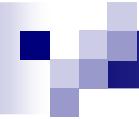


Introducción a los microarrays

(No tan) nuevas aproximaciones
al estudio de la actividad de los
genes



Esquema de la sesión

- Presentación
- Introducción
- Microarrays de expresión
- Experimentos con microarrays
- Análisis de los datos
- Ejercicios prácticos

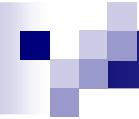


Presentación



Objetivos

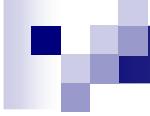
- Conocer la tecnología de experimentación con microarrays
- Comprender sus posibilidades y limitaciones
- Familiarizarse con el proceso de experimentacion basado en los microarrays
- Saber donde acudir para aprender más



Contenidos

- Introducción
 - Antecedentes históricos: El cambio de paradigma
 - Que es un microarray
 - Que tipos de microarrays existen
 - Aplicaciones de los microarrays
- Experimentos con microarrays
 - Cómo funciona un microarray de expresión
 - El ciclo de vida de un experimento con microarrays
- De los números a la interpretación biológica
 - Preprocesado
 - Análisis de los datos

Introducción

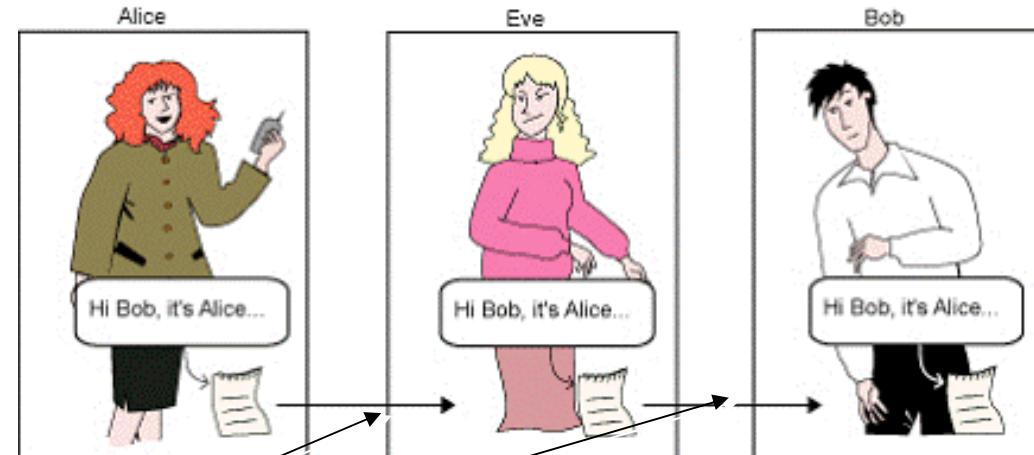


Antecedentes históricos

- La biología molecular dispone de múltiples técnicas para medir los niveles de ARN, ADN, proteínas o metabolitos
 - Northern Blot, differential display, SAGE
 - Southern Blott: [similar a los microarrays]
- Lo que caracteriza la era post genómica no es lo que se puede medir sino la cantidad de mediciones simultaneas que se pueden realizar

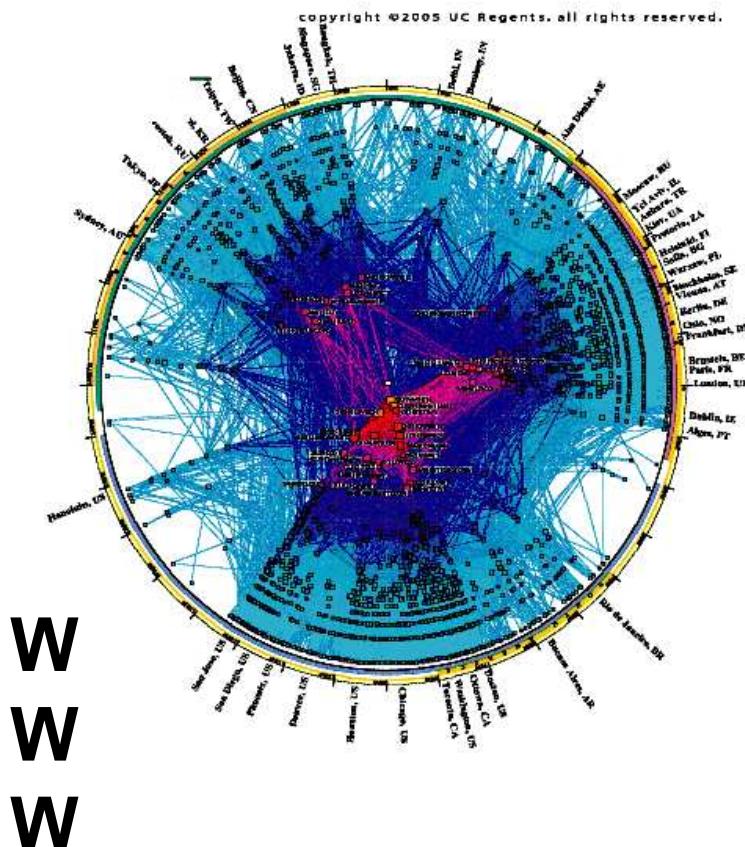
Una analogía

- En la era pre-genómica la biología “espiaba” los genes
 - Individualmente, de uno en uno
 - Cada gen se podía estudiar a fondo



Una analogía (y 2)

- En la era post-genómica se pueden estudiar muchos genes a la vez
- Pero, como separamos el grano de la paja?



Lo he oído
todo



El cambio de paradigma (J. Dopazo)

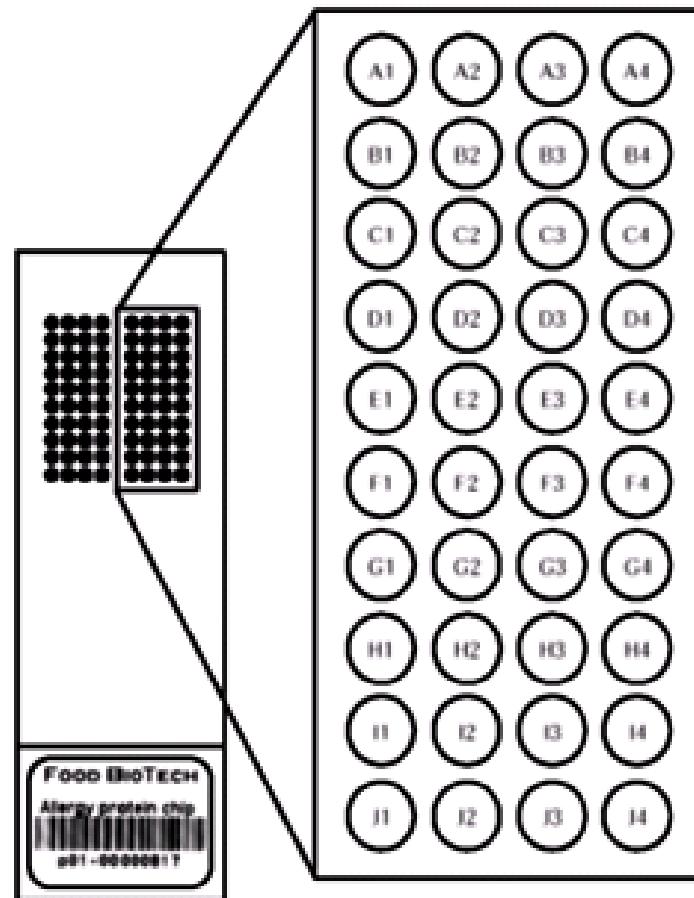


**Con los mismos recursos
Obtenemos una imagen de
menor resolución pero con
una perspectiva más
general**



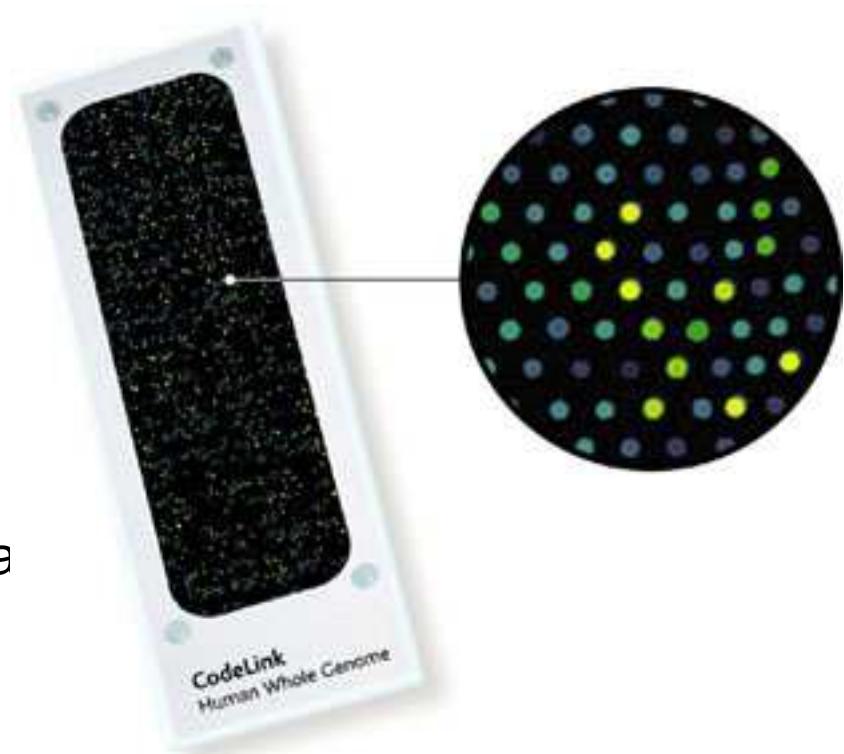
Pero, ¿Qué es un microarray?

- Un formato experimental,
- basado en la síntesis o fijación de **sondas**, que representan los genes (o proteínas, o metabolitos),
- sobre un sustrato sólido (cristal, plástico, silice,...),
- y expuestos a las moléculas **diana** (la muestra).



