

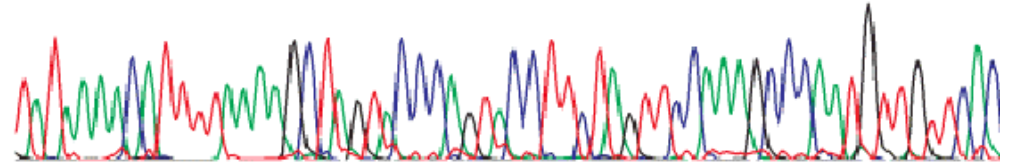
**MICROARRAYS:
CARATTERISTICHE,
APPLICAZIONI E
LIMITI**

- Cosa è un microarrays un po di storia
- La tecnologia: arrays **spotted** e sintetizzati *in situ* (spotted, Agilent, GeneChips Affymetrix, Illumina)
- Disegno dei **probes** e marcatura del **target**
- Cenni su analisi dei dati e validazione
- Applicazioni: analisi su scala genomica e ritratti molecolari (applicazioni allo studio dei tumori)
- Limiti: controllo trascrizionale e post-trascrizionale, quantificazione comparativa non assoluta
- Microarrays e splicing alternativo
- Non solo espressione genica (SNPs e genomi)
- miRNA microarrays e protein microarrays

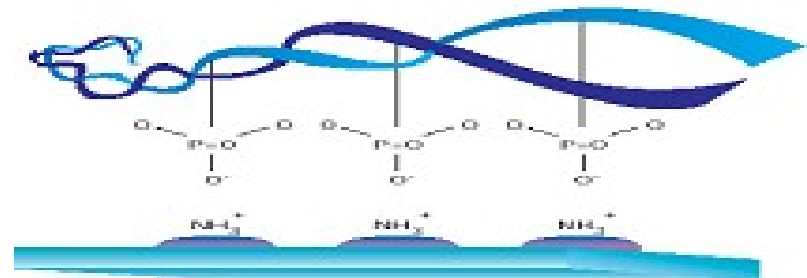
Microarray: una tecnologia post-genomica

genomica

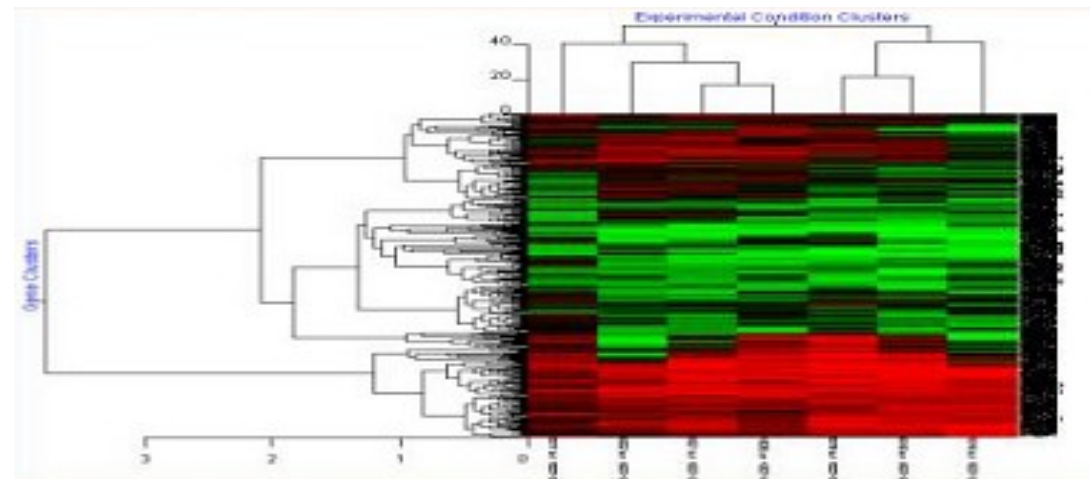
TATAAAACATTTTAAAAAGC TAGTAGCCAGTAGCTTCTAGTTCCAAAGCCCAATGTTGTTGAC
140 150 160 170 180 190 200



chimica delle superfici



bioinformatica

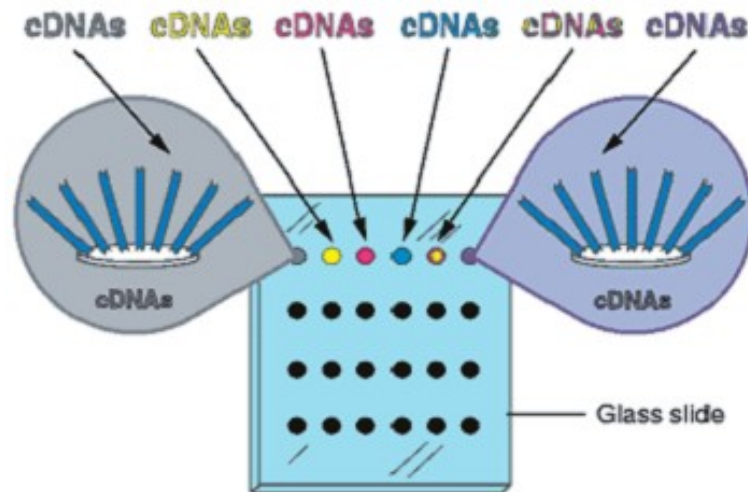


Che cos'è un microarray

- potente tecnologia: consente l'analisi comparativa (e simultanea) dei livelli di attività di migliaia di geni
- in cosa consiste: ibridazione ad alta specificità fra

1 "PROBE"

sequenze geniche (DNA amplificato o oligonucleotidi sintetizzati *in silico*) immobilizzate su superfici solide (vetro)

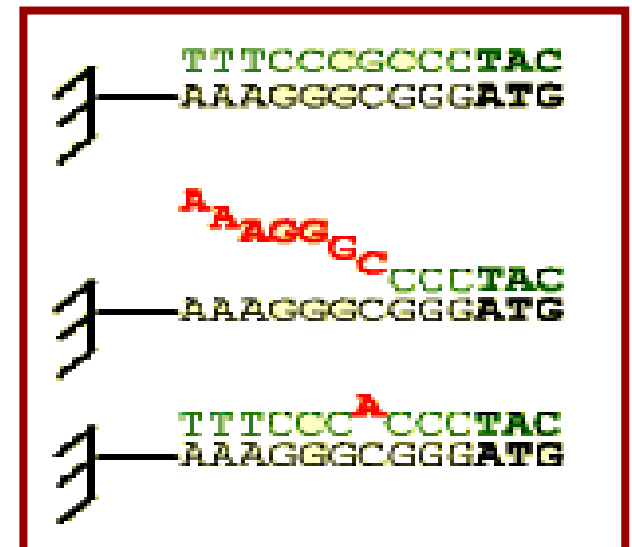
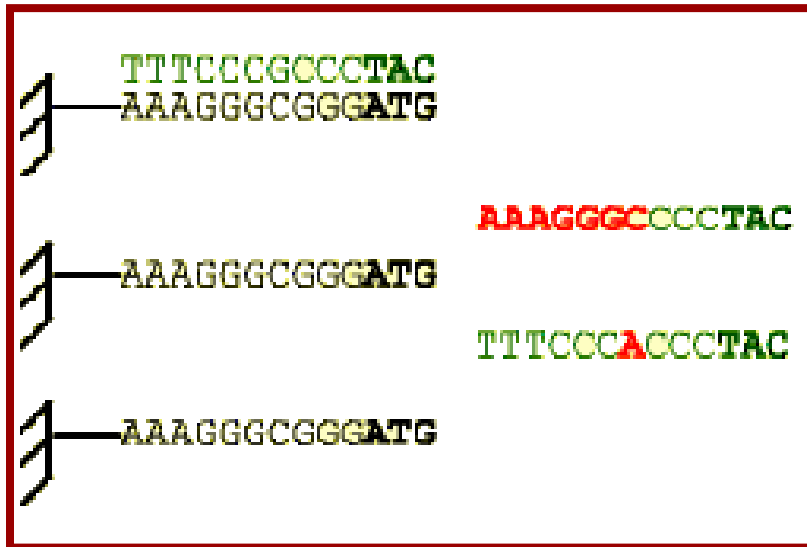


2 "TARGET"

miscela di ibridazione contenente acidi nucleici marcati (cDNA fluorescente o cRNA biotinilato)

metodica basata su alta specificita' di ibridazione

la specificità è influenzata dalle condizioni sperimentali



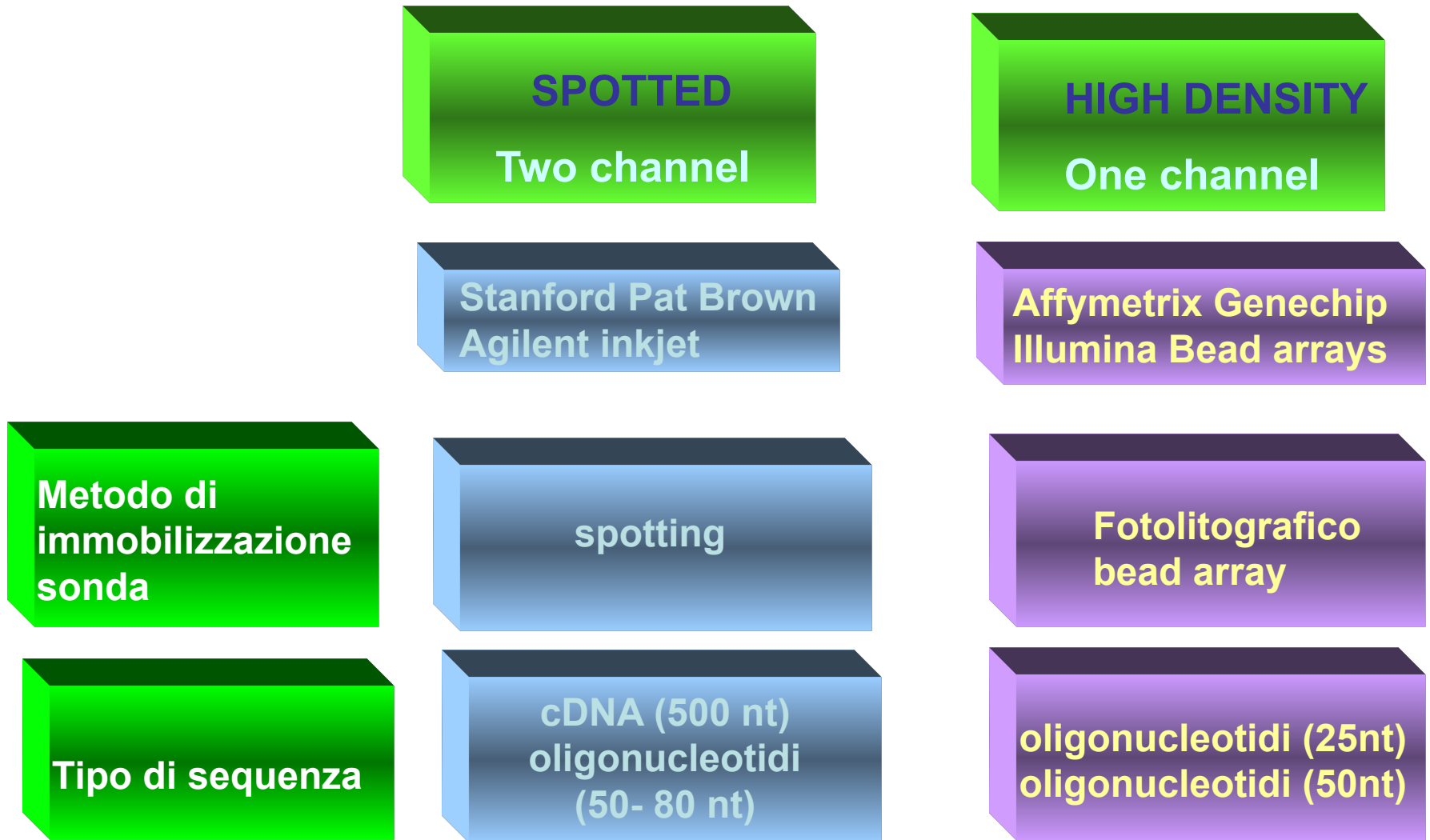
alta stringenza



bassa stringenza

temperatura
miscela di ibridazione
lunghezza del probe

TIPI DI ARRAYS



Microarrays commerciali danno risultati più riproducibili (alto controllo qualità industriale)

