in cancer pathogenesis. *Nat Rev Cancer* 6; 776-88, 2006.
9. Leng RP et al. Perh2, a p53-indiced ubiquitin-protein ligase, promotes p53 degradation. *Cell* 112; 779-91, 2003.
10. Dornan D et al. The ubiquitin ligase COP1 is a critical negative regulator of p53. *Nature* 429; 86-92, 2004.
11. Chen D et al. ARF-BP1/mule is a critical mediator of the ARF tumor suppressor. *Cell* 121; 1071-83, 2005.
12. Marine C and Jochemsen AG. Mdmx as an essential regulator of p53 activity. *Bioche Biophys Res Commun* 381; 750-60, 2005.
13. Parant J et al. Rescue of embryonic lethality in Mdm4-null mice by loss of Tp53 suggest an overlapping pathway with MDM2 to regulate p53. *Nat Genet* 29; 92-5: 2001.
14. Bond G et al. MDM2 is a central node in the p53 pathway: 12 years and counting. *Curr Cancer Drug Targets* 5; 3-8: 2005.
15. Chene P Inhibiting the p53-MDM2 interaction: an important target for cancer therapy. *Nat Rev Cancer* 3; 102-9, 2003.
16. Zhang R et al. Novel antisense anti-MDM2 mixed-backbone oligonucleotides: proof of principle, in vitro and in vivo activities, and mechanisms. *Curr Cancer Drug Targets* 5; 43-9: 2005.
17. Zhang Z et al. MDM2 is a negative regulator of p21WAF1/CIP1 independent of p53. *J Biol Chem* 279; 16000-6, 2004.

18. Yang Y et al. Small molecule inhibitors of MDM2 ubiquitin ligase activity stabilize and activate p53 in cells. *Cancer Cell* 7; 547-59, 2005.
19. Kussie PH et al. Structure of the MDM2-oncoprotein bound to the p53 tumor suppressor transactivation domain. *Science* 274; 948-53, 1996.
20. Vassilev T. p53 activation by small molecules: application in oncology. *J Med Chem* 48; 4492-9, 2005.
21. Vassilev T, et al. In vivo activation of the p53 pathway by small-molecule antagonists of MDM2. *Science* 303; 844-8, 2004.
22. Tovar C et al. Small-molecule MDM2 antagonists reveal aberrant p53 signalling in cancer: implications for therapy. *PNAS* 103; 1888-93, 2006.
23. Grasberger BL et al. Discovery and cocrystal structure benzodiazepine MDM2 antagonists activate p53 in cells. *J Med Chem* 48; 909-12, 2005.
24. Issaeva N et al. Small molecule RITA binds p53, blocks p53-MDM2 interaction and activates p53 function in tumors. *Nat Med* 10; 1321-8, 2004.
25. Levine AJ. p53, the cellular gatekeeper for growth and division. *Cell* 88; 323-31; 1997.
26. Ahrendt SA et al. Rapid p53 sequence analysis in primary lung cancer using an oligonucleotide probe array. *Proc Natl Acad Sci USA* 96; 7382-7, 1999.
27. Vousden KH and Prives C. p53 and prognosis: new insights and further complexity. *Cell* 120; 7-10, 2005.
28. Patton JC et al, Levels of MDMX expression dictate the sensitivity of normal and transformed cells to Nutli-3. *Cancer Res* 66; 3169-73, 2006.
29. Gudkov AV and Komarova EA. The role of p53 in determining sensitivity to radiotherapy. *Nat Rev Cancer* 3; 117-29, 2003.
30. Appella E and Anderson CW. Post-translational modifications and activation of p53 by genotoxic agents. *Eur J Biochem* 268; 2764-72, 2001.
31. Thompson T et al. Phosphorylation of p53 on key serines is dispensable for transcriptional activation of apoptosis. *J Biol Chem* 279; 53015-22, 2004.
32. Stuhmer T et al. Nongenotoxic activation of the p53 pathway as a therapeutic strategy for multiple myeloma. *Blood* 106; 3609-17, 2005.
33. Kojima K et al. MDM2 antagonists induce p53 apoptosis in AML: implications for leukaemia therapy. *Blood* 106; 3150-9, 2005.
34. Secchiero P et al. Functional integrity of the p53-mediated apoptotic pathway induced by the nongenotoxic agent nutlin-3 in B-cell chronic lymphocytic leukemia (B-CLL) *Blood* 107; 4122-29, 2006.
35. Blagosklonny MV and Pardee AB. Exploiting cancer cell cycling for selective protection of normal cells. *Cancer Res* 61; 4301-5, 2001.
36. Caravajal D et al. Activation of p53 by MDM2 antagonists can protect proliferating cells from mitotic inhibitors. *Cancer Res* 65; 1928-24, 2005.

**ELEMENTI
DI BIOINFORMATICA**

Informatica biomedica: lo scenario internazionale

Riccardo Bellazzi

Gruppo di Bioinformatica e Data Mining, Laboratorio di Informatica Biomedica,
Dipartimento di Informatica e Sistemistica, Università di Pavia

Gli ultimi decenni hanno visto un enorme sviluppo della genetica, della ricerca clinica e delle tecnologie dell'informazione. La convergenza di questi settori sembra portare nuova linfa alla ricerca biomedica, con la promessa di rendere praticabile la "genomic medicine" che consentirà di trasformare radicalmente la medicina e la sanità grazie a nuovi approcci personalizzati, preventivi e predittivi. L'Informatica Biomedica si occupa specificatamente di alcuni aspetti di questa convergenza (Tarczy-Hornoch, 2007). La Bioinformatica e la biologia computazionale sono state fondamentali nell'aiuto alla gestione ed alla comprensione di grandi quantità di dati genomici messi a disposizione dai progetti di sequenziamento. L'Informatica Medica è stata altrettanto fondamentale nell'applicare gli sviluppi dell'Information Technology alla Sanità e nello studiare e rappresentare dati e conoscenza clinica. A questo proposito, i ricercatori nel campo e le associazioni scientifiche di settore, quali l'International Medical Informatics Association (IMIA, <http://www.imia.org>), hanno negli anni fortemente spinto verso l'impiego delle tecnologie informatiche orientate non solo all'ottimizzazione delle procedure amministrative ma anche al miglioramento della qualità delle cure. In questo contesto l'Informatica Biomedica si propone come ponte fra l'Informatica Medica e la Bioinformatica.

L'esperienza condotta nel campo della Bioinformatica e dell'Informatica Medica sembra oggi di fondamentale importanza per permettere di colmare le distanze ancora esistenti fra i sistemi informativi biologici e quelli clinici, e per diventare un fattore abilitante per la genomic medicine. In questo contesto, è di particolare rilevanza l'iniziativa posta in essere dall'Associazione Americana di Informatica Medica (AMIA, <http://www.amia.org>).

L'AMIA ha recentemente definito fra le sue linee strategiche prioritarie la ricerca nel settore della bioinformatica traslazionale, definita come lo sviluppo di metodi di archiviazione, analisi e interpretazione di dati biomedici, al fine di trasformarli in conoscenza immediatamente utile nella pratica clinica (<http://www.amia.org/inside/stratplan/>). La bioinformatica traslazionale riguarda lo sviluppo di nuove tecniche per l'integrazione di dati biologici e clinici e la tra-

sformazione delle metodologie proprie dell'informatica medica per utilizzare nei processi di decisione clinica le osservazioni ottenute a livello molecolare. Il prodotto finale della ricerca nel campo della bioinformatica traslazionale è la nuova conoscenza che verrà generata, formalizzata e quindi disseminata ad una varietà di operatori che comprendono i ricercatori dell'area biomedica, i clinici ed i pazienti. Questo approccio sembra fondamentale per cogliere e sfruttare pienamente le implicazioni della rivoluzione copernicana che in questi ultimi anni ha cambiato e sta cambiando radicalmente la ricerca biomedica. Le iniziative nel campo della Bioinformatica traslazionale dovranno dunque necessariamente volgere la loro attenzione sia allo sviluppo di competenze ed alla messa a disposizione di risorse computazionali per un accesso sempre più facilitato alla conoscenza biomedica, sia allo studio di nuovi metodi ed algoritmi che permettano la soluzione di problemi complessi, legati alle applicazioni della ricerca alle problematiche medico/cliniche. Queste linee di ricerca sono state confermate dalle linee di intervento strategico identificate dal National Institute of Health (NIH) degli Stati Uniti tramite la Biomedical Information Science and Technology Initiative (<http://www.bisti.nih.gov/nbc/>).

Questo intervento ha finanziato la nascita di 7 centri sul territorio statunitense con lo scopo di creare del software innovativo in grado di consentire ai ricercatori in ambito biomedico ed ai clinici di integrare, analizzare, modellizzare, simulare e condividere dati sulla salute e sulle malattie dell'uomo (Tabella 1).

In questa iniziativa possiamo scorgere alcune fra le linee di ricerca che appaiono più promettenti nell'immediato futuro: l'informatica biomedica "integrativa", la systems biology e lo studio dei sistemi multi-scala, la rappresentazione della conoscenza biomedica e le ontologie ad essa collegate, la simulazione di sistemi complessi, le tecniche di imaging dedicate alle applicazioni su scala molecolare, l'integrazione fra dati clinici e molecolari e la biologia computazionale. È evidente come il filo conduttore di tutte queste iniziative riguardi la capacità di sistemi informativi di supportare la ricerca e le sue ricadute mediante la capacità di integrare fattivamente dati e conoscenza. Nel seguito verranno evidenziate alcune problematiche di interesse, con particolare riferimento alla necessità di implementare approcci integrativi ed alla tipologia di dati che devono essere appunto "integrati", e verrà descritto in maggior livello di dettaglio il centro legato al progetto i2b2, che presenta gli aspetti di maggior coerenza con le idee di ricerca traslazionale precedentemente descritti.

Approcci integrativi e problematiche computazionali

L'approccio integrativo a supporto della ricerca

Uno degli aspetti più interessanti dei mutamenti in atto nel campo della ricerca biomedica è legato al cambiamento del paradigma sperimentale, che dai tradizionali studi cellulari basati su approcci riduzionisti è oggi passato ad approcci integrativi. Questi ultimi si basano sull'osservazione che sistemi complessi quali quelli cellulari non possano essere interamente compresi tramite lo studio di componenti individuali isolati. L'approccio classico alla ricerca biologica è stato

Tab. 1 - I centri della Biomedical Information Science and Technology Initiative.

Nome	Principal Investigator	Istituzione	Sito web
National Center for Integrative Biomedical Informatics	Brian D. Athey, Ph.D.	University of Michigan	http://www.ncibi.org/
National Center for the Multi-Scale Analysis of Genomic and Cellular Networks	Andrea Califano, Ph.D.	Columbia University	http://magnet.c2b2.columbia.edu/
National Center for Biomedical Ontology	Mark A. Musen, M.D., Ph.D.	Stanford University	http://bioontology.org/
Physics-based Simulation of Biological Structures	Russ Altman, M.D., Ph.D.	Stanford University	http://cbmc-web.stanford.edu/simbios/
National Alliance for Medical Imaging Computing	Ron Kikinis, M.D.	Brigham and Women's Hospital	http://www.na-mic.org/
Informatics for Integrating Biology and the Bedside	Isaac Kohane, M.D., Ph.D.	Brigham and Women's Hospital	http://www.partners.org/i2b2
Center for Computational Biology	Arthur Toga, Ph.D.	University of California at Los Angeles	http://www.loni.ucla.edu/CCB/

infatti rivolto all'identificazione dei componenti individuali (geni, proteine, metaboliti, ecc.) responsabili di un particolare fenomeno in un organismo (e.g. attività metabolica, risposta a stimoli esterni). Questo approccio si è dimostrato efficace nell'elucidare i componenti chiave dei sistemi viventi, portando ad una varietà di importanti applicazioni in medicina. È tuttavia ora chiaro come l'informazione ad un solo livello (genoma o proteoma) non possa spiegare completamente il comportamento di uno specifico sistema biologico (Aggarwal H, 2003; Reif DM, 2004). Per studiare e comprendere un sistema biologico è quindi necessario un approccio integrato basato sull'uso di misurazioni simultanee di parametri genetici, di genomica funzionale e di proteomica. Approcci "integrativi" allo studio di sistemi biologici sono stati proposti già dalla fine degli anni '60: la mancanza di informazioni a livello molecolare su larga scala ne ha tuttavia limitato l'utilizzo. Grazie alla disponibilità delle sequenze genomiche complete di diversi organismi e all'emergere delle tecnologie high-throughput per la misura dell'espressione genica, dell'espressione proteica e delle interazioni tra proteine e tra proteine e DNA, è ora possibile misurare simultaneamente un vasto numero di componenti cellulari e studiare sistemi biologici a livello molecolare. Gli approcci integrativi richiedono necessariamente l'utilizzo di risorse computazionali, che divengono parte integrante della ricerca scientifica. Queste risorse si possono oggi dividere in due grandi categorie, da una parte grandi basi di dati che

raccogliono, integrano e sistematizzano la conoscenza disponibile (e.g. le banche dati NCBI, EBI, JDDDB) e dall'altra lo sviluppo di algoritmi e metodologie in grado di permettere l'interpretazione dei dati e la scoperta "in silico" di associazioni, relazioni, strutture funzionali tra componenti biologici.

L'approccio integrativo e la necessità di gestire dati clinici

È ben evidente come attualmente nel paradigma integrativo non debbano essere visti soltanto gli aspetti legati alla ricerca, ma anche alla clinica. È pertanto di particolare interesse studiare le metodologie necessarie ad integrare dati clinici con i dati provenienti dalla ricerca.

Le risorse come OMIM (McKusick VA, 1998) e PharmGKB (Klein TE, 2001) sono tools molto utili per recuperare informazioni ad un certo livello di dettaglio sulle problematiche cliniche e biologiche coinvolte in una certa patologia, ma non sono strutturate per supportare la cura dei pazienti in un contesto operativo. Nonostante alcuni studi pilota, l'integrazione dei dati clinici e genomici non è ancora di larga diffusione. Attualmente, le cartelle cliniche informatizzate e i sistemi informativi clinici non sono progettati per supportare una gestione ottimale dell'informazione genomica. In particolare, molte cartelle cliniche gestiscono i dati di diagnostica di laboratorio come dati transienti o di follow-up. In realtà, i dati genomici (che verranno brevemente descritti nel seguito) possono avere sia caratteristiche simili all'anagrafica dei pazienti (nel caso dei dati di genotipo), che, viceversa, essere considerati alla stregua di indagini diagnostiche particolarmente costose e complesse (nel caso dei dati di proteomica o di espressione genica), come le immagini radiografiche. Assumono particolare criticità, anche dal punto di vista dei requisiti di privacy connessi, i dati di genotipo, il cui utilizzo può condizionare le decisioni terapeutiche anche dei familiari dei pazienti. Inoltre, i dati di tipo genomico debbono necessariamente essere interpretati mediante l'accesso alla conoscenza biologica in costante evoluzione; ciò richiede di ripensare le cartelle cliniche elettroniche esistenti, trasformandole in uno strumento "dinamico", in grado di collegare in modo proattivo diverse sorgenti di informazione. Come riportato in (Hoffman, 2007) vi sono almeno tre direzioni che sarà necessario sviluppare per la realizzazione di un sistema informativo che possa essere considerato "Genome-enabled": nuovi strumenti per collegare dinamicamente i dati e la conoscenza disponibile in campo genomico, un adeguato vocabolario controllato (e le connesse ontologie) per dare una descrizione delle evidenze cliniche legate alle informazioni genetiche, e applicazioni in grado di supportare il processo decisionale dei clinici. Attualmente, molti dei sistemi utilizzati nei laboratori di diagnostica e citogenetica non permettono di gestire in modo strutturato gli output di dati genomici e genetici, e quando questo avviene, è solitamente il risultato di uno sviluppo di soluzioni "ad-hoc", dipendenti da specifici aziende o società di software. La sfida sarà quindi quella di proporre degli standard condivisi che recepiscano i tentativi di standardizzazione fatti nel campo della ricerca (come lo standard MIAME per i microarrays, <http://www.mged.org/Workgroups/MIAME/miame.html>) e li inseriscano nella standardizzazione semantica e sintattica in atto nel campo dell'Informatica Medica.

