

UN TOOL PER LA LETTURA DELLE IMMAGINI DI ESPRESSIONE DA MICROARRAY

Raffaella Rizzi(*), Luciano Milanese(*), Luigi Mussio(**)

(*) CNR-ITBA, via Fratelli Cervi 93 20090 Segrate (Milano), milanese@itba.mi.cnr.it

(**) DIAR-Politecnico di Milano, Piazza Leonardo da Vinci 32 20133 Milano

Riassunto

Nel campo della genetica e della biologia molecolare, i saggi di espressione genica su *microarray* costituiscono un passo in avanti verso la comprensione del comportamento dei geni in un organismo vivente e della causa primaria delle malattie. Le sequenze geniche *target*, depositate sul *microarray*, reagiscono per ibridazione con gli RNA messaggeri estratti dalle popolazioni cellulari oggetto del saggio. Le sostanze fluorescenti che marcano gli mRNA sviluppano un insieme di *spot* di segnale che, letti da dispositivi di scansione, vengono trasformati in immagine digitale. La lettura dell'immagine permette di estrarre il dato di espressione relativo ad ogni *target* del *microarray*. *SpotView* è un *software* sviluppato a tale scopo. Mediante lettura attraverso griglie regolari di aree di interesse, *SpotView* fornisce, per ogni *target*, un dato grezzo di espressione ed una stima del rumore locale di fondo per ogni banda dell'immagine.

Abstract

In the field of genetics and molecular biology, microarray-based gene expression assays are shedding light on the gene behavior in a living organism and on the onset of diseases. The deposited target sequences hybridize to the messenger RNAs from the involved cellular populations. The mRNA-labeling fluorescent substances develop signal spots, which are read by scanning devices and are transformed into a digital image. Image reading allows obtaining expression data for each microarray target. *SpotView* is a *Java*-written computer program developed to this aim. By means of grids of interest areas, *SpotView* produces raw expression data and estimates the local background for each target and for each signal band.

Introduzione

Il prodotto di un saggio di espressione su *microarray* è un'immagine digitale, il cui particolare *pattern* a *spot* di segnale deve essere recuperato se si vuole associare, ad ogni cellula coinvolta nell'esperimento, il profilo di espressione dei *target*. La lettura dell'immagine consente di recuperare dati grezzi di espressione da sottoporre a normalizzazione rispetto a *target* noti.

Dato il particolare *pattern* a *spot* di segnale, risulta immediato leggere i dati per mezzo di griglie regolari di aree di interesse, posizionate sull'immagine. Se la griglia segue il *pattern*, è possibile leggere un valore di integrazione delle intensità dei *pixel*, come dato grezzo di espressione, ed effettuare una stima del rumore di *background* locale allo *spot* (per ogni banda di segnale e, di conseguenza, per ogni popolazione cellulare). *SpotView* è un programma, scritto in codice *Java*, pensato a tale scopo. Esso permette all'utente di effettuare il posizionamento semiautomatico delle griglie regolari, introdotte sopra, e di effettuare la lettura del *pattern* di espressione dell'immagine.

Griglie di lettura del pattern di espressione

Una griglia di lettura delle immagini di espressione da microarray è, in sostanza, composta di aree di interesse disposte secondo uno schema a matrice. La loro regolarità ne semplifica notevolmente la definizione ed il posizionamento. Un'immagine può essere letta utilizzando più griglie indipendenti che vengono, in seguito, riunite in un unico profilo di lettura, come descritto nei paragrafi seguenti. Ogni singolo elemento della griglia è composto di un'area di interesse del segnale di espressione e di un'area di stima del rumore di fondo locale allo *spot*.

