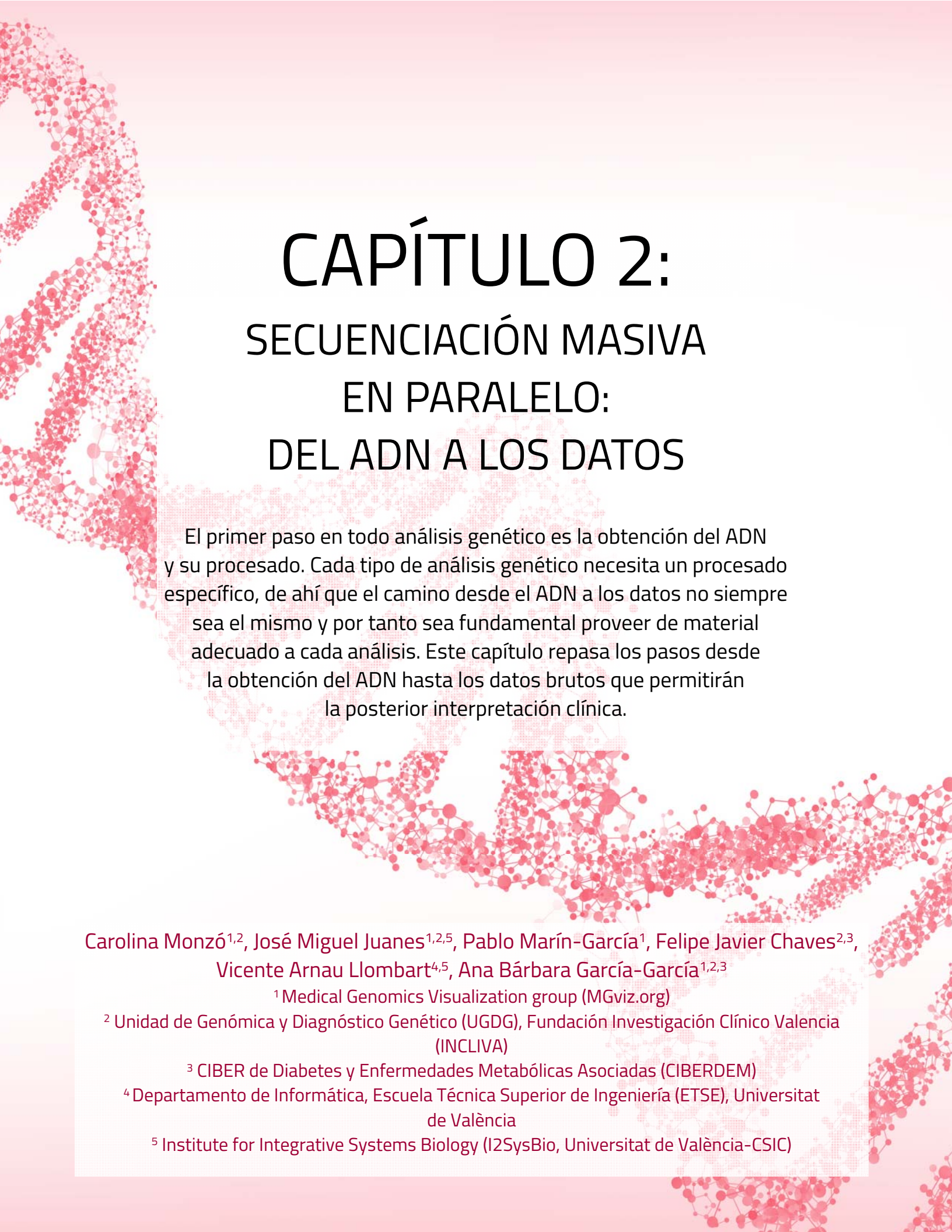




# Genómica en Medicina

## Una Guía Práctica





# CAPÍTULO 2:

## SECUENCIACIÓN MASIVA EN PARALELO: DEL ADN A LOS DATOS

El primer paso en todo análisis genético es la obtención del ADN y su procesado. Cada tipo de análisis genético necesita un procesado específico, de ahí que el camino desde el ADN a los datos no siempre sea el mismo y por tanto sea fundamental proveer de material adecuado a cada análisis. Este capítulo repasa los pasos desde la obtención del ADN hasta los datos brutos que permitirán la posterior interpretación clínica.