I dati e la conoscenza da integrare

Mentre i dati clinici sono ormai stati ampiamente studiati ed hanno dato luogo a cartelle informatizzate con caratteristiche commerciali, i dati di tipo genomico rappresentano quindi l'attuale frontiera per la realizzazione di sistemi informativi completamente integrati. Lo sviluppo di metodi computazionali per la loro analisi è poi l'altro aspetto con cui l'informatica biomedica si va confrontando. La disponibilità di nuovi strumenti quindi, a cavallo di ricerca e clinica, sembra essere la nuova ed autentica sfida per l'informatica dei prossimi anni. Si ritiene utile quindi fornire alcuni spunti sulle problematiche relative alla tipologia di alcuni dei dati sperimentali che sono oggi disponibili e che dovranno un domani essere integrati nei sistemi informativi.

Dati genomici e di genotipo

La genotipizzazione, ovvero la determinazione della sequenza genomica di un individuo (genotipo), è al giorno d'oggi estremamente facilitata dalla disponibilità di tecniche di misura su larga scala, quali i CGH arrays e gli SNP arrays.

Grazie alla genotipizzazione, possono essere messe in luce le numerose tipologie di modificazioni genetiche che possono avvenire nel genoma di un individuo, quali mutazioni, ricombinazioni, o variazioni nel numero di copie (Brown TA, 2002).

Gli array-CGH rappresentano una importante tecnica diagnostica e prognostica delle patologie collegate a variazioni nel numero di copie di segmenti genomici (Solinas-Toldo S, 1997; Pinkel D, 1998). Tali array sono una evoluzione della CGH ("comparative genomic hybridization") convenzionale, rispetto alla quale permettono una maggior risoluzione consentendo di individuare delezioni ed amplificazioni anche di piccoli segmenti genomici.

Negli array CGH, il DNA genomico proveniente da un campione di test e da un campione di riferimento normale (controllo) vengono sottoposti a ibridazione competitiva: l'analisi fornisce i rapporti tra il segnale di test e il segnale di riferimento lungo ciascun cromosoma, offrendo una visione globale di guadagni e perdite (amplificazioni e delezioni) di segmenti genomici lungo l'intero genoma analizzato. L'obiettivo principale dell'utilizzo di metodi computazionali per l'analisi dei dati provenienti da esperimenti di array-CGH consiste nell'identificazione delle regioni in cui si verificano tali guadagni e perdite. Una volta individuati questi segmenti, che possono essere di varie ampiezze, l'interesse si sposta sulla stima del corretto numero di copie di ciascun segmento (Pollack JR, 2002; Hodgson G, 2001, Wang P, 2005, Fridlyand J, 2004, Autio R, 2003; Hupe P, 2004; Olshen AB, 2004; Picard F, 2005).

Un'altra tecnica largamente utilizzata per la genotipizzazione è quella degli SNP array, tramite i quali è attualmente possibile analizzare in parallelo centinaia di migliaia di "single nucleotide polymorphisms" (polimorfismi a singola base). L'applicazione principale degli SNP arrays è nello studio di malattie genetiche complesse (Mayeux R, 2005). Nello studio di malattie complesse, i metodi di linkage presentano delle limitazioni, in quanto i tratti complessi hanno bassa eredi-

tabilità (Hirschhorn JN, 2005). Gli studi di associazione sono stati proposti per identificare le varianti più comuni che causano tratti complessi: in generale tali studi confrontano la frequenza allelica tra individui malati (casi) e individui sani (controlli). In generale gli studi di associazione sono preceduti da analisi preliminari perlopiù finalizzate ad effettuare un controllo di qualità dei dati e a selezionare l'insieme di SNPs su cui focalizzare l'attenzione (Balding DJ, 2006). Esempi di queste pre-analisi sono test di deviazione dall'equilibrio di Hardy-Weinberg (Wigginton JE, 2005; Weir BS, 2004), inferenze degli aplotipi (Marchini J, 2006; Scheet P, 2006) e SNP tagging (International Hap Map Consortium, 2005; Huang W, 2006; Gonzalez-Neira A, 2006; Sebastiani P, 2003).

Questi studi preliminari sono seguiti da test di associazione, basati su SNPs singoli, multipli o su aplotipi (Freidlin B, 2002; Lunn DJ, 2006; Waldron ER, 2006; Wang H, 2005; Lin DY, 2006; Tzeng JY, 2006). Grazie alla attuale disponibilità di vaste banche dati di SNP (<http://www.ncbi.nlm.nih.gov/projects/SNP/>), di tools per analizzare (<http://snpper.chip.org/>) e di SNP array capaci di analizzare migliaia di SNPs alla volta, è oggi possibile affrontare studi di associazione su scala genomica (genome-wide association studies), ovvero studi che considerano l'intero genoma per l'analisi di varianti genetiche alla base di malattie complesse. Questi studi sono particolarmente promettenti poiché non richiedono la formulazione a priori di ipotesi sulla posizione genomica delle varianti causali (Verzilli CJ, 2006; Wang H, 2006; Skol AD, 2006).

Dati di trascrittomica

Le tecnologie high-throughput per la misura dell'espressione genica (DNA Microarrays) rappresentano un potente strumento per la caratterizzazione dei cambiamenti che hanno luogo durante vari processi biologici, quali stati di malattia, stadi di sviluppo, e risposte a perturbazioni genetiche o indotte da farmaci. I profili di espressione possono essere raccolti sia osservando le cellule in diverse condizioni sperimentali (dati statici), sia osservando l'evoluzione di un particolare processo nel tempo (dati dinamici). L'elevata dimensionalità dei dati raccolti richiedono lo sviluppo di opportuni metodi bioinformatici in grado di estrarre l'informazione in essi contenuti.

In letteratura si ritrovano diversi metodi per l'analisi di profili di espressione genica, rivolti alla analisi differenziale, al raggruppamento dei profili in cluster ed alla ricostruzione di reti di regolazione tramite metodi di reverse engineering. L'obiettivo dell'analisi differenziale è l'identificazione di geni differenzialmente espressi in diverse condizioni sperimentali. Nell'analisi di dati statici, diversi metodi sono stati proposti in letteratura e implementati in applicazioni software: alcuni si basano sull'uso di test d'ipotesi (Reich M, 2004, Tusher VG, 2001), altri sfruttano metodi Bayesiani (Sebastiani P, 2006). Una volta identificati i geni di interesse, i dati vengono spesso analizzati attraverso tecniche di clustering, che ha l'obiettivo di partizionare i geni in gruppi omogenei e ben separati, sulla base dei profili di espressione. L'assunzione chiave alla base di questo approccio è che geni che presentano profili di espressione simili (geni co-espressi) siano funzio-

nalmente correlati. Uno degli approcci più usati è il clustering gerarchico agglomerativo proposto da Eisen et al. (Eisen MB, 1998), ma interessanti alternative sono rappresentate da metodi Bayesiani (Ramoni MF, 2002; Ferrazzi F, 2005), e da metodi basati su una descrizione qualitativa dei profili (Hvidsten TR, 2003; Sacchi L, 2005). Il traguardo più ambizioso nell'analisi dei dati provenienti da esperimenti con DNA microarray è realizzare il "reverse engineering" di reti di regolazione a partire da dati di espressione, ovvero la ricostruzione dei meccanismi di controllo genico.

Dati di proteomica

Nell'ultimo decennio è cresciuto esponenzialmente l'interesse della ricerca biomedica verso la proteomica, vale a dire la scienza che studia l'insieme delle proteine espresse dal genoma di un organismo. I principali motivi di tale interesse possono essere così sintetizzati:

- Nell'uomo i geni sono stimati essere circa 35.000, mentre il numero totale di proteine è valutato fra 250 e 500 mila; è pertanto ormai evidente che le sole informazioni genomiche sono insufficienti a spiegare gli eventi biologici, almeno negli organismi superiori.
- È ormai chiaro che i profili di espressione del mRNA non si correlano sempre con il livello di espressione proteica, principalmente perché non tutti gli mRNA vengono effettivamente tradotti in proteine, e poiché il contenuto proteico della cellula è determinato sia dalla velocità di sintesi di nuove proteine sia da quella di degradazione di quelle già esistenti.
- A seguito dei progressi delle tecnologie analitiche la ricerca biologica si è spostata dallo studio del singolo gene all'analisi dell'espressione dell'intero genoma. La messa a punto di nuove tecniche di misura high-throughput richiede tuttavia una sempre maggiore integrazione con risorse bioinformatiche, al fine di rendere interpretabile la mole di dati generata.

La ricerca in proteomica sta andando in varie direzioni, costituendo vari settori caratterizzati da differenti obiettivi (Timischl B, 2004); tra questi ricordiamo la separazione di proteine, l'identificazione di proteine, la proteomica strutturale, l'interattomica, il protein profiling e l'analisi delle sequenze proteiche con predizioni in-silico.

Il progetto i2b2

Il centro che maggiormente coglie le esigenze della ricerca traslazionale è i2b2 (Informatics for Integrating Biology and the Bedside), che è collocato presso il Partners Health Care System a Boston; il centro è costituito da sette "cores" che coinvolgono ricercatori degli ospedali affiliati ad Harvard, il MIT, l'Harvard School of Public Health, il Joslin Diabetes Center, l'Harvard Medical School e l'Harvard/MIT Division di Health Sciences and Technology. Il centro i2b2 sta sviluppando un ambiente di Information Technology che ha l'obiettivo di integrare le basi di dati cliniche presenti presso gli ospedali affiliati e le grandi basi

di dati biologiche per migliorare la comprensione dei tratti genetici complessi. Gli obiettivi di alta integrazione di i2b2 vengono resi evidenti dal seguente esempio, tratto dal sito del centro:

Un medico esperto di una rara patologia neuromuscolare è a conoscenza che un gene, Y, un fattore di trascrizione, ha alcune mutazioni che sono responsabili per il 45% di tutti i casi della malattia. La sua ipotesi è che il gene contribuisca ad aumentare il rischio di un tipo comune di cardiomiopatia, in cui l'età di insorgenza sia dopo i 40 anni. Per esplorare questa ipotesi vuole sapere se vi sono associazioni in uno degli organismi modello fra il gene Y e le patologie cardiache. I2b2 mette¹ a disposizione un motore di ricerca che permette di effettuare una ricerca del costruito "cardiac and gene Y" e di ottenere un report contenenti tutte le informazioni relative agli organismi modello, integrando database sia gene-centrici, che locus-centrici che infine malattie-centrici. Le informazioni fanno esplicitamente riferimento ad ontologie e vocabolari in grado di integrare funzioni biologiche e localizzazioni anatomiche. Nel nostro esempio, supporremo che il medico abbia trovato evidenza di una relazione fra Y e le malattie cardiovascolari nel topo e che si sappia che Y è espresso nel tessuto cardiaco. Incoraggiato da questi risultati il medico decide di pianificare uno studio di associazione per quel gene. Grazie ad una funzionalità i2b2 il medico consulta il database HapMap (www.hapmap.org) per determinare quali siano gli aplotipi esistenti per quel gene. Sulla base dell'esistenza di aree di omologia conservate lungo un ampio tratto dell'albero filogenetico i2b2 suggerisce che lo studio di associazione segua una larga porzione di DNA per includere le possibili regioni di regolazione. Il tool consente di evidenziare una mappa della regione con il minimo numero di SNPs che permettono di identificare l'aplotipo e di recuperare tutte le annotazioni disponibili. I2b2 suggerisce anche che in ogni braccio dello studio vengano arruolati 5,500 individui. Grazie alla descrizione dei pazienti fornita ad uno specifico tool di i2b2 per una query fra tutti gli istituti ospedalieri coinvolti, è possibile sapere se vi è la disponibilità dei pazienti richiesti. Il risultato mostra quale istituzione ha i pazienti, se sono disponibili dei campioni per l'analisi del DNA e se vi sono dei tessuti cardiaci per le analisi genomiche e proteomiche. Vengono inoltre fornite altre informazioni sulla fattibilità dello studio, come ad esempio le liste dei prezzi delle piattaforme necessarie per la genotipizzazione. I dati ottenuti dalle cartelle cliniche elettroniche, vengono integrati con i dati sperimentali ottenuti dalle analisi dei genotipi. I dati vengono successivamente analizzati mediante strumenti statistici adeguati. Lo studio dimostra che un blocco aplotipico in una regione conservata del gene Y è in linkage disequilibrium altamente significativo con il fenotipo di cardiomiopatia precoce. Questo blocco aplotipico è grande e contiene 35 siti con varianti comuni. Per risolvere queste varianti il sistema i2b2 presenta diverse opzioni che includono l'elenco dei transcriptional factor binding sites vicini a queste varianti, una valutazione strutturale della probabilità che le varianti contribuiscano a questo fenotipo, la proposta di collaboratori esperti del gene Y da coinvolgere per un test in-vitro o in-vivo su organismi modello.

¹In realtà solo alcuni dei servizi sono già disponibili, mentre la maggior parte sono attualmente in fase di sviluppo.

Questo esempio mostra come ricerca e clinica vengano integrati in un solo grande contesto, in cui la disponibilità di dati clinici diviene uno strumento abilitante per condurre ricerca genetica. La realizzazione di i2b2 coinvolge tuttavia grandi sfide di tipo informatico, sia di integrazione di dati, che di rappresentazione della conoscenza che di gestione dell'organizzazione del lavoro di ricerca.

Attualmente i2b2 stà affrontando con le metodologie proposte quattro ambiti complessi: malattie respiratorie, ipertensione, Diabete Mellito di tipo 2, malattia di Huntigton.

Conclusioni

Ai grandi avanzamenti scientifici della ricerca biomedica si stanno affiancando grandi trasformazioni delle tecnologie di supporto alla ricerca ed alla pratica clinica, quali l'informatica biomedica. In particolare, è ormai evidente che l'informatica sia divenuta uno strumento indispensabile sia per la ricerca di base che per il lavoro quotidiano, e che debba essere considerata un elemento ineludibile della società dell'informazione in cui ciascuno di noi si trova oggi ad operare.

Tuttavia, l'abbondanza di informazioni può trasformarsi da fattore abilitante in fattore di confondimento quando non si attuino strategie opportune per affrontarne l'impatto.

La sfida dell'intero settore dell'informatica biomedica si colloca proprio nel fornire ai ricercatori ed ai clinici nuovi strumenti per muovere da una babele disomogenea di dati ed informazioni ad una gestione coerente della conoscenza necessaria per il progresso scientifico.

