

Bibliotecas de ADN

Una biblioteca de ADN es esencialmente una representación de todo el conjunto de ADN (ya sea genómico o complementario a los ARN).

Una biblioteca de ADN genómico se puede definir como la representación de todo el ADN genómico de un organismo, incluyendo regiones codificantes y no codificantes por igual.

La biblioteca de ADNc, por otro lado, se construye mediante la transcripción inversa (retrotranscripción) de los ARNm, y por lo tanto solo representa las regiones codificantes de proteínas (CDS) del genoma.

Por lo tanto, las dos bibliotecas se utilizan para diferentes propósitos que se describirán en detalle. Según el tipo de biblioteca,

(a) la fuente de ADN,

(b) el tipo de vector utilizado,

(c) las metodologías experimentales

El principio básico de todas las bibliotecas de ADN implica:

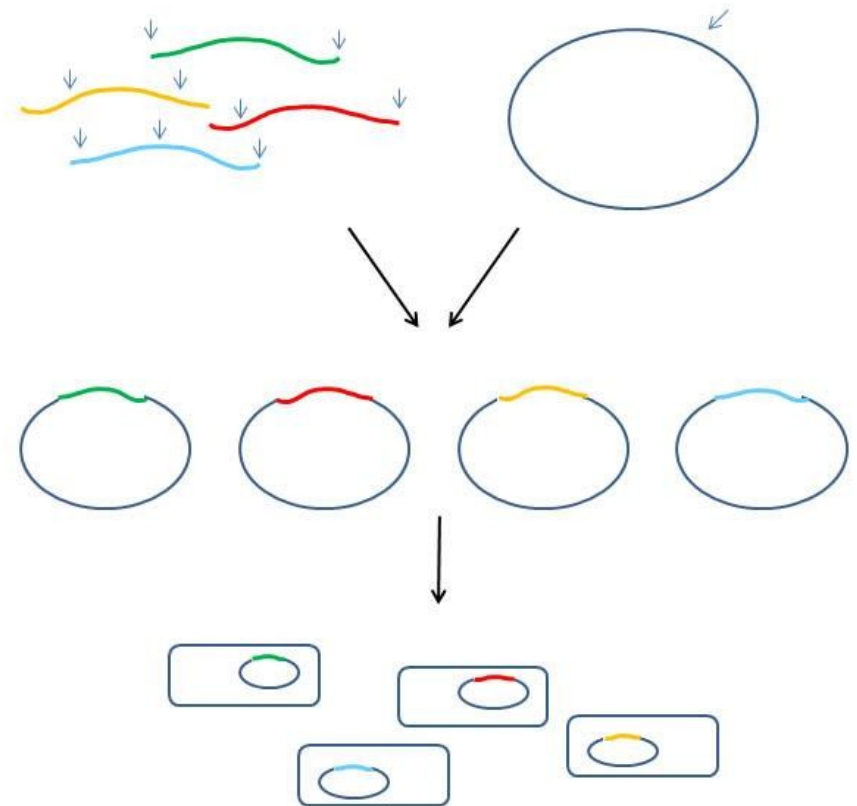
- Preparación de fragmentos de ADN representativos;
- Preparación del vector apropiado;
- Clonado y transformación;
- Selección y screening.

Bibliotecas de ADN genómico

Los fragmentos de ADN genómico contendrán no solo secuencias de codificación, pero también intrones, regiones promotoras, otros tipos de regiones reguladoras, varios elementos repetitivos etc. Por lo tanto, tales bibliotecas serían útiles en la detección de regiones no codificantes de ADN.

Figura 3.1. Una representación esquemática de las bibliotecas de ADN genómico. Esencialmente, todo el genoma se utiliza para la fragmentación y clonado en diferentes vectores. Los productos fragmentados se ligan en los vectores y éstos se transforman en bacterias. Idealmente, cada clon bacteriano debe incluir un diferente fragmento genómico.

Cómo se fragmenta el ADNg???