Carolina Monzó<sup>1,2</sup>, José Miguel Juanes<sup>1,2,5</sup>, Pablo Marín-García<sup>1</sup>, Felipe Javier Chaves<sup>2,3</sup>,  
Vicente Arnau Llombart<sup>4,5</sup>, Ana Bárbara García-García<sup>1,2,3</sup>

<sup>1</sup> Medical Genomics Visualization group (MGviz.org)

<sup>2</sup> Unidad de Genómica y Diagnóstico Genético (UGDG), Fundación Investigación Clínico Valencia (INCLIVA)

<sup>3</sup> CIBER de Diabetes y Enfermedades Metabólicas Asociadas (CIBERDEM)

<sup>4</sup> Departamento de Informática, Escuela Técnica Superior de Ingeniería (ETSE), Universitat de València

<sup>5</sup> Institute for Integrative Systems Biology (I2SysBio, Universitat de València-CSIC)

## INTRODUCCIÓN

Han pasado más de 10 años desde la finalización del Proyecto Genoma Humano (International Human Genome Sequencing Consortium, 2004) y gracias al rápido progreso de la secuenciación masiva en paralelo y su abaratamiento, hoy en día es fácil ver el uso rutinario del diagnóstico genético de muchas enfermedades en el ámbito hospitalario.

Aunque el uso de la genómica médica es muy prometedor, todavía está en sus inicios. Cada año se desarrollan nuevas tecnologías, métodos y secuenciadores que hacen del diagnóstico genético una realidad (Goodwin, 2016). La denominación más precisa para la secuenciación de nueva generación sería Secuenciación Masiva en Paralelo (MPS, de su nombre en inglés Massive Parallel sequencing). No obstante, la expresión más utilizada, que adoptamos también en esta guía, es la de Next Generation Sequencing (o NGS). La secuenciación de nueva generación ha supuesto un salto cualitativo y cuantitativo que ha cambiado las reglas del juego en el panorama del diagnóstico genético, permitiendo pasar de la genética clínica a la genómica médica y en un futuro cercano, a la medicina de precisión (Green, 2011).

Durante los últimos años, se han realizado grandes proyectos públicos internacionales de secuenciación masiva, como el proyecto 1000 genomas (Birney y Soranzo, 2015), ENCODE (ENCODE Project Consortium, 2012), 100K genomas del Genome England (Genomics England, 2017), ExAC y gnomAD (Monkol, 2016), Genome Asia 100K (<http://www.genomeasia100k.com/>), AllOfUs/Precision Medicine Initiative (NIH) (Sankar, 2016) y algunos privados como AstraZeneca (Ledford, 2016). A partir de este tipo de proyectos y otros a menor escala, se han abaratado los costes de procesado de ADN, estandarizado los protocolos de secuenciación y análisis y se han desarrollado máquinas de secuenciación lo suficientemente simples, rápidas y baratas

que, junto a los avances en las herramientas de análisis y potencia de cálculo de los ordenadores actuales, han permitido la incorporación de la secuenciación del ADN en los sistemas de Salud y la obtención de diagnósticos basados en la genómica de forma rutinaria.

Estos avances están permitiendo definir los componentes genéticos de muchas enfermedades, caracterizar comprensivamente genomas de cáncer, crear sistemas prácticos para la informática de la genómica clínica e incluir el microbioma en los modelos de salud y enfermedad humana.

En este capítulo introduciremos las diferentes técnicas de secuenciación y análisis del ADN, explorando todos los pasos del camino que se recorre desde el ADN hasta los datos brutos que permitirán la interpretación clínica de los resultados de la secuenciación masiva.

## TENEMOS UN PACIENTE, ¿Y AHORA QUÉ?

### *Preparación de las muestras*

Para proceder a la correcta extracción del material genético a analizar hay que tener claro el estudio que se quiere realizar. En la mayoría de los casos se analiza el ADN genómico para la identificación de las variantes causantes de una enfermedad. Por ejemplo, en enfermedades hereditarias se suele obtener ADN a partir de sangre periférica, pero si se quiere ver la presencia de mutaciones en un tumor, hay que hacer una biopsia y analizar el ADN del mismo. También se puede analizar el ADN circulante para diagnóstico prenatal no invasivo (Figura 1) e incluso el de la carga vírica.

La extracción del ADN es un proceso sencillo que se puede hacer incluso en casa y consiste básicamente en extraer los leucocitos de la sangre (o células de tejidos como la mucosa bucal, biopsias o heces), romper su pared celular y la del núcleo, y precipitar el ADN mediante sales y alcohol. Cuestiones importantes en este proceso son: inhibir las proteínas que

degradan el ADN cuando se lisan las células, obtener solo ADN o ARN (o los dos juntos según se necesite), o únicamente el ADN mitocondrial, y que el ADN obtenido sea lo más puro posible y no contenga trazas de los productos usados para la lisis y extracción del ADN (detergentes, fenol, cloroformo, etc.), que interferirían en la PCR y procesos de secuenciación (Psifidi A, 2015). Según el tipo de análisis posterior es muy importante que el ADN esté lo menos fraccionado posible. Algunos robots de extracción automática de ADN no servirían para estos análisis, así como tampoco el ADN de muestras sometidas a procesos de conservación de tejidos fijados con formaldehído (bloques de parafina), que también fraccionan el ADN en tamaños de entre 125 y 200 pares de bases.