Bibliografia

1. Aggarwal K, Lee KH. Functional genomics and proteomics as a foundation for systems biology. *Brief Funct Genomic Proteomic*. 2003 Oct; 2(3): 175-84.
2. Autio R, Hautaniemi S, Kauraniemi P, Yli-Harja O, Astola J, Wolf M, Kallioniemi A. CGH-Plotter: MATLAB toolbox for CGH-data analysis. *Bioinformatics*. 2003 Sep 1; 19(13): 1714-5.
3. Balding DJ. A tutorial on statistical methods for population association studies. *Nat Rev Genet*. 2006 Oct; 7(10) 781-91.
4. Brown TA, Genomes. BIOS Scientific Publishers Ltd, 2002.
5. Di Camillo B, Sanchez-Cabo F, Toffolo G, Nair SK, Trajanoski Z, Cobelli C. A quantization method based on threshold optimization for microarray short time series. *BMC Bioinformatics*. 2005 Dec 1; 6 Suppl 4: S11.
6. Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci USA*. 1998 Dec 8; 95(25): 14863-8.
7. Ferrazzi F, Magni P, Bellazzi R. Random walk models for Bayesian clustering of gene expression profiles. *Appl Bioinformatics*. 2005; 4(4): 263-76.
8. Freidlin B, Zheng G, Li Z, Gastwirth JL. Trend tests for case-control stu-

- dies of genetic markers: power, sample size and robustness. *Hum Hered.* 2002; 53(3): 146-52.
9. Fridlyand J, Snijders AM, Pinkel D, Albertson DG, Jain AN. Hidden Markov Models approach to the analysis of array CGH data. *Journal of Multivariate Analysis.* 2004; 90(1): 132-153.
 10. Gonzalez-Neira A, Ke X, Lao O, Calafell F, Navarro A, Comas D, Cann H, Bumpstead S, Ghori J, Hunt S, Deloukas P, Dunham I, Cardon LR, Bertranpetit J. The portability of tagSNPs across populations: a worldwide survey. *Genome Res.* 2006 Mar; 16(3): 323-30.
 11. Hirschhorn JN, Daly MJ. Genome-wide association studies for common diseases and complex traits. *Nat Rev Genet.* 2005 Feb; 6(2): 95-108. Review.
 12. Hodgson G, Hager JH, Volik S, Hariono S, Wernick M, Moore D, Nowak N, Albertson DG, Pinkel D, Collins C, Hanahan D, Gray JW. Genome scanning with array CGH delineates regional alterations in mouse islet carcinomas. *Nat Genet.* 2001 Dec; 29(4): 459-64.
 13. Hoffman MA. The genome-enabled electronic medical record. *J Biomed Inform.* 2007 Feb; 40(1): 44-6.
 14. Huang W, He Y, Wang H, Wang Y, Liu Y, Wang Y, Chu X, Wang Y, Xu L, Shen Y, Xiong X, Li H, Wen B, Qian J, Yuan W, Zhang C, Wang Y, Jiang H, Zhao G, Chen Z, Jin L. Linkage disequilibrium sharing and haplotype-tagged SNP portability between populations. *Proc Natl Acad Sci U S A.* 2006 Jan 31; 103(5): 1418-21.
 15. Hupe P, Stransky N, Thiery JP, Radvanyi F, Barillot E. Analysis of array-CGH data: from signal ratio to gain and loss of DNA regions. *Bioinformatics.* 2004 Dec 12; 20(18): 3413-22.
 16. Hvidsten TR, Laegreid A, Komorowski J. Learning rule-based models of biological process from gene expression time profiles using Gene Ontology. *Bioinformatics.* 2003 Jun 12; 19(9):1116-23.
 17. International HapMap Consortium. A haplotype map of the human genome. *Nature.* 2005 Oct 27; 437(7063): 1299-320.
 18. Klein TE, Chang JT, Cho MK, Easton KL, Fergerson R, Hewett M, Lin Z, Liu Y, Liu S, Oliver DE, Rubin DL, Shafa F, Stuart JM, Altman RB. Integrating Genotype and Phenotype Information: An Overview of the PharmGKB Project. *The Pharmacogenomics Journal.* 2001; 1: 167-170.
 19. Lin DY, Zeng D. Likelihood-based inference on haplotype effects in genetic association studies. *J Am Stat Assoc.* 2006; 101(473): 89-104.
 20. Lunn DJ, Whittaker JC, Best N. A Bayesian toolkit for genetic association studies. *Genet Epidemiol.* 2006 Apr; 30(3): 231-47.
 21. Marchini J, Cutler D, Patterson N, Stephens M, Eskin E, Halperin E, Lin S, Qin ZS, Munro HM, Abecasis GR, Donnelly P. International HapMap Consortium. A comparison of phasing algorithms for trios and unrelated individuals. *Am J Hum Genet.* 2006 Mar; 78(3): 437-50.
 22. Mayeux R. Mapping the new frontier: complex genetic disorders. *J Clin Invest.* 2005 Jun; 115(6): 1404-7.
 23. McKusick VA. Mendelian Inheritance in Man. *A Catalog of Human Genes*

- and Genetic Disorders. Baltimore: Johns Hopkins University Press, 1998 (12th edition).
24. Olshen AB, Venkatraman ES, Lucito R, Wigler M. Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics*. 2004 Oct; 5(4): 557-72.
 25. Picard F, Robin S, Lavielle M, Vaisse C, Daudin JJ. A statistical approach for array CGH data analysis. *BMC Bioinformatics*. 2005 Feb 11; 6: 27.
 26. Pinkel D, Seagraves R, Sudar D, Clark S, Poole I, Kowbel D, Collins C, Kuo WL, Chen C, Zhai Y, Dairkee SH, Ljung BM, Gray JW, Albertson DG. High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. *Nat Genet*. 1998 Oct; 20(2): 207-11.
 27. Pollack JR, Sorlie T, Perou CM, Rees CA, Jeffrey SS, Lonning PE, Tibshirani R, Botstein D, Borresen-Dale AL, Brown PO. Microarray analysis reveals a major direct role of DNA copy number alteration in the transcriptional program of human breast tumors. *Proc Natl Acad Sci USA*. 2002 Oct 1; 99(20): 12963-8.
 28. Ramoni MF, Sebastiani P, Kohane IS. Cluster analysis of gene expression dynamics. *Proc Natl Acad Sci USA*. 2002 Jul 9; 99(14): 9121-6.
 29. Reich M, Ohm K, Angelo M, Tamayo P, Mesirov JP. GeneCluster 2.0: an advanced toolset for bioarray analysis. *Bioinformatics*. 2004 Jul 22; 20(11): 1797-8.
 30. Reif DM, White BC, Moore JH. Integrated analysis of genetic, genomic and proteomic data. *Expert Rev Proteomics*. 2004 Jun; 1(1): 67-75.
 31. Sacchi L, Bellazzi R, Larizza C, Magni P, Curk T, Petrovic U, Zupan B. TA-clustering: cluster analysis of gene expression profiles through Temporal Abstractions. *Int J Med Inform*. 2005 Aug; 74(7-8): 505-17.
 32. Scheet P, Stephens M. A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. *Am J Hum Genet*. 2006 Apr; 78(4): 629-44.
 33. Sebastiani P, Lazarus R, Weiss ST, Kunkel LM, Kohane IS, Ramoni MF. Minimal haplotype tagging. *Proc Natl Acad Sci U S A*. 2003 Aug 19; 100(17): 9900-5.
 34. Sebastiani P, Ramoni MF, Nolan V, Baldwin CT, Steinberg MH. Genetic dissection and prognostic modeling of overt stroke in sickle cell anemia. *Nat Genet*. 2005 Apr; 37(4): 435-40.
 35. Sebastiani P, Xie H, Ramoni MF. Bayesian analysis of comparative microarray experiments by model averaging. *Bayesian analysis* 1, Number 1. 2006: 1-26.
 36. Skol AD, Scott LJ, Abecasis GR, Boehnke M. Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. *Nat Genet*. 2006 Feb; 38(2): 209-13.
 37. Solinas-Toldo S, Lampel S, Stilgenbauer S, Nickolenko J, Benner A, Dohner H, Cremer T, Lichter P. Matrix-based comparative genomic hybridization:

- biochips to screen for genomic imbalances. *Genes Chromosomes Cancer*. 1997 Dec; 20(4): 399-407.
38. Tarczy-Hornoch P, Markey MK, Smith JA, Hiruki T. Bio*Medical informatics and genomic medicine: Research and training, Editorial, *J Biomed Inform*. 2007 Feb; 40(1): 1-4.
 39. Timischl B. *An Introduction to Mass Spectrometry in Proteomics*. 2004.
 40. Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci USA*. 2001 Apr 24; 98(9): 5116-21.
 41. Tzeng JY, Wang CH, Kao JT, Hsiao CK. Regression-Based Association Analysis with Clustered Haplotypes through Use of Genotypes. *Am J Hum Genet*. 2006 Feb; 78(2): 231-42.
 42. Verzilli CJ, Stallard N, Whittaker JC. Bayesian graphical models for genome-wide association studies. *Am. J. Hum. Genet*. 2006; 79:100-112.
 43. Waldron ER, Whittaker JC, Balding DJ. Fine mapping of disease genes via haplotype clustering. *Genet Epidemiol*. 2006 Feb; 30(2): 170-9.
 44. Wang H, Zhang YM, Li X, Masinde GL, Mohan S, Baylink DJ, Xu S. Bayesian shrinkage estimation of quantitative trait loci parameters. *Genetics*. 2005 May; 170(1): 465-80.
 45. Wang H, Thomas DC, PÉer I, Stram DO. Optimal two-stage genotyping designs for genomewide association scan. *Genet. Epidemiol*. 2006; 30: 356-368.
 46. Wang P, Kim Y, Pollack J, Narasimhan B, Tibshirani R. A method for calling gains and losses in array CGH data. *Biostatistics*. 2005 Jan; 6(1): 45-58.
 47. Weir BS, Hill WG, Cardon LR; SNP Consortium. Allelic association patterns for a dense SNP map. *Genet Epidemiol*. 2004 Dec; 27(4): 442-50.
 48. Wigginton JE, Cutler DJ, Abecasis GR. A note on exact tests of Hardy-Weinberg equilibrium. *Am J Hum Genet*. 2005 May; 76(5): 887-93.

Banche dati e recupero delle informazioni

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 (Martin-Sanchez et al., 2003; Maffezzoli e Masseroli, 2005).

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. Queste banche dati rappresentano la maggior fonte di informazione su sequenze biomolecolari (DNA, RNA, proteine), dati sperimentali, conoscenze biomediche, e referenze bibliografiche 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 (Figura 1), al febbraio 2007 si conosceva la sequenza genomica di più di 2000 organismi (inclusi 1.275 virus, 421 batteri e 47 eucarioti), e la sequenza genomica di 426 genomi risultava completa. I genomi più studiati sono quelli dell'uomo, moscerino della frutta *Drosophila melanogaster*, topo, ratto, pesce zebra, della piantina da fiore

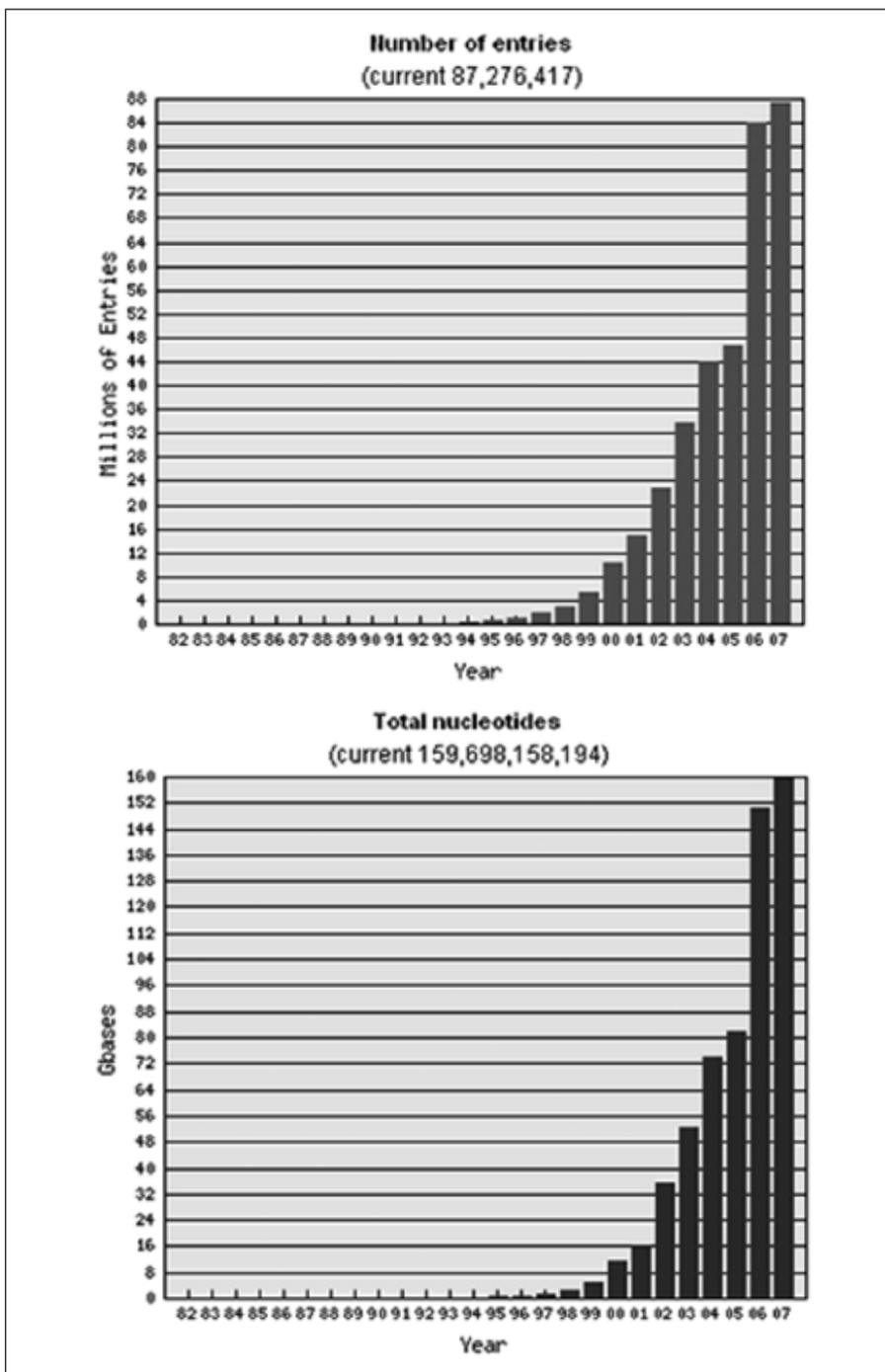


Fig. 1 - Andamento del numero di sequenze nucleotidiche ed entries dal 1982 al 2007 presso la banca dati EMBL (<http://www3.ebi.ac.uk/Services/DBStats/>)

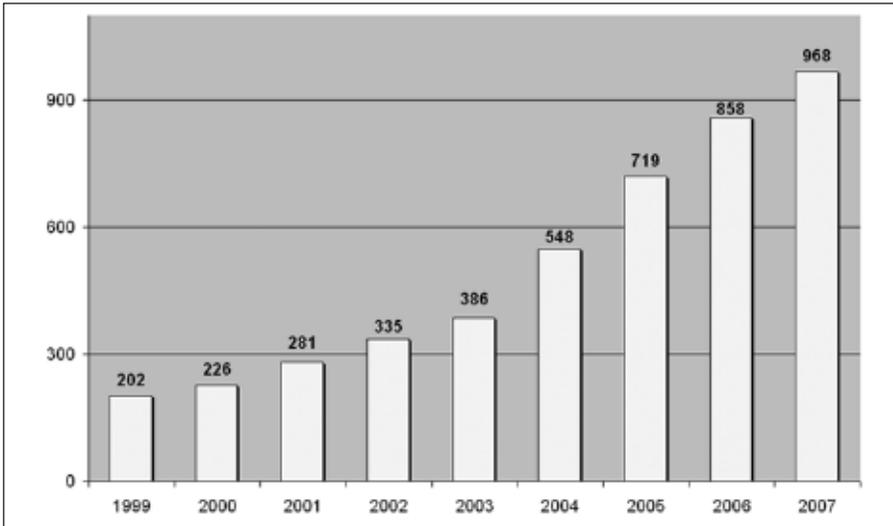


Fig. 2 - Istogramma raffigurante il numero di banche dati genomiche disponibili per anno, dal 1999 al 2007

Arabidopsis thaliana, del batterio escherichia coli, lievito, pisello, mais, e 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 accessibili. Il numero del 2007 riporta una lista di 968 banche dati, 110 in più del precedente anno (Figura 2) (Galperin, 2007).

Tipi di banche dati

Dato l’elevato e crescente numero di banche dati genomiche biomolecolari 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 di queste banche dati 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.

Le principali banche dati primarie di aminoacidi sono:

- Swiss-Prot/TrEMBL (<http://www.expasy.org/sprot/>).
- PIR (Protein Identification Resource) (<http://pir.georgetown.edu/>).
- UniProt (Unified Protein Resource) (<http://www.pir.uniprot.org/>).

Le *banche dati derivate* o *specializzate* raccolgono informazioni molto utili derivate 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 (NCBI RefSeq <http://www.ncbi.nlm.nih.gov/RefSeq/>, Entrez Gene <http://www.ncbi.nih.gov/entrez/query.fcgi?db=gene>, Swiss-Prot);
- 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 di 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) <http://www.renabi.fr/article203.html>, 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 multi-allineate*, come nella banca dati Ribosomal RNA Database (rRNA) (<http://www.psb.ugent.be/rRNA/index.html/>);
- *un set di informazioni complementari a quelle nelle banche dati primarie e specifiche per una ben definita classe di sequenze*, come ad esempio 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://gdbwww.gdb.org/>);
- *banche dati Integrazionali*, 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/>), SOURCE (<http://source.stanford.edu/>) e Harvester (<http://harvester.embl.de/index.html>).

Accessibilità delle banche dati e recupero di informazioni

Le banche dati genomiche sono accessibili tramite diverse modalità (Masseroli, 2004a):

- accesso *diretto*: è raramente permesso per motivi di sicurezza;
- accesso tramite *server FTP*: 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, e i risultati dell'interrogazione sono per lo più forniti in formato HTML.