I parametri geometrici definiscono una griglia di lettura indipendentemente dalla sua posizione sul sistema immagine e, più in dettaglio, sono il numero di elementi per riga e per colonna, lo *spacing* lungo le righe e lungo le colonne, la forma e la dimensione delle due aree di interesse componenti il generico elemento della griglia. La figura 1 fornisce un esempio di griglia di lettura (utilizzata in *SpotView*) in cui le aree di interesse hanno forma circolare: l'area di interesse del segnale di espressione è il cerchio interno, mentre l'area di stima del rumore locale è la corona circolare trasparente. Per ognuno dei *target* del pannello, si possono produrre dati grezzi di espressione in ognuna delle popolazioni cellulari coinvolte nell'esperimento. Se A_s è l'area di interesse del segnale di espressione, $s(i, b)$ è l'intensità del segnale che è funzione del pixel i e della banda b , si può trovare un valore di integrazione I_b dato dalla seguente relazione:

$$I_b = \sum_{i \in A_s} s(i, b) \quad [1]$$

Inoltre si suppone che il rumore di fondo abbia un valore costante nell'area di interesse del segnale di espressione. Tale valore viene stimato come la media dell'intensità del segnale nell'area di stima del rumore. Se A_n è l'area di stima del rumore di fondo, C_n il numero di *pixel* in A_n , il rumore medio è dato dalla seguente relazione:

$$\bar{n}_b = \frac{1}{C_n} \sum_{i \in A_n} s(i, b) \quad [2]$$

Il rumore di fondo dello *spot* generico risulta quindi:

$$N_b = \bar{n}_b \times A_s \quad [3]$$

Sottraendo il rumore di fondo, così stimato, dal valore in [1], il dato di integrazione I_b è purificato dal rumore e si ottiene il dato grezzo di espressione E_b :

$$E_b = I_b - N_b \quad [4]$$

Posizionamento di una griglia di lettura

La posizione della griglia viene data nel sistema immagine che ha il *pixel* (0,0) nell'angolo superiore sinistro, l'asse X lungo le righe verso destra e l'asse Y lungo le colonne verso il basso (figura 1). I parametri di posizione sono le coordinate (x,y) del centro dell'elemento in alto a sinistra e l'angolo di rotazione oraria a partire dalla posizione positiva dell'asse X . L'elemento attorno a cui la griglia ruota è l'ancora della griglia e può essere qualsiasi elemento di essa. La rotazione può avvenire rigidamente attorno all'ancora, o attorno alla colonna passante per l'ancora, o infine alla riga passante per l'ancora (figura 2). Il posizionamento della griglia è compiuto seguendo una procedura semiautomatica consistente in un posizionamento manuale approssimato, sulla base della conoscenza del processo di acquisizione dell'immagine, e in un successivo affinamento

automatico, basato sul *pattern* di segnale mediante massimizzazione (o minimizzazione) di una misura di segnale, calcolata per la griglia intera. Il raffinamento automatico aggiorna progressivamente i parametri della griglia fino a raggiungere una situazione d'equilibrio permanente, cosicché la misura di segnale raggiunge il suo massimo (o minimo) valore. Più precisamente, il posizionamento automatico aggiorna, in una serie ordinata e predefinita, ogni grado di libertà della griglia: due traslazioni lungo gli assi X e Y , una rotazione, due espansioni, due contrazioni (figura 3). Il gruppo di movimenti scelto (numero, tipo e ordine), la banda di segnale e la misura di segnale determinano la posizione finale di equilibrio della griglia. Il processo di aggiustamento per il generico movimento M induce la griglia a compiere un'oscillazione, limitata in un range R , attorno alla posizione corrente. Esisterà una posizione di massimo (o minimo) locale dalla quale ripartire con una nuova oscillazione dello stesso movimento, finché l'equilibrio permanente sarà stato raggiunto per il movimento M . *SpotView* utilizza, come misura di segnale, l'integrazione totale delle intensità dei *pixel* nelle aree di interesse del segnale di espressione, data dalla seguente relazione:

$$I_{Tb} = \sum_{j \in G} \sum_{i \in A_{sj}} s_j(i, b) \quad [5]$$

dove G è la collezione degli elementi della griglia e A_{sj} è l'area di interesse del segnale di espressione del j -esimo elemento.

Letture del pattern di espressione: *SpotView*

SpotView è un *software* scritto in codice *Java* che recupera i dati grezzi di espressione da un'immagine da *microarray*. *SpotView* è strutturato in tre sezioni di base: visualizzazione dei profili di segnale per facilitare il posizionamento manuale delle griglie, posizionamento semiautomatico delle griglie, lettura e registrazione dei dati. Una quarta sezione riguarda l'*enhancement* dell'immagine.