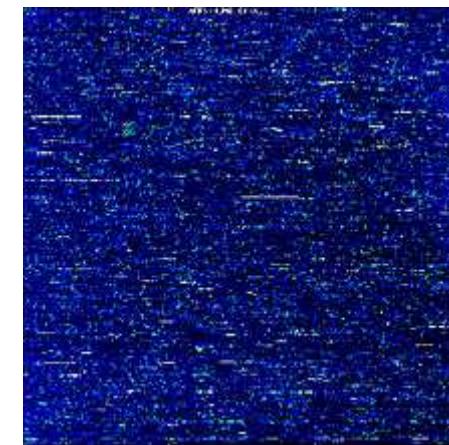
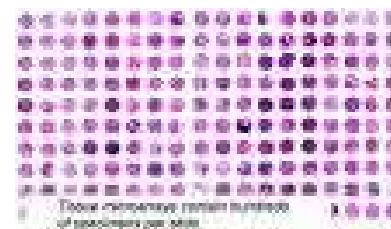
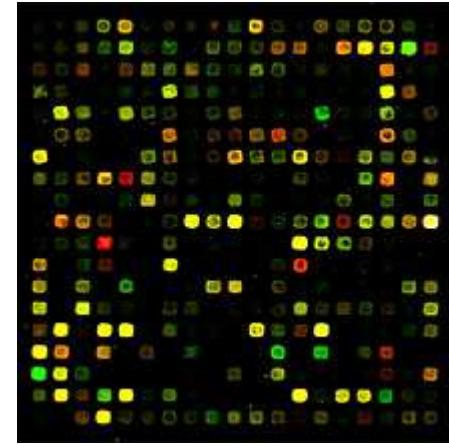
Cómo funciona un microarray

- El nivel de hibridación entre
 - la sonda específica (**probe**) y
 - la molécula diana (**target**)
- se indica generalmente
 - mediante **fluorescencia** y se
 - mide por **análisis de imagen**
- se indica el **nivel de expresión del gen** correspondiente a la sonda en la muestra problema



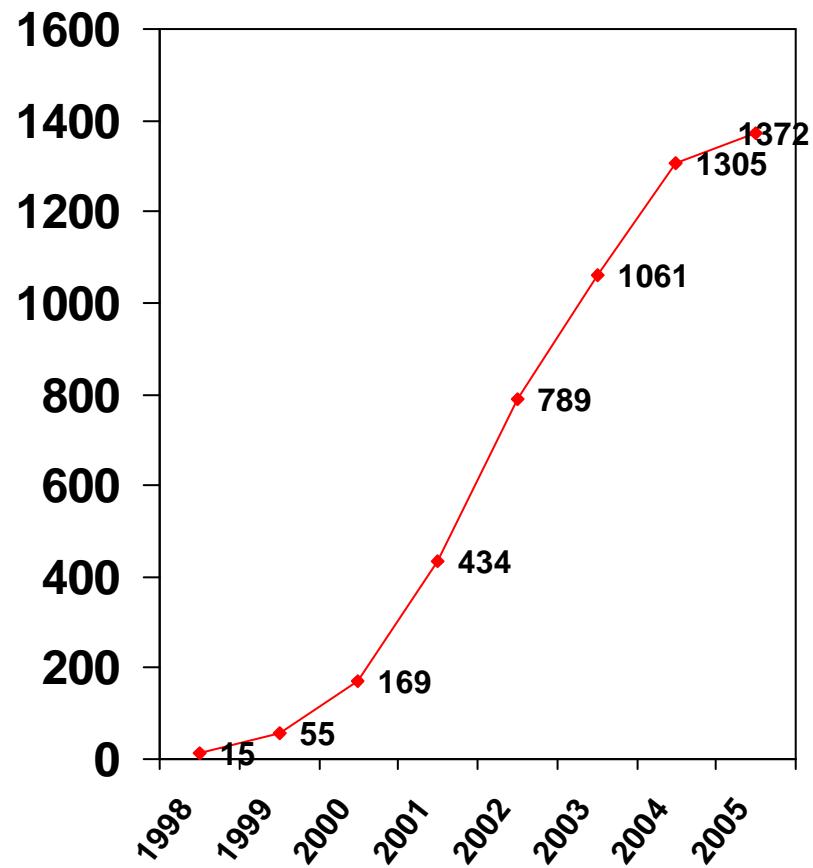
Que tipos de microarrays existen

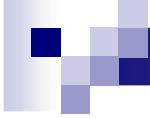
- De Proteínas
- De Tejidos
- De DNA
 - Arrays de CGH
 - SNPs
- De Expresión
 - De cDNA
 - De oligonucleótidos:
 - GeneChip® Affymetrix
 - Otras marcas



Aplicaciones de los microarrays

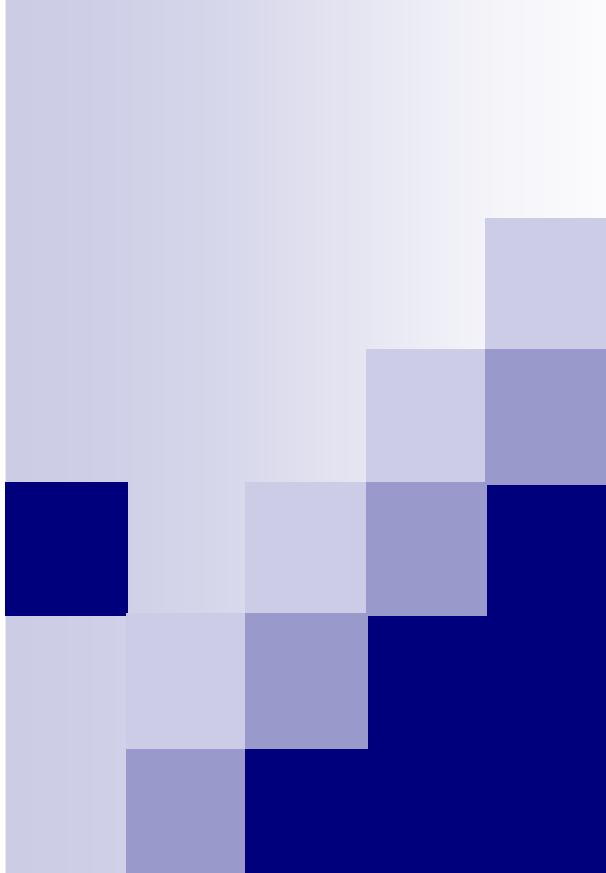
- Los microarrays se han aplicado al estudio de casi cualquier tipo de problema biológico
- El numero de publicaciones anuales con la palabra microarray en el título es muy alto y continua creciendo (?)



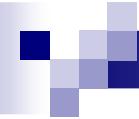


Aplicaciones de los microarrays (2)

- Estudio de genes que se expresan diferencialmente entre varias condiciones
 - Sanos/enfermos, mutantes/salvajes, tratados/no tratados
- Clasificación molecular en enfermedades complejas
- Identificación de genes característicos de una patología (*firma o “signature”*)
- Predicción de respuesta a un tratamiento
- Detección de mutaciones y polimorfismos de un único gen (SNP)
- Etc, etc, etc...

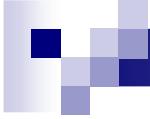


Construcción y uso de los microarrays de expresión



Microarrays de expresión

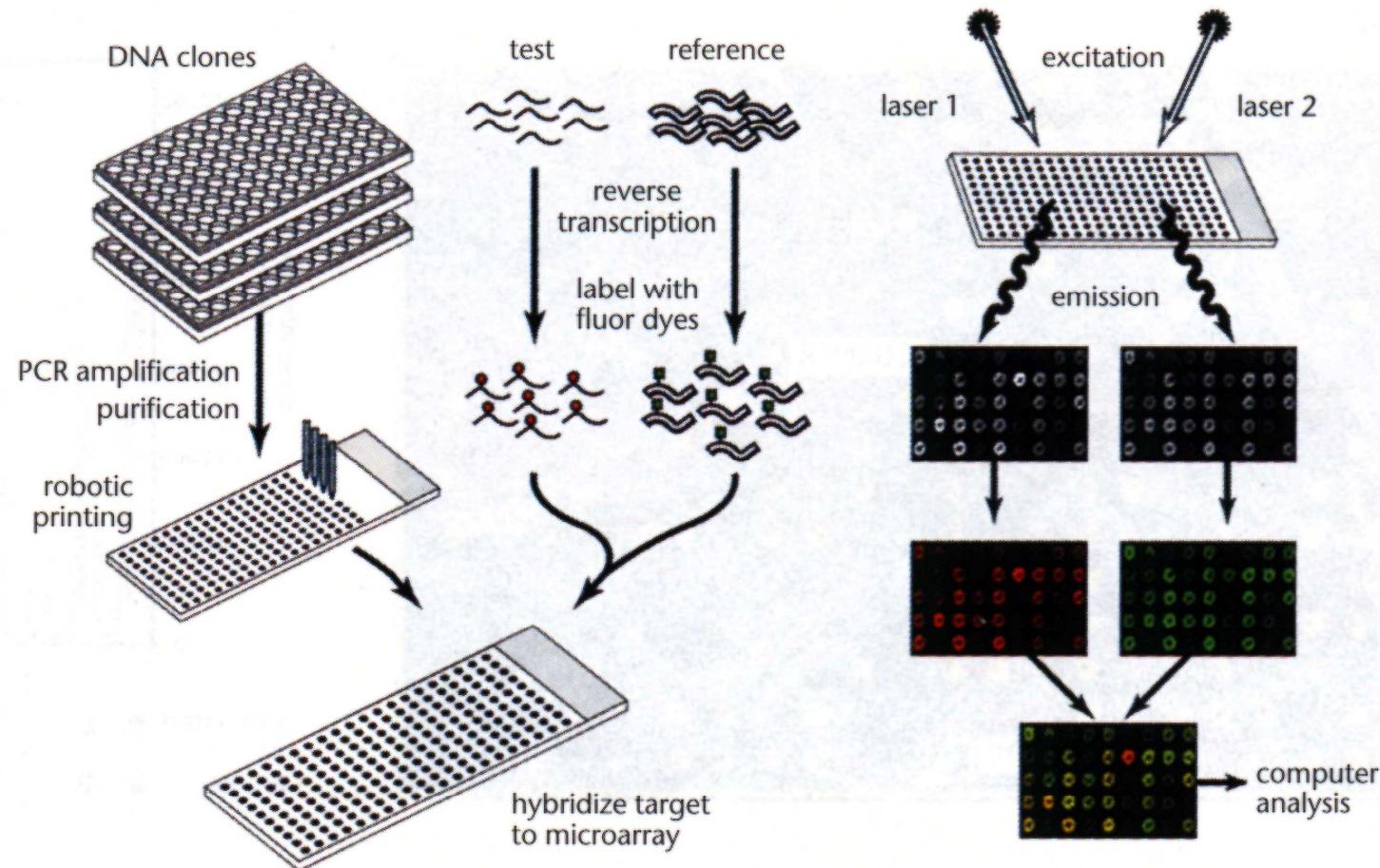
- Existen muchos tipos de microarrays
- Los principios en que se basan son similares
- Los detalles de su funcionamiento varían de uno a otro caso
- En este primer contacto nos centraremos en los arrays de expresión
 - *Arrays de 2 colores (spotted)*
 - *Arrays de oligonucleótidos sintetizados in situ*