Per ricercare adeguatamente nelle molte banche dati genomiche biomolecolari le informazioni di interesse in esse raccolte è innanzitutto necessario conoscere e 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 molecolari 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 descrivano 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 biomolecolari 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 dei dati

Le diverse modalità di accesso disponibili alle banche dati genomiche biomolecolari non sono funzionali all'efficiente utilizzo delle informazioni fornite per le attuali necessità dei ricercatori medici e biologi di analisi di liste di geni. Infatti, in genere i ricercatori biomedici hanno bisogno di avere in forma aggregata i soli dati genomici di interesse per lo specifico esperimento o caso in esame, in modo da poterli visualizzare, "sfogliare", e riorganizzare facilmente, e realizzare su di loro interrogazioni articolate per evidenziare eventuali informazioni nascoste e cercare di scoprire nuova conoscenza. Le principali difficoltà derivano dal fatto che i dati raccolti sono distribuiti all'interno di numerose banche dati eterogenee per schema, piattaforme software utilizzate, e contenuti, e che generalmente possono essere interrogate pubblicamente solo via web, per una singola sequenza genomica alla volta, e i cui risultati d'interrogazione sono usualmente 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 interpretare risultati sperimentali ottenuti mediante tecnologie biomolecolari high-throughput di analisi di molte sequenze molecolari in parallelo.

Nonostante gli sforzi per integrare varie annotazioni genomiche fornite da diverse fonti, dati genomici rilevanti sono ancora raccolti in modo distribuito tra banche dati eterogenee. Conseguentemente, l'efficace utilizzo del crescente quantitativo di informazioni disponibili richiede nuovi approcci per riassumere, visual-

lizzare, e confrontare le annotazioni genomiche con lo scopo di rendere possibile svelare nuova conoscenza biomedica.

Tools bioinformatici per il recupero e l'analisi di informazioni

Per l'efficace utilizzo dei dati raccolti nelle banche dati genomiche biomolecolari esistono e si stanno sviluppando vari strumenti bioinformatici implementati per eseguire ricerche ed estrarre dai dati immagazzinati l'informazione più significativa per successive analisi (Masseroli et al., 2004b). 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 molecolare fornita dall'utilizzatore e ognuna delle sequenze registrate all'interno di una banca dati, al fine di trovare tutte le sequenze con una similarità con la sequenza data superiore a una certa soglia. Si basano principalmente sui seguenti tre algoritmi:
 - *Smith-Waterman* (<http://decypher2.stanford.edu/>)
 - *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
 - output che include un range di soluzioni
 - ogni matching (sequenza trovata simile alla sequenza fornita dall'utilizzatore) è accompagnato da una stima statistica della significatività della sua similarità.
- *Text-based tools*: permettono di realizzare ricerche per parole chiave di annotazioni di sequenze molecolari 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 con annotazioni nella banca dati contenenti le parole chiave ricercate. Esempi di tali strumenti sono: *SRS 6* (<http://srs6.ebi.ac.uk/>) e *DBGET/LinkDB* (http://www.genome.jp/dbget-bin/www_linkdb/).

Tuttavia 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, come avviene quando si vogliono interpretare i risultati di esperimenti biomolecolari high-throughput, per esempio ottenuti con microarray, che forniscono liste numerose di geni 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 gene nella lista per poter fare emergere significative caratteristiche comuni, o diverse, esistenti tra tali geni. Per questo fine fondamentale è il ruolo delle annotazioni basate su vocabolari controllati e ontologie, che permettono di raggruppare i diversi geni annotati in categorie funzio-

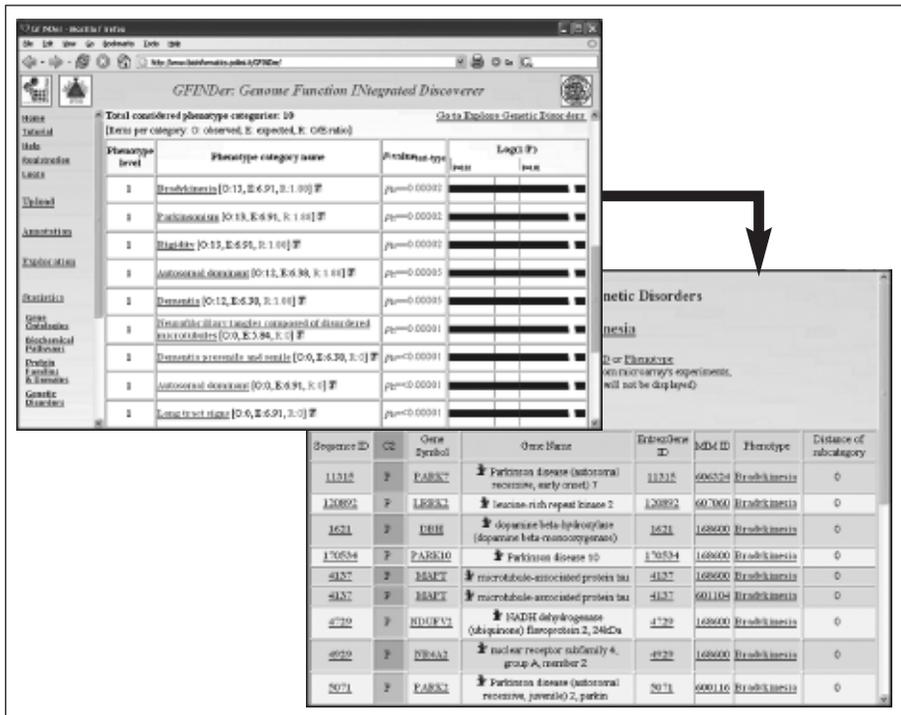


Fig. 3 - Esempio d'utilizzo del modulo "Statistics" "Genetics Disorders" di GFINDER: fenotipi di patologie ereditarie significativamente associati ad una lista di 33 geni correlati con la patologia del Parkinson (18 geni) o dell'Alzheimer (15 geni), e 8 dei 13 di tali geni associati al fenotipo "Bradykinesia".

nali che possono essere analizzate statisticamente (Masseroli e Pinciroli, 2006). Principali esempi di tali strumenti che usano informazioni genomiche pubblicamente disponibili sono: DAVID, GFINDER, FatiGO+, e Onto-Tools. Tra questi, *GFINDER* (Genome Function Integrated Discoverer <http://www.bioinformatics.polimi.it/GFINDER/>) 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 (Figura 3) (Masseroli et al., 2004b; Masseroli et al., 2005).

Bibliografia

1. Galperin MY. "The Molecular Biology Database Collection: 2007 update". *Nucleic Acids Res.* 2007; 35 (Database issue): D3-D4.
2. Maffezzoli A, Masseroli M. "Banche genomiche". In: Pinciroli F, Masseroli M, editors. *Elementi di Informatica BioMedica*. ISBN 88-7398-0171. Milano, IT: Polipress, 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". *BIOINFOMED 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): 731-747.

Modelli per la predizione della struttura genica e scoperta di siti di splicing

Paolo Magni

Dipartimento di Informatica e Sistemistica, Università di Pavia

Il completamento del Progetto Genoma umano (o Human Genome Project - HGP), avvenuto nell'aprile 2003, è stato indubbiamente uno dei più straordinari successi fino ad ora ottenuti nell'ambito della biologia molecolare. Tra gli obiettivi fissati all'inizio del progetto, c'è quello di compilare una lista completa dei geni che codificano le proteine, così da avere "un manuale per poter comprendere completamente il corpo umano". Tuttavia, di questo manuale si conoscono solo i 3 miliardi di lettere che lo compongono, ossia le 4 basi azotate: l'adenosina (A), la guanina (G), la citosina (C) e la timina (T), e una parte dei circa 30 mila capitoli in cui è suddiviso, ossia i geni codificanti le proteine. Dall'HGP sono stati messi in luce due aspetti particolari: il primo che circa il 90% del DNA è formato da sequenze "junk" (o spazzatura), cioè sequenze che non codificano alcuna proteina, ed il secondo che non vi è una corrispondenza biunivoca tra il numero di geni e le centinaia di migliaia di proteine. È stata inoltre confermata negli eucarioti la presenza del meccanismo di splicing, attraverso il quale dalla copia del DNA, detta pre-mRNA, vengono eliminate porzioni di gene non codificanti le proteine (introni) da quelle codificanti (esoni) per formare l'RNA messaggero finale (o mRNA) che verrà tradotto in proteina dai ribosomi. Questo processo non si verifica sempre nello stesso modo ma può variare dando luogo al fenomeno di splicing alternativo. È questo il motivo che consente di ottenere da uno stesso gene proteine differenti che con la loro presenza/assenza possono influire sul funzionamento della cellula e dell'intero organismo, portando, nei casi patologici, a disfunzioni cellulari e quindi a malattie. Il processo di splicing è molto complesso e fino ad oggi non si sono ancora completamente identificati gli elementi che lo regolano. Tuttavia si ritiene che all'interno del gene siano presenti particolari sottosequenze (dette "motivi di base") che consentono ai meccanismi cellulari di splicing di individuare con precisione il confine esone/introne. Da qui nasce l'esigenza di costruire procedure bioinformatiche e modelli matematico/statistici in grado di predire con sufficiente precisione questi siti e quindi la struttura del gene a partire dalla sequenza del DNA.

La sintesi proteica e il processo di splicing

L'inizio del processo di trascrizione avviene grazie al riconoscimento di particolari regioni di DNA, dette promotori e situate prima del punto di trascrizione, alle

quali si lega l'enzima RNA polimerasi II, responsabile del processo. Le regioni, anche chiamate box, sono identificate da specifiche sequenze di basi quali CAAG, GC e TATA. Successivamente un solo filamento del DNA viene usato come stampo per ottenere il pre-mRNA. Solo un decimo del pre-mRNA negli eucarioti più evoluti contiene l'informazione necessaria per la sintesi proteica. Il pre-RNA subisce un processo di maturazione o splicing in cui viene prodotto l'mRNA o RNA messaggero, attraverso la rimozione degli introni (regioni non codificanti) e la conservazione degli esoni. Le sequenze consenso che definiscono i confini di ogni introne sono chiamate: giunzioni di Donor (inizio introne) identificate generalmente dalla coppia GT e giunzioni Acceptor (fine introne) generalmente identificate dalla coppia AG. La rimozione di questi viene effettuata da un complesso di montaggio enzimatico nucleare formato da proteine e da delle piccole molecole di RNA nucleare, le snRNA, che formano questo complesso detto spliceosoma. Le snRNA, identificate dalle lettere U1, U2, U4, U5 e U6 svolgono le seguenti azioni coordinate:

- prima U1 si lega all'estremità 5' dell'introne;
- successivamente U2 si fissa all'adenosina in un punto detto punto di ramificazione o Branch-point;
- quindi entrano in gioco il complesso U4/U6 e U5;
- poi U4 si dissocia, permettendo così a U6 di accoppiarsi con U2;
- si forma il lazzo e la separazione dell'introne dall'esone;
- infine gli esoni si accoppiano, lo spliceosoma si dissocia e l'introne viene scomposto in monomeri.

Una volta ottenuto l'mRNA, questo migra dal nucleo della cellula e viene convertito o tradotto in sequenze aminoacidiche che costituiscono le proteine.

Metodi computazionali per la predizione delle sequenze biologiche

Ci sono due approcci bioinformatici fondamentali all'individuazione di possibili geni a partire dalla sequenza del DNA: il primo consiste nella ricerca per similarità, mentre il secondo (detto predizione ab initio) usa le conoscenze statistiche sulla struttura dei geni per costruire modelli matematico/statistici predittivi.

Il metodo di ricerca per similarità si basa sostanzialmente sulla ricerca di somiglianze tra le sequenze genomiche da analizzare e le EST (Expressed Sequence Tag), le proteine note e altre sequenze genomiche annotate provenienti anche da organismi diversi. Questo approccio si fonda sull'ipotesi che le regioni codificanti (esoni) siano fortemente conservate durante il processo di evoluzione rispetto alle regioni non codificanti (introni e regioni intergeniche). Una volta trovata una somiglianza tra una particolare regione genomica ed una EST, una sequenza di DNA di un gene che codifica per una proteina nota, o una proteina, l'informazione ottenuta può essere usata per dedurre la struttura di un gene. Tuttavia questo metodo di ricerca presenta lo svantaggio di non riuscire a predire l'intera struttura del gene, ma solo una piccola frazione di esso. L'altra famiglia di metodi per l'identificazione automatica dei geni si basa sull'approccio ab initio, in cui

si cerca di scoprire i geni e la loro struttura a partire dalle conoscenze ricavate dai geni sperimentalmente validati. La predizione viene fatta utilizzando due tipi di informazioni presenti: la differente composizione delle sequenze codificanti e non codificanti e le regioni consenso dei confini introne/esone o di altri segnali nella sequenza quali quelli della regione identificante il punto di ramificazione o Branch-Point. Il metodo di predizione *ab initio* si basa fundamentalmente su tre passaggi consecutivi:

- definizione della struttura di un modello statistico/matematico in grado di descrivere sequenze di DNA che rappresentano un gene o una sua porzione caratteristica (ad esempio i siti di splicing, il Branch-Point, ecc.);
- addestramento del modello su sequenze di cui è nota la struttura e la suddivisione introni/esoni per catturare l'informazione caratteristiche necessarie per effettuare predizioni e classificazioni il più possibilmente "corrette";
- data una sequenza di cui si ignora la struttura, predizione della struttura o della posizione di punti caratteristici (es. siti di splicing, codone di start, ecc.).

Gli algoritmi maggiormente utilizzati per la predizione *ab initio* si basano sulla programmazione dinamica, su gli Hidden Markov Models, sulle Neural Network, sulle Bayesian Network e, recentemente, sulle Support Vector Machines. Il problema fondamentale legato a tutti questi modelli di predizione è l'elevato numero di falsi positivi, cioè la bassa specificità dei vari metodi. Questo generalmente accade perché l'informazione riassunta nelle sequenze consenso dei vari metodi non è abbastanza dettagliata e specifica.

L'altro aspetto fondamentale comune a tutte le metodologie ed in generale ai problemi di machine learning è che tutti i modelli devono essere addestrati su set di esempi, possibilmente vasti. La disponibilità di un buon training set rappresenta tutt'oggi un problema. Infatti nonostante l'informazione sulla struttura dei geni e la loro divisione in introni/esoni sia disponibile per molti organismi nelle varie banche dati biologiche sia primarie che derivate anche grazie ai vari progetti genoma, tale informazione è spesso inaccurata e inconsistente e ridondante.

Pertanto un punto cruciale spesso sottovalutato in letteratura è la costruzione di training set accurati e possibilmente di buone dimensioni. Alcune ricerche che abbiamo condotto hanno consentito di costruire insieme affidabili di esempi a partire da due banche dati del NCBI: la banca dati REFSEQ e la banca dati dei geni GENE. Infatti, la banca dati REFSEQ include un subset non ridondante di sequenze dei vari organismi. Inoltre ad ogni sequenza è associata una proprietà detta Status Key che definisce il livello di accuratezza dell'informazione associata ai geni. I possibili status, periodicamente aggiornati, sono qui di seguito riportati.

- Lo status INFERRED che viene attribuito quando non ci sono supporti sperimentali relativi all'intera sequenza. Si possono però avere supporti tramite omologia.
- Lo status MODEL che rappresenta i record la cui struttura è stata predetta dall'analisi della sequenza genomica tramite una predizione *ab initio* oppure una predizione per omologia.

- Lo status **PREDICTED** che viene attribuito quando il record è predetto e non soggetto ad una revisione individuale. Il trascritto può rappresentare una predizione ab initio o derivare da altri trascritti precedentemente definiti. In entrambi i casi, la proteina è predetta. Il supporto per il trascritto può includere l'esistenza di cloni cDNA ed EST (Expressed Sequence Tag).
- Lo status **PROVISIONAL** è associato quando il record non è ancora stato soggetto ad una revisione individuale. Si può pensare comunque che sia ben supportato e che rappresenti una valida proteina. Provisional è lo status di default applicato ai genomi per i quali non ci sono chiare informazioni sul metodo usato per definirne la sequenza.
- Lo status **REVIEWED** caratterizza i record che sono stati rivisti dallo staff del NCBI. In questo frangente è stato effettuato un confronto con le sequenze omologhe disponibili e spesso è stato validato anche dalla letteratura. Alcune volte le caratteristiche del record sono state definite da progetti esterni.
- Lo status **VALIDATED** che individua i record che hanno subito tutti i processi di annotazione tra cui una revisione individuale da parte dello staff di NCBI. È il livello di accuratezza migliore.