MICROARRAYS: alcuni aspetti importanti

- ✓ il problema biologico in esame deve essere connesso alla trascrizione
- ✓ il disegno sperimentale richiede particolare cura: repliche biologiche (3 o più) e repliche tecniche

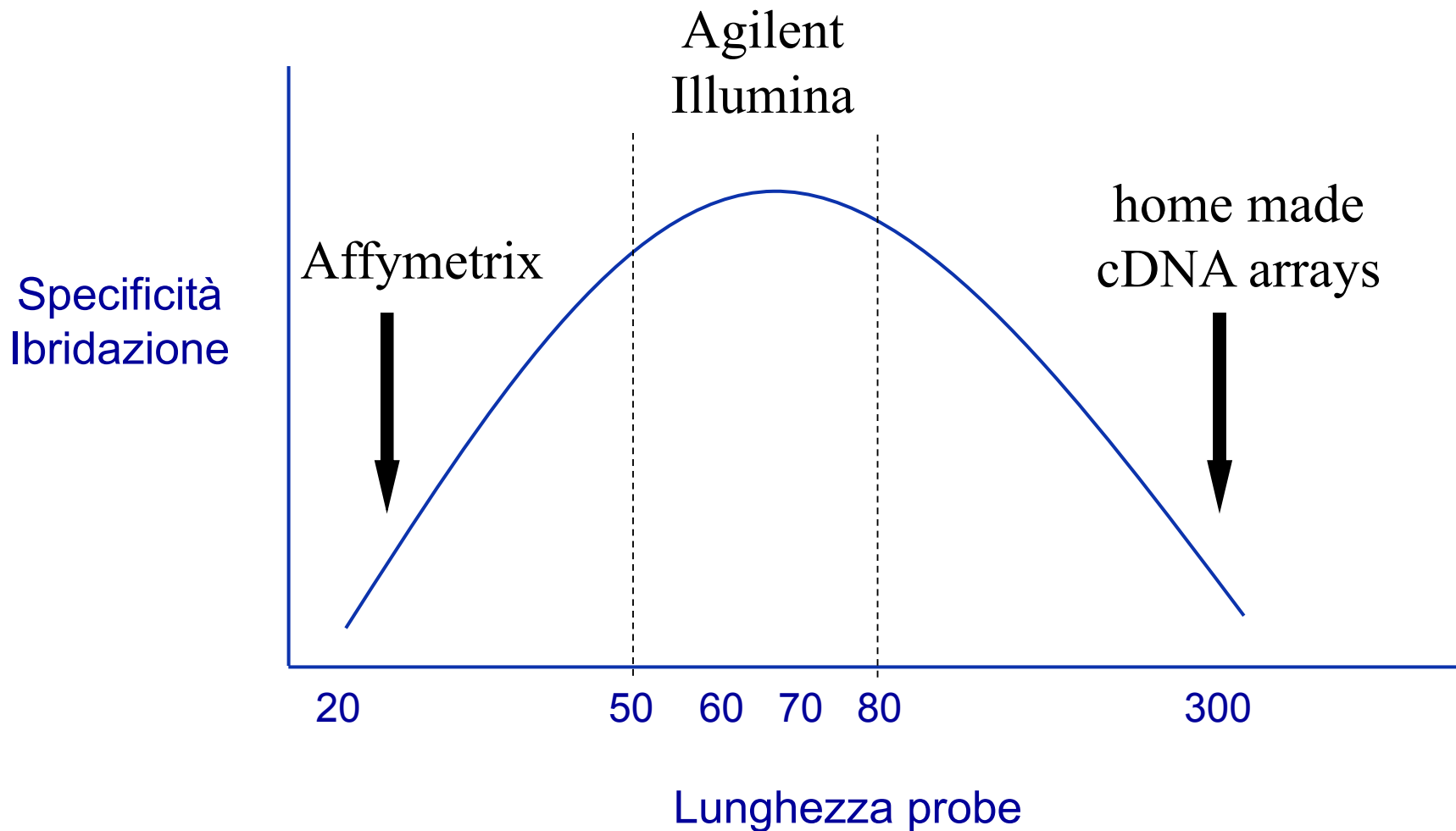
Identificazione di:

- geni differenzialmente espressi (profilo trascrizionale di 2 o più punti sperimentali)
- marcatori diagnostici e prognostici (profilo trascrizionale di campioni normali e patologici)
- pathways trascrizionali (esperimenti in serie temporali, *time-course*)

MICROARRAYS: il disegno dei probes

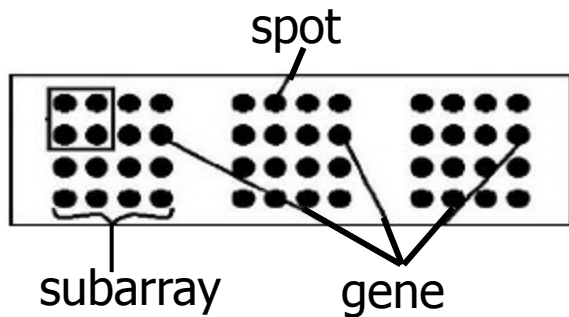
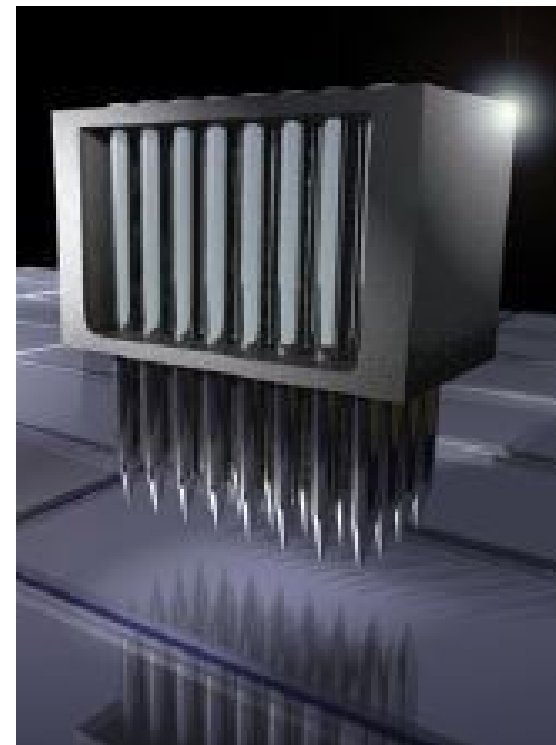
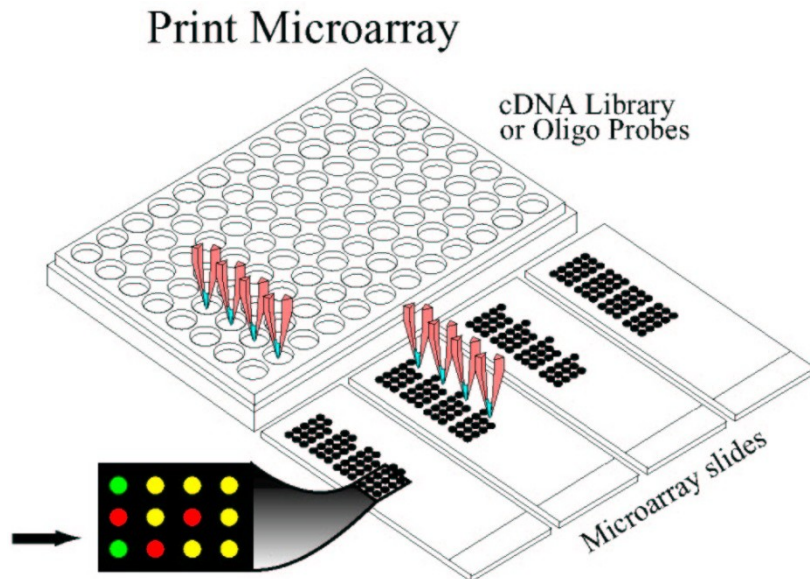
- il disegno degli oligonucleotidi richiede la conoscenza delle sequenze codificanti i geni di interesse
- ogni oligonucleotide deve identificare un determinato gene in maniera univoca (non assomigliare a nessuna altra sequenza del genoma in esame)
- gli oligonucleotidi non devono contenere sequenze ripetute
- il disegno dei probes avviene generalmente grazie a specifici algoritmi ed all'ausilio della bioinformatica
- i probes vengono spesso validati sperimentalmente

Probes fra i 50 e gli 80 mers danno le migliori specificita' di ibridazione



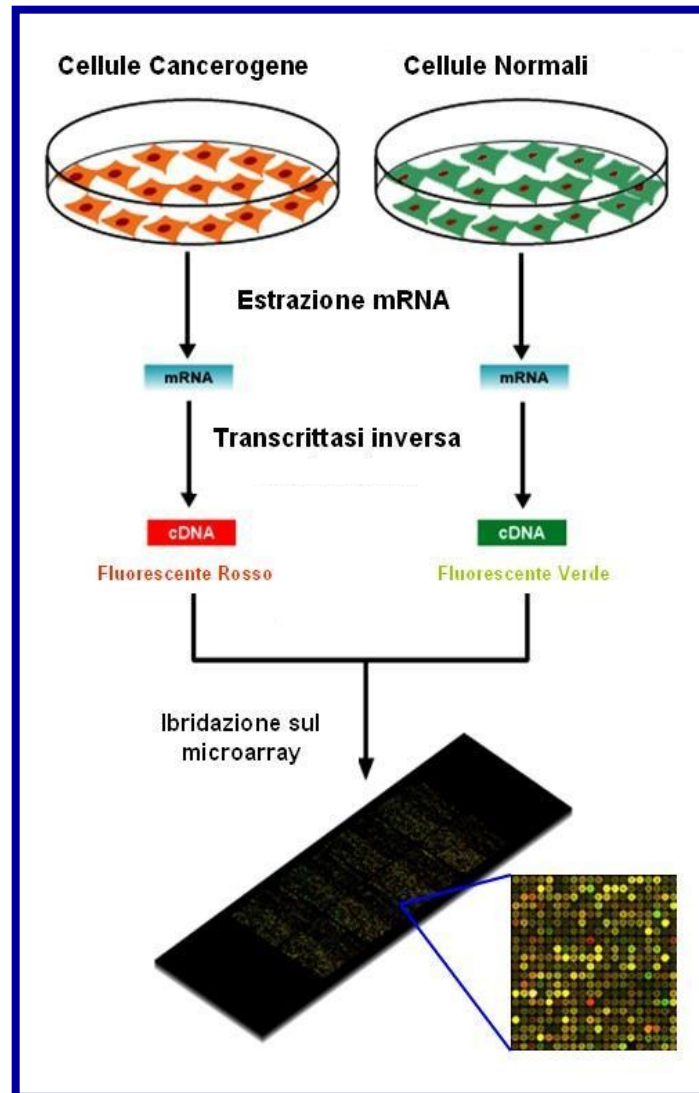
Microarray two channels – Spotted Agilent

SPOTTED o PRINTED ARRAYS: MICRODEPOSIZIONE



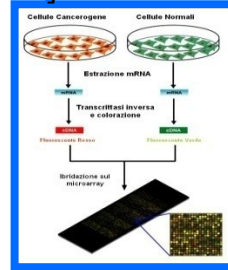
Densità moderata
~10000-30000

SPOTTED ARRAYS (due canali)



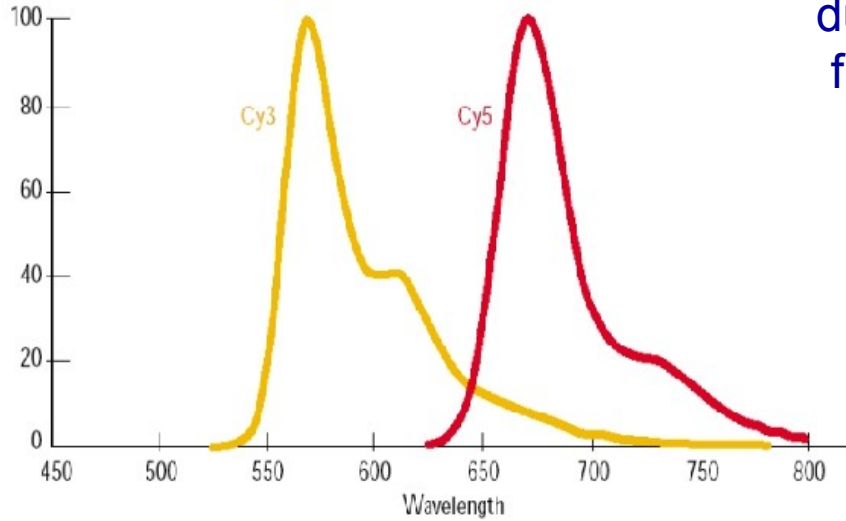
SPOTTED ARRAYS: SCANSIONE LASER

spotted



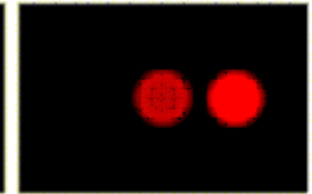
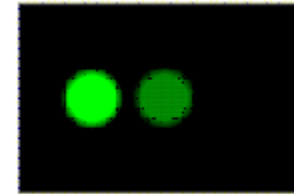
Emission Spectra