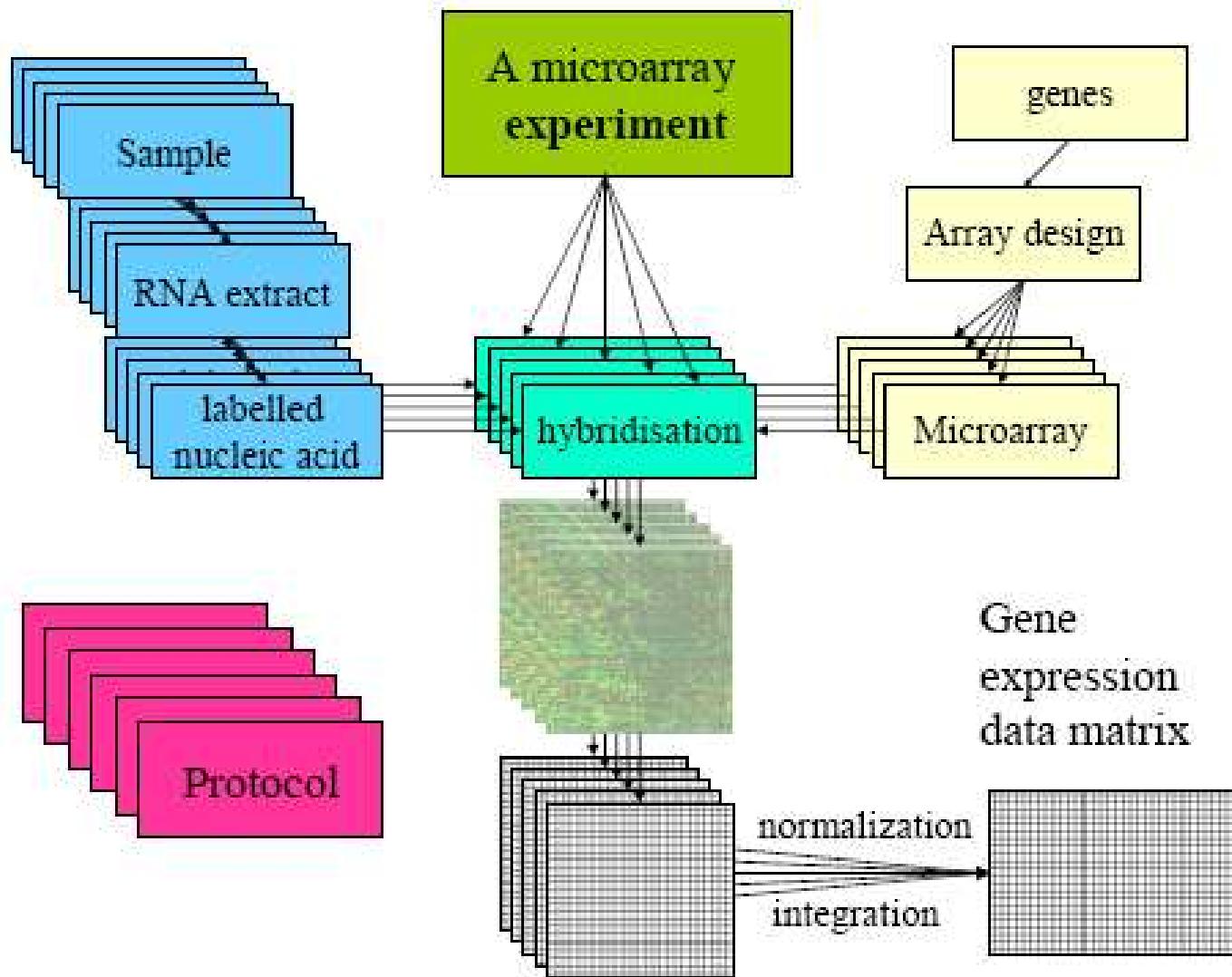
Microarrays de 2 colores (spotted)

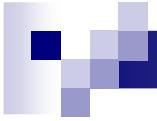
- Diseño y producción del chip
- Preparación de la muestra
- Hibridación
- Escaneado del chip
- Análisis de la imagen

Visión general del proceso



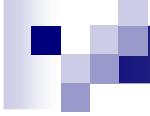
Pulse [este enlace](#) para visualizar una animación del proceso





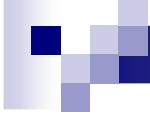
Microarrays de oligos sintetizados *in situ*

- Diseño más avanzado que los de 2 colores
- Utilizan tecnologías desarrolladas en el entorno de la microelectrónica
- Algunos rasgos distintivos
 - No se basan en hibridación competitiva: cada chip contiene muestras de un solo tipo ($\leftarrow\rightarrow$ "1 color")
 - Las sondas se sintetizan directamente sobre el chip en vez de sintetizarlas *in vitro* y adherirlas después
 - Cada gen esta representado por un grupo de sondas cortas en vez de por una solo



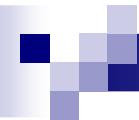
Los GeneChips de Affymetrix

- Affymetrix (www.affymetrix.com) es la compañía líder en este tipo de chips
- Se denominan genericamente *GeneChips*
- Cada gen esta representado por un conjunto de secuencias cortas que lo caracterizan
- Algunos chips contienen genomas completos con más de 50.000 grupos de sondas!
- NOTA: Grupos de sondas = *Probesets*



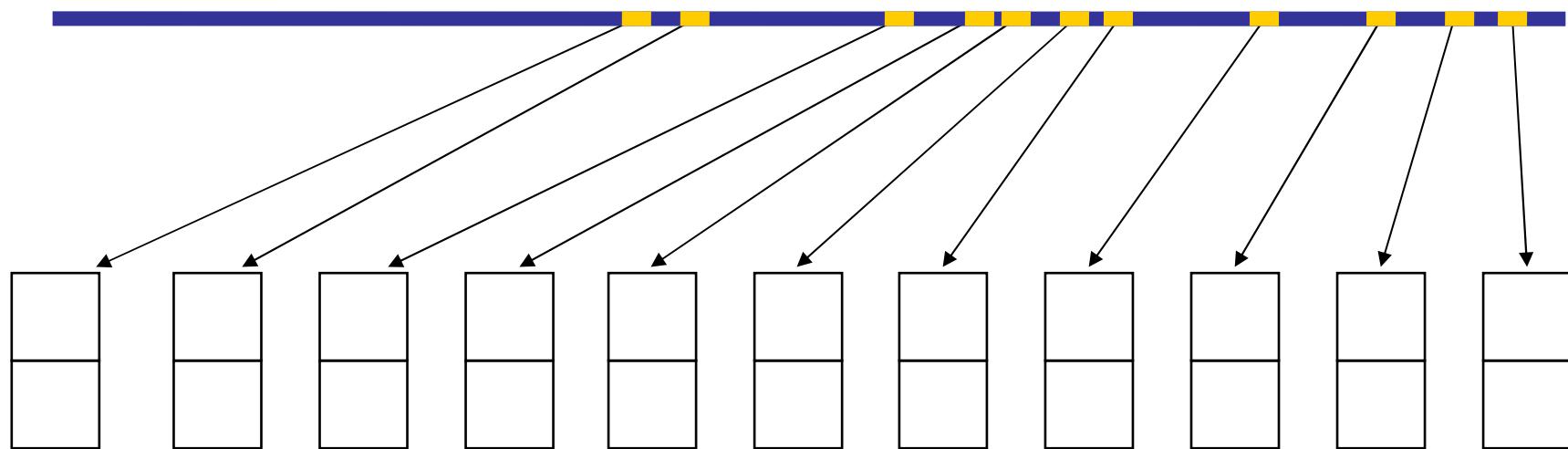
Probesets, probes, PM & MM

- Un *grupo de sondas* se utiliza para medir niveles de mRNA de un único gen
- Cada grupo (*probeset*) consta de múltiples pares de celdas (*probe cells*)
 - Con millones de copias de un oligo de 25bp
 - Organizadas en parejas (*probe pairs*) con un *Perfect Match (PM)* y un *Mismatch (MM)*
 - **PM:** coincide exactamente con una parte del gen
 - **MM:** idéntico al PM excepto en el nucleótido central reemplazado por su complementario