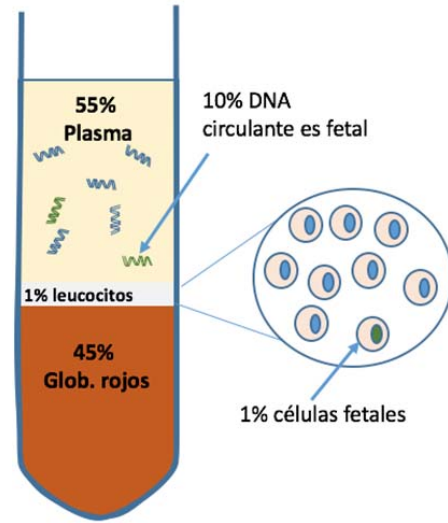


Figura 1: El test prenatal no invasivo o NIPT (non invasive prenatal test) se hace a partir de ADN libre circulante (cfDNA) en el plasma materno (Vermeesch, 2016).

**PRINCIPALES OPCIONES DE TÉCNICAS DE SECUENCIACIÓN.**

SECUENCIACIÓN CLÁSICA		
<b>Sanger</b>		Secuencias de 500 a 100 pares de bases (una secuencia de un solo individuo a la vez). Paralelización 96x.
MPS (MILLONES DE SECUENCIAS EN PARALELO)		
<b>Lecturas cortas</b>	<b>Illumina</b>	Secuencias de entre 75 y 150 pb. Error aleatorio, pero según secuenciador (los que usan dos colores) mayor sesgo en llamado de Gs y problemas con regiones ricas en AT y GC.
	<b>Ion torrent</b>	Secuencias entre 200 y 400 pb. Mayor tasa de error en homopolímeros.
<b>Lecturas largas</b>	<b>PacBio</b>	Secuencias de entre 8 y 30 Kb. Tasas de error bajas a 1% con muchos pases de la misma secuencia (circular consensus read).
	<b>Oxford nanopore</b>	Secuencias de 200 Kb (en continua mejora). Tasa de error entre un 2 y un 10% según metodología.
	<b>Lecturas largas sintéticas</b>	Secuencias 'ligadas' de longitud virtual de 100 Kb. Uso de código de barras únicos para fragmentos largos antes de fraccionamiento para preparar la genoteca y 'unir sintéticamente' las lecturas. 10X Genomics.

## YA TENEMOS EL ADN ¿Y AHORA QUÉ?

Una vez extraído el ADN, el siguiente paso es ver qué método de secuenciación es el más adecuado para el experimento planteado.

A la hora de secuenciar el ADN la primera decisión es si necesitamos secuenciar una secuencia larga y continua (para el tipado de HLA, discriminar la secuencia de un gen de sus pseudogenes, o estudiar grandes reordenamientos del genoma) o nos sirven fragmentos pequeños. Otro factor a tener en cuenta es si queremos secuenciar muchos individuos o muchos genes a la vez, en cuyo caso la secuenciación Sanger sería sustituida por la secuenciación masiva en paralelo (MPS) por su rapidez y precio.

## SECUENCIACIÓN MASIVA

La secuenciación masiva en paralelo consiste en fragmentar el ADN en pequeños trozos, amplificarlos mediante PCR y procesarlos todos a la vez (Figura 2). Esto permite analizar cualquier ADN aunque no se tenga conocimiento de la secuencia a priori, ya que no se van a diseñar cebadores para amplificar regiones específicas del ADN, como se hace clásicamente con la tecnología Sanger o de electroforesis capilar.

El truco de la secuenciación masiva está en la paralelización, es decir, en secuenciar todos los fragmentos de ADN a la vez. Los 3 mil millones de bases de un genoma se pueden secuenciar en dos días, y teniendo en cuenta que se pueden poner varias muestras a la vez cada una con una 'etiqueta', los secuenciadores actuales como el NovaSeq de Illumina admiten hasta 16 genomas a la vez por carrera y los secuencian todos en 40 horas. A esto hay que añadir otras 40 horas para procesar los datos bioinformáticamente en un centro de alta computación que permita la paralelización masiva de los procesos de análisis. Este análisis bioinformático se puede hacer en ordenadores de sobremesa con 32 GB de RAM y se puede analizar un exoma en 3 horas.

Cómo conseguir la paralelización y cómo se leen las bases es lo que diferencia a las diferentes plataformas

A día de hoy hay dos modelos principales de técnicas de secuenciación MPS: secuenciación por síntesis (SBS) que implica fragmentos cortos y Single-Molecule Real-Time Sequencing (SMRT) que permite fragmentos de varias kilobases pero con mayor tasa de error.

