ISSN 0394 3291

Caleidoscopio



A cura di Anna Di Lonardo, Corrado Fagnani, Simonetta Pulciani

I Microarray



Direttore Responsabile Sergio Rassu



21-01-2009 9:28:59 221 | Microarray.indd 1



Società verificata e risultata conforme alla Norma UNI EN ISO 9001:2000

Il Sistema di Gestione per la Qualitàè applicato alla: Progettazione ed erogazione di corsi di formazione in campo sanitario. Settore EA: 37





Restless Architect of Human Possibilities s.a.s. (R.A.H.P. sas)

..... dalla Pedagogia all'Andragogia



Educazione Continua in Medicina





Restless Architect of Human Possibilities s.a.s. (R.A.H.P.) Sede Legale Via Pietro Nenni, 6 - 07100 Sassari
Tel/Fax 079 270464; - e-mail: rahp80@yahoo.it - http://rahp80.googlepages.com
Sede Regione Lombardia: Via Mauro Macchi, 73 - 20124 Milano
P. IVA 01991360908





A cura di Anna Di Lonardo, Corrado Fagnani, Simonetta Pulciani

Istituto Superiore di Sanità. Roma



Direttore Responsabile **Sergio Rassu**

221

... il futuro ba il enore antico MEDICAL SYSTEMS SPA

ISTRUZIONI PER GLI AUTORI



INFORMAZIONI GENERALI. Caleidoscopio pubblica lavori di carattere monografico a scopo didattico su temi di Medicina. La rivista segue i requisiti consigliati dall'International Committee of Medical Journal Editors. Gli Autori vengono invitati dal Direttore Responsabile. La rivista pubblica anche monografie libere, proposte direttamente dagli Autori, redatte secondo le regole della Collana.

TESTO. La monografia deve essere articolata in paragrafi snelli, di rapida consultazione, completi e chiari. I contenuti riportati devono essere stati sufficientemente confermati. E' opportuno evitare di riportare proprie opinioni dando un quadro limitato delle problematiche. La lunghezza del testo può variare dalle 60 alle 70 cartelle dattiloscritte ovvero 100-130.000 caratteri (spazi inclusi). Si invita a dattilografare su una sola facciata del foglio formato A4 con margini di almeno 25 mm. Usare dovunque doppi spazi e numerare consecutivamente. Ogni sezione dovrebbe iniziare con una nuova pagina.

Frontespizio. Deve riportare il nome e cognome dell'Autore(i) -non più di cinque- il titolo del volume, conciso ma informativo, la Clinica o Istituto cui dovrebbe essere attribuito il lavoro, l'indirizzo, il nome e l'indirizzo dell'Autore (compreso telefono, fax ed indirizzo di E-mail) responsabile della corrispondenza.

BIBLIOGRAFIA. Deve essere scritta su fogli a parte secondo ordine alfabetico seguendo le abbreviazioni per le Riviste dell'Index Medicus e lo stile illustrato negli esempi:

1) Björklund B., Björklund V.: Proliferation marker concept with TPS as a model. A preliminary report. J. Nucl. Med. Allied. Sci 1990 Oct-Dec, VOL: 34 (4 Suppl), P: 203.

2 Jeffcoate S.L. e Hutchinson J.S.M. (Eds): The Endocrine Hypothalamus. London. Academic Press, 1978. Le citazioni bibliografiche vanno individuate nel testo, nelle tabelle e nelle legende con numeri arabi tra parentesi.

TABELLE E FIGURE. Si consiglia una ricca documentazione iconografica (in bianco e nero eccetto casi particolare da concordare). Figure e tabelle devono essere numerate consecutivamente (secondo l'ordine di citazione nel testo) e separatamente; sul retro delle figure deve essere indicato l'orientamento, il nome dell'Autore ed il numero. Le figure realizzate professionalmente; è inaccettabile la riproduzione di caratteri scritti a mano libera. Lettere, numeri e simboli dovrebbero essere chiari ovunque e di dimensioni tali che, se ridotti, risultino ancora leggibili. Le fotografie devono essere stampe lucide, di buona qualità. Gli Autori sono responsabili di quanto riportato nel lavoro e dell'autorizzazione alla pubblicazione di figure o altro. Titoli e spiegazioni dettagliate appartengono alle legende, non alle figure stesse. Su fogli a parte devono essere riportate le legende per le figure e le tabelle.

UNITÀ DI MISURA. Per le unità di misura utilizzare il sistema metrico decimale o loro multipli e nei termini dell'International system of units (SI).

ABBREVIAZIONI. Utilizzare solo abbreviazioni standard. Il termine completo dovrebbe precedere nel testo la sua abbreviazione, a meno che non sia un'unità di misura standard.

Presentazione della monografia. Riporre il dattiloscritto, le fotografie, una copia del testo in formato doc oppure .rtf, ed copia di grafici e figure in formato Tiff con una risoluzione di almeno 240 dpi, archiviati su CD in buste separate.

Il dattiloscritto originale, le figure, le tabelle, il dischetto, posti in busta di carta pesante, devono essere spediti al Direttore Responsabile con lettera di accompagnamento. L'autore dovrebbe conservare una copia a proprio uso. Dopo la valutazione espressa dal Direttore Responsabile, la decisione sulla eventuale accettazione del lavoro sarà tempestivamente comunicata all'Autore. Il Direttore responsabile deciderà sul tempo della pubblicazione e conserverà il diritto usuale di modificare lo stile del contributo; più importanti modifiche verranno eventualmente fatte in accordo con l'Autore. I manoscritti e le fotografie se non pubblicati non si restituiscono.

L'Autore riceverà le bozze di stampa per la correzione e sarà Sua cura restituirle al Direttore Responsabile entro cinque giorni, dopo averne fatto fotocopia. Le spese di stampa, ristampa e distribuzione sono a totale carico della Medical Systems che provvederà a spedire all'Autore cinquanta copie della monografia. Inoltre l'Autore avrà l'opportunità di presentare la monografia nella propria città o in altra sede nel corso di una serata speciale.

L'Autore della monografia cede tutti i pieni ed esclusivi diritti sulla Sua opera, così come previsti dagli artt. 12 e segg. capo III sez. I L. 22/4/1941 N. 633, alla Rivista Caleidoscopio rinunciando agli stessi diritti d'autore (ed acconsentendone il trasferimento ex art. 132 L. 633/41).

Tutta la corrispondenza deve essere indirizzata al seguente indirizzo:

Restless Architect of Human Possibilities sas Via Pietro Nenni, 6 07100 Sassari







Caleidoscopio

Editoriale

'editoriale di questo volume sarà esclusivamente dedicato alla presentazione di tutti coloro che hanno collaborato alla Sua realizzazione. Si tratta di un volume del tutto particolare e speciale, sia per l'argomento trattato che per il numero degli Autori che, grazie all'opera di coordinamento dei tre che compaiono in copertina, hanno realizzato una monografia unica. Un ringraziamento mio particolare va poi alla dottoressa Simonetta Pulciani che con pazienza ha interpretato un ruolo chiave nella realizzazione di questa opera.

Il Dottor Corrado Fagnani, matematico, ha una posizione di ricercatore presso il Centro Nazionale di Epidemiologia, Sorveglianza e Promozione della Salute (CNESPS) dell'Istituto Superiore di Sanità (ISS) di Roma, e fa parte del gruppo di ricerca che gestisce il Registro Nazionale Gemelli. Il Dr. Fagnani si occupa di sviluppo ed applicazione di modelli statistici per la stima dell'ereditabilità di tratti complessi tramite lo studio dei gemelli. E' coinvolto in numerosi progetti di epidemiologia genetica, alcuni dei quali in collaborazione con i Registri Gemelli di altri Paesi europei.

La Dottoressa Simonetta Pulciani ha conseguito la laurea in Fisica presso l'Università degli Studi di Roma "La Sapienza", svolgendo una tesi sperimentale sotto la direzione del Prof. E. Calef. Dopo la laurea ha svolto attività di ricerca, come ospite, nel Laboratorio di Malattie Batteriche e Virali dell'Istituto Superiore nel gruppo del Prof. G.B. Rossi, studiando i meccanismi di trasformazione dei retrovirus murini. All'inizio degli anni '80 ha trascorso un periodo di studio presso il Laboratorio di Biologia Molecolare e Cellulare National Cancer Institute (N.C.I.) National Institute of Health (NIH) - Bethesda USA sotto la supervisione del Dr. Mariano Barbacid. In questo periodo, ha evidenziato mediante tecniche di trasfezione di DNA oncogeni della famiglia Ras in cellule tumorali umane. Come ricercatrice presso l'Istituto Superiore di Sanità (ISS) ha continuato le sue ricerche sul ruolo degli oncogeni nello sviluppo di neoplasie spontanee ed indotte da carcinogeni in modelli sperimentali animali. Si è interessata, anche, all'analisi delle interazioni HIV-ospite. In seguito, ha avuto l'incarico di trasferire



presso l'ISS le competenze teoriche e tecnico relative alla Tecnologia dei Microarray. Più recentemente si è interessata di epidemiologia genetica e le malattie rare. Attualmente svolge attività di carattere istituzionale, presso il Centro per la Ricerca e la Valutazione dei prodotti Immunobiologici, occupandosi di problematiche legate all'uso dei farmaci ricombinanti. E' autore di diverse pubblicazioni su riviste a carattere nazionale e internazionale.

La Dottoressa Anna Di Lonardo ha conseguito presso l'Università degli Studi di Roma "La Sapienza" la laurea in Scienze biologiche e la specializzazione in Patologia clinica. Nel corso della sua carriera si è occupata prevalentemente di Virologia, affrontando sia ricerche di base che traslazionali/cliniche. Ha lavorato per un lungo periodo come ricercatore nel laboratorio di Virologia dell'Istituto Nazionale Tumori "Regina Elena" di Roma, partecipando a studi, programmi di ricerca e progetti di rilevanza internazionale volti allo studio dell'infezione da Papillomavirus. In particolare si è occupata dei meccanismi di cancerogenesi indotti da papillomavirus umani, e nel corso dei suoi studi ha dimostrato, per la prima volta, il coinvolgimento (fino allora sconosciuto) di papillomavirus nei tumori della mammella. Nell'ambito delle collaborazioni internazionali ha lavorato presso il Beatson Institute for Cancer Research di Glasgow, portando a termine gli studi sulla risposta sierologica in pazienti affette da lesioni precancerose o francamente tumorali della cervice uterina associate ad infezione da HPV. Attualmente è ricercatore presso il Centro per la Ricerca e la Valutazione dei prodotti Immunobiologici dell'Istituto Superiore di Sanità. In questo Istituto la sua attività è stata rivolta allo studio delle infezioni da enterovirus, con particolare attenzione ai poliovirus e coxsackievirus. Inoltre ha incarichi di carattere istituzionale quali il controllo del vaccino antipoliomielitico, l'elaborazione di pareri per vaccini e la consulenza per la stesura di linee guida a livello Europeo per il controllo del vaccino antipoliomielitico. E' particolarmente interessata allo sviluppo delle nuove tecnologie, quali i DNA microarray e alle sue potenzialità in campo applicativo. E' autore di diverse pubblicazioni su riviste a carattere nazionale e internazionale.

Il Signor Alessandro Spurio si occupa di grafica/disegno impaginazione di documentazione tecnico-scientifica tradizionale (opuscoli, rapporti ecc.) presso l'Istituto Superiore di Sanità – Settore Attività Editoriali. Inoltre, collabora alla realizzazione di prodotti audio/video prodotti dall'Istituto. Sporadicamente si occupa di corsi di grafica ed impaginazione rivolti al personale di ruolo dell'Istituto Superiore di Sanità.

Il Dottor Alessandro Giuliani si occupa di applicazione di modelli matematici e statistici ai sistemi biologici, con particolare attenzione alla statistica multidimensionale e ai metodi non-lineari per lo studio dei segnali. Si è occupato di temi di ricerca molto diversi fra loro come la chimica farmaceu-



a ann fi ann air canad

tica, la fisiologia, l'ecologia, la biologia cellulare, la psicologia, la biochimica, la biologia strutturale e la fisica dei sistemi complessi. E' attualmente primo ricercatore presso l' Istituto Superiore di Sanità e membro del collegio dei docenti del dottorato di ricerca in Biofisica dell' Università 'La Sapienza' di Roma. E' autore di circa 170 pubblicazioni in riviste 'peer-reviewed' e del libro 'Simplicity: The Latent Order of Complexity' apparso nel 2007.

La Dottoressa Censi Federica ha conseguito la laurea in Scienze Biologiche presso l'Università di Roma "Roma Tre" ed in Bioinformatica (applicazioni biomediche e Farmaceutiche) presso l'Università degli Studi di Roma "La Sapienza". Inoltre è in possesso della Specializzazione in Genetica Medica presso la Prima Facoltà di Medicina e Chirurgia dell'Università degli studi di Roma "La Sapienza". Dal settembre 2002 svolge attività di ricerca presso il Centro Nazionale Malattie Rare dell'Istituto Superiore di Sanità, partecipando allo studio dei tumori rari e al controllo esterno di qualità dei test genetici.

La Signora Stella Lanni è laureanda in Biotecnologie Mediche, Molecolari e Cellulari presso l'Università degli Studi di Roma "La Sapienza", e sta svolgendo, presso il Centro Nazionale Malattie Rare dell'Istituto Superiore di Sanità, una tesi sull'analisi citogenetica dei tumori delle ghiandole salivari.

La Dottoressa Giovanna Floridia ha conseguito la Laurea in Scienze Biologiche presso l'Università di Catania, la Specializzazione in Citogenetica Umana presso l'Università di Pavia ed il titolo di Dottore di Ricerca in Patologia Umana presso l'Università di Pavia. Ha svolto attività di ricerca presso il CRC, Chromosome Molecular Biology Group, Dipartimento di Biochimica, Università di Oxford e attualmente è ricercatrice presso il Centro Nazionale Malattie Rare dell'Istituto Superiore di Sanità. Ha una ricca esperienza nel campo della diagnostica di malattie genetiche e nella ricerca della Genetica Medica Umana, essendosi occupata di analisi del genotipo/fenotipo in pazienti con anomalie cromosomiche costituzionali, meccanismi molecolari alla base di riarrangiamenti cromosomici ricorrenti nella popolazione. Si è occupata, anche, della caratterizzazione molecolare di un neocentromero umano del cromosoma Y e dell'analisi di eventi epigenetici coinvolti nell'attivazione e inattivazione centromerica. Presso il Centro Nazionale Malattie rare dell'Istituto Superiore di Sanità si occupa di controllo esterno di qualità dei test genetici e della patogenesi di tumori rari mediante strategie di citogenetica molecolare.

La Dottoressa Maria Letizia Polci si è laureata in Chimica all'Università "La Sapienza" di Roma con una tesi sperimentale in HPLC accoppiata alla



spettrometria di massa per la determinazione di antibatterici in matrici alimentari. Ha proseguito i suoi studi a Milano presso il DIBIT dell'Ospedale San Raffaele, ed ha acquisito esperienza in Biochimica e Spettrometria di massa di proteine. Ha perfezionato la sua formazione negli Stati Uniti alla George Mason University (Virginia), occupandosi di identificazione di marcatori proteici di tumori allo stadio precoce da siero umano.

La dottoressa Tamara Corinna Petrucci si è laureata in Chimica presso l'Università di Roma 'La Sapienza'. Ha svolto attività di ricerca presso il Medical Research Council di Londra e la Yale University, New Haven, Connecticut, USA. E' dirigente di ricerca presso il Dipartimento di Biologia Cellulare e Neuroscienze dell'Istituto Superiore di Sanità. Si è occupata di studi su: biochimica delle proteine del frumento, proteine del citoscheletro di eritrociti umani e di anfibi, proteine neuronali interagenti con l'actina, proteine sinaptiche. Attualmente, è interessata ai seguenti argomenti: ruolo di proteine citoscheletriche nell'organizzazione strutturale e funzionale di compartimenti sinaptici, complesso proteico di membrana delle proteine associate alla distrofina nel sistema nervoso ruolo funzionale ed implicazioni in patologie neurodegenerative, interazione proteina-proteina, meccanismi di trasmissione di segnali tra matrice extracellulare e citoscheletro. E' autore di numerose pubblicazioni di elevata qualità su riviste nazionali ed internazionali.

Il Dottor Carlo Petrini si è laureato in Scienze Biologiche a Torino e in Bioetica presso l'Università Cattolica del Sacro Cuore di Roma. Svolge la sua attività come primo ricercatore, incaricato dell'Unità di Bioetica presso la Presidenza dell'Istituto Superiore di Sanità. Nel settore della bioetica è membro di comitati e commissioni nazionali e responsabile di progetti di ricerca nazionali ed internazionali. E' autore di numerose pubblicazioni su riviste nazionali ed internazionali.

La Dottoressa Sabina Gainotti si è laureata in bioetica all'Università Cattolica di Roma. Collabora con l'Unità di bioetica della Presidenza e il reparto di Farmacoepidemiologia dell'Istituto Superiore di Sanità. I suoi interessi di ricerca sono l'etica della sanità pubblica e la tutela dei soggetti nella sperimentazione clinica. Insegna metodologia della ricerca sociale alla scuola di specializzazione in igiene e medicina preventiva e al corso di laurea specialistica per dirigenti dell'assistenza infermieristica presso l'Università Cattolica di Roma

Sergio Rassu





Dalla doppia elica ai microarray

Anna Di Lonardo¹, Simonetta Pulciani¹ e Corrado Fagnani²

¹Centro per la Ricerca e la Valutazione dei Prodotti ImmunoBiologici (CRIVIB), Istituto Superiore di Sanità

²Centro Nazionale di Epidemiologia, Sorveglianza e Promozione della Salute (CNESPS), Istituto Superiore di Sanità

Il ventesimo secolo ha rappresentato un momento di svolta nella scienza biologica. La scoperta della struttura del DNA, avvenuta nel 1953, ha definito finalmente la molecola della vita ed ha segnato l'inizio della biologia e genetica molecolare (Watson J.D. e Crick F.H.C., 1953, Watson J.D., 1990).

Questo traguardo non sarebbe stato raggiunto senza il contributo delle conoscenze e delle tecnologie proprie della fisica e della chimica e le osservazioni della genetica classica.

La scoperta della struttura del DNA ha rappresentato il punto di partenza per la conoscenza dei meccanismi di trasmissione e regolazione dell'informazione genetica. Fino alla metà degli anni '80, gli approcci metodologici permettevano l'analisi di singoli geni come consentito dalle tecnologie a disposizione, con un interesse prevalente verso il clonaggio e sequenziamento di singoli geni o di interi frammenti cromosomici (Miles J.S. e Wolf C.R., 1989, Southern E.M., 1975).

"Partire dalla doppia elica per approdare al genoma umano". Questa frase, pronunciata da James Watson agli inizi degli anni '90, presenta la sua personale opinione del Progetto Genoma Umano (PGU), ed è indicativa del percorso che era stato intrapreso nella comprensione del codice genetico.

Il PGU rappresenta lo sforzo congiunto di diversi gruppi di ricercatori a livello internazionale che ha permesso il sequenziamento del genoma umano, punto di partenza per definirne poi l'organizzazione e la regolazione.

Esso è stato avviato nella seconda metà degli anni '80, e la sua prima fase è stata completata nel 2001, a circa cinquant'anni di distanza dalla scoperta del DNA.

Il PGU è soltanto uno degli innumerevoli Progetti Genoma, poiché sono stati sequenziati anche i patrimoni genetici di altri organismi, quali, ad esempio, il ratto, il lievito, la drosofila, etc. (Borsani G. et al., 1998, Spieth J. e Lawson D., 2006).

I Progetti Genoma, per la rivoluzione apportata nella sperimentazione biologica, possono essere paragonati alla scoperta del DNA (McPherson J.D. et al 2001, Pennisi E., 2003). Essi sono ancora in fase di perfezionamento, in quanto rimane ancora da stabilire l'associazione tra sequenze geniche e loro funzio-



ni, e la determinazione dei vari polimorfismi tra gli individui (McPherson J.D. et al 2001, Borsani G. et al., 1998, Wheeler D.A., 2008). Queste ricerche sono state già avviate e contraddistinguono la cosiddetta era *Post-Genomica*. L'insieme di questi studi permetterà anche di evidenziare i fondamenti genetici ed il ruolo dei fattori ambientali implicati nello sviluppo di patologie semplici e complesse.

Caratteristica peculiare dell'era *Post-Genomica* è una sperimentazione ad ampio raggio che non si limita più all'analisi di geni singoli, ma coinvolge l'intero genoma, rendendo le tecniche e i protocolli della biologia molecolare classica inadeguate (Amaral P.P. et al., 2008, Chambers G. et al., 2000, Southern E.M., 1975, VanGuilder H.D. et al., 2008).

Di pari passo, nell'era Post-Genomica il vocabolario scientifico si è dovuto arricchire di nuovi termini. Per esempio sono stati coniati nuovi termini quali *Transcriptoma* e *Proteoma* per indicare rispettivamente l'insieme delle molecole di RNA e delle proteine presenti in una cellula o organismo in un preciso istante, e le scienze che studiano questi componenti cellulari e i processi di trascrizione e traduzione hanno preso il nome di *Transcriptomica* e *Proteomica* (Chambers G. et al., 2000, Kiechle F.L.et al., 2004).

Come già avvenuto nel passato, di pari passo con le conoscenze scientifiche, si sono sviluppate nuove tecnologie che hanno permesso continui ed inarrestabili avanzamenti nella biologia molecolare.

Le informazioni e i risultati acquisiti con i Progetti Genoma hanno, infatti, permesso di ideare e realizzare sofisticate tecnologie per l'analisi strutturale e funzionale dei geni clonati e sequenziati.

Senza dubbio, tra queste tecnologie, i *Microarray* rappresentano una di quelle che hanno avuto maggior riscontro e consenso nella comunità scientifica (Jares P., 2006, Seidel M. e Niessner R., 2008, Trevino V. et al., 2007, Wiltgen M. e Tilz G.P., 2007).

Tale metodica deve il proprio nome ai micro-schieramenti di sonde realizzati su un vetrino.

Nel caso in cui le sonde siano costituite da sequenze di DNA, ci troviamo di fronte alla tecnica denominata *DNA-Microarray* (Schena M., Seidel M.).

I microscopici allineamenti di sequenze di DNA, immobilizzati sul vetrino, fungono da bersaglio d'ibridazione per gli acidi nucleici e permettono l'analisi simultanea di migliaia di geni in un singolo esperimento. Il principio base dei *DNA-Microarray* deriva dai protocolli di *Southern* e *Northen blot* elaborati per l'analisi di sequenze note di DNA o di RNA, mediante ibridazione con sonde radioattive o chemioluminescenti (Southern E.M., 1975).

Le sonde possono essere costituite da frammenti di cDNA ottenuti da tessuti e sottoposti ad amplificazione genica (PCR) oppure oligonucleotidi. Tutti i tipi di sonde formate da cDNA sono sintetizzate *in vitro* e quindi depositate con sistemi automatici su vetrini trattati chimicamente. Nel caso degli oli-



gonucleotidi la sintesi può essere anche realizzata direttamente sul vetrino, mediante reazioni fotochimiche o chimiche (McGall G. et al. 1996, Okamoto T. et al., 2000, Pease A.C. et al., 1994). I vetrini, preparati con reazioni fotochimiche, vengono chiamati *GeneChips*; essi sono stati sviluppati e vengono prodotti dalla società *Affymetrix* (Pease A.C., 1994).

La tecnologia dei *Microarray* deve il suo successo anche ai traguardi raggiunti dalla bioinformatica e al supporto che essa le fornisce. Infatti, alla fase sperimentale deve far seguito l'analisi dei dati che dà delle indicazioni sul significato biologico dei risultati ottenuti (Giuliani A., 2008, Kuehn H. et al., 2008, Saal L.H. et al., 2002, Zhu M.M., 2008).

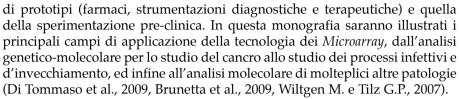
Dalla loro nascita, i *DNA-Microarray* hanno subito numerose evoluzioni sia sotto il profilo tecnologico che applicativo, consentendo lo studio dell'organizzazione strutturale e funzionale del genoma umano (Karsten S.L. et al., 2002, Hewett R. e Kijsanayothin P., 2008).

L'applicazione più comune dei *DNA-Microarray* è senza dubbio l'analisi dell'espressione genica, che ha permesso di ottenere informazioni sui meccanismi di regolazione genica e su diverse vie metaboliche. Ciò è stato reso possibile dal confronto dei profili di espressione relativi a differenti condizioni di crescita (ad esempio specifici elementi nutritivi, tossici, farmaci, ecc) (Gaudi S. et al., 2009, Gulletta E. et al., 2008).

I *DNA-Microarray* sono attualmente utilizzati anche per la caratterizzazione di amplificazioni e delezioni geniche, per lo studio dei siti di legame dei fattori di trascrizione e per l'analisi di mutazioni epigenetiche (Bonetta et al., 2009, Floridia et al., 2008, Polci et al., 2008). Inoltre, grazie ai *Gene-chips* è possibile analizzare l'esistenza di polimorfismi ed eventuali mutazioni puntiformi in sequenze già presenti in banche dati (McGall et al., 1996, Müller A. et al., 2008).

In virtù della sua versatilità e delle sue peculiari caratteristiche, la tecnologia dei *Microarray* si è rilevata uno strumento di straordinaria potenzialità negli studi di medicina traslazionale mirati al trasferimento alla clinica delle conoscenze scientifiche e metodologiche acquisite dalla ricerca biomedica (Hutchison C.A., 2007). Il concetto di scienza traslazionale ha fatto la sua apparizione agli inizi degli anni '90, quando le ricerche nel campo biologico-molecolare erano prevalentemente rivolte alla scoperta di nuovi farmaci nella terapia dei tumori. Negli anni successivi tale concetto è stato esteso ad altre aree della medicina: psichiatria, cardiologia, neurologia, ecc., e, affinché fosse ben compreso, sono state coniate diverse definizioni, di cui la più esplicativa è senza dubbio "from the bench to the bedside". Secondo la definizione più recente fornita dall'Istituto statunitense per il controllo dei farmaci e degli alimenti (FDA), la scienza traslazionale rappresenta l'anello di congiunzione tra la ricerca di base e lo sviluppo di nuovi protocolli clinici. Al suo interno comprende le fasi della scoperta e messa a punto





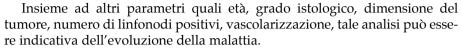
I DNA-Microarray contribuiscono alla diagnosi delle patologie tumorali, attraverso l'analisi molecolare dell'espressione genica e/o di alterazioni strutturali cromosomiche, quali delezioni o amplificazioni geniche; inoltre forniscono valutazioni sullo stadio del tumore, sulla prognosi e sull'efficacia delle terapie (Aricò E. et al., 2009, Bagg A., 2005, Cheang M.C. et al., 2008).

Le ricerche effettuate da Golub e i suoi collaboratori sui tumori del sistema emopoietico rappresentano un esempio emblematico delle applicazioni dei *DNA-Microarray* in campo oncologico. Questi ricercatori, analizzando gli RNA isolati da cellule del midollo osseo di pazienti affetti da leucemia mieloide acuta (AML) e leucemia linfoide acuta (ALL), con *DNA-Microarray* di circa 6800 geni umani, hanno evidenziato una espressione differenziale di circa 50 geni tra i due tipi di tumori. In maniera analoga, i *DNA-Microarray* sono stati utilizzati come valido strumento per suddividere i linfomi in sottotipi, e verificare la loro suscettibilità alle terapie (Hewett R.e Kijsanayothin P., 2008, Hoefnagel J.J. et al., 2005, Ma C.e Staudt L.M., 2001).

La tecnica dei *DNA-Microarray* ha fornito contributi notevoli anche alla classificazione molecolare di melanomi e tumori della ghiandola mammaria. I carcinomi mammari sono altamente eterogenei per le loro caratteristiche morfologiche e genetiche. I metodi convenzionali per la classificazione in sottogruppi solo parzialmente riescono a sottolineare le differenze cliniche della malattia. Infatti, tumori della ghiandola mammaria, simili all'esame istologico, possono presentare sostanziali differenze genetiche, tali da consentire, attraverso la tecnologia dei *DNA-Microarray*, la formulazione di una nuova tassonomia su basi molecolari (Dalgin G.S. et al., 2007, van der Vegt B. et al., 2008).

La più importante suddivisione dei tumori mammari riguarda la presenza di recettori per gli estrogeni (ER) e mutazioni nei geni BCRA1 e BCRA2. Grazie ai *DNA-Microarray* è stato possibile ottenere, attraverso i profili di espressione, un'ulteriore classificazione dei gruppi ER positivi e negativi in sottogruppi (Kristensen V.N. et al., 2007, Lønning P.E. et al., 2008). Questi sottogruppi sono caratterizzati dall'espressione di un particolare set di geni, e ciò si è rivelato importante non solo ai fini della diagnosi, ma anche per la prognosi e la risposta alla terapia. Infatti, è stato osservato in diversi studi che la sopravvivenza a 10 anni può essere stimata attraverso l'analisi dell'espressione di 70 geni (Bueno-de-Mesquita J.M. et al., 2008, van de Vijver M.J. et al., 1999, van 't Veer L.J. et al. 2002, Weigelt B. et al. 2005).





Ragguardevole è il lavoro di van't Veer e dei suoi collaboratori che, attraverso l'analisi del profilo di espressione di 117 tumori mammari, hanno identificato una correlazione tra profilo di espressione e prognosi della malattia. Ciò è determinante nell'individuare le terapie adiuvanti a cui sarà necessario sottoporre i pazienti. L'eterogeneità dei tumori mammari comporta che tumori allo stesso stadio possano non solo evolvere diversamente ma anche rispondere diversamente alle terapie (Cheang M.C. et al., 2008, Dalgin G.S. et al., 2007, Rody A. et al., 2008).

Indagini molecolari sulla natura dei processi neoplastici vengono effettuate anche attraverso l'analisi delle delezioni e amplificazioni geniche con la tecnica definita *Comparative Genomic Hybridization Microarray* (*CGH-Microarray*). In questo numero, un capitolo verrà dedicato esclusivamente ad illustrare tale tecnica e le sue applicazioni (Floridia G. et al., 2008).

La tecnologia dei *DNA-Microarray* è utilizzata anche nel campo delle patologie infettive per analizzare i profili d'espressione sia dell'ospite che del patogeno (Maynard J.A. et al., 2007, Mikhailovich V. et al., 2008, Robertson B.H. e Nicholson J.K., 2005).

I risvolti clinici di tale approccio sono considerevoli, poiché possono portare a definire le basi genetiche della patogenicità e sviluppare nuove strategie preventive e terapeutiche.

Campioni sia clinici che ambientali possono essere sottoposti ad analisi per individuare la presenza di sequenze batteriche e virali conservate, fattori specifici di virulenza, ceppi mutati per la resistenza ad antimicrobici e ceppi emergenti (Mikhailovich V. et al., 2008).

Significativo è l'esempio dei Virochips, cioè *Microarray* costituiti da un reticolo di 22000 nucleotidi, rappresentanti circa 277000 sequenze di genomi virali presenti nel database GeneBank. Utilizzando tali *array* è stato possibile identificare una vasta gamma di virus da RNA estratti da colture cellulari infette o da campioni umani. Caso emblematico dell'utilizzo dei Virochips è stata l'identificazione e caratterizzazione di una sequenza di 1000 basi di un Corona Virus (SARS-CoV), oggi riconosciuto come l'agente eziologico della Sindrome Respiratoria Acuta Severa (Wang D. et al., 2008).

In campo virologico, altre applicazioni dei Virochips sono la diagnosi di infezioni del sistema nervoso centrale e l'apparato respiratorio. In batteriologia, i Virochips rappresentano una risorsa per l'identificazione di ceppi batterici e per l'individuazione di geni e/o mutazioni responsabili della resistenza agli antibiotici. Inoltre, alcuni di essi, come ad esempio i Flu-Chips, vengono impiegati per la sorveglianza ed i programmi di epidemiologia molecolare dell'influenza (Townsend M.B. et al., 2006). Il monitoraggio di mutazioni genetiche a maggiore frequenza può fornire, così, utili informazioni per lo



sviluppo di nuovi vaccini. Come approfondimento dell'applicazione dei *Microarray* nelle patologie infettive, uno dei capitoli di questa monografia sarà dedicato allo studio del virus dell'epatite C (Ciccaglione A.R., 2009).

Nella ricerca immunologica così come in quella oncologica viene impiegato il cosiddetto *Microarray* "Lymphochip", costituito da un set di cDNA di geni rappresentativi di geni con funzioni rilevanti nel sistema immunitario (Ma C.e Staudt L.M., 2001).

Questo tipo di *Microarray* può essere di aiuto nello studio dei fenomeni di de-regolazione delle malattie autoimmuni, delle risposte infiammatorie e dei tumori.

Ad esempio, è possibile caratterizzare la risposta dei monociti/macrofagi ad infezioni batteriche gram negative, ed è stato osservato che, in risposta all'infezione, circa duecento geni vengono attivati e circa cento vengono repressi. I primi includono citochine (per esempio IL-1 α , IL-1 β , IL-6, IL-10, TNF α), chemochine (IL-8, MIP-1 α , MIP-1 β), recettori di superficie e ligandi (per esempio CD40 e il suo ligando) (Boldrick J.C. et al., 2002).

Maas e collaboratori hanno studiato il profilo dell'espressione differenziale di cellule mononucleate del sangue periferico di soggetti sani e sottoposti a vaccinazione antiinfluenzale.

Questo profilo è stato confrontato con i profili di espressione di pazienti affetti da malattie autoimmuni (lupus eritematoso, diabete di tipo 1, artrite reumatoide, sclerosi multipla).