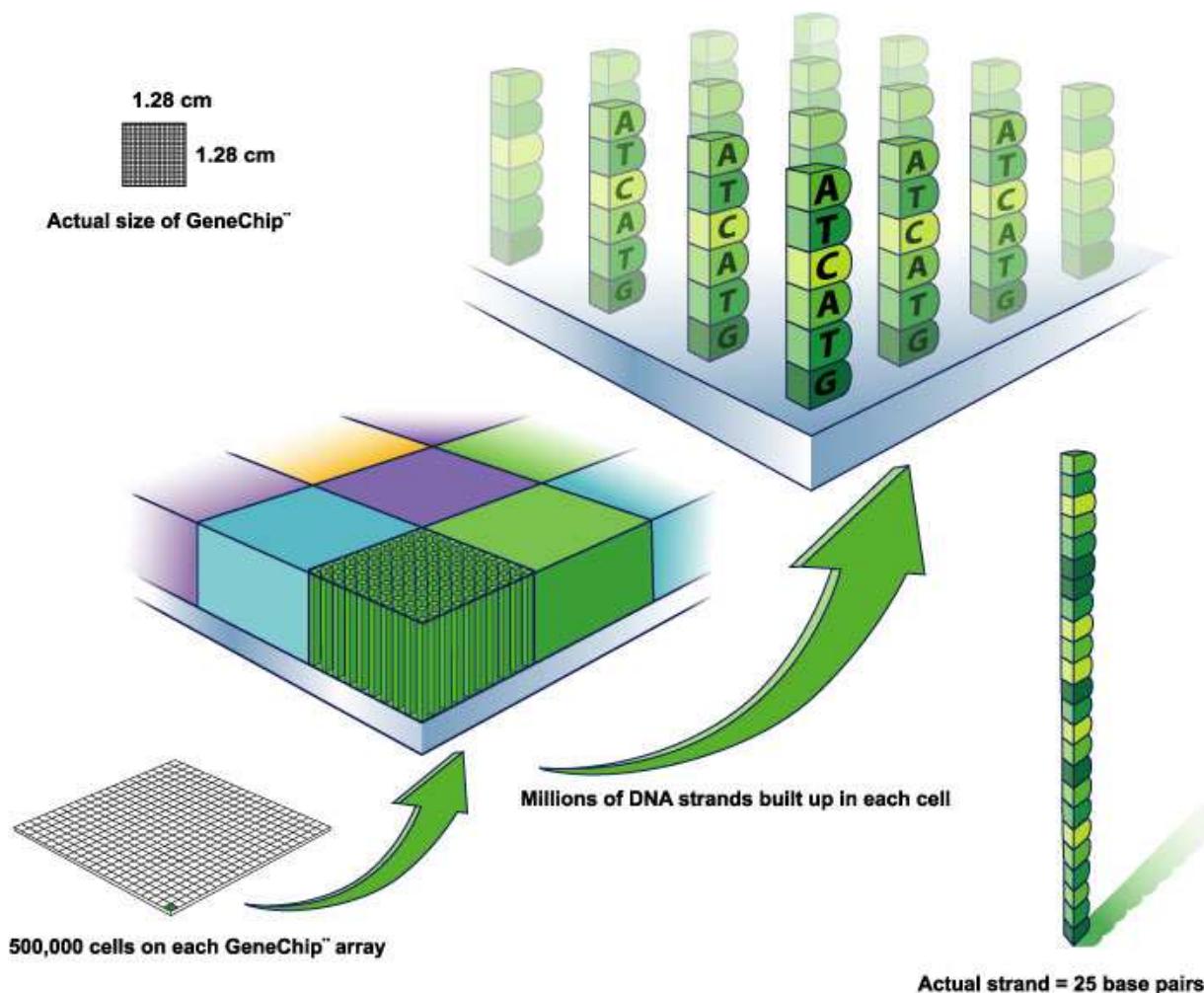
Distintos “Pares de Sondas” representan partes distintas del mismo gen (1 gen=1 grupo de sondas)

Secuencia del gen



Las sondas se seleccionan para ser específicas del gen que representan y para tener buenas propiedades de hibridación

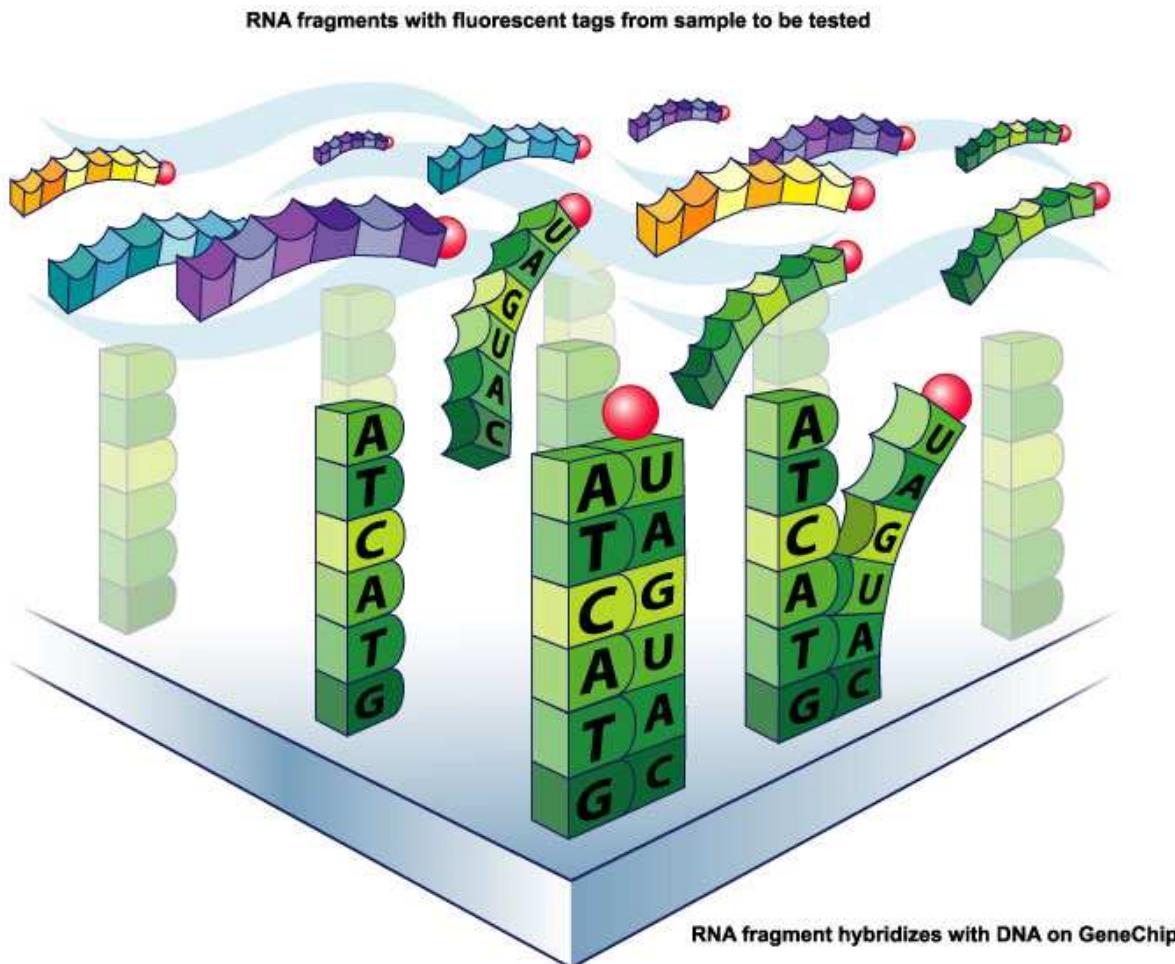
Resultado de la síntesis de oligos en el chip



Cada celda contiene múltiples copias de la misma secuencia

Image courtesy of Affymetrix.