En el lado de los secuenciadores SBS, las dos tecnologías más asentadas son Illumina (detecta la adición de las bases una a una mediante fluoróforos) e Ion Torrent, que detecta grupos de bases iguales y

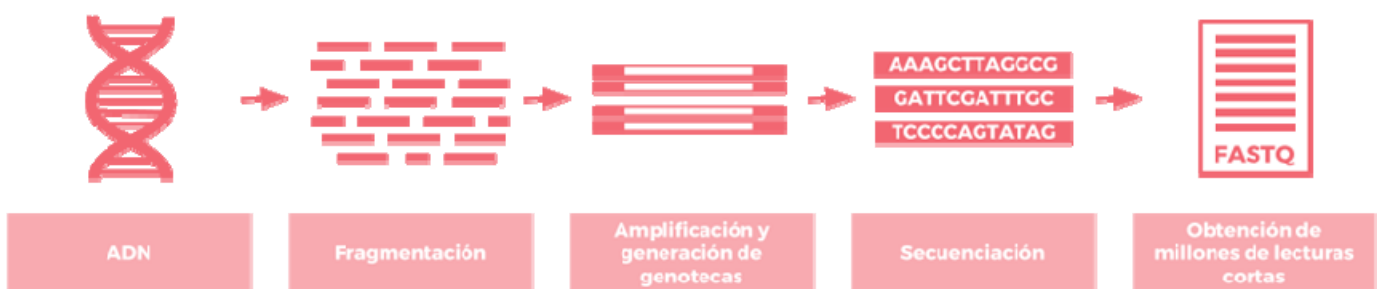


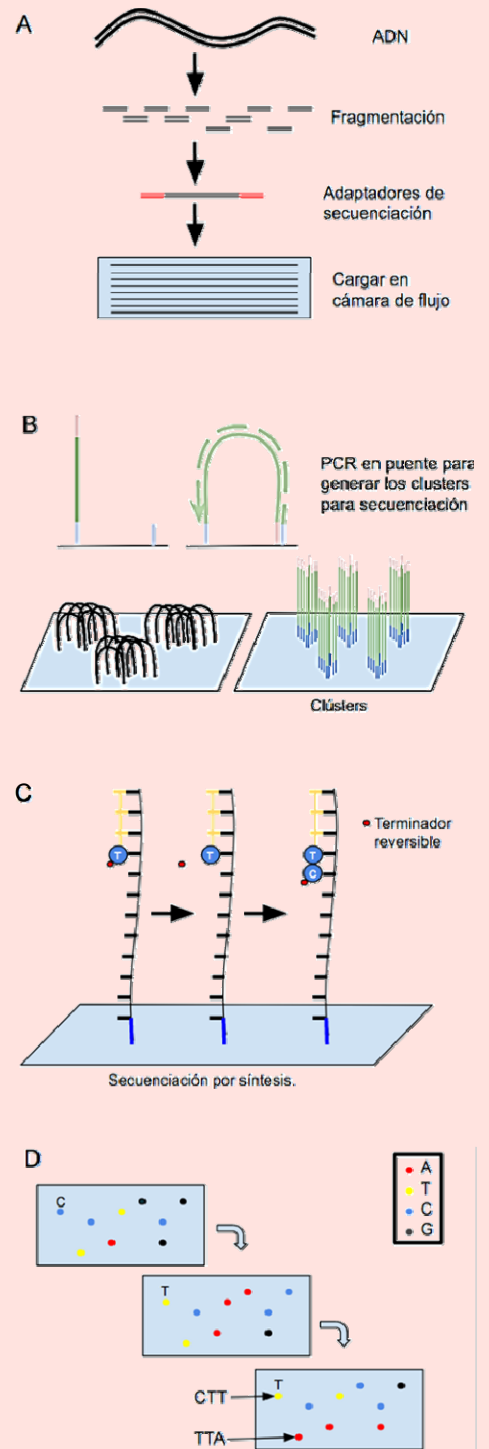
Figura 2. Sucesión simplificada de pasos a llevar a cabo en estudios de secuenciación masiva.

que mide la señal mediante semiconductores que evalúan cambios de concentración de protones (pH).

En cuanto a los secuenciadores de secuencias largas y detección en tiempo real de moléculas únicas, los dos exponentes principales son PacBio de Pacific Biosciences y Minlon de Oxford Nanopore. PacBio lee secuencias largas en tiempo real midiendo la emisión de luz del fluoróforo liberado tras la incorporación de cada nucleótido. Minlon detecta las bases de la secuencia midiendo cambios de corriente eléctrica en la membrana del poro a medida que pasa la secuencia de cadena simple del DNA.