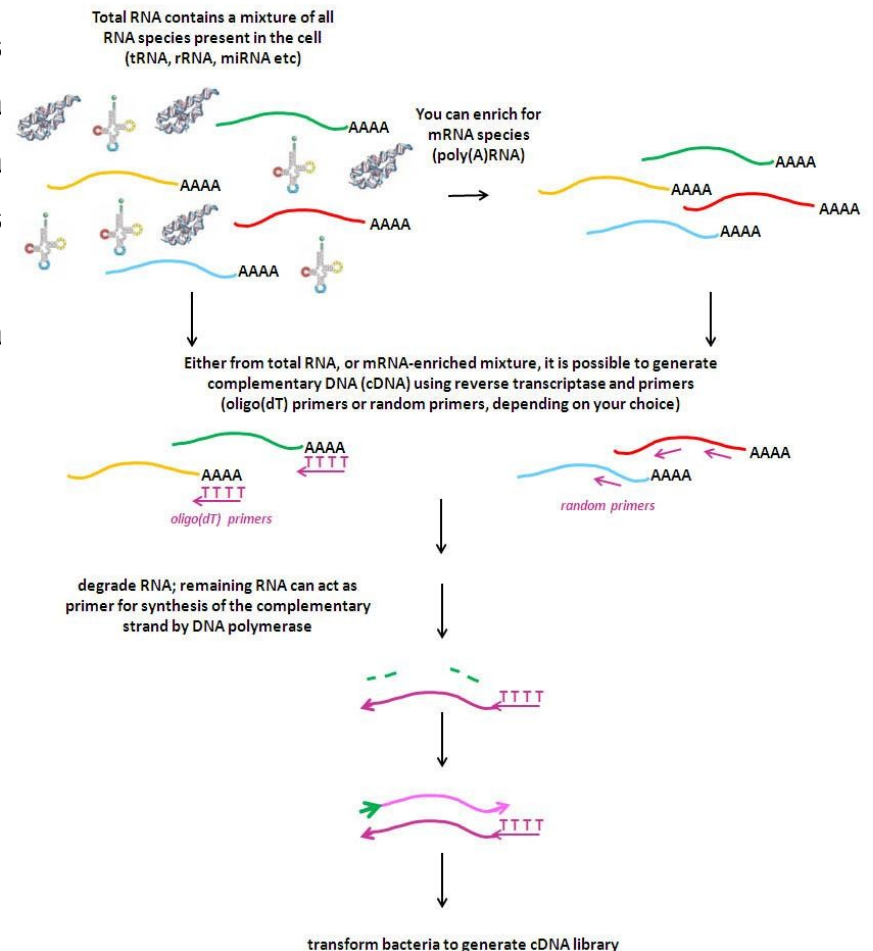


Bibliotecas de ADN copia o complementario

El ADNc es la abreviatura de ADN complementario o copia, es decir, ADN que ha sido sintetizado como complementario al ARNm maduro, representando solo las regiones de codificante o exones. Por lo tanto, la biblioteca de ADNc se refiere a una representación de la mayoría, o todas, las transcripciones de ARNm maduras aisladas de un tipo de célula, o en otras palabras, representativo de todos los genes expresados en una célula dada en un momento dado y un condición dada.

Figura 3.2. Representación de la síntesis de ADNc y la biblioteca de ADNc.

Para clonar solo los genes expresados, primero se debe convertir el ARNm (que es altamente susceptible a la degradación por RNasas) a cDNA (que es resistente a la degradación de RNasas -que son mucho más abundantes que las DNAsas) mediante el uso de una transcriptasa inversa (existen 2, la más utilizada es la del virus MLMV) y primers (cuales??). El procedimiento de síntesis de ADNc podría comenzar a partir de ARN total o de un enriquecimiento de ARNm (como??).



Screening de bibliotecas

Una vez que se obtiene la biblioteca es necesario encontrar un gen particular de interés, generalmente usando un ADN homólogo como sonda, que debe "marcarse".

Esta sonda podría ser el homólogo del fragmento de ADN de una especie diferente (como la región de codificante de un gen humano usado como una sonda para seleccionar una biblioteca de ADNc de ratón del tejido de interés), o parte de una región de codificación para "buscar" un región promotora, o parte de un gen de un tejido para detectar homólogos en un tejido diferente, etc. El principio detrás del screening basado en sondas es que las especies de ADN o ARN monocatenarias se hibridarán a secuencias con cierto porcentaje de identidad en la biblioteca.

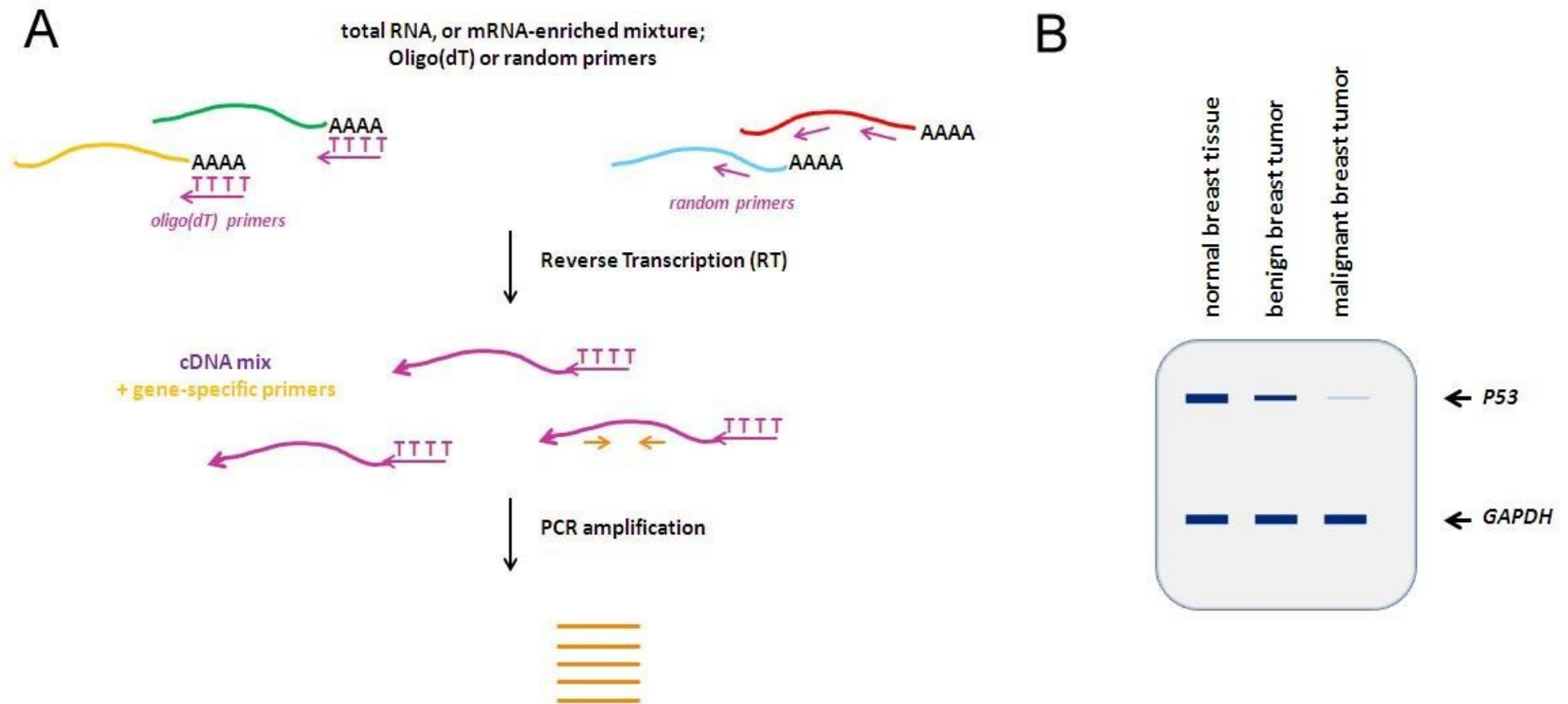
PCR de transcripción inversa (RT-PCR) ES CUALITATIVA

no debe confundirse con la PCR en tiempo real; qPCR

Como el ARN es bastante menos estable que el ADN, en otras palabras, es más susceptible a la degradación por RNasas que son muy abundantes, es más difícil trabajar con las especies de ARNm para comparar los niveles de expresión. Por lo tanto, los investigadores a menudo se utilizan el ADNc que se sintetiza a partir del ARNm por transcripción inversa para comparar de niveles de expresión entre diferentes muestras. Como con cualquier otro experimento, se requiere un control interno para mantener cierto nivel de cuantificación: se utiliza genes que se llaman estándares (antiguamente llamados housekeeping como la **gliceraldehído-3-fosfato deshidrogenasa** (GAPDH o G3PDH) o el gen b-tubulina se utilizan para la normalización entre diferentes muestras. La mezcla de ADNc (generalmente la 1ª cadena es suficiente para este propósito) generada a partir de ARNm de diferentes muestras luego se someten a amplificación por PCR mediante primers específicos para el gen de interés.