Tra le principali caratteristiche di REFSEQ vi sono la possibilità di avere collegamenti espliciti alla sequenza nucleotidica e proteica di ogni elemento; un aggiornamento continuo e controllato delle informazioni presenti nel database; un formato consistente e sintetico delle sequenze rappresentate (FASTA format); la possibilità di visualizzare le versioni precedenti e gli aggiornamenti relativi ad ogni record. A partire da queste considerazioni nella relazione verranno illustrate i principi fondamentali sottostanti ad alcuni dei modelli di predizione ab initio più semplici ed importanti.

Bibliografia

1. B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, P. Walter. *Biologia molecolare della cellula*. Quarta edizione. Zanichelli, 2004.
2. ST. Eskesen, FN. Eskesen, A. Ruvinsky. Natural selection affects frequencies of AG and GT dinucleotides at the 5' and 3' ends of exons. *Genetics*, 2004, 167(1): 543.
3. M.B. Shapiro and P. Senapathy. RNA splice junction of different classes of eukaryotes: sequence statistic and functional implications in gene expression. *Nucleic Acids Research*, 1987, 15(17): 7155-7174.
4. Z. Wang, Y. Chen, Y. Li. A brief review of computational gene prediction methods. *Genomics Proteomics Bioinformatics*, 2004, (4): 216-21.
5. K.D. Pruitt, T. Tatusova, D.R. Maglott. NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Research*, 2005, 1(33): 501.
6. W.H. Majoros and S.L. Salzberg. An empirical analysis of training protocols for probabilistic gene finders. *BMC Bioinformatics*, 2004, 5: 206.
7. M. Burset, I.A. Seledtsov, V.V. Solovyev. Analysis of canonical and non-

- canonical splice sites in mammalian genomes. *Nucleic Acids Research*, 2000, 28(21): 4364-75.
8. T.M. Chen, C.C.Lu, and W.H. Li. Prediction of splice sites with dependency graphs and their expanded bayesian networks. *Bioinformatics*, 2005, 21(4): 471-482.
 9. D.Castanotto, J.Rossi. Small sequence insertions within the branch point region dictate alternative sites of lariat formation in a yeast intron. *Nucleic Acids Research*, 1992, 20(24): 6649-55.
 10. J.W.Fickett. Recognition of protein coding regions in DNA sequences. *Nucleic Acids Research*, 1982, 10(17): 5303-18.
 11. G. Aggarwal, R.Ramaswamy. Ab initio gene identification: prokaryote genome annotation with GeneScan and GLIMMER. *The Journal of Biosciences*, 2002, 27(1 Suppl 1): 7-14.
 12. Sven Dergroeve, Yvan Saeys, Bernard De Beats, Pierre Rouz , Yves Van de Peer. SpliceMachine: predicting splice sites from high-dimensional local context representations. *Bioinformatics*, 2005, 21(8): 1332-1338.
 13. Zhuo wang, Yazhu Chen, Yixue Li, A brief review of computational gene prediction methods. *Genomics Proteomics Bioinformatics*, 2004, (4): 216-221.
 14. Pierre Rouz , Gene prediction Genome annotation, <http://www.psb.rug.ac.be>.
 15. Bill Majoros, Eukaryotic Gene Prediction, <http://www.tigr.org>.
 16. M. Pertea, X. Lin, S.L. Salzberg, Genesplicer: a new computational method for splice site prediction. *Nucleic Acid Research*, 2001, (29): 1185-1190.
 17. L.P. Lim, C.B. Burge, A computational analysis of sequence features involved in recognition of short introns. *Proceeding of the National Academy of Science*, 2001, 98: 11193-11198.
 18. M. Bocassini. Creazione di basi di dati non ridondanti di sequenze di DNA per l'analisi di siti di splicing nell'uomo, 2005, Universit  di Pavia, Italy.
 19. D. Bianchini. Sviluppo e implementazione di un modello per la predizione di siti di splicing nel pre-mRNA, 2006, Universit  di Pavia.
 20. GB. Hutchinson, MR. Hayden, SORFIND: a computer program that predicts exons in vertebrate genomic DNA. *Proceedings of the Second International Conference on Bioinformatics, Supercomputing and Complex Genome Analysis*, 1993: 513-520.
 21. T. Alun, M.H. Skolnick, A probabilistic model for detecting coding regions in DNA sequences. *IMA Journal of Mathematics Applied in Medicine and Biology*, 1994, 11, 149-160.
 22. G. Aggarwal, R. Ramaswamy, Ab initio gene identification: prokaryote genome annotation with GeneScan and GLIMMER. *The Journal of Bioscience*, 2002, 27(1 Suppl 1): 7-14.
 23. M. Stanke, R. Steinkamp, S. Waak1, B. Morgenstern, AUGUSTUS: a web server for gene finding in eukaryotes. *Nucleic Acids Research*, 2004, 32:309-312.
 24. National Human Genome Research Institute, <http://www.genome.gov/>.
 25. GRAIL web site, <http://avalon.epm.ornl.gov>.
 26. National Center for Biotechnology Information, <http://www.ncbi.nlm.nih.gov/>.

Analisi di dati di microarrays di espressione: predizione di funzione, studio di pathways e reti di regolazione

Riccardo Bellazzi, Lucia Sacchi

Gruppo di Bioinformatica e Data Mining, Laboratorio di Informatica Medica,
Dipartimento di Informatica e Sistemistica, Università di Pavia

Il sequenziamento del genoma e il parallelo sviluppo di nuove tecniche computazionali in biologia molecolare hanno permesso di intraprendere un'analisi sistematica dei meccanismi micromolecolari alla base dei sistemi biologici. In particolare, lo sviluppo della tecnologia delle matrici ad alta densità, o microarray, di oligonucleotidi e di cDNA ha reso possibile estendere all'intero genoma la misura del livello di espressione genica, che precedentemente era limitata ad un ridotto numero di geni selezionati in base ad ipotesi a priori, dando così la possibilità di “fotografare” il livello di trascrizione di tutti i geni, in un dato istante, in un dato tessuto o in una specifica condizione fisiologica.

Per analizzare l'enorme quantità di dati resa disponibile dalla tecnologia dei microarray è fondamentale identificare gli strumenti di data mining più appropriati tra i molti disponibili dalla statistica, dall'intelligenza artificiale, dall'analisi del segnale, dalla pattern recognition e, se necessario, mettere a punto nuove metodologie di analisi. Tipicamente i dati da analizzare nell'ambito di un disegno sperimentale provengono da più microarray, relativi a diversi fenotipi, ad esempio di soggetti appartenenti a popolazioni diverse (es. normali e patologiche) oppure studiati in condizioni diverse (es. dopo diversi trattamenti farmacologici), oppure provenienti da tessuti diversi.

Per lo studio dei meccanismi di regolazione trascrizionale, si stanno inoltre diffondendo disegni sperimentali che producono serie temporali di dati, relativi al profilo di espressione genica nella transizione tra due diversi stati stazionari. Grazie a queste osservazioni, è possibile inferire delle relazioni temporali fra le espressioni dei geni e, quindi, formulare delle ipotesi sulla struttura delle reti di regolazione. Questa attività è chiamata “reverse engineering”, in quanto affronta il problema inverso di formulare delle ipotesi sul sistema dinamico che ha generato i dati sulla base dell'osservazione dei dati stessi. In questa relazione verranno descritte le più recenti metodologie bioinformatiche impiegate per la predizione di funzione e per derivare reti di regolazione a partire da dati di microarrays.

Analisi di dati di microarrays

Le tecnologie

Come già sottolineato, la crescente diffusione dei microarray è data dal fatto che essi consentono di monitorare contemporaneamente tutta l'informazione genetica di un organismo con svariati scopi, ad esempio scoprire nuovi geni, caratterizzare il profilo genetico di un organismo, o collegare la risposta terapeutica ad un farmaco con il profilo genetico del paziente. Lo sviluppo dei microarray è iniziato nel 1995 con l'introduzione delle due principali tecnologie che rimangono attualmente le più utilizzate: i microarrays ad oligonucleotidi (Lockhart DJ, 1996) e a cDNA (o *spotted*) (Schena M, 1996). Entrambe le tecnologie sfruttano la complementarità che caratterizza RNA e DNA e i due filamenti di DNA, mentre si differenziano per il processo di fabbricazione. Nel caso dei cDNA microarrays singoli o doppi filamenti di DNA pre-sintetizzati vengono depositati sull'array (Schena M, 1996; De Risi JL, 1997), mentre per gli array ad oligonucleotidi gruppi di basi (o *oligo-meri*) sono sintetizzati direttamente *in situ* su wafer di vetro mediante una tecnica fotolitografica (Chee M, 1996; Lockhart DJ, 1996). I cDNA microarrays possono essere prodotti in laboratorio o acquistati da fornitori commerciali specializzati, mentre i microarray ad oligonucleotidi vengono realizzati solo industrialmente; questi ultimi sono oggi disponibili su diverse piattaforme tecnologiche, anche se gli array più diffusi rimangono i GeneChip® arrays prodotti da Affymetrix (Harrington CA, 2000). Sulle piattaforme a cDNA i geni vengono generalmente rappresentati da singoli frammenti di DNA lunghi parecchie centinaia di basi. I campioni di cDNA retro-trascritti dagli mRNA di cellule trattate e non trattate (controlli) vengono ibridizzati sull'array e sono nella maggior parte dei casi marcati utilizzando dei nucleotidi etichettati con coloranti fluorescenti (generalmente Cy3- and Cy5-dUTP). Una volta marcati, i cDNA vengono ibridizzati sull'array; le quantità relative di uno specifico trascritto nei due campioni (trattato e controllo) vengono determinate misurando le intensità del segnale per ciascun colorante e calcolandone il rapporto. Sui GeneChip® array (Affymetrix, Inc, Santa Clara, CA), ogni gene viene rappresentato da 11-20 gruppi di 25 oligonucleotidi che servono come detettori sequenza-specifici. Ogni oligonucleotide, chiamato perfect match (PM) viene scelto sulla base della sequenza del gene, e deve essere estremamente sensibile e specifico per il gene stesso. Un elemento di controllo, chiamato mismatch (MM), viene accoppiato ad ogni PM ed è utilizzato per monitorare l'ibridazione non specifica. Una sequenza di mismatch consiste in una sequenza identica alla controparte PM eccetto che per la base centrale (la tredicesima). La presenza dei mismatch permette che la cross-ibridazione e il background locale possano essere stimati e sottratti al segnale di interesse. Anche in questo caso l'intensità del segnale misurato è un indice della quantità di mRNA presente nella cellula per uno specifico gene. Come vedremo nel seguito, esistono diverse metodologie per estrarre il segnale dalle misure di PM e MM. Indipendentemente dall'algoritmo utilizzato per l'estrazione del segnale, la misura che si ottiene tramite microarrays ad oligonucleotide è una misura assoluta, e non un rapporto come invece si ha nel caso di microarrays a cDNA.

Il problema del pre-processing e della normalizzazione

L'analisi di dati provenienti da DNA microarrays inizia subito dopo l'ibridazione del campione all'array. Le prime fasi dell'analisi comprendono lo scanning dell'array per rilevarne l'immagine, la quantificazione delle intensità, la pre-elaborazione e la normalizzazione dei dati.

Analisi e acquisizione dell'immagine

Il primo passo dell'acquisizione dell'immagine è il processo di scanning, che consiste nell'acquisizione degli spettri di emissione dei vari pixel dell'array. In un processo che viene identificato come *segmentazione*, vengono identificate le zone dell'array in cui sono posizionate le sonde di nucleotidi specifiche per ciascun gene, i pixel che corrispondono effettivamente al segnale (foreground) e quelli che invece corrispondono al background. Il passo successivo alla segmentazione è il processo di *intensity extraction*, durante il quale le intensità dei pixel del background e del foreground vengono elaborate al fine di determinare la misura di intensità del segnale corrispondente a ciascuna sonda posizionata sull'array. È bene sottolineare che, a seconda della tecnologia impiegata, il processo di estrazione del segnale dall'immagine avviene attraverso metodologie differenti. Esistono diversi software, sia commerciali che freeware, per l'analisi dell'immagine da microarray (Scan Analyse, GenePix e QuantArray tra i più popolari) e si sta sviluppando in questo settore un ampio spazio di ricerca, relativo allo studio di metodologie atte a minimizzare le fonti di rumore.

Pre-processing

I dati ottenuti dopo l'analisi dell'immagine vengono pre-elaborati allo scopo di eliminare i segnali di scarsa qualità o di scarsa intensità rispetto al background e di calcolare, a partire dal segnale delle sonde corrispondenti ad un determinato gene, il segnale di intensità di espressione genica. Anche questo tipo di analisi dipende dalla tecnologia utilizzata; nel caso di microarray Affymetrix, ad esempio, lo scopo è quello di estrarre una sorta di misura che sintetizzi il segnale che viene emesso dalle 11-20 sonde che rappresentano uno specifico gene. A questo scopo sono state proposte diverse metodologie, che si basano su diversi modelli del segnale e del background corrispondente. Esempi di questi algoritmi sono MAS5.0 (Hubbell E, 2002), il software direttamente integrato nel tool di analisi Affymetrix, il Model Based Expression Index (MBEI) (Li C, 2001), ed RMA (Robust Multichip Average) (Irizarry RA, 2003), un interessante algoritmo che racchiude in sé le tre fasi di rimozione del background, normalizzazione tra più chip ed estrazione del segnale.

Normalizzazione

L'ultimo step della pre-elaborazione riguarda la normalizzazione, ovvero la rimozione degli errori sistematici, che possono avere origine da diverse fonti quali l'efficienza di incorporazione del fluorescente, diversi setting nei parametri dello scanner, diverse efficienze di ibridazione in zone diverse dell'array, diverse effi-

cienze di ibridazione in array diversi, ecc.. Il processo di normalizzazione è pertanto indispensabile nel caso si debbano paragonare valori di espressione in array diversi o, nel caso degli array a cDNA, nel medesimo array. Diversi metodi sono stati proposti in letteratura per quanto riguarda la normalizzazione; questi metodi si basano su ipotesi diverse che dipendono in larga misura anche dalla tecnologia considerata. Il metodo noto come Global Normalization, ad esempio, assume che l'espressione della maggior parte dei geni monitorati non vari significativamente nelle condizioni sperimentali analizzate; i dati vengono quindi normalizzati in modo che la mediana (o la media) abbia lo stesso valore in array diversi (o nello stesso array nel caso dei microarray a due canali). Una volta normalizzati, i dati vengono generalmente espressi in scala logaritmica per avere una rappresentazione efficace di valori che tipicamente variano in un range molto ampio. Per gli array Affymetrix, tale scelta è anche motivata dal fatto che ci sono evidenze di una proporzionalità del segnale al logaritmo dell'espressione genica misurata (Irizarry RA, 2003).

Selezione di geni differenzialmente espressi

La fase successiva alla pre-elaborazione è quella che si concentra sulla selezione dei geni la cui espressione varia significativamente nella condizione sperimentale considerata rispetto ad una situazione di 'controllo'; i geni estratti durante questo processo vengono detti differenzialmente espressi. Anche per quanto riguarda il tema della selezione genica, molti sono stati i metodi proposti in letteratura, di cui daremo una breve illustrazione nel seguito.

L'approccio più comunemente utilizzato in letteratura fin dalle prime pubblicazioni sui microarray e che viene tuttora largamente impiegato nel caso in cui nell'esperimento non siano presenti repliche è quello basato su *fold-change* (FC) (Sreekumar R, 2002). Secondo questa tecnica, un gene viene dichiarato differenzialmente espresso se il suo livello di espressione varia per più di un fattore costante, fissato in modo empirico (tipicamente $FC=2$) nelle situazioni sperimentali analizzate. Sebbene sia semplice e molto intuitivo, questo metodo rivela anche molti svantaggi, tra cui quello di non considerare la dipendenza dall'intensità della variabilità sperimentale penalizzando i geni espressi ad alta intensità rispetto a quelli, più rumorosi, a bassa intensità di espressione.