In questo modo è stato possibile riconoscere due differenti gruppi di geni non coinvolti nella risposta immunitaria, ma implicati nella apoptosi, nella progressione del ciclo cellulare e differenziazione cellulare (Maas K. et al., 2002).

Una delle più recenti applicazioni dei *Microarray* interessa l'analisi molecolare delle malattie rare, cioè quelle patologie che, per definizione della Comunità Europea, colpiscono meno di 1 persona su 2000. Bisogna comunque precisare che, in contraddizione con questa definizione, il numero degli individui colpiti non è esiguo, dal momento che esistono tra 5000 e 8000 differenti malattie rare; da ciò ne deriva che tra il 6% e 8% della popolazione mondiale è affetto da queste patologie (Salvatore M. et al., 2009). L'80% delle malattie rare sono di origine genetica ed interessano tra il 3% e il 4% delle nascite.

Altre malattie rare sono il risultato di infezioni e allergie o dovute a cause degenerative e proliferative. Lo studio delle malattie rare risulta difficile per vari motivi: innanzitutto il ridotto numero di pazienti da seguire e la scarsità di conoscenze a riguardo che conduce ad una diagnosi tardiva, inaccurata e poco esaustiva.

L'applicazione della tecnologia dei *Microarray* alla malattie rare è solo agli inizi, e per le comprensibili difficoltà può definirsi anch'essa "rara" (Salvatore M. et al., 2009).



Ad oggi le malattie rare studiate con la tecnologia dei *Microarray* sono essenzialmente la sindrome di Bardet-Biedl, l'ipertensione polmonare idiopatica e alcuni tumori rari cui sarà dedicato un capitolo successivo (Salvatore M. et al., 2009). Infine, i *Microarray* sono stati utilizzati per studiare anche modelli murini della sindrome di Rett e di Nieman-Pick.

Qui descriviamo un esempio di applicazione dei microarray alle ipertensioni polmonari.

Le ipertensioni polmonari sono caratterizzate da una diffusa ostruzione dei vasi polmonari e da una rapida progressione della sintomatologia. La forma idiopatica è molto rara, e si può stimare che ogni anno ne vengano colpite una o due persone per milione di abitanti. Più frequenti sono le situazioni in cui l'ipertensione polmonare è associata ad altre patologie, come malattie del connettivo del fegato, cardiopatie congenite e i tromboemboli polmonari recidivanti.

Bull e collaboratori hanno condotto studi sull'espressione differenziale in prelievi bioptici di pazienti affetti da ipertensione polmonare, includendo anche la forma idiopatica (Bull T.M. et al., 2007.

I ricercatori hanno analizzato i profili di espressione di cellule mononucleate del sangue periferico di 15 pazienti con ipertensione arteriosa polmonare e di 6 soggetti sani per evidenziare le differenze di espressione che potrebbero essere correlate alla patologia.

Sono stati così identificati un gruppo di 106 geni la cui espressione è modulata nei pazienti con ipertensione arteriosa polmonare rispetto agli individui sani: ad esempio, il gene codificante per la adrenomedullina ed il gene per il fattore di crescita-1 per le cellule endoteliali (ECGF-1)(Bull T.M. et al., 2007).

Questi esperimenti indicano uno dei possibili percorsi che potrebbero in futuro permettere di discriminare gli individui sani da quelli affetti da ipertensione polmonare e facilitare la diagnosi, lo screening, e la patogenesi di questa malattia. Lo scopo finale dei ricercatori sarà anche quello di definire le differenze nell'espressione genica tra le diverse forme di ipertensione polmonare.

Infine, non si può tralasciare di menzionare l'applicazione della tecnica dei *Microarray* allo studio dei gemelli, che permette di ampliare le potenzialità sperimentali di questo modello. I gemelli hanno da sempre rappresentato un modello ideale per stabilire le influenze genetiche ed ambientali nella predisposizione alle malattie (Fagnani C. et al., 2009). La tecnologia dei *Microarray* potrà offrire l'opportunità di indagare, tramite i gemelli, l'eziologia di patologie complesse, risultanti dall'interazione di due o più geni (Pulciani S. et al., 2006).

Qui di seguito, forniamo due esempi di applicazione della tecnica dei *Microarray* allo studio di gemelli monozigoti (MZ) discordanti per patologia,



e per una trattazione più completa dell'argomento rimandiamo al capitolo dedicato (Fagnani C. et al., 2009).

Brevemente, utilizzando piattaforme di *Microarray* (tipo HG-U95av2) della ditta *Affymetrix* che contengono più di 12000 geni umani, è stato studiato il profilo di espressione di plasmacellule CD138+ e CD138- isolate dal midollo osseo di una coppia di gemelli MZ discordanti per mieloma multiplo (MM) (Munshi N.C. et al., 2004). Nelle cellule CD138+ del gemello affetto da MM, si è osservata una sovraespressione di duecentonovantasei geni, tra cui geni anti-apoptotici (FLIP, dad-1, mcl-1), oncogeni e fattori trascrizionali (Jun D, Xbp-1, FGFR-3) e geni collegati con l'attività ribosomale.

Inoltre, centotre geni, tra cui alcuni con attività pro-apoptotica e di riparo e ricombinazione del DNA (RAD51), sono risultati sottoespressi. La modulazione genica osservata concorda con le alterazioni che hanno luogo in una cellula trasformata, confermando l'adeguatezza dei *Microarray* applicati allo studio dei gemelli (Munshi N.C. et al., 2004)

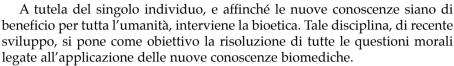
Un altro esempio di applicazione della tecnica dei *Microarray* allo studio dei gemelli è la malattia bipolare, patologia psichiatrica di cui non si conoscono ancora le cause.

Dall'analisi dell'espressione genica differenziale di cellule linfoblastoidi provenienti da una coppia di gemelli sani e due coppie di gemelli MZ discordanti per disturbi bipolari è stata rilevata una ridotta espressione di alcuni geni collegati al funzionamento del reticolo endoplasmatico, tra cui i geni XBP1 e HSPA5 (Kakiuchi C. et al., 2003). Bisogna sottolineare che questo tipo di approccio per l'analisi molecolare dei gemelli discordanti è ancora agli inizi e dovrà essere in futuro perfezionato per determinare le alterazioni genetiche, siano esse strutturali o epigenetiche, correlate ai processi di trascrizione. Inoltre, poiché il fenotipo è determinato dalle proteine biologicamente attive, lo studio sarà completato soltanto quando verranno analizzati anche i processi di traduzione e maturazione proteica. Infatti, gli studi sull'espressione genica danno un'indicazione delle proteine potenzialmente espresse, ma non danno informazioni su quali giungeranno a maturazione, contribuendo così a determinare il fenotipo, patologico o no. I Microarray per le proteine saranno in grado di dirimere questo interrogativo. I Microarray di proteine ed il loro impiego nella ricerca clinica sarà oggetto di un capitolo successivo (Polci M. L. e Petrucci T. C., 2008).

La tecnologia dei *Microarray*, come tutte le tecniche che fanno uso di informazioni genetiche, pongono importanti interrogativi riguardo al loro uso e divulgazione (Gainotti S.e Petrini C., 2008, Collins F.S. et al., 2003).

L'utilizzo della tecnica dei *Microarray* permette di raccogliere informazioni di carattere genetico-molecolare, e potenzialmente indicare l'eventuale predisposizione allo sviluppo di una patologia e/o alla sua evoluzione.





La bioetica si avvale del contributo di diverse figure professionali, quali biologi molecolari, medici, avvocati, teologi, psicanalisti, sociologi, ecc, che, operando in sinergia, cercano di indicare il miglior uso delle emergenti acquisizioni scientifiche nel rispetto della dignità dell'essere umano. A sottolineare l'importanza della bioetica, la conclusione di questa monografia verterà proprio sugli aspetti etici legati all'utilizzo dei test genetici (Gainotti S.e Petrini C., 2008, Lunshof J.E. et al., 2008).

In conclusione, sono stati brevemente presentati alcuni degli argomenti su cui verterà questa monografia, e per un approfondimento nello specifico si rimanda il lettore ai capitoli successivi.

E' importante precisare che, data la complessità e la vastità delle informazioni sui *Microarray*, la monografia non può essere esaustiva, ma si limita a fornire un quadro indicativo delle potenzialità di questa moderna tecnologia in ambito biomedico.

Bibliografia

- Amaral P.P., Dinger M.E., Mercer T.R., Mattick J.S.: The eukaryotic genome as an RNA machine. Science 2008 Mar, VOL:319(5871), P:1787.
- Aricò E., Parlato S., Castiello L., Canini I., Gabriele L.:La Tecnologia dei Microarray nella Ricerca Oncologica Biomedica e Clinica. questo volume
- Bagg A.: Molecular diagnosis in lymphoma. Curr Hematol Rep 2005 Jul, VOL:4(4),P:313.
- Boldrick J.C., Alizadeh A.A., Diehn M., Dudoit S., Liu C.L., Belcher C.E., Botstein D., Staudt L.M., Brown P.O., Relman D.A.: Stereotyped and specific gene expression programs in human innate immune responses to bacteria. Proc Natl Acad Sci U S A. 2002 Jan, VOL:99(2), P:972.
- Bonetta L.: Epigenomics: Detailed analysis. Nature 2008 Aug, VOL:454(7205), P:795.
- Borsani G., Ballabio A., Banfi S.: A practical guide to orient yourself in the labyrinth of genome databases. Hum Mol Genet 1998, VOL:7(10), P:1641.
- Brunetto B., Pini C., Tinghino R., Iacovacci P.: Impiego degli allergeni ricombinanti nello sviluppo di sistemi microarray per la diagnosi delle patologie allergiche. Caleidoscopio in preparazione 2009





- Bueno-de-Mesquita J.M., Linn S.C., Keijzer R., Wesseling J., Nuyten D.S., van Krimpen C., Meijers C., de Graaf P.W., Bos M.M., Hart A.A., Rutgers E.J., Peterse J.L., Halfwerk H., de Groot R., Pronk A., Floore A.N., Glas A.M., Van't Veer L.J., van de Vijver M.J.: Validation of 70-gene prognosis signature in node-negative breast cancer. Breast Cancer Res Treat. 2008 Sep, VOL: 26. [Epub ahead of print]
- Bull T.M., Coldren C.D., Geraci M.W., Voelkel N.F.: Gene expression profiling in pulmonary hypertension. Proc Am Thorac Soc 2007 Jan, VOL:4(1), P:117.
- Chambers G., Lawrie L., Cash P., Murray G.I.: Proteomics: a new approach to the study of disease. J Pathol 2000 Nov, VOL:192(3), P:280.
- Cheang M.C., van de Rijn M., Nielsen T.O.: Gene expression profiling of breast cancer. Annu Rev Pathol 2008, VOL:3, P:67.
- Ciccaglione A.R., Marcantonio C., Tritarelli E., Tataseo P., Ferraris A., Bruni R., Dallapiccola B., Gerosolimo G., Costantino A., Geraci L., Rapicetta M.: Microarray e Virus dell' Epatite C. Caleidoscopio in preparazione 2009
- Collins F.S., Morgan M., Patrinos A.: The Human Genome Project: lessons from large-scale biology. Science 2003 Apr, VOL:300(5617), P:286.
- Dalgin G.S., Alexe G., Scanfeld D., Tamayo P., Mesirov J.P., Ganesan S., DeLisi C., Bhanot G.: Portraits of breast cancer progression.BMC Bioinformatics 2007 Aug, VOL:8, P:291.
- Di Tommaso C., Magrì M., Capanna A., Tataseo P., Gerosolimo G., D'Alesio Y., Di Fonso A., Rispoli F., Bascietto F.: Microarray e Celiachia. questo volume
- Fagnani C., Di Lonardo A., Pulciani S.: I Microarray come tecnologia emergente per lo studio dei gemelli. Caleidoscopio in preparazione 2009
- Floridia G., Censi F., Lanni S.:CGH array ed applicazioni. Questo volume Gainotti S., Petrini C.: Problematiche Etiche collegate Ai Test Genetici. Questo volume
- Gaudi S., Macciardi F., Carlini F., Arcieri R., Vella S.: Farmacogenomica: dal progetto genoma umano alla terapia personalizzata. Caleidoscopio in preparazione 2009
- Giuliani A.: Analisi dei dati negli esperimenti di microarray Questo volume Golub T.R., Slonim D.K., Tamayo P., Huard C., Gaasenbeek M., Mesirov J.P., Coller H., Loh M.L. Downing J.R., Caligiuri M.A., Bloomfield C.D., Lander E.S.: Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science 1999 Oct, VOL:286(5439),P:531.
- Gulletta E., P. Foti D., Galliera E., Corsi M. M. Citochine e Chemochine. Caleidoscopio, Giugno 2008
- Hewett R., Kijsanayothin P.: Tumor classification ranking from microarray data. BMC Genomics 2008 Sep, VOL:9 (Supp 1 2), P:S21.





- Hoefnagel J.J., Dijkman R., Basso K., Jansen P.M., Hallermann C., Willemze R., Tensen C.P., Vermeer M.H.: Distinct types of primary cutaneous large B-cell lymphoma identified by gene expression profiling. Blood 2005 May, VOL:105(9), P:3671.
- Hutchison C.A.:3rd. DNA sequencing: bench to bedside and beyond.. Nucleic Acids Res. 2007, VOL:35(18), P:6227.
- Jares P.: DNA microarray applications in functional genomics. Ultrastruct Pathol 2006 May-Jun, VOL:30(3), P:209.
- Kakiuchi C., Iwamoto K., Ishiwata M., Bundo M., Kasahara T., Kusumi I., Tsujita T., Okazaki Y., Nanko S., Kunugi H., Sasaki T., Kato T.: Impaired feedback regulation of XBP1 as a genetic risk factor for bipolar disorder. Nat Genet 2003 Oct, VOL:35(2), P:171.
- Karsten S.L., Geshwind D.H.: Gene expression analysis using cDNA microarrays. Curr Protoc Neurosci 2002 Nov, Chapter 4:Unit 4.28.
- Kiechle F.L., Zhang X., Holland-Staley C.A.: The -omics era and its impact. Arch Pathol Lab Med 2004 Dec, VOL:128(12), P:1337.
- Kristensen V.N., Sørlie T., Geisler J., Langerød A., Yoshimura N., Kåresen R., Harada N., Lønning P.E., Børresen-Dale A.L.: Gene expression profiling of breast cancer in relation to estrogen receptor status and estrogen-metabolizing enzymes: clinical implications. Clin Cancer Res 2005 Jan, VOL:11(2 Pt 2), P:878s.
- Kuehn H., Liberzon A., Reich M., Mesirov J.P.: Using Gene Pattern for gene expression analysis. Curr Protoc Bioinformatics 2008 Jun, Chapter 7:Unit 7.12.
- Lønning P.E., Sørlie T., Børresen-Dale A.L.: Genomics in breast cancer-therapeutic implications. Nat Clin Pract Oncol 2005 Jan, VOL:2(1), P:26.
- Lunshof J.E., Chadwick R., Vorhaus D.B., Church GM.: From genetic privacy to open consent.Nat Rev Genet. 2008 May, VOL:9(5),P:406.
- Ma C., Staudt L.M.: Molecular definition of the germinal centre stage of B-cell differentiation. Philos Trans R Soc Lond B Biol Sci 2001 Jan, VOL:356(1405),P:83.
- Maas K., Chan S., Parker J., Slater A., Moore J., Olsen N., Aune T.M.: Cutting edge: molecular portrait of human autoimmune disease. J Immunol. 2002 Jul, VOL:169(1), P:5.
- Maynard J.A., Myhre R., Roy B.: Microarrays in infection and immunity. Curr Opin Chem Biol 2007 Jun VOL:11(3), P:306.
- McGall G., Labadie J., Brock P., Wallraff G., Nguyen T., Hinsberg W.: Light-directed synthesis of high-density oligonucleotide arrays using semiconductor photoresists. *Proc Natl Acad Sci USA* 1996 Nov, VOL:93 (24), P:13555.
- McPherson J.D., Marra M., Hillier L., Waterston R.H., Chinwalla A., Wallis J., Sekhon M., Wylie K., Mardis E.R., Wilson R.K., Fulton R., Kucaba T.A.,



Wagner-McPherson C., Barbazuk W.B., Gregory S.G., Humphray S.J., French L., Evans R.S., Bethel G., Whittaker A., Holden J.L. McCann O.T., Dunham A., Soderlund C., Scott C.E., Bentley D.R., Schuler G., Chen H.C., Jang W., Green E.D., Idol J.R., Maduro V.V., Montgomery K.T., Lee E., Miller A., Emerling S., Kucherlapati, Gibbs R., Scherer S., Gorrell J.H., Sodergren E., Clerc-Blankenburg K., Tabor P., Naylor S., Garcia D., de Jong P.J., Catanese J.J., Nowak N., Osoegawa K., Qin S., Rowen L., Madan A., Dors M., Hood L., Trask B., Friedman C., Massa H., Cheung V.G., Kirsch I.R., Reid T., Yonescu R., Weissenbach J., Bruls T., Heilig R., Branscomb E., Olsen A., Doggett N., Cheng J.F., Hawkins T., Myers R.M., Shang J., Ramirez L., Schmutz J., Velasquez O., Dixon K., Stone N.E., Cox D.R., Haussler D., Kent W.J., Furey T., Rogic S., Kennedy S., Jones S., Rosenthal A., Wen G., Schilhabel M., Gloeckner G., Nyakatura G., Siebert R., Schlegelberger B., Korenberg J., Chen X.N., Fujiyama A., Hattori M., Toyoda A., Yada T., Park H.S., Sakaki Y., Shimizu N., Asakawa S., Kawasaki K., Sasaki T., Shintani A., Shimizu A., Shibuya K., Kudoh J., Minoshima S., Ramser J., Seranski P., Hoff C., Poustka A., Reinhardt R., Lehrach H.; International Human Genome Mapping Consortium. A physical map of the human genome. Nature 2001, VOL:409(6822), P:934.

Mikhailovich V., Gryadunov D., Kolchinsky A., Makarov A.A., Zasedatelev A.: DNA microarrays in the clinic: infectious diseases. Bioessays 2008 Jul, VOL:30(7), P:673.

Miles JS, Wolf CR. Principles of DNA cloning. BMJ. 1989 Oct 21;299(6706):1019.

Müller A., Holzmann K., Kestler H.A.: Visualization of genomic aberrations using Affymetrix SNP arrays. Bioinformatics 2007 Feb, VOL:23(4), P:496.

Munshi N.C., Hideshima T., Carrasco D., Shammas M., Auclair D., Davies F., Mitsiades N., Mitsiades C., Kim R.S., Li C., Rajkumar S.V., Fonseca R., Bergsagel L., Chauhan D., Anderson K.C.: Identification of genes modulated in multiple myeloma using genetically identical twin samples. Blood 2004 Mar, VOL:103(5), P:1799.

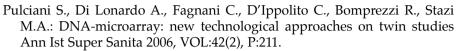
Okamoto T., Suzuki T., Yamamoto N.: Microarray fabrication with covalent attachment of DNA using bubble jet technology Nat Biotechnol 2000 Apr, VOL:18(4), P:438.

Pease A.C., Solas D., Sullivan E.J., Cronin M.T., Holmes C.P., Fodor S.P.A.: Light-generated oligonucleotide arrays for rapid DNA sequence analysis. Proc Natl Acad Sci USA 1994 May, VOL:91(11), P:5022.

Pennisi E.: DNA's cast of thousands. Science 2003 Apr, VOL:300(5617), P:282.

Polci M. L. & Petrucci T. C.: Array di proteine Caleidoscopio in preparazione 2009





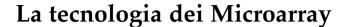
- Robertson B.H., Nicholson J.K.: New microbiology tools for public health and their implications. Annu Rev Public Health 2005, VOL:26, P:281.
- Rody A., Karn T., Ruckhäberle E., Müller V., Gehrmann M., Solbach C., Ahr A., Gätje R., Holtrich U., Kaufmann M.: Gene expression of topoisomerase II alpha (TOP2A) by microarray analysis is highly prognostic in estrogen receptor (ER) positive breast cancer. Breast Cancer Res Treat 2008 Mar 14 [Epub ahead of print].
- Saal L.H., Troein C., Vallon-Christersson J., Gruvberger S., Borg A., Peterson C.: BioArray Software Environment (BASE): a platform for comprehensive management and analysis of microarray data. Genome Biol 2002 Jul, VOL:3(8):SOFTWARE0003. Epub 2002 Jul 15.
- Salvatore M., Viganotti M., Lorenzetti S.: Studio di tumori pediatrici rari attraverso l'impiego di cDNA microarray . Caleidoscopio in preparazione 2009
- Schena M., Shalon D., Davis R.W., Brown P.O.: Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science 1995 Oct 20, VOL:270(5235), P:467.
- Seidel M., Niessner R.: Automated analytical microarrays: a critical review. Anal Bioanal Chem 2008 Jul, VOL:391(5), P:1521.
- Shoemaker D.D., Schadt E.E., Armour C.D., He Y.D., Garrett-Engele P., McDonagh P.D., Loerch P.M., Leonardson A., Lum P.Y., Cavet G., Wu L.F., Altschuler S.J., Edwards S., King J., Tsang J.S., Schimmack G., Schelter J.M., Koch J., Ziman M., Marton M.J., Li B., Cundiff P., Ward T., Castle J., Krolewski M., Meyer M.R., Mao M., Burchard J., Kidd M.J., Dai H., Phillips J.W., Linsley P.S., Stoughton R., Scherer S., Boguski M.S.: Experimental annotation of the human genome using microarray technology. Nature 2001 Feb, VOL:409(6822), P:922.
- Southern E.M.: Detection of specific sequences among DNA fragments separated by gel electrophoresis. J Mol Biol 1975, VOL:98(3), P:503.
- Spieth J., Lawson D.: Overview of gene structure. WormBook 2006 Jan, VOL:18, P:1-10.
- Townsend M.B., Dawson E.D., Mehlmann M., Smagala J.A., Dankbar D.M., Moore C.L., Smith C.B., Cox N.J., Kuchta R.D., Rowlen K.L.: Experimental evaluation of the FluChip diagnostic microarray for influenza virus surveillance. J Clin Microbiol 2006 Aug, VOL:44(8), P:2863.
- Trevino V., Falciani F., Barrera-Saldaña HA.: DNA microarrays: a powerful genomic tool for biomedical and clinical research. Mol Med 2007 Sep-Oct, VOL:13(9-10), P:527.





- van de Vijver M.J., He Y.D., van't Veer L.J., Dai H., Hart A.A., Voskuil D.W., Schreiber G.J., Peterse J.L., Roberts C., Marton M.J., Parrish M., Atsma D., Witteveen A., Glas A., Delahaye L., van der Velde T., Bartelink H., Rodenhuis S., Rutgers E.T., Friend S.H., Bernards R.: A gene-expression signature as a predictor of survival in breast cancer. N Engl J Med 2002 Dec, VOL:347(25), P:1999.
- van der Vegt B., de Bock G.H., Hollema H., Wesseling J.:Microarray methods to identify factors determining breast cancer progression: Potentials, limitations, and challenges. Crit Rev Oncol Hematol 2008 Oct, vol: 8. [Epub ahead of print]
- van 't Veer L.J., Dai H., van de Vijver M.J., He Y.D., Hart A.A., Mao M., Peterse H.L., van der Kooy K., Marton M.J., Witteveen A.T., Schreiber G.J., Kerkhoven R.M., Roberts C., Linsley P.S., Bernards R., Friend S.H.: Gene expression profiling predicts clinical outcome of breast cancer. Nature 2002 Jan, VOL:415(6871), P:530.
- VanGuilder H.D., Vrana K.E., Freeman W.M.: Twenty-five years of quantitative PCR for gene expression analysis. Biotechniques 2008 Apr, VOL:44(5), P:619.
- Virtanen C., Woodgett J.: Clinical uses of microarrays in cancer research. Methods Mol Med 2008, VOL:141, P:87.
- Wang D., Urisman A., Liu Y.T., Springer M., Ksiazek T.G., Erdman D.D., Mardis E.R., Hickenbotham M., Magrini V., Eldred J., Latreille J.P., Wilson R.K., Ganem D., DeRisi J.L.: Viral discovery and sequence recovery using DNA microarrays. PLoS Biol 2003 Nov, VOL:1(2), P:E2.
- Watson J.D., Crick F.H.C.: Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. Nature 1953 Apr, VOL:171(4356), P:737.
- Watson JD. The human genome project: past, present, and future. Science 1990 Apr, VOL:248(4951), P:44.
- Weigelt B., Hu Z., He X., Livasy C., Carey L.A., Ewend M.G., Glas A.M., Perou C.M., Van't Veer L.J.: Molecular portraits and 70-gene prognosis signature are preserved throughout the metastatic process of breast cancer. Cancer Res 2005 Oct, VOL:65(20), P:9155.
- Wheeler D.A., Srinivasan M., Egholm M., Shen Y., Chen L., McGuire A., He W., Chen Y.J., Makhijani V., Roth G.T., Gomes X., Tartaro K., Niazi F., Turcotte C.L., Irzyk G.P., Lupski J.R., Chinault C., Song X.Z., Liu Y., Yuan Y., Nazareth L., Qin X., Muzny D.M., Margulies M., Weinstock G.M., Gibbs R.A., Rothberg J.M.. The complete genome of an individual by massively parallel DNA sequencing. Nature 2008 Apr, VOL:452(7189), P:872.
- Wiltgen M., Tilz G.P.: DNA microarray analysis: principles and clinical impact. Hematology 2007 Aug, VOL:12(4), P:271.
- Zhu M.M., Wu Q.: Transcription network construction for large-scale microarray datasets using a high-performance computing approach. BMC Genomics 2008, VOL9(Suppl 1), P:S5.





¹Simonetta Pulciani, ¹Anna Di Lonardo, ²Alessandro Spurio, ³Corrado Fagnani

¹Centro per la Ricerca e la Valutazione dei Prodotti ImmunoBiologici, Istituto Superiore di Sanità. ²Settore attività editoriali, Istituto Superiore di Sanità. ³Centro Nazionale di Epidemiologia, Sorveglianza e Promozione della Salute, Istituto Superiore di Sanità

I Microarray

La tecnologia dei *Microarray* è nata poco più di dieci anni fa e fin dall'inizio ha destato un grande interesse nella comunità scientifica per le sue potenzialità nella ricerca di base e biomedica. Sin dal loro esordio i *Microarray* hanno subito numerose evoluzioni sia sotto il profilo tecnologico che applicativo, consentendo lo studio dell'organizzazione strutturale e funzionale del genoma umano e fornendo nuovi strumenti alla ricerca scientifica e traslazionale.

La tecnologia dei Microarray si fonda su sofisticate basi scientifiche e tecnologiche, ma il suo disegno sperimentale è semplice e lineare: consiste, praticamente, in microallineamenti di sonde su un unico vetrino. I Microarray, infatti, prendono proprio il loro nome dalle micro matrici formate dalle sonde deposte sul vetrino; tali sonde permettono di analizzare contemporaneamente un singolo campione per migliaia di "caratteristiche".

La prima pubblicazione relativa all'applicazione della tecnica dei Microarray risale al 1995, allorquando Pat Brown e collaboratori utilizzarono migliaia di microscopici allineamenti di sequenze di DNA, immobilizzati sul vetrino, per l'analisi dell'espressione genica in cellule di lievito, durante le varie fasi cellulari del ciclo cellulare (Schena M.). Dal momento che le micromatrici erano costituite da sequenze di DNA, la tecnica è stata denominata DNA-Microarray.

I DNA-Microarray rappresentano l'evoluzione delle tecniche di *Southern* e *Northen blot* sviluppate, circa trent' anni fa, sulla base della peculiare caratteristica degli acidi nucleici di formare ibridi a doppia elica in presenza di filamenti singoli complementari (Southern EM). Prima di addentrarci nella descrizione dei DNA-Microarray ci sembra opportuno dare dei brevi cenni sulla tecnica di *Southern Blot*. Tale tecnica fu ideata da Edwin Southern per rilevare la presenza di sequenze specifiche in un DNA campione (Southern EM). Il DNA da esaminare è trattato con enzimi di restrizione e successivamente sottoposto ad elettroforesi su gel d'agarosio o di poliacrilammide in modo da separare i



frammenti secondo il loro peso molecolare. Questi sono in seguito trasferiti e legati covalentemente su membrane di Nitrocellulosa, per capillarità o per elettroporazione. La membrana di nitrocellulosa, quindi, viene posta a contatto con una soluzione contenente le sequenze marcate in vario modo (fluorescenza, radioattività, ecc..). Queste ultime, ibridizzando con le sequenze complementari presenti sulla membrana di nitrocellulosa, ne permettono l'identificazione e la caratterizzazione. Una delle differenze più significative tra le tecniche di *Southern | Northern blot* e la tecnologia dei DNA Microarray risiede proprio nella preparazione dei campioni da analizzare e delle sonde. Precisiamo che, nel caso dei DNA Microarray, con il termine sonda o *probe* si definisce la sequenza complementare ai geni da analizzare. Nelle tecniche di *Southern | Northern blot* i campioni vengono legati ad una membrana e le sonde sono libere nella miscela di reazione, mentre nei DNA Microarray la situazione è capovolta, ossia i campioni sono contenuti in forma libera nelle miscele di reazione e le sonde sono disposte su un supporto solido (vetrino).

Altra differenza fondamentale tra le tecniche di blotting e i DNA Microarray è nel numero di sonde utilizzabili in ogni singolo esperimento: una sola nel primo caso, migliaia nel secondo (Yauk CL).

Nella seconda metà degli anni novanta due gruppi di ricercatori, operando indipendentemente uno dall'altro ed utilizzando strategie differenti, hanno realizzato la strumentazione della tecnologia dei DNA Microarray e messo a punto le relative procedure.

Il primo gruppo diretto da Pat Brown sviluppò, nei laboratori dell'Università di Stanford, una strumentazione capace di depositare automaticamente su vetrini da microscopio migliaia di sequenze geniche precedentemente sottoposte ad amplificazione genica (PCR Polymerase Chain Reaction). Questa procedura di deposizione permette di creare un micro reticolo costituito da micro spot di sequenze geniche a singola catena su un unico vetrino. Il vetrino così preparato permetterà l'analisi multigenica dei campioni che verranno ad esso ibridati. In particolare, questa tecnica fu utilizzata la prima volta per studiare il profilo di espressione di colture di lievito in stati fisiologici diversi, utilizzando campioni di mRNA marcati selettivamente con molecole fluorescenti (Schena M). Il secondo gruppo di ricercatori, diretto da Lipshutz e Fodor, prendendo spunto da tecnologie in uso nell'industria dei semiconduttori, mise a punto una tecnica fotolitografica idonea a sintetizzare oligonucleotidi direttamente su una superficie solida di quarzo (Lipshutz RJ 1995). In quest'ultimo approccio tecnologico, la formazione degli oligonucleotidi avviene per reazione fotochimica aggiungendo sequenzialmente le singole basi e permettendo la produzione di microarray contenenti circa 400.000 sonde su una superficie di pochi centimetri quadrati (Lipshutz RJ 1999). Gli approcci di Pat Brown e Lipshutz e Fodor hanno dato il via alle due principali tecniche metodologiche per la produzione dei DNA Microarray (Trevino V).



La metodologia proposta da Pat Brown utilizza sonde sintetizzate *in vitro* che vengono depositate su un supporto solido, generalmente un vetrino da microscopia trattato chimicamente. Per queste caratteristiche tale processo viene detto *spotting* e i vetrini prodotti con questo tipo di processo sono comunemente chiamati DNA Microarray. La metodologia proposta da Lipshutz e Fodor utilizza tecnologie che permettono di sintetizzare sonde di oligonucleotidi direttamente sul supporto solido. Questi microarray, prodotti "in situ" prendono il nome di *GeneChips* (Lipshutz RJ 1999).

I "GeneChips" vengono prodotti tramite tecnologie derivate dall'industria dei semiconduttori, sviluppate dalla società Affymetrix. La sintesi delle sonde avviene direttamente sulla superficie del supporto solido adottando tecnologie quali la fotolitografia e la sintesi diretta in fase solida di oligonucleotidi. Il silicio costituisce il supporto solido su cui viene innescata la reazione fotochimica per la sintesi della catena di oligonucleotidi (figura 1).

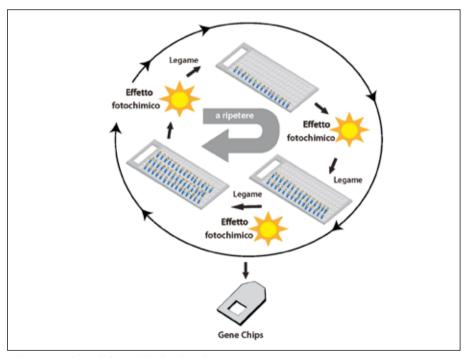


Figura 1. Sintesi fotochimica in situ.

Nella sintesi fotochimica in situ le sonde sono rappresentate esclusivamente da oligonucleotidi. Essi vengono sintetizzati per aggiunta progressiva di basi che vengono legate covalentemente per effetto fotochimico. Il processo si ripete più volte fino a formare il microreticolo (gene chips).



Maggiori dettagli su tale tecnica sono reperibili, principalmente, sul sito web della società *Affymetrix* e nella letteratura scientifica (Lipshutz RJ 1999).

I *GeneChips* hanno dei costi elevati, e data la strumentazione complessa e i relativi brevetti sono realizzabili soltanto dall'industria.

Inoltre, sono state sviluppate tecnologie che permettono di sintetizzare *in situ* oligonucleotidi mediante reazioni chimiche ed utilizzando un metodo di deposizione detto a *getto* o *Ink-jet* (Hughes TR 2001). La metodologia a *getto* o *Ink-jet* può essere utilizzata anche per depositare su vetrini pretrattati oligonucleotidi pre-sintetizzati in vitro (Okamoto T).