Emissioni diverse dei due coloranti fluorescenti



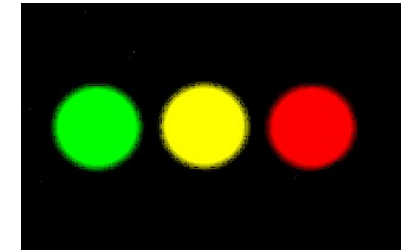
Cy3

Cy5



canale verde

canale rosso



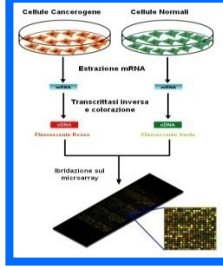
Cy3 (assorbe a ~550, emette a ~570, VERDE)
Cy5 (assorbe a ~650, emette a ~670, ROSSO)

PARAMETRI CRITICI

- Lo **scanner** è importante (genera il segnale che deve essere quantificato)
- **Software per analisi di immagine** critico per la produzione di valori affidabili di intensità

SPOTTED ARRAYS: DISEGNO SPERIMENTALE

spotted

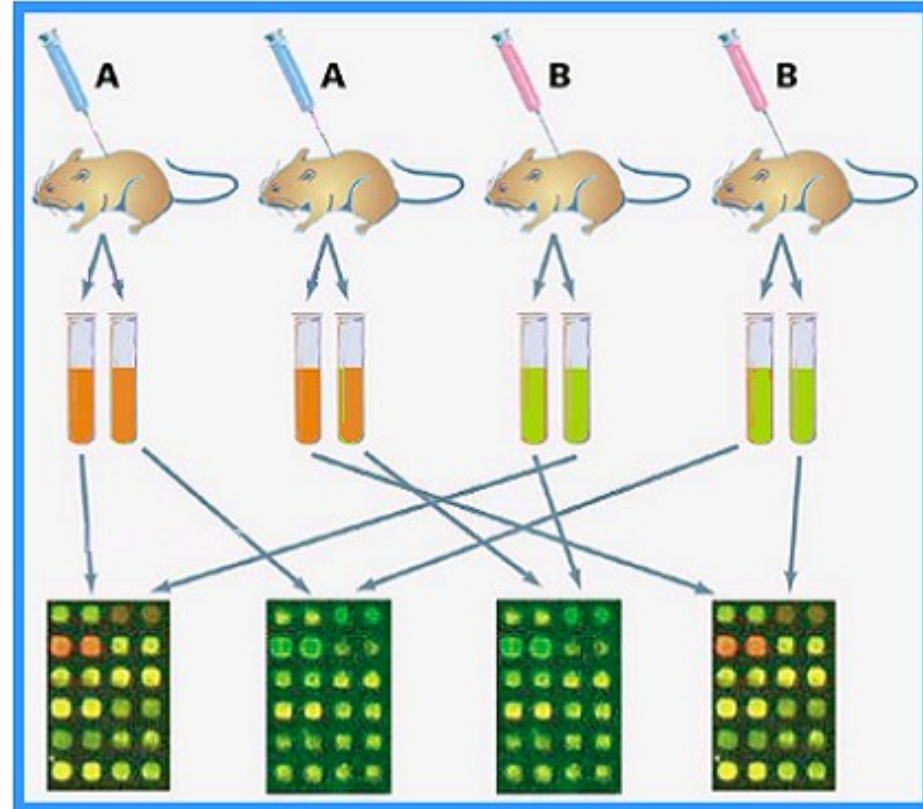
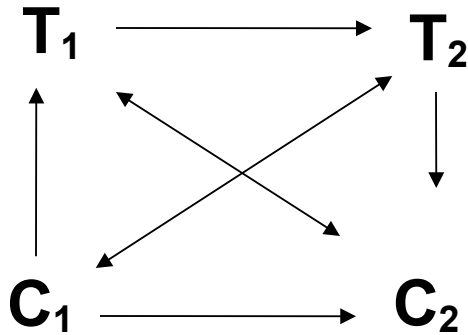


scelta del sistema biologico

scelta delle repliche
tecniche (numero, tipo)

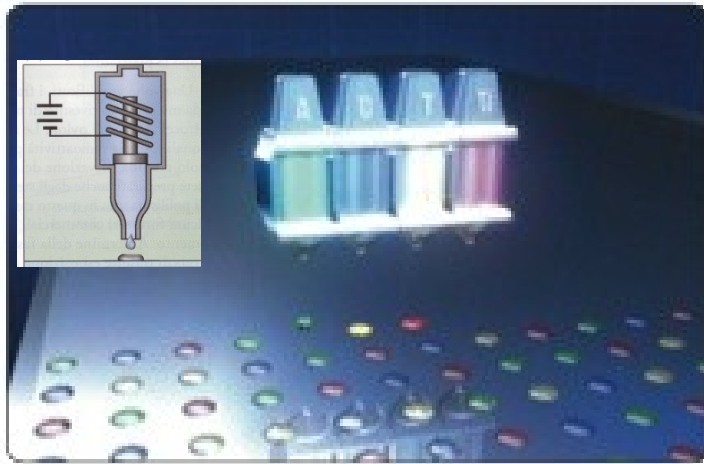
confronto tra due trattamenti

- replicato biologico
- replicato tecnico



Spesso esperimenti circolari in modo da acquisire più informazioni possibili per identificare con ragionevole sicurezza la presenza di geni differenzialmente espressi

SPOTTED ARRAYS: LA TECNOLOGIA INKJET AGILENT

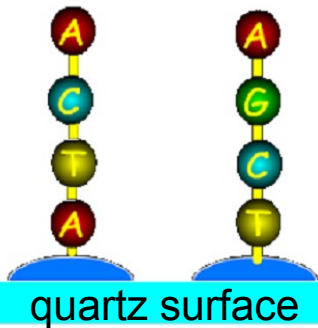


- Tecnologia tipo inkjet: no contatto con i vetrini.
- Probes estesi in situ mediante rounds ripetuti di spotting: prodotto finale oligo di 60-mer
- Microarrays con minori difetti, quindi più uniformi e consistenti.

**Alta Densità
fino a ~250000**

**Microarray one channel – GeneChip
Affymetrix**

MICROARRAYS AFFYMETRIX: GENECHIPS



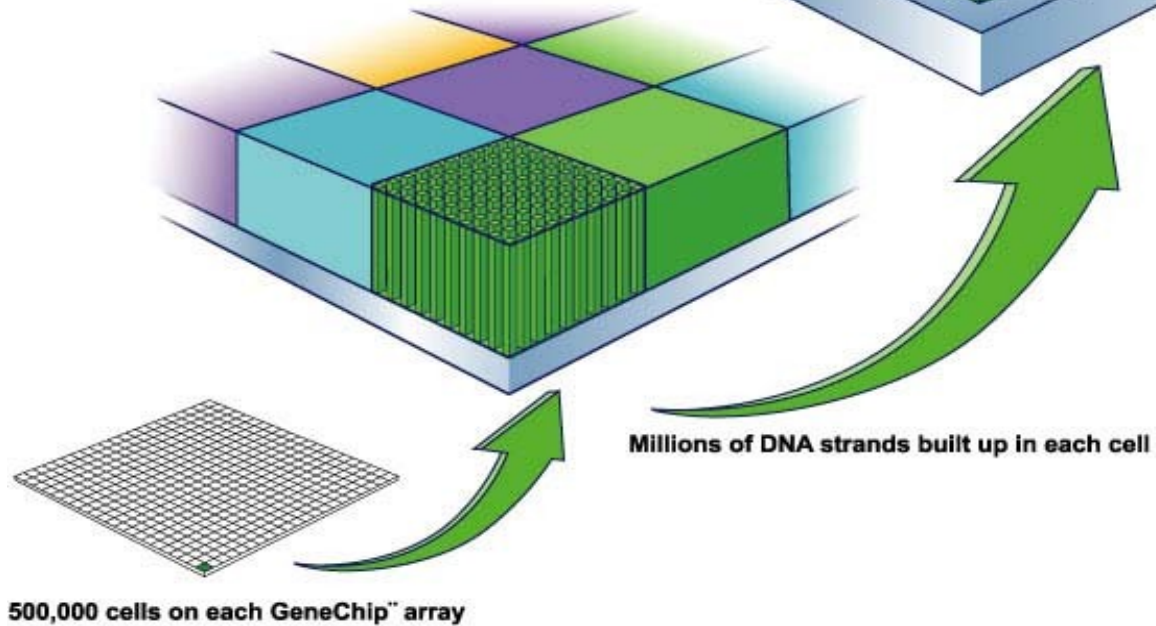
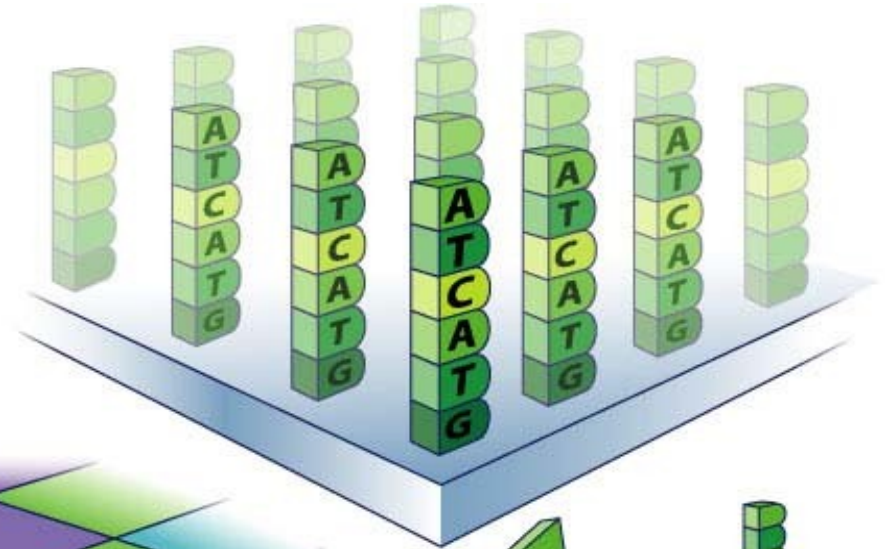
- Sintesi fotolitografica su supporto di quarzo
- Nucleotidi inseriti in successione (oligo 25-mer)
- Reazioni successive di deprotezione ed aggiunta guidate da una maschera fotolitografica.
- Probe sets: vari oligo 25-mer disegnati su uno stesso gene in posizioni diverse
- Oligo perfect match e mismatch
- Tecnologia a singolo canale: campioni di mRNA processati separatamente invece che a coppie
- Misurazioni di espressione genica combinando informazioni da più ibridazioni separate.
- Genechip 1.28 x 1.28 cm

Alta Densità ~500000 (fino a 10^6)