Per la selezione di geni differenzialmente espressi in presenza di repliche delle misure, cioè nel caso in cui siano disponibili misure multiple, provenienti da più microarray, per ciascuna delle condizioni analizzate, sono stati spesso usati dei test statistici di significatività, tra cui il t-test (utilizzato per il confronto tra due popolazioni) e il test ANOVA (utilizzato per il confronto in più di due popolazioni). Anche questi metodi hanno mostrato tuttavia alcuni punti di debolezza, il primo tra i quali è l'assunzione di gaussianità delle variabili analizzate. Un secondo problema deriva dal fatto che le poche repliche in genere disponibili non permettono di stimare in maniera accurata la variabilità sperimentale, monitorandone ad esempio la dipendenza dall'intensità. Infine, il fatto di applicare un test statistico a migliaia di geni contemporaneamente rende necessaria l'applicazione di una correzione sul

livello di significatività. Infatti, quante più ipotesi vengono testate contemporaneamente, tanto più aumenta la probabilità di selezionare falsi positivi, ovvero di giudicare come differenzialmente espressi geni che non lo sono (errore di tipo I). Se lo scopo dell'esperimento con i microarray è quello di identificare dei geni come "marcatori" di uno stato fisiologico con un basso margine di incertezza, l'errore di tipo I viene definito come la probabilità di selezionare almeno un gene come differenzialmente espresso quando non lo è e, di conseguenza, al livello di significatività α viene applicata la correzione di Bonferroni. Un controllo così forte dell'errore di tipo I non è invece giustificabile quando lo scopo è, ad esempio, lo screening del profilo molecolare in due o più condizioni sperimentali. In tal caso si utilizzano criteri che tengano conto del rapporto tra potenza del test ed errore di tipo I o tra numero di geni selezionati ed errore di tipo I, si veda (Dudoit S, 2002) per una review sull'argomento. In letteratura sono state proposte numerose alternative ai test classici, tra cui spicca l'usattissimo algoritmo Significance Analysis of Microarrays (SAM). Questo metodo si basa sull'utilizzo di un t-test modificato per tener conto dell'alta variabilità a bassi livelli di espressione. Per trovare i geni differenzialmente espressi, l'algoritmo utilizza una strategia di permutazione e ricampionamento dei dati e un criterio di soglia basato sulla *false discovery rate*, ovvero il rapporto tra falsi positivi e geni selezionati sul data set originale. Esempi di applicazione di SAM al confronto tra due e tra tre gruppi si possono trovare in (Tusher VG, 2001), e (Hansel et al. 2003).

Una discussione a parte merita il problema della selezione dei geni a partire da esperimenti temporali. In tali esperimenti, in cui l'espressione genica viene monitorata durante la transizione tra due stati stazionari indotta da una specifica perturbazione, l'obiettivo è infatti quello di selezionare i geni la cui espressione presenti in uno o più campioni temporali delle variazioni significative rispetto ad un livello di baseline. Per selezionare solo i geni che variano per effetto della perturbazione e non per effetto di altri fattori, il disegno sperimentale spesso prevede la misura dei profili temporali sia in cellule sottoposte alla perturbazione che in cellule non trattate. In questi casi l'obiettivo si sposta sulla selezione dei geni la cui espressione presenti in uno o più campioni temporali delle variazioni significative rispetto al livello di espressione misurato nelle cellule di controllo a tempi corrispondenti. Nella maggior parte dei casi, i data set di serie temporali di dati non sono provvisti di repliche in numero sufficiente da poter applicare un test statistico, per cui spesso si ricorre all'utilizzo di una fold-change.

Nell'ambito della selezione genica in esperimenti temporali, recentemente sono stati fatti alcuni sforzi per integrare in maniera più efficace la dimensione temporale; in (Di Camillo B, 2005), ad esempio, viene descritto un metodo per la selezione di geni differenzialmente espressi basato sull'area sottesa dalle serie temporali di trattati e controlli.

Predizione di funzione

Una volta selezionati i geni il cui profilo di espressione risulta rilevante ai fini dell'esperimento, le fasi successive dell'analisi di dati di microarray si pongono

nell'ottica di soddisfarne uno degli obiettivi principali, ovvero la comprensione del ruolo biologico dei geni selezionati. Nel caso ad esempio di esperimenti in cui si analizzano i profili di espressione di un set di geni nelle due condizioni sperimentali di trattato e controllo, far luce sui processi in cui i geni differenzialmente espressi siano coinvolti può portare a trarre importanti conclusioni sull'effetto del trattamento. Attraverso specifici tool i vari geni vengono *annotati*, cioè associati alle diverse funzioni biologiche, organizzate in una vera e propria tassonomia chiamata *Gene Ontology*; prima di fare questo, risulta utile effettuare una fase di *clustering*. Il clustering è un potente strumento esploratorio che ha l'obiettivo di partizionare i geni in gruppi omogenei e ben separati, sulla base dei loro profili di espressione. L'assunzione chiave alla base di questo approccio è che geni che presentano profili di espressione simili (geni co-espressi) siano funzionalmente correlati.

Uno dei metodi oggi maggiormente utilizzati nelle analisi di DNA microarray è il clustering gerarchico agglomerativo, una cui versione per dati temporali è stata proposta da Eisen et al. (Eisen MB, 1998). In questo algoritmo i geni 'simili' vengono identificati sulla base del calcolo del coefficiente di correlazione standardizzato; secondo una strategia gerarchica, ogni passo consiste nell'unione di due cluster sulla base del coefficiente di similarità prescelto fino a che tutti i geni non appartengano ad un unico gruppo. Il risultato viene visualizzato attraverso un grafico ad albero detto *dendrogramma*, in cui l'unione di due elementi è rappresentata da due rami dell'albero che si uniscono in un nodo e l'altezza di rami è inversamente proporzionale alla similarità fra gli elementi: quanto più i cluster sono simili, tanto più il ramo è corto, e viceversa. Oltre al clustering basato su distanza come quello gerarchico appena introdotto, altre interessanti alternative sono rappresentate da metodi basati su modello, capaci di tenere specificamente conto della natura dinamica dei profili e di trovare il numero ottimo di clusters sulla base dei dati a disposizione (Ramoni MF, 2002; Ferrazzi F, 2005), e da metodi basati su una descrizione qualitativa dei profili, che consentono una rappresentazione intuitiva dei risultati ottenuti (Hvidsten TR, 2003; Sacchi L, 2005a). Nei metodi di clustering basati su modello si assume che i dati siano generati in modo casuale a partire da una popolazione costituita da un certo numero di sottopopolazioni (i cluster), ciascuna caratterizzata da una distribuzione di probabilità. Il modello è rappresentato dall'insieme dei cluster e dalle loro distribuzioni probabilistiche. La scelta del modello, cioè la scelta del numero di cluster e le distribuzioni, viene affrontata tramite tecniche di inferenza statistica. Nel caso delle serie temporali, ciascun cluster viene caratterizzato da un diverso processo stocastico che si suppone abbia generato le serie in esso contenute.

Un importantissimo passaggio che segue la fase dell'estrazione dei risultati attraverso metodi computazionali, come ad esempio gli algoritmi di clustering, è quello del cosiddetto *gene enrichment*, ossia l'integrazione delle informazioni estratte con elementi di conoscenza biologica, quali ad esempio l'annotazione dei geni coinvolti, la determinazione del loro coinvolgimento in determinati pathways, ecc., al fine di estrarre informazione utile per far luce sui complessi meccanismi biologici coinvolti (Subramanian A, 2005).

Come precedentemente descritto, l'ontologia di maggior successo per quanto riguarda l'organizzazione della conoscenza in genomica funzionale è la *Gene Ontology* (GO) (<http://www.geneontology.org/>), un vocabolario che definisce in dettaglio le funzioni molecolari che i prodotti genici possono avere, i processi biologici cui possono prendere parte e le localizzazioni cellulari nelle quali possono essere attivi.

La GO è un esempio di Ontologie Biologiche Aperte (OBO), un insieme di ontologie libere che possono essere utilizzate finché le loro fonti sono riconosciute ed utilizzate dalla comunità scientifica. Ogni termine della GO rappresenta un nodo in una visualizzazione grafica ottenuta attraverso un DAG (Directed Acyclic Graph), un grafico simile ad un albero ma con la differenza che in un DAG è possibile che un nodo sia generato da più di un "genitore". Ciascun gene segue un percorso di annotazione che lo porta dalla categoria più generale, per esempio *processo biologico*, attraverso sottocategorie intermedie (ad es. *ciclo cellulare*), fino al processo più specifico in cui il gene è coinvolto (ad es. *apoptosi*). Tanto maggiore è la conoscenza disponibile sul ruolo biologico di un gene, tanto più dettagliato sarà il suo percorso lungo l'albero. Esistono numerosissimi tool che consentono di associare queste categorie ai risultati delle analisi, con lo scopo principale di confrontare classi di geni per valutare la significatività dei risultati ottenuti sia dal punto di vista dei geni coinvolti che delle proteine da essi prodotti che dei pathway di cui fanno parte. Tra i numerosi tool presenti in rete possiamo ricordare DAVID (Database for Annotation, Visualization and Integrated Discovery, <http://david.abcc.ncifcrf.gov/>), i tool disponibili nel pacchetto GeneTools (<http://www.genetools.microarray.ntnu.no/>) (Beisvag V, 2006), il tool Panther (<http://www.pantherdb.org/>) (Thomas PD, 2003) ed i tool resi disponibili da Affymetrix (NetAffx™ Analysis Center, <http://www.affymetrix.com>).

Inferenza di reti di regolazione

Oltre agli studi di genomica funzionale, nel corso degli ultimi anni sono stati proposti molti metodi per la generazione di reti di regolazione a partire da dati di DNA microarray (de Jong H, 2002; Bellazzi R, 2004), attraverso un processo noto come "*reverse engineering*". Gli "ingredienti" della maggior parte degli algoritmi per l'inferenza di reti di regolazione sono una rete di interazioni, che esprime quali siano i legami fra i geni, ed un modello delle interazioni, in grado di descrivere la dinamica del sistema. Una rete di interazioni tra geni viene solitamente rappresentata tramite un *grafo*. Per quanto riguarda la modellizzazione della dinamica sono state invece proposte diverse soluzioni, che comprendono sistemi di equazioni differenziali ed alle differenze, modelli stocastici e modelli qualitativi. Alcuni approcci proposti in letteratura si concentrano sull'apprendimento del grafo (Butte AJ, 2000; Sacchi L, 2005b), mentre metodologie più avanzate per l'apprendimento automatico sia della struttura delle reti di regolazione sia della loro dinamica sono costituite da reti Booleane (Liang S, 1998; Shmulevich I, 2002), reti Bayesiane (Friedman N, 2000; Segal E, 2003a; Ferrazzi F, 2006) e reti basate su equazioni differenziali (D'haeseleer P, 1999).

Sebbene i dati provenienti da esperimenti con DNA microarray consentano una visione di insieme dell'attività dei geni sino ad oggi impensabile, la caratterizzazione completa di una rete di regolazione genetica è complicata dal grande numero di variabili coinvolte, che possono comprendere sia grandezze intracellulari, come la quantità di mRNA sintetizzata a partire da un gene, sia grandezze extracellulari, come gli ormoni secreti da altri organi per comunicare "segnali" alle cellule sia altre grandezze, quali le "condizioni ambientali" esterne. L'inferenza di reti di regolazione genetica a partire dai dati dei DNA microarrays e la loro successiva "validazione" sperimentale rimangono pertanto un problema aperto. In particolare, l'integrazione di dati dinamici genomici e proteomici rappresenta un passo cruciale nello sviluppo di approcci di "systems biology". La maggior parte degli studi presenti in letteratura considera l'interazione solo in condizioni statiche, ricorrendo generalmente ad analisi di correlazione (Chen G, 2002; MacKay VL, 2004; Tian Q, 2004). Di recente, alcuni autori hanno proposto un approccio più complesso, basato su un modello concettuale probabilistico per l'integrazione di dati funzionali genomici di differente estrazione basato su un punteggio di verosimiglianza derivato mediante l'utilizzo di un approccio di tipo statistico-Bayesiano (Lee I, 2004).

Un'altra interessante direzione di ricerca è rappresentata nell'integrazione fra i dati e la conoscenza a priori disponibile sui processi cellulari durante la costruzione della rete di regolazione. L'esplicita considerazione delle informazioni provenienti da diverse sorgenti di conoscenza nel processo di analisi dei dati può permettere di prevenire la scoperta di relazioni banali, complementare le nuove ipotesi generate con ciò che è già noto, evitare di estrarre relazioni con eccessiva confidenza e, quindi, costruire in modo incrementale nuova conoscenza. Fra i diversi approcci proposti in merito è possibile citare i lavori pubblicati sulle reti Bayesiane dinamiche. Ad esempio, Bernard and Hartemink (Bernard A, 2005) hanno proposto un modello che permette loro di costruire una rete di relazioni probabilistiche fra le variabili partendo da una rete iniziale che si basa su un database di fattori di trascrizione, mentre Nariai e colleghi (Nariai N, 2005) hanno proposto un approccio per apprendere in modo congiunto reti di regolazione geniche e reti di interazioni fra proteine.

Conclusioni

In questo capitolo abbiamo evidenziato come l'analisi dei dati di microarray possa dar luogo alla generazione di ipotesi su geni candidati di patologie e come l'utilizzo di metodi matematici possa essere un utile strumento nello studio dei sistemi di regolazione genica. Da un punto di vista metodologico, il principale problema connesso all'applicazione delle tecniche matematico statistiche è rappresentato dalla dimensionalità tipica degli esperimenti con i microarray, in cui sono presenti "pochi" esperimenti e molte variabili (i geni). Un ulteriore limite di queste tecniche deriva dalla natura stessa dei dati di espressione genica, che non consentono di ricavare l'intero insieme di processi regolatori, ma soltanto di evidenziare delle relazioni "fenomenologiche" (ad es. ogni volta che il gene A è

sovraespresso il gene B è sottoespresso). Queste relazioni devono quindi essere verificate sulla base delle conoscenze disponibili (i geni A e B sono già funzionalmente annotati come facenti parte di un medesimo processo?), tenendo anche conto delle informazioni provenienti dalle altre reti di regolazione (la proteina espressa dal gene A e quella dal gene B sono coinvolte nello stesso pathway metabolico?). Le relazioni estratte devono poi essere validate mediante appositi esperimenti mirati, che spesso necessitano la creazione di cellule/organismi mutanti con geni soppressi o l'utilizzo di metodiche con la RNA-interference. I metodi di reverse engineering costituiscono comunque un importante primo passo verso la comprensione dei meccanismi di regolazione genica (Zupan B, 2003). In particolare, come precentemente sottolineato, una strada di notevole interesse vedrà l'integrazione di informazioni provenienti da DNA microarray con dati di espressione proteica (Segal E, 2001) e con informazioni sui processi metabolici coinvolti (Cavaliere D, 2002; Segal E, 2003b), con l'obiettivo di studiare i sistemi di regolazione con un approccio "a livello di sistema" (Kitano H, 2002). Nell'ambito della systems biology gli obiettivi sono quelli di modellizzare le reti di regolazione superando approcci riduzionismi e di essere in grado quindi di costruire dei simulatori realistici che permettano di fare delle analisi di tipo "what-if". Lo sviluppo di questo settore rappresenta una delle attuali frontiere della ricerca nel campo della bioinformatica, della biologia computazionale e dell'informatica biomedica.