In questa rassegna non tratteremo dei sistemi *Affymetrix* e *Ink-jet*, ma focalizzeremo la nostra attenzione sui vetrini Microarray realizzati per deposizione di sonde preparate *in vitro*. Per semplicità nella descrizione, chiameremo tali vetrini DNA Microarray.

DNA Microarray

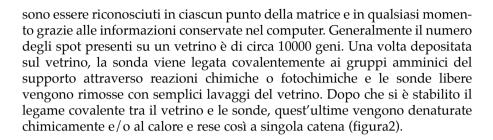
La metodologia degli DNA Microarray, ideata da Pat Brown, utilizza sistemi automatici capaci di deporre in maniera ben ordinata e riproducibile un alto numero di specifiche sonde precedentemente sintetizzate in vitro su un vetrino chimicamente trattato.

In questa metodica le sonde possono essere rappresentate da sequenze di cDNA o DNA clonati ed amplificati in PCR o da oligonucleotidi sintetizzati secondo sequenze conosciute. Le sonde derivate da amplificazione genica di cDNA o DNA possono essere note nella loro sequenza quando sono acquisite da banche dati genomiche, quali, ad esempio, GenBank, dbEST o UniGene.

Tali banche forniscono, oltre alle sequenze, informazioni utili per la successiva valutazione dei profili di espressione. Le sonde possono essere anche sconosciute, in quanto provenienti da librerie genetiche non ancora del tutto sequenziate, ma in ogni caso debbono essere note le estremità del frammento di DNA utilizzato per effettuare l'amplificazione genica. La deposizione delle sonde avviene tramite strumenti di fabbricazione industriale che sono chiamati *Arrayer* o *Printer* o *Spotter* (Hegde P.). Questi sono dei robot controllati da computer dotati di programmi contenenti informazioni sia sulle sonde (provenienza, letteratura scientifica, ecc) che sullo schema di deposizione.

Il braccio meccanico del robot dispone di speciali aghi che raccolgono le sonde dal piatto dove sono conservate e, con precisione micrometrica, le depositano sul vetrino nella posizione e nell'ordine stabiliti, tramite diretto contatto con la superficie del vetrino stesso. Si formano dei micro-*spots* di circa 80-150 µm di diametro, distanziati l'uno dall'altro 200-250 µm che pos-





Protocollo di Ibridazione per gli DNA Microarray

Una volta pronti, i vetrini di DNA Microarray potranno essere utilizzati per svariate analisi quali: profili di espressione, polimorfismi, amplificazioni

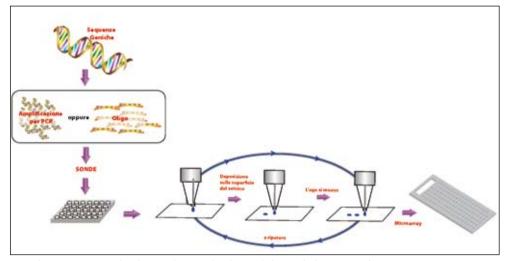


Figura 2. Tecnologie per la produzione dei vetrini DNA Microarray. Le sonde di DNA derivano da sequenze geniche clonate da vari tessuti e sequenziate. Esse possono essere cDNA amplificati per PCR o oligonucleotidi di sequenza nota, e vengono depositate sui vetrini a formare il reticolo del microarray mediante sistemi automatici.

Le sonde prodotte in vitro vengono aspirate da microaghi e deposte per contatto su vetrini chimicamente trattati. L'ago si muove e il processo si ripete più volte fino a formare il micro reticolo.





e delezioni geniche, ed anche variazioni epigenetiche (Floridia G., Sulewska A).

Nel nostro caso, ci limitiamo a descrivere il protocollo base elaborato per l'utilizzo dei DNA Microarray in una delle loro più comuni applicazioni: l'analisi dell'espressione differenziale.

E' fondamentale ed essenziale sottolineare che questo tipo di microarray sfrutta l'ibridazione delle sonde con cDNA preparati da due campioni, uno dei quali con funzione di controllo. Tali campioni sono marcati con due differenti fluorofori e confrontati.

Come esempio semplificativo della procedura sperimentale, descriviamo in figura 3 l'analisi dell'espressione genica differenziale tra la linea cellulare di ratto NT-R, non trasformata, e la linea cellulare T-R, trasformata. La linea cellulare di ratto trasformata, T-R, è stata ottenuta dalla linea cellulare non trasformata, NT-R, mediante trattamento con mutageni chimici.

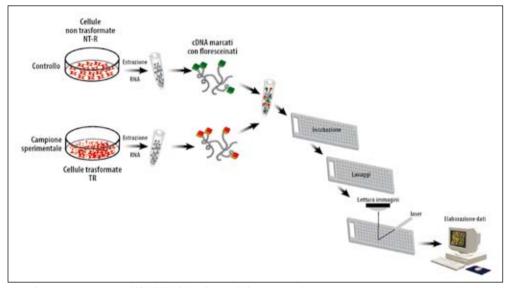


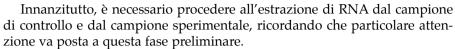
Figura 3. Protocolli di ibridazione dei DNA Microarray.

Gli RNA estratti dalle cellule trasformate e da quelle non trasformate vengono convertiti in cDNA, e marcati con i fluorofori Cy3 e Cy5. I due cDNA, così ottenuti, vengono entrambi ibridati su uno stesso vetrino di DNA-microarray.

Dopo il processo di ibridazione il vetrino viene lavato per eliminare sequenze non perfettamente appaiate. Successivamente viene letto con degli strumenti (Scanner) capaci di rilevare i segnali dei due fluorofori ed acquisire i dati della loro intensità.

I dati ottenuti vengono analizzati ed elaborati con programmi informatici appropriati per determinarne il significato biologico.





Specifici protocolli sono stati sviluppati per l'estrazione e conservazione degli RNA, per i quali consigliamo di consultare i manuali di laboratorio (Sambrook J) e le schede tecniche fornite dalle ditte produttrici di Kit di estrazione.

Gli RNA verranno miscelati con reagenti specifici e due differenti fluorofori, tipicamente le cianine *Cy3* e *Cy5*, in modo da ottenere la loro retrotrascrizione in cDNA e la loro marcatura. I fluorofori permetteranno di seguire i campioni nelle fasi successive al processo di ibridazione sul vetrino di microarray. I campioni di cDNA così preparati verranno dapprima uniti, quindi ibridizzati ad un singolo vetrino microarray ed infine incubati per tutta la notte a 65° gradi o a temperature diverse, in funzione dell'indagine da eseguire (Trevino V).

Come accennato precedentemente, la reazione di ibridazione si baserà sul principio di complementarietà tra le sonde presenti sul vetrino e i cDNA presenti nei campioni utilizzati.

Completata la fase di ibridazione i vetrini verranno sottoposti ad adeguati lavaggi, in modo da eliminare tutto ciò che si è legato in maniera aspecifica al vetrino microarray.

Esistono una serie di problemi legati all'uso dei fluorofori quali: la ridotta dimensione del Cy3 rispetto al Cy5 che induce un diverso livello di incorporazione dei due fluorofori, e la maggiore labilità di Cy5 rispetto a Cy3. Per ovviare a tutti questi inconvenienti di natura tecnica si effettua il cosiddetto "scambio di colori", per cui il campione sperimentale viene marcato in un primo esperimento con il Cy3 ed in un secondo esperimento con Cy5, o viceversa. Terminata tale fase, il vetrino verrà sottoposto a scansione mediante degli appositi strumenti (Scanner), generalmente dotati di due fasci di luce laser, a lunghezza d'onda specifica, in grado di eccitare i due fluorofori (figura 3). I segnali fluorescenti emessi dai due fluorofori hanno diversa lunghezza d'onda e sono raccolti da rilevatori in due canali separati. Alla fine della lettura, lo strumento fornirà due immagini del vetrino in esame, ognuna delle quali è derivata dall'acquisizione delle emissioni di ciascun fluoroforo, e gli spot saranno rappresentativi dell'intensità dei segnali. Le immagini, memorizzate su un computer, verranno analizzate con specifici software. L'intensità del segnale ottenuta dall'intero vetrino verrà sottratta ai valori di fondo, e verrà calcolato il rapporto tra l'intensità dei due fluorofori. Alla fine i dati, così elaborati, verranno convertiti tramite algoritmi in un'immagine che riprodurrà il reticolo del vetrino. Ogni spot avrà uno pseudo-colore indicativo del rapporto tra il livello di espressione del campione rispetto al controllo. Poiché il campione sperimentale è generalmente marcato con il



fluoroforo Cy5, gli spot aventi una scala di pseudo-colori con gradazioni del rosso indicheranno un'aumentata espressione del gene corrispondente nel campione. Viceversa, le varie gradazioni del verde indicheranno più elevati valori di espressione nel controllo. Infine, il colore giallo indicherà un'eguale espressione tra il campione e il controllo.

L'insieme dei dati sarà poi sottoposto ad analisi nel contesto sia del singolo esperimento che di esperimenti multipli. Per la grossa mole di dati ottenuti e la complessità di quesiti a cui rispondere, una parte critica ed essenziale della metodica dei microarray è l'analisi statistica.

Quest'ultima delicata e complessa fase della sperimentazione sarà oggetto di un capitolo successivo di questo numero monografico, cui rimandiamo il lettore (Cheung V.G., Giuliani A., Rogers S.).

Per un approfondimento sulla metodica e sui protocolli dei DNA Microarray si invita il lettore a consultare, ad esempio, il sito web dell'Università dello Utah "DNA MICROARRAY VIRTUAL LAB" all'indirizzo: http://learn.genetics.utah.edu/content/labs/microarray/.

E' necessario sottolineare che i dati ottenuti dalla tecnologia dei microarray necessitano sempre di ulteriori conferme sperimentali. Nel caso dell'analisi dell'espressione differenziale, tale conferma viene eseguita tramite la Real Time PCR (Vanguilder H.).

Infatti, la tecnologia dei DNA microarray dà una visione globale dei geni sovraespressi o sottoespressi, mentre con la Real Time PCR vengono analizzati quei geni di effettivo interesse nell'ambito dello studio, per verificare la correttezza dei risultati e avere delle stime quantitative della differente espressione.

In questo paragrafo ci siamo limitati alla trattazione della metodica dei DNA Microarray. Due paragrafi successivi tratteranno le metodiche e le applicazioni dei CGH–array e degli array di proteine (Floridia G., Polci M.L.). I CGH–array sono dedicati all'analisi di amplificazioni e delezioni geniche.

Per le tecnologie ed i protocolli Microarray elaborati per l'analisi delle mutazioni puntiformi e delle variazioni epigenetiche, rimandiamo alla letteratura scientifica consigliata in bibliografia (Bonetta L., Waddell N.), così come per i Microarray disegnati per lo studio dei microRNA e dei tessuti (Ach R.A., Lueking A.).

Gli sviluppi e le applicazioni della tecnologia dei Microarray sono innumerevoli, e sono stati messi a punto specifici protocolli per ogni tipo di applicazione.

Il protocollo qui da noi descritto fornisce solo delle indicazioni di carattere generale, che potranno costituire un punto di partenza per ulteriori approfondimenti per il lettore interessato.





Bibliografia

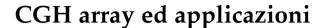
- Ach R.A., Wang H., Curry B.: Measuring microRNAs: comparisons of microarray and quantitative PCR measurements, and of different total RNA prep methods. BMC Biotechnol 2008 Sep. VOL: 8,P:69.
- Bonetta L.: Epigenomics: Detailed analysis. Nature 2008 Aug, VOL: 454(7205), P:795.
- Cheung V.G., Morley M., Aguilar F., Massimi A., Kucherlapati R., Childs G.: Making and reading microarrays. Nat Genet 1999 Nov, VOL:21(1 Suppl), P:15.
- Floridia G., Censi F., Lanni S.: CGH array ed applicazioni Questo volume 2008
- Giuliani A.: Analisi dei dati negli esperimenti di microarray Questo volume 2008
- Hegde P., Qi R., Abernathy K., Gay C., Dharap S., Gaspard R., Hughes J.E., Snesrud E., Lee N., Quackenbush J.: A concise guide to cDNA microarray analysis. Biotechniques 2000 Sep, VOL:29(3), P:548.
- Hughes T.R., Mao M., Jones A.R., Burchard J., Marton M.J., Shannon K.W., Lefkowitz S.M., Ziman M., Schelter J.M., Meyer MR., Kobayashi S, Davis C., Dai H., He Y.D., Stephaniants S.B., Cavet G., Walker W.L., West A., Coffey E., Shoemaker D.D., Stoughton R., Blanchard A.P., Friend S.H., Linsley P.S.: Expression profiling using microarrays fabricated by an ink-jet oligonucleotide synthesizer. Nat Biotechnol 2001 Apr, VOL:19(4), P:342.
- Lipshutz R.J., Fodor S.P., Gingeras T.R., Lockhart D.J.: High density synthetic oligonucleotide arrays. Nat Genet 1999 Jan, VOL:21(1 Suppl), P:20.
- Lipshutz R.J., Morris D., Chee M., Hubbell E., Kozal M.J., Shah N., Shen N., Yang R., Fodor S.P.: Using oligonucleotide probe arrays to access genetic diversity. Biotechniques 1995 Sep, VOL:19(3), P:442.
- Lueking A., Beator J., Patz E., Müllner S., Mehes G., Amersdorfer P.: Determination and validation of off-target activities of anti-CD44 variant 6 antibodies using protein biochips and tissue microarrays. Biotechniques 2008 Oct, VOL: 45(4), P:Pi-v.
- Okamoto T., Suzuki T., Yamamoto N.: Microarray fabrication with covalent attachment of DNA using bubble jet technology. Nat Biotechnol 2000 Apr, VOL:18(4), P:438.
- Polci M. L. & Petrucci T. C.: Array di proteine Questo volume 2008
- Rogers S., Cambrosio A.: Making a new technology work: the standardization and regulation of microarrays. Yale J Biol Med 2007 Dec, VOL:80(4), P:165.





- Sambrook J., MacCallum P., Russell D. (Eds): Molecular Cloning: A Laboratory Manual Cold Spring Harbor Laboratory Press, 2001.
- Schena M., Shalon D., Davis R.W., Brown P.O.: Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science 1995 Oct 20, VOL:270(5235), P:467.
- Southern E.M.: Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J Mol Biol* 1975, VOL:98(3), P:503.
- Sulewska A., Niklinska W., Kozlowski M., Minarowski L, Naumnik W., Niklinski J., Dabrowska K., Chyczewski L.: Detection of DNA methylation in eucaryotic cells. Folia Histochem Cytobiol 2007, VOL:45(4), P:315.
- Trevino V., Falciani F., Barrera-Saldaña HA.: DNA microarrays: a powerful genomic tool for biomedical and clinical research. Mol Med 2007 Sep-Oct, VOL:13(9-10), P:527.
- Vanguilder H., Vrana K., Freeman W.: Twenty-five years of quantitative PCR for gene expression analysis. Biotechniques 2008 Apr, VOL:44(5), P:619.
- Waddell N.: Microarray-based DNA profiling to study genomic aberrations. IUBMB Life 2008 Jul, VOL: 60(7), P:437.
- Yauk C.L., Berndt M.L.: Review of the literature examining the correlation among DNA microarray technologies. Environ Mol Mutagen 2007 Jun, VOL:48(5), P:380.





Giovanna Floridia, Federica Censi, Stella Lanni Centro Nazionale Malatte Rare - Istituto Superiore di Sanità, Roma

Dalla CGH alla CGH array

Alterazioni nel numero di copie di cromosomi o di regioni cromosomiche nell'intero genoma sono eventi frequenti in patologie tumorali e costituzionali. La "Comparative Genomic Hybridization (CGH)" fu sviluppata agli inizi degli anni 90 per una caratterizzazione delle variazioni del numero di copie, soprattutto nei tumori (Kallioniemi A. et al. 1992).

In questa tecnica due DNA genomici totali, provenienti da cellule da testare e da cellule di controllo normali, sono marcati con fluorocromi diversi e ibridati su cromosomi metafasici normali. Differenze nel rapporto di fluorescenza tra i due DNA sono, quindi, quantizzate per rilevare i cambiamenti del numero di copie nel DNA di interesse; il rapporto tra le due fluorescenze risulterà pari a 1 se non vi sono alterazioni nel numero di copie (Kallioniemi A. et al. 1992).

La CGH array (a-CGH) è una metodologia, sviluppata successivamente, che ha sostituito i cromosomi metafasici con array di sequenze genomiche; vantaggio principale rispetto alla CGH su metafase è quello di eseguire una analisi della variazione del numero di copie con una risoluzione molto più alta (Fig.1).

In generale i metodi basati su CGH array sono attualmente focalizzati sul rilevamento del cambiamento del numero di copie in un DNA da testare rispetto ad un DNA di riferimento; riarrangiamenti quali traslocazioni bilanciate o inversioni non sono rilevabili (Emanuel B.S. e Saitta S.C., 2007).

Nonostante la metodologia della CGH array sia robusta e abbia un ottimo livello di risoluzione secondo una prospettiva diagnostica, la maggiore difficoltà risiede nella assegnazione di causalità e significato clinico alle alterazioni rilevate; le CNV (Copy Number Variations), definite come variazioni del numero di copie di segmenti del DNA maggiori di 1Kb, esistono, infatti, in individui apparentemente normali (Emanuel B.S. e Saitta S.C., 2007).

Le CNV comprendono inserzioni, delezioni, cambiamenti complessi con guadagni e perdite nello stesso locus, e complesse CNV in siti multipli del genoma (Emanuel B.S. e Saitta S.C., 2007).





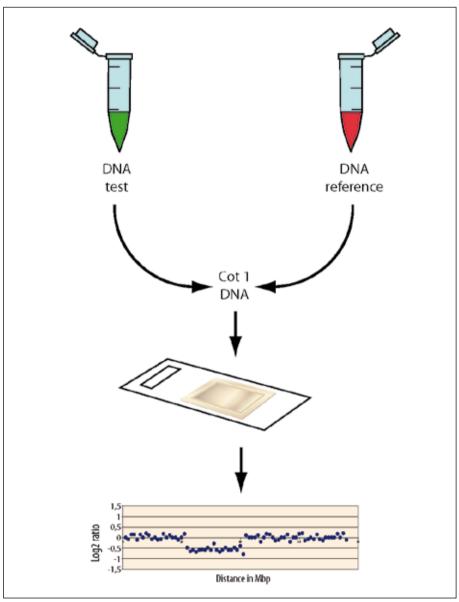


Figura 1. Array-CGH. Il DNA genomico in esame e un DNA genomico proveniente da un paziente sano sono marcati con fluorocromi diversi e coibridati su un array in presenza di Cot-1 DNA, in modo da sopprimere le sequenze ripetute. Un apposito software calcola, poi, i rapporti di fluorescenza per ciascuno spot dell'array e li normalizza.





Come discuteremo nelle sezioni relative alla diagnostica, banche dati con informazioni su variazioni normali in diversi gruppi etnici e test sui genitori sani dei pazienti permettono di comprendere se un cambiamento del numero di copie rilevato dalla CGH array è verosimilmente legato alla patologia.

Caratteristiche dei CGH array

La risoluzione della CGH array varia a seconda della densità e delle sonde spottate sull'array che possono variare da cloni genomici (80-200 kb) ad oligonucleotidi (25-85 bp) (Emanuel B.S. e Saitta S.C., 2007).

Tutte le piattaforme sviluppate si basano sul principio comune di rilevare alterazioni nel numero di copie confrontando due campioni e variano secondo la grandezza della sonda "spottata" e della copertura del genoma.

Le prime piattaforme di CGH array sono state sviluppate usando cloni BAC (Bacterial Artificial Chromosome), YAC (Yeast Artificial Chromosome) o PAC (P1-derived Artificial Chromosome).

In seguito sono stati spottati cloni più corti di cosmidi e fosmidi e molecole di DNA a singola elica di 130-600 bp sino ad arrivare all'uso di oligonucleotidi a singolo filamento di 25-85 mer (Ylstra B. et al., 2006).

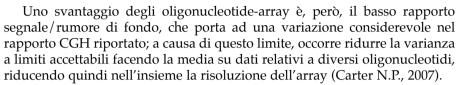
I cloni BAC hanno, in genere, una grandezza di 80-200 Kb e consentono sia una rappresentazione del genoma migliore rispetto agli altri cloni o prodotti PCR usati per gli array che una ibridazione con basso background; tuttavia, a causa delle loro dimensioni, non permettono l'identificazione di alterazioni più piccole di 50 kb. I cloni di cosmidi e fosmidi sono di circa 40 Kb, hanno un buon rapporto segnale/rumore di fondo, e permettono l'identificazione di CNV fino a 20 Kb (Carter N.P., 2007).

Molti gruppi hanno utilizzato array con cDNA ma, per il rilevamento di variazioni del numero di copie, la distribuzione disomogenea dei geni produce una risoluzione variabile lungo il genoma e il non appaiamento tra cDNA e DNA genomico durante l'ibridazione riduce il segnale e i cambiamenti del rapporto.

Una risoluzione più alta può essere ottenuta utilizzando prodotti PCR da DNA genomico ma il rapporto segnale/rumore può essere basso e i costi per la generazione delle sonde, soprattutto per gli array a grande scala, sono alti. Una limitazione degli array con cloni e prodotti PCR è che essi sono generalmente costruiti facendo uno spotting meccanico di soluzioni di DNA su vetrini; la copertura dell'intero genoma è, quindi, limitata dal fatto che non possono essere solitamente spottati più di 60.000 cloni (Carter N.P., 2007).

I CGH array con oligonucleotidi hanno il più alto potere risolutivo; recentemente sono stati sviluppati array commerciali in cui gli oligonucleotidi sono sintetizzati direttamente sul vetrino.





Per migliorare il rapporto segnale/rumore può essere utilizzato un metodo di marcatura, detto ROMA (Representational Oligonucleotide Microarray Analysis), che riduce la complessità del DNA genomico usato per l'ibridazione (Lucito R. et al., 2003). Tale metodo prevede uno step di digestione enzimatica, seguito da una PCR che consente l'amplificazione dei frammenti più piccoli (fino a circa 1.2 kb); tuttavia, il rapporto segnale/rumore di fondo rimane più basso dei BAC array, per cui occorre fare una media tra almeno tre oligonucleotidi per ridurre la varianza (Carter N.P., 2007).

Inoltre, data la frequenza di variazione del numero di copie nel genoma, array con una più alta densità di oligonucleotidi richiederanno una valutazione attenta del disegno della sonda per una genotipizzazione accurata del numero di copie in regioni complesse del genoma (Emanuel B.S. e Saitta S.C., 2007).

Piattaforme commerciali basate sugli oligonucleotide-array sono fornite da varie aziende tra cui Agilent Technologies (http://www.agilent.com), OGT (http://www.ogt.co.uk) PerkinElmer (http://www.perkinelmer.com/default.htm) e NimbleGen (http://www.nimblegen.com).

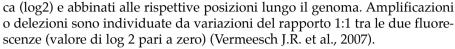
Oltre ad array che coprono l'intero genoma, vi sono array disegnati per una o più regioni specifiche del genoma al fine di studiare un cromosoma specifico o una regione cromosomica o identificare e valutare specifiche anomalie nel numero di copie in individui con sospette sindromi o riarrangiamenti (Bejjani B.A. e Shaffer L.G., 2006).

Affymetrix e Illumina utilizzano sia oligonucleotidi che rappresentano SNPs (short-nucleotide polymorphisms) che sonde selezionate polimorfiche, al fine di rilevare, in un unico esperimento, variazioni del numero di copie, perdita di eterozigosità e disomia uniparentale (www.affymetrix. com; www. illumina.com); l'ibridazione non viene effetuata mediante coibridazione dei due DNA, ma ibridando un singolo DNA.

Analisi dei dati

Nell' array-CGH il DNA target e il DNA di riferimento sono marcati con fluorocromi diversi, la cui intensità di fluorescenza è letta da un apposito scanner per ogni singola sequenza dell'array. I rapporti di fluorescenza tra il DNA target e quello usato come riferimento sono normalizzati al fine di evitare eventuali errori dovuti a motivi tecnici, quindi riportati in scala logaritmi-





Con l'incremento delle applicazioni dell'array CGH e lo sviluppo di piattaforme differenti è stata sviluppata una serie di software dedicati alla visualizzazione dei dati e all'analisi statistica (Lockwood W.W. et al., 2006).

CGH array e studio dei tumori

Il genoma di un tumore contiene numerose alterazioni a livello dei cromosomi, della cromatina e dei nucleotidi; si ha, quindi, una attivazione/inibizione di diversi eventi biologici quali l'angiogenesi, l'evasione immunitaria, la metastatizzazione ed una alterazione di crescita, apoptosi e metabolismo cellulare (Chin L. Gray J.W., 2008).

La CGH array permette di analizzare cambiamenti nel DNA tumorale mediante una comparazione con un DNA di riferimento che può derivare da cellule sane del paziente o da un altro individuo. L'identificazione di regioni sistematicamente delete o amplificate in un gruppo di tumori permette, quindi, di identificare geni coinvolti nella progressione dei tumori, quali oncosoppressori ed oncogeni.

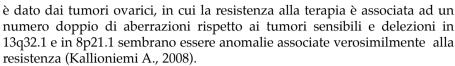
L'uso di array con oligonucleotidi ha, inoltre, aumentato la risoluzione di mappatura permettendo una definizione estremamente precisa delle bande e dei punti di rottura con una risoluzione teorica di meno di una kilobase. Grazie a questa metodologia sono state identificate piccole delezioni omozigoti, in cui mappano geni oncosoppressori ed è stata definita la reale natura di alcune traslocazioni apparentemente bilanciate.

Una metanalisi eseguita su 373 studi di aCGH su tumori ha confermato che vi sono patterns specifici per diversi tumori da utilizzare per una efficiente classificazione; inoltre, profili del numero di copie genomiche possono distinguere sottogruppi all'interno di entità istologicamente definite (Kallioniemi A., 2008).

Inoltre, dalla comparazione di campioni tumorali che rappresentano differenti stadi dello sviluppo del tumore, quali lesioni premaligne o in situ, tumori invasivi e metastatici, è stato dimostrato che nell'insieme il numero di cambiamenti del numero di copie aumenta durante la progressione tumorale (Kallioniemi A., 2008).

Un numero alto di aberrazioni del numero di copie è legato anche ad una prognosi sfavorevole e l'identificazione di aberrazioni genetiche specifiche associate con la prognosi è stato l'obiettivo di numerosi studi. Un esempio





Infine specifiche aberrazioni genetiche sono state associate a precise caratteristiche clinicopatologiche, quali progressione della patologia, risposta alla terapia o prognosi del paziente. La CGH array fornisce, quindi, un eccellente punto di partenza per l'identificazione di geni coinvolti nei tumori; tuttavia sono necessarie analisi funzionali per stabilire il reale contributo di tali geni nella patogenesi (Kallioniemi A., 2008).

CGH array e studio di disordini genetici

La a-CGH si è dimostrata molto utile nella identificazione di aberrazioni non identificabili mediante l'analisi citogenetica tradizionale in molti disordini genetici. Variazioni del numero di copie sono state ridefinite nella Sindrome Cri-du-chat e nella Sindrome di Prader-Willi; delezioni e duplicazioni submicroscopiche sono state identificate in pazienti citogeneticamente normali e che presentavano ritardo mentale e dismorfismi (Lockwood W.W. et al., 2006).

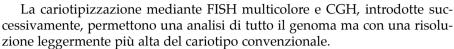
Inoltre, regioni candidate in cui sono localizzati geni putativi responsabili di disordini genetici umani, sono state definite mediante a-CGH ad alta risoluzione; un esempio è quello del lavoro di Vissers e coautori, i quali, analizzando due pazienti con Sindrome di CHARGE con delezioni sovrapponibili in 8q12, hanno ridotto il numero di geni candidati ad otto, tra cui CHD7, le cui mutazioni sono riportate nel 90% dei pazienti (Bergman J.E. et al., 2008, Vissers L.E. et al., 2004).

CGH array nella diagnosi di patologie costituzionali

L'analisi citogenetica convenzionale nella diagnostica permette il rilevamento di anomalie cromosomiche e riarrangiamenti in tutto il genoma ma con alcuni limiti: a) richiede da quattro a dieci giorni, b) il cariotipo ha una risoluzione di 5-10 Megabasi, c) sono richieste competenze specifiche per l'analisi del cariotipio con bandeggio (Vermeesch J.R. et al., 2007).

L'applicazione di tecniche molecolari, quali la FISH, la PCR quantitativa e la MLPA (multiplex ligation dependent probe amplification), permette di identificare anomalie cromosomiche e variazioni del numero di copie ad una risoluzione submicroscopica su un numero limitato di regioni genomiche.





La CGH array è stata recentemente ottimizzata e applicata nel campo della diagnostica citogenetica costituzionale per rilevare riarrangiamenti sbilanciati su tutto il genoma: lo screening delle variazioni del numero di copie ad un'alta risoluzione in un contesto clinico è stato definito cariotipo molecolare (Vermeesch J.R. et al., 2007).

L'analisi dei dati della CGH array nella citogenetica clinica porta alla suddivisione delle CNV in a) benigne, b) patogeniche, c) di significato clinico non noto (Lee C. et al., 2007).

Le CNV che si sovrappongono a regioni associate a sindromi note da microdelezione o microduplicazione (o ad altre regioni genomiche definite significative dal punto di vista clinico, quali quelle subtelomeriche) sono probabilmente patogeniche in natura. Tali sbilanci, a seconda della natura e della grandezza, sono spesso validati con una tecnologia di rilevamento molecolare alternativa quale la FISH interfasica o su metafasi, la MLPA, la PCR quantitativa, la genotipizzazione dei microsatelliti, la long range PCR e, in alcuni casi, mediante un'altra piattaforma di CGH array o di genotipizzazione (Lee C. et al., 2007).

Nel caso in cui le CNV ritrovate sono inizialmente di significato clinico sconosciuto, occorre determinare se esse sono ereditate o de novo testando entrambi i genitori nella stessa piattaforma o utilizzando altre metodologie (Lee C. et al., 2007).

La presenza dello sbilancio genomico sia nell'individuo affetto che nel genitore sano indica, probabilmente, la non patogenicità della CNV; in assenza di materiale dei genitori, si possono testare parenti sani e ammalati per cercare di capire il significato (Lee C. et al., 2007).

Nel caso di uno sbilancio genomico di CNV non ereditate o de novo occorre consultare banche dati pubbliche, tra cui il Database of Genomic Variants (http://projects.tcag.ca/variation), in cui sono accumulati dati su CNV di centinaia di individui sani.

CNV de novo, non ritrovate nel database delle varianti, dovrebberero essere cercate in banche dati di CNV patogeniche ritrovate in individui ammalati quali il DECIPHER (http://www.sanger.ac.uk/PostGenomics/decipher/), il CAD (http://www.ukcad.org.uk/cocoon/ukcad/), the Mendelian Cytogenetics Network Online Database e l'European Cytogeneticists Association Register of Unbalanced Chromosome Aberrations - ECARUCA (http://www.ecaruca.nt) (Lee C. et al., 2007).

Le CNV de novo, specialmente quelle mai osservate tra individui sani e che non coincidono con CNV ricorrenti osservate in individui con caratteristiche cliniche simili, possono essere valutate usando diversi parametri (Lee C. et al., 2007).



Criteri maggiori	Caratteristica di CNV	Caratteristica di CNV
	patogenica	benigna
1. a. CNV ereditata da un genitore sano		X
b. CNV ereditata da un genitore	X	
ammalato		
2. a. CNV simile a CNV in un parente sa	no	X
b. CNV simile a CNV in un parente	X	
ammalato		
3. a. CNV che si sovrappone ad uno		
sbilancio genomico in un database		
per individui sani (per es. Database		
di varianti Genomiche)		X
b. CNV che si sovrappone ad uno		
sbilancio genomico in un database di		
CNV di individui malati		
(per es. DECIPHER)	X	
4. CNV che contiene geni morbosi OMIN		
5. a. CNV ricca di geni	Χ	
b. CNV povera di geni		X
Criteri minori	CNV patogenica	CNV benigna
1. a. la CNV è una delezione	X	
b. la CNV è una delezione omozigote	X	
2. a. la CNV è una duplicazione		X
b. la CNV è una amplificazione	X	
(acquisto di più di una copia)		
3. CNV > di 3 Mb	X	
4. la CNV è priva di noti elementi regola	tori	X
Legenda: CNV: Variazione del numero di copie; copy number variation		

Tabella 1. Fattori che influenzano la valutazione di rischio di una CNV (Tabella tradotta da C. Lee, A.J. Iafrate e A.R.Brothman, 2007)

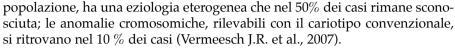
Nella tabella 1 sono riassunti criteri utili per la valutazione della potenziale patogenicità di una CNV (Lee C. et al., 2007).

CGH array e diagnosi postnatale

La CGH array è una metodologia di accertata validità clinica nella diagnosi di ritardo mentale e di ritardo dello sviluppo.

Il ritardo mentale, con o senza anomalie congenite, si ritrova nel 3% della





Molti studi, rivisti da Vermesh et al., hanno dimostrato l'efficacia della CGH array nel rilevare sbilanci patogenici in individui con cariotipo normale; dati ottenuti usando array con diversa risoluzione hanno mostrato che nell'insieme il 7-11% dei pazienti avevano sbilanci cromosomici criptici patogenici submicroscopici. Inoltre, in coorti di pazienti non analizzate per sbilanci subtelomerici, si ritrovano riarrangiamenti subtelomerici nel 5% dei casi (Vermeesch J.R. et al., 2007).