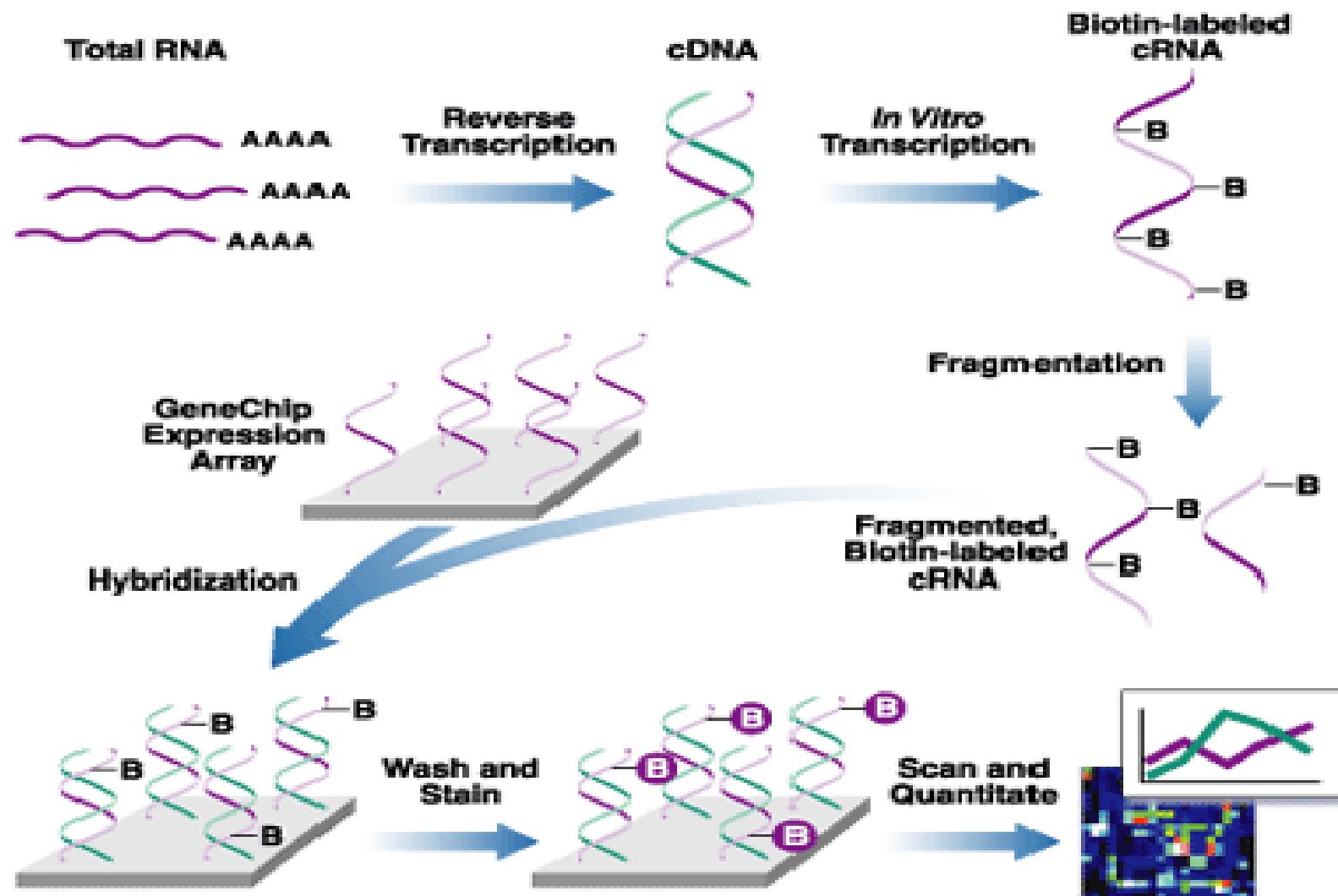
Proceso de hibridación



- Tras la síntesis de los “oligos” se realiza la hibridación, depositando el mRNA **marcado** del tejido a estudiar sobre cada chip

Image courtesy of Affymetrix.

Visión general del proceso (Affy)



@Affymetrix

Comparación entre los 2 tipos de chips

Microarrays de cDNA

- VENTAJAS
 - Económicos
 - Flexibilidad en el diseño experimental
 - Elevada intensidad de señal (secs largas)
- DESVENTAJAS
 - Baja Reproducibilidad
 - Hibridación cruzada (baja especificidad)
 - Elevada manipulación manual (Posibilidad de contaminación)

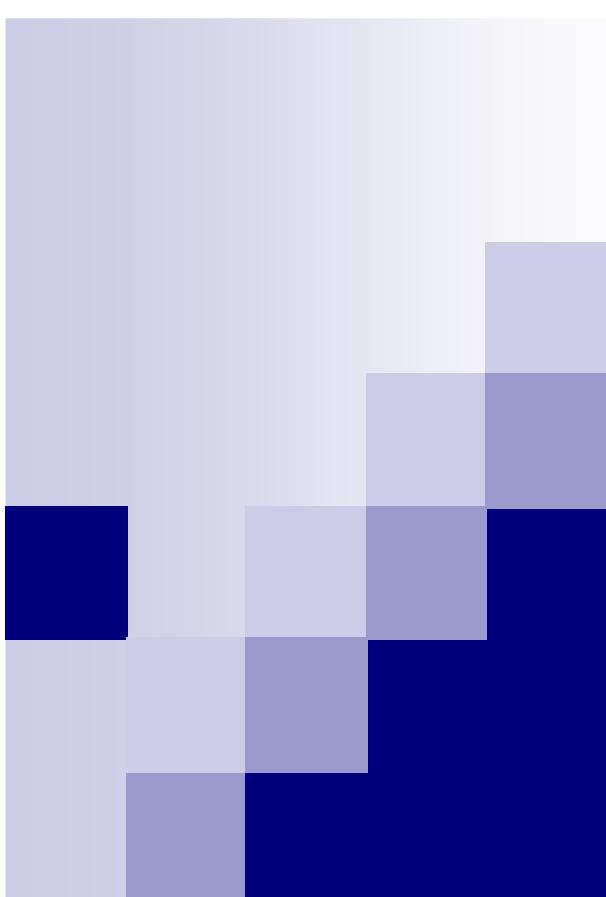
Microarrays de Oligonucleótidos

VENTAJAS

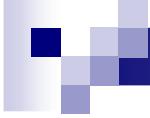
- Fabricación Rápida y más robotizada
- Elevada Reproducibilidad
- Elevada especificidad (secuencias cortas)
- Utiliza muchas sondas/gen

DESVENTAJAS

- Requiere equipamiento más especializado
- Caros
- Poca flexibilidad

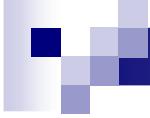


Experimentos con microarrays



Experimentos con microarrays

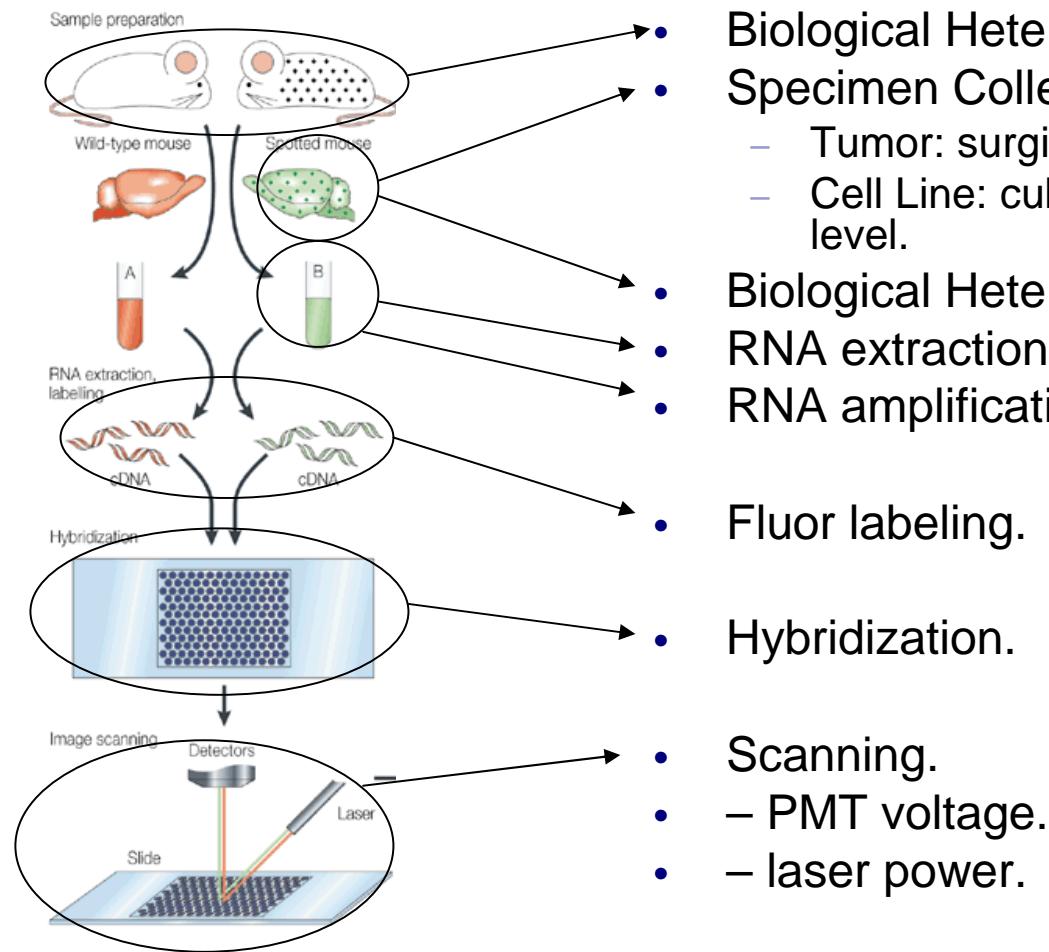
- Fuentes de variabilidad y su control
- Ciclo de vida de un experimento con microarrays
- El diseño del experimento
 - Tipos de cuestiones que se desea responder
 - Factores que debemos tener en cuenta
- Preprocesado: de los datos crudos al análisis
 - Control de calidad
 - Normalización



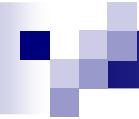
Experimentos con microarrays

- Tal y como su nombre indica un experimento con microarrays es un experimento, es decir:
 - Se lleva a cabo para determinar si ciertas hipótesis previas son ciertas o falsas (*aun cuando también puede llevar a generar nuevas hipótesis*)
- Como todo experimento está sujeto a errores que pueden provenir de múltiples fuentes y ser de tipos distintos
 - Aleatorios
 - Sistemáticos

Fuentes de variabilidad



(Geschwind, *Nature Reviews Neuroscience*, 2001)