Bibliografia

1. Beisvag V, Junge FK, Bergum H, Jolsum L, Lydersen S, Gunther CC, Ramampiaro H, Langaas M, Sandvik AK, Laegreid A. GeneTools-application for functional annotation and statistical hypothesis testing. *BMC Bioinformatics*. 2006 Oct 24; 7: 470.
2. Bellazzi R, Bicciato S, Cobelli C, Di Camillo B, Ferrazzi F, Magni P, Sacchi L, Toffolo G. Analisi di dati di DNA microarray: reti di regolazione. In *Fondamenti di elaborazione dei segnali biomedici e introduzione ai metodi avanzati*, Patron editore. 2004; 413-429.
3. Bernard A, Hartemink AJ. Informative structure priors: joint learning of dynamic regulatory networks from multiple types of data. *Pac Symp Biocomput*. 2005: 459-70.
4. Butte AJ, Tamayo P, Slonim D, Golub TR, Kohane IS. Discovering functional relationships between RNA expression and chemotherapeutic susceptibility using relevance networks. *Proc Natl Acad Sci USA*. 2000 Oct 24; 97(22).
5. Cavaliere D, De Filippo C, Grosu P, Biggeri A. Making sense of whole genome expression data. *Minerva Biotecnologica*. 2002; 14: 291-304.
6. Chee M, Yang R, Hubbell E, Berno A, Huang XC, Stern D, Winkler J, Lockhart DJ, Morris MS, Fodor SPA. Accessing genomic information with high-density arrays. *Science*. 1996; 274: 610-614.
7. Chen G, Gharib TG, Huang CC, Taylor JM, Misek DE, Kardias SL, Giordano

- TJ, Iannettoni MD, Orringer MB, Hanash SM, Beer DG. Discordant protein and mRNA expression in lung adenocarcinomas. *Mol Cell Proteomics*. 2002 Apr; 1(4): 304-13.
8. D'haeseleer P, Wen X, Fuhrman S, Somogyi R. Linear modeling of mRNA expression levels during CNS development and injury. *Pac Symp Biocomput*. 1999; 41-52.
 9. de Jong H. Modeling and simulation of genetic regulatory systems: a literature review. *J Comput Biol*. 2002; 9(1): 67-103.
 10. De Risi JL, Iyer VR, Brown PO. Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science*. 1997; 278: 680-686.
 11. Di Camillo B, Sanchez-Cabo F, Toffolo G, Nair SK, Trajanoski Z, Cobelli C. A quantization method based on threshold optimization for microarray short time series. *BMC Bioinformatics*. 2005 Dec 1; 6 Suppl 4: S11.
 12. Dudoit S, Fridlyand J. A prediction-based resampling method for estimating the number of clusters in a dataset. *Genome Biol*. 2002 Jun 25; 3(7): RESEARCH0036. Epub 2002 Jun 25.
 13. Eisen, MB, Spellman, P.T., Botstein, D. and Brown P.O. 1998. Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. USA*, 95: 14863-14868.
 14. Ferrazzi F, Magni P, Bellazzi R. Random walk models for Bayesian clustering of gene expression profiles. *Appl Bioinformatics*. 2005; 4(4): 263-76.
 15. Ferrazzi F, Sebastiani P, Kohane IS, Ramoni MF, Bellazzi R. Dynamic Bayesian Networks in Modelling Cellular Systems: a Critical Appraisal on Simulated Data. 19th IEEE Symposium on Computer-Based Medical Systems. 2006; 544-549.
 16. Friedman N, Linial M, Nachman I, Pe'er D. Using Bayesian networks to analyze expression data. *J Comput Biol*. 2000;7(3-4): 601-20.
 17. Hansel DE, Rahman A, Hidalgo M, Thuluvath PJ, Lillemoe KD, Shulick R, Ku JL, Park JG, Miyazaki K, Ashfaq R, Wistuba II, Varma R, Hawthorne L, Geradts J, Argani P, Maitra A. Identification of novel cellular targets in biliary tract cancers using global gene expression technology. *Am J Pathol*. 2003; 163(1): 217-29.
 18. Harrington CA, Rosenow C, Retief J. Monitoring gene expression using DNA microarrays. *Curr. Opin. Microbiol*. 2000; 3: 285-291.
 19. Hubbell E, Liu W-M, Mei R. Robust estimators for expression analysis. *Bioinformatics*. 2002; 18: 1585-1592.
 20. Hvidsten TR, Laegreid A, Komorowski J. Learning rule-based models of biological process from gene expression time profiles using Gene Ontology. *Bioinformatics*. 2003 Jun 12; 19(9): 1116-23.
 21. Irizarry RA, Hobbs B, Collin F., Beazer-Barclay Y.D., Antonellis K.J., Scherf U., and Speed T.P.. Exploration, normalization and summaries of high density oligonucleotide array probe level data. *Biostatistics*. 2003; 4(2): 249-264.
 22. Kitano H. Systems Biology: a brief overview. *Science*. 2002; 295: 1662-1664.

23. Lee I, Date SV, Adai AT, Marcotte EM. A probabilistic functional network of yeast genes. *Science*. 2004 Nov 26; 306(5701): 1555-8.
24. Li C, Wong WH. Model-based analysis of oligonucleotide arrays: expression index computation and outlier detection. *Proc Natl Acad Sci USA*. 2001 Jan 2; 98(1): 31-6.
25. Liang S, Fuhrman S, Somogyi R. Reveal, a general reverse engineering algorithm for inference of genetic network architectures. *Pac Symp Biocomput*. 1998; 18-29.
26. Lockhart DJ, Dong H, Byrne MC, Follettie MT, Gallo MV, Chee MS, Mittmann M, Wang C, Kobayashi M, Horton H, Brown EL. Expression monitoring by hybridization to highdensity oligonucleotide arrays. *Nat. Biotechnol*. 1996; 14(13): 1675-80.
27. MacKay VL, Li X, Flory MR, Turcott E, Law GL, Serikawa KA, Xu XL, Lee H, Goodlett DR, Aebersold R, Zhao LP, Morris DR. Gene expression analyzed by high-resolution state array analysis and quantitative proteomics: response of yeast to mating pheromone. *Mol Cell Proteomics*. 2004 May; 3(5): 478-89.
28. Nariai N, Tamada Y, Imoto S, Miyano S. Estimating gene regulatory networks and protein-protein interactions of *Saccharomyces cerevisiae* from multiple genome-wide data. *Bioinformatics* 2005;21 Suppl 2: ii206-ii212.
29. Ramoni MF, Sebastiani P, Kohane IS. Cluster analysis of gene expression dynamics. *Proc Natl Acad Sci USA*. 2002 Jul 9; 99(14): 9121-6.
30. Sacchi L, Bellazzi R, Larizza C, Magni P, Curk T, Petrovic U, Zupan B. TA-clustering: cluster analysis of gene expression profiles through Temporal Abstractions. *Int J Med Inform*. 2005 Aug; 74(7-8): 505-17.
31. Sacchi L, Bellazzi R, Porreca R, Larizza C, Magni P. Precedence Temporal Networks from Gene Expression Data. In *Proc. of the 18th IEEE Symposium on Computer-Based Medical Systems (CBMS 2005)*. 2005: 109-114.
32. Schena M, Shalon D, Davis RW, Brown, PO. Quantitative monitoring of gene expression pattern with a complementary DNA microarray. *Science*. 1996; 270:467-470.
33. Segal E, Taskar B, Gasch A, Friedman N, Koller D. Rich probabilistic models for gene expression. *Bioinformatics*. 2001; 17(1): S243-252.
34. Segal E, Shapira M, Regev A, Pe'er D, Botstein D, Koller D, Friedman N. Module networks: identifying regulatory modules and their condition-specific regulators from gene expression data. *Nat Genet*. 2003; 34(2): 166-76.
35. Segal E, Wang H, Koller D. Discovering molecular pathways from protein interaction and gene expression data. *Bioinformatics*. 2003; 19(1): i274-i272.
36. Shmulevich I, Dougherty ER, Kim S, Zhang W. Probabilistic Boolean Networks: a rule-based uncertainty model for gene regulatory networks. *Bioinformatics*. 2002 Feb; 18(2): 261-74.
37. Sreekumar R, Halvatsiotis P, Schimke JC, Nair KS Gene Expression Profile

- in Skeletal Muscle of Type 2 Diabetes and the Effect of Insulin Treatment. *Diabetes* 2002; 51(6): 1913-20.
38. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*. 2005 Oct 25; 102(43): 15545-50.
 39. Thomas PD, Campbell MJ, Kejariwal A, Mi H, Karlak B, Daverman R, Diemer K, Muruganujan A, Narechania A. PANTHER: a library of protein families and subfamilies indexed by function. *Genome Res*. 2003 Sep; 13(9): 2129-41.
 40. Tian Q, Stepaniants SB, Mao M, Weng L, Feetham MC, Doyle MJ, Yi EC, Dai H, Thorsson V, Eng J, Goodlett D, Berger JP, Gunter B, Linseley PS, Stoughton RB, Aebersold R, Collins SJ, Hanlon WA, Hood LE. Integrated genomic and proteomic analyses of gene expression in Mammalian cells. *Mol Cell Proteomics*. 2004 Oct; 3(10): 960-9.
 41. Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci U S A*. 2001 Apr 24; 98(9): 5116-21.
 42. Zupan B, Demsar J, Bratko I, Juvan P, Halter JA, Kuspa A, Shaulsky G. GenePath: a system for automated construction of genetic networks from mutant data. *Bioinformatics*. 2003; 19(3): 383-389.

Analisi di dati di spettrometria di massa finalizzata alla scoperta di biomarcatori

Paolo Magni

Dipartimento di Informatica e Sistemistica, Università di Pavia

Nell'ultimo decennio è cresciuto esponenzialmente l'interesse della ricerca biomedica verso la proteomica, vale a dire la scienza che studia l'insieme delle proteine espresse dal genoma di un organismo. I principali motivi di tale interesse possono essere così sintetizzati:

- L'analisi del proteoma, in particolare nell'uomo, ha attirato su di sé una considerevole attenzione dopo che la ricerca genomica ha evidenziato che il genoma è costituito da molti meno geni rispetto al numero di proteine. Nell'uomo i geni sono stimati essere circa 35.000, mentre il numero totale di proteine è valutato fra 250 e 500 mila. Soltanto una parte di esse è stata identificata e classificata. È pertanto ormai evidente che le sole informazioni genomiche sono insufficienti a spiegare gli eventi biologici, almeno negli organismi superiori, e a consentire l'individuazione di nuovi bersagli e terapie farmacologiche.
- È ormai chiaro che i profili di espressione del mRNA non si correlano sempre con il livello di espressione proteica. Infatti l'analisi del trascrittoma fornisce un'indicazione accurata di quali siano i geni attivi in una particolare cellula, ma dà indicazioni meno accurate circa le proteine che sono espresse. Ci sono diverse ragioni per questa mancanza di equivalenza tra trascrittoma e proteoma, tra cui le più importanti sono:
 - non tutti gli mRNA vengono effettivamente tradotti in proteine;
 - il contenuto proteico della cellula è determinato sia dalla velocità di sintesi di nuove proteine sia da quella di degradazione di quelle già esistenti.
- A seguito dei progressi delle tecnologie analitiche la ricerca biologica si è spostata dallo studio del singolo gene all'analisi dell'espressione del intero genoma. La disponibilità dei dati della sequenza genomica e le nuove tecniche di ionizzazione in spettrometria di massa hanno fortemente contribuito a rendere possibile questo spostamento delle metodologie di studio anche a livello del proteoma.

Gli obiettivi primari della proteomica sono la determinazione della sequenza aminoacidica delle proteine, la definizione della loro struttura tridimensionale, l'individuazione della loro funzione e l'indicazione del modo in cui interagiscono tra loro e con altre molecole.

Accanto a queste aree si sta sviluppando recentemente la *proteomica clinica* che si prefigge di sviluppare tecnologie per la cura e la diagnosi di malattie (soprattutto tumori) dei pazienti attraverso l'analisi e il confronto del proteoma di cellule sane e patologiche. L'obiettivo è individuare proteine, nel sangue, nell'urina, o nei tessuti patologici, che possano essere utilizzate come marker precoci oppure che possano predire la risposta alla terapia o la probabilità di recidiva dopo un trattamento. In particolare uno dei campi in cui la proteomica clinica potrebbe segnare una svolta è quello dell'individuazione di nuovi biomarcatori per la diagnosi precoce dei tumori. Marker di questo tipo sono già stati proposti, ma non sono ancora abbastanza efficaci. Tra questi possiamo ricordare il PSA, associato al tumore alla prostata (la cui utilità è stata recentemente messa in discussione), oppure il CA-125 per il tumore alle ovaie. La proteomica si prefigge di capire meglio quali sono le proteine che caratterizzano le cellule tumorali rispetto a quelle sane per poter scoprire così precocemente se vi sia un'evoluzione tumorale in corso.

Tra gli approcci più promettenti finalizzati alla scoperta di biomarcatori vi è l'analisi di profili proteomici di fluidi biologici ottenuti tramite spettrometria di massa.

La spettrometria in proteomica

La spettrometria di massa (MS) è lo strumento di analisi per determinare la massa delle molecole con la maggiore precisione possibile. Di solito è composta da tre parti principali:

- una sorgente di ionizzazione che vaporizza e carica elettricamente le molecole;
- un analizzatore di massa che in generale ordina gli ioni molecolari secondo il rapporto massa - carica applicando un campo elettromagnetico;
- il detettore che rileva la presenza dello ione corrente e trasmette il segnale amplificato al sistema di acquisizione.

L'output di un'analisi spettrometrica di massa è costituito da uno spettro di massa, dove i rapporti massa-carica (m/z) degli ioni identificati sono assegnati all'asse delle x, mentre l'intensità assoluta, corrispondente alla quantità di ioni giunti al detettore, è rappresentata sull'asse delle y. Spesso si utilizza l'intensità relativa ottenuta assegnando valore 100 al picco maggiore.

Le proprietà intrinseche degli analizzatori di massa permettono solo la misura del rapporto massacarica.

Ogni molecola può generare diversi picchi a seconda di quante cariche la abbiano ionizzata. Infatti il rapporto massa carica è solitamente determinato dalla seguente formula $m/z = (m_1 + n) / n$ con m_1 la massa della molecola e n il numero di cariche ionizzanti. Di conseguenza la massa reale delle molecole è indicata solo dalla specie con carica unitaria.

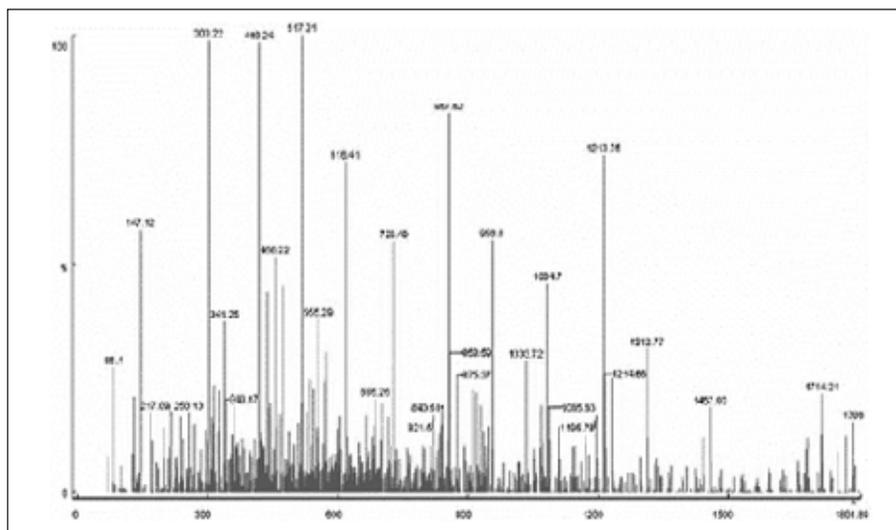


Fig. 1 - Esempio di uno spettro di una mistura di polipeptidi.

Di solito gli spettri di massa, soprattutto quelli rappresentanti misture di più macromolecole, sono molto complessi (Figura 1) e le loro analisi richiedono rilevanti capacità computazionali.

Esistono differenti sorgenti di ionizzazione ma tutte vaporizzano e caricano elettricamente il campione da analizzare. Infatti solo gli ioni carichi possono essere misurati dalla MS.

Le principali metodologie di ionizzazione sviluppate sono:

- ad impatto elettronico (EI);
- ionizzazione chimica (CI);
- a bombardamento con atomi veloci (FAB);
- ionizzazione elettrospray (ESI);
- ionizzazione per desorbimento laser assistito da matrice (MALDI).

La ESI è insieme alla tecnica MALDI il metodo di ionizzazione più utilizzato in proteomica. Infatti tale tecnica è adatta per analizzare molecole con massa tra 100 e 10⁶ Da (questo è all'incirca il range di variabilità di tutte le proteine conosciute). Le due caratteristiche essenziali dell'ESI sono che la ionizzazione si verifica a pressione atmosferica e che spesso deposita una carica multipla. Pertanto lo spettro ESI anche di una stessa molecola si presenta come una distribuzione di picchi corrispondenti alla distribuzione dei valori m/z dei diversi ioni multicarica.

Nella MALDI il campione è associato a matrici assorbenti energia. Prevede l'irradiazione della miscela campione-matrice con impulsi laser che provoca l'e-spulsione e la successiva ionizzazione del campione. A differenza del metodo ESI la singola molecola acquisisce nella maggior parte dei casi una sola carica. È tuttavia possibile, ma meno probabile, l'acquisto di due o tre cariche.