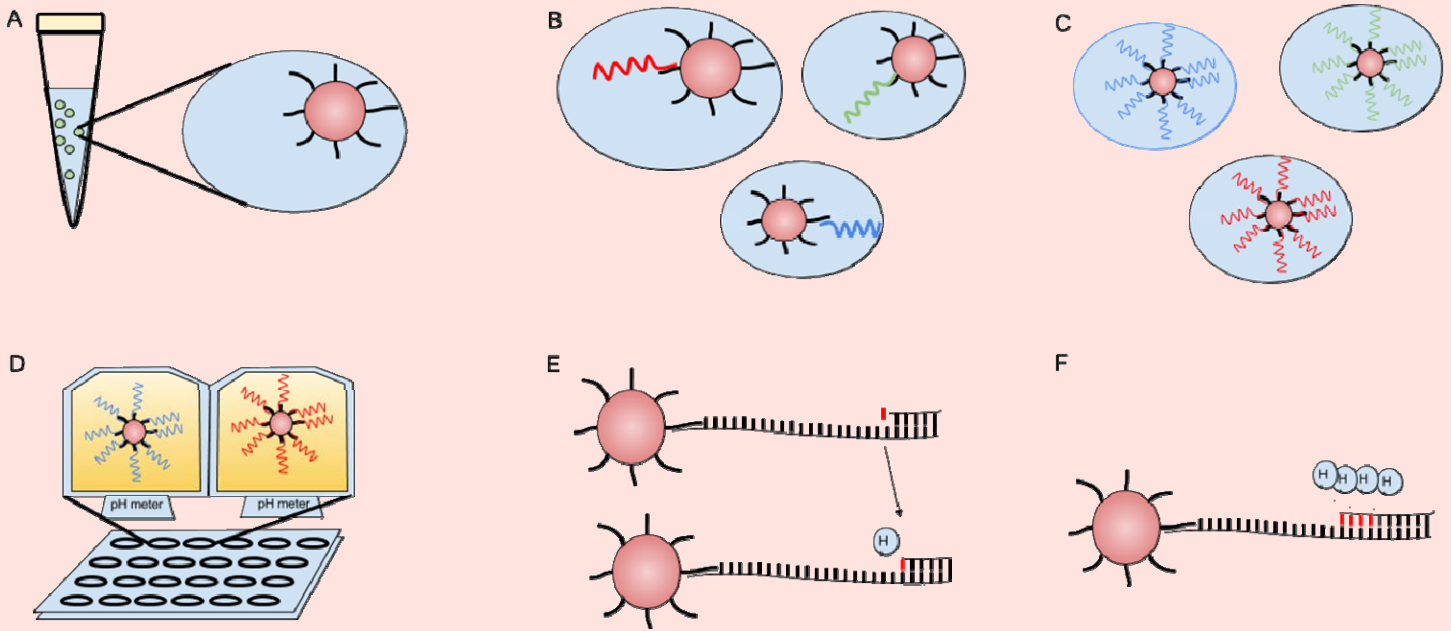
Veamos más detalladamente en qué consiste cada técnica:

- **Ion Torrent.** La técnica de secuenciación de Ion Torrent inicia su procesamiento con una PCR en emulsión con microgotas (Nyrén, 1985) y usa semiconductores para detectar los H<sup>+</sup> desprendidos en la incorporación de los dNTPs. Figura 3
- **Illumina.** Una de las mayores aportaciones de la tecnología de secuenciación de Solexa-Illumina, es la PCR puente para la generación de clústeres, y el método de la terminación cíclica reversible para la secuenciación por síntesis. En cada ciclo, se une un dNTP marcado, se toma una fotografía y se retira para empezar de nuevo (Bentley, 2008). Figura 4
- Tanto Ion Torrent como Illumina, generan secuencias cortas. Illumina de 75 a 300 pares de bases; Ion Torrent hasta 400. Si se necesitan lecturas más largas, los secuenciadores de **Pacific Biosciences** (Rhoads, 2015) y **Oxford Nanopore Technologies** (Haque, 2013) son capaces de producir lecturas que superan los 1.000-10.000 pares de bases. Estas tecnologías, facilitan la secuenciación de regiones del ADN que contienen alta cantidad de nucleótidos GC, y el alineamiento de secuencias que contienen repeticiones. Además, al no requerir un paso previo de amplificación por PCR, evitan los errores de la enzima polimerasa. Figura 5.