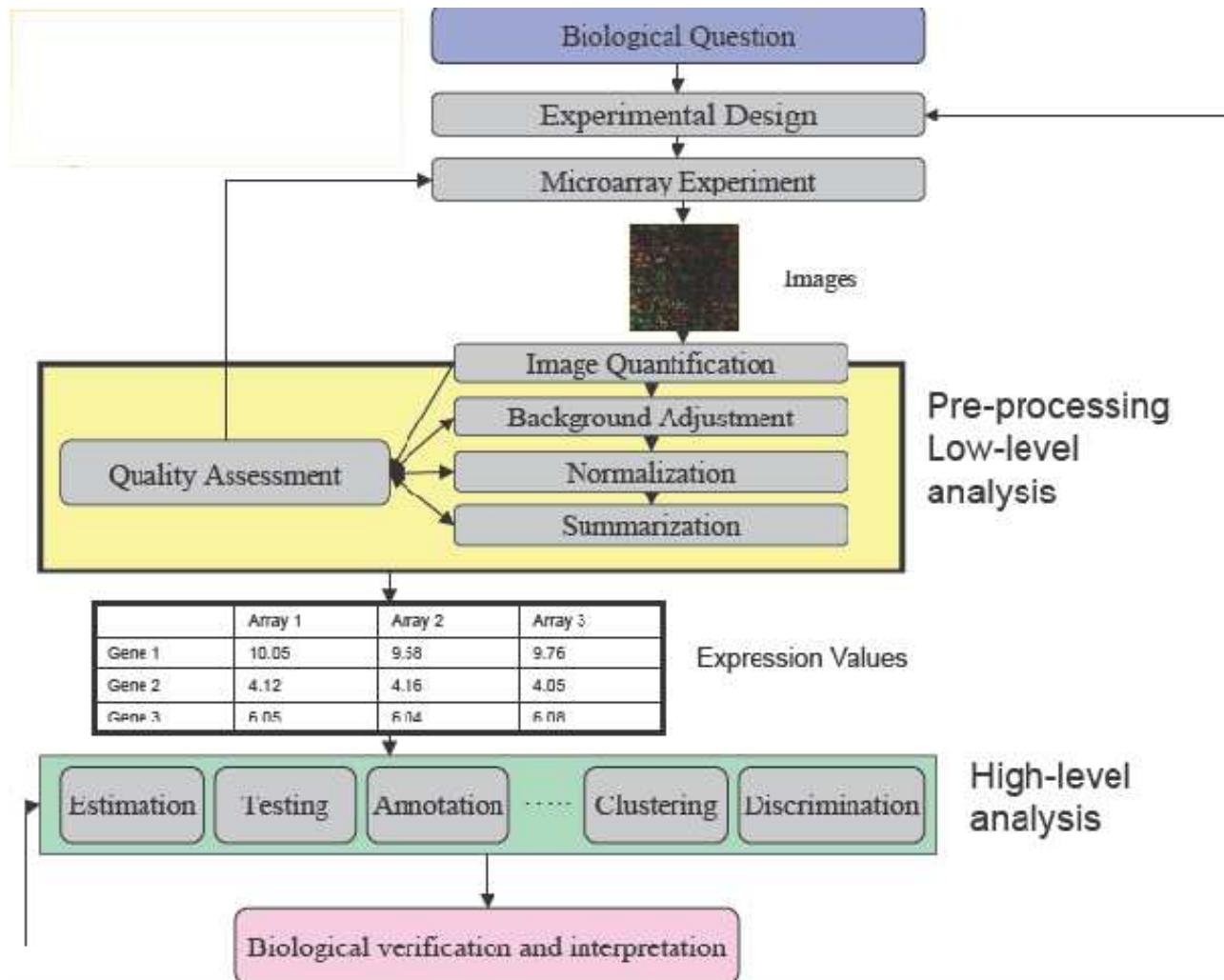
Tipos de variabilidad

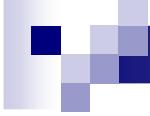
- La variabilidad **sistemática** es aquella que afecta de manera similar a todas las mediciones
 - Cantidad de material disponible
 - Instrumental de laboratorio
- La variabilidad **aleatoria** puede afectar de forma distinta a cada componente del experimento
 - Calidad del material
 - Eficiencia de los procedimientos de laboratorio

Cómo se afronta la variabilidad

- Cada tipo se trata de forma distinta
 - Variabilidad Sistemática
 - Podemos estimar las correcciones necesarias a partir de los datos: **NORMALIZACION o CALIBRACIÓN**
 - Variabilidad Aleatoria
 - Suponemos ciertos modelos de error (e.g. $e_i \sim N(0, \sigma^2)$) y recurrimos al
 - **DISEÑO EXPERIMENTAL** *Para controlarla*
 - **INFERENCIA ESTADÍSTICA** *para extraer conclusiones en su presencia*
- Todos estos procedimientos se integran en un **flujo de trabajo** (“*pipeline*”) o **ciclo de vida** de un experimento con microarrays

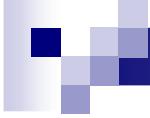
El ciclo de vida de un experimento





De la cuestión biológica al experimento

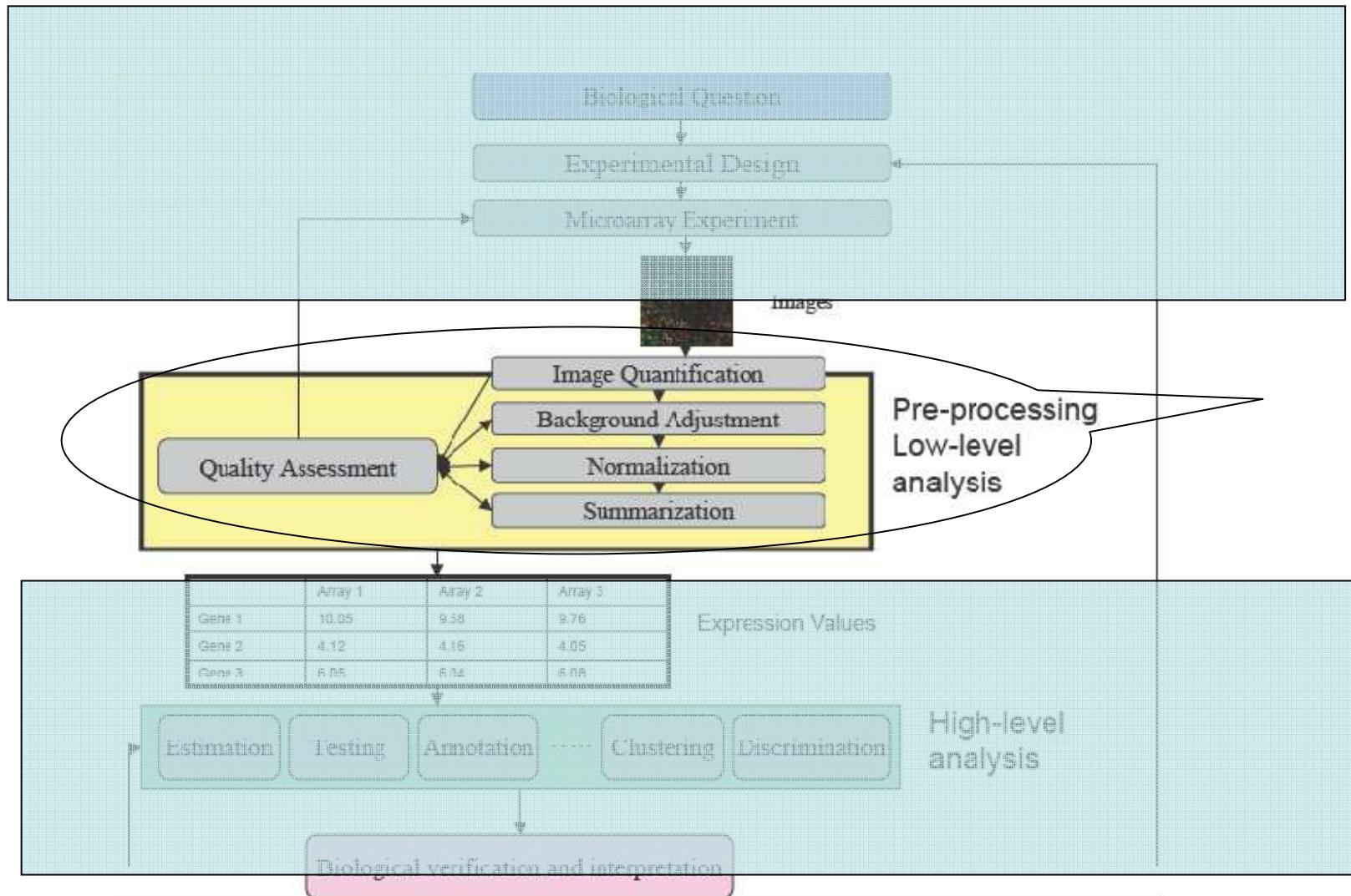
- Una vez planteada una cuestión los implicados en el estudio deberían planearlo conjuntamente
Researchers / Core Facility/ Statisticians
- Es preciso especificar
 - Cual es el propósito del estudio
 - Que objetivos persigue
 - Que limitaciones y de que tipo presenta
- A partir de aquí podrá elaborarse el diseño experimental adecuado



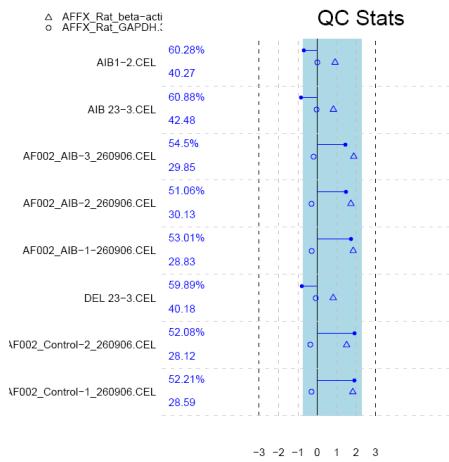
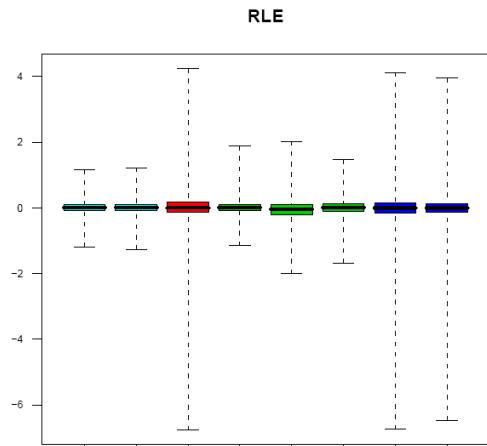
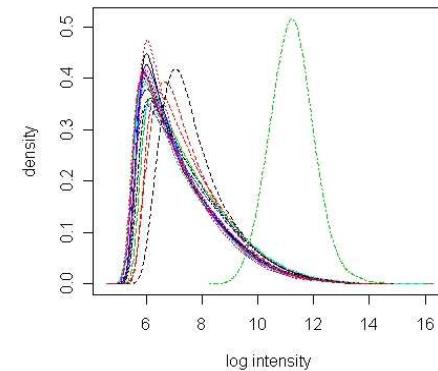
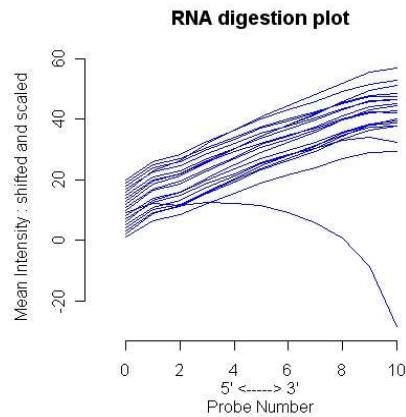
Diseño experimental

- Deben tomarse decisiones relativas a aspectos diversos implicados en el experimento
 - Tipos de muestras
 - Mezcladas (“pooled”) o individuales
 - Con réplicas independientes o sin ellas
 - Limitaciones físicas (coste)
 - Número de arrays necesarios/posibles
 - Cantidad de material necesaria/disponible
- De aquí saldrá
 - La forma en que se realizará el experimento
 - Los métodos estadísticos que debemos aplicar

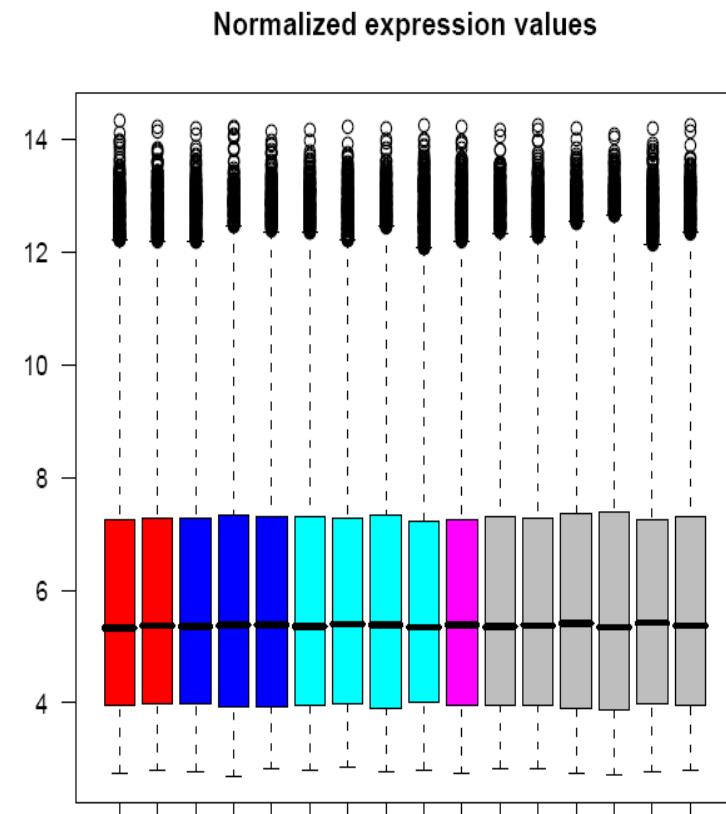
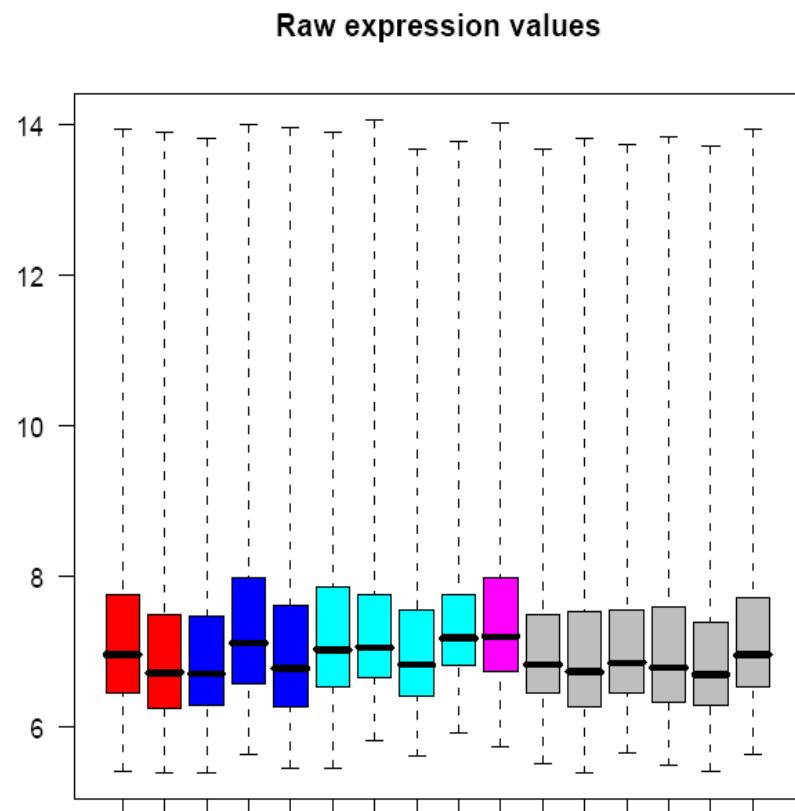
El preprocessado de los datos



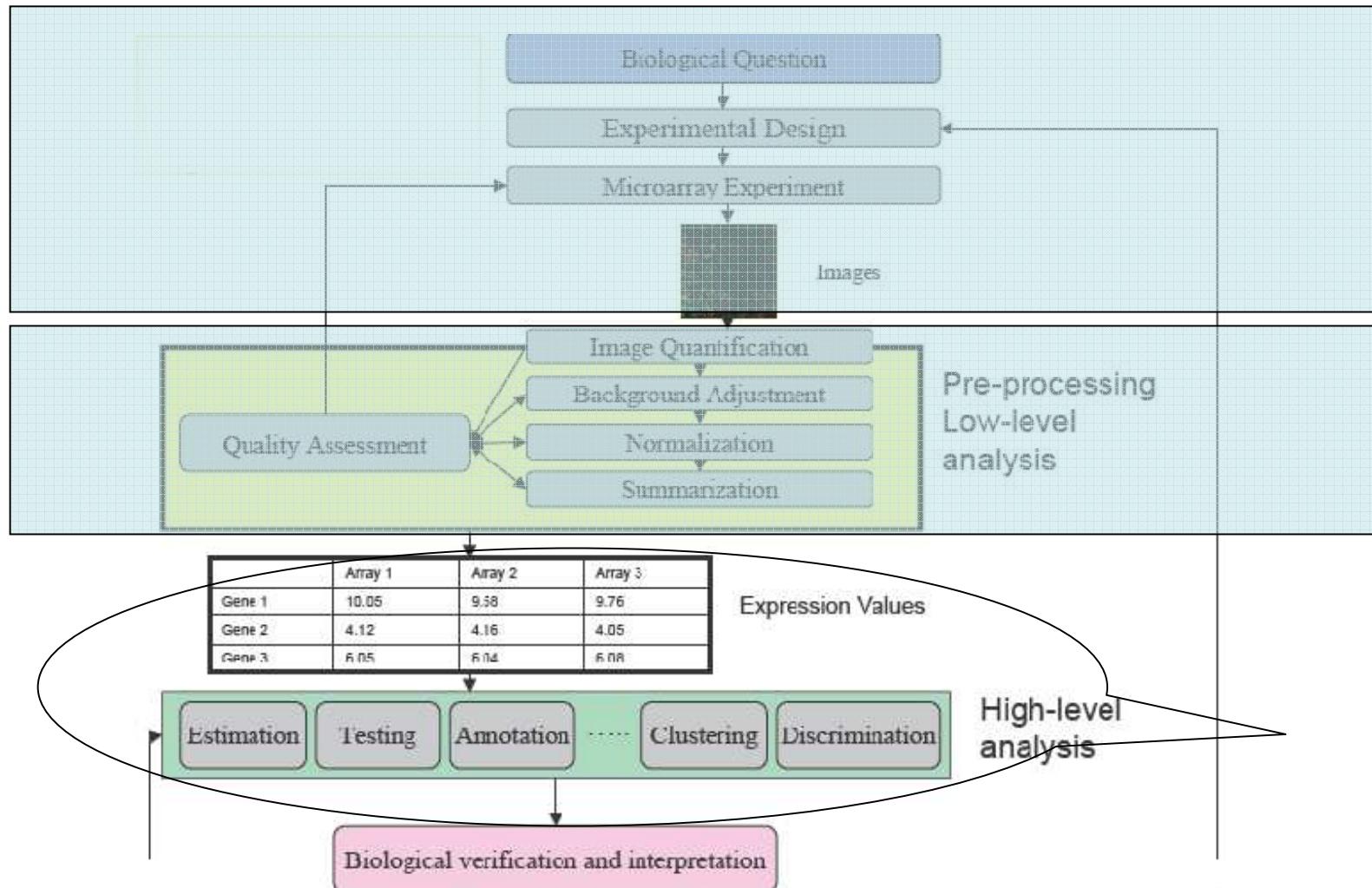
Preprocesado (1) El control de calidad

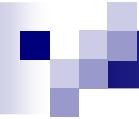


Preprocesado (2) Normalización



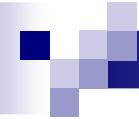
El análisis de los datos





Tipos de análisis

- Los investigadores suelen estar interesados en distintos tipos de cuestiones:
 - Encontrar genes **diferencialmente expresados** entre dos o más condiciones o a lo largo del tiempo.
 - Identificar **nuevos subtipos** en una población
 - Descubrir **patrones de expresión** característicos.
 - **Predecir la respuesta al tratamiento** or **clasificar un nuevo individuo** utilizando información molecular.
 - **Identificar genes co-regulados** o expresándose en la misma **ruta metabólica**.