Il posizionamento semiautomatico si compone della fase manuale e della fase di aggiustamento della posizione. La prima fase viene compiuta attraverso un *wizard* (figura 5) che guida l'utente attraverso il posizionamento del centro dell'elemento superiore sinistro della griglia (tramite semplice *click* del *mouse* nel punto desiderato), la definizione della forma e delle dimensioni delle aree di composizione degli elementi, la definizione del numero di colonne e di righe e del loro *spacing*. Ogni griglia aggiunta all'immagine è gestita attraverso un *Grid Manager* (figura 6) che permette di impostare i colori degli elementi significativi delle griglie, di muovere manualmente la griglia (traslazione lungo X e Y e rotazione), di aggiustare automaticamente la posizione sulla base della banda di segnale e della sequenza dei gradi di libertà prescelti, di rimuovere le griglie aggiunte, di leggere il *pattern* di espressione per una banda di segnale, di salvare i dati letti in formato importabile in *database* relazionali, di salvare e di ricaricare un profilo di lettura per una riletture del *pattern*.

Un profilo di lettura è una sequenza ordinata di griglie di lettura alle quali l'utente associa una precisa modalità di lettura. La modalità di lettura della generica griglia definisce l'ordine in cui gli elementi (o *spot*) devono essere presi e dipende dal verso secondo cui la griglia viene scandita (ad esempio, prima da sinistra a destra e poi dall'alto verso il basso). Ogni elemento di una griglia del profilo ottiene un indice progressivo, necessario all'utente per collegare i dati salvati all'informazione contenuta in un eventuale *database* esterno.

Un esempio di lettura del *pattern* di figura 5 è riportato nella tabella 7.

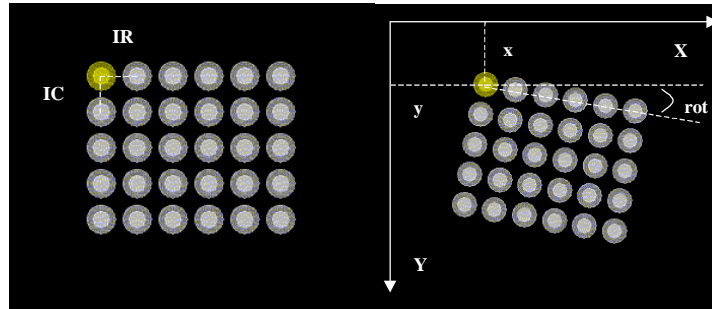


Figura 1 – Griglia di lettura a elementi circolari utilizzata da SpotView. I parametri di posizione sono sulla destra

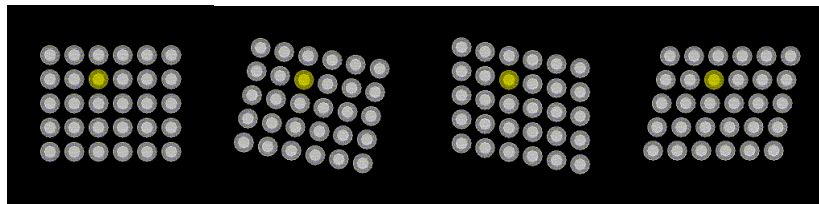


Figura 2 – Modalità di rotazione di una griglia di lettura attorno all'elemento d'ancora

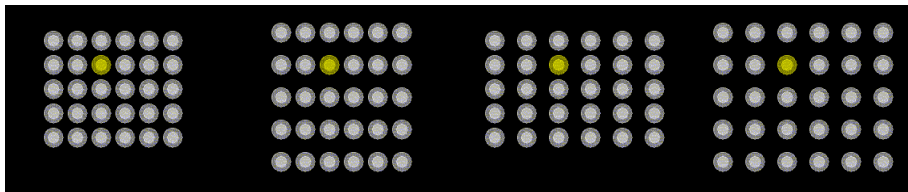


Figura 3 – Espansioni (contrazioni) di una griglia: lungo le righe e lungo le colonne