Lo spettrometro di massa in modalità MS, poiché è costituito da una sorgente di

ioni, da un solo analizzatore e da un rivelatore, di solito da solo informazioni sul rapporto massa-carica delle molecole d'interesse. Nella maggior parte dei casi queste informazioni non bastano ad identificare inequivocabilmente la molecola. Le informazioni strutturali possono essere acquisite tramite diverse tecniche coinvolgendo più analizzatori in sequenza tramite la cosiddetta MS tandem (o MS/MS). Le funzioni che vengono svolte da tali componenti, oltre a quella di analizzatore di massa, sono di selettore dello ione d'interesse e di camera di collisione nella quale le molecole vengono frammentate. L'ordine nel quale tali funzioni vengono svolte dipende dal metodo di scansione utilizzato.

Poiché anche semplici organismi esprimono numerose proteine alla volta ad ogni istante della loro esistenza, per ottenere spettri di massa interpretabili è necessario analizzare solo un numero limitato di proteine. Per questo sono stati messi a punto vari tipi di prefrazionamento; in particolare sono stati sviluppati numerosi metodi di separazione delle proteine che posso precedere la spettrometria di massa quali l'elettroforesi su gel (1-D, 2-D), la cromatografia liquida (es. HPLC), il ProteinChip. Il SELDI (Surface enhanced laser desorption and ionization) è il chip più utilizzato in proteomica.

SELDI è un metodo per mezzo del quale le proteine contenute in complesse misture vengono purificate, essendo catturate da superfici cromatografiche specifiche a loro volta immobilizzate su un supporto solido. Dopo vari lavaggi per rimuovere le proteine che non sono richieste e altre tipi di molecole, i polipeptidi che rimangono legati vengono mischiati con la matrice necessaria per una ionizzazione di tipo MALDI e quindi analizzate direttamente con uno spettrometro di massa. Gli spettrometri di massa ad oggi possono analizzare assieme una serie di array di alluminio (10 mm x 80 mm), ognuno composto da otto o sedici spot (all'interno del quale vi è il campione di un paziente) aventi un diametro di 2 mm con superfici capaci di catturare proteine chimicamente o biochimicamente. Il pregio principale di tale metodo è che l'intero processo dall'acquisizione del campione alla MS è veloce e facilmente automatizzabile, poiché la preparazione necessaria è molto semplice. Dato che è possibile utilizzare ogni spot per un paziente diverso, si possono ottenere enormi quantità di dati facilmente. Inoltre la quantità di materiale necessario per l'analisi è estremamente ridotta.

La principale difficoltà è rappresentata dallo sviluppo di superfici altamente specifiche che permettano di selezionare le proteine di interesse. Infatti se l'interazione tra superficie e proteine non è abbastanza specifica, è stata osservata una scarsa riproducibilità dell'esperimento.

Proteomica clinica e proteomica serica

La proteomica in chiave diagnostica può prevedere l'analisi di tessuti solidi oppure del sangue. Nel primo caso si parla di proteomica tissutale, nel secondo di proteomica serica.

La proteomica tissutale analizza i proteomi di singole cellule prelevate dai tessuti normali/patologici. Il problema in questo caso consiste nel disporre di strumenti in grado di analizzare singole cellule con certezza sane oppure malate. Ciò è possibile tramite un dissettore laser, uno strumento in grado di prendere il tessuto

ottenuto da una biopsia, poi prelevare e analizzare, una per una, le singole cellule. Attraverso questo tipo di analisi è possibile ricostruire con precisione le catene di eventi chimici in cui sono coinvolte le diverse proteine e quindi studiare i passaggi mal funzionanti responsabili della malattia.

Nel caso della proteomica serica si analizzano le proteine presenti nel siero. Oggi infatti la formazione della maggior parte delle patologie, come per esempio i tumori, è vista sempre di più legata all'alterazione del rapporto tra le cellule e l'ambiente circostante. Le cellule scambiano continuamente messaggi con i tessuti che hanno intorno rilasciando e ricevendo proteine. È ovvio quindi pensare che le cellule che diventano tumorali, rilascino proteine diverse, almeno in parte, da quelle normali. Queste proteine possono immettersi passivamente nel vaso sanguigno più vicino alla cellula e quindi entrare in circolazione.

Quindi al proteoma serico, che normalmente contiene le proteine rilasciate dalle cellule sanguigne e una grande abbondanza di carrier (albumina, immunoglobuline...), si aggiunge una piccola quantità di polipeptidi provenienti dai tessuti circostanti. Le proteine di grandi dimensioni possono entrare in circolo solo se i vasi diventano permeabili a causa di una malattia. Comunque la digestione effettuata dagli enzimi incontrati nel tragitto tra la cellula e il vaso sanguigno genera dei piccoli frammenti che possono entrare passivamente in circolo. Di conseguenza la regione del proteoma serico a basso peso molecolare (LMW), vale a dire all'incirca quella comprendente i polipeptidi sotto i 15.000 Dalton, risulta contenere sia piccole proteine intatte che frammenti di grosse proteine, ed entrambi possono contenere informazioni diagnostiche finora ignorate.

Inoltre solo ultimamente si è dimostrato che la dimensione del polipeptide circolante determina la velocità con cui viene filtrato a livello renale. Le molecole più piccole sono eliminate in poche ore, mentre le proteine più grosse restano in circolo più a lungo; per esempio l'albumina viene rimossa dopo circa 19 giorni. A causa di tale processo, un polipeptide di ridotte dimensioni può rimanere in circolo solo se si lega ad uno dei carrier circolanti. Di conseguenza è fondamentale, prima dell'analisi, non eliminare dal siero le proteine più abbondanti. La concentrazione di un possibile biomarcatore a basso peso molecolare è quindi strettamente dipendente dalla sua capacità di legarsi ai trasportatori sanguigni.

Il siero non è l'unico fluido biologico analizzato per fini diagnostici; si utilizzano anche plasma, fluido aspirato dal seno, liquido sinoviale e fluido cerebrospinale. Inoltre proprio per il fenomeno della clearance renale anche le urine contengono parte del LMW potenzialmente diagnostico; per questo motivo alcuni ricercatori effettuano analisi anche sulle urine.

Tecnologia SELDI-TOF

I polipeptidi a basso peso molecolare risultano essere al di sotto del limite misurabile tramite le tecniche convenzionali in proteomica, come l'elettroforesi su gel, poiché non possono essere separati in maniera efficiente. È per questo motivo che la ricerca si sta rivolgendo alla spettrometria di massa, la quale possiede performance ottimali nel LMW.

Sono stati sviluppati vari metodi basati sulla MS per valutare il proteoma a fini diagnostici, tuttavia la tecnologia SELDI-TOF risulta la più utilizzata. Tale approccio combina la selettività cromatografica dei chip proteici con la misura tramite MS TOF, fornendo il profilo proteomico di un qualsiasi campione biologico. Recenti studi, che hanno impiegato approcci basati su MS SELDI-TOF, hanno dimostrato che i profili proteomici del siero sono utili ai fini di una diagnosi precoce del cancro alle ovaie, alla prostata, alla cervice, al seno, ai polmoni, alla vescica, al colon e al pancreas. Oltre al cancro la tecnica SELDI-TOF è stata anche adoperata per studiare le proteine legate ad altre patologie come l'Alzheimer, l'artrite reumatoide e l'AIDS.

La strumentazione

La maggior parte degli studi che prevedono l'utilizzo di pattern proteomici per la diagnosi hanno impiegato la medesima piattaforma di analisi con differenti metodi di preparazione e usando diversi strumenti bioinformatici. La piattaforma analitica adoperata è il sistema PBS-II (ProteinChip® Biomarker System-II, Ciphergen Biosystems), un spettrometro di massa SELDI-TOF a bassa risoluzione.

Tuttavia uno dei fattori che ha limitato il successo dei classificatori sviluppati è la bassa risoluzione presentata da tale strumento. Quindi per produrre spettri con una risoluzione maggiore sono stati impiegati spettrometri di massa a configurazione ibrida QqTOF (QSTAR pulsar i, Applied Biosystems) accoppiati alla sorgente di ionizzazione SELDI.

Tale dispositivo è costituito da una sorgente di ionizzazione SELDI, da tre quadrupoli e da un analizzatore di tipo TOF a riflessione. Per accertarsi dell'aumento in termini di sensibilità e specificità della classificazione ottenuta tramite l'introduzione dello strumento ad alta risoluzione, è stato eseguito uno studio comparativo da Conrads et al. (2004).

Analisi dati

I lavori finora eseguiti sugli spettri di massa, finalizzati alla diagnosi tramite pattern proteomici e alla scoperta di nuovi biomarcatori, presentano analisi molto varie non esistendo una procedura standardizzata; nonostante tale fatto esse si sviluppano secondo un disegno abbastanza omogeneo che prevede lo svolgimento di una precisa successione di fasi: controllo di qualità, preprocessing, riduzione degli attributi, selezione degli attributi, classificazione, identificazione delle proteine. Si noti che non tutti gli studi eseguono tutti i passi.

Inoltre la maggior parte degli studi eseguiti nell'ambito della diagnostica tramite spettri di massa hanno portato alla scoperta di un pattern proteomico costituito da una serie di rapporti m/z . Poiché la diagnostica tramite pattern proteomici si basa sulla convinzione che l'identificazione delle proteine associate ai picchi selezionati sia utile ma non indispensabile, solo alcuni degli studi riportati si sono occupati anche di identificare a quali proteine questi rapporti massa-carica fossero associati (Rai et al., 2002; Ye et al., 2003; Carrette et al., 2003; Zhang et al., 2004). In questi casi sono stati impiegati i due approcci per l'identificazione delle

proteine: separazione su gel unita a PMF oppure sequenziamento tramite MS tandem. Tale fase quindi viene solitamente effettuata tramite ulteriori esperimenti di spettrometria di massa. È stato solo recentemente proposto un metodo valido, puramente bioinformatico, che permetta di associare i vari picchi spettrometrici a elenchi di possibili proteine conosciute (Barbarini et al. 2006).

Aspetti critici legati alla proteomica clinica

L'idea che la spettrometria di massa possa fornire dei pattern proteomici capaci di distinguere tra pazienti con una certa patologia e individui sani, ha generato, soprattutto nell'ambito della ricerca sul cancro, un enorme interesse. Infatti tale approccio potrebbe portare, in teoria, allo sviluppo di un test sanguigno semplice e rapido per qualsiasi tipo di tumore. Recentemente però alcuni ricercatori hanno messo in dubbio la validità di questo metodo (Diamandis et al., 2004; Baggerly et al., 2004; Grizzle et al., 2005). Si è sviluppato un vero e proprio dibattito tra i sostenitori e i detrattori di tale approccio. Vengono ora descritte le principali critiche poste.

- Non è possibile classificare i soggetti tramite picchi spettrometrici, senza sapere quali proteine essi rappresentino. Secondo i sostenitori invece l'identificazione delle proteine associate ai picchi è utile ma non necessaria a fini diagnostici.
- Gli esperimenti sono difficilmente riproducibili da un macchinario all'altro e nel corso dei giorni. Coloro che stanno lavorando nel campo della diagnostica tramite pattern proteomici garantiscono però che con l'avanzare delle tecnologie e il raggiungimento di una procedura standard si potrà in futuro raggiungere una perfetta riproducibilità.
- I pattern diagnostici scoperti dai vari gruppi di ricerca sono tutti differenti. In parte la causa di tale fatto è che evidentemente non tutti i gruppi hanno utilizzato lo stesso tipo di chip per le analisi. A chip differenti si legano proteine diverse; di conseguenza il profilo proteomico rilevato è strettamente dipendente dalla superficie cromatografica selezionata, e quindi rappresenta la parte di proteoma che il chip cattura.
- Il profilo proteomico di una persona non è costante, ma varia in ogni momento. Per questo i sostenitori di tale approccio affermano che tramite numerosi test di validazione è possibile trovare un pattern discriminante costituito da quelle parti del proteoma che non sono soggette a continue variazioni.
- I sistemi di misura utilizzati sono fortemente influenzati dalla presenza, per esempio nel siero, di proteine molto abbondanti che possono saturare i legami di uno spot prevenendo l'attacco delle proteine presenti in più bassa concentrazione.
- Gli spettri risultano falsati dal fenomeno della soppressione ionica, vale a dire la tendenza di alcuni ioni ad inibire la ionizzazione e quindi la misura di altri. I ricercatori interessati rispondono però che questo non è un problema poiché ciò che si vuole volta per volta analizzare, non è l'intero proteoma, ma una parte del proteoma in modo riproducibile.

- La contaminazione del campione e i movimenti del piatto su cui il fluido biologico è posto possono causare la generazione di artefatti; è d'altronde vero che tali artefatti, dato che riguardano in modo casuale le varie analisi, sicuramente non verranno scelti come porzioni valide per la classificazione.
- Infine vi è una critica fatta a tale approccio non più per quanto riguarda il metodo di analisi ma circa la validità da un punto di vista biologico dei risultati. Infatti i picchi che effettivamente risultano differenzialmente espressi tra i campioni di pazienti sani e quelli di individui malati non sono necessariamente associati a proteine che hanno un ruolo nel meccanismo che genera la patologia. Tali picchi spettrometrici, per esempio, possono essere associati a prodotti legati agli aspetti infiammatori della patologia, non potendo così fornire alcuna informazione diretta sui meccanismi che provocano la malattia.

Bibliografia

1. Adam, BL, Qu Y, Davis JW, Ward MD, Clements MA, Cazares LH, Semmes OJ, Schellhammer, PF, Yasui Y, Feng Z, Wright Jr, GL. Serum protein fingerprinting coupled with a patternmatching algorithm distinguishes prostate cancer from benign prostate hyperplasia and healthy men. *Cancer Res.* 2002; 62: 3609-3614.
2. Baggerly KA, Morris JS, Coombes R. Reproducibility of SELDI-TOF protein patterns in serum: comparing datasets from different experiments. *Bioinformatics* 2004; 20: 777-785.
3. Barbarini N, Magni P, Bellazzi R. A new approach for the analysis of mass spectrometry data for biomarker discovery. *Proceedings of the 2006 Annual Symposium of the American Medical Informatics Association. AMIA 2006;* 26-30.
4. Conrads TP, Fusaro VA, Ross S, Johann D, Rajapakse V, Hitt BA, Steinberg SM, Kohn EC, Fishman DA, Whitely G, Barrett JC, Liotta LA, Petricoin EF 3rd, Veenstra TD. Highresolution serum proteomic features for ovarian cancer detection. *Endocr Relat Cancer* 2004; 11(2): 163-78.
5. Diamandis EP. Mass spectrometry as a diagnostic and a cancer biomarker discovery tool: opportunities and potential limitations. *Mol Cell Proteomics.* 2004; 3(4): 367-78.
6. Grizzle WE, Semmes OJ, Bigbee W, Zhu L, Malik G, Oelschlager DK, Manne B, Manne U. The Need for Review and Understanding of SELDI/MALDI Mass. *Cancer Informatics:* 2005; 1 (1) 86-97.
7. Simone C, Fishman DA, Kohn EC, Liotta LA. Use of proteomic patterns in serum to identify ovarian cancer. *Lancet* 2002; 359, 572-577.
8. Omenn GS, States DJ, Adamski M, Blackwell TW, Menon R, Hermjakob H, Apweiler R, Haab BB, Simpson RJ, Eddes JS, Kapp EA, Moritz RL, Chan DW, Rai AJ, Admon A, Aebersold R, Eng J, Hancock WS, Hefta SA, Meyer H, Paik YK, Yoo JS, Ping P, Pounds J, Adkins J, Qian X, Wang R, Wasinger V, Wu CY, Zhao X, Zeng R, Archakov A, Tsugita A, Beer I, Pandey A,

- Pisano M, Andrews P, Tammen H, Speicher DW, Hanash SM, Overview of the HUPO Plasma Proteome Project: results from the pilot phase with 35 collaborating laboratories and multiple analytical groups, generating a core dataset of 3020 proteins and a publicly-available database. *Proteomics* 2005; 5(13): 3226-45.
9. Wright Jr GL. Computational protein biomarker prediction: a case study for prostate cancer. *BMC Bioinform.* 2004; 5, 26.



Tipografia Viscontea
Via Riviera, 39 - 27100 Pavia
Tel. 0382/526253 r.a. - Fax 0382/423120