Actualmente se sabe que todos los genes pueden variar dependiendo del fenómeno biológico que se estudie, por lo que hay que determinar experimentalmente que gen será el correcto estándar en cada experimento y no dar por sentado que GAPDH o b-tubulina serán los adecuados.

Figura 3.3. Reacción en cadena de la polimerasa de transcripción inversa (RT-PCR). (a) Un resumen esquemático de la RT-PCR (b) un ejemplo hipotético de un resultado de RT-PCR en gel de agarosa, que muestra amplificaciones con primers específicos para control interno GAPDH y transcripción del gen P53 en tres muestras de tejido diferentes.



Northern blots

La transferencia *Northern* es una técnica utilizada para el análisis de ARN, y más comúnmente ARNm, entre muestras, y fue desarrollado por James Alwine, George Stark y David Kemp.

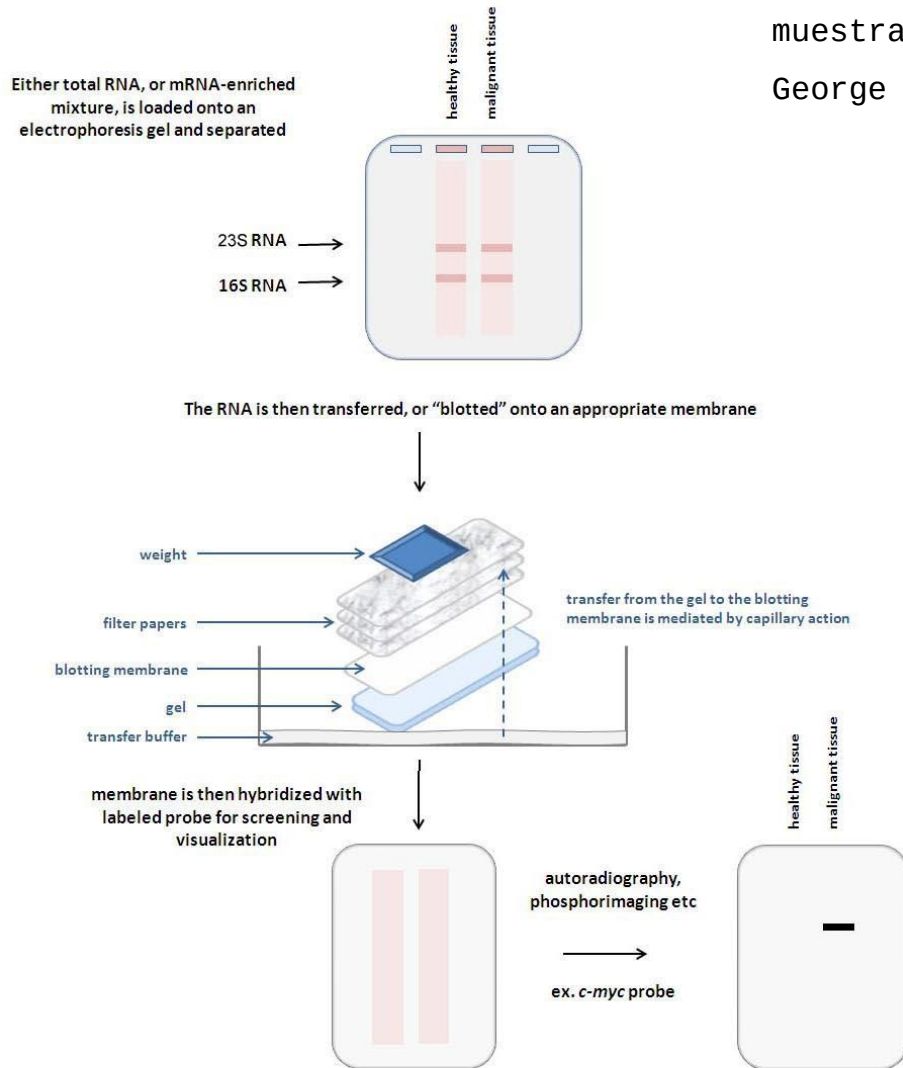


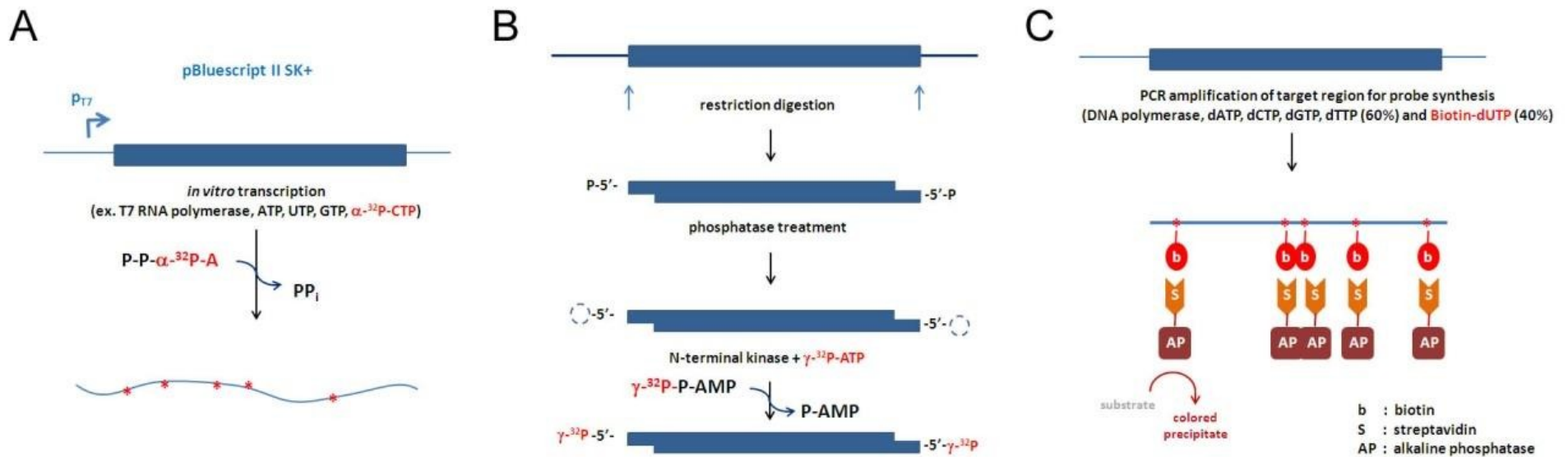
Figura 3.6.