La detection rate degli sbilanci patogenici aumenta usando array a più alta risoluzione; infatti screening di pazienti con ritardo mentale/malformazioni congenite hanno identificato che il 20% degli sbilanci sono più piccoli di 1 Mb (Vermeesch J.R. et al., 2007).

La CGH array, inoltre, è molto utile per mappare e misurare con precisione la grandezza di aberrazioni grazie all'uso di array ad alta risoluzione o array specifici contenenti sequenze di regioni cromosomiche di interesse al fine di identificare le regioni o i geni responsabili di un determinato fenotipo (Sanlaville D. et al., 2005).

Infine, in portatori di traslocazioni apparentemente bilanciate de novo od ereditate con fenotipo anomalo, è possibile identificare sbilanciameti cripici mediante l'uso di CGH array ad alta risoluzione (Sanlaville D. et al., 2005).

CGH array e diagnosi prenatale

Una piccola ma significativa proporzione dei test, eseguita in diagnosi prenatale, rivela uno sbilancio cromosomico, responsabile di fenotipi complessi tra cui ritardo mentale e difetti alla nascita.

Anomalie più frequentemente riscontrate sono le trisomie dei cromosomi 21, 13 e 18, e aneuploidie dei cromosomi sessuali; riarrangiamenti strutturali, mosaicismi, poliploidie e marker cromosomici rappresentano solo un piccolo numero delle alterazioni citogenetiche identificate (Rickman L. et al., 2005).

L'analisi mediante CGH array permette l'analisi ad alta risoluzione dell'intero genoma in tempi brevi; il cariotipo molecolare rileverà sia sbilanci cromosomici associati a note patologie dello sviluppo, sia sbilanci di cui non si conoscono ancora le conseguenze fenotipiche (Vermeesch J.R. et al., 2007).

La rilevanza clinica di CNV benigne catalogate richiede ulteriori studi ed inoltre, per ogni nuovo sbilancio mai riportato, è difficile predire le conseguenze fenotipiche e la penetranza in assenza di tanti dati familiari e di correlazioni genotipo/fenotipo (Vermeesch J.R. et al., 2007).



Una soluzione a questo problema potrebbe essere lo sviluppo di array con un numero limitato di target ben noti (Vermeesch J.R. et al., 2007).

Array commerciali, quali quelli forniti dalla Spectral Genomics (http://www.spectralgenomics.com), sono attualmente disponibili per il rilevamento di anomalie cromosomiche, riarrangiamenti cromosomici e telomerici associati con sindromi genetiche note e ritardo mentale.

La CGH array ha la potenzialità di sostituire l'analisi del cariotipo convenzionale in diagnosi prenatale, ma sono necessari ulteriori studi al fine di raggiungere un consensus sull'optimum di un array per uso clinico (Rickman L. et al., 2006).

Conclusioni

La CGH array è una tecnologia che sta rapidamente evolvendosi con lo sviluppo di tante piattaforme ed è sempre più applicata in diverse aree della genetica umana.

I continui avanzamenti tecnologici e banche dati con un crescente numero di informazioni pemetteranno sempre più l'ampliamento d'uso dell'aCGH sia nella diagnostica che nella ricerca (Lockwood et al., 2006).

Bibliografia

- Bejjani B.A., Shaffer L.G.: Application of array-based comparative genomic hybridization to clinical diagnostics. J Mol Diagn. 2006 Nov, VOL: 8 (5), P: 528.
- Bergman J.E., de Wijs I., Jongmans MC, Admiraal RJ, Hoefsloot LH, van Ravenswaaij-Arts CM.: Exon copy number alterations of the CHD7 gene are not a major cause of CHARGE and CHARGE-like syndrome. Eur J Med Genet. 2008 Sept-Oct, VOL: 51(5), P: 417.
- Chin L., Gray J.W.: Translating insights from the cancer genome into clinical practice Nature. 2008 Apr, VOL: 452 (7187), P: 553.
- Carter N.P.: Methods and strategies for analyzing copy number variation using DNA microarrays. Nat Genet 2007 Jul, VOL: 39 (7 Suppl), P:S16.
- Emanuel B.S., Saitta S.C.: From microscopes to microarrays: dissecting recurrent chromosomal rearrangements. Nat Rev Genet. 2007 Nov, VOL: 8 (11), P: 869.





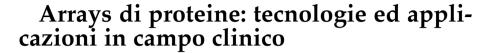
- Kallioniemi A., Kallioniemi O.P., Sudar D., Rutovitz D., Gray J.W., Waldman F., Pinkel D.: Comparative Genomic Hybridization for molecular cytogenetic analysis of solid tumors. Science 1992 Oct, VOL: 258 (5083), P: 818.
- Kallionemi A.: CGH microarray and cancer. Curr Opin in Biotech 2008, Feb, VOL:19 (1), P: 36.
- Lee C., Iafrate A.J., Brothman A.R.: Copy number variations and clinical cytogenetic diagnosis of constitutional disorders. Nat Genet. 2007 Jul, VOL: 39 (7 Suppl), P:S48.
- Lockwood W.W., Chari R., Chi B., Lam W.L.: Recent advances in array comparative genomic hybridization technologies and their applications in human genetics. Eur J Hum Genet. 2006 Feb, VOL:14 (2), P: 139.
- Lucito R., Healy J., Alexander J., Reiner A., Esposito D., Chi M., Rodgers L., Brady A., Sebat J., Troge J., West J.A., Rostan S., Nguyen K.C., Powers S., Ye K.Q., Olshen A., Venkatraman E., Norton L., Wigler M.: Representational oligonucleotidi microarray analysis: a high-resolution method to detect copy number variation Genom Res. 2003 Oct, VOL: 13 (10), P: 2291.
- Rickman L., Fiegler H., Carter N.P., Bobrow M.: Prenatal diagnosis by array-CGH. Eur J Med Genet. 2005 Jul-Sep, VOL: 48 (3), P:232.
- Rickman L., Fiegler H., Shaw-Smith C., Nash R., Cirigliano V., Voglino G., Ng B.L., Scott C., Whittaker J., Adinolfi M., Carter N.P., Bobrow M.: Prenatal detection of unbalanced chromosomal rearrangements by array CGH. J Med Genet. 2006 Apr, VOL: 43 (4), P: 353.
- Sanlaville D., Lapierre J.M., Turleau C., Coquin A., Borck G., Colleaux L., Vekemans M., Romana S.P.: Molecular karyotyping in human constitutional cytogenetics Eur J Med Genet. 2005 Jul-Sep, VOL: 48 (3), P: 214.
- Vermeesch J.R., Fiegler H., de Leeuw N., Szuhai K., Schoumans J., Ciccone R., Speleman F., Rauch A., Clayton-Smith J., Van Ravenswaaij C., Sanlaville D., Patsalis P.C., Firth H., Devriendt K., Zuffardi O.: Guidelines for molecular karyotyping in constitutional genetic diagnosis. Eur J of Hum Genet 2007 Nov, VOL 15(11), P: 1105.
- Vissers L.E., van Ravenswaaij C.M., Admiraal R., Hurst J.A., de Vries B.B., Janssen I.M., van der Vliet W.A., Huys E.H., de Jong P.J., Hamel B.C., Schoenmakers E.F., Brunner H.G., Veltman J.A., van Kessel A.G.: Mutations in a new member of the chromodomain gene family cause CHARGE syndrome Nat Genet. 2004 Sep, VOL: 36 (9), P: 955.
- Ylstra B., van den Ijssel P., Carvalho B., Brakenhoff R.H., Meijer G.A.: BAC to the future! or oligonucleotides: a perspective for micro array comparative genomic hybridization (array CGH). Nucleic Acids Res. 2006 Jan, VOL: 34(2), P: 445.



I Microarray







Maria Letizia Polci¹ e Tamara Corinna Petrucci²

¹Dipartimento Ambiente e Connessa Prevenzione, ²Dipartimento di Biologia
Cellulare e Neuroscienze, Istituto Superiore di Sanità - Roma

Gli array di proteine

Gli array di proteine, conosciuti anche come biochip, sono strumenti utilizzati in campo biomedicale per analisi qualitative e quantitative di proteine in campioni biologici. Lo sviluppo dei biochip nasce da tecniche già in uso in genomica riadattate alla proteomica: gli array di DNA puntano principalmente ad identificare potenziali difetti genetici che causano alterazioni delle vie di segnalazione cellulare; gli array di proteine, d'altro canto, forniscono informazioni cruciali sullo stato funzionale di queste vie alterate. La conoscenza del network di proteine in un particolare stato cellulare è il primo passo verso l'identificazione di potenziali molecole diagnostiche e di bersagli farmacologici. L'integrazione di questi due strumenti permette, dunque, di raggiungere una visione completa del macchinario cellulare alterato alla base di una patologia.

Un array di proteine è costituito da un supporto inerte (lastrine di vetro, membrane o biglie) sul quale sono immobilizzati anticorpi monoclonali, proteine, frammenti proteici, peptidi o aptameri. Queste molecole dette "esca" vengono scelte *ad hoc* per identificare interazioni proteina-proteina, substrati di chinasi o target di molecole piccole biologicamente attive. La rivelazione degli array di proteine si ottiene attraverso un passaggio di incubazione con anticorpi, ligandi, siero o lisati cellulari che contengono un determinato generatore di segnale. L'intensità del segnale sarà proporzionale alla quantità della molecola che si è legata alla molecola esca (Figura 1).

Il design dei biochip garantisce facilità di utilizzo, automazione e grande rapidità, sensibilità, e possibiltà di generare un gran numero di dati in un singolo esperimento. Nel formato più comune di un array di proteine, degli anticorpi sono depositati sul chip e usati come molecole esca per analizzare proteine presenti in lisati cellulari. Questi array che seguono il principio dei saggi ELISA sono destinati ad accelerare le tecniche di immunodiagnostica. Ekins negli anni '80 ha introdotto gli immunosaggi con micro-gocce di campione (micro-spot) dimostrandone il guadagno in sensibilità e selettività





(Ekins RP). Oltre alle applicazioni in campo diagnostico, la tecnologia degli array di proteine può aiutare la ricerca di base nello studio delle interazioni proteina-proteina e permettere di conoscere (studiare) il profilo dell' espressione proteica a partire da un numero limitato di proteine fino ad un'analisi proteomica globale, mentre nell'industria farmaceutica gli array possono essere utilizzati per l'identificazione di bersagli di farmaci e nella validazione di processi. Tra i principali usi dei biochip risalta l'utilizzo per l'individuazione del profilo delle modifiche post-traduzionali di proteine: monitorare l'entità delle modifiche post-traduzionali nel tempo, prima e dopo un trattamento farmacologico, o tra stato patologico e non patologico, può dare informazioni in tempo reale sui livelli di attività di una particolare proteina.

Tecnologie, chimica e formati degli array di proteine

La tecnologia alla base di un array di proteine deve superare dei limiti analitici molto significativi che per contro non esistono per gli array di geni. Un biochip deve innanzitutto essere in grado di rilevare un ampio intervallo di concentrazione di analiti: le proteine di interesse possono essere presenti in una cellula in un intervallo dinamico di concentrazione che può variare anche di un fattore 10^{10} . La seconda sfida è la ricerca di adeguate sensibilità e specificità nell'identificazione di proteine poco abbondanti che generalmente è inficiata da un rapporto segnale/rumore inaccettabile dovuto alla co-presenza di molecole contaminanti che danno reazioni incrociate. Per l'identificazione di proteine poco abbondanti coinvolte in processi di segnalazione cellulare sono richiesti sistemi con sensibilità nell'intervallo di poche femtomoli. In ultimo, all'array di proteine viene richiesta la capacità di bloccare molecole endogene presenti nel sistema in analisi che possono interferire con la chimica di amplificazione del segnale.

Forward e reverse-phase arrays

Esistono due diversi approcci nel designo di un array di proteine: il forward-phase e il reverse-phase array (Zong Y, Espina V 2007). Il forward-phase array contiene varie molecole esca, tipicamente anticorpi, marcate in modo opportuno ed immobilizzate sul substrato e quindi, dopo incubazione con un singolo campione, fornisce informazioni su molteplici proteine di inte-



resse in esso contenute (Figura 1). Gli analiti catturati sono rivelati con una seconda molecola marcata o marcando direttamente gli analiti.

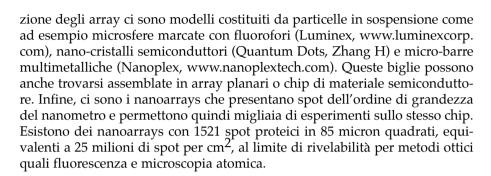
I reverse-phase array prevedono l'immobilizzazione degli analiti stessi presenti in una miscela. In ogni reverse-phase array, sono quindi depositati molti campioni e l'incubazione avviene con una singola molecola esca che rivela un singolo analita in molteplici campioni nelle stesse condizioni sperimentali. Questo formato viene, ad esempio, spesso usato per determinare la forma fosforilata della proteina d'interesse presente in un campione e può, quindi, fornire informazioni sull'entità della fosforilazione di una data proteina in diversi pazienti (Espina V 2003). Il tipico formato di un reverse-phase array consiste nell'immobilizzazione dei campioni ottenuto da un certo numero di pazienti su un singolo array. Spesso sono campioni ottenuti dagli stessi pazienti, prima e dopo un trattamento, oppure campioni normali contro campioni patologici. Ciascun array viene ibridato con un anticorpo, consentendo così un confronto sia per la presenza sia per la quantità di uno stesso analita nei diversi campioni.

Tipi di chip

Gli array di proteine sono stati inizialmente disegnati come una miniaturizzazione di saggi immunologici, quali ELISA e dot blotting, che spesso utilizzano una lettura a fluorescenza e presentano sistemi di rivelazione robotizzati per garantire l'esecuzione di analisi multiple. I supporti utilizzati nei biochip sono: vetro, silicone, microwells, membrane di nitrocellulosa o PVDF e biglie (magnetiche e non). I requisiti fondamentali per un substrato di un array di proteine sono: 1) grande capacità di legame 2) tendenza a non alterare la struttura delle proteine 3) basso background. Le lastrine di vetro ricoperte di nitrocellulosa sono il substrato più comunemente usato nella costruzione di un array di proteine (FAST slides, Schleicher & Schuell Bio-Sciences, Dassel, Germany). Questo formato permette di immobilizzare lo stesso campione su tante lastrine, massimizzando quindi il numero dei differenti agenti esca da utilizzare per una stessa serie di array. La nitrocellulosa inoltre rispetta i criteri di elevata capacità di legame con proteine, con valori pari a 75–150 mg/cm² in un volume di 0.3–2 nL/spot.

Nel formato più diffuso i micro-spot di proteine vengono depositati su superfici planari, ma esistono delle strutture alternative che prevedono sistemi di centrifugazione (Gyros, www.gyros.com) o design molto specializzati come microcanali su piastra (The Living Chip, www.biotrove.com) e superifici siliconiche (Zyomyx, www.zyomyx.com). Nella nuova genera-





Molecole esca

Le molecole esca più comunemente utilizzate nei biochip sono anticorpi monoclonali che rappresentano gli agenti meglio caratterizzati e più efficaci per la cattura di proteine (Figura 2). Gli anticorpi usati negli array vengono generalmente prodotti secondo procedure convenzionali di immunizzazione o come frammenti ricombinanti espressi in E. Coli. Recentemente si sta diffondendo anche l'impiego di molecole esca non proteiche, ad esempio gli aptameri, per studiare le interazioni di proteine con DNA che permetterebbero di ovviare ai problemi di specificità che possono presentare alcuni anticorpi in commercio. Gli aptameri vengono selezionati da librerie di oligonucleotidi e la loro interazione con le proteine di interesse può essere migliorata attraverso incorporazione di deossiuridina brominata o mediante ligandi UV-attivati (fotoaptameri). Il cosiddetto "photocrosslinking" riduce reazioni secondarie degli aptameri grazie ad impedimento sterico. Il grosso vantaggio degli aptameri risiede essenzialmente nella facilità di produzione e nella stabilità del DNA.

Inoltre, con l'immobilizzazione su chip di acidi nucleici, recettori, enzimi e proteine oggi è permessa un'ampia gamma di esperimenti per identificare interazioni proteina-proteina e i substrati che si legano alle proteine di interesse.

Produzione degli arrays di proteine

Nei forward array, le molecole esca possono essere sintetizzate esternamente, opportunamente purificate ed infine immobilizzate sull'array. In



alternativa esiste una sintesi peptidica in situ che evita il passaggio di immobilizzazione. La costruzione dei reverse-phase array prevede il seguente protocollo: i) estrazione di proteine da cellule, sieri o tessuti; ii) immobilizzazione su supporto solido; iii) incubazione con anticorpi selezionati ed infine rivelazione. Le tecniche di immobilizzazione sono state prese in prestito dagli array di geni, sebbene la struttura dell'array di proteine risulti completamente diversa rispetto ad un tipico array di DNA. Entrambe le tecnologie di immobilizzazione trasferiscono il campione fluido su un substrato con una grande area superficiale, generalmente una lastrina di vetro ricoperta. Gli array di DNA si presentano come spot ad alta densità dello stesso campione Il substrato è tipicamente vetro ricoperto da gruppi amminici che permettono l'adesione del DNA carico negativamente al substrato positivo. Vengono utilizzati sistemi di rivelazione fluorometrici o a luminescenza caratterizzati da un adeguato rapporto segnale/rumore. Il metodo di immobilizzazione scelto deve essere riproducibile e automatizzabile per garantire analisi multiple in parallelo. La scelta del pennino e del tipo di strumento che si usa per l'immobilizzazione su supporto è cruciale per la costruzione di un buon array. Parametri fondamentali da considerare nella scelta del sistema di immobilizzazione sono: il volume di campione a disposizione, la viscosità, il numero degli array richiesti per ogni campione e il tipo di substrato. La fase di immobilizzazione (printing) nella costruzione di un array di proteine può far uso di diverse tecniche (Schena M) a seconda del tipo di campione da immobilizzare, del suo volume e della capacità dell'array in analisi. Metodi di fabbricazione di biochip includono printing robotizzato a contatto, inkjetting, spotting piezoelettrico e fotolitografia. Esistono in commercio degli strumenti in grado di preparare array in modo automatizzato o equipaggiamenti per una costruzione manuale. In alternativa, colonie batteriche possono essere roboticamente distribuite su membrane di polyvinylidene difluride-(PVDF) per l'induzione dell'espressione di una proteina in situ.

Le tecnologie di printing si dividono essenzialmente in strumenti a contatto e a non-contatto. Il printing a contatto viene realizzato attraverso il contatto diretto tra la testina contenente il campione da immobilizzare e il substrato. Nel printing a non-contatto un volume minimo di campione viene rilasciato sul substrato. Esempi di formati che fanno uso di printing a contatto e a non-contatto sono riportati in tabella 1.

La struttura tipica di un array di proteine consiste di una serie di 4–6 spot in orizzontale che rappresentano una mini curva di diluizione per ciascun campione (Figura 3). L'uso di queste diluizioni assicura che la proteina di interesse abbia una concentrazione nell'intervallo dinamico del saggio, a seconda della sensibilità e dell'affinità dell'anticorpo usato.





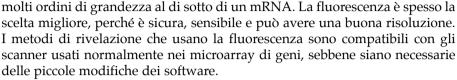
Considerazioni

Parametri fondamentali per la bontà dell'immobilizzazione sono le caratteristiche chimico-fisiche delle molecole da immobilizzare e la natura della superficie su cui aderiranno. Un buon supporto per un array di proteine è chimicamente stabile al processo di immobilizzazione, risulta in una buona morfologia degli spot, presenta un minimo legame non specifico, non contribuisce al background ed è compatibile con diversi sistemi di rivelazione. Il metodo di immobilizzazione usato deve essere riproducibile, applicabile a proteine con differenti proprietà (dimensione, idrofilicità, idrofobicità), e automatizzabile. Diverse chimiche di immobilizzazione sono state descritte per la fabbricazione degli array di proteine. L'assorbimento passivo, o immobilizzazione non-covalente, è metodologicamente molto semplice ma adatto a piccole quantità di campione, non permette un buon controllo dell'orientamento delle molecole rilasciate, potrebbe alterare le proprietà funzionali delle proteine e ha riproducibilità ed efficienza molto variabili. L'immobilizzazione covalente risulta invece in legami stabili, può essere applicata ad un ampio insieme di proteine ed ha buona riproducibilità, sebbene la derivatizzazione chimica potrebbe alterare la funzione delle proteine. I substrati utilizzati nella immobilizzazione covalente includono lastre di vetro ricoperte con reagenti silanici contenenti gruppi amminici o aldeidici. Ne è un esempio il sistema Versalinx, nel quale il legame covalente è dovuto all'interazione tra proteine derivatizzate con acidi fenilboronico e salicilidrossamico immobilizzati sulla superficie del supporto. Questo formato ha un background limitato, una bassa fluorescenza intrinseca e non altera la funzione delle proteine. Un legame non-covalente di proteine non modificate avviene all'interno di strutture porose come HydroGel (Saaem I) che consistono in gel di poliacrilamide tridimensionali. Per questi tipi di array si riporta un bassissimo background, un'alta capacità e conservazione della funzione delle proteine. Esistono metodi di immobilizzazione ampiamente usati che consistono nello sfruttare le forti e specifiche interazioni biotina/streptavidina o esa-istidina/Nichel dopo aver opportunamente modificato le proteine. La biotina può essere conjugata ad una catena di poli-lisine immobilizzata su supporti tipo diossido di titanio o pentossido di tantalio.

Sistemi di rivelazione

I sistemi di rivelazione usati nei protein array devono tenere conto della concentrazione di una proteina in un campione biologico, che può essere





La marcatura con fluorofori e il loro uso nei metodi di rivelazione che vediamo negli array di DNA può essere quindi estesa anche agli array di proteine. Ad esempio, si possono incubare dei forward-phase array con proteine fluorescenti provenienti da due diversi stati cellulari rappresentati da lisati coniugati con differenti fluorofori (Cy-3, Cy-5) e miscelati, così che i due colori rendano visibile delle variazioni nell'abbondanza della proteina bersaglio. La lettura della fluorescenza può essere amplificata da 10 a 100 volte mediante amplificazione del segnale mediante l'uso di tiramide (Tyramide Signal Amplification, TSA) (Morozov VN, Schweitzer B, King G, Kukar T, Wiese R).

Negli array con anticorpi gli analiti vengono rivelati direttamente o attraverso anticorpi secondari in un sistema a sandwich. Il metodo diretto è usato per il confronto di campioni diversi con colori differenti. Nel caso in cui siano disponibili coppie di anticorpi in grado di legare lo stesso analita, i sistemi a sandwich sono invece preferiti perché molto specifici e sensibili e rappresentano il sistema migliore per l'analisi di proteine poco abbondanti come citochine, o di modifiche postraduzionali di proteine.

Nel caso dei reverse-phase array, ogni spot ha in genere un diametro di alcune centinaia di micron e può rappresenta l'intero proteoma cellulare o complessi fluidi biologici. Le strategie di rivelazione richiedono una buona amplificazione del segnale quale prerequisito per raggiungere la sensibilità richiesta per l'analisi di proteine poco abbondanti. Un metodo affidabile sfrutta reazioni enzimatiche già collaudate nei saggi immunologici delle analisi cliniche. (In particolare si usa la deposizione del complesso biotinatiramide sul substrato in cui sarà presente il complesso anticorpo biotinilatoligando).

Esistono poi sistemi ad elevata sensibilità che utilizzano particelle in sospensione (Luminex) e le proprietà di semiconduttori dei nanocristalli (Quantum Dots). Oggi sono disponibili varie alternative nuove nell'arena delle biotecnologie commercializzate. Queste includono variazioni della "surface plasmon resonance", amplificazione con DNA circolare, spettrometria di massa con sorgente SELDI (Surface Enhanced Laser Desorption Ionization). In quest'ultima metodologia vengono usati dei chip per la purificazione di proteine accoppiati alla spettrometria di massa SELDI-TOF che rivela le molecole isolate. I chip per la purificazione contengono fasi solide per cromatografia destinate a legare proteine con caratteristiche di idrofobicità o di carica simili, presenti in miscele complesse quali plasma o estratti proteici di



cellule o tessuti tumorali. Questa tecnologia differisce dagli array di proteine perché non prevede immobilizzazione di molecole esca e l'identificazione di proteine mediante interazioni con specifici ligandi.

Aree di applicazione

E' possibile suddividere le applicazioni dei biochip in tre principali aree, ognuna con uno specifico formato e relativa metodologia di lettura.

- Diagnostica: identificazione di antigeni e anticorpi nel sangue; visualizzazione del profilo proteico del siero per individuare nuovi potenziali marcatori tumorali; monitoraggi in matrici alimentari e ambientali.
- 2) Proteomica: determinazione dell'espressione proteica; costruzione di array specifici per un organo o per una patologia; analisi funzionale di proteine (interazioni proteina-proteina, identificazione di ligandi di recettori, determinazione di attività enzimatica, mappatura degli epitopi di anticorpi).
- Studi farmacologici: saggi di competizione in cui piccole molecole potenzialmente utilizzabili come inibitori biologici sono testate nel legame con chinasi specifiche in presenza di ligandi fluorescenti.

Diagnostica

I biochip nei vari formati analizzati forward-phase e reverse-phase arrays hanno ampio utilizzo in diagnostica per effettuare saggi immunologici in grande scala.

Quale strumento in campo clinico, l'array di proteine ha il grande vantaggio di poter effettuare simultaneamente molte analisi sia di diversi campioni sia ripetute dello stesso campione. Gli array di anticorpi sono principalemte utilizzati nell'identificazione e quantificazione di citochine; quelli di antigeni sono impiegati invece nell'analisi di anticorpi presenti nel siero per la diagnosi di infezioni, malattie autoimmuni e allergie. La sensibilità dei biochipe e la possibilità di effettuare analisi multiple in parallelo permette l'identificazione di marcatori tumorali in estratti da piccole quantità di materiale bioptico, aprendo nuove possibilità nella diagnostica e nella terapia dei tumori. L'identificazione di nuovi autoanticorpi è oggi resa possibile dall'analisi di sieri utilizzando proteine umane immobilizzate su un array. La specificità



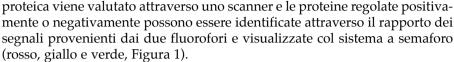
degli autoanticorpi e il loro coinvolgimento in processi patologici apre a questi biochip un importante ruolo nella diagnosi e prognosi delle malattie autoimmuni (AutoImmune Diseases, AID). Poche AID come il diabete di tipo 1 sono oggi diagnosticate in vitro attraverso profili di autoanticorpi, ma per molte altre come la sclerosi multipla o l'artrite idiopatica giovanile gli antigeni coinvolti sono sconosciuti. Recentemente è stata presentata una strategia per identificare nuovi marcatori di AID che prevede una combinazione della tecnologia degli array di proteine con l'uso delle librerie di cDNA. Questa strategia è stata utilizzata in uno studio sull'artrite idiopatica giovanile, l'artrite reumatoide e la sclerosi multipla (Lueking A). Un'altra applicazione dei biochip in diagnostica è il monitoraggio della SARS (Severe Acute Respiratory Syndrome), in cui un array di proteine da coronavirus fa da esca per le proteine specifiche per la SARS. Questi biochip sono stati usati per lo screening di 400 sieri canadesi durante l'epidemia da SARS, inclusi campioni di casi accertati di SARS e di pazienti affetti da patologie respiratorie. E' stato poi sviluppato un algoritmo per predizioni di diagnosi della SARS e quindi utilizzato per monitorare 206 sieri di pazienti cinesi con febbre alta. I risultati ottenuti correlavano benissimo con i test di immunofluorescenza indiretta. I biochip si sono dimostrati strumento efficace, rapido e sensibile per l'identificazione di anticorpi specifici per virus nel siero (Zhu H).

Proteomica

L'analisi di proteine in cellule e tessuti e il confronto tra differenti stati (normale, patologico, stati diversi di differenziazione, e prima e dopo trattamento farmacologico) è l'obiettivo centrale della proteomica. In proteomica i biochip vengono usati principalmente per studi qualitativi e quantitativi di proteine in campioni patologici e sani (ad es. nella visualizzazione del profilo di espressione proteica, in cui i biochip competono con l'elettroforesi su gel bidimensionale). Nei saggi funzionali *in vitro* vengono invece studiate le interazioni proteina-proteina, proteina-DNA, proteina-farmaco, recettoreligando, enzima-substrato.

Nel formato usato per ottenere il profilo proteico differenziale, i biochip sono incubati con una miscela di campioni (ad es. due estratti proteici ottenuti da tessuto normale e patologico) e gli analiti di interesse vengono catturati dalle specifiche esche. Analogamente al confronto tra tessuti normali e patologici mediante array di DNA o elettroforesi su gel bidimensionale, i campioni vengono marcati con differenti fluorofori, Cy3 e Cy5, miscelati, gel filtrati e incubati sul chip di anticorpi. L'aumento o la diminuzione dell'espressione





Per studiare le interazioni proteina-proteina i biochip sono un'alternativa *in vitro* ai sistemi cellulari e risultano molto più efficaci come ad esempio per le interazioni che coinvolgono proteine secrete o proteine con ponti disolfuro.

Esempi di analisi di attività biologica con i biochip sono stati descritti per chinasi del proteoma di lievito (Ptacek J).

Studi farmacologici

La principale applicazione dei biochip in studi farmacologici è rappresentata da saggi di competizione di chinasi. Procognia (http://www.procognia.com/) ha presentato recentemente un saggio competitivo in cui piccole molecole, potenziamente inibitori di chinasi, vengono studiate nel legame con varie chinasi in presenza di ligandi universali marcati con fluorofori. Questo saggio può essere utilizzato per caratterizzare l'interazione di specifici composti con centinaia di chinasi nello stesso giorno.

Esiste, inoltre, un'interessante applicazione in cui i biochip supportano un doppio strato fosfolipidico adatto a studi di elettrofisiologia e applicazioni di biosensori. In particolare, lo studio si focalizza sul possibile uso delle proteine di membrana per studi farmacologici. Un esempio ne è la ricostruzione di un canale di potassio in proteoliposomi che vengono poi fusi nel doppio strato (Zagnoni M).

Conclusioni

Gli array di proteine stanno diventando rapidamente uno strumento molto efficace per identificare proteine, valutare il loro livello di espressione e investigare le loro interazioni e funzioni. Si sta assistendo ad un rapido progresso e alla crescita dell'interesse verso questa emergente area biotecnologica. L'obiettivo di questa tecnologia è il raggiungimento di una buona sensibilità ed efficacia in analisi di proteine che prevedano molte determinazioni in parallelo mediante procedure automatizzate. Il formato caratteristico di un array nasce con l'avvento dell'analisi dell'espressione genica ma questa





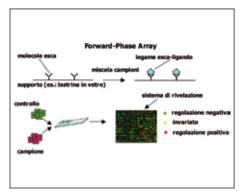


Figura 1. Disegno di un esperimento di un array di proteine secondo l'approccio "forward-phase".

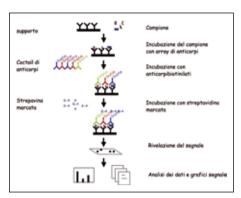


Figura 2. Anticorpi monoclonali utilizzati come molecole esca nei "biochip"

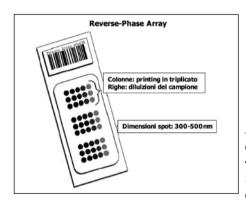


Figura 3. Struttura tipica di un array di proteine, ralizzata con una serie di 4–6 spot in orizzontale che rappresentano una mini curva di diluizione per ciascun campione.

	Printing a contatto	Printing a non-contatto
Tecnologia	Contatto diretto col substrato	Contatto fluido
Tipi	Quill, ink jet, pin and ring	Piezoelettrico, solenoide
Volume di campione	0.3-2.0 nl	0.1-0.3 nl piezoelettrico
rilasciato		4-8 nl solenoide
Determinazione dimensione dello spot	Diametro pennino	Volume della goccia
-		

Tabella 1. Esempi di formati che fanno uso di printing a contatto e a non-contatto.

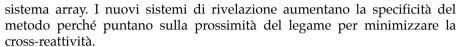


versione presenta difetti nell'indicare l'effettiva abbondanza proteica e non fornisce informazioni sulle modifiche post-traduzionali e sulle interazioni proteina-proteina. Concepiti originariamente come miniaturizzazione di saggi immunologici, gli array di proteine hanno subito numerosi sviluppi con il miglioramento delle tecniche di espressione di proteine ricombinanti. Oggi la tecnologia dei biochip permette di effettuare in parallelo lo screening di migliaia di interazioni quali: proteina-proteina, proteina-anticorpo, proteina-ligando, proteina-farmaco, enzima-substrato, così come saggi diagnostici multianalita. Il grande vantaggio dei biochip è la capacità di identificare inoltre con buona sensibilità ed accuratezza bassi livelli di proteine richiedendo una minima conoscenza tecnica da parte degli utilizzatori. Per tutti questi motivi si prevede che presto diventeranno una piattaforma affidabile ed efficace per un'analisi molecolare più completa che non abbia i limiti dei chip di DNA.