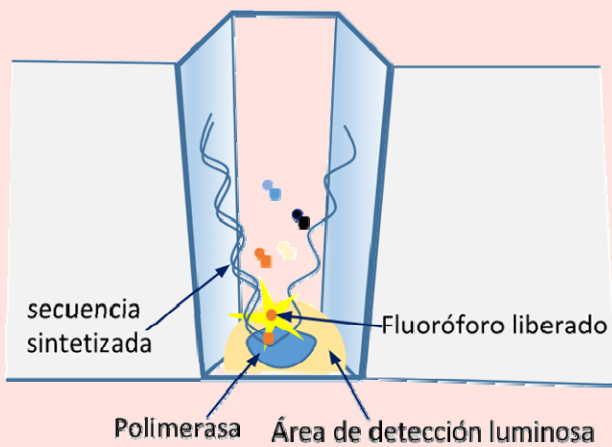
### TECNOLOGÍA ILLUMINA

- El ADN se fragmenta y se une a dos adaptadores, uno en cada extremo, y se fijan sobre una lámina de vidrio sobre la que hay oligonucleótidos complementarios a los adaptadores.
- PCR en puente para amplificación clonal. Cuando se han formado los clústeres de secuencias iguales.
- Secuenciación. Esta técnica utiliza dNTPs con distintos fluoróforos y terminadores reversibles.
- Cada ciclo, una nueva base es añadida al clúster y sacado una foto. Por ejemplo en el clúster superior izquierdo primero se ha añadido una C luego una T y finalmente una T, quedando la secuencia CTT.



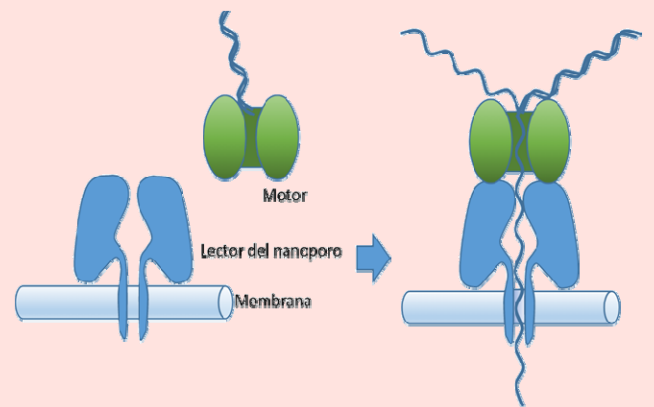
**TECNOLOGÍA ION TORRENT**

- A. Microgotas con perlas y reactivos de PCR.
- B. Unión de un único fragmento por perla.
- C. Amplificación del fragmento en cada microgota de PCR individual.
- D. distribución de una perla por pocillo.
- E. Secuenciación por síntesis. Cada incorporación de un nucleótido libera un protón que mide un pHmetro.
- F. Si hay varias bases idénticas consecutivas da una señal proporcional al número de bases consecutivas, pero debido a problemas de saturación de señal esto genera problemas con homopolímeros de más de 7 bases dando tasas de error alto en estos casos.



**TECNOLOGÍA PacBio**

Detecta mediante un microscopio confocal la luz liberada tras cada incorporación de cada nucleótido a la cadena de ADN que va sintetizando



**TECNOLOGÍA MinIon**

Detecta cambios en la corriente que pasa por el poro a medida que la secuencia va avanzando.

### ¿Cuál ha sido la revolución de la secuenciación masiva?

La importancia de la paralelización está en todos los pasos del proceso. Los métodos actuales de secuenciación han triunfado porque son capaces de paralelizar y automatizar el análisis a gran escala, no sólo la adquisición de la secuencia. Esto se ha conseguido gracias a un punto clave que ha facilitado todos los pasos posteriores. Cada secuenciación en paralelo parte de una sola hebra de ADN, no de cada una de las copias de cada cromosoma. Por tanto, a diferencia de la secuenciación Sanger, donde una variación en heterocigosis implica la detección en cada lectura de las dos variantes a la vez, aquí se tiene una secuencia distinta para cada alelo, o copia de ADN. Esto implica lo que se conoce como una lectura "digital", o lo que es lo mismo, es posible contar cuántas lecturas hay de un alelo y cuantas del alternativo e inferir si esas proporciones corresponden a un heterocigoto (50% de lecturas aproximadamente de cada alelo), mutaciones somáticas (en cáncer) o a un artefacto de secuenciación.

### La MPS es más rápida, pero la secuenciación Sanger es mejor ¿No?

No necesariamente. Hoy en día, la secuenciación masiva de lecturas cortas como la de Illumina, tiene la misma o más calidad que la secuenciación mediante el método Sanger (patrón estándar actual) si se realiza a una cobertura de 200x. Además, tiene la ventaja de que se pueden automatizar todos los pasos del análisis, sobre todo el de llamado de variantes, lo que permite poder analizar un exoma completo en el mismo tiempo que se haría un panel de 10 o 15 genes siguiendo el método estándar. Hay que resaltar que es verdad que la secuenciación masiva tiene un una mayor tasa de error por lectura, pero eso no es ningún problema ya que su potencia viene de leer la misma secuencia entre 30 y 200 veces y sacar un consenso.