Un esquema simplificado de estrategias de preparación de sondas.

(a) Preparación de la sonda de ARN por transcripción *in vitro*. Para el marcado radiactivo, uno de los nucleótidos que se incorpora a la transcripción debe contener un radioisótopo.

(b) marcado final de un fragmento de ADNc. El fosfato gamma del ATP se transfiere al fragmento de ADN.

(c) Generación de la sonda basada en PCR usando desoxinucleótido biotilado. La afinidad de la biotina/estreptavidina es cercana a la de un enlace covalente. Una enzima como la fosfatasa alcalina se puede conjugar a estreptavidina, que luego convierte un sustrato incoloro en un precipitado coloreado para la detección.



Análisis de microarrays

Tanto en RT-PCR como en *Northern blots*, uno debe tener una idea sobre qué gen podría expresarse para poder preparar las sondas específicas para el ARNm en estudio. Sin embargo, si se desea abordar un tema más general e imparcial cuestión de "qué genes se pueden expresar de manera diferente entre las muestras", es necesario un método más general. La tecnología de microarrays o chips de ADN es bastante útil para este propósito y se ha convertido en un estándar para estudiar perfiles globales de expresión génica.

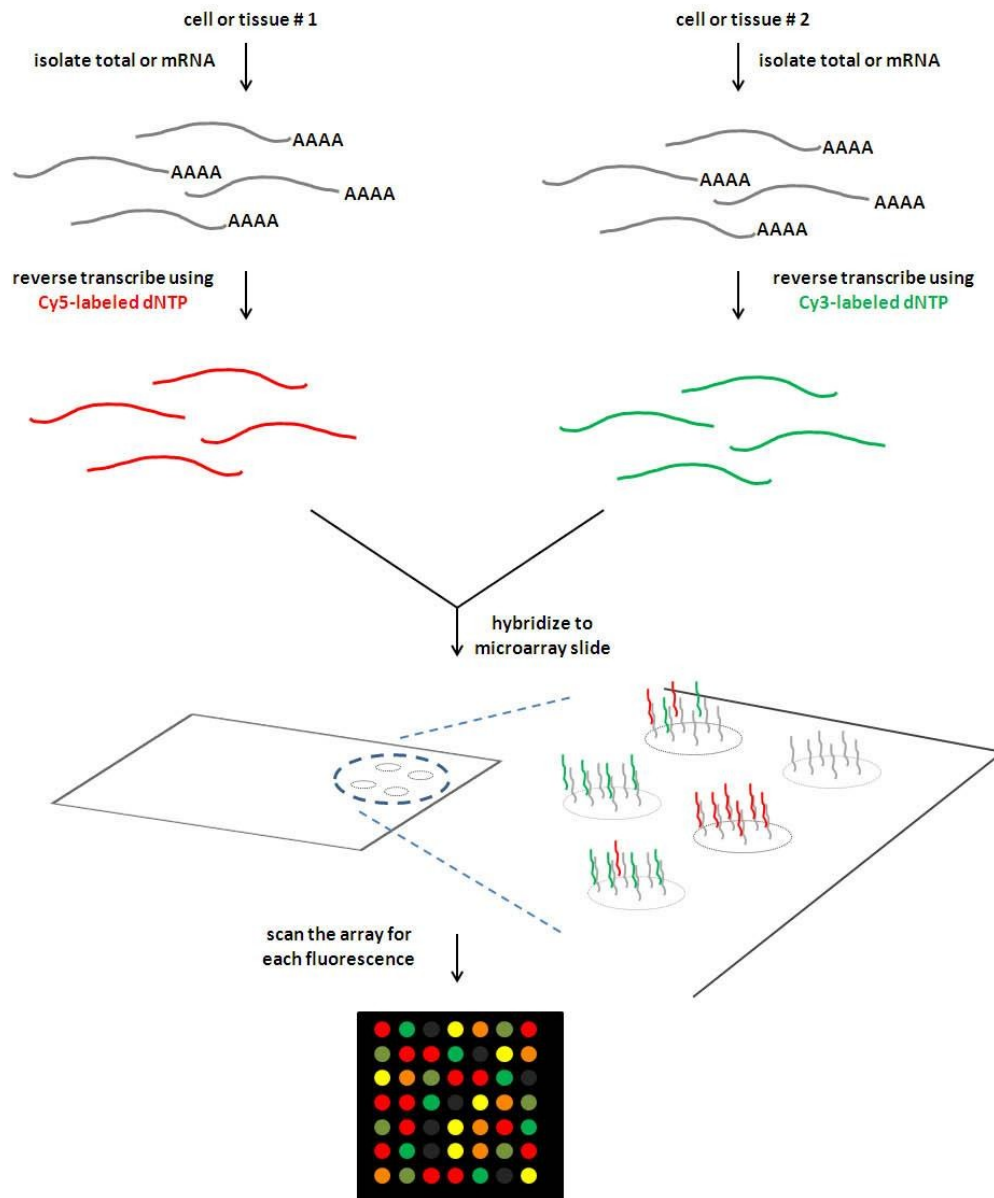


Figura 3.7. Una visión general del análisis de microarrays para perfiles de expresión de genes. Por lo general, ARN de dos diferentes muestras se aíslan y se transcriben de forma inversa usando desoxinucleótidos marcados fluorescentemente (con Cy3 (verde) y Cy5 (rojo)). Estos ADNc marcados se mezclan e hibridan con el microarrays. Cada punto en microarrays corresponde a un gen diferente (mostrado aquí como líneas onduladas grises en el inset), y dependiendo de qué genes se transcriben en cada muestra (y en qué cantidad), los ADNc etiquetados hibridizan de manera diferente a cada lugar. Cuando la matriz se escanea para cada fluorescencia y se combina la información, cada punto no solo mostrará un color diferente (diferentes escalas de rojo, verde o amarillo), pero también tienen una medición de intensidad diferente registrada en un archivo Excel o similar para un análisis posterior.