L'attuale tecnologia alla base degli array di proteine presenta degli aspetti tecnici importanti, alcuni specifici delle proteine in generale, altri caratteristici dei sistemi multi-analisi che devono essere risolti per arrivare ad esplicare la massima potenzialità di questi strumenti. Si tratta in particolare di problemi legati alla costruzione di array, alla selezione di esche per ottenere un profilo completo di espressione e di funzione proteiche e alla scelta di sistemi di rivelazione che siano sensibili e con un buon intervallo dinamico di concentrazione. Per quel che riguarda la costruzione di biochip, la difficoltà maggiore è oggi rappresentata dalla produzione dell'enorme numero di proteine e di molecole esca che saranno poi depositate sull'array e ne rappresenteranno l'elemento chiave per sensibilità e specificità di funzionamento. Per superare tale limite è richiesta una buona collezione di cloni di alta qualità per la purificazione di proteine e devono essere sviluppati dei sistemi che permettano produzione di molte proteine in parallelo. L'obiettivo è quello di arrivare ad un metodo di espressione globale, così che potenzialmente tutte le proteine diventino disponibili per la costruzione di un array sempre più specifico. Un'ulteriore considerazione riguarda il folding corretto e la conservata funzionalità delle proteine depositate sull'array. Questa caratteristica richiede validazioni puntuali, difficili da effettuare in caso di proteine con funzione ancora sconosciuta. La chimica delle proteine è un altro punto in cui la costruzione dei biochip presenta delle difficoltà, soprattutto riguardo all'esistenza di frequenti e differenti modifiche post-traduzionali da incorporare negli array.

Il design dei forward array che utilizzano anticorpi come molecole esca, in particolare nel caso di analisi di miscele complesse come plasma ed estratti da tessuto, deve superare il problema della cross-reattività che si presenta soprattutto nel caso di analisi multiple. Infine, i sistemi di rivelazione devono essere migliorati per raggiungere i requisiti di sensibilità e specificità del





La complessità del proteoma umano è di gran lunga maggiore di quella del genoma, se si pensa che eventi di gene splicing e di modificazioni post-traduzionali fanno salire di almeno un ordine di grandezza il numero di proteine umane rispetto a quello dei geni. Lo sviluppo degli array di proteine che ha permesso di migliorare le analisi dell'espressione proteica e dell'interazione tra proteine, è di grande supporto per le nuove scoperte della genomica funzionale. Le analisi proteomiche convenzionali che usano elettroforesi su gel bidimensionale e spettrometria di massa, sebbene molto efficaci, non hanno la capacità di identificare e poter quantificare molte proteine poco abbondanti spesso di grandissimo interesse diagnostico. Manca a tutt'oggi una tecnologia sensibile, di facile uso e di analisi multiple che permetta l'identificazione di proteine, la loro quantificazione e l'analisi differenziale dell'espressione proteica in stati patologici verso normali. Per questa ragione, gli array di proteine stanno generando un enorme interesse sia nella ricerca di base che nelle biotecnologie, soprattutto nelle applicazioni cliniche.

La realizzazione del profiling molecolare di campioni clinici ha il potenziale di cambiare la diagnostica di importanti patologie per le quali c'è scarsità di tecniche disponibili. L'utilità dei biochip è stata dimostrata in studi fatti su campioni di siero e di tessuto (Knezevic V, Miller JC). L'inserimento di comuni tecniche di biologia molecolare e di analisi clinica nelle procedure dei reverse-phase array, ha reso questi biochip facilmente utilizzabili e molto riproducibili. La loro applicazione nella valutazione dell'efficacia di un trattamento già dai primi giorni di azione potrà portare ad un intervento precoce e quindi ad un possibile decorso positivo della malattia.

Bibliografia

Ekins R.P., Chu F.W.: Multianalyte Microspot Immunoassay-Microanalytical "Compact Disk" of the Future. Clin. Chem 1991 Nov, VOL: 37 (11), P: 1955.

Espina V., Mehta A.I., Winters M.E., Calvert V., Wulfkuhle J., Petricoin E.F., Liotta L.A.: Protein microarrays: Molecular profiling. Proteomics 2003 Nov, VOL: 3 (11), P: 2091.

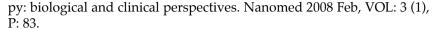
Espina V, Wulfkuhle J.D, Calvert V.S, Petricoin E.F 3rd. Liotta L.A.: Reverse phase protein microarrays for monitoring biological responses Methods Mol, Biol 2007, VOL: 383, P: 321.





- King G., Payne S., Walker F., Murray, G.I.: A highly sensitive detection method for immunohistochemistry using biotinylated tyramine. J. Pathol 1997 Oct, VOL: 183(2), P: 237.
- Kukar T., Eckenrode S., Gu Y., Lian W. et al.: Protein microarrays to detect protein-protein interactions using red and green fluorescent proteins. Anal. Biochem 2002 Jul, VOL: 306 (1), P: 50.
- Knezevic V., Leethanakul C., Bichsel V.E., Worth J.M., Prabhu V.V., Gutkind J.S., Liotta L.A., Munson P.J., Petricoin E.F. 3rd, Krizman D.B.: Proteomic profiling of the cancer microenvironment by antibody arrays. Proteomics 2001 Oct, VOL: 1(10), P: 1271–1278.
- Lueking A., Huber O., Wirths C., Schulte K., Stieler K.M., Blume-Peytavi U.,
 Kowald A., Hensel-Wiegel K., Tauber R., Lehrach H., Meyer H.E. e Cahill
 D.J: Profiling of alopecia areata autoantigens based on protein microarray technology. Mol. Cell. Proteomics 2005 Sep, VOL:4 (9), P: 13823.
- Miller J.C., Zhou H., Kwekel J., Cavallo R., Burke J., Butler E.B., Teh B.S., Haab B.B.: Antibody microarray profiling of human prostate cancer sera: antibody screening and identification of potential biomarkers. Proteomics 2003 Jan, VOL: 3(1), P: 56–63.
- Morozov V.N., Gavryushkin A.V., Deev A.A.: Direct detection of isotopically labeled metabolites bound to a protein microarray using a charge-coupled device. J. Biochem. Biophys. Methods 2002 Mar, VOL: 51 (1), P: 57.
- Ptacek J., Devgan G., Michaud G., Zhu H., Zhu X-, Fasolo J-, Guo H., Jona G., Breitkreutz A., Sopko R., McCartney R.R., Schmidt M.C., Rachidi N., Lee S.J., Mah A.S., Meng L., Stark M.J., Stern D.F., De Virgilio C., Tyers M., Andrews B., Gerstein M., Schweitzer B., Predki P.F., Snyder M.: Global analysis of protein phosphorylation in yeast. Nature. 2005 Dec, VOL: 438 (7068), P: 679.
- Saaem I., Papasotiropoulos V., Wang T., Soteropoulos P., Libera M.: Hydrogelbased protein nanoarrays. J. Nanosci. Nanotechnol 2007 Aug, VOL: 7 (8), P: 2623.
- Schena M.: Microarray Biochip Technology, Eaton. Natick 2000.
- Schweitzer B, Roberts S, Grimwade B, Shao W, Wang M, Fu Q, Shu Q, Laroche I, Zhou Z, Tchernev VT, Christiansen J, Velleca M, Kingsmore SF.: Multiplexed protein profiling on microarrays by rolling-circle amplification. Nat. Biotechnol 2002 Apr, VOL: 20 (4), P: 359.
- Wiese R.: Analysis of several fluorescent detector molecules for protein microarray use. Luminescence 2003 Jan-Feb, VOL: 18 (1), P: 25–30.
- Zagnoni M., Sandison M.E., Marius P., Lee A.G., Morgan H.: Controlled delivery of proteins into bilayer lipid membranes on chip. Lab. Chip 2007 Sep, VOL: 7 (9), P: 1176.
- Zhang H., Yee D., Wang C.: Quantum dots for cancer diagnosis and thera-





- Zhu H., Hu S., Jona G., Zhu X., Kreiswirth N., Willey B.M., Mazzulli T., Liu G., Song Q., Chen P., Cameron M., Tyler A., Wang J., Wen J., Chen W., Compton S., Snyder M.: Severe acute respiratory syndrome diagnostics using a coronavirus protein microarray Proc. Natl. Acad. Sci 2006 Mar, VOL: 103 (11), P: 4011–4016.
- Zong Y., Zhang S., Chen H.T., Zong Y., Shi Y.: Forward-phase and reverse-phase protein microarray. Methods Mol. Biol 2007, VOL: 381, P: 363.



I Microarray





Analisi dei dati negli esperimenti di microarray

Alessandro Giuliani

Dipartimento Ambiente e Prevenzione Primaria, Istituto Superiore di Sanità, Roma

Il grande numero di informazioni simultanee impone un ripensamento profondo nell'elaborazione dei dati scientifici

I dati provenienti dai cosiddetti *Microarray* sull'attivazione differenziale di migliaia di geni, ottenuti simultaneamente sullo stesso campione biologico costringono il ricercatore a farsi delle domande che coinvolgono il senso profondo della sua attività di ricerca ed ad acquisire consapevolezza della non neutralità della scelta della procedura statistica da utilizzare. Questa affermazione verrà qui sostanziata da una carrellata sulle principali strategie di analisi dei dati utilizzate per lo studio delle tecniche *'high-throughput'* o comunque ad alto parallelismo.

Nel giro di pochi anni la tecnica dei Microarray è passata da essere una procedura di frontiera della ricerca genetica, a rappresentare lo standard praticamente ubiquitario nella ricerca biomedica in campo molecolare. Parallelamente, la metodologia statistica per l'analisi dei dati prodotti da questa tecnica è andata da semplici tecniche di visualizzazione a nuovi algoritmi sviluppati ad hoc per trattare un genere di informazione altamente inusuale per le tecniche più note a medici e biologi ma, come vedremo, assolutamente in linea con l'esperienza di campi disciplinari apparentemente molto lontani come l' analisi dei segnali, la meccanica statistica e la meteorologia. La ricerca dei 'metodi più efficienti' per trattare i dati provenienti dai Microarray ha costretto la comunità scientifica ad avere un rapporto molto diverso di quello che biologi e medici usualmente intrattengono con le tecniche statistiche. Qui non era possibile semplicemente selezionare un piccolo insieme di tecniche da usare acriticamente come una ricetta di cucina (cosa che succede normalmente in epidemiologia o negli studi clinici, tanto che le riviste scientifiche di quei settori a volte riportano le 'tecniche consigliate' nelle istruzioni per gli autori). Al contrario l'elaborazione quantitativa del dato è diventata parte integrante della interpretazione biologica e la scelta di una strada invece di un'altra implica molto di più di una convenienza di analisi corrispondendo



ad una precisa scelta di campo sulle nostre idee generali del funzionamento di un sistema biologico. In questo, le scienze biomediche si sono ritrovate in una condizione del tutto nuova che in qualche modo le porta allo statuto epistemologico della fisica, dove la matematica utilizzata è parte integrante della teoria e non una semplice appendice.

In questo articolo cercherò di fornire, in modo forzatamente parziale ma spero almeno chiaro, le linee principali dei problemi aperti dalla necessità di trarre informazioni rilevanti dalle grandi masse di dati provenienti dai microarray.

La forma dei dati

La prima domanda che ci si deve porre in qualsiasi analisi statistica ha a che vedere con la forma generale dei dati e corrisponde a chiedersi 'Quali sono i miei casi? Quali sono le mie variabili?' (Siegel S., 1980). Questa domanda nella sperimentazione biomedica classica non creava soverchi problemi ed era suggerita dal tipo di osservazione, per cui se si voleva studiare l'effetto di un farmaco sulla pressione arteriosa i casi (unità statistiche) erano i pazienti arruolati secondo certi criteri di inclusione e le variabili la misura scelta di pressione arteriosa e tutte le informazioni di contorno che lo sperimentatore riteneva utili per il suo studio (età del paziente, sesso, abitudini alimentari, fumo, patologie concomitanti...). Le eventuali difficoltà si risolvevano con l'usuale buon senso statistico ed operando a differenti livelli di analisi (AA.VV.2006). Rimaneva però saldo l'assunto alla base della statistica classica: il numero di osservazioni indipendenti (unità statistiche, individui, casi) era comunque superiore a quello delle variabili (Benigni R. e Giuliani A., 1994).

Questo assunto è di importanza vitale, la statistica inferenziale, quella che ci consente di aggiungere quel tanto (troppo) idolatrato p < 0.05 ai nostri risultati aprendo la strada della pubblicazione, deriva direttamente dal controllo di qualità e a questa origine dobbiamo per un momento andare se vogliamo capire a fondo l'entità delle nuove problematiche introdotte dall'uso delle tecniche di *Microarray* (Allison D.B. et al. 2006, Berrar D.P. et al., 2003).

Student, quello del famoso t, è lo pseudonimo usato dallo statistico inglese William Gossett che più di cento anni fa lavorava alla fabbrica di birra Guinness a Dublino. Egli si era applicato a questo problema pratico: perché la birra mantenesse le sue qualità organolettiche il più possibile intatte era importantissimo che la concentrazione di malto non si allontanasse 'trop-





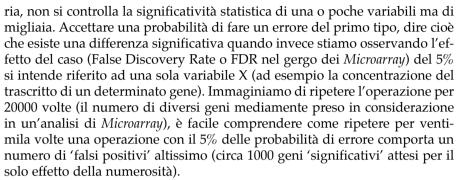
po' da quella prefissata. Per controllare che tutto andasse per il meglio, un campione costituito da un certo numero n di bottiglie veniva prelevato da ogni lotto e la concentrazione relativa di malto saggiata per ogni bottiglia. Se il saggio forniva dei risultati soddisfacenti (la concentrazione non era significativamente diversa dall'atteso) allora la produzione continuava senza interruzioni, se invece si riscontravano anomalie 'significative' si doveva dare l'ordine di interrompere tutto per andare a controllare che cosa andasse storto e rimettere le cose a posto.

Gosset aveva di fronte un problema classico della ricerca operativa: la ricerca del compromesso migliore tra due esigenze opposte: fermare troppo spesso la catena di produzione con falsi allarmi avrebbe comportato dei costi eccessivi per l'azienda, fermarla troppo di rado sarebbe ugualmente stato deleterio in quanto il prodotto Guinness sarebbe stato troppo spesso non all'altezza del nome con evidenti danni d'immagine (e alla lunga anche economici).

Per arrivare al migliore compromesso fra queste due esigenze Gosset costruì empiricamente una carta di controllo in cui andava a registrare quante volte una certa concentrazione osservata di malto corrispondesse ad un effettivo problema di produzione (casi positivi) e quante volte invece si era trattato di un falso allarme (casi negativi). Nel primo caso fermare la produzione era la scelta giusta da fare, nel secondo era meglio lasciar perdere: non fermare la produzione quando invece sarebbe stato necessario è quello che noi chiamiamo 'errore del primo tipo' che nelle scienze sperimentali corrisponde a 'accettare l'ipotesi nulla quando è falsa' (cioè a dire che il nostro risultato non è significativo quando invece un effetto c'è), fermare la catena quando non sarebbe stato necessario (falsi allarmi) è invece l'errore di secondo tipo e corrisponde a giudicare significativa un'evidenza quando è invece il frutto del caso. Chiaramente i due errori sono collegati, per cui dei criteri troppo rigidi aumenteranno l'errore del primo tipo e diminuiranno quello del secondo tipo, mentre l'opposto accadrà per criteri troppo laschi.

Lo scienziato inglese scoprì che accettare un errore del primo di tipo con una probabilità del 5% (p < 0.05) era il miglior compromesso fra i due rischi contrapposti, migliore nel senso che la sua azienda perdeva meno soldi possibile. Questa soglia era chiaramente ottimale per il particolare caso in questione, era un risultato sperimentale, non esisteva nessun motivo necessario per cui la soglia del 5% dovesse risultare ottimale anche in altri ambiti. Ma, essendo un risultato elegante e già pronto per l'uso, la scienza non ha fatto altro che appropriarsi di questo 'ottimo aziendale' trasformandolo in una sorta di costante universale come la carica dell'elettrone o la velocità della luce. Non dobbiamo quindi stupirci più di tanto se in molti casi l' approccio inferenziale classico arriva a degli esiti paradossali. Nel caso dei *Microarray* il paradosso deriva essenzialmente dal fatto che, a differenza della formulazione origina-





Ecco allora spiegato perché la statistica inferenziale classica non può essere applicata tal quale su sperimentazioni coinvolgenti grandissimi numeri di variabili come nei *Microarray*. Semplicemente siamo fuori dal dominio di plausibilità della metodica come il dr. Gosset l'aveva pensata per la sua birra, è necessario abbandonare un uso acritico e strumentale della statistica (ricetta 'consensus' applicata automaticamente) per una compenetrazione dello strumento d'analisi con le domande che ci poniamo e la concezione del mondo con cui guardiamo ai fenomeni studiati. Insomma, ci piaccia o no, l'analisi dei dati dei *Microarray* costringe la biologia ad assumere uno sguardo autenticamente quantitativo e non semplicemente limitarsi a qualche calcolo eseguito alla fine della sperimentazione per andare a vedere cosa è successo (l'esperimento è venuto bene / non è venuto bene) (Metha T. et al., 2004).

Nell' acquisizione di questo sguardo critico possiamo assumere due atteggiamenti principali, uno di tipo 'pessimista' ed uno che possiamo definire 'ottimista' o 'speranzoso'.

La via del pessimista (la maledizione della dimensionalità)

L'atteggiamento più conservatore, nei confronti della 'sfida quantitativa' posta dai microarray all'usuale modo di ragionare delle scienze biomediche, è quello di cercare in ogni caso di ottenere le informazioni che si sono da sempre ritenute rilevanti eliminando il disturbo della molteplicità delle variabili (Metha T. et al., 2004). Ciò corrisponde a individuare quali geni hanno significativamente alterato il loro livello di espressione a seguito della modifica di condizione imposta, od osservata, sia esso un trattamento farmacologico, una mutazione, uno stato patologico, riducendo al minimo il confondimento



dovuto all'elevato numero di variabili. In questo caso la numerosità estrema di geni di un chip di Microarray completo (attorno a ventimila) è veramente una maledizione (il termine 'curse of dimensionality', la maledizione della dimensionalità è di uso abbastanza frequente nella letteratura specialistica). La maledizione deriva dal carattere probabilistico dell'informazione ricavabile dai test inferenziali che rende elevatissimo il rischio di falsi positivi. Il primo atteggiamento utilizzato dalla comunità scientifica di fronte a questo scoglio fu di netto rifiuto di ogni elaborazione statistica: si decise di utilizzare esclusivamente il cosiddetto Fold Change (FC), in altre parole senza alcuna considerazione legata alla significatività statistica si accettavano come 'buoni' i geni che 'almeno raddoppiassero' (o decuplicassero a seconda delle tecniche di normalizzazione adottate) l'espressione passando da un gruppo sperimentale all'altro. Il sistema sembrò all'inizio molto efficiente in quanto di estrema semplicità, dopo poco tempo però ci si dovette ricredere: i risultati non erano per nulla ripetibili, i geni 'buoni' di un esperimento erano diversi da quelli 'buoni' in una sua replica ed il raggiungimento o meno di una soglia critica dipendeva fortemente dalla variabilità e numerosità del campione. Si era insomma barattata una 'probabilità d'errore nota' con una probabilità d'errore variabile, un pessimo affare a ben vedere!

Il passaggio successivo di questo attacco frontale contro la dimensionalità fu l'utilizzo di strategie di analisi di test 'alla Bonferroni'. In altre parole, seguendo le indicazioni di un metodo classico di valutazione dei test multipli, risalente allo statistico italiano Carlo Emilio Bonferroni che lo aveva proposto con grande successo negli anni '30 del secolo scorso, nell'ambito delle assicurazioni. Il livello di significatività 'accettato' in un test si sarebbe dovuto dividere per il numero di confronti eseguiti: misurando contemporaneamente l'espressione di 10 geni, sarà accettato come statisticamente significativo ogni risultato corrispondente ad una 'p' di 0.05/10 = 0.005, per 100 sarà significativo il livello pari a 0.05/100 = 0.0005 e via di seguito. Questa procedura è chiaramente efficace nel limitare il numero di falsi positivi, ma lascia piuttosto scontenti i biologi che riescono a 'vedere' solo gli effetti estremi di un trattamento che può scartare informazioni potenzialmente interessanti. Oltre a ciò c'è da considerare il fatto che, alla base delle strategie 'alla Bonferroni', è l'assunzione che i test siano tutti indipendenti fra loro e quindi lo siano anche i geni collegati. Immaginare una regolazione cellulare in cui ogni gene 'fa gioco a sé', senza alcuna relazione con gli altri geni, non è chiaramente biologicamente plausibile laddove insito nel concetto stesso di organismo è l'azione simultanea e coordinata di molti elementi (geni, proteine, cellule, tessuti, ecc..) (Alter O. et al., 2009).

A questo punto la sfida alla dimensionalità ha preso una strada meno 'automatica' ed ha cercato di 'aggirare l'ostacolo' piuttosto che 'prenderlo di petto'. Le strategie cosiddette GCT (gene-class testing) possono essere



considerate una risposta 'più articolata' alla sfida. Alla base di queste strategie, tuttora molto in voga ed utilizzate nella ricerca, è l'idea di fare un uso esplicito della conoscenza *a priori* dei ricercatori per la valutazione statistica dei risultati. Secondo le strategie GCT il raggiungimento di una significatività statistica, per quel che riguarda un particolare gene, è solo un primo passaggio dell'analisi. La, di solito lunga, lista di geni, 'significativamente diversi in termini di espressione' tra i gruppi studiati, è ulteriormente analizzata in termini di 'arricchimento differenziale' in certe 'classi di geni', dove per classi di geni si intendono famiglie di geni legati ad una certa funzione fisiologica: sistema immunitario, metabolismo, riparazione del DNA, apoptosi, ecc.... Le classi di geni si basano sulle cosiddette categorie GO (Gene Ontology), che forniscono una complessa stratificazione gerarchica: da classi molto generali a funzioni molto specifiche con cui catalogare i differenti geni (Camon E. et al., 2003).

A questo punto il 'soggetto' dell'analisi non è più il singolo gene, ma 'l'intera lista di geni significativi' di cui viene studiata la distribuzione in termini di particolari categorie GO. Così, ad esempio, se la lista risulta 'significativamente arricchita' di geni legati al funzionamento del sistema immunitario rispetto a quanto atteso per puro effetto del caso, possiamo affermare che il particolare trattamento 'coinvolge il sistema immunitario'.

Il 'carattere polisemico' dell' informazione biomedica rende questo approccio piuttosto difficoltoso; infatti, uno stesso gene può essere classificato in molte differenti classi di attività biologica. Inoltre, potremmo immaginare di operare classificazioni alternative prendendo in considerazione la sua localizzazione cellulare o addirittura la sua omologia di sequenza con altri geni. La risposta alla domanda generica, il gene X è più simile al gene Y o al gene Z? è quindi, quasi invariabilmente, un insoddisfacente 'dipende': dipende dal livello di descrizione che stiamo considerando, da che cosa ci interessa dimostrare, a quanti diversi aspetti della caratterizzazione dei geni vogliamo prendere in considerazione. Chiaramente la presenza di due geni nella stessa famiglia, a tutti i possibili livelli di rappresentazione, è indice di interscambiabilità completa dei due geni in questione, come avviene per le isoforme di uno stesso gene. Una loro completa mancanza di relazione, a tutte le scale, è ugualmente indice di completa separazione, ma questo avviene assai di rado, essendo il caso di gran lunga più comune un mutevole grado di sovrapposizione fra una qualsiasi coppia di geni, a seconda dei livelli ontologici presi in esame.

Il particolare carattere conservatore delle scienze biomediche, rispetto alla fisica o alla chimica, che spesso appare insopportabilmente pedante ai cultori di altre scienze più matematizzate, deriva proprio dalla delicatezza e dipendenza dal contesto dei concetti espressi, a tutte quelle sfumature di significato che non sono percepibili a chi non è del campo. Questo è un punto



importantissimo ed ha a che vedere con le trappole insite negli approcci GCT. Di fatto scegliendo un'ontologia ad un certo livello di definizione ed assumendo il criterio di 'plausibilità biologica' oltre che statistica della soluzione (vengono accettate solo quelle soluzioni che mostrano un senso da un punto di vista funzionale) il ricercatore intreccia fortemente il piano dell'analisi dei dati con quello della sua personale interpretazione dei fatti. Il nome generico di questo atteggiamento epistemologico è 'approccio Bayesiano', dal nome del reverendo e statistico inglese Thomas Bayes (1702-1761) che propose una definizione di probabilità comprendente anche le 'aspettative precedenti all'esperimento'. In altre parole, il peso che una determinata osservazione acquisisce viene a dipendere non solo dall'osservazione in sé, ma anche dalla conoscenza pregressa secondo la formula:

$$P(A \land B) = P(A/B) / P(B)$$
 (formula 1)

Nella formula 1 P (A ^ B) indica il verificarsi dell'evento A insieme all' evento B, P(A/B) è la probabilità del verificarsi di A se B è già noto essere avvenuto, e P(B) è la probabilità a priori di B. Senza entrare nel dettaglio della formula possiamo immaginare come la probabilità di pioggia (A) sia modificata dal sapere che la giornata è molto nuvolosa (B) rispetto al non sapere nulla di cosa avviene fuori dalla finestra. I due eventi A (pioggia) e B (nuvolosità) sono infatti collegati tra di loro e scambiano informazione tra di loro. Nel caso dei *Microarray*, un risultato significativo 'corroborato' da altri risultati significativi in geni 'equiparabili in quanto a funzione' vale più di un risultato significativo in un gene 'isolato'. Ovviamente, a differenza del semplice caso pioggia/nuvole, l'enorme molteplicità classificatoria raggiungibile da significati ricchi e sfaccettati come quelli biologici è molto rischiosa e provoca un insidioso riproporsi della maledizione della dimensionalità (Crescenzi M. e Giuliani A., 2001).

Immaginiamo di avere a disposizione un software che cerchi, all'interno degli alberi di classificazione multipla di geni, la combinazione di *pathways* noti, che meglio corrisponda con la lista di geni riconosciuti come statisticamente significativi in un esperimento di *Microarray*. Un *pathway* è la rappresentazione di un processo biologico con l'usuale grafica a 'scatole-e-linee' dove le scatole indicano i geni e le linee dei 'legami di relazione' tra i geni. Questi legami di relazione sono frasi codificate del tipo 'il gene A attiva il gene B' o 'il gene B interviene sul prodotto del gene A' e così via. In figura 1 è schematizzato parte del pathway dell'apoptosi, o morte cellulare programmata, come rappresentato da KEGG (Kyoto Enciclopedia of Genes and Genomes), consultabile al sito http://www.genome.ad.jp/kegg/).

Navigando nel web, si possono trovare migliaia di questi schemi riferiti a diversi fenomeni fisiologici che costituiscono l'informazione *a priori* con cui





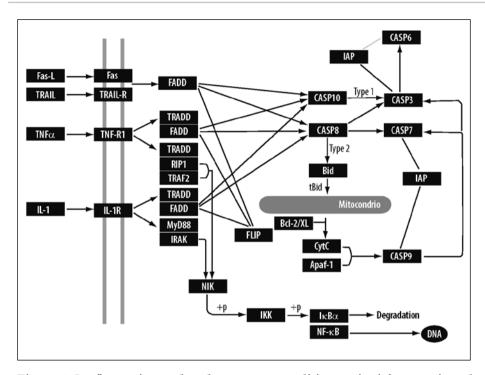


Figura 1. La figura riporta lo schema presunto di interazioni fra proteine ed altri costituenti cellulari che porta al fenomeno dell'apoptosi o morte cellulare programmata. Lo schema deriva dal grande data base KEGG (Kyoto Encyclopedia of Gene and Genomics) consultabile al sito: http://www.genome.ad.jp/kegg/.

In questo sito si trovano migliaia di processi cellulari descritti in termini di nodi (di volta in volta geni, proteine, metaboliti) e frecce rappresentanti le interazioni fra questi nodi. Questi schemi derivano dall'analisi critica di grandi masse di letteratura e in qualche modo rappresentano la 'Biblioteca di Alessandria' dell'odierna biologia molecolare, ogni passaggio descritto è l'oggetto di uno o più lavori scientifici, l'integrazione in schemi unici avviene per pura combinazione lineare.

andiamo ad analizzare in maniera Bayesiana i dati dei microarray. A questo punto il problema di 'quale classificazione scegliere' è a sua volta un problema risolubile in maniera algoritmica: semplicemente si esplorano le basi di dati che raccolgono i diversi pathway cercando le combinazioni di questi pathway, che si accordano meglio con i geni effettivamente riconosciuti come modificati in termini di espressione(Kitano H., 2002).



La molteplicità praticamente infinita di combinazioni, derivate dall'informazione scientifica accumulata negli anni e per di più tutte plausibili, fa sì che praticamente ogni analisi di questo tipo fornisca una risposta 'ragionevole in termini biologici'. Ciò rischia di rendere l'esperimento di *Microarray* una noiosa ripetizione del già noto, una stucchevole esercitazione in cui ci si bea di ritrovare 'in bella forma' quello che già si sospettava esserci, siamo tutti contenti, si pubblica un articolo e via così, nel più puro manierismo, senza aumentare di nulla la nostra conoscenza sul mondo anzi, al contrario, rendendo un pochino più difficile scalzare quanto già si credeva noto con nuove idee e paradigmi.

Una certa disaffezione degli ambienti di ricerca più accorti nei confronti dei microarray deriva proprio da questa esagerata plasticità, permessa dagli strumenti della bioinformatica, che si limitano a produrre complicate reti di regolazione e danno una certa impressione di scienza 'avanzata', ma che non tardano a mostrare la loro futilità. Peccato, perché le premesse di questa via di 'venire a patti con la molteplicità' presenta degli aspetti saggi e promettenti. Vediamo allora se è possibile andare ulteriormente avanti.

La via dell'ottimista (la benedizione della dimensionalità)

Una domanda che si pone spontanea: "Ma possibile mai che avere tante informazioni sia peggio che averne poche?" La cosa appare a prima vista piuttosto singolare ed, infatti se ci pensiamo bene, i guai non sono tanto legati al fatto di avere tante informazioni, di per sé, quanto all'uso e al valore che noi assegniamo all'insieme delle nostre misure. Nel caso del modello riduzionista classico la cellula è paragonabile ad una enorme pulsantiera del genere di quelle presenti nei moderni jet di linea in cui i geni corrispondono alle singole manopole. Ogni gene 'statisticamente significativo' rappresenta una manopola portata su un livello differente nella cellula sottoposta ad una certa condizione sperimentale, quindi la molteplicità diventa problematica. In questo modello i geni sono 'quelli che fanno avvenire le cose' in quanto entità autonome, attori del gioco, ed è quindi di fondamentale importanza conoscere con la massima precisione possibile nome e cognome dei geni effettivamente coinvolti nelle particolari condizioni sperimentali utilizzate. La totalità, o quasi, dei ricercatori negherà decisamente di avere un tale modello dei sistemi viventi, che sicuramente giudica incredibilmente rozzo e poco realistico, ma quando bramano di avere 'la lista dei geni significativi', come base d'appoggio per le loro elucubrazioni, di fatto utilizzano impli-



citamente il modello a pulsantiera di jet. L'addolcimento di questa visione attraverso l'uso della conoscenza *a priori*, che abbiamo discusso nel paragrafo precedente, non è altro che lo sforzo di riportare i risultati ottenuti ad una sorta di "pulsantiera condivisa", ma sempre di pulsantiera si tratta.

In questi ultimi anni, il termine 'Systems Biology' si sta velocemente diffondendo nel campo delle scienze biomediche con il corollario di riviste dedicate, istituti di ricerca e carriere accademiche. L'idea di fondo, a ben vedere, non è affatto nuova ed è quella di considerare gli organismi per quello che sono: unità funzionali in cui le parti costituenti collaborano per costruire qualcosa di più e di diverso dalla somma delle sue parti. Il vecchio termine di 'fisiologia' indicava appunto questa consapevolezza e l'enfasi era posta, in opposizione all'anatomia che si occupava dell'aspetto strutturale di queste relazioni, sulle mutue regolazioni dinamiche tra i tessuti.

Ciò che differenzia la *Systems Biology'* dalla fisiologia è la mole di dati su cui si basa, la disponibilità di una grande potenza di calcolo a basso costo, fornita dalla rivoluzione informatica, ed il poter portare l'approccio regolativo a livello molecolare. L'apparato metodologico della *Systems Biology'* complementa i metodi classici della statistica con la cultura dell'analisi dei segnali e dei servomeccanismi di derivazione ingegneristica. Ciò fa sì che l'enfasi si sposti dai geni presi singolarmente ai flussi di correlazione tra livelli di espressione di geni diversi (Kitano H,. 2002).

Prima di andare a vedere i geni 'significativamente interessati' all'effetto di una mutazione o di un trattamento farmacologico sarebbe necessario soffermarsi sulla struttura di correlazione dei geni ed individuare i flussi di attivazioni comuni. Indipendentemente da qualsiasi considerazione sul trattamento dovremmo prima di tutto conoscere come i geni si 'muovono' di concerto, per cui ci interesserà scoprire se ad una sovraespressione del gene A corrisponde anche una elevata espressione del gene B e come si comporta il gene C.