La secuenciación Sanger sigue siendo útil en regiones del genoma donde hagan falta secuencias largas para poder alinear/amplificar de forma unívoca una

“LA SECUENCIACIÓN SANGER SE USA PARA VALIDAR LOS RESULTADOS DE LA SECUENCIACIÓN MASIVA PORQUE ES UNA TÉCNICA 'ORTOGONAL', ES DECIR, UNA TECNOLOGÍA CON UN TIPO DE ERROR TOTALMENTE DIFERENTE.”



región y asegurarse de que las variantes que se van a llamar, son realmente las de esa zona y no las de un pseudogen o región homóloga.

Si la secuenciación Sanger no es mejor ¿Por qué se usa para validar los datos de secuenciación masiva?

La secuenciación Sanger se usa para validar los resultados de la secuenciación masiva porque es una técnica 'ortogonal', es decir, una tecnología con un tipo de error totalmente diferente al de la metodología original, de forma que si el resultado de la primera era un artefacto de la técnica, podría ser un error intrínseco del método y volvería a salir en una repetición. Usando una tecnología completamente diferente, con fuentes de error distintas, nos reafirma en que un resultado coincidente sea real y no un artefacto.

## PREPARACIÓN DE GENOTECAS PARA SECUENCIACIÓN MASIVA (MPS)

En este apartado trataremos solo la preparación de genotecas de Illumina, que es el método más extendido en el ámbito clínico. Los conceptos son básicamente los mismos para otras plataformas. El primer paso para poder secuenciar con las tecnologías de MPS es procesar el ADN creando una genoteca (fragmentos pequeños de ADN de unos 200 a 400 pares de bases con unas secuencias en los extremos llamadas adaptadores, que nos permitirán realizar la secuenciación). Para ello, se usan diferentes técnicas de fragmentación, ya sea mecánica (sonicación), química (temperatura y pH) o enzimática (fragmentasas o transposasas).

El sistema de fragmentación a elegir es muy importante porque es uno de los pasos donde hay que poner el primer control de calidad, ya que cualquier fallo puede incurrir en un gran sesgo en la calidad de los resultados. Técnicamente, la mejor fragmentación y la mayor calidad se consiguen con un método especial de sonicación llamado de focalización adaptativa (Covaris es la principal casa comercial que lo

distribuye). Este método consigue un fraccionamiento del DNA muy homogéneo y normalmente se obtiene un rango de fragmentos con un pico a 250 nucleótidos, a los que luego se les añade por ligación un adaptador a cada lado. Este método, aunque muy preciso, es laborioso y hoy en día hay protocolos más cortos con métodos enzimáticos.

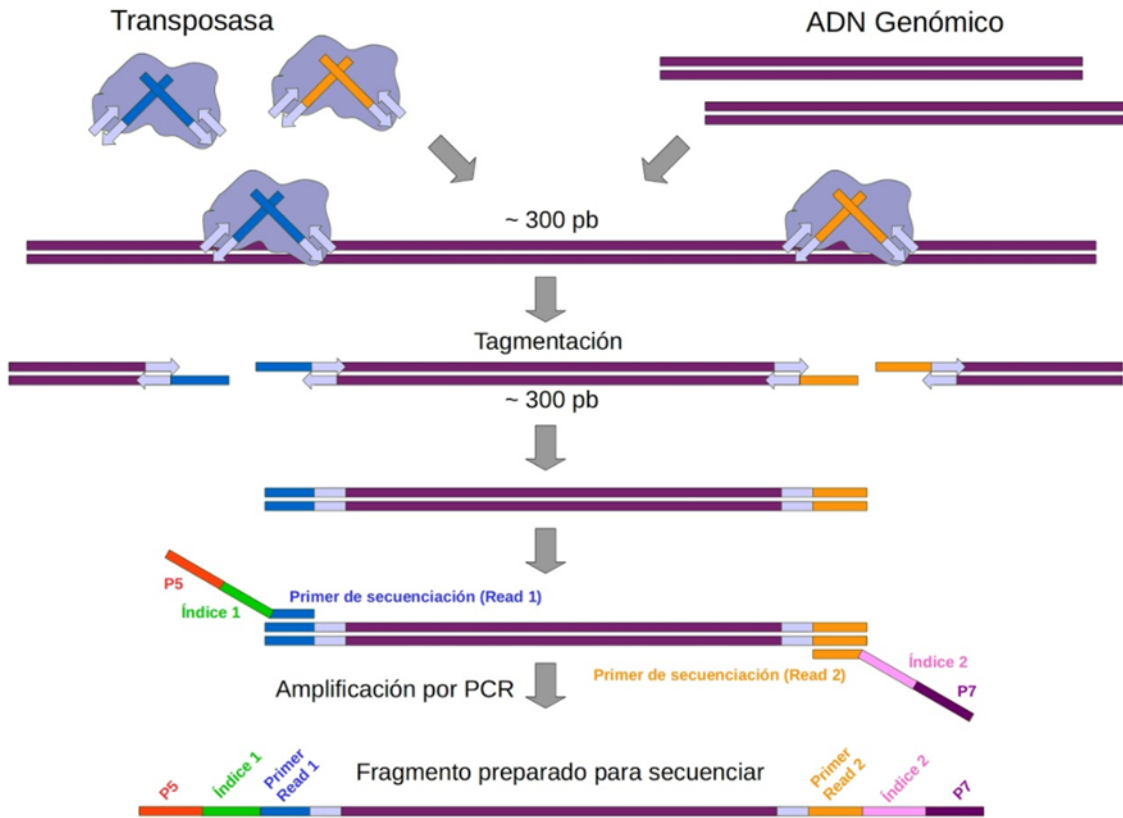
Los métodos enzimáticos se basan en enzimas derivadas de elementos transponibles llamadas transposasas. Este método añade ya una parte del adaptador al hacer el corte y luego se añade el resto del adaptador por PCR (Figura 6). La fragmentación basada en transposasas (tagmentación), aunque es mucho más rápida, no tiene tanta calidad como la sonicación de foco adaptativo. Nextera (Illumina) y QXT (Agilent) son dos productos comerciales basados en esta tecnología.