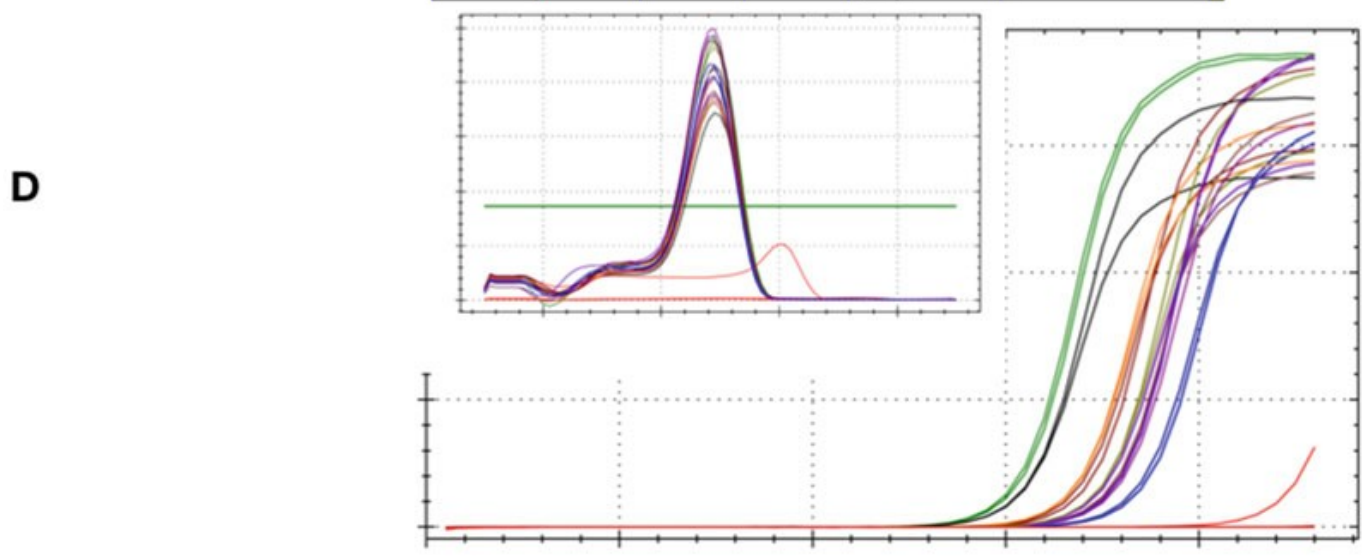
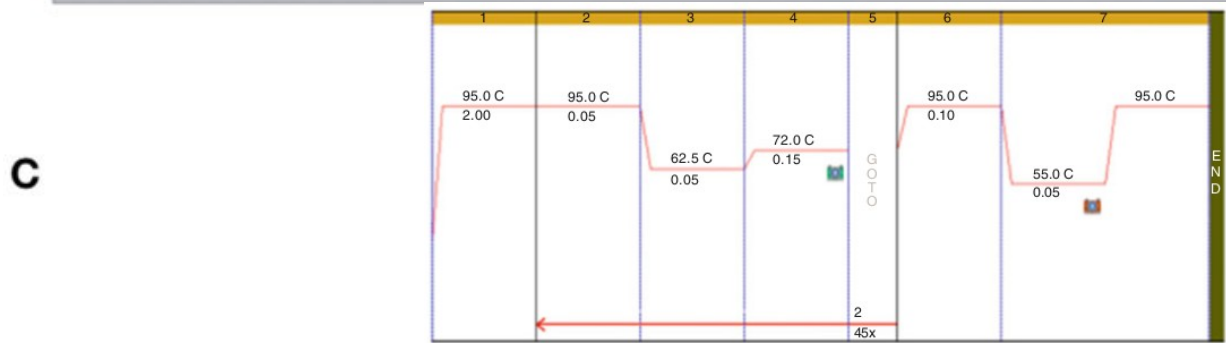
qPCR

A

CD1: GCAATAGATTTAGCTAAAGAGGATATT CD2: CTCTACGTTGATAGAGTAGCTAAAC
 CD3: CAATAGATTTAGCTAAAGAGGATATTTCT CD4: TCTACGTTGATAGAGTAGCTAAACAT
 CD5: GATTTAGCTAAAGAGGATATTTCTCT CD6: TCCTCTACGTTGATAGAGTAGCT

B

5' -AGCAGCAATAGATTTAGCTAAAGAGGATATTTCTCTTAAGTCAATAGAAATAAATTTATTAGGATGTAATATGTTTAGCTACTCTATCAACGTAGAGGAGACT-3'
 3' -TCGTCGTTATCTAAATCGATTCTCCTATAAAGAGGATTCAGTTATCTTTATTAAATAATCCTACATTATACAAATCGATGAGATAGTTGCATCTCCTCTGA-5'



E

	CD1/CD2	CD1/CD4	CD1/CD6	CD3/2	CD3/4	CD3/6	CD5/2	CD5/4	CD5/6
Cq1	29.54	32.58	29.74	33.72	32.43	33.79	34.68	35.66	33.75
Cq2	29.67	33.23	29.97	33.96	32.50	34.73	34.69	35.92	34.45
Average Cq	29.60	32.91	29.85	33.84	32.47	34.26	34.69	35.79	34.10
Cq range					6.19				

(a) Selección de primers dirigidos al gen de interés (son muy diferentes a los de PCR comunes).