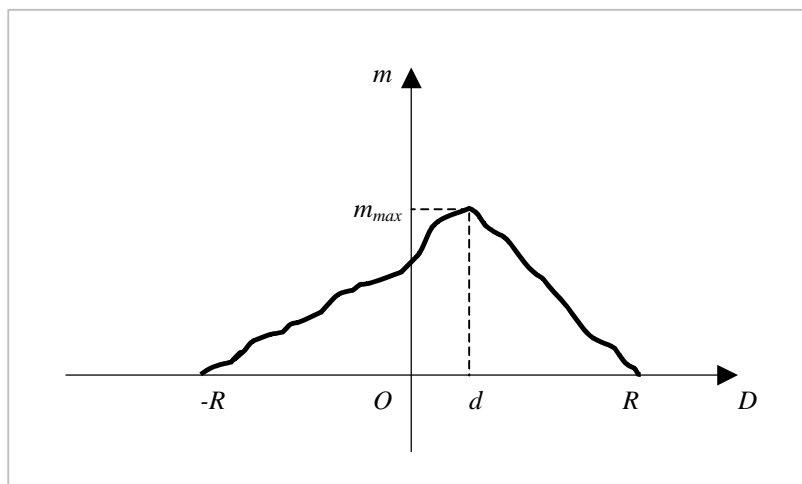


Figura 4 – Diagramma della misura di segnale, associata all'intera griglia, in funzione dell'oscillazione del generico movimento nel range R attorno alla posizione corrente (origine O); l'asse X rappresenta la distanza D dell'oscillazione dalla posizione corrente; l'asse Y rappresenta la misura di segnale m per la griglia e per una banda dell'immagine; il valore m_{max} fornisce la distanza di oscillazione d alla quale aggiornare la posizione della griglia per il movimento