GENECHIPS ARRAYS



1.28 cm
1.28 cm
Actual size of GeneChip™



500,000 cells on each GeneChip™ array

Millions of DNA strands built up in each cell

Actual strand = 25 base pairs

GENECHIPS ARRAYS: PROBE SET

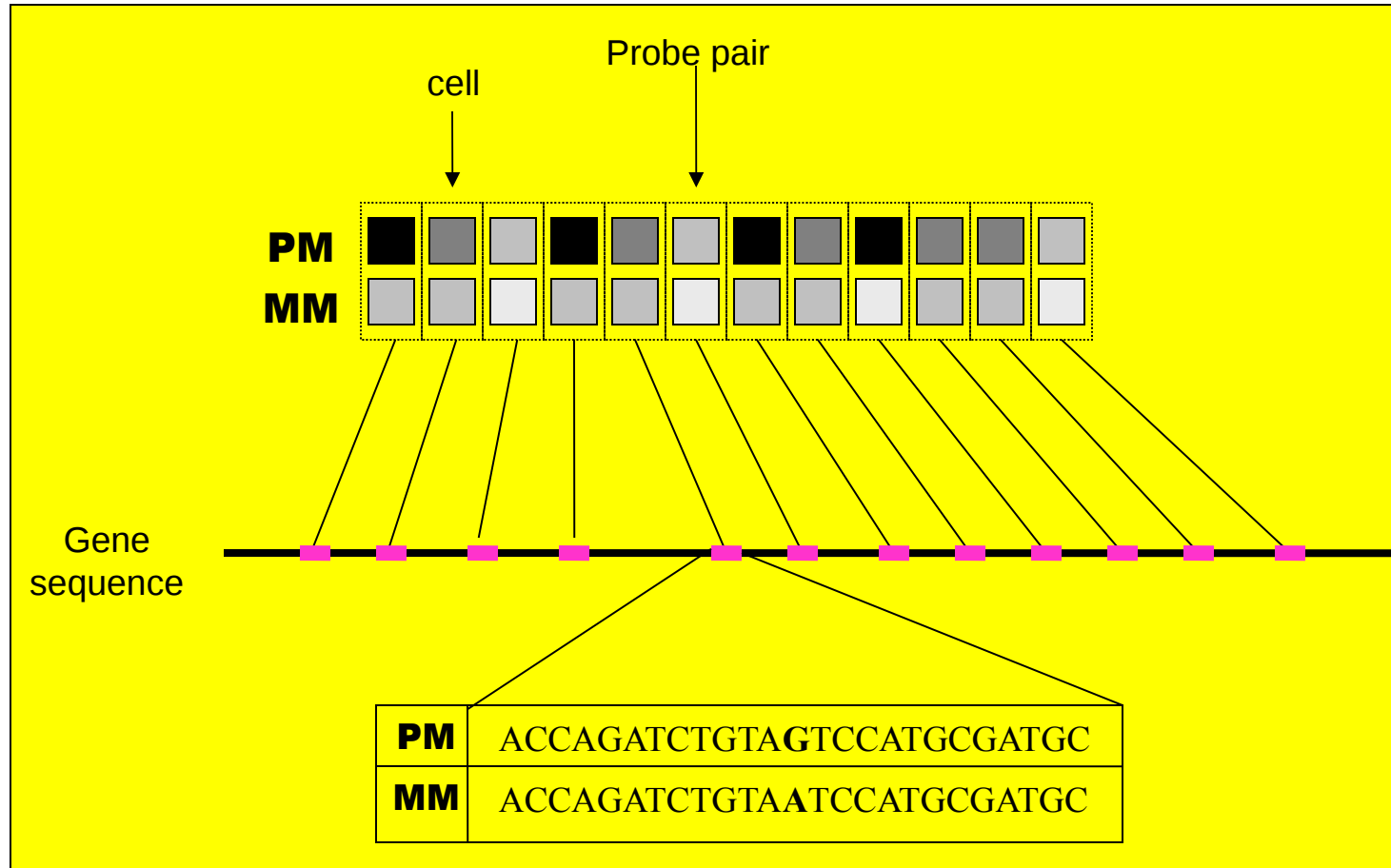
RIDONDANZA. Uso di oligo multipli con sequenza differente ma disegnati per ibridare in zone differenti al 3' di uno stesso RNA



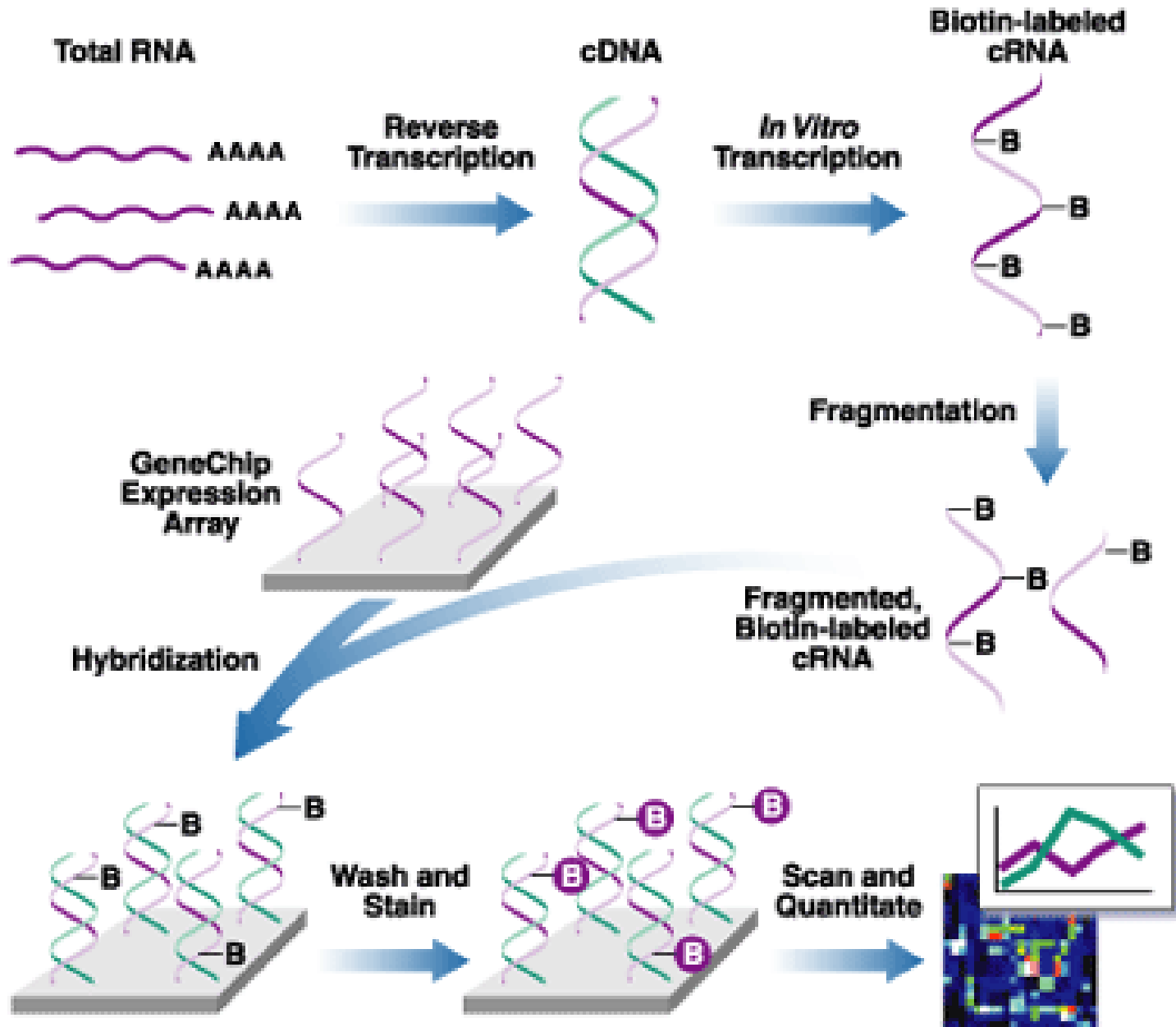
presenza di segnali multipli per ciascun gene

- maggiore accuratezza
- mitiga effetti cross-ibridazione
- riduce falsi positivi

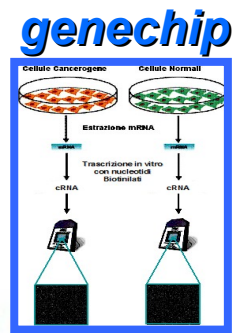
Probe set (Affymetrix)



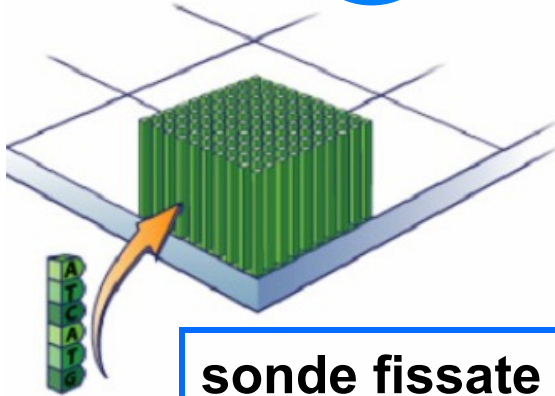
GENECHIPS: MARCATURA



GENECHIPS: IBRIDAZIONE E RIVELAZIONE



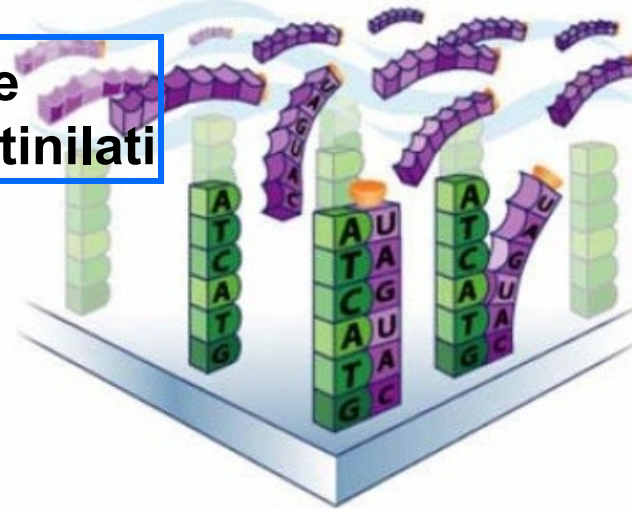
1



sonde fissate su una cella dell'array

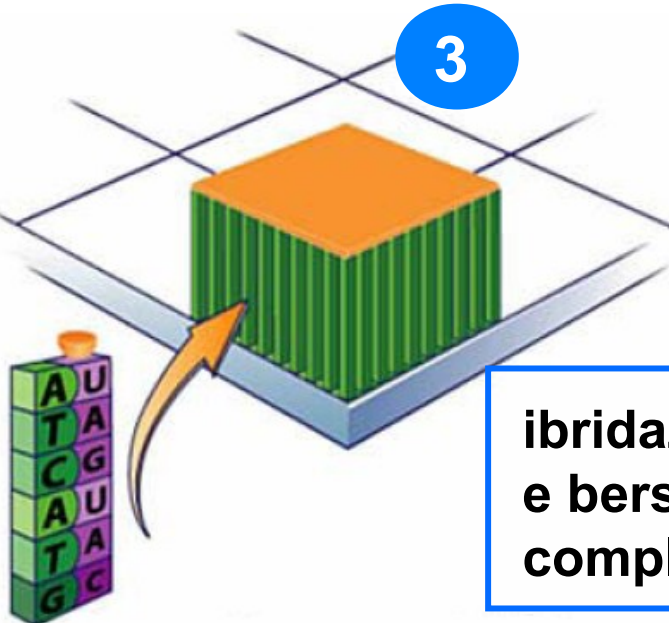
2

introduzione bersagli biotinilati



3

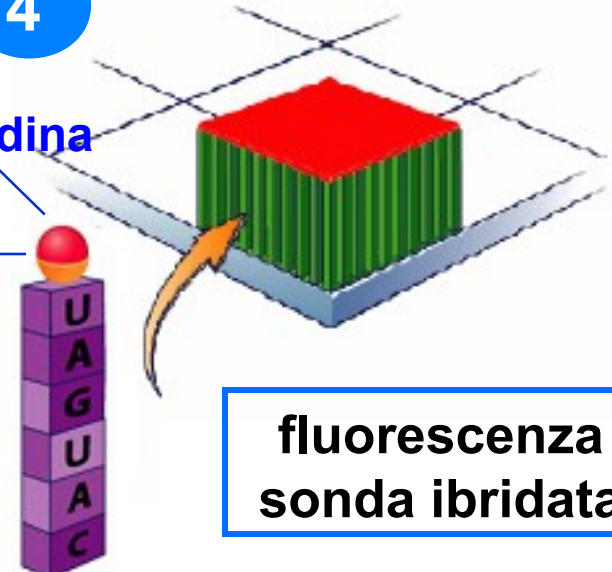
ibridazione tra sonda e bersaglio complementari

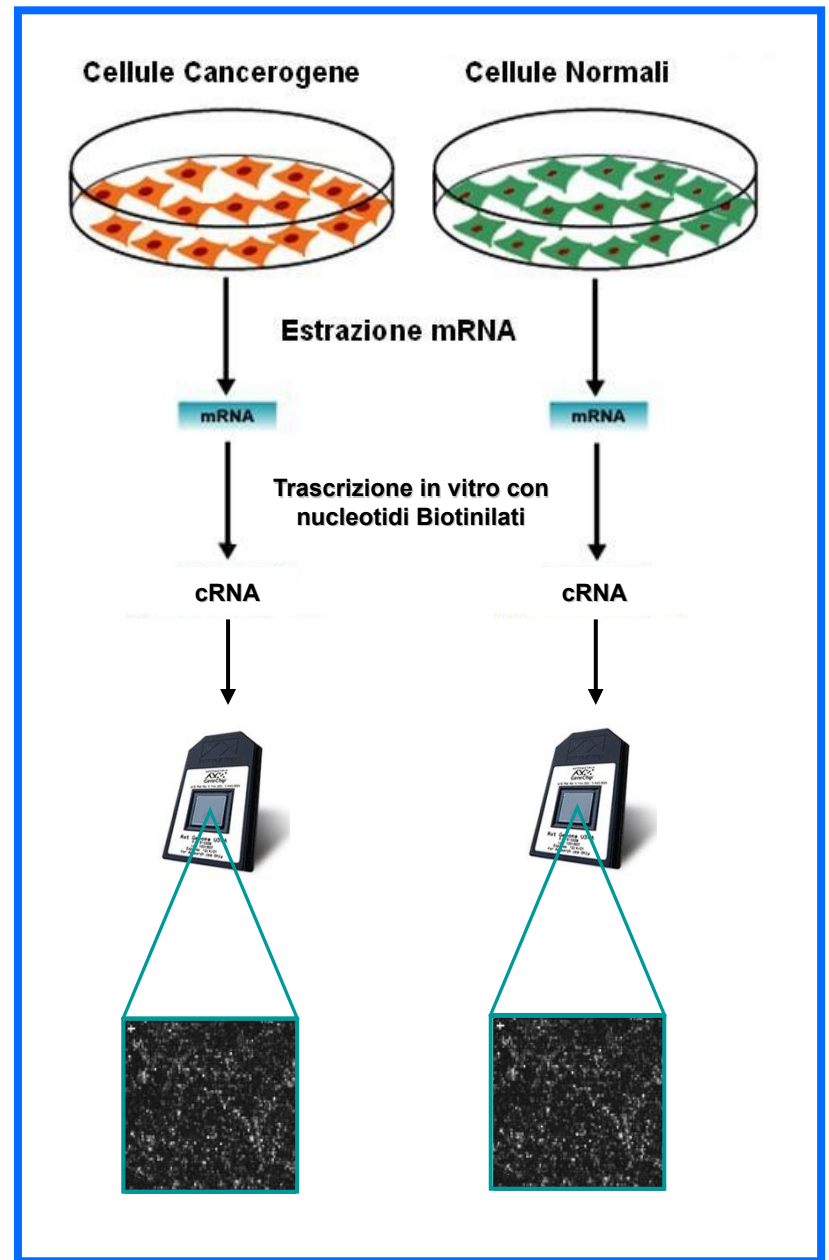
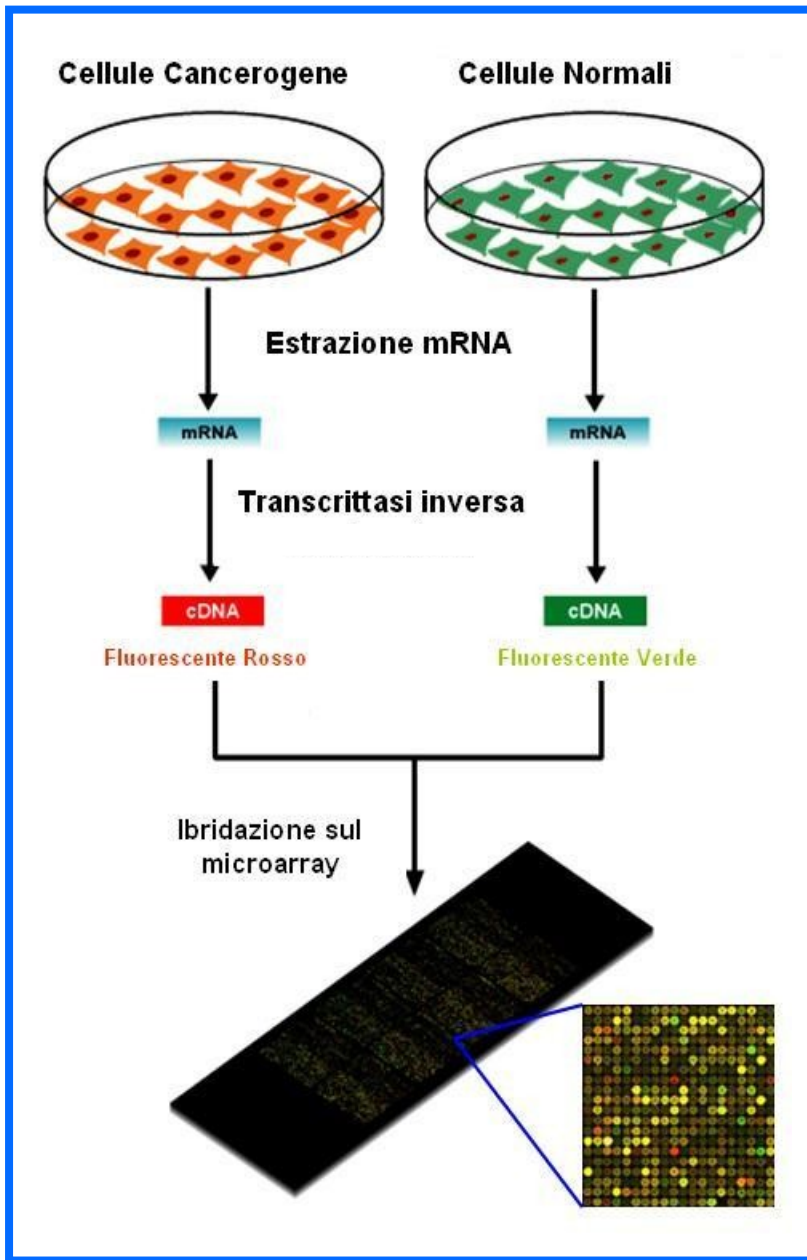


4

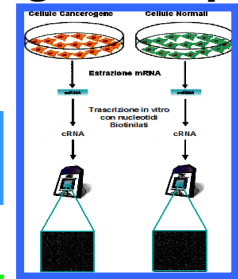
streptavidina
biotina

fluorescenza sonda ibridata





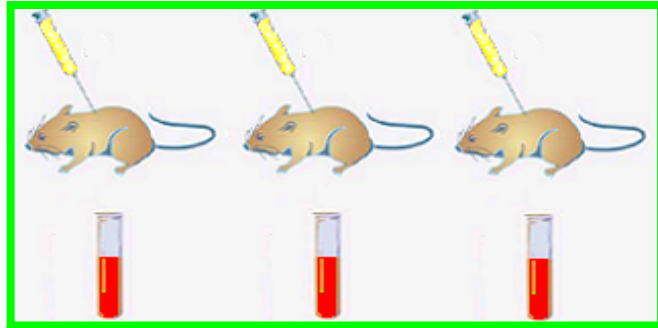
GENECHIPS: DISEGNO SPERIMENTALE



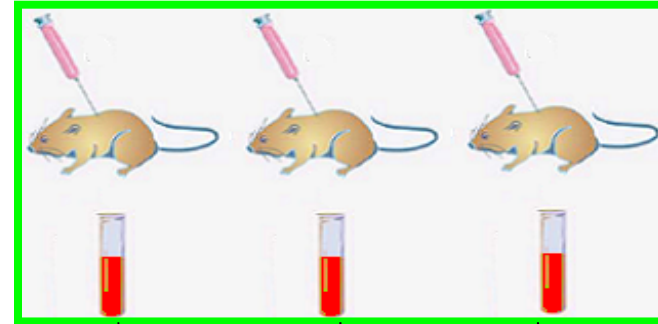
scelta del sistema biologico

scelta delle repliche (numero)

controllo



trattato



repliche biologiche



no repliche tecniche

- elevata riproducibilità dei risultati
- assenza di errore di misura (associato con la lettura del segnale fluorescente)
- misure multiple per ogni RNA

numero di repliche biologiche

- linee cellulari (3)
- ceppi inbred di animali (3-5)
- campioni umani (8-10) (deve essere valutato per ogni esperimento)

GeneChip® Platform

Design
Experiment



Prepare
Sample



Hybridize



Hybridization
Oven

Wash
&
Stain

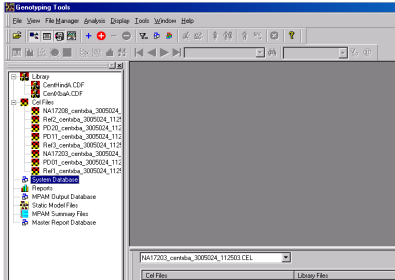


Fluidics
Station

Scan



Scanner



Data Analysis
Software



Probe
Array

CAPACITA' ATTUALE ED EVOLUZIONI FUTURE



Evolution



Feature size	18 μ m	11 μ m	8 μ m	5 μ m	1 μ m
Probes	500,000	1,300,000	2,600,000	6,500,000	165,000,000

Commercial expression Arrays

U133 A+B

Commercial expression Arrays

U133 Plus
10K SNP

Commercial genotyping Arrays

100K SNP

New Applications

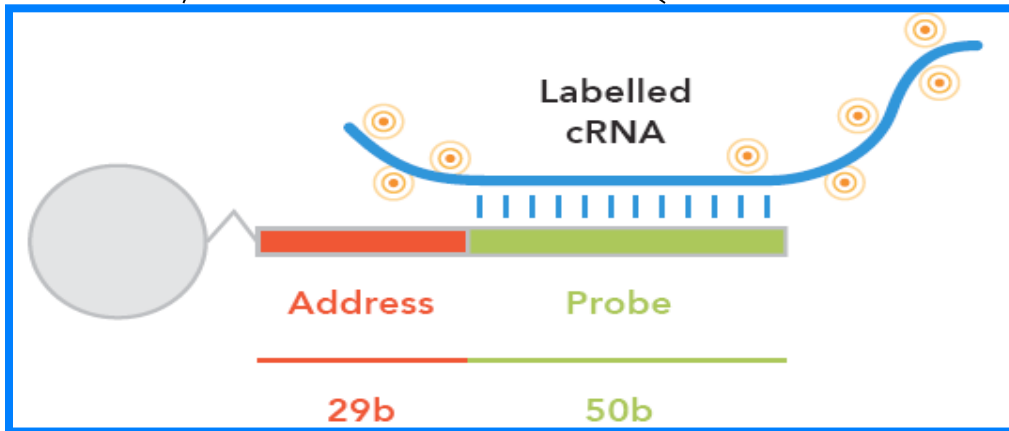
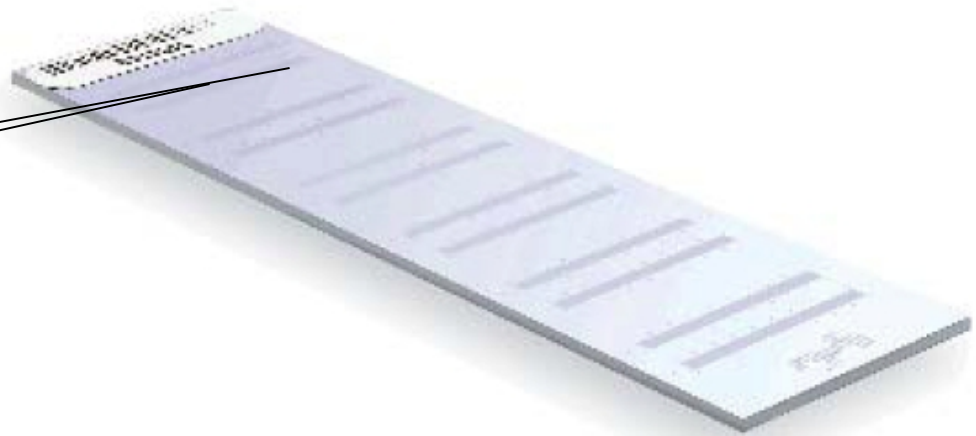
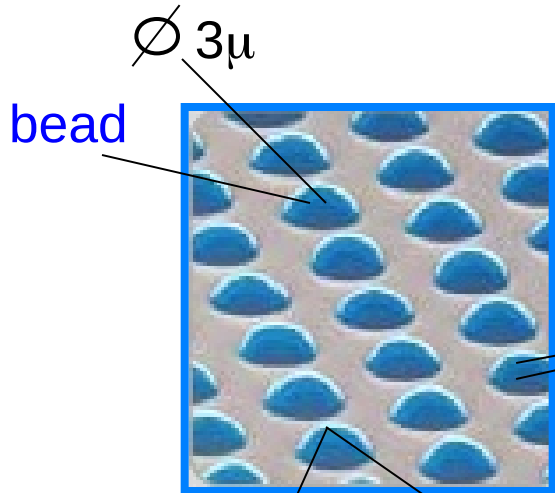
500K SNP Exon arrays
Tiling arrays