El tipo de muestra de partida es determinante en la elección del método de fragmentación. A nivel práctico, cuando hay que procesar gran cantidad de muestras de sangre o material no embebido en parafina, la tagmentación es una buena opción. No obstante, para muestras de parafina, es altamente recomendable usar sonicación ya que estos ADN ya están fragmentados o deteriorados por el proceso de fijación y la mayoría de los fragmentos de ADN tienen tamaños por debajo de los 250 pares de bases, lo que dificulta el corte por las transposasas.

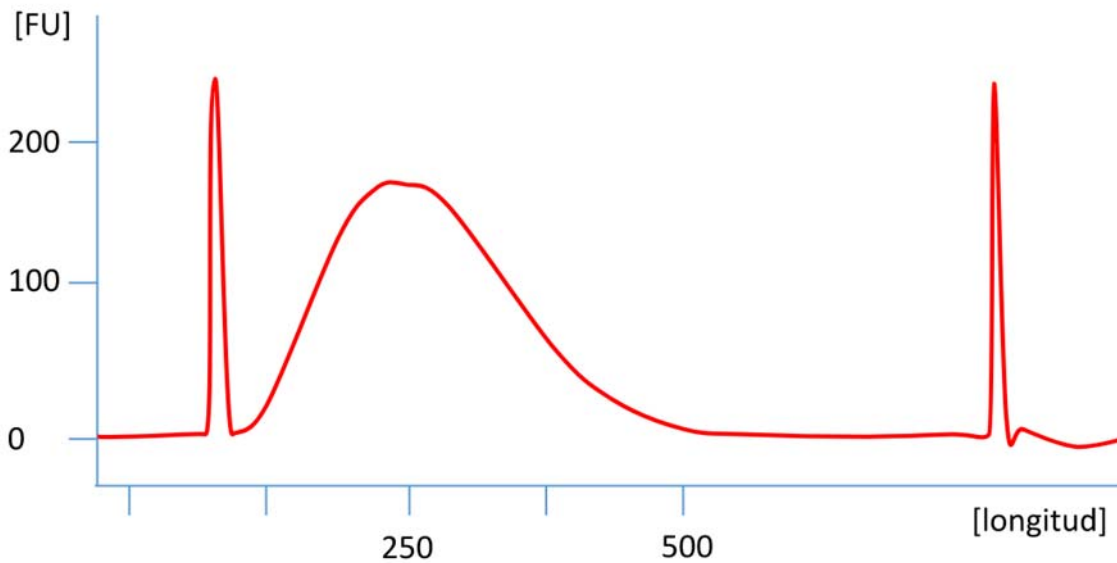
La fragmentación requiere una comprobación de calidad para ver que ha sido correcta. Este chequeo consiste en ver el perfil de fragmentación en un Bioanalyzer, TapeStation o QIAxcel.

Otra forma de obtener la genoteca es por amplificación de fragmentos en vez de por fragmentación de todo el genoma (Ion Torrent). Cuando el número de genes a estudiar es bajo, es más rentable diseñar PCRs para todos los exones de los genes y secuenciar los fragmentos, que hacer un exoma o diseñar sondas de captura y realizar todo el proceso de enriquecimiento.

Hay productos comerciales que amplifican por PCR



Tagmentación: Fragmentación del ADN con transposasas modificