



# Tecnologías de Secuenciación Masiva: Características generales

<b>Apellidos, nombre</b>	Vilanova Navarro, Santiago ( <a href="mailto:sanvina@upv.es">sanvina@upv.es</a> ) Gadea Vacas, José ( <a href="mailto:jgadeav@ibmcp.upv.es">jgadeav@ibmcp.upv.es</a> )
<b>Departamento</b>	Departamento de Biotecnología
<b>Centro</b>	Universitat Politècnica de València



## 1 Resumen de las ideas clave

La aparición de las tecnologías de secuenciación masiva ha hecho posible que los proyectos de secuenciación de genomas estén al alcance de todos. Hoy en día, cualquier laboratorio con presupuesto ajustado pueden aventurarse a realizar este tipo de proyectos en sus líneas de investigación. Las propuestas son normalmente subcontratadas a empresas de secuenciación, que realizan el trabajo y parte del análisis bioinformático, dejando en manos del investigador la interpretación final de los datos.

Para poder seleccionar el tipo de proyecto de secuenciación más adecuado a nuestros objetivos, es esencial conocer la base de estas tecnologías. Solo con este conocimiento podremos discutir con los responsables de las empresas de secuenciación el diseño experimental y la tecnología más adecuada para nuestros fines

En este artículo se pretende revisar las características básicas de las tecnologías de secuenciación masiva de segunda generación, de manera que el alumno pueda ser consciente de los diferentes pasos del proceso, para que les sea más fácil la toma de decisiones cuando tenga que enfrentarse a proyectos de esta índole.

## 2 Objetivos

Una vez que el alumno haya leído con detenimiento este documento, será capaz de:

1. Entender el cambio de paradigma en estas tecnologías que permite la paralelización de secuencias
2. Describir y explicar los pasos a seguir para realizar una secuenciación masiva con tecnologías de segunda generación
3. Entender el concepto de librería en el contexto de la secuenciación masiva
4. Entender la necesidad de una amplificación por PCR para generar miles de moléculas idénticas, anteriormente a la secuenciación
5. Entender la química de la secuenciación masiva basada en nucleótidos terminadores reversibles.

## 3 Introducción

Empecemos entendiendo cómo las tecnologías de secuenciación masiva (NGSs) suponen un cambio radical en el proceso de secuenciación que utiliza la tecnología de Sanger.

La secuenciación del genoma humano de referencia, terminada en el año 2001, costó 2.7 billones de dólares y tardó una década en terminarse. La tecnología utilizada fue la tecnología de secuenciación de Sanger, la única estrategia disponible en los últimos 30 años para secuenciar ácidos nucleicos. Un año después de este esfuerzo, Craig Venter (el director de la iniciativa privada de esta secuenciación) ofreció 500.000 dólares como estímulo para quien fuera capaz de ofrecer una tecnología capaz de secuenciar un genoma humano por menos de 1000 dólares.



Diferentes empresas, como Roche o Illumina, se ponen manos a la obra. premio, a través del X Archon Genomics Price Foundation, asciende pronto de recompensa.

En abril de 2017, Dante Labs ofrece la secuencia de un genoma humano completo por 900 dólares. También en 2017, Beijing Genomics Institute lo ofrece por 600 dólares. In 2018, en Amazon Prime Day, Dante Labs lo ofrece por 349, y en noviembre, por el Black Friday, por menos de 200 dólares. Veritas Genetics, por el mismo precio, lo proporciona en dos días. En marzo del mismo año, Illumina anuncia que lo ofrece por 100 dólares.

Esta revolución en tan poco tiempo ha sido posible gracias a un cambio de tecnología: el nacimiento de las tecnologías de secuenciación masiva de segunda generación, donde la secuenciación de todos los fragmentos se realiza simultáneamente, mientras que en la tecnología anterior de Sanger se realizaba fragmento a fragmento. Esta paralelización es uno de los aspectos clave del éxito de estas tecnologías, el más importante para la secuenciación de grandes proyectos genoma en tiempos cortos.

## 4 Desarrollo

Para entender el funcionamiento de estas tecnologías, es importante tener en cuenta los pasos básicos de que consta este proceso de secuenciación.

Inicialmente, tres empresas desarrollaron tecnologías de secuenciación que permitían la paralelización que permitía desbancar a la tecnología de Sanger para realizar proyectos de secuenciación genómica. Las tres tienen en común un paso de amplificación de cada fragmento inicial mediante PCR. Estas tecnologías se llaman de Segunda Generación. La tecnología que sigue el método de Sanger se denomina en este contexto de Primera Generación.

Estos sistemas de secuenciación comparten los mismos pasos en la realización de todo el proceso:

1. Generación de una colección de fragmentos a secuenciar (Esto es lo que se suele llamar la generación de "librerías", aunque esto sea una traducción desafortunada del término library en inglés. Como esta mala traducción está muy extendida en la comunidad científica de habla española, lo utilizaremos en este contenido docente).

2. Amplificación de los fragmentos mediante PCR y deposición individualizada en un soporte sólido (Estos dos pasos pueden cambiar de orden dependiendo de cada tecnología)

3. Reacción de secuenciación masiva mediante ciclos de adición y retirada de reactivos.

De las tres tecnologías iniciales, la desarrollada por la empresa Illumina es la más utilizada. Por tanto, vamos a ilustrar el proceso explicando cómo funciona este sistema. Esto es importante porque una muy alta proporción de los proyectos de secuenciación masiva actuales se realizan con esta tecnología. Sobre la base de Illumina ilustraremos los aspectos más relevantes de las tecnologías de secuenciación masiva de segunda generación.

## 4.1. Generación de una colección de fragmentos a secuenciar

El primer paso, como hemos dicho, es generar una colección de fragmentos que quieren ser secuenciados (la "librería"). Para ello, lo primero es fragmentar el genoma hasta el tamaño de fragmento requerido y posteriormente, "reparar" molecularmente los fragmentos: esto es, hacer que los extremos de cada fragmento queden romos, sin nucleótidos sobresaliendo en uno u otro extremos (algo normal cuando el método de fragmentación es físico), y añadir una Adenina a cada fragmento en una de las cadenas, que facilitará el acoplamiento con adaptadores en los que sobresalga una Timina. Estos dos pasos se consiguen fácilmente mediante enzimas y reacciones in vitro.

El siguiente paso va a ser añadir pequeños fragmentos de DNA diseñados por Illumina (llamados adaptadores) para que se unan a los extremos de cada uno de los fragmentos a secuenciar. El diseño de los mismos permite que se "acoplen" a cada uno de los extremos de cada fragmento en una orientación concreta, de tal modo que, si imaginamos una de las dos cadenas de un fragmento (los fragmentos resultantes de la fragmentación de un genoma serán de doble cadena), a un extremo del mismo quede una zona del adaptador y al otro extremo del fragmento otra de secuencia distinta. Fíjate en los colores verde y naranja de esta figura, y lo verás fácilmente.



**Figura 1.** Características generales de los adaptadores. Se indica la reparación de los fragmentos de ADN, dejando el grupo fosfato (P) accesible, la adición de adeninas (A) en un extremo para facilitar a ligación con timinas (T) sobresalientes en el extremo del adaptador. Los diferentes colores del adaptador indican diferentes zonas para la amplificación (naranja, verde) y secuenciación (rojo) del fragmento (gris).

La preparación de una librería lleva muchísimos pasos, que puedes consultar en cualquier manual de Illumina. No vamos a ver en detalle el proceso de generación de una librería, ya que ese no es el objetivo de este artículo docente. Tras todo el proceso, los fragmentos quedan en cadena simple, algo parecido a lo que vemos en esta imagen. Estas moléculas serán las que inicien el proceso de amplificación por PCR en el soporte sólido.