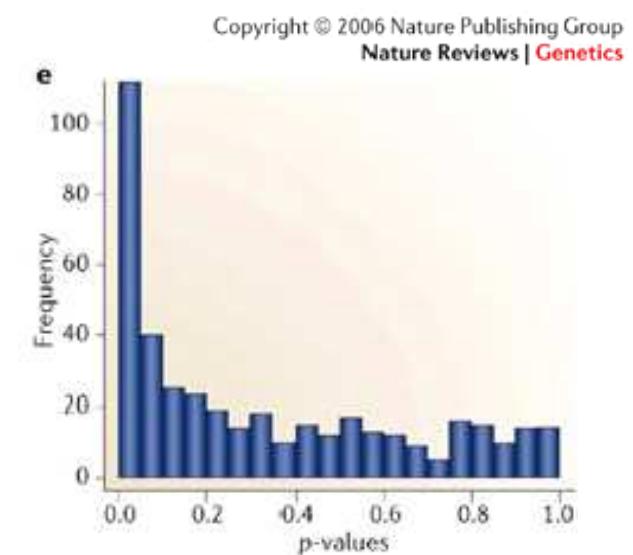
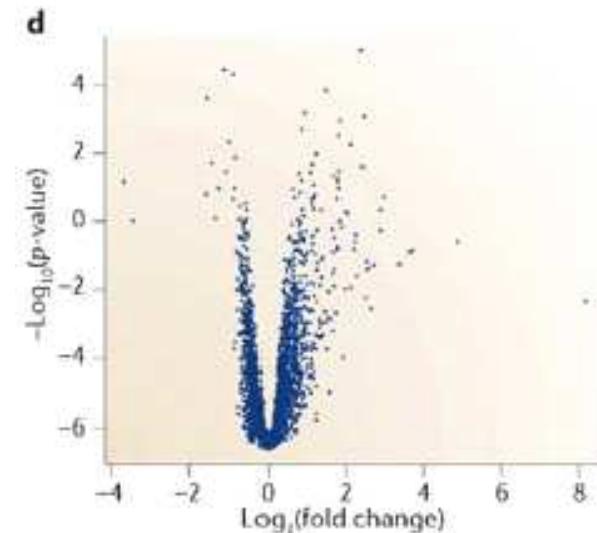


Métodos de análisis

- Para cada problema existen múltiples métodos
 - **Modelos lineales, pruebas-t con shrinkage** para estudios de expresión diferencial
 - Distintos tipos de **análisis de conglomerados** (“**clustering**”) para descubrir patrones de corregulación
 - **Métodos de clasificación** tradicionales (**LDA, kNN**) y modernos (**SVM, PAM**) para construir predictores
 - Métodos de **análisis basados en la GO (GSEA)** para buscar significación biológica
 - Y muchos más ...

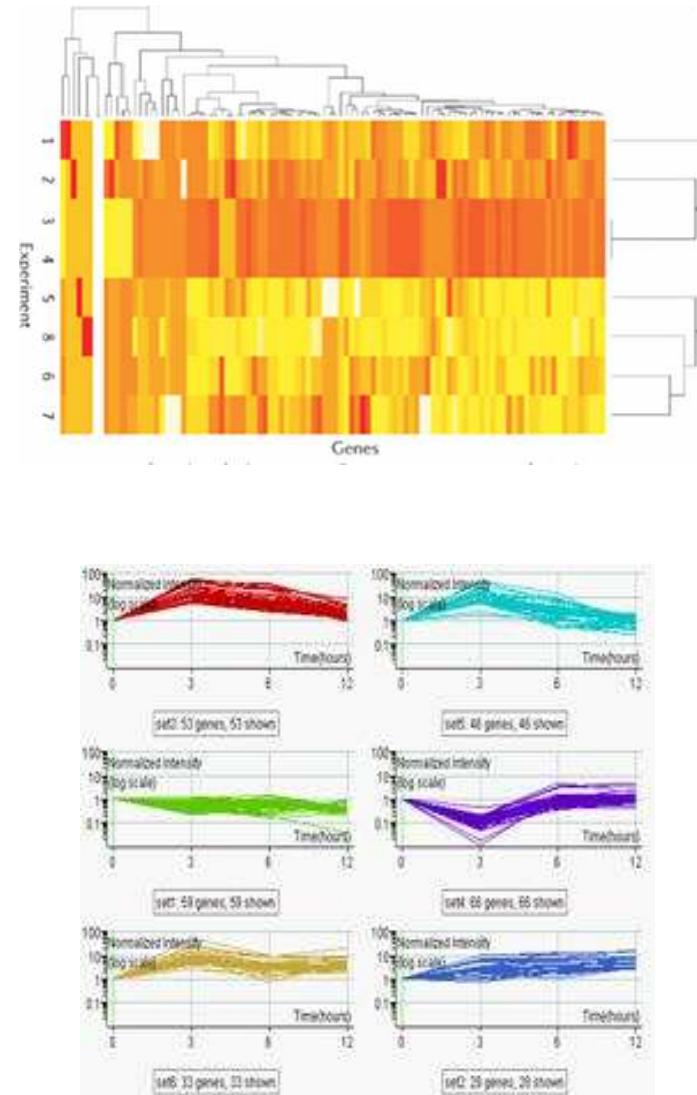
Tests para expresión diferencial

- Para comparar dos o más grupos:
extensiones del test t
 - El tamaño muestral suele ser ↓
 - Se compensa estimando la varianza de cada gen a partir de la de todos los genes
 - SAM, Empirical Bayes, ...
- Para cada gen se hace un test
→ Problema de multiplicidad
 - Es preciso hacer ajustes para múltiple testing
 - O estimar la tasa de falsos positivos (FDR)



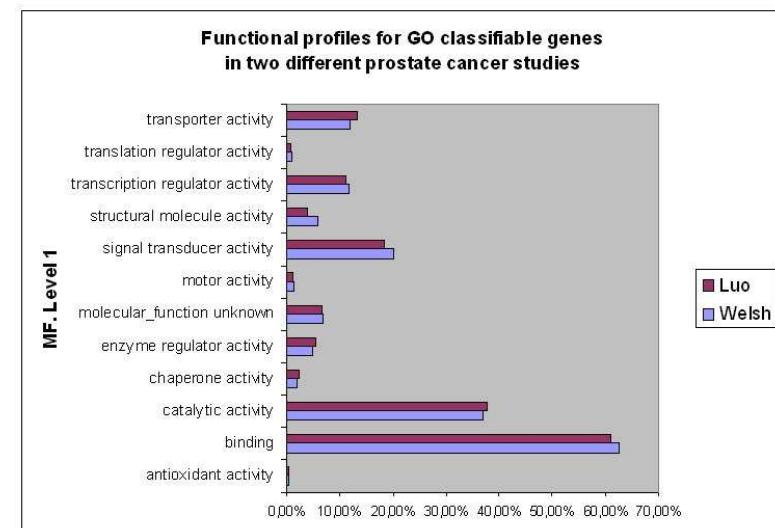
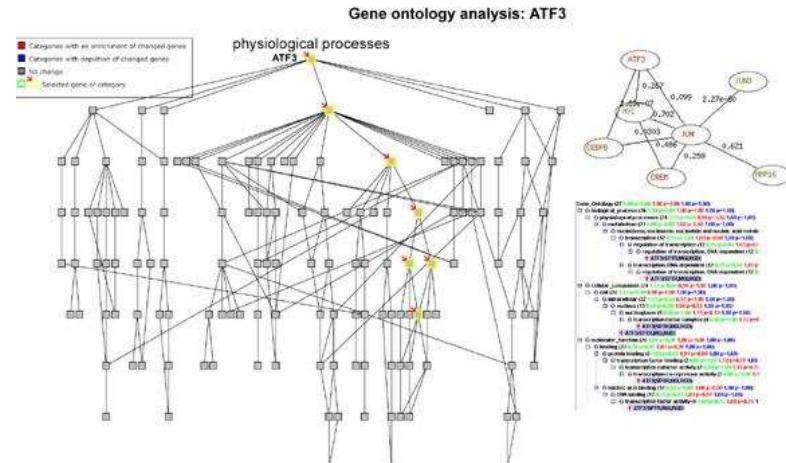
Análisis de conglomerados

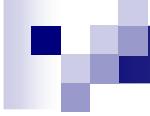
- Los genes no varían de forma independiente
- El análisis de conglomerados permite descubrir grupos de genes que varían de forma similar
- Puede utilizarse también para agrupar muestras: (fenotipos similares) → descubrimiento de subclases



Análisis basados en la GO

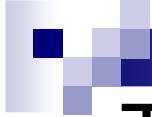
- Los resultados de los estudios de microarrays suelen ser largas listas de genes
- Para contribuir a su interpretación podemos
 - Proyectarse en bases de datos de anotaciones como la GO o KEGGS
 - Estudiar si hay clases funcionales enriquecidas entre los genes seleccionados
 - Agrupar los genes por su similitud funcional





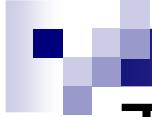
Conclusiones y perspectivas

- Los experimentos con microarrays han revolucionado el estudio de la *genómica funcional*
 - Mejorando el conocimiento de la función de los genes a partir de la similitud de patrones de expresión
 - Mejorando el conocimiento de las familias de genes:
 - Permiten incluir nuevos genes en las familias
 - Descubren patrones de expresión coordinados
 - Aumenta el número de familias conocidas de genes
- Como toda tecnologías los tiene sus limitaciones
 - Algunas como la baja reproducibilidad o la calidad del genoma se solucionaran con el tiempo
 - Otras como el uso adecuado de sus posibilidades dependen del buen (o mal) uso que se haga de ellas



The Promise of Microarray Technology in Treating Disease ([NCBI](#)) (1)

- *Now that you understand the concept behind array technology, picture this:*
- *A hand-held instrument that a physician could use to quickly diagnose cancer or other diseases during a routine office visit.*
- *What if that same instrument could also facilitate a personalized treatment regimen-exactly right for you?*



The Promise of Microarray Technology in Treating Disease ([NCBI](#)) (2)

- ***Personalized drugs, Molecular diagnostics and***
 - ***Integration of diagnosis and therapeutics***
- *These are the long-term promises of microarray technology*
- *Maybe not today or even tomorrow, but someday*
- *For the first time, arrays offer hope for obtaining global views of biological processes by providing a systematic way to survey DNA and RNA variation*