La presenza di gruppi di geni co-regolati è implicita nella considerazione della cellula come un sistema, cioè non una semplice congerie di geni indipendenti, la pulsantiera, ma un' entità vivente. Perché questa affermazione passi da una semplice posizione di principio ad una prassi di ricerca abbiamo, quindi, per prima cosa bisogno di una metrica che ci consenta di apprezzare l'entità dell'accoppiamento fra geni e quindi l'intensità del loro legame funzionale. La metrica più diffusa è sicuramente il coefficiente di correlazione di Pearson, il coefficiente r, sviluppato dallo statistico inglese Karl Pearson ai primi del '900. Possiamo descrivere graficamente il coefficiente r utilizzando la Figura 2 dove gli assi rappresentano le variabili studiate (ad esempio due geni) ed i punti i campioni sperimentali. I pannelli a sinistra ed a destra indicano due situazioni di massima correlazione (rispettivamente negativa e positiva) di due geni, il pannello centrale la situazione di massimo disaccoppiamento (le fluttuazioni del gene r000 sono indipendenti da quelle del gene r100 seniore di massimo disaccoppiamento (le fluttuazioni del gene r100 sono indipendenti da quelle del gene r100 seniore di massimo disaccoppiamento (le fluttuazioni del gene r210 sono indipendenti da quelle del gene r210 seniore di massimo disaccoppiamento (le fluttuazioni del gene r210 seniore di massimo disaccoppiamento (le fluttuazioni del gene r210 seniore di massimo disaccoppiamento (le fluttuazioni del gene r210 seniore di massimo disaccoppiamento (le fluttuazioni del gene r31 seniore di massimo disaccoppiamento (le fluttuazioni del gene r31 seniore di massimo disaccoppiamento (le fluttuazioni del gene r31 seniore di massimo disaccoppiamento (le fluttuazioni del gene r31 seniore di massimo disaccoppiamento (le fluttuazioni del gene r31 seniore di massimo disaccoppiamento (le fluttuazioni del gene r31 seniore di massimo disaccoppiamento (le fluttuazioni del gene r31 seniore di massimo di massimo di massimo di



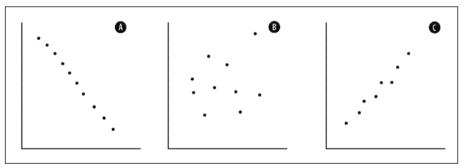


Figura 2. I tre 'casi estremi' di correlazione tra due variabili sono riportati nei tre pannelli di questa figura in cui gli assi rappresentano le due variabili di cui si studia la relazione reciproca ed i punti sono le unità statistiche su cui sono state eseguite le osservazioni. Ogni unità statistica ha quindi una collocazione nello spazio corrispondente alle sue coordinate (valori) nella variabile X (ascissa) e nella variabile Y (ordinata).l pannello a) mostra la massima correlazione negativa tra le variabili in cui a valori elevati della X corrispondono valori bassi della Y e viceversa. Il valore del coefficiente r di Pearson è in questo caso uguale a -1.

Il pannello b) mostra l'assenza di relazione tra X ed Y, la conoscenza di X non comporta nessuna conoscenza aggiuntiva sulla Y, le due variabili non hanno alcuna relazione lineare, il valore di r è pari a 0.

Il pannello c) mostra la massima correlazione positiva fra le variabili X ed Y, valori elevati di X corrispondono a valori elevati di Y e così per i valori bassi. Il coefficiente r è in questo caso pari a +1.

Le due condizioni a) e c) sono entrambe indice di una forte relazione fra le due variabili (ad esempio due geni inseriti nello stesso circuito di regolazione), mentre la situazione b) mostra due variabili indipendenti.

Nel caso **a** la correlazione è negativa: a valori grandi della X (a destra nel grafico) corrispondono valori piccoli della Y (in basso nel grafico), la situazione **c** invece corrisponde ad una correlazione positiva: valori grandi della X vanno insieme a valori grandi della Y e viceversa. Entrambe le situazioni sono perfettamente compressibili, il piano (due dimensioni) e' cioè riconducibile ad una retta (una sola dimensione).

Il caso **b** rappresenta invece l'assenza di correlazione: conoscere la posizione di un punto sull' asse X non e' di nessun aiuto per diminuire l'incertezza riguardo alla sua posizione.

Una buona misura di correlazione dovrà allora riconoscere come massimi (di segno opposto) le situazioni $\bf a$ e $\bf c$ ed avere come minimo la situazione $\bf b$, il coefficiente $\bf r$ ottiene questo risultato costruendo una metrica continua di correlazione con due massimi $\bf a$ -1 (massima correlazione negativa, caso $\bf a$) e +1 (massima correlazione positiva, caso $\bf c$) ed un minimo a $\bf r$ = 0 (assenza



di correlazione, caso **b**). Risulta così stabilita una metrica che consente di apprezzare il grado di correlazione di due variabili e di giudicare quanto forte sia il legame che le unisce.

Ciò dimostra come il coefficiente di correlazione ci permetta di stabilire una metrica intuitiva che misuri il grado di interrelazione fra due geni.
Immaginiamo ora di allargare il discorso a tutte le possibili coppie di geni
presenti in un esperimento di *Microarray*: otterremo una matrice detta matrice di correlazione avente come righe e colonne tutti i geni presi in considerazione e, all'incrocio tra la riga i e la colonna j il valore del coefficiente r per
la coppia i, j (i valori della diagonale principale di questa matrice avranno
valori unitari e la matrice sarà triangolare simmetrica cioè la posizione i, j sarà
identica alla posizione j,i).

Questa matrice può essere immaginata (in primissima approssimazione) come una rappresentazione della 'rete di relazioni' intercorrente fra i differenti geni del sistema come descritto nella figura 3, dove per semplicità abbiamo immaginato elementi legati da correlazioni massime (1) o nulle (0).

La matrice A i cui elementi sono le correlazioni tra l'elemento di riga ed il corrispondente elemento di colonna, è un'immagine fedele del grafico a

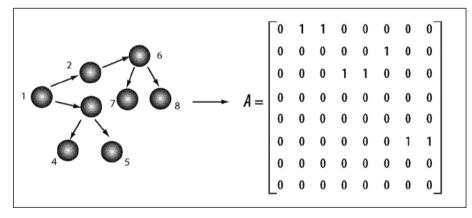


Figura 3. A sinistra è riportato un albero (una particolare forma di grafo) che esplicita le relazioni fra gli elementi (nodi) di un determinato sistema. Potrebbe essere indifferentemente l'organizzazione gerarchica di un ufficio così come una cascata enzimatica o uno schema di distribuzione di elettricità. La stessa informazione è riportata in forma matriciale come matrice di adiacenza (A) sulla destra. Questo tipo di matrici ha solo due valori possibili: 0 per assenza di relazioni dirette fra i nodi corrispondenti alla posizione di riga e di colonna della matrice A, ed 1 nel caso in cui fra i due nodi sussista un legame diretto. La matrice A consente di operare delle statistiche e comunque delle operazioni matematiche sui grafi (o reti che dir si voglia).



sinistra dove i cerchi corrispondono ai geni e le frecce ai loro legami funzionali, queste immagini sono comuni nei libri e negli articoli di biologia. Qui abbiamo immaginato legami unidirezionali, nel caso della correlazione avremo legami in due direzioni ed in più ogni nodo, gene, è collegato con sé stesso: il punto da tenere a mente è che una metrica di correlazione è, come dicono i matematici, 'isomorfa' ad una rete di relazione fra gli elementi. Si può cioè passare da una rappresentazioni a nodi e frecce (molto in voga nelle scienze biomediche) ad una matrice (utile per fare calcoli) senza perdere informazione. In figura 4 è rappresentato come apparirebbe la rete in un caso più realistico, dove l'entità delle correlazioni corrisponde all'intensità dello spigolo che lega i nodi.

Nel caso in cui i nodi, geni, presi in considerazione sono pochi una rappresentazione come quella della figura 4 è più che sufficiente per dare un'idea generale della struttura del sistema analizzato. Invece, nel caso che i nodi sono molti, una descrizione di questo tipo è praticamente inutilizzabile, come si può intuire dalla figura 5, dove è rappresentata una frazione del 'metaboloma' del lievito cioè l'insieme delle reazioni biochimiche che avvengono nella cellula di lievito. I nodi sono i metaboliti e gli spigoli indicano la presenza di un enzima che trasforma il metabolita a nel metabolita b. Una tale rappresentazione rende possibile fare delle inferenze generali, ad esempio individuare se le mutazioni letali si situano in particolari zone del grafico (Palombo M.C. et al., 2005).

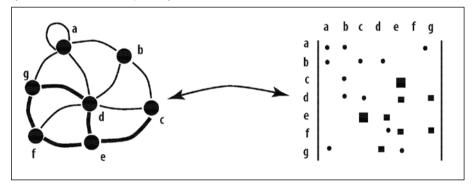


Figura 4. La figura riporta il caso più generale di rappresentazione matriciale di una rete. A differenza del grafo riportato nella figura precedente, qui i legami fra i nodi non sono tutti uguali ma hanno una diversa intensità, sono cioè dei legami più forti o più deboli tra le variabili (nodi). Questa gradazione di intensità è riportata nel pannello di sinistra con linee di diverso spessore e nella forma matriciale a destra con simboli di diverse dimensioni. Con questo tipo di formalismo la matrice di adiacenza diventa una matrice di correlazione, l'intensità del legame fra i nodi non è infatti niente di diverso dal coefficiente di correlazione fra le variabili (nodi) corrispondenti.







Figura 5. Una frazione del metaboloma di lievito. Ogni nodo corrisponde ad un determinato metabolita, cioè ad una molecola organica prodotta e/o utilizzata dal metabolismo della cellula del lievito, la presenza di una freccia (arco, nel linguaggio della teoria dei grafi) tra due metaboliti indica la presenza nella cellula di un enzima che trasforma la molecola corrispondente al nodo di arrivo in quella corrispondente al nodo di partenza. Alcuni nodi hanno associata una cifra che corrisponde ad un indice di un elenco che ne riporta la formula di struttura, i nodi di differente colore indicano la posizione di mutazioni essenziali, in cui il metabolita in blu deve essere fornito come nutriente al lievito mutante di uno o più enzimi situati nella sua via di sintesi per garantire la sopravvivenza del lievito stesso. La descrizione ed il coomento di questa figura, così come il modo di ricavare la relativa importanza di un enzima dalla topologia della rete metabolica (metaboloma) del lievito è riportata in: Palumbo M.C., Colosimo A., Giuliani A., Farina L. (2005) Functional Essentiality from Topology features in metabolic networks: a case study in yeast. FEBS Letters. (579): 4642-4646.

Non è possibile individuare i così detti 'moduli', ossia i gruppi di metaboliti geni (come nel caso dei microarray) correlati fra di loro e formanti delle unità funzionali.



Individuare questi moduli è di importanza cruciale per l'interpretazione dei dati. Ma perché non utilizzare direttamente le ontologie come abbiamo visto nel paragrafo precedente? Semplicemente perché, nel caso dei geni, le alleanze sono variabili ed il tipo di correlazione presente in una certa situazione non è identica a quella che si può verificare in un'altra. La plasticità del sistema di regolazione, fa sì che ogni particolare situazione esalti certi legami e ne deprima altri, un po' come succede al nostro cervello (non a caso si parla di reti neurali) quando deve processare uno stimolo (Feng X.J. et al., 2004).

Il metodo principe per 'raggruppare' le relazioni e quindi individuare i geni che formano delle unità funzionali co-regolate è l'analisi in componenti principali (PCA = Principal Component Analysis).

Senza entrare in dettagli matematici, che implicherebbero una certa conoscenza dell'algebra lineare, rimandiamo alla figura 6. Nel pannello di sinistra vediamo uno spazio avente come assi le variabili V1 e V2 tra le quali esiste una forte relazione. Questa forte relazione fa sì che sia possibile riassumere l'informazione legata alla posizione di ogni punto nello spazio inizialmente dispersa su due coordinate (V1 e V2) in un unico indice chiamato PC1 (prima

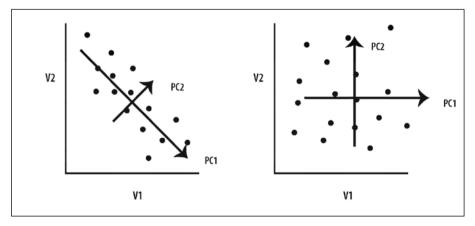


Figura 6. Nel pannello di sinistra notiamo come la forte correlazione fra le variabili V1 e V2 consente di riassumere l'informazione congiunta delle due variabili in una sola componente principale (PC1) funzione di V1 e V2, la seconda componente principale (PC2) è in questo caso molto meno importante della prima e riporta la 'porzione discordante di informazione' fra le due variabili originali. Nel pannello di sinistra invece, l'assenza di correlazione fra V1 e V2 fa sì che non sia possibile 'riassumere' l'informazione originariamente connessa alle due variabili in un indice sintetico: PC1 e PC2 sono entrambe sullo stesso piano di contenuto informativo, il sistema si dice in questo caso 'non comprimibile'.



componente principale), combinazione lineare di V1 e V2 secondo una formula del tipo: PC1 a*V1 + b*V2 (con a e b costanti reali).

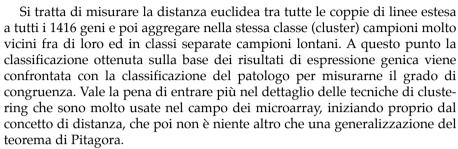
Chiaramente il riassunto non sarà perfetto, perderemo una parte di informazione che sarà associata alla seconda componente principale (PC2) ortogonale (e quindi indipendente) a PC1. Nel pannello di sinistra considerando solo PC1 si perde poca informazione e quindi il riassunto (o meglio la compressione) operato dalle componenti principali è efficiente. Nel pannello di destra, ove non c'è nessuna correlazione fra V1 e V2, le due componenti hanno invece più o meno la stessa importanza, quindi il sistema non è utilmente compressibile.

Immaginiamo di applicare questo ragionamento in un caso con migliaia di variabili; è intuibile che una compressione di circa 1400 variabili (geni) in uno spazio di 15 componenti principali, mantenendo il 65% dell'informazione iniziale, risulta un ottimo dal punto di vista di interpretazione dei dati sperimentali. Semplicemente, dovremmo gestire 15 variabili invece di 1400, in più sappiamo che queste variabili costituiscono la parte 'robusta' dell'informazione, quella cioè confermata dai valori coerenti di molti geni e quindi utilmente analizzabile in termini funzionali.

Sarebbe così risolto velocemente il problema della molteplicità dei test analizziamo ora nel dettaglio un caso specifico di analisi dei dati di *Microarray* con questo tipo di strategia che fa un uso costruttivo dell' alta dimensionalità sfruttando le correlazioni fra geni. Qui abbiamo come unità statistiche 60 linee cellulari derivanti da altrettanti tumori appartenenti a 9 classi differenti (seno, sistema nervoso centrale, colon, polmone, leucemia, melanoma, ovaio, prostata, rene) e come variabili i valori di espressione di 1416 geni. Scopo di questa analisi è quello di giungere ad una classificazione il più possibile accurata dei tumori attraverso i loro differenti profili di espressione genica.

Chiaramente qui non si cerca di costruire uno strumento diagnostico (le classiche tecniche di patologia servono benissimo allo scopo) ma piuttosto di scoprire una sorta di 'caratterizzazione funzionale' dei diversi tumori capace di offrire informazioni rilevanti sulla loro biologia e quindi, indirettamente, su eventuali trattamenti farmacologici. La molteplicità delle classi impedisce qualsiasi tentativo di statistica inferenziale gene-per-gene che in ogni caso sarebbe stata di difficile interpretazione per la difficoltà nel confrontare le varie classi di tumori., quindi, in prima istanza ci si è posti questo quesito minimo: 'indipendentemente da quali siano i geni importanti per la classificazione, le linee derivanti dalla stessa classe di tumori sono più simili fra di loro rispetto a linee provenienti da tipi tumorali differenti?' La risposta a questa domanda è stata affermativa ed è stata data dai ricercatori utilizzando una tecnica di Cluster Analysis (Ross D.T. et al., 2000).





Nella figura 7 vediamo il significato geometrico della distanza tra due punti A e B (d(ab)) in uno spazio bidimensionale (nel nostro caso A e B sono due linee tumorali, i due assi X ed Y due geni e le coordinate di A e B su X ed Y i livelli di espressione relativi alle due linee tumorali).

L'applicazione del teorema di Pitagora ci porta ad una semplice formula per la distanza:

$$D(ab) = \sqrt{(X(a) - X(b))^2 + (Y(a) - Y(b))^2}$$
 (formula 2)

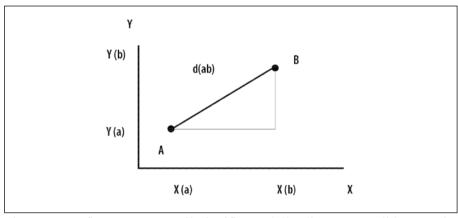


Figura 7. La figura 7 mostra il significato della distanza Euclidea tra due punti in uno spazio bidimensionale. La distanza tra i punti a) e b) corrisponde all'ipotenusa del triangolo avente cateti di lunghezza rispettivamente (X(b)-X(a)), e (Y(b)-Y(a)). In accordo con il teorema di Pitagora quindi, la distanza tra a e b (d(ab)) risulta essere: $d(ab) = \sqrt{(X(b)-(X(a))2+(Y(b)-Y(a))2)}$ (Eq.1) Se a) e b) sono due unità statistiche definite invece cha da due sole variabili da n variabili in uno spazio ad n dimensioni con n >> 2, potremo allo stesso modo calcolare la loro distanza reciproca semplicemente aggiungendo altri addendi all' eq.1 nella forma:

 $d(ab) = \sqrt{(X(b)-(X(a))^2 + (Y(b)-Y(a))^2 + (Z(b)-Z(a))^2} + (J(b)-J(a))^2 +$ (Eq.2)



A cura di Anna Di Lonardo,

D(ab) corrisponde all'ipotenusa del triangolo avente come cateti (X(a) -X(b)) e (Y(a) - Y(b)).

La formula (2) può essere generalizzata a quante dimensioni si vuole, semplicemente aggiungendo addendi sotto alla radice quadrata, nel nostro caso quindi la distanza tra due linee A e B sarà un numero proveniente da una somma di 1416 addendi e non solo da due come nella (formula 2).

Dopo aver calcolato le distanze per tutte le coppie distinte di linee dobbiamo risolvere il problema di costruire dei gruppi (cluster) di linee, di raggruppare cioè le nostre unità statistiche in classi formate da elementi simili (poco distanti) tra di loro.

Esistono molte strategie per risolvere questo problema che possono essere fatte ricadere in due grandi classi: a) metodi di clustering gerarchico e b) metodi di clustering non gerarchico.

I metodi di clustering gerarchico si ispirano alle classificazioni biologiche e partono dal presupposto che esistano differenti livelli di similitudine che appunto costituiscono una gerarchia del tipo specie, genere, classe, ecc... La matrice di distanze tra tutte le coppie di elementi viene quindi analizzata con vari approcci euristici.

Un approccio molto comune è quello detto 'average linkage' che si basa su questi principi:

- Al primo passaggio le due unità A e B più vicine (distanza minima) vengono fuse in un primo gruppo che viene considerato come una unità singola AB.
- La matrice di distanze viene ricalcolata su n-1 elementi (A e B ora contano solo per una unità AB e non per due come all'inizio), la distanza per ogni singola unità (i) dalla nuova unità fusa AB è la media delle distanze da A e da B: d(i,AB) = (d(i,A) + d(i,B)) / 2.

Si ripetono i passaggi precedenti fino a che tutte le unità risultino aggregate fra loro.

Chiaramente alla fine della procedura tutte le unità statistiche saranno state aggregate insieme, ma a livelli di distanza progressivamente maggiori corrispondenti alle diverse gerarchie di similitudine. L'intero processo di aggregazione tiene quindi traccia della classificazione in termini di livello di distanza a cui ogni unità si è aggregata alle altre. Questo processo è riassunto da un albero i cui rami hanno una lunghezza proporzionale al livello di aggregazione (distanza a cui avviene la fusione) come riportato nella figura 8, riferito alle 60 linee tumorali.

Osservando il livello a cui avviene l'aggregazione delle diverse linee (la cui distanza è stata calcolata sull'intero insieme di 1416 geni) possiamo individuare 6 cluster indicati sulla destra dalle cifre da 1 a 6. Il nome delle unità statistiche è riportato nella figura, le prime lettere individuano la classe dei tumori: BR = seno, CNS = sistema nervoso centrale, CO = colon, LE = leucemia, LC = polmone, ME = melanoma, OV = ovaio, PR = prostata, RE = rene.



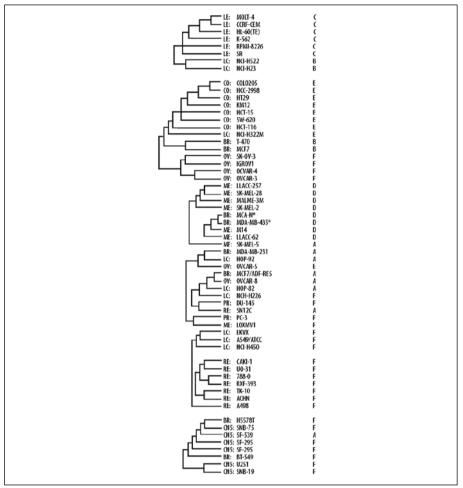


Figura 8. Sulla sinistra della figura è riportato il risultato dell'analisi gerarchica dei cluster eseguita su 60 diverse linee tumorali, sulla base delle rispettive distanze Euclidee calcolate sull'intero insieme di migliaia di geni. La lettera sulla destra riporta invece il gruppo di appartenenza di ogni linea generato da un'analisi dei cluster (questa volta non gerarchica) applicata sulle distanze tra linee questa volta però calcolata non su migliaia di dimensioni ma sulle sole cinque dimensioni corrispondenti alle prime componenti principali del campo di dati. Come si nota le due classificazioni sono molto simili nonostante la drastica riduzione di dimensionalità apportata dall'analisi in componenti principali. Questo esperimento è descritto in: M. Crescenzi and A. Giuliani (2001) The main biological determinants of tumor line taxonomy elucidated by a principal component analysis of microarray data. FEBS Letters (507) 1: 114-118.



Come si può notare le similitudini globali tra linee in termini di espressione genica si conservano; linee di uguale provenienza tendono ad andare insieme ma la corrispondenza fra classificazione patologica e genetica funzionale non è certamente stringente. Questa non perfetta concordanza deriva da una molteplicità di fattori che vanno dalla variabilità genetica delle linee cellulari stabilizzate in vitro rispetto fino alla eterogeneità delle popolazioni cellulari eterogenee presenti nella coltura primaria.

Passiamo ora a discutere delle strategie di clustering non gerarchico da cui derivano le lettere dalla A alla F disposte sulla destra della figura.

Discuteremo la procedura cosiddetta *k-means* che è di gran lunga la tecnica di clustering non gerarchico più utilizzata. In questo caso la procedura segue questo criterio:

'La migliore partizione di un insieme di n unità in k classi è quella che costruisce classi il più possibile compatte al loro interno e distanti fra loro', matematicamente questo corrisponde a cercare il minimo della misura:

C = (Varianza entro classi) / (Varianza tra classi) (formula3).

La cosa è piuttosto intuitiva, è quello che normalmente fa il nostro occhio quando individua dei gruppi su un piano: si cercano gli oggetti vicini fra loro e ben distinti dagli altri come nella figura 9.

La procedura k-means raggiunge un massimo del coefficiente di clustering (formula 3) attraverso una procedura di prove ed errori che parte da una prima scelta casuale di centri di aggregazione a cui le unità vengono asse-

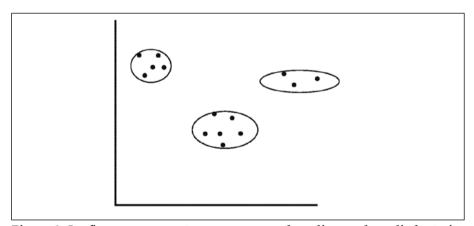


Figura 9. La figura rappresenta un caso esemplare di procedura di clustering non gerarchico (k-means): la clusterizzazione ottimale è quella che consente di avere delle classi (cluster) il più possibili compatte al loro interno (punti vicini fra loro) e separate tra di esse (cluster non sovrapposti).



gnate sulla base della minima distanza dai centri e poi prosegue riallocando le unità a nuovi centri questa volta non più casuali ma corrispondenti ai centroidi (medie nello spazio multivariato) dei gruppi successivamente formati, la procedura termina quando nessuna riallocazione varia il valore di C.

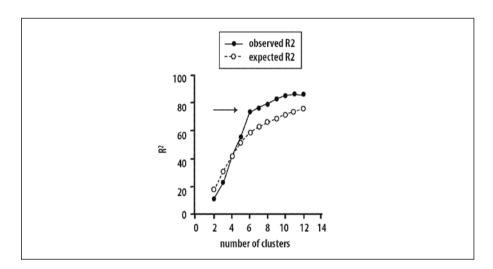


Figura 10. Il rapporto tra la varianza spiegata dal modello e la varianza totale corrisponde al quadrato del coefficiente di correlazione di Pearson (R2). Nel caso di una procedura di clustering la varianza spiegata dal modello è quella che si misura sostituendo, per ogni unità statistica, al valore effettivamente misurato, il valore corrispondente a quello della media della sua classe (cluster) di appartenenza, mentre la varianza totale è chiaramente quella effettivamente misurata come sommatoria degli scarti quadrati dei valori misurati dalla media. Al crescere del numero di classi la varianza spiegata dal modello tendere a coincidere in maniera meccanica con la varianza totale, questo si comprende in modo intuitivo se pensiamo che per un numero di classi uguale al numero di unità statistiche ogni classe avrà come valore medio il valore effettivo dell'unità statistica. Il criterio di decisione del numero ottimale di classi non può quindi essere la pura massimizzazione di R2, ma il numero di classi in cui il valore di R2 effettivamente osservato è il più differente possibile da quello atteso in una distribuzione casuale di valori, posto che si è arrivati ad una fase di saturazione dell'indice R2, che implica il raggiungimento della soglia in cui la sua crescita avviene per pura deriva meccanica. Questo è il senso della scelta del numero di classi di tumori nel caso dell'analisi dei microarray per le linee tumorali riportato in figura (Crescenzi and A. Giuliani (2001) The main biological determinants of tumor line taxonomy elucidated by a principal component analysis of microarray data. FEBS Letters (507) 1: 114-118).



Il caso riportato in figura 9 è estremamente semplice e mostra una evidente soluzione a K=3 (3 cluster presenti). Oltre a questo caso, l'insieme in studio potrebbe mostrare diversi valori di K che rendono ottimale la soluzione come ad esempio nel caso di una gerarchia livelli multipli. Si sceglieranno, quindi in questo caso, i valori di K da utilizzare per cui il valore K mostrerà un massimo locale o una netta separazione con quanto atteso per puro caso (coefficiente di clustering corrispondente ad un valore K, il quadrato del coefficiente di Pearson). Questa strategia è stata eseguita sul nostro insieme di dati (le K0 linee tumorali) ottenendo come ottimale ancora una volta una struttura a K1 cluster come si osserva in figura K2. Teniamo presente che i circoli vuoti corrispondono al clustering atteso per una distribuzione gaussiana, i circoli pieni il coefficiente di clustering osservato sui nostri dati. Come si vede dalla figura la soluzione a K2 è quella più parsimoniosa (minor numero di classi) e dove la distribuzione si allontana nettamente dal puro caso.

I sei cluster ottenuti con la procedura k-means sono quelli da A ad F riportati a fianco della classificazione gerarchica di figura 8. Come si può notare la sovrapposizione tra le procedure gerarchica e non gerarchica è molto forte. La cosa è ancora più rilevante se si pensa che la procedura k-means non si basa sull'intero insieme dei 1416 geni, preso come tale, ma filtrato attraverso le sue prime 5 componenti principali. In altre parole i cluster derivano dalle similitudini fra linee basate su solo 5 informazioni (valori distinti). L'informazione è stata compressa in maniera molto spinta (da 1416 a 5) e ciò nonostante, con questo insieme ridottissimo di variabili, corrispondenti alle prime cinque componenti principali, la struttura globale della classificazione è rimasta praticamente invariata (anzi si avvicina di più a quella basata sulla patologia). Questo risultato ci indica l'esistenza di fortissime relazioni fra i geni.

La figura 11 ci consente di approfondire la natura di queste relazioni. A questo livello di dettaglio (3 variabili e non più come nel caso dei geni 1416) possiamo agevolmente capire il 'messaggio biologico' che ci proviene dalla sperimentazione.

Si noti ad esempio come il cluster D (triangoli vuoti) sia molto distanziato dagli altri ed in netta evidenza ad un estremo della seconda componente. Il cluster D e' formato da tutte le linee provenienti da melanoma e da due linee di tumore al seno (figura 8). Osservando i geni che hanno un coefficiente più alto sulla seconda componente troviamo tutti geni legati ai processi di endocitosi, esocitosi e secrezione. Questo risultato è coerente con il fatto che i melanociti sono delle cellule la cui attività e strettamente legata ai processi di secrezione (produzione di melanina) che coinvolgono esocitosi ed endocitosi; analogamente l'attività secretoria è strettamente legata alla funzione della ghiandola mammaria.



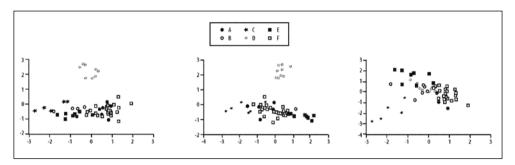


Figura 11. La figura riporta la distribuzione dei cluster nel piano delle prime componenti principali. I cluster che si trovano nelle posizioni estreme degli assi nei differenti grafici permettono un'interpretazione biologica delle componenti estratte, in maniera complementare, i geni massimamente correlati con le stesse componenti consentono di fornire un 'profilo di espressione genomica' alle classi di tumori.

Osservando la prima componente notiamo che ad una estremità si collocano gli elementi del cluster C (leucemie) mentre al lato opposto le cellule del cluster F (rene e sistema nervoso centrale). I geni maggiormente rilevanti per la componente 1 sono implicati con la motilità cellulare, ciò è coerente con il fatto che le cellule leucemiche sono disperse in un fluido e quindi non migrano su una matrice solida. Invece le cellule del sistema nervoso e del rene hanno una rilevante attività di motilità attraverso matrici. Inoltre la prima componente è fortemente correlata (r di Pearson = 0.82) con la successione da 1 a 6 dei cluster gerarchici calcolati sull'intero insieme di geni.

Questo ci fornisce una indicazione importantissima per l'intera biologia dei tumori: l'aspetto più importante delle differenze biologiche tra le linee tumorali è la loro capacità di migrare, cioè la loro invasività. Queste osservazioni a carattere generale potrebbero essere di spunto per la progettazione di farmaci, strategie terapeutiche o quant'altro.

Da quanto detto si evince la ricchezza di dati scaturiti dai *Microarray*, che solo se utilizzati criticamente e senza forzature potranno fornire informazioni scientificamente valide.

Chiaramente l'analisi degli esperimenti di *Microarray* sarà "intelligente" se operata di concerto, da un esperto di statistica ed da un esperto del fenomeno biologico in esame (Crecenzi M. e Giuliani A., 2001).

La descrizione di questo caso paradigmatico ci porta all'ultimo paragrafo di questo articolo che dimostra come l'uso di dati altamente paralleli modifica il nostro modo di fare scienza.



A cura di Anna Di Conardo

Come l'uso di dati altamente paralleli modifica il nostro modo di fare scienza

Fino ad ora l'attività dello scienziato sperimentale (fisico, chimico o biologo) si basava su poche misure eseguite sulle unità statistiche (animali, estratti organici, miscele gassose, ecc..). Le misure scelte erano accuratamente selezionate secondo la teoria vigente tra quelle con più possibilità di fornire risultati 'interessanti' cioè di essere notevolmente influenzate dal fenomeno che si era intenzionati a studiare. Per cui un fisiologo interessato alla distribuzione di aritmie misurava dei tracciati ECG, un fisico interessato alle instabilità dell'idrodinamica si concentrava sulle velocità dei flussi e sulla viscosità di liquidi scelti opportunamente, un chimico che avesse voluto studiare la chiralità si sarebbe concentrato su delle misure spettroscopiche di cristalli formati da una molecola con uno o più centri chirali.

Allo stesso modo un biologo molecolare individuava i tre o quattro prodotti genici che reputava (per considerazioni legate alla conoscenza pregressa) rilevanti per il fenotipo in esame e, attraverso opportune tecniche cromatografiche (elettroforesi, PCR...), cercava le correlazioni tra il grado di espressione dei geni analizzati con il fenotipo osservato.

Con l'avvento di tecniche altamente parallele come i *Microarray* la prospettiva cambia in maniera brusca: a questo punto il ricercatore ha la possibilità di analizzare simultaneamente sullo stesso campione migliaia di prodotti genici, senza disporre di una teoria che possa riunire in un unico quadro meccanicistico le relazioni fra questa moltitudine di prodotti e la sua specifica ricerca. Non solo questa teoria non esiste ma, anche se esistesse, sarebbe una teoria improduttiva, in primo luogo perché con migliaia di gradi di libertà potrebbe dar luogo a migliaia di interpretazioni tra l'altro non gestibili in maniera operativa, e pertanto ci si fermerebbe alla semplice descrizione del fenomeno (Rzhetsy E. et al., 2006).