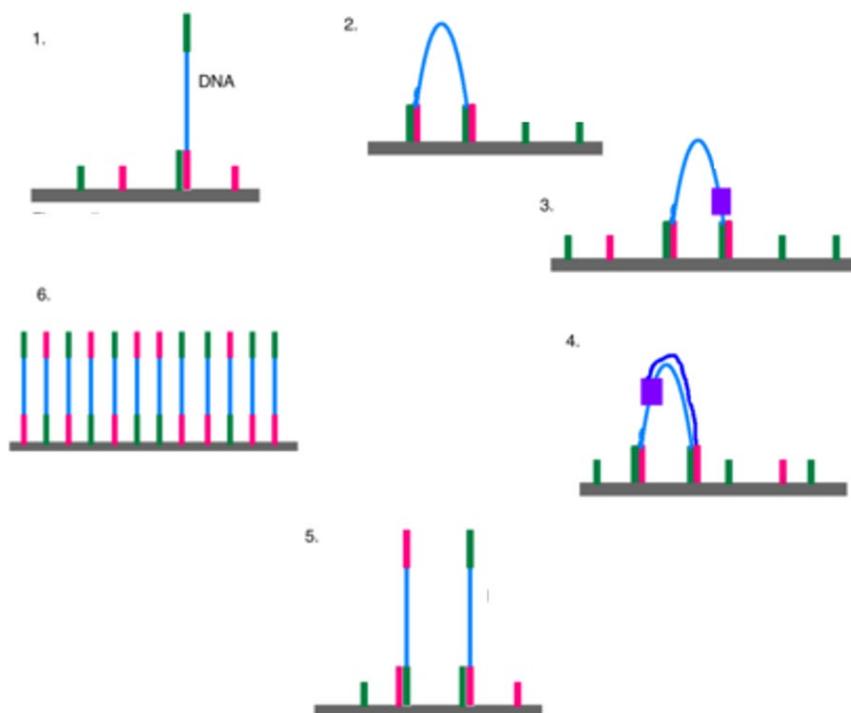
**Figura 2.** Esquema característico de un fragmento de ADN con los adaptadores, antes de su anclaje al soporte sólido para su amplificación por PCR

## 4.2. Amplificación de los fragmentos mediante PCR y deposición individualizada en un soporte sólido.

Todos los fragmentos a secuenciar, en simple cadena y con los adaptadores adecuados a cada extremo, se inyectan en una cámara especialmente diseñada por Illumina para que cada uno de ellos quede "enganchado" al soporte, por medio de oligonucleótidos con secuencia complementaria a la de uno de los extremos del adaptador. Cada fragmento quedará así "enganchado" en una zona diferente del soporte. Millones de fragmentos diferentes pueden distribuirse por un soporte de Illumina.

El siguiente paso es la amplificación de cada uno de esos fragmentos. La amplificación es necesaria porque la detección final se realizará con dispositivos ópticos que requieren un umbral de señal que no proporciona un fragmento individual. La amplificación de cada fragmento, y la secuenciación simultánea de todos los clones de un mismo fragmento asegura una señal suficiente, tras la reacción de secuenciación, para ser detectada por el dispositivo óptico.

El modo de amplificar los fragmentos en el soporte de Illumina es el PCR-puente, que genera los grupos (clusters) de fragmentos idénticos que serán secuenciados. La explicación detallada, y cómo queda preparado el soporte con los fragmentos a secuenciar la podéis en esta figura.



**Figura 3.** Esquema característico del proceso de amplificación de fragmentos y generación de clones de moléculas en el sistema Illumina. 1. DNA proveniente de las librerías anclado al soporte. 2. Puente y anclaje a oligo complementario. 3 y 4. Adición de primer para generación de cadena complementaria. 5.



*Cadenas directa y reversa unidas al soporte. 6. Generación tras varios ciclos de una población clonal de moléculas.*

### 4.3. Proceso de secuenciación masiva.

Una vez preparado el soporte, empieza el proceso de secuenciación. Recuerda que la preparación de la muestra y el proceso de amplificación se ha realizado en una única reacción para todos los fragmentos a secuenciar. La reacción de secuenciación también se va a realizar de modo simultáneo para todos los fragmentos del soporte.

El proceso consiste en la adición cíclica de los reactivos necesarios para iniciar y terminar la reacción enzimática de incorporación de un único nucleótido a cada uno de los fragmentos (recuerda que de cada fragmento inicial hay millones tras la amplificación por PCR, pero nosotros llamaremos "fragmento" a cada cluster, realmente). Esta es la clave del proceso: la reacción de secuenciación puede iniciarse, pero en cada fragmento se paraliza tras la adición a la cadena naciente de un único nucleótido. ¿Por qué?

La razón es que los nucleótidos utilizados no disponen del extremo 3-OH del azúcar para reaccionar con el siguiente nucleótido, de modo que la DNA polimerasa sólo puede añadir un nucleótido a la cadena naciente. El siguiente no puede ya incorporarse, porque el enlace fosfodiéster no puede realizarse sin el grupo 3-OH.

Realmente esto también ocurría en la tecnología de Sanger con los dideoxinucleótidos ¿no es cierto?. Sí, pero en Sanger el nucleótido estaba químicamente construido sin el -OH. En el sistema de Illumina, el grupo OH está bloqueado por un grupo químico, y tras la adición de un reactivo, se produce una reacción que lo restaura. Una vez restaurado el grupo -OH, ya podría incorporarse un segundo nucleótido, si se añadieran al soporte. De nuevo, si estos nuevos nucleótidos están también bloqueados en 3-OH, sólo podrá unirse uno por fragmento.

Empezamos a ver el sentido de los "ciclos" en estos sistemas de secuenciación: Se añaden los nucleótidos bloqueados al soporte, junto con la DNA polimerasa y el cebador necesario para iniciar la síntesis de la cadena complementaria a cada fragmento del soporte. Cada fragmento incorpora un nucleótido, Y SOLO UNO, ya que dichos nucleótidos están bloqueados en 3-OH. Las reacciones se paralizan en todos los fragmentos tras la adición de un nucleótido a cada cadena naciente secuenciada. Ese será el primer nucleótido de nuestra secuencia.

Eliminamos los nucleótidos sobrantes del soporte, añadimos el reactivo que desbloquea los grupos -OH de los nucleótidos (un nucleótido por fragmento, recuerda) recién añadidos en cada fragmento, de tal modo que los fragmentos ya están preparados para incorporar un segundo nucleótido. Añadimos, pues, de nuevo, los nucleótidos.

Pero estos nucleótidos están bloqueados, así que, de nuevo, sólo se incorporará uno de ellos a la cadena naciente de cada fragmento. Ese será el segundo nucleótido secuenciado del fragmento.



De nuevo, eliminamos los nucleótidos sobrantes, añadimos el reactivo que desbloquea, y volvemos a tener accesible el grupo 3-OH del último nucleótido incorporado en cada fragmento. Y vuelta a empezar.

Pero...¿por qué queremos que la secuenciación PARE cada nucleótido incorporado? El objetivo es poder parar la reacción de secuenciación en cada uno de los fragmentos después de la adición de un único nucleótido, PARA PODER SABER qué nucleótido (A, T, C o G) se ha incorporado/secuenciado en cada fragmento.

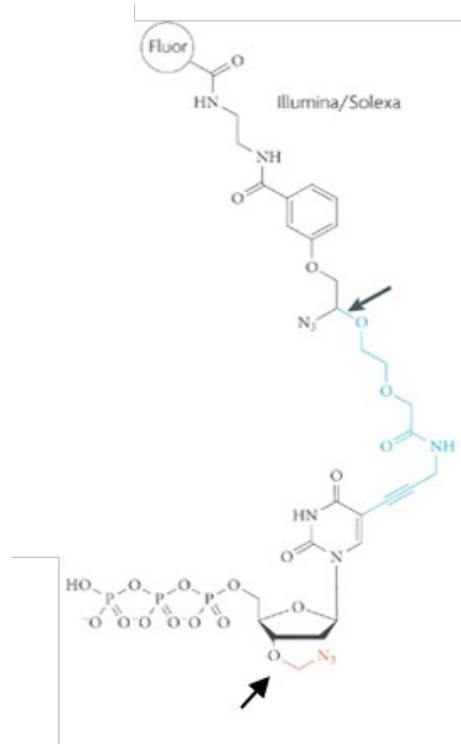
Esto se consigue porque cada nucleótido lleva químicamente anclado un fluoróforo diferente (los verás como rojo, verde, azul y amarillo en el vídeo). De este modo, tras cada ciclo, y con la reacción parada, se CAPTURA una imagen digital que muestra un color en cada una de las posiciones del soporte (en cada fragmento). Esa captura contiene la información de secuencia de la PRIMERA BASE DE CADA FRAGMENTO. La primera base es la que está más cerca del cebador, Tras el segundo ciclo, se realiza una segunda captura de imagen. Esta segunda captura de imagen contiene la información de secuencia de la SEGUNDA BASE incorporado DE CADA FRAGMENTO.