(b) región de ADN de interés.

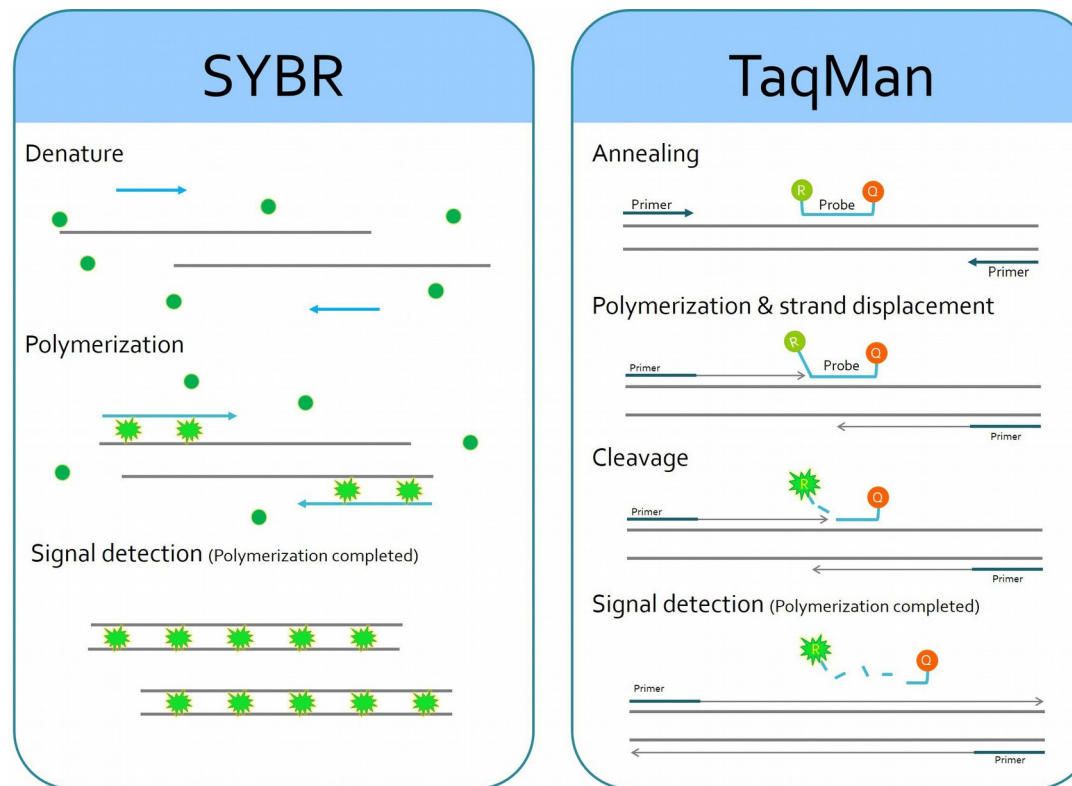
(c) Condiciones de amplificación y curva de melting.

(d) resultados de la curva de melting y de amplificación

(e) Cq

qPCR

- Cuantitativa absoluta vs relativa (más usada)
- SybrGreen vs Taqman
- Cálculo de Cq (umbral vs 2da derivada)
- Cálculo de la eficiencia
- Genes de referencia



Cuantificación relativa.

La cuantificación relativa no requiere estándares con concentraciones determinadas. Esta técnica se utiliza para obtener la magnitud de los cambios en los niveles de expresión genética de un gen en estudio en comparación con uno o más genes de referencia. Los cálculos en cuantificación relativa de expresión genética se basan en la comparación de los valores Ct utilizando la eficiencia de la reacción de la PCR como factor de corrección. Sin embargo, hay un modelo que no requiere la eficiencia de la reacción para acceder a un factor de corrección. Este modelo supone una eficiencia óptima e idéntica (correspondiente al 100%) en la eficiencia de reacción en las PCR en tiempo real tanto del gen en estudio como del gen de referencia. Este es el método delta-delta Ct que sólo es aplicable para una estimación rápida de la proporción relativa de la expresión genética en estudio. El método delta-delta Ct expresa la proporción obtenida de la relación entre los valores Ct de la muestra y los valores Ct del control tal y como se muestra en la siguiente ecuación:

$$\text{ratio} = 2^{-[\Delta\text{CP}_{\text{sample}} - \Delta\text{CP}_{\text{control}}]}$$

$$\text{ratio} = 2^{-\Delta\Delta\text{CP}}$$

Otro modelo. Las diferentes eficiencias de la PCR tanto para los genes en estudio como para los genes de referencia se toman en cuenta como se muestra en la siguiente ecuación:

$$\text{ratio} = \frac{(E_{\text{target}})^{\Delta\text{CP}_{\text{target}}(\text{control} - \text{sample})}}{(E_{\text{ref}})^{\Delta\text{CP}_{\text{ref}}(\text{control} - \text{sample})}}$$

En esta ecuación la diferencia del gen en estudio se expresa en una muestra frente a un control en comparación con un gen de referencia. E_{target} representa la eficiencia de la PCR en tiempo real del amplicon en estudio; E_{ref} representa la eficiencia de la PCR en tiempo real del gen de referencia; $\Delta\text{CP}_{\text{target}}$ es la desviación en Ct del control menos la muestra del gene en estudio; y $\Delta\text{CP}_{\text{ref}}$ es la desviación en Ct del control menos la muestra del gen de referencia.