Gli scienziati, secondo il loro temperamento e le loro idee, mostrano differenti atteggiamenti verso la sfida del parallelismo: alcuni semplicemente lo ignorano e si concentrano su *array* di pochi geni, tutti legati ad una certa funzione (riparazione del DNA, risposta immunitaria, metabolismo dei grassi...) che continuano a studiare in maniera classica. Questo è un atteggiamento assolutamente comprensibile e, in termini epistemologici, corrisponde alla 'scienza del tempo normale'. Tale modello descritto da Kuhn corrisponde ad un periodo in cui i paradigmi (teorie, modelli di interpretazione) sono abbastanza solidi da poterli considerare assodati e quindi è possibile concentrarsi su aspetti locali della scienza, su aggiustamenti ed applicazioni specifiche. Un altro atteggiamento, anche questo in qualche modo definibile da Kuhn 'scienza normale' è quello che abbiamo descritto in questo capitolo come 'la



vista del pessimista'. Il parallelismo è messo in condizione di 'non nuocere' con artifizi statistici o ricorrendo alle ontologie (e quindi ancora riferendosi in maniera acritica alla scienza pregressa) (Kuhn T., 1969). Un'altra posizione possibile è quella che accetta il parallelismo come una ricchezza, azzera per quanto possibile le premesse teoriche iniziali e si concentra sulla costruzione di un modello parsimonioso a pochi gradi di libertà. Esso, indipendentemente dalle sue motivazioni teoriche, fornisce una catalogazione efficiente dei risultati sperimentali. Soltanto dopo aver elaborato struttura efficace e consolidata del modello, è possibile trarre delle conclusioni circa lo studio in esame. Questo tipo di atteggiamento è quello seguito nello studio riportato al paragrafo precedente e corrisponde a qualcosa che Kuhn non aveva esplicitamente previsto, una sorta di 'grado minore' della scienza che potremmo definire 'approccio strutturalista' che si concentra sulla consistenza metodologica e la solidità delle relazioni evidenziate più che su una loro interpretazione teorica (Zbilut J. e Giuliani A., 2008). Come ultimo modello abbiamo la scelta definita da Kuhn 'scienza rivoluzionaria' o 'cambio di paradigma", che suggerisce di non considerare il gene come livello rilevante di spiegazione dei fenomeni biologici ed assumere, invece, il livello dei fenomeni collettivi. In questo caso si tratta di immaginare che le componenti principali (o qualsiasi altro costrutto matematico utilizzato per definire la dinamica congiunta di molti elementi differenti come geni, proteine, metaboliti ecc.) rimangano sostanzialmente invariate nella loro natura e comportamento al variare del grado di coinvolgimento dei geni particolari nella specifica componente. In questo modello, quindi, l'identità dei singoli geni viene tralasciata e si approfondisce la natura dei fenomeni collettivi. La cosa potrebbe sembrare assurda ma segue un ragionamento assolutamente naturale ed usato normalmente non solo in scienza ma anche nella vita di tutti i giorni. Immaginiamo di osservare le dinamiche del traffico su un' arteria che porta ad una grande città, ad esempio la A24 che porta a Roma; ogni giorno migliaia e migliaia di macchine la percorrono dall'estesa periferia orientale e dai paesi del Lazio e dell' Abruzzo. Possiamo essere praticamente sicuri, in modo quasi deterministico, che, in un giorno feriale, la mattina tra le otto e mezza e le nove, le corsie della A24 in ingresso a Roma dal raccordo anulare verso il centro saranno afflitte da un traffico intensissimo (se non completamente bloccato) e sicuramente immensamente superiore a quello osservabile alle tre del mattino o anche alla stessa ora ma in senso opposto. Questa sicurezza, fornitaci dalle osservazioni, viene efficientemente spiegata da considerazioni legate alla presenza di molti luoghi di lavoro nelle zone centrali di Roma (ministeri, scuole, uffici vari, negozi) rispetto alla periferia ed alla provincia e dalla generale importanza della città come centro di servizi per tutto il territorio. Sicuramente a nessuno verrà in testa di cercare di spiegare il flusso con delle motivazioni da ricercarsi nella vita dei singoli automobilisti (corrispondenti



ai singoli geni). Forniamo il seguente esempio: il signor X di Vicovaro una mattina si sveglia con una forte influenza e decide di non andare a Roma al lavoro (e quindi si esclude dal flusso, dal fenomeno collettivo), probabilmente sarà sostituito sulla A24 dalla signora Y che da Tivoli deve recarsi a Roma, perché il giorno prima è stata avvertita da un ospedale che si è liberato un posto per l'intervento programmato da mesi. Queste contingenze sono assolutamente irrilevanti per ciò che riguarda la consistenza e la natura del flusso e quindi non costituiscono un livello interessante per il suo studio, le periodicità del traffico rimarranno esattamente le stesse vanificando il nostro rincorrere le singole motivazioni.

L'odierna fisica teorica sta attraversando un periodo rivoluzionario di questo tipo: le frontiere della ricerca di base si stanno decisamente spostando dalle particelle elementari a nuovi fenomeni considerati fondamentali (e a volte chiamati quasi-particelle in nostalgico ricordo del passato) che hanno però un carattere collettivo, non diversamente dai flussi di traffico sull' autostrada A24 e sono quindi indipendenti dai particolari. Il lettore interessato all'argomento è rimandato al bellissimo libro divulgativo del premio Nobel della fisica Robert Laughlin del 1998. Esiste qualcosa del genere per la regolazione genica che ci autorizzi ad abbandonare il livello genico predominante negli ultimi decenni per flussi collettivi e aspecifici di espressione?

Forse qualcosa si sta affacciando all'orizzonte ma è ancora troppo presto per decidere se si tratta di curiosità episodiche o delle prime avvisaglie di un deciso cambio di paradigma. In ogni caso esistono indizi, supportati da evidenze sperimentali molto solide, di qualcosa che non trova spiegazione nel quadro interpretativo odierno della regolazione genica. La prima e più importante evidenza è legata all'esistenza di un livello caratteristico di espressione di ogni gene, tipico di una certa linea cellulare, fortemente invariante anche quando sottoposto a modificazioni rilevanti (mutazioni di geni importanti, trattamenti farmacologici). La figura 12 riporta due colture cellulari di macrofagi murini; sull' ascissa una coltura wild type, sull'ordinata una coltura di macrofagi murini mutante per l'assenza di un gene cruciale per l'immunità spontanea (Myd88). I punti rappresentano i livelli di espressione di ogni singolo elemento del genoma (23456 geni) nelle due linee cellulari (Tsuchyia M. et al., 2007). Come si osserva nella figura, esiste una convergenza quasi perfetta tra l'espressione di migliaia di geni che si estende su quattro ordini di grandezza di espressione (da 0 a 16000 unità di fluorescenza). Un ordine così perfetto che copre un raggio così vasto di livelli di espressione è qualcosa di assolutamente inusitato in natura e necessita di una spiegazione. Questo fenomeno non viene sufficientemente studiato dai biologi, dando come qualcosa di scontato o indimostrabile (se è così vuol dire che così deve essere). Anzi si cerca di normalizzare le espressioni geniche per nascondere queste differenze di ordini di grandezza. Il fenomeno però rima-



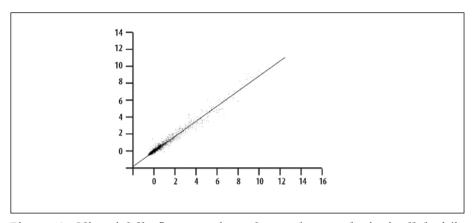


Figura 12. Gli assi della figura corrispondono a due popolazioni cellulari (in questo caso macrofagi murini): l'asse delle ascisse deriva da una popolazione 'wild type', l'asse delle ordinate da una popolazione in cui un gene molto importante per la risposta immunitaria (Myd88) è stato soppresso (knock-out). I punti corrispondono ai livelli di espressione nei due ceppi di 23000 geni. I due campioni sono stati analizzati in maniera indipendente e nonostante ciò mostrano una correlazione strettissima (r = 0.98), vicina al massimo assoluto (r=1). La correlazione vale per un numero impressionante di punti (i 23000 geni) e copre quattro ordini di grandezza (i valori di espressione riportati sugli assi). Un tale livello di integrazione e relazione reciproca tra i geni presuppone una rete strettissima di relazioni fra i geni dell'organismo che quindi non possono in nessun caso essere considerati come 'agenti autonomi'.

ne ed è osservabile ad ogni esperimento di microarray, e qualcuno prima o poi dovrà affrontarlo. Offrire una risposta scientifica a questa incredibile capacità organizzativa della regolazione genica, a mio avviso, porterebbe la biologia a confrontarsi direttamente con la meccanica statistica aprendo delle potenzialità immense di avanzamento.

La figura 13 mostra un altro interessantissimo fenomeno che indica la possibile esistenza di modi collettivi largamente indipendenti dai singoli geni coinvolti. Si tratta delle oscillazioni del livello di espressione genica che si osservano in colture di cellule non sincronizzate (e che quindi secondo la teoria corrente non dovrebbero mostrare alcun comportamento collettivo tempo-dipendente). Nella figura 13 si riporta la prima componente dell'espressione genica calcolata per insiemi di geni che vanno dall'intero genoma a pesche casuali di poche centinaia di geni a selezioni specifiche di geni unificati per comune funzione. Questa figura 13 ha nell'ascissa il tempo e nell'ordinata le variazioni correlate di espressione genica, questi cicli sono molto ripetibili ed indipendenti dai geni coinvolti e ci fanno comprendere



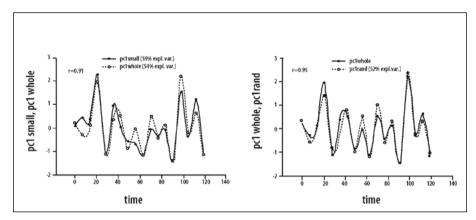


Figura 13. I grafici riportati in figura rappresentano l'andamento temporale della prima componente principale (pc1#) dell'espressione genica misurata su colture di lievito non esplicitamente sincronizzate. Se l'ipotesi ergodica (ogni cellula di una popolazione è indipendente dalle altre per cui i dati ricavati da una coltura cellulare si possono riferire ad una cellula media) fosse vera noi non osserveremmo alcun andamento temporale regolare ma solo fluttuazioni casuali. Al contrario non solo si osservano andamenti regolari, ma questi andamenti sono sempre gli stessi con differenti selezioni di geni (whole = intero insieme dei geni implicati nella costruzione dei ribosomi (256 geni), small = sottoinsieme di geni ribosomi ali (14 geni), random = estrazione casuale di 256 geni dall'intero insieme di trascritti). L'indipendenza dei ritmi osservati dalle selezioni di geni indica che qui stiamo osservando un'attività globale del genoma legata all'organizzazione spaziale e temporale delle cellule in coltura che formano quindi non un insieme casuale di elementi ma una popolazione molto organizzata. I dati qui riportati provengono da: M. Tsuchyia, S.T. Wong, Z. X. Yeo, A. Colosimo, M.C. Palumbo, L. Farina, M. Crescenzi, A. Mazzola, R. Negri, M.M. Bianchi, K. Selvarajoo, M. Tomita and A.Giuliani (2007) Gene expression waves: cell cycle independent collective dynamics in cultured cells FEBS J. 274: 2874-2886.

due cose molto importanti: la prima è che una colonia cellulare non è un assortimento casuale di cellule che fanno ognuna il suo 'gioco' indipendentemente dall'altra, in questo caso non avremmo nessun ciclo, la seconda è che i geni si muovono di concerto indipendentemente dalla loro funzione specifica. Queste dinamiche fanno pensare ad una sorta di 'metronomo comune' che mette in fase le diverse cellule di una coltura, una sorta di battito cardiaco delle cellule di cui non conosciamo l'origine ed il senso, ma di cui abbiamo molte evidenze sperimentali che vanno dallo studio del metabolismo, alla produzione di proteine e alla regolazione dell'espressione genica. L'esistenza di queste dinamiche svuota di significato l'usuale interpretazione dei dati



provenienti dalle colture cellulari come ascrivibili ad una sorta di 'cellula media' e quindi riportati invariabilmente al livello molecolare.

In futuro conosceremo l'esito di queste sperimentazioni che ora sono appena agli esordi, qui ci preme far notare come questi nuovi sviluppi siano stati propiziati dalla necessità di sfruttare tutte le potenzialità di una nuova tecnologia. La tecnologia usata per derivare le nostre misure sperimentali non è mai indifferente rispetto al nostro modo di pensare e di porci i problemi. Quando una tecnologia ispirata ad una filosofia nuova come l'alto parallelismo nel caso dei *Microarray* compare sulla scena non è certamente saggio far finta che niente di nuovo sia accaduto.

Bibliografia

- Elementi di statistica descrittiva ed inferenziale, statistica multidimensionale
- AA.VV.(2006) Statsoft Electronic Textbook: http://www.statsoft.com/textbook/stathome.html (Risorsa on-line, consultabile gratuitamente, con spiegazioni chiare ed aggiornate su tutte le principali tecniche statistiche e spiegazioni approfondite di analisi in componenti principal: factor analysis e cluster analysis).
- Benigni R., Giuliani A.: Quantitative modeling and biology: the multivariate approach. Am J Physiol. 1994 May; VOL: 266(5 Pt 2):R, P:1697. (Presentazione generale delle tecniche multivariate in ambito biomedico).
- Siegel S.: Statistica non parametrica per le scienze del comportamento; Organizzazioni Speciali, Firenze 1980. (Bellissima trattazione della metodologia statistica con un occhio attento al senso profondo delle tecniche utilizzate e non solo alla loro applicazione).
- Applicazioni della Statistica ai microarray
- Allison D.B., Cui X., Page G.P.: Sabripour M. Microarray data analysis: from disarray to consolidation and consensus. Nature Reviews Genetics 2006 Jan, VOL: 7(1), P:55. (Rassegna sullo stato dell'arte dei diversi approcci statistici all'analisi dei microarray).
- Alter O., Brown P.O., Botstein D.: Singular value decomposition for genome-wide expression data processing and modeling. Proc Natl Acad Sci U S A. 2000 Aug 29, VOL:97(18), P:10101. (*Trattazione sull'applicazione delle tecniche multidimensionali all'analisi dei dati di microarray*).





- Berrar D.P., Dubitzky W., Granzow M. (Eds): A practical approach to Microarray Data Analysis. Kluwer, (2003). (Libro dedicato ad una trattazione dettagliata ed esauriente di tutte le tecniche per effettuare un'analisi dettagliata di un esperimento di microarray. Necessita di una certa formazione matematica per la sua comprensione, anche se scritto con sufficiente chiarezza).
- Camon E., Magrane M., Barrell D., Binns D., Fleischmann W., Kersey P., Mulder N., Oinn T., Maslen J., Cox A., Apweiler R.: The Gene Ontology Annotation (GOA) Project: Implementation of GO in SWISS-PROT, TrEMBL, and InterPro. Genome Res 2003 Apr, VOL:13(4), P:662. (questa è la presentazione fatta dal gruppo di lavoro dell' EMBL (European Molecular Biology Laboratory) dei criteri guida utilizzati per la costruzione di ontology unificate che fornissero una descrizione il più possibili e accurate ed efficienti dei prodotti genici. L'interesse dell'articolo è anche legato ai link con molte ontologie in linea che il lettore potrà utilmente consultare).
- Crescenzi M., Giuliani A.: The main biological determinants of tumor line taxonomy elucidated by a principal component analysis of microarray data. FEBS Lett. 2001 Oct 19, VOL:507(1), P:114. (*lavoro descritto nell'esempio*).
- Mehta T., Tanik M., Allison DB.: Towards sound epistemological foundations of statistical methods for high-dimensional biology. Nat Genet. 2004 Sep, VOL:36(9), P:943. (articolo chiaro ed esauriente sul problema del parallelismo e dell'alta dimensionalità).
- Ross D.T., Scherf U., Eisen M.B., Perou C.M., Rees C., Spellman P., Iyer V., Jeffrey S.S., Van de Rijn M., Waltham M., Pergamenschikov A., Lee J.C., Lashkari D., Shalon D., Myers T.G., Weinstein J.N., Botstein D., Brown P.O.: Systematic variation in gene expression patterns in human cancer cell lines. Nat Genet. 2000 Mar, VOL:24(3), P:227. (fonte dei dati originali descritti nell'esempio svolto).

Systems Biology

- Feng X.J., Hooshangi S., Chen D., Li G., Weiss R., Rabitz H.: Optimizing genetic circuits by global sensitivity analysis. Biophys J. 2004 Oct, VOL:87(4), P:2195. (articolo tecnico che merita di essere inserito nei riferimenti bibliografici,poiché rispecchia pregi ed difetti di un approccio puramente ingegneristico alla biologia.).
- Kitano H.: Systems biology: a brief overview. Science. 2002 Mar 1, VOL: 295(5560), P:1662. (Hiroaki Kitano è uno dei nomi maggiori esperti in questa nuova scienza e questa breve rassegna e di fondamentale interesse).
- Palumbo M.C., Colosimo A., Giuliani A., Farina L.: Functional essentiality from topology features in metabolic networks: a case study in yeast. FEBS Lett. 2005 Aug 29, VOL:579(21), P:4642.(La figura n 5 di questo saggio è tratta da questo articolo che descrive la previsione del carattere letale delle mutazioni nella rete metabolica del lievito.).



Questioni Epistemologiche

Kuhn T.: La struttura delle rivoluzioni scientifiche. Torino. Einaudi, 1969. Laughlin R.: Un universo diverso. Reinventare la fisica da cima a fondo. Codice, 2005.

Rzhetsky A., Iossifov I., Loh J.M., White K.P.: Microparadigms: chains of collective reasoning in publications about molecular interactions. Proc Natl Acad Sci U S A. 2006 Mar 28, VOL:103(13), P:4940. (articolo su argomentazioni statistiche, le questioni trattate sono di grande importanza ed attualità).

Tsuchiya M., Wong S.T., Yeo Z.X., Colosimo A., Palumbo M.C., Farina L., Crescenzi M., Mazzola A., Negri R., Bianchi M.M., Selvarajoo K., Tomita M., Giuliani A.: Gene expression waves. Cell cycle independent collective dynamics in cultured cells. FEBS J. 2007 Jun, VOL: 274(11), P:2878. (fonte degli ultimi grafici del testo. Articolo della dedicato alla regolazione genica, provvisto di una buona bibliografia che il lettore interessato può consultare per approfondire l'argomento).

Zbilut J, Giuliani A.: The Latent Order of Complexity. Nova Publisher New York, 2008. (ibro dedicato all'analisi di come l'approccio scientifico stia cambiando, spostando i suoi interessi primari verso i sistemi complessi. Il libro è di carattere divulgativo e non necessita di alcuna conoscenza specifica).

NB: L'autore di questa rassegna ha scelto di segnalare i riferimenti bibliografici divisi secondo gli argomenti trattati. Inoltre, l'autore acclude alcuni commenti specifici, per accompagnare il lettore nella scelta dei testi per ulteriori approfondimenti.



I Microarray







Problematiche etiche collegate ai test genetici

Sabina Gainotti¹, Carlo Petrini²

 $^{
m 1}$ Centro Nazionale di Epidemiologia, Sorveglianza e Promozione della Salute Istituto Superiore di Sanità - Roma.

²Unità di Bioetica, Presidenza, Istituto Superiore di Sanità - Roma.

Test Genetico

Il termine "test genetico" è ampio e può indicare esami diagnostici e predittivi che hanno implicazioni etiche diverse.

Alcuni test genetici possono indicare la probabilità che una persona sviluppi una certa malattia. Attraverso i test è anche possibile scoprire se una persona è portatrice di una mutazione genetica, condizione che comporta un maggior rischio per i suoi figli di sviluppare la malattia. Con la realizzazione del progetto Genoma Umano, la cartografia del genoma umano e le conoscenze dei geni associati a diverse malattie è destinato ad aumentare, insieme con la disponibilità di test genetici. I test ad oggi disponibili possono essere classificati in test e screening genetici diagnostici e predittivi (Tabella 1).

I problemi etici sollevati dall'uso dei test genetici sono numerosi. Senza pretendere di essere esaustivi elencheremo alcune problematiche già individuate nella letteratura, relativa all'esecuzione dei soli test genetici. Non affronteremo invece il tema degli screening.

Affidabilità del/dei test

Un primo problema fondamentale riguarda l'affidabilità di test e il loro valore predittivo, in particolare il rischio di generare falsi positivi o falsi negativi. I falsi positivi infatti comportano inutili preoccupazioni per chi si sottopone al test. Si può persino essere indotti a chiedere interventi chirurgici radicali (ad esempio una mastectomia per prevenire un tumore al seno). I falsi negativi comportano invece una falsa sicurezza che impedisce al soggetto di prendere adeguate misure di prevenzione (un trattamento, un cambiamento di stile di vita, ecc.).





TEST DIAGNOSTICI: Riguardano le malattie che hanno una forte componente genetica, e possono essere effettuati in vari momenti della vita, sia prima che dopo la nascita.

SCREENING DIAGNOSTICI

Prenatali: Riguardano i cromosomi principali e i difetti di sviluppo del sistema nervoso; per questi test vengono utilizzati dei sieri in combinazione con immagini ad ultrasuoni. In caso di risultato positivo la conferma della diagnosi è effettuata utilizzando test citogenetici e analisi dei cromosomi. Neonatali per malattie curabili come la fenilchetonuria e l'ipotiroidismo; nella maggior parte dei casi si tratta di test biochimici a basso costo su campioni di sangue presi da bambini appena nati conosciuti con il nome di test di Guthrie.

Test per altre patologie, come la galattosemia.

Postnatali: (per malattie come la fibrosi cistica). È stato introdotto in Inghilterra ed è stato preso in considerazione in altri paesi, incluse la Francia e la Repubblica Ceca.

TEST PREDITTIVI Rientrano in questa categoria test che ricoprono una ampia gamma di malattie e condizioni acquisite e possono aiutare a predire il futuro stato di salute di un individuo. In realtà la linea che divide le condizioni ereditate da quelle acquisite non è chiaramente definita.

Esistono due gruppi di test predittivi presintomatici:

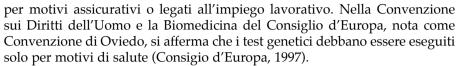
- test che individuano la presenza di difetti in particolari geni o prodotti genetici e possono predire (a volte anche con certezza) la probabilità di sviluppare una malattia negli anni futuri (ad es. la Corea di Huntington, alcune forme di tumore e le forme genetiche della malattia di Alzheimer)
- test di predisposizione a malattie in cui il difetto di un unico gene è considerato come un fattore di rischio per lo sviluppo di malattie (ad es. il cancro al seno o alle ovaie)
- test farmacogenetici che indicano le predisposizioni individuali nella rea zione ai farmaci.
- test per monitorare gli effetti di alcune malattie che possono essere usati al momento della diagnosi per prevedere le possibili evoluzioni della malattia (ad esempio per alcune forme tumorali).

Tabella 1. Test e screening genetici.

Obiettivo del/dei test

Prima di effettuare un test genetico è necessario capire il motivo della richiesta. Un test genetico potrebbe essere fatto per motivi strettamente medici o anche, soprattutto in paesi che non hanno un servizio sanitario pubblico,





Tuttavia i "motivi di salute" devono essere a loro volta definiti. La salute può definirsi non solo come assenza di malattia ma anche come speranza di vita. Gli "obiettivi di salute" inoltre non escludono la presenza di altri obiettivi.

Prima di effettuare un test per motivi legati alla salute sarebbe dunque corretto interrogarsi sui seguenti punti:

- La malattia che si vuole ricercare è grave (mortale o comportante un handicap grave)?
- La malattia può essere curata?
- I sintomi della malattia possono essere alleviati?
- È possibile agire in modo preventivo sulla malattia?
- La persona che richiede il test è preoccupata?
- La malattia è già presente nella famiglia?

Informazione del soggetto e consenso informato

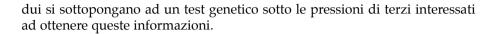
Il contenuto delle informazioni genetiche, l'importanza e l'eventuale impatto di queste informazioni sulla vita del paziente e dei suoi familiari condizionano i modi di ottenimento del consenso informato all'esecuzione dei test. Quando il contenuto dell'informazione è a basso impatto (ad esempio nelle analisi per il gruppo sanguigno) il consenso può essere addirittura presunto, o dato solo verbalmente.

In alcune istituzioni, in altri paesi, per poter usare i tessuti umani e i dati collegati a scopo di ricerca ci si avvale del cosiddetto "blanket consent" o "consenso generico", che consiste nell'informare il paziente dell'utilizzo a scopo di ricerca dei suoi tessuti e delle informazioni mediche ad essi collegate (Murray T.H. et al., 2000).

Al contrario alcuni test genetici, per via del loro impatto e per via delle possibili pressioni da parte di terzi, richiedono sempre un consenso informato scritto e dettagliato.

Il consenso informato infatti deve permettere agli individui di comprendere tutte le implicazioni del test, incluse la possibilità di apprendere di essere portatore sano, l'eventualità di dover adattare il proprio stile di vita, le possibili conseguenze del test a livello familiare e sociale. Le informazioni devono dunque essere estese e comprensive, e devono evitare che gli indivi-





Il consenso informato "del futuro"

In alcune ricerche genetiche può accadere che i ricercatori scoprano informazioni clinicamente rilevanti per il paziente-donatore di tessuti. Tali scoperte possono porre domande difficili, soprattutto se sono inattese (è accaduto ad esempio di trovare una forte correlazione tra il gene apo-E4 e la malattia di Alzheimer nello studio del collegamento tra il gene apo-E4, l'ipercolesterolemia e le malattie cardiache,) (Kolata G., 1995). La domanda centrale in questi casi è quando sia giusto restituire i risultati della ricerca ai soggetti partecipanti che non hanno dato il loro consenso per quel tipo di ricerca. Esistono argomenti favorevoli e argomenti contrari al "dovere di contattare" il paziente in seguito a risultati inattesi.

Gli argomenti a favore richiamano il dovere del clinico di ricontattare i pazienti per dare loro tutte le informazioni disponibili sul loro trattamento e l'obbligo del ricercatore di procurare un beneficio al paziente quando ciò sia possibile.

Gli argomenti contrari riguardano l'aumento del peso finanziario ed amministrativo della pratica di ricontattare il paziente, l'incertezza di alcuni risultati della ricerca che potrebbero non essere pienamente capiti o validati in studi successivi e gli eventuali danni psicologici, sociali ed economici per i soggetti coinvolti se i risultati sono comunicati senza un previo consenso ed un counselling adeguato.

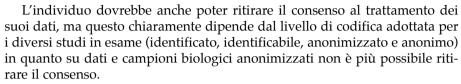
Il consenso per la raccolta e la conservazione dei tessuti e dei dati collegati

Un problema che merita attenzione al momento del consenso informato riguarda la classificazione e l'uso dei campioni prelevati dalla persona per effettuare ricerche mediche o scientifiche.

La "fonte" del materiale biologico, ovvero la persona, deve poter decidere sulla conservazione del proprio materiale biologico e dei suoi usi futuri.

Gli eventuali usi futuri del materiale biologico "donato" devono essere ben spiegate alla persona in questione anche se non sempre è possibile prevedere gli sviluppi di una ricerca nel tempo.





Esistono proposte contrastanti sul consenso informato per la ricerca genetica e la conservazione dei tessuti e dei dati ad essi collegati. Ad un estremo troviamo, per esempio, la posizione di George Annas, secondo cui: "Nessun raccolta di campioni di DNA dovrebbe essere permessa senza una autorizzazione scritta dell'individuo che a) stabilisca gli obiettivi per cui il materiale è raccolto b) stabilisca tutti gli usi consentiti del campione, inclusi quelli commerciali; c) garantisca all'individuo di avere un continuo accesso al campione e ai dati collegati nonché il diritto assoluto di richiedere la distruzione del campione in qualunque momento e d) garantisca la distruzione del campione nel caso in cui la banca dati che lo conserva chiuda o cambi identità" (Annas G.J., 1994).

Sempre secondo Annas i campioni andrebbero usati unicamente per gli scopi per cui sono stati raccolti, mentre l'uso di moduli "aperti" per il consenso informato e l'utilizzo dei campioni da parte di terze parti non dovrebbero essere permessi.

Altri autori hanno fatto notare che un consenso informato elaborato, simile a quello adottato nella ricerca medica più "interventista" non sarebbe appropriato per la ricerca genetica, che non comporta rischi altrettanto rilevanti per il soggetto. L'uso di un consenso particolarmente elaborato potrebbe spaventare i soggetti prospettando danni spaventosi che sono in realtà estremamente improbabili, portando i soggetti ad esprimere un "dissenso disinformato" (Korn D., 1999).

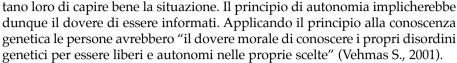
Un'ipotesi alternativa riguarda la possibilità di esprimere una "direttiva anticipata" all'uso dei propri tessuti, aggiornabile e modificabile dai soggetti (Weir R.F., 1998). Un documento di questo tipo potrebbe consentire l'utilizzo dei propri tessuti per ogni tipo di ricerca, limitarne l'uso per ricerche su malattie specifiche, o impedirne ogni tipo di utilizzo.

Diritto del paziente di sapere o di non sapere

La ricerca genetica offre nuove informazioni alle persone: aumentando le informazioni aumentano anche le possibilità di fare scelte autonome e "responsabili".

Questo potrebbe implicare nuovi "doveri" per gli individui, ad esempio il dovere di mantenersi informati sulla ricerca genetica. Per fare scelte autonome le persone hanno necessità di ricevere informazioni adeguate, che permet-





Ogni scelta di vita in effetti comporta una responsabilità morale, e tale responsabilità può essere impedita in due modi: con la coercizione oppure con l'omissione di informazioni rilevanti. In realtà anche scegliere di non sapere può essere un comportamento pienamente cosciente e responsabile, nonché un diritto, non si può sottovalutare il rischio di "danno informativo" che potrebbe derivare dal rivelare i risultati di un test genetico.

Eppure il diritto di sapere o di non sapere coinvolge non solo la sfera personale degli individui, ma anche quella familiare (McNally E. e Cambon -Thomsen A., 2004). Il diritto di sapere di un membro della famiglia può entrare in conflitto con il diritto di non sapere di un altro membro della famiglia. Ad esempio, se un parente deceduto ha sofferto di un problema genetico, suo nipote potrà essere interessato ad effettuare un test di portatore sano per fare scelte a livello procreativo, ma la sua scelta potrebbe rivelare informazioni non volute anche ai suoi genitori. Queste situazioni sono inevitabili, ma in principio si risolvono se il motivo per cui la persona richiede il test è sufficientemente valido e importante.

Infine, il diritto di una persona di non sapere può entrare in conflitto con il diritto all'integrità fisica propria e di altre persone.

Dovere del medico di avvertire

L'informazione genetica impone nuove scelte alle persone: decidere se rivelare o meno i risultati di un test genetico ai propri familiari, condividere con loro le informazioni o non condividerle: queste nuove questioni possono creare conflitti all'interno delle famiglie, e mettere il medico di fronte a problemi etici e legali rilevanti.

Come regola generale, la possibilità che i familiari siano interessati ai risultati del test dovrebbe essere discussa in un counselling pre-test, dove si parlerà anche dei modi di un loro eventuale coinvolgimento.

Esistono approcci legali diversi per quanto riguarda il passaggio d'informazione ai familiari. In molti paesi i medici sono tenuti al rispetto del segreto professionale, e prima di dare informazioni ai familiari, è necessario il consenso dell'individuo interessato. Solo in casi eccezionali, quando la comunicazione può prevenire danni gravi ai familiari le regole di tutela della riservatezza possono venir meno. In alcuni casi si può informare un familiare



a rischio per proporgli di effettuare un test genetico: questo potrebbe evitare dei danni alla sua salute futura.

Inevitabilmente però, iniziative di questo tipo vanno contro l'eventuale "diritto di non sapere" del familiare. Evidentemente, con l'informazione genetica, la privacy individuale potrà creare conflitti di interessi tra gruppi di persone.

Distinto dal rivelare è il dovere di avvertire. Negli Stati Uniti esistono diverse norme che riguardano il dovere di avvertire terze parti nel caso in cui la loro salute sia a rischio; questo può accadere quando il problema genetico di una persona può mettere in pericolo la salute propria e altrui. In queste situazioni il dovere di avvertire può prevalere sulle linee guida di pratica professionale e di rispetto della riservatezza. La paura di essere denunciati può rendere ancor più problematico per i medici il problema del segreto professionale, e in alcuni casi avvertire i familiari può sembrare l'opzione meno rischiosa.

Gli studi sulle popolazioni e il "consenso di gruppo" o di comunità

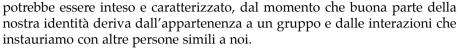
Il 90% delle variazioni genetiche presenti nel genoma umano è comune a tutte le popolazioni, ma alcune variazioni, causate da mutazioni o da altri meccanismi, sono più comunemente riscontrabili in persone che vivono in particolari aree geografiche.

Lo studio dei diversi gruppi di persone è molto utile a scopo comparativo, ma questi studi potrebbero essere usati in modo inappropriato e discriminatorio.

I diritti e il benessere dei gruppi sociali e delle persone che ne fanno parte stanno acquisendo un'importanza sempre maggiore nell'ambito della ricerca biomedica e genetica.

Possiamo definire un gruppo "... un insieme di individui legati culturalmente o etnicamente, che è probabile o possibile che condividano alcune caratteristiche genetiche. Spesso i gruppi si caratterizzano con delle etichette demografiche, ad esempio afroamericani, indiani americani, ebrei askenaziti, ecc." (Apert S., 2000). Con i progressi della ricerca genetica molecolare, i gruppi composti da persone geneticamente collegate diventano "unità d'analisi" sempre più ambite, soprattutto quando nel tratto genetico si ipotizza una componente di "gruppo", ovvero quando quel tratto è prevalente in alcuni individui geneticamente collegati. Non ci si deve stupire del fatto che i membri di alcuni gruppi si preoccupino del modo in cui il loro gruppo





A questo proposito Gostin ha scritto: "Le informazioni degradanti associate ad un gruppo possono provocare danni importanti contro i membri del gruppo stesso, come le discriminazioni in ambito lavorativo, assicurativo o abitativo. Le informazioni degradanti possono anche provocare danni intangibili ai gruppi come un abbassamento dell'autostima o dell'orgoglio razziale o culturale" (Gostin L., 2001).

Preoccupazioni importanti riguardano il possibile impatto sui gruppi etnici dei geni associati con la personalità, l'intelligenza, il comportamento. Quando un gruppo o una comunità identificabile diventa una "unità d'analisi" di una ricerca genetica, i ricercatori dovrebbero coinvolgere i membri della comunità nel processo stesso di ricerca, dall'arruolamento alla pubblicazione dei risultati.