In development

**Microarray one channel – GeneChip
Illumina**

Arrays illumina (beads array)

ciascun array contiene micropozzetti che vengono riempiti con biglie ricoperte con migliaia di copie di uno stesso probe (50 bp)

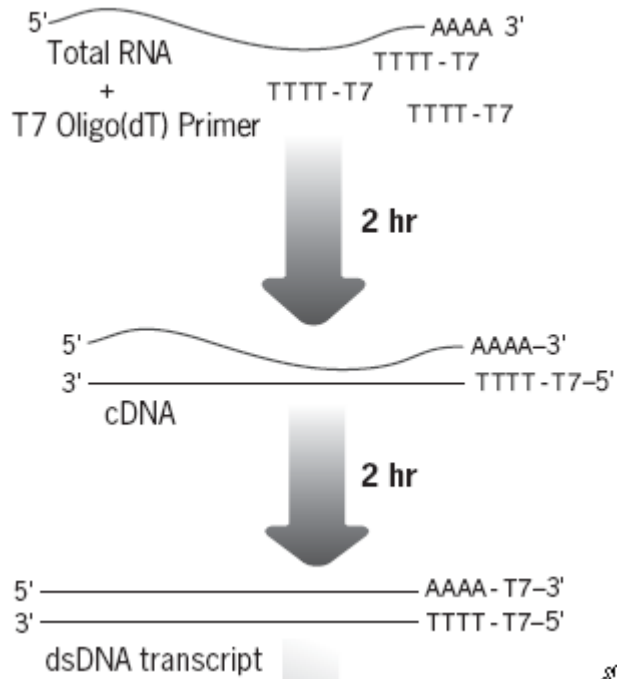


durante la produzione ad ogni probe viene legata una corta sequenza (barcode o address sequence) che serve per identificare la bead

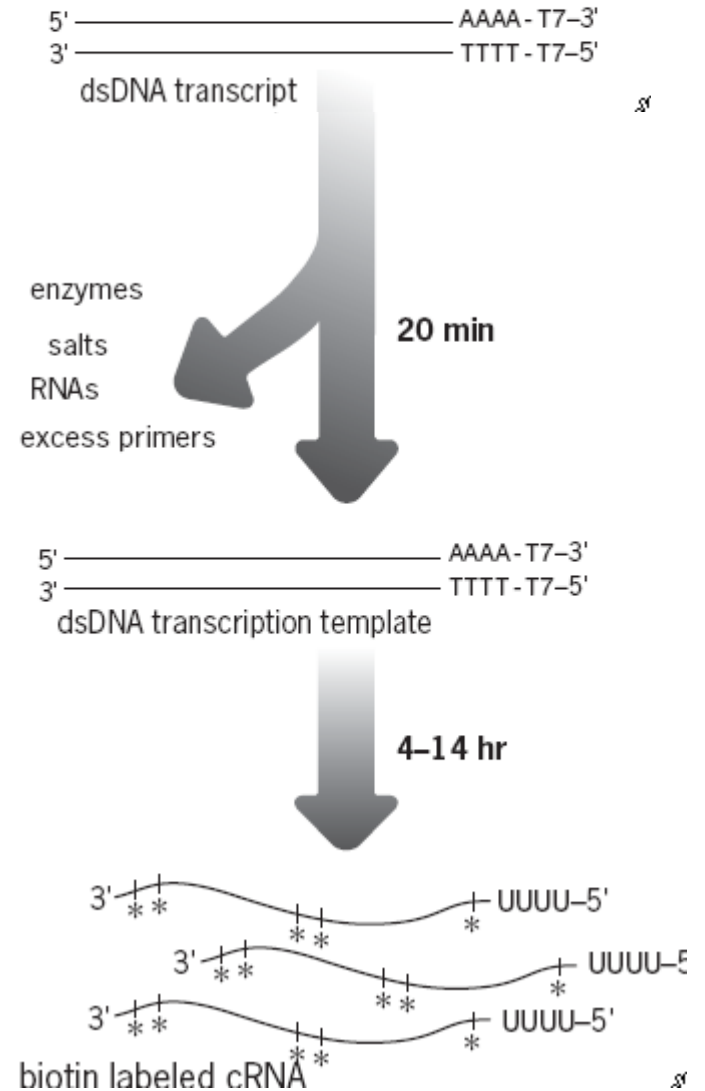
Alta Densità fino a 10^6

Illumina labeling

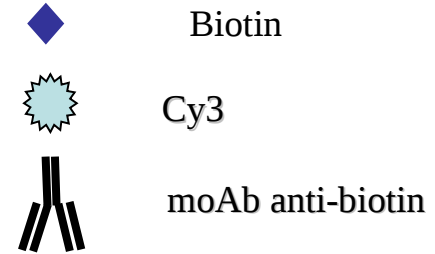
A



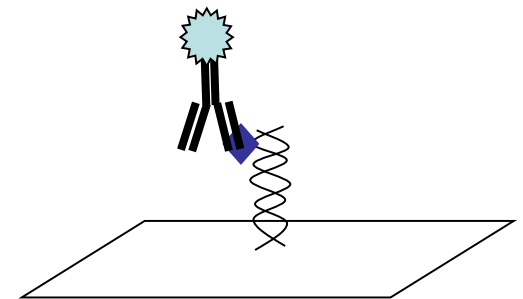
B



Detection



- Hybridization detection is performed using anti-biotin moAb labeled with Cy3.
- The fluorescence signal is read by a laser based scanner.

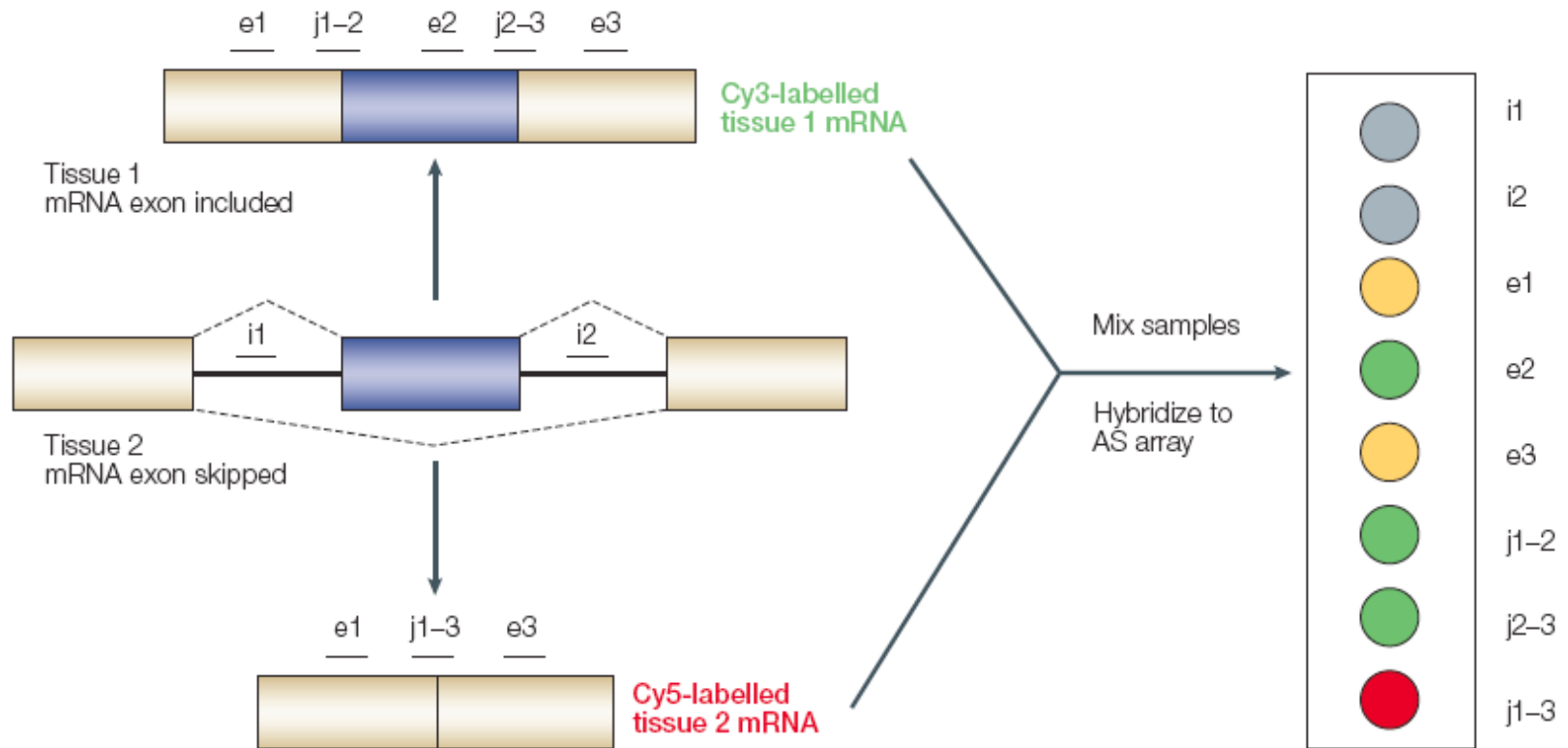


New Types of Microarrays

Alternative splicing microarrays

Si basano su probes oligonucleotidici (25-60 nt) in grado di ibridare a regioni isoforma-specifiche degli mRNA

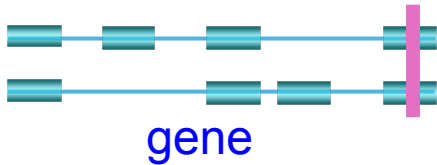
- probes esone-specifici: $e1, e2, e3, \dots$:
- probes giunzione-specifici: $j1-2, j2-3, j1-3, \dots$:
- probes intronici: $i1, i2, \dots$:



Gli Exon Arrays

Gene-Level Expression Profiling

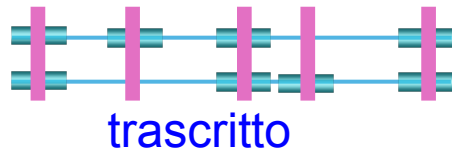
3' expression array



per uomo topo ratto e tutti i maggiori organismi modello; rappresentano la maggior collezione di array whole-genome

Alternative Splicing

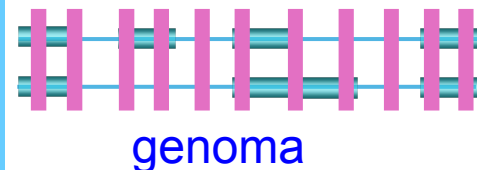
exon array



permettono di analizzare in un solo array tutti i geni a livello di esoni dato che contengono un totale di 1.000.000 di esoni che virtualmente rappresentano tutti i geni codificanti

Examination of Protein/DNA Interactions

tiling array



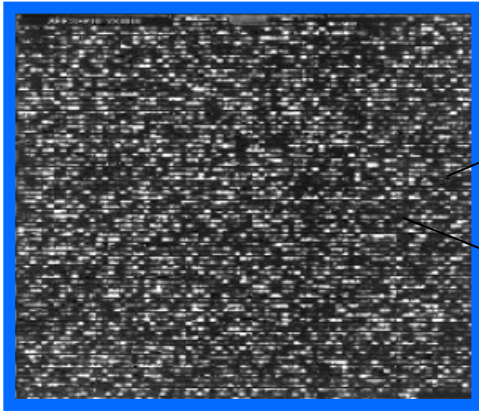
permettono di analizzare regioni del genoma inesplorate (es. promotori) e di studiare i processi di regolazione (interazioni DNA - proteina o Chromatin immunoprecipitation - ChIP on Chip)

DNA ARRAYS: SNPs

Consentono analisi genetiche complesse (analisi genome-wide per identificazione polimorfismi associati a patologie)