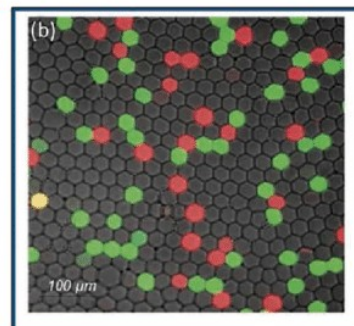
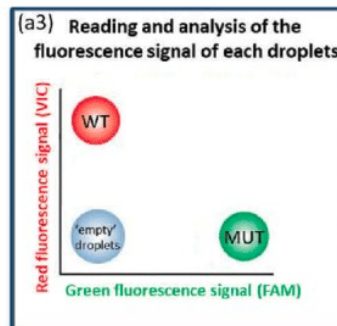
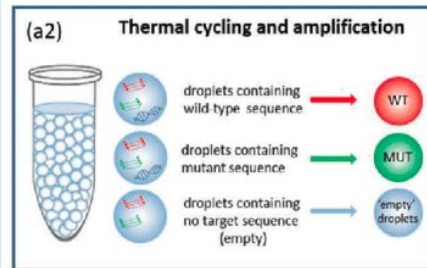
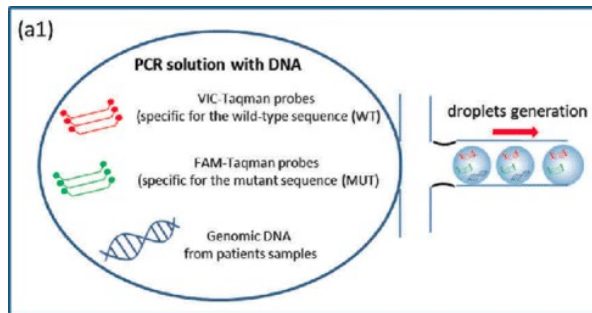
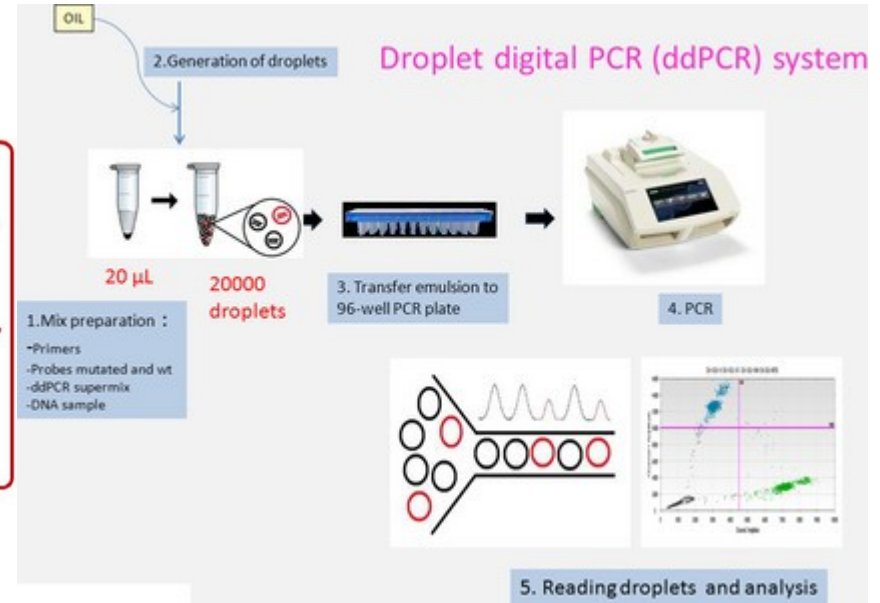
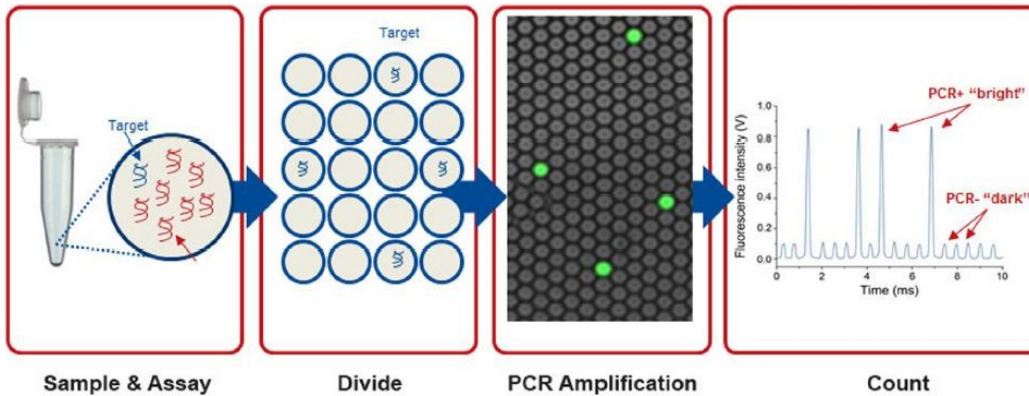
Como se mencionó, en este modelo también es necesario conocer la eficiencia de PCR de cada gen estudiado. Las eficiencias de la PCR en tiempo real se calculan a partir de las pendientes de la curva estándar obtenidas después de realizar diluciones seriadas con las reacciones de la PCR en tiempo real de acuerdo a la siguiente fórmula: $E=10[-1/\text{slope}]-1$.

Se ha observado que la cuantificación relativa de ARNm tiene algunas limitaciones. En primer lugar, se puede introducir un sesgo estadístico importante cuando hay grandes diferencias en los niveles de expresión del gen en estudio ($Ct_{\text{min}} \geq 20$ $Ct_{\text{max}} \leq 30$) y del gen normalizador lo que pueden conducir a una interpretación biológica equivocada y en segundo lugar, es difícil encontrar genes de referencia adecuados. Por estas razones es recomendable la utilización de más de un gen de referencia.

Cuantificación absoluta.