Per tutti questi motivi è importante ricercare il "consenso informato delle comunità" nelle ricerche che riguardano popolazioni. Esistono proposte per il consenso informato negli studi di popolazione (Beskow L.M. et al., 2001). Il "consenso informato di comunità" rimane comunque un concetto problematico. Nei progetti sui gruppi la domanda: "chi parla per il gruppo?" solleva un problema importante, e di difficile risposta. Mentre è possibile trovare persone che facilitino il dialogo (leader religiosi), alcuni gruppi sono troppo popolosi e dispersi per poter identificare dei leader che parlino a nome del gruppo.

Nel descrivere gli scopi di uno studio poi, non sempre è facile spiegare se, e come, i risultati ottenuti saranno utili a migliorare la salute delle popolazioni in questione, né si può dire quale sarà il contributo di una conoscenza della storia biologica dell'umanità.

Questa complessità di obiettivi non è sempre facile da spiegare in modo chiaro, soprattutto in contesti culturali "diversi".

Gli orizzonti della ricerca genetica possono risultare in nuove dimensioni societarie. In alcuni paesi esistono collezioni di dati molto ampie (la biobanca inglese ad esempio include campioni e dati provenienti da circa 500 mila persone), e la loro gestione ed utilizzo è un problema che riguarda le società in senso lato. Il consenso di gruppo può essere appropriato per popolazioni di dimensioni ristrette, anche se in alcuni contesti lo stesso consenso di gruppo è difficile da raggiungere.

Per le popolazioni più ampie il consenso di gruppi ristretti risulta meno adatto; solo un dibattito allargato e democratico potrà aiutare gli individui a prendere decisioni importanti, prima ancora dell'inizio di progetti su larga scala e prima della richiesta di consensi individuali.





- Alpert S.: Privacy and the analysis of stored tissues. In Research involving human biological materials; ethical issues and policy guidance; vol II, Commissioned Papers. National Bioethics Advisory Commission, Rockville, MD, 2000.
- Annas G.J.: Rules for gene banks: protecting privacy in the genetics age. In Murphy T.F., Lappé A.M. (Eds) Justice and the Human Genome Project. Berkeley. University of California Press, 1994.
- Beskow L.M., Burke W., Merz J.F. et al.: Informed consent for population-based research involving genetics. JAMA 2001 Nov, VOL:286(18), P:2315.
- Consiglio d'Europa. Convention for the Protection of Human Rights and Dignity of the Human Being with regard to the Application of Biology and Medicine: Convention on Human Rights and Biomedicine. Oviedo, 4.IV.1997.
- Gostin L.: Ethical principles for the conduct of human subject research: population-based research and ethics. Law Med Health Care 1991, VOL:19(3-4), P:191.
- Kolata G.: New York Times, October 24, pp A1, C6. 1995.
- Korn D.: Genetic privacy, medical information privacy, and the use of human tissue specimens in research. in genetic testing and the use of information. Washington, DC. The AEI Press, 1999.
- McNally E. e Cambon-Thomsen A.: Ethical, legal and social aspects of genetic testing: research, development and clinical applications. 2004; Belgium, European Communities (p.60). Disponibile per la consultazione al sito web: http://ec.europa.eu/research/conferences/2004/genetic/pdf/report_en.pdf
- Murray T.H. e Mehlman M.J. (Eds). Encyclopedia of ethical, legal and policy issues in biotechnology. Vol 1. New York. John Wiley & Sons, 2000.
- Weir R.F. Stored tissue samples: Ethical, legal, and public policy implications. Iowa City. University of Iowa Press, 1998.
- Vehmas S.: Just Ignore It? Parents and Genetic Information. Theor Med Bioethics 2001 Sep, VOL:22(5), P:473.



Si ringrazia il Signor Alessandro Spurio per il suo prezioso contributo nella realizzazione della parte grafica





Indice

Editorialepa	.g. 3
Dalla doppia elica ai microarray (Anna Di Lonardo, Simonetta Pulciani, Corrado Fagnani)	» 7
La tecnologia dei Microarray (Simonetta Pulciani, Anna Di Lonardo, Alessandro Spurio, Corrado Fagnani)	» 21
CGH array ed applicazioni (Giovanna Floridia, Federica Censi, Stella Lanni)	» 31
Arrays di proteine: tecnologie ed applicazioni in campo clinico (Maria Letizia Polci e Tamara Corinna Petrucci)	» 43
Analisi dei dati negli esperimenti di microarray (Alessandro Giuliani)	» 59
Problematiche Etiche Collegate ai Test Genetici (Sabina Gainotti e Carlo Petrini)	» 91
Indica	101



Caleidos copio

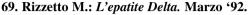
... il futuro ha il cuore artico MEDICAL SYSTEMS SpA

- 1. Rassu S.: Principi generali di endocrinologia. Gennaio '83
- 2. Rassu S.: L'ipotalamo endocrino. Giugno '83
- 3. Rassu S.: L'ipofisi. Dicembre '83
- 4. Alagna., Masala A.: La prolattina. Aprile '84
- 5. Rassu S.: Il pancreas endocrino. Giugno '84
- 6. Fiorini I., Nardini A.: Citomegalovirus, Herpes virus, Rubella virus (in gravidanza). Luglio '84.
- 7. Rassu S.: L'obesita'. Settembre '84
- 8. Franceschetti F., Ferraretti A.P, Bolelli G.F., Bulletti C.: Aspetti morfofunzionali dell'ovaio. Novembre '84.
- 9. Kubasik N.P.: Il dosaggio radioimmunologico (1). Dicembre '84.
- 10. Kubasik N.P.: Il dosaggio radioimmunologico (2) parte prima. Gennaio'85.
- 11. Kubasik N.P.: Il dosaggio radioimmunologico (2) parte seconda. Febbraio '85.
- 12.Kubasik N.P.: Il dosaggio radioimmunologico (3) parte prima. Aprile '85.
- 13. Nacamulli D, Girelli M.E, Zanatta G.P, Busnardo B.: Il TSH. Giugno '85.
- 14. Facchinetti F. e Petraglia F.: La β-endorfina plasmatica e liquorale. Agosto '85.
- 15. Baccini C.: Le droghe d'abuso (1). Ottobre '85.
- 16. Kubasik N.P.: Il dosaggio radioimmunologico (3) parte seconda. Dicembre '85.
- 17. Nuti R.: Fisiologia della vitamina D: Trattamento dell'osteoporosi post-menopausale. Febbraio '86
- 18. Cavallaro E.: Ipnosi: una introduzione psicofisiologica. Marzo '86.
- 19. Fanetti G.: AIDS: trasfusione di sangue emoderivati ed emocomponenti. Maggio '86.
- 20. Fiorini I., Nardini A.: Toxoplasmosi, immunologia e clinica. Luglio '86.
- 21. Limone P.: *Il feocromocitoma*. Settembre '86.
- 22. Bulletti C., Filicori M., Bolelli G.F., Flamigni C.: *Il Testicolo. Aspetti morfo-funzionali e clinici*. Novembre '86.
- 23. Bolcato A.: Allergia. Gennaio '87.
- 24. Kubasik N.P.: Il dosaggio enzimoimmunologico e fluoroimmunologico. Febbraio '87.
- 25. Carani C.: Patologie sessuali endocrino-metaboliche. Marzo '87.
- 26. Sanna M., Carcassi R., Rassu S.: Le banche dati in medicina. Maggio '87.
- 27. Bulletti C., Filicori M., Bolelli G.F., Jasonni V.M., Flamigni C.: L'amenorrea. Giugno '87.
- 28. Zilli A., Pagni E., Piazza M.: Il paziente terminale. Luglio '87.
- 29. Pisani E., Montanari E., Patelli E., Trinchieri A., Mandressi A.: *Patologie prostatiche*. Settembre '87.
- 30. Cingolani M.: Manuale di ematologia e citologia ematologica. Novembre '87.
- 31. Kubasik N.P.: Ibridomi ed anticorpi monoclonali. Gennaio '88.



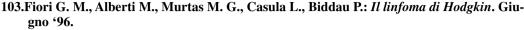
- 32. Andreoli C., Costa A., Di Maggio C.: Diagnostica del carcinoma mammario. Febbraio '88.
- 33. Jannini E.A., Moretti C., Fabbri A., Gnessi L., Isidori A.: Neuroendocrinologia dello stress. Marzo '88.
- 34. Guastella G., Cefalù E., Carmina M.: La fecondazione in vitro. Maggio '88.
- 35. Runello F., Garofalo M.R., Sicurella C., Filetti S., Vigneri R.: Il gozzo nodulare. Giugno '88.
- 36. Baccini C.: Le droghe d'abuso (2). Luglio '88.
- 37. Piantino P., Pecchio F.: Markers tumorali in gastroenterologia. Novembre '88.
- 38. Biddau P.F., Fiori G.M., Murgia G.: Le leucemie acute infantili. Gennaio '89.
- 39. Sommariva D., Branchi A.: Le dislipidemie. Febbraio '89.
- 40. Butturini U., Butturini A.: Aspetti medici delle radiazioni. Marzo '89.
- 41. Cafiero F., Gipponi M., Paganuzzi M.: Diagnostica delle neoplasie colo-rettali. Aprile '89.
- 42. Palleschi G.: Biosensori in Medicina. Maggio '89.
- 43. Franciotta D.M., Melzi D'Eril G.V. e Martino G.V.: HTLV-I. Giugno '89.
- 44. Fanetti G.: Emostasi: fisiopatologia e diagnostica. Luglio '89.
- 45. Contu L., Arras M.: Le popolazioni e le sottopopolazioni linfocitarie. Settembre '89.
- 46. Santini G.F., De Paoli P., Basaglia G.: Immunologia dell'occhio. Ottobre '89.
- 47. Gargani G., Signorini L.F., Mandler F., Genchi C., Rigoli E., Faggi E.: *Infezioni opportunistiche in corso di AIDS*. Gennaio '90.
- 48. Banfi G., Casari E., Murone M., Bonini P.: La coriogonadotropina umana. Febbraio '90.
- 49. Pozzilli P., Buzzetti R., Procaccini E., Signore E.: L'immunologia del diabete mellito. Marzo '90.
- 50. Cappi F.: La trasfusione di sangue: terapia a rischio. Aprile '90.
- 51. Tortoli E., Simonetti M.T.: I micobatteri. Maggio '90.
- 52. Montecucco C.M., Caporali R., De Gennaro F.: Anticorpi antinucleo. Giugno '90.
- 53. Manni C., Magalini S.I. e Proietti R.: Le macchine in terapia intensiva. Luglio '90.
- 54. Goracci E., Goracci G.: Gli allergo-acari. Agosto '90.
- 55. Rizzetto M.: L'epatite non A non B (tipo C). Settembre '90.
- 56. Filice G., Orsolini P., Soldini L., Razzini E. e Gulminetti R.: *Infezione da HIV-1: patogenesi ed allestimento di modelli animali*. Ottobre '90.
- 57. La Vecchia C. Epidemiologia e prevenzione del cancro (I). Gennaio '91.
- 58. La Vecchia C. Epidemiologia e prevenzione del cancro (II). Febbraio '91.
- 59. Santini G.F., De Paoli P., Mucignat G., e Basaglia G., Gennari D.: Le molecole dell'adesività nelle cellule immunocompetenti. Marzo '91.
- 60. Bedarida G., Lizioli A.: La neopterina nella pratica clinica. Aprile '91.
- 61. Romano L.: Valutazione dei kit immunochimici. Maggio '91.
- 62. Dondero F. e Lenzi A.: L'infertilità immunologica. Giugno '91.
- 63. Bologna M. Biordi L. Martinotti S.: Gli Oncogèni. Luglio '91.
- 64. Filice G., Orsolini P., Soldini L., Gulminetti R., Razzini E., Zambelli A. e Scevola D.: *Infezione-malattia da HIV in Africa*. Agosto '91.
- 65. Signore A., Chianelli M., Fiore V., Pozzilli P., Andreani D.: L'immunoscintigrafia nella diagnosi delle endocrinopatie autoimmuni. Settembre '91.
- 66. Gentilomi G.A.: Sonde genetiche in microbiologia. Ottobre '91.
- 67. Santini G.F., Fornasiero S., Mucignat G., Besaglia G., Tarabini-Castellani G. L., Pascoli L.: Le sonde di DNA e la virulenza batterica. Gennaio '92.
- 68. Zilli A., Biondi T.: Il piede diabetico. Febbraio '92.





- 70. Bracco G., Dotti G., Pagliardini S., Fiorucci G.C.: Gli screening neonatali. Aprile '92.
- 71. Tavani A., La Vecchia C.: Epidemiologia delle patologie cardio e cerebrovascolari. Luglio '92.
- 72. Cordido F., Peñalva A., De la Cruz L. F., Casanueva F. F., Dieguez C.: L'ormone della crescita. Agosto '92.
- 73. Contu L., Arras M.: Molecole di membrana e funzione immunologica (I). Settembre '92.
- 74. Ferrara S.: Manuale di laboratorio I. Ottobre '92.
- 75. Gori S.: Diagnosi di laboratorio dei patogeni opportunisti. Novembre '92.
- 76. Ferrara S.: Manuale di laboratorio II. Gennaio '93.
- 77. Pinna G., Veglio F., Melchio R.: Ipertensione Arteriosa. Febbraio '93.
- 78. Alberti M., Fiori G.M., Biddau P.: I linfomi non Hodgkin. Marzo '93.
- 79. Arras M., Contu L.: Molecole di membrana e funzione immunologica (II). Aprile '93.
- 80. Amin R.M., Wells K.H., Poiesz B.J.: Terapia antiretrovirale. Maggio '93.
- 81. Rizzetto M.: L'epatite C. Settembre '93.
- 82. Andreoni S.: Diagnostica di laboratorio delle infezioni da lieviti. Ottobre '93.
- 83. Tarolo G.L., Bestetti A., Maioli C., Giovanella L.C., Castellani M.: Diagnostica con radionuclidi del Morbo di Graves-Basedow. Novembre '93.
- 84. Pinzani P., Messeri G., Pazzagli M.: Chemiluminescenza. Dicembre '93.
- 85. Hernandez L.R., Osorio A.V.: Applicazioni degli esami immunologici. Gennaio 94.
- 86. Arras M., Contu L.: *Molecole di Membrana e funzione immunologica. Parte terza: I Infociti* B. Febbraio '94.
- 87. Rossetti R.: Gli streptoccocchi beta emolitici di gruppo B (SGB). Marzo '94.
- 88. Rosa F., Lanfranco E., Balleari E., Massa G., Ghio R.: Marcatori biochimici del rimodel-lamento osseo. Aprile '94.
- 89. Fanetti G.: Il sistema ABO: dalla sierologia alla genetica molecolare. Settembre '94.
- 90. Buzzetti R., Cavallo M.G., Giovannini C.: Citochine ed ormoni: Interazioni tra sistema endocrino e sistema immunitario. Ottobre '94.
- 91. Negrini R., Ghielmi S., Savio A., Vaira D., Miglioli M.: Helicobacter pylori. Novembre '94.
- 92. Parazzini F.: L'epidemiologia della patologia ostetrica. Febbraio '95.
- 93. Proietti A., Lanzafame P.: Il virus di Epstein-Barr. Marzo '95.
- 94. Mazzarella G., Calabrese C., Mezzogiorno A., Peluso G.F., Micheli P, Romano L.: *Immunoflogosi nell'asma bronchiale*. Maggio '95.
- 95. Manduchi I.: Steroidi. Giugno '95.
- 96. Magalini S.I., Macaluso S., Sandroni C., Addario C.: Sindromi tossiche sostenute da principi di origine vegetale. Luglio '95.
- 97. Marin M.G., Bresciani S., Mazza C., Albertini A., Cariani E.: Le biotecnologie nella diagnosi delle infezioni da retrovirus umani. Ottobre '95.
- 98.La Vecchia C., D'Avanzo B., Parazzini F., Valsecchi M.G.: *Metodologia epidemiologica e sperimentazione clinica*. Dicembre '95.
- 99.Zilli A., Biondi T., Conte M.: Diabete mellito e disfunzioni conoscitive. Gennaio '96.
- 100.Zazzeroni F., Muzi P., Bologna M.: Il gene oncosoppressore p53: un guardiano del genoma. Marzo '96.
- 101.Cogato I. Montanari E.: La Sclerosi Multipla. Aprile '96.
- 102.Carosi G., Li Vigni R., Bergamasco A., Caligaris S., Casari S., Matteelli A., Tebaldi A.: *Malattie a trasmissione sessuale.* Maggio '96.





104. Marcante R., Dalla Via L.: Il virus respiratorio sinciziale. Luglio '96.

105. Giovanella L., Ceriani L., Roncari G.: Immunodosaggio dell'antigene polipeptidico tissutale specifico (TPS) in oncologia clinica: metodologie applicative. Ottobre '96.

106.Aiello V., Palazzi P., Calzolari E.: Tecniche per la visualizzazione degli scambi cromatici (SCE): significato biologico e sperimentale. Novembre '96.

107. Morganti R.: Diagnostica molecolare rapida delle infezioni virali. Dicembre '96.

108. Andreoni S.: Patogenicità di Candida albicans e di altri lieviti. Gennaio '97.

109. Salemi A., Zoni R.: Il controllo di gestione nel laboratorio di analisi. Febbraio '97.

110.Meisner M.: Procalcitonina. Marzo '97.

111. Carosi A., Li Vigni R., Bergamasco A.: Malattie a trasmissione sessuale (2). Aprile '97.

112.Palleschi G. Moscone D., Compagnone D.: *Biosensori elettrochimici in Biomedicina*. Maggio '97.

113. Valtriani C., Hurle C.: Citofluorimetria a flusso. Giugno '97.

114.Ruggenini Moiraghi A., Gerbi V., Ceccanti M., Barcucci P.: *Alcol e problemi correlati*. Settembre '97.

115.Piccinelli M.: Depressione Maggiore Unipolare. Ottobre '97.

116.Pepe M., Di Gregorio A.: Le Tiroiditi. Novembre '97.

117. Cairo G.: La Ferritina. Dicembre '97.

118.Bartoli E.: Le glomerulonefriti acute. Gennaio '98.

119.Bufi C., Tracanna M.: Computerizzazione della gara di Laboratorio. Febbraio '98.

120. National Academy of Clinical Biochemistry: Il supporto del laboratorio per la diagnosi ed il monitoraggio delle malattie della tiroide. Marzo '98.

121. Fava G., Rafanelli C., Savron G.: L'ansia. Aprile '98.

122.Cinco M.: La Borreliosi di Lyme. Maggio '98.

123. Giudice G.C.: Agopuntura Cinese. Giugno '98.

124.Baccini C.: Allucinogeni e nuove droghe (1). Luglio '98.

125.Rossi R.E., Monasterolo G.: Basofili. Settembre '98.

126. Arcari R., Grosso N., Lezo A., Boscolo D., Cavallo Perin P.: Eziopatogenesi del diabete mellito di tipo 1. Novembre '98.

127.Baccini C.: Allucinogeni e nuove droghe (11). Dicembre '98.

128.Muzi P., Bologna M.: Tecniche di immunoistochimica. Gennaio '99.

129.Morganti R., Pistello M., Vatteroni M.L.: *Monitoraggio dell'efficacia dei farmaci antivirali*. Febbraio '99.

130.Castello G., Silvestri I.: Il linfocita quale dosimetro biologico. Marzo '99.

131.AielloV., Caselli M., Chiamenti C.M.: *Tumorigenesi gastrica Helicobacter pylori - correlata*. Aprile '99.

132.Messina B., Tirri G., Fraioli A., Grassi M., De Bernardi Di Valserra M.: *Medicina Termale e Malattie Reumatiche*. Maggio '99.

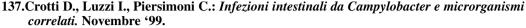
133.Rossi R.E., Monasterolo G.: Eosinofili. Giugno '99.

134.Fusco A., Somma M.C.: NSE (Enolasi Neurono-Specifica). Luglio '99.

135. Chieffi O., Bonfirraro G., Fimiani R.: La menopausa. Settembre '99.

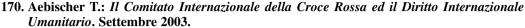
136.Giglio G., Aprea E., Romano A.: Il Sistema Qualità nel Laboratorio di Analisi. Ottobre '99.





- 138. Giovanella L.: Tumori Neuroendocrini: Diagnosi e fisiopatologia clinica. Dicembre '99.
- 139. Paladino M., Cerizza Tosoni T.: *Umanizzazione dei Servizi Sanitari: il Case Management*. Gennaio 2000.
- 140.La Vecchia C.: Come evitare la malattia. Febbraio 2000.
- 141.Rossi R.E., Monasterolo G.: Cellule dendritiche. Marzo 2000.
- 142.Dammacco F.: Il trattamento integrato del Diabete tipo 1 nel bambino e adolescente (I). Aprile 2000.
- 143.Dammacco F.: Il trattamento integrato del Diabete tipo 1 nel bambino e adolescente (II). Maggio 2000.
- 144. Croce E., Olmi S.: Videolaparoscopia. Giugno 2000.
- 145.Martelli M., Ferraguti M.: AllergoGest. Settembre 2000.
- 146. Giannini G., De Luigi M.C., Bo A., Valbonesi M.: TTP e sindromi correlate: nuovi orizzonti diagnostici e terapeutici. Gennaio 2001.
- 147.Rassu S., Manca M.G., Pintus S., Cigni A.: L'umanizzazione dei servizi sanitari. Febbraio 2001.
- 148. Giovanella L.: I tumori della tiroide. Marzo 2001.
- 149.Dessì-Fulgheri P., Rappelli A.: L'ipertensione arteriosa. Aprile 2001.
- 150. The National Academy of Clinical Biochemistry: Linee guida di laboratorio per lo screening, la diagnosi e il monitoraggio del danno epatico. Settembre 2001.
- 151. Dominici R.: Riflessioni su Scienza ed Etica. Ottobre 2001.
- 152.Lenziardi M., Fiorini I.: Linee guida per le malattie della tiroide. Novembre 2001.
- 153. Fazii P.: Dermatofiti e dermatofitosi. Gennaio 2002.
- 154.Suriani R., Zanella D., Orso Giacone G., Ceretta M., Caruso M.: Le malattie infiammatorie intestinali (IBD) Eziopatogenesi e Diagnostica Sierologica. Febbraio 2002.
- 155. Trombetta C.: Il Varicocele. Marzo 2002.
- 156.Bologna M., Colorizio V., Meccia A., Paponetti B.: Ambiente e polmone. Aprile 2002.
- 157. Correale M., Paradiso A., Quaranta M.: I Markers tumorali. Maggio 2002.
- 158. Loviselli A., Mariotti S.: La Sindrome da bassa T3. Giugno 2002.
- 159. Suriani R., Mazzucco D., Venturini I., Mazzarello G., Zanella D., Orso Giacone G.: *Helicobacter Pylori: stato dell'arte*. Ottobre 2002.
- 160. Canini S.: Gli screening prenatali: marcatori biochimici, screening nel 1° e 2° trimestre di gravidanza e test integrato. Novembre 2002.
- 161. Atzeni M.M., Masala A.: La β-talassemia omozigote. Dicembre 2002.
- 162. Di Serio F.: Sindromi coronariche acute. Gennaio 2003.
- 163. Muzi P., Bologna M.: Il rischio di contaminazione biologica nel laboratorio biosanitario. Febbraio 2003.
- 164. Magni P., Ruscica M., Verna R., Corsi M.M.: Obesità: fisiopatologia e nuove prospettive diagnostiche. Marzo 2003.
- 165. Magrì G.: Aspetti biochimici e legali nell'abuso alcolico. Aprile 2003.
- 166. Rapporto dello Hastings Center: Gli scopi della medicina: nuove priorità. Maggio 2003.
- 167. Beelke M., Canovaro P., Ferrillo F.: Il sonno e le sue alterazioni. Giugno 2003.
- 168. Macchia V., Mariano A.: Marcatori tumorali nel cancro della vescica. Luglio 2003.
- 169. Miragliotta G., Barra Parisi G., De Sanctis A., Vinci E.: *La Turbercolosi Polmonare: Diagnostica di Laboratorio*. Agosto 2003.





- 171. Martino R., Frallicciardi A., Tortoriello R.: Il manuale della sicurezza. Ottobre 2003.
- 172. Canigiani S. e Volpini M.: Infarto acuto del miocardio: biochimica del danno cellulare e marcatori di lesione. Novembre 2003.
- 173. La Brocca A. Orso Giacone G. Zanella D. Ceretta M.: Laboratorio e clinica delle principali affezioni tiroidee. Dicembre 2003.
- 174. Savron G.: Le Fobie. Gennaio 2004.
- 175. Paganetto G.: Evoluzione storica del rischio di patologie umane per contaminazione chimica ambientale. Febbraio 2004.
- 176. Giovanella L.: Iperparatiroidismo e tumori paratiroidei. Marzo 2004.
- 177. Severino G., Del Zompo M.: Farmacogenomica: realtà e prospettive per una "Medicina Personalizzata". Aprile 2004.
- 178 Arigliano P.L.: Strategie di prevenzione dell'allergia al lattice nelle strutture sanitarie. Maggio 2004.
- 179. Bruni A.: Malattia di Alzheimer e Demenza Frototemporale. Giugno 2004.
- 180. Perdelli F., Mazzarello G., Bassi A.M., Perfumo M., Dallera M.: Eziopatogenesi e diagnostica allergologica. Luglio 2004.
- 181. Franzoni E., Gualandi P. Pellegrini G.: I disturbi del comportamento alimentare. Agosto 2004.
- 182. Grandi G., Peyron F.: La toxoplasmosi congenita. Settembre 2004.
- 183. Rocca D.L., Repetto B., Marchese A., Debbia E.A: *Patogeni emergenti e resistenze batteriche*. Ottobre 2004.
- 184. Tosello F., Marsano H.: Scientific English Handout. Novembre 2004.
- 185. La Brocca A., Orso Giacone G., Zanella D.: Ipertensione arteriosa secondaria: clinica e laboratorio. Dicembre 2004.
- 186. Paganetto G.: Malattie Neoplastiche: dalla Paleopatologia alle Fonti Storiche. Gennaio 2005.
- 187. Savron G.: La sindrome dai mille tic: il disturbo di Gilles de la Tourette. Febbraio 2005.
- 188. Magrì G., Baghino E., Floridia M., Ghiara F.: Leishmania. Marzo 2005.
- 189. Lucca U., Forloni G., Tiraboschi P., Quadri P., Tettamanti M., PasinaL.: *Invecchiamento, deterioramento cognitivo e malattia di Alzheimer*. Aprile 2005.
- 190. Volpe G., Delibato E., Orefice L., Palleschi G.: Tossinfezioni alimentari e metodiche recenti ed innovative per la ricerca dei batteri patogeni responsabili. Maggio 2005.
- 191. Mazzarello M.G., Albalustri G., Audisio M., Perfumo M., L. Cremonte G.: Aerobiologia ed allergopatie. Giugno 2005.
- 192. Scalabrino G., Veber D., Mutti E.: Nuovi orizzonti biologici per la vitamina B12. Luglio 2005.
- 193. Zepponi E.: Guida pratica per gli utenti del laboratorio analisi. Settembre 2005.
- 194. Faricelli R., Esposito S., Martinotti S.: *La sindrome da anticorpi anti-fosfolipidi*. Ottobre 2005.
- 195. Baccini C., Bezzi F., Conti M., Tazzari V.: Doping e antidoping nello sport. Novembre 2005.
- 196. Lozzi M.: La Mediazione pacifica dei conflitti. Una risorsa socio-relazionale in ambito medico-sanitario. Dicembre 2005.
- 197. Bracco G.: Progettare un Laboratorio di Analisi. Gennaio 2006.





- 199. Commissione Tecnica sul Rischio Clinico: Risk management in Sanità. Il problema degli errori. Marzo 2006
- 200. Casati G., Marchese E., Roberti V., Vichi M.C.: La gestione dei processi clinico assistenziali per il miglioramento delle prassi. Aprile 2006.
- 201. Zanella D., Ceretta M., Orso Giacone G.: Peptidi natriuretici: nuove frontiere in cardiologia? Maggio 2006.
- 202. Cicala M., Dal Lago U., Vinci P., Maggiorotti M.: L'accusa di malpractice in ambito medico. Giugno 2006.
- 203. Martino R.: Manuale Qualità UNI EN ISO 9001. Luglio 2006.
- 204. Mazzarello M.G., Arata M., Perfumo M., Marchese A., Debbia E.A.: *Tubercolosi e micobatteri*. Settembre 2006.
- 205. Matrullo R.: Anoressia: la negazione della sessualità come difesa narcisistica. Ottobre 2006.
- 206. Crotti D.: Le parassitosi intestinali ed uro-genitali. Novembre 2006.
- 207. Orso Giacone G., Zanella D., Ceretta M.: Il referto interpretativo in infettivologia. Dicembre 2006.
- 208. Baghino E., Magrì G., Nicoletti L., Novaro G., Vignale C., Mazzei C.: Stato dell'arte delle aneuploidie fetali, dall'indagine clinica prenatale alla diagnosi anatomopatologica. Gennaio 2007.
- 209. Mazzarello M.G., Brunetti R., Perfumo M., Torriglia A.M., Montresor G.: Principali Tecniche Analitiche in uso nei Laboratori di Analisi Chimico Cliniche e Microbiologiche. Febbraio 2007.
- 210. Orso Giacone G., Zanella D., Ceretta M.: Celiachia dalla A alla Z. Marzo 2007.
- 211. Cingolani M., Sparviero E.: Decidere ora per allora: il testamento biologico (dichiarazioni anticipate di trattamento). Aprile 2007.
- 212. Barletta G., Pastacaldi V., Peracino A.P.: La misura dei processi nella medicina di laboratorio. Maggio 2007.
- 213. Rassu S., Masia L., Delussu P., Chessa P., Demartis M.G., Moroso G.: *Manuale per il supporto vitale di base e la defibrillazione precoce (BLS-D)*. Giugno 2007.
- 214. Anchisi R., M. Gambotto Dessy: Il Burnout del personale sanitario. Marzo 2008.
- 215. Gulletta E., Orrico F., Foti D.P.: Clinical Governance nel Laboratorio Biomedico. Aprile 2008.
- 216. Rochira V., Scaltriti S., Zirilli L., Carani C.: Il ruolo degli estrogeni nel maschio. Maggio 2008.
- 217. Gulletta E., Foti D.P., Corsi M.M., Galliera E.: Citochine e Chemochine. Giugno 2008.
- 218. Zambotto F.M.: La biotecnologia transgenica utilizzata nella produzione degli alimenti di origine vegetale. Settembre 2008
- 219. Cavallini M.: Tecniche di Ringiovanimento del viso. Ottobre 2008
- 220. Morra A., Odetto L., Bozza C., Bozzetto P., Agostinis S., Bariona M.: Compendio di Medicina delle Grandi Emergenze. Novembre 2008





I volumi disponibili su Internet nel sito www.medicalsystems.it sono riportati in nero mentre in grigio quelli non ancora disponibili su Internet.



Inoltre sono disponibili un limitato numero di copie di alcuni numeri del Caleidoscopio che ormai sono "storiche". Qualora mancassero per completare la collana potete farne richiesta al collaboratore Medical Systems della Vostra zona. I numeri sono: Caleidoscopio 14, 18, 33, 40, 48, 49, 50, 54, 65, 68, 84, 100, 106, 118, 121, 126, 129, 130, 131, 132, 133, 134. I volumi verranno distribuiti sino ad esaurimento e non verranno ristampati se non in nuove edizioni.





Caleidoscopio

Rivista mensile di Medicina anno 25, numero 221

Direttore Responsabile

Sergio Rassu Tel. mobile 338 2202502 E-mail: sergiorassu@yahoo.it

Responsabile Ufficio Acquisti Giusi Cunietti

Servizio Abbonamenti Maria Grazia Papalia Laura Cecchi

Progettazione e Realizzazione

Restless Architect of Human Possibilities s.a.s. Via Pietro Nenni, 6 07100 Sassari

Tel. 079 270464 http://rahp80.googlepages.com/

Consulenti di Redazione

Giancarlo Mazzocchi ed Angelo Maggio

Segretaria di Direzione Maria Speranza Giola

EDITORE

... il futuro ba il cuore antico MEDICAL SYSTEMS SpA

Via Rio Torbido, 40 16165 Genova (Italy) Tel. 010 83401 Numero Verde 800 801005 (senza prefisso); Telefax 010/8340310-809070.

Internet URL: http://www.medicalsystems.it La Medical Systems pubblica anche le seguenti riviste: Caleidoscopio Illustrato, Caleidoscopio Letterario, Guida Pratica Immulite[®], Guida Pratica Città, Journal of Clinical Ligand Assay, Pandora, Tribuna Biologica e Medica.

Stampa

Tipolitografia Nuova ATA Via Gelasio Adamoli, 281 - Genova Tel. 010 513120 - Fax 010 503320 - info@nuovaata.com - www.nuovaata.com

Registrazione Tribunale di Genova n. 34 del 31/7/1996 Iscrizione al Registro Nazionale della Stampa nº 2661 del 2 Settembre 1989 Iscrizione al Registro degli Operatori di Comunicazione (ROC) nº 1188

> Finito di stampare: Dicembre 2008 Sped. in Abb. Post. 45%

Pubblicazione protetta a norma di legge dall'Ufficio proprietà letteraria, artistica e scientifica della Presidenza del Consiglio dei Ministri, dedicata all'aggiornamento professionale continuo e riservata ai medici.

> Caleidoscopio viene anche letto e rilanciato da: "L'ECO DELLA STAMPA" Via Compagnoni, 28 - Milano

ISSN 0394 3291

Caleidoscopio



Antonio Morra, Lorenzo Odetto, Carmine Bozza, Pierangelo Bozzetto, Stefano Agostinis, Marioluca Bariona



Direttore Responsabile Sergio Rassu

220

... il futuro ha il cuore antico MEDICAL SYSTEMS SpA





In caso di mancato recapito, pregasi ritornare al mittente che pagherà la tassa dovuta.



)

21-01-2009 9:29:11