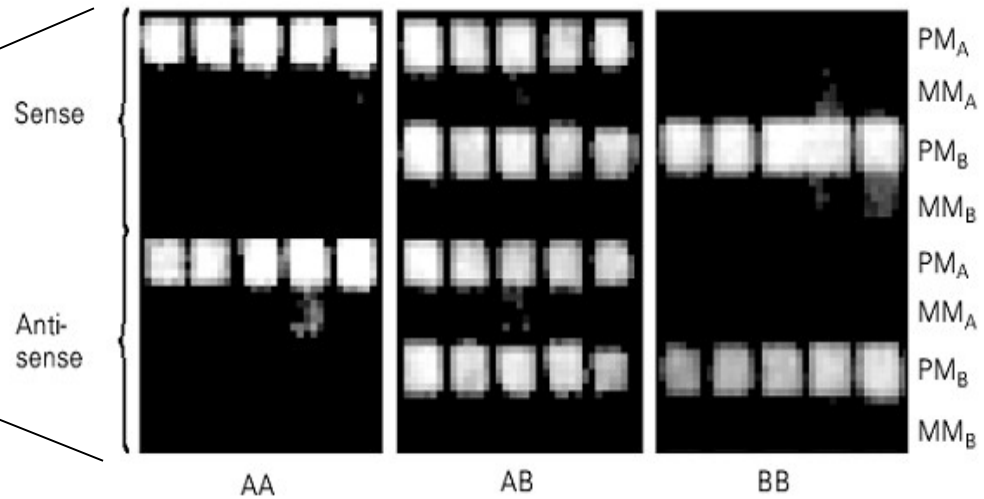
allele A
 TTTGTCACAAG PM
 TTTGTCGCAAG MM

allele B
 TTTGTCACGAG PM
 TTTGTCGCGAG MM

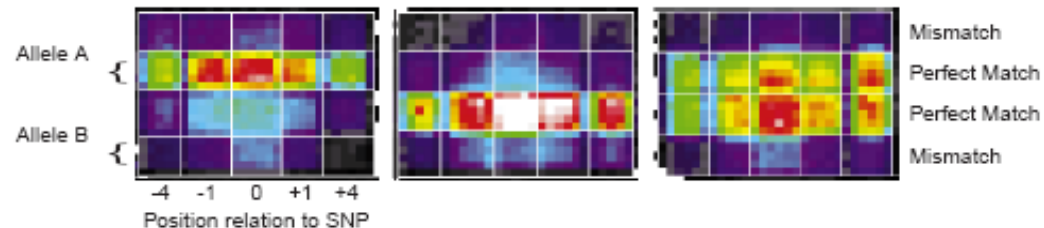


Genotyping Array

Oligo design in a SNP array



A detail of a SNP array output



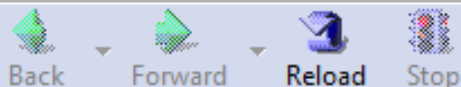
gli SNP array permettono genotipizzazioni complesse

Alcuni limiti nell'utilizzo dei microarrays

- 1** Richiede la conoscenza, almeno parziale, del genoma o trascrittoma dell'organismo/sistema in oggetto
in assenza necessario usare altre metodiche per es DD-PCR
- 2** E' una metodica quantitativa però comparativa non assoluta
se si vuole analisi quantitativa rigorosa allora per es. SAGE
- 3** Permette di determinare variazioni dell'espressione genica solo al livello trascrizionale
Non consente di ottenere informazioni su tutto ciò che avviene post-trascrizionalmente:
 - controllo traduzionale
 - stabilità mRNA
 - silenziamento
 - splicing alternativo

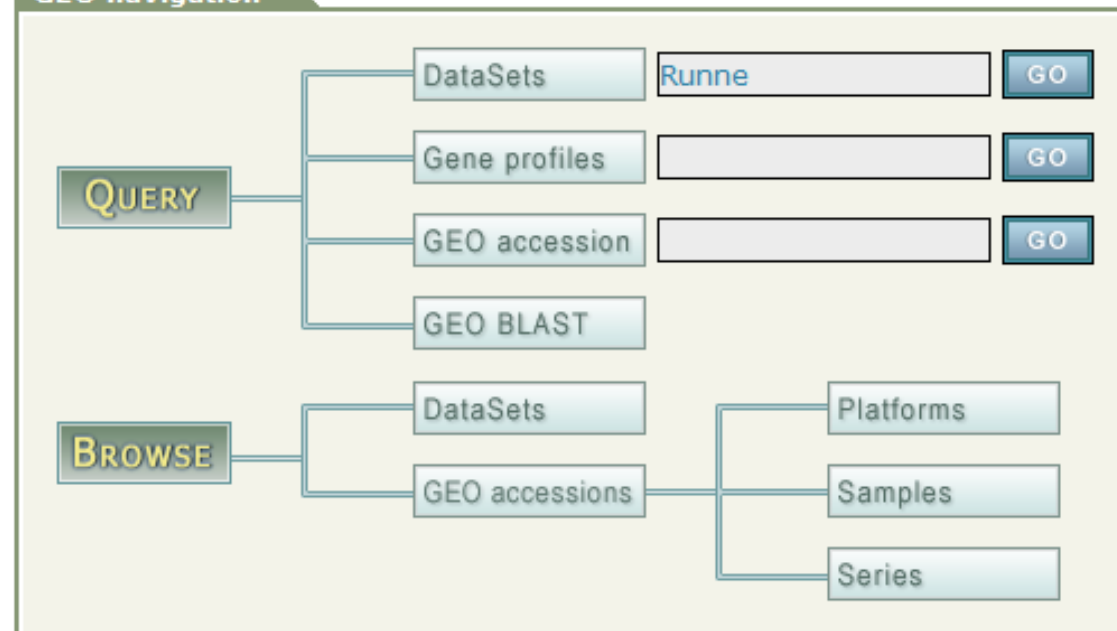
Come posso ottenere i dati sperimentali?

- Recentemente per l'accettazione di un articolo su riviste internazionali viene richiesto che dati siano depositati su banche dati pubbliche:
 - Europa: arrayexpress
 - USA: GEO



Gene Expression Omnibus: a gene expression/molecular abundance repository supporting [MIAME compliant](#) data submissions, and a curated, online resource for gene expression data browsing, query and retrieval.

GEO navigation



Search for

[Save Search](#)

Display Show Sort By Send to

Items 1 - 4 of 4

One page.

1: GSE17409 record: Pregnancy changes expression in peripheral blood mononuclear cells of healthy donors [*Homo sapiens*] [Links](#)

Summary: (Submitter supplied) Background: pregnancy is associated with reduced activity of multiple sclerosis (MS). However, the biological mechanisms underlying this pregnancy-related decrease in disease activity are poorly understood. This data series contains the subset of data used to generate a healthy donors signature comparing female healthy specimens before pregnancy with respect to female healthy specimens at ninth month pregnancy.

[1 related Platform](#)

Type: Expression profiling by array

Supplementary Files: CEL [download...](#)


Samples: 11


- [GSM434518](#): 21 preN
- [GSM434521](#): 31 preN
- [GSM434524](#): 33 preN
- [GSM434723](#): CF4 GRA9n
- [GSM434512](#): VC1 preN
- [GSM434718](#): 24_MO4 GRA9n

Scope: Format: Amount: GEO accession:

Series GSE17393
[Query DataSets for GSE17393](#)

Status	Public on Jan 10, 2010
Title	Transcription signature of Multiple Sclerosis in peripheral blood mononuclear cells.
Organism	Homo sapiens
Experiment type	Expression profiling by array
Summary	Background: pregnancy is associated with reduced activity of multiple sclerosis (MS). However, the biological mechanisms underlying this pregnancy-related decrease in disease activity are poorly understood. This data series contains the subset of data used to generate a MS signature comparing female healthy specimens with respect to MS patients
Overall design	Subjects were followed in the outpatients clinic and blood was collected before pregnancy and at the following time points during pregnancy: first trimester (gestational age at sampling 12 weeks), second trimester (24 weeks), and third trimester (36 weeks). Before-pregnancy samples were obtained in a treatment-free period and after anticonceptual drug withdrawal. Peripheral blood mononuclear cells (PBMCs) obtained from 15 women (8 MS patients and 7 healthy controls) were analyzed by oligonucleotide microarray technology.
Contributor(s)	Gilli F , Lindberg R , Valentino P , Marnetto F , Malucchi S , Sala A , Capobianco M , di Sapio A , Sperli F , Kappos L , Calogero R , Bertolotto A
Citation(s)	Gilli F, Lindberg RL, Valentino P, Marnetto F et al. Learning from nature: pregnancy changes the expression of inflammation-related genes in patients with multiple sclerosis. <i>PLoS One</i> 2010 Jan 29;5(1):e8962. PMID: 20126412

Download family	Format
SOFT formatted family file(s)	SOFT ?
MINiML formatted family file(s)	MINiML ?
Series Matrix File(s) 	TXT ?

Supplementary files	File type
GSE1869_RAW.tar  Raw data provided as supplementary file	TAR (of CEL)

E' possibile scaricare i dati:

1. in formato tipo excel (tabulato) contenente tutte le informazioni dell'esperimento
2. le immagini dell'array (in questo caso i .CEL files dell'Affymetrix)

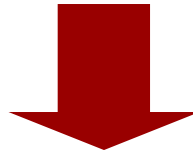
!Series_title	Transcription signature of Multiple Sclerosis in peripheral blood mononuclear cells.					
!Series_geo_accession	GSE17393					
!Series_status	Public on Jan 10 2010					
!Series_submission_date	Jul 29 2009					
!Series_last_update_date	Apr 11 2010					
!Series_pubmed_id	20126412					
!Series_summary	Background: pregnancy is associated with reduced activity of multiple sclerosis (MS). However, the biological mech					
!Series_summary	This data series contains the subset of data used to generate a MS signature comparing female healthy specimens v					
!Series_overall_design	Subjects were followed in the outpatients clinic and blood was collected before pregnancy and at the following tim					
!Series_overall_design	Peripheral blood mononuclear cells (PBMCs) obtained from 15 women (8 MS patients and 7 healthy controls) were					
!Series_type	Expression profiling by array					
!Series_contributor	F,,Gilli					
!Series_contributor	RLP,,Lindberg					
!Series_contributor	P,,Valentino					
!Series_contributor	F,,Marnetto					
!Series_contributor	S,,Malucchi					
!Series_contributor	A..Sala					
!Series_contribu	!Sample_contact_zip/postal_code	10126	10126	10126	10126	10126
!Series_contribu	!Sample_contact_country	Italy	Italy	Italy	Italy	Italy
!Series_contribu	!Sample_contact_web_link	www.bioinfo	www.bioinfo	www.bioinfo	www.bioinfo	www.bioinfo
!Series_contribu	!Sample_supplementary_file	ftp://ftp.ncbi	ftp://ftp.ncbi	ftp://ftp.ncbi	ftp://ftp.ncbi	ftp://ftp.ncbi
!Series_contribu	!Sample_data_row_count	22277	22277	22277	22277	22277
!Series_sample_	!series_matrix_table_begin					
!Series_contact	ID_REF	GSM434504	GSM434505	GSM434506	GSM434507	GSM434509
!Series_contact	1007_s_at	5.3259982	5.56098118	5.82862377	5.40299335	5.81924684
	1053_at	6.27633553	5.88714462	5.82917371	5.93963556	5.19605033
	117_at	7.74118892	6.80088963	6.61506377	6.62906164	5.65883036
	121_at	7.11639447	6.97002243	7.28293919	6.90636939	6.94414601
	1255_g_at	2.50464489	2.56132031	2.8202719	2.57717801	2.68799653
	1294_at	8.19030324	8.37414622	8.36360465	8.05901279	7.92724214
	1316_at	4.36599255	4.39691213	4.71645052	4.40575418	4.59439182

Header Matrix series file

Data Analysis

Molteplici applicazioni per i microarrays

- ✓ analisi di specifici pathways
- ✓ identificazione di geni associati a fasi del ciclo vitale
- ✓ analisi della risposta a un farmaco
- ✓ analisi di meccanismi complessi coinvolti in processi biologici
- ✓ etc....