Pero...si el primer nucleótido incorporado en un determinado fragmento es, digamos A (azul) y el segundo es, digamos, T (verde), al capturar la imagen del segundo ciclo tendremos una mezcla de señal azul y verde....! Por eso, tras cada ciclo, y tras la captura de imagen, se añade otro reactivo químico que rompe el enlace que une el fluoróforo del nucleótido sin perjudicar sus capacidades para formar un nuevo enlace fosfodiéster. Así, el fluoróforo del nucleótido incorporado en cada ciclo se elimina tras la captura de imagen, y tras el segundo ciclo se captura información de señal SOLO del fluoróforo del segundo nucleótido.

El proceso sería el siguiente: añadimos nucleótidos, se incorpora uno por fragmento, se eliminan (se lavan) los nucleótidos restantes que queden por el soporte, se captura la señal óptica, se desbloquea el grupo 3-OH del nucleótido recién incorporado, se elimina el fluoróforo del nucleótido recién incorporado, y se inicia el ciclo de nuevo, añadiendo de nuevo los nucleótidos.

Si se realizan 50 ciclos, se habrán incorporado 50 nucleótidos (uno a uno) a cada uno de los fragmentos del soporte, es decir, habremos secuenciado 50 nucleótidos de dicho fragmento. Tendremos 50 imágenes, una por ciclo, con información de cada uno de los nucleótidos incorporado en cada fragmento en cada ciclo. Tras la finalización de los ciclos, la información se procesará para dar la secuencia de cada uno de los fragmentos.

Estos nucleótidos especiales que permiten todo este proceso se denominan Nucleótidos Terminadores Reversibles. Aquí tenéis su estructura química, simplemente para que veáis los puntos de bloqueo y adición del fluoróforo. Las flechas indican los enlaces químicos de corte de los grupos de bloqueo del -OH y del fluoróforo.



**Figura 4.** Esquema representativo de un nucleótido terminador reversible utilizado en secuenciación masiva por la tecnología Illumina. En rojo se observa el grupo que bloquea el enlace 3-OH necesario para la adición del siguiente nucleótido. En azul, el grupo espaciador que enlaza con el fluoróforo.

## 5 Cierre

A lo largo de este objeto de aprendizaje hemos visto las tecnologías de secuenciación masiva permiten paralelizar la secuenciación de millones de fragmentos, permitiendo así la secuenciación de genomas en poco tiempo.

Utilizando como ejemplo la tecnología de Illumina hemos visto como el proceso se inicia con la construcción de una librería de fragmentos, que son inmovilizados en un soporte sólido para su amplificación por PCR, previamente a la secuenciación propiamente dicha, posible gracias al uso de nucleótidos terminadores reversibles añadidos de manera cíclica.

Para ampliar este conocimiento, investiga el modo de secuenciación de otra tecnología de segunda generación (Ion Torrent), identifica los tres pasos básicos que has aprendido aquí, señalando semejanzas y diferencias entre ambas tecnologías.



## 6 Bibliografía

### 6.1 Libros:

Primrose, S., & Twyman, R. (2003). Principles of Genome Analysis and Genomics. Tema 2.

### 6.2 Artículos de revisión

Liu L, Li Y, Li S, Hu N, He Y, Pong R, Lin D, Lu L, Law M. Comparison of next-generation sequencing systems. J Biomed Biotechnol. 2012;2012:251364. doi:10.1155/2012/251364.

Metzker ML. Sequencing technologies - the next generation. Nat Rev Genet. 2010,Jan;11(1):31-46. doi: 10.1038/nrg2626.

Mardis ER. Next-generation DNA sequencing methods. Annu Rev Genomics Hum Genet. 2008;9:387-402. doi: 10.1146/annurev.genom.9.081307.164359.