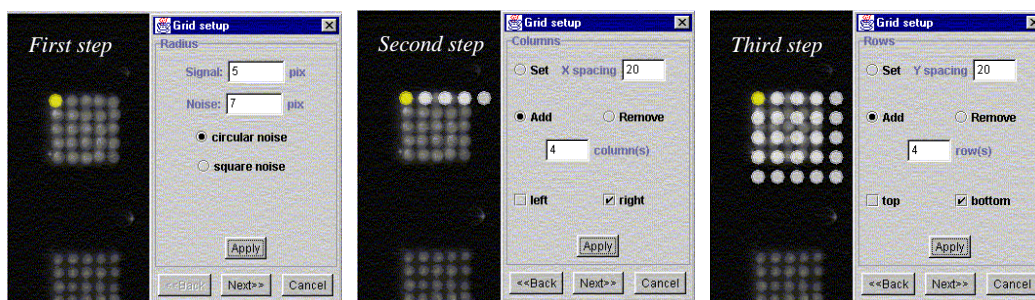


Figura 5 – Aggiunta di una nuova griglia di lettura all'immagine

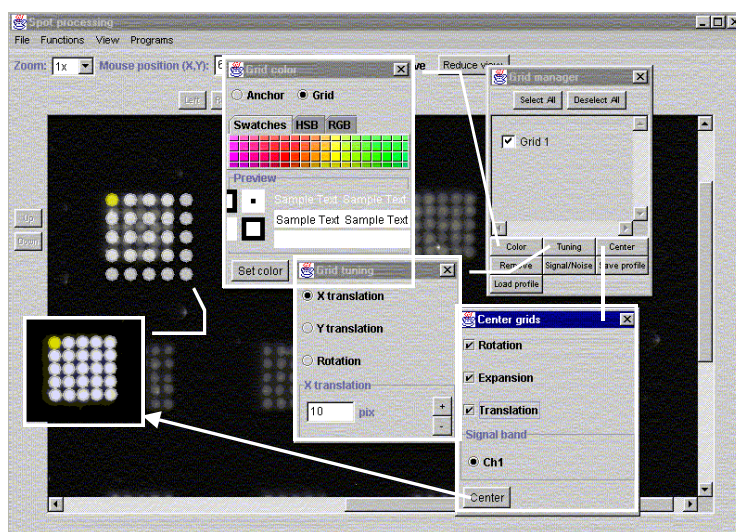


Figura 6 – Funzioni offerte dal Grid Manager

Indice dello spot	Valore di integrazione nell'area di interesse di segnale	Stima del rumore di fondo	Dato grezzo di espressione	Saturazione
1	2102816	1369422	733394	0.0
2	1997968	1410170	587798	0.0
3	2047104	1367803	679301	0.0
4	2058272	1303617	754655	0.0
5	1828368	1245416	582952	0.0
6	2361712	1421119	940593	0.0
7	2238768	1475194	763574	0.0
8	2149344	1509081	640263	0.0
9	2148288	1345145	803143	0.0
10	1989664	1245416	744248	0.0
11	2212368	1485265	727103	0.0
12	2280880	1495336	785544	0.0
13	2289120	1510610	778510	0.0
14	2287472	1414497	872975	0.0
15	1965216	1309603	655613	0.0
16	2121440	1416890	704550	0.0
17	2278480	1428055	850425	0.0
18	2280704	1510610	770094	0.0
19	2352976	1414497	938479	0.0
20	2003424	1271489	731935	0.0
21	2169632	1416890	752742	0.0
22	2154320	1421276	733044	0.0
23	2203984	1461696	742288	0.0
24	2208288	1405037	803251	0.0
25	1669232	1271489	397743	0.0

Tabella 7 – Risultati ottenuti dalla lettura della griglia presentata nella figura 5. Il rumore di background è stimato come la mediana dei valori dello spot corrente e degli otto spot attorno per eliminare l'effetto di spot vicini eccessivamente luminosi

Conclusioni

SpotView è un programma per *computer* che permette il recupero dei dati di espressione grezza da immagini da *microarray*. La regolarità delle griglie di lettura rende semplice e immediata la loro definizione e il loro posizionamento. Inoltre il salvataggio dei dati, in formato importabile in *database* relazionali, consente la costruzione di un sistema di informazioni accessibili e riproducibili. Infine, esso fornisce statistiche per misurare l'affidabilità del singolo *spot* nel fornire dati di espressione. Mediante tale programma, è possibile la lettura di un elevato numero di immagini da *microarray* con determinate caratteristiche: *pattern* di segnale regolare con *spot* disposti in righe e colonne.

Al momento attuale, *SpotView* presenta problemi aperti tra i quali il raffinamento della stima del rumore di fondo locale allo *spot* (soprattutto per quanto riguarda l'influenza del segnale da *spot* vicini), l'estensione a diversi formati di immagine, oltre al BMP e TIFF, e l'indirizzamento dell'utente nella scelta delle dimensioni delle aree di interesse, senza dover fare affidamento su conoscenze del processo di acquisizione dell'immagine.

Ringraziamenti

Il presente lavoro deriva da una Tesi di Dottorato di Ricerca in Scienze Geodetiche e Topografiche svolta presso l'ITBA-CNR (Istituto di Tecnologie Biomediche Avanzate) del CNR di Segrate (MI) sul tema Sviluppo di metodologie per l'analisi di immagini di *microarray*.

Bibliografia

Bernard K., Auphan N., Granjeaud S., Victorero G., Schmitt-Verhulst A. M., Jordan B. R., Nguyen C. (1996), "Multiplex messenger assay: simultaneous, quantitative measurement of expression for many genes in the context of T cell activation", *Nucleic Acid Research*, 24: 1453-1443.

Chen J. J. W., Wu R., Yang P-C., Huang J-Y., Sher Y-P., Han M-H., Kao W-C., Lee P-J., Chiu T. F., Chang F., Chu Y-W., Wu C-W., Peck K. (1998), "Profiling expression patterns and isolating differentially expressed genes by cDNA *microarray* system with colorimetry detection", *Genomics*, 51: 313-324.

Lockhart D. J., Dong H., Byrne M. C., Follettie M. T., Gallo M. V., Chee M. S., Mittmann M., Wang C., Kobayashi M., Horton H., Brown E. L. (1996), "Expression monitoring by hybridization to high-density oligonucleotide arrays", *Nature Biotechnology*, 14: 1675-1680.

Schena M., Shalon D., Heller R., Chai A., Brown P. O., Davis R. W. (1996), "Parallel human genome analysis: *microarray*-based expression monitoring of 1000 genes", *Proceedings of the National Academy of Sciences USA*, 93: 10614-10619.

Umbaugh, S. E. (1998). *Computer vision and image processing – a practical approach using CVIPtools*. Prentice Hall PTR.

Wodicka L., Dong H., Mittmann M., Ho M-H., Lockhart D. J. (1997), "Genome-wide expression monitoring in *Saccharomyces cerevisia*", *Nature Biotechnology*, 15: 1359-1367.