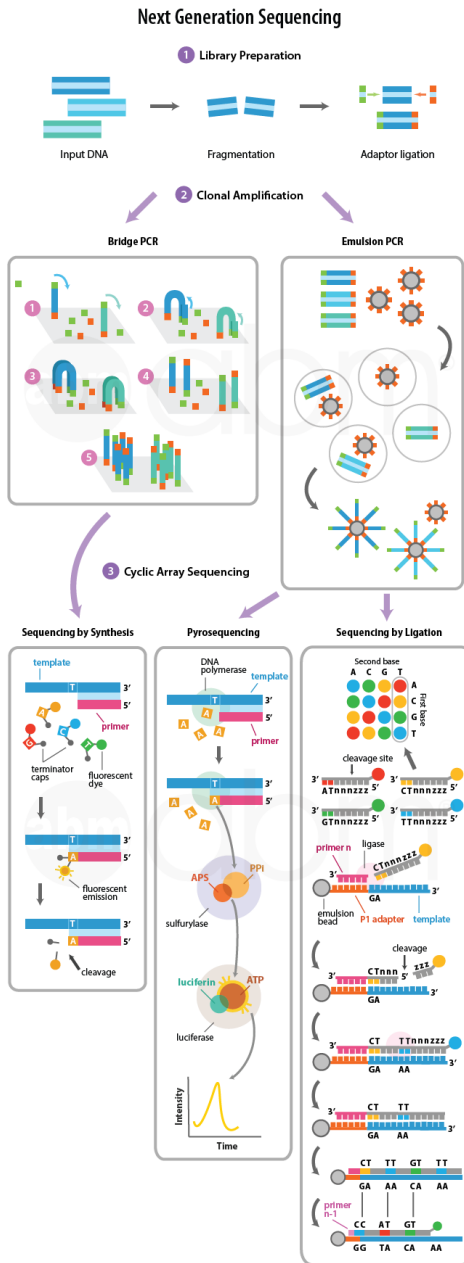
Las curvas de amplificación se comparan con las curvas donde el número exacto de moléculas del ADN target son utilizadas para la amplificación. Como requiere tener esto último casi no es utilizado.

Digital PCR



El principio básico implica la dilución y la división extrema de la muestra en millones de unidades separadas que idealmente no contienen partículas o una sola partícula. Cada unidad contiene todos los reactivos necesarios para una reacción de PCR, y básicamente funciona como un reactor de micro-PCR. Si la unidad contiene la plantilla de interés, la amplificación por PCR produce una señal positiva. Si no hay una plantilla, no hay señal. La cuantificación es binaria, de ahí el término "digital". Si se conoce el número de unidades de partición, la cantidad inicial de moléculas objetivo puede estimarse conociendo el número total de señales positivas a negativas [2]. En una dilución extrema, esto simplemente representa la relación de las unidades positivas a negativas, ya que no hay plantilla o una plantilla única dentro de la unidad.

Next Generation Sequencing/Deep Sequencing/NGS



Pirosecuenciación

En la pirosecuenciación, la reacción de secuenciación se controla mediante la liberación del pirofosfato durante la incorporación de nucleótidos. Se agrega un solo nucleótido al chip de secuenciación que conducirá a su incorporación de una manera dependiente del molde. Esta incorporación dará como resultado la liberación de pirofosfato que se utiliza en otras reacciones químicas cuyo producto generan luz. La emisión de luz es detectada por una cámara. Cualquier base no incorporada es degradada por la pirofosfatasa antes de la adición del siguiente nucleótido. Este ciclo continúa hasta que se completa la reacción de secuenciación. Desventajas: Alto costo de los reactivos y alta tasa de error en cadenas de 6 o más nucleótidos iguales.

Secuenciación por síntesis:

La secuenciación por síntesis utiliza la incorporación de a una vez de nucleótidos fluorescentes y modificados específicamente para que solo se incorpore 1 nucleotido por vez en la reacción de secuenciación de ADN. Es utilizada por las plataformas Illumina NGS. Los nucleótidos utilizados en este método se han modificado de dos maneras:

1) cada nucleótido se une de forma reversible a una molécula fluorescente única con longitudes de onda de emisión únicas

2) cada nucleótido también se bloquea de forma reversible asegurando que solo se incorporará un solo nucleótido por ciclo. Los cuatro nucleótidos se agregan al chip de secuenciación y después de la incorporación de nucleótidos, las bases de ADN restantes se eliminan por lavado. La señal fluorescente es fotografiada (literalmente); tanto la molécula fluorescente como el grupo terminador se escinden y se lavan. Este proceso se repite hasta que se completa la reacción de secuenciación. Este sistema puede superar las desventajas del sistema de pirosecuenciación, pero igual tiene desventajas: A medida que avanza la reacción de secuenciación, la tasa de error de la máquina también aumenta. Esto se debe a la eliminación incompleta de la señal fluorescente que conduce a niveles más altos de ruido de fondo. Por lo general no se puede generar lecturas de más de 100 o 120 nt. Entonces como leo un genoma? O ARNm de más de 120nt?

Secuenciación por ligación. Antes de la secuenciación, el ADN se amplifica por PCR en emulsión. Las perlas resultantes, cada una con copias individuales de la misma molécula de ADN, se depositan en un portaobjetos de vidrio. Aquí, un grupo de todos los oligonucleótidos posibles de una longitud fija se y marcados de colores específicos se ponen en contacto con el ADN a secuenciar y un primer complementario a un secuencia común adicionada durante la construcción de la biblioteca. Los oligos son reconocidos y ligados. La ligación (por la ADN ligasa) solo se da para secuencias coincidentes y da como resultado una señal informativa del nucleótido en esa posición. Luego se corta el ADN en la posición +5 y se repite el proceso. Finalmente se reemplaza el primer por otro en posición -1 y se repite todo el proceso. 5 rondas completan el proceso.

Se ha informado que este tipo de secuenciación tiene problemas para secuenciar secuencias palindrómicas.

Paired-End Sequencing:

Secuencia desde ambos extremos de un fragmento mientras se realiza un seguimiento de los datos emparejados. Con este método, la reacción de secuenciación comenzará desde un extremo del fragmento. Una vez completado, el fragmento se desnaturaliza y un primer de secuenciación se hibrida a la otra hebra. El fragmento se secuencia de nuevo. El uso de este método permitirá una confirmación adicional de la precisión de la secuencia o podría usarse para aumentar la longitud total de la lectura.

Secuenciación de semiconductores iónicos. La secuenciación de semiconductores de iones utiliza la liberación de iones de hidrógeno durante la reacción de secuenciación para detectar la secuencia de un grupo. Cada grupo está ubicado directamente encima de un transistor semiconductor que es capaz de detectar cambios en el pH de la solución. Durante la incorporación de nucleótidos, se libera un solo H^+ en la solución y el semiconductor lo detecta. La reacción de secuenciación en sí procede de manera similar a la pirosecuenciación pero a una fracción del costo. Desventajas: Alta tasa de error sobre tramos homopoliméricos.