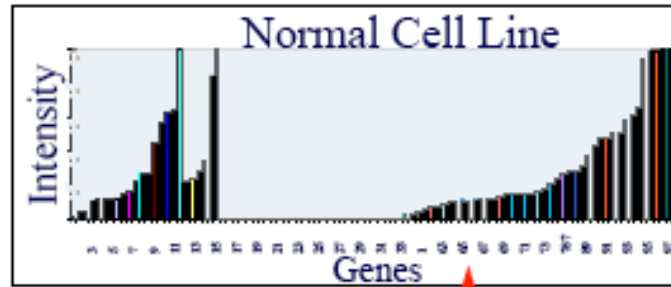
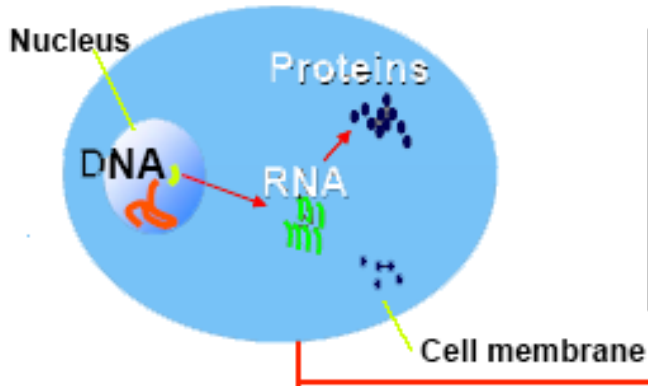


IL RITRATTO MOLECOLARE

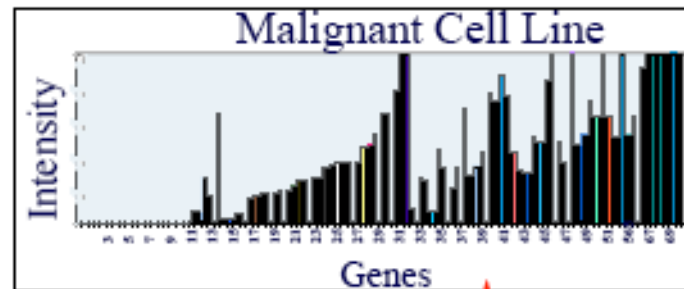
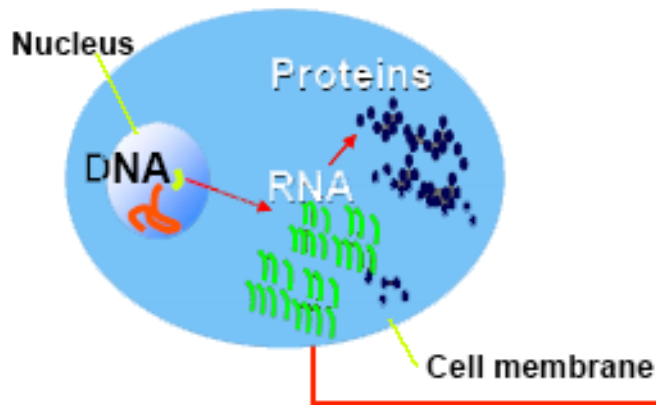
Classificazione dei tumori in base al profilo genico

trasformazione di una cellula normale in cellula tumorale dovuta ad accumulo di cambiamenti molecolari che determinano variazioni del pattern di espressione genica

Tumori e gene profiling



profilo di espressione identico in cellule normali di uno stesso tessuto

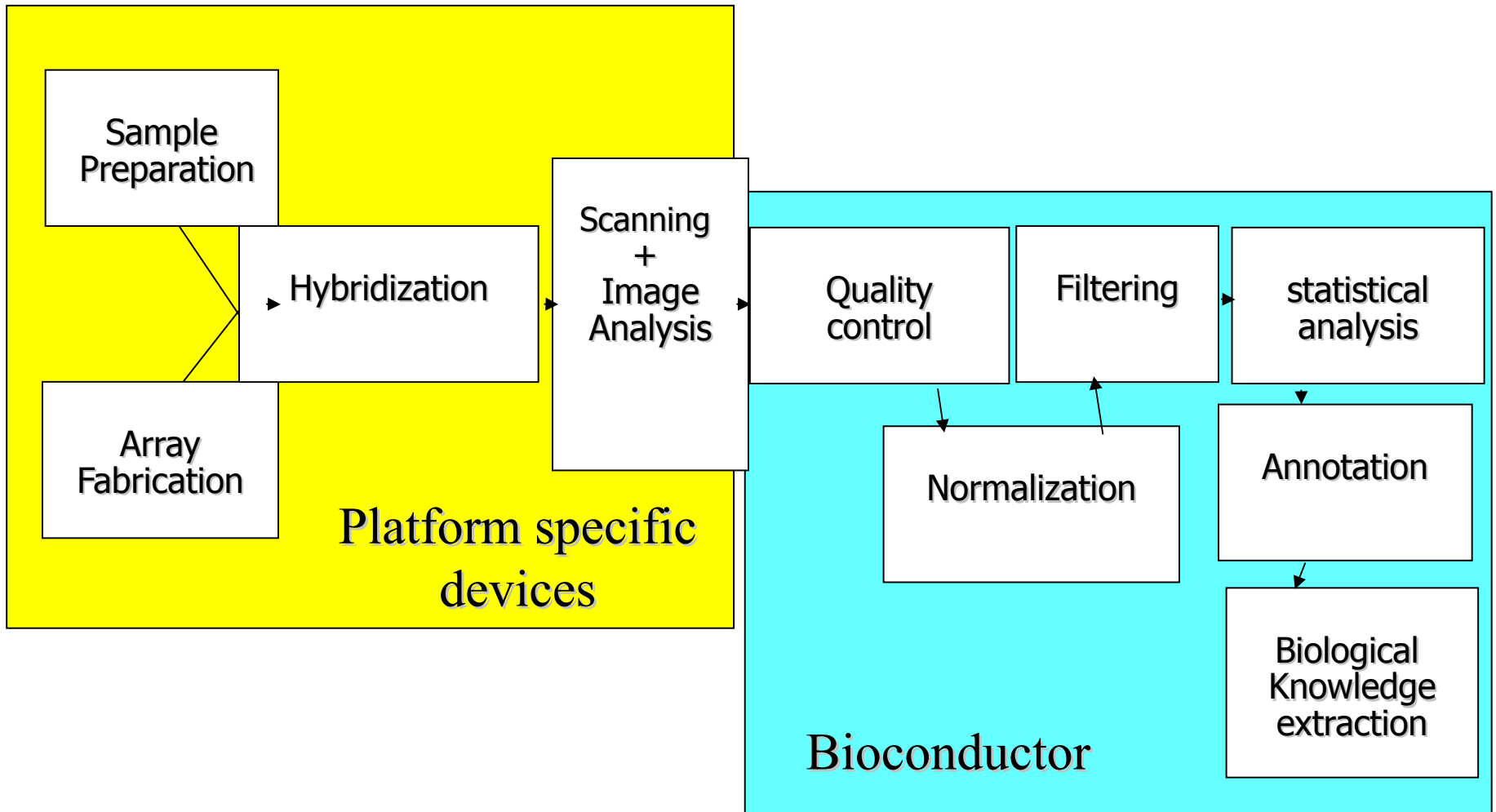


profilo di espressione varia quando insorge il tumore

Per analizzare i dati di microarray è necessario disporre di software dedicati

- I dati da microarray non possono essere analizzati con un semplice foglio excel ma necessitano di strumenti statistici alquanto sofisticati.
- Esistono software commerciali od open-source.
 - Bioconductor

Analysis pipe-line



Disegno sperimentale

- I dati che si ottengono dai microarray indicano le espressioni di un gene relative alla sonda che è stata utilizzata per quel gene
- Quali sono i pericoli legati ad un disegno sperimentale sbagliato?
 - Potrebbe non rispondere direttamente alla domanda sperimentale
 - I dati contengono una quantità di errore tale da compromettere l'interpretazione del risultato

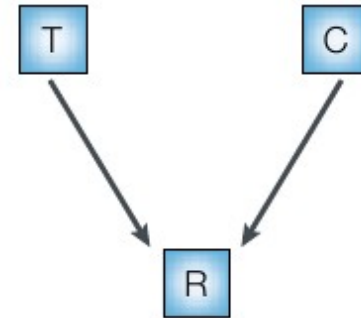
Un fattore chiave è scegliere se usare una comparazione diretta o indiretta tra due campioni cellulari.

Diretta: i due campioni vengono misurati sulla stessa slide



$\log(T/C)$ con varianza $\sigma^2/2$

Indiretta: i livelli di espressione sono misurati su slide differenti



$\log(T/C) = \log(T/R) - \log(C/R)$
con varianza $2\sigma^2$

Dye-swap experiments: coinvolgono due ibridazioni con due campioni in cui il colore assegnato è capovolto rispetto alla prima ibridazione

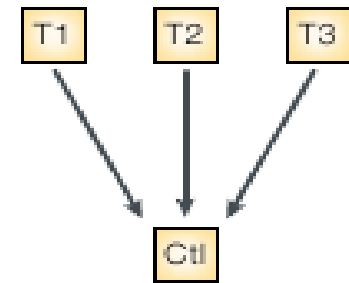
La scelta del disegno sperimentale è dettata dallo scopo dello studio e dall'efficienza statistica che ovviamente ha un effetto sulla scelta del disegno.

Case: Utilizzo di un significativo controllo biologico (crt).

Campioni: tessuto da topi trattati con un farmaco specifico e non trattati

Domande: (1) quali sono i geni che hanno una espressione diversa tra trattati e controllo

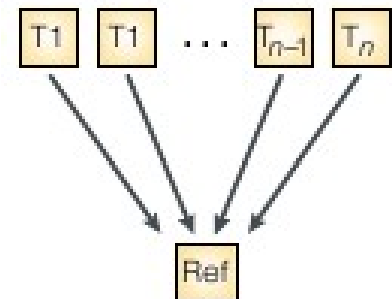
(2) quali son i geni che rispondono in maniera simile tra 2 o più trattamenti



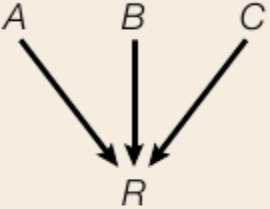
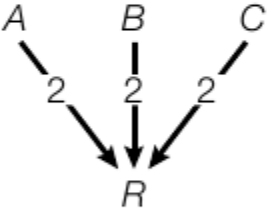
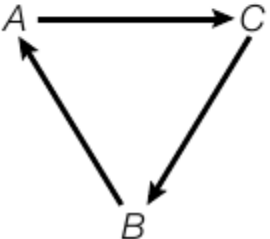
Case: Utilizzo di un reference universale (ref).

Campioni: tessuto da diversi tumori



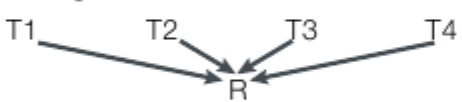
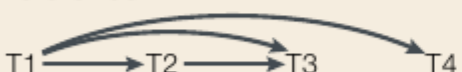


Domande: (1) quali sono i sottopiti tumorali?



Considerate un esperimento in cui tre mRNA, ottenuti da tre sorgenti diversi, sono comparabili e i cui confronti sono di uguale interesse

Design choices	Number of slides	Units of material (number of samples)	Average variance
Indirect designs			
Design I 	3	$A = B = C = 1$	2.00
Design II 	6	$A = B = C = 2$	1.00
Direct design			
Design III 	3	$A = B = C = 2$	0.67

Esperimenti di time-course: il risultato più preciso dipende dalla quantità dei punti temporali

Design choices	t versus t + 1			Comparisons t versus t + 2		t versus t + 3	Average variance
	t_1/t_2	t_2/t_3	t_3/t_4	t_1/t_3	t_2/t_4	t_1/t_4	
Design I — T1 as common reference 	1.00	2.00	2.00	1.00	2.00	1.00	1.5
Design II — direct: sequential 	1.00	1.00	1.00	2.00	2.00	3.00	1.67
Design III — common reference 	2.00	2.00	2.00	2.00	2.00	2.00	2.00
Design IV — T1 as common reference 	0.67	0.67	1.67	0.67	1.67	1.00	1.06
Design V — direct: loop 	0.75	0.75	0.75	1.00	1.00	0.75	0.83
Design VI — direct: mixed 	1.00	0.75	1.00	0.75	0.75	0.75	0.83

Repliche

Le repliche biologiche (e se necessario) tecniche degli esperimenti di microarray, riducono la variabilità tra gli esperimenti.

Quindi i replicati sono essenziali per offrire un risultato significativo.

Senza replicati nessuna analisi statistica può essere eseguita incrementando il numero di falsi positivi e falsi negativi nel risultato finale

Pool di mRNA —————> quando è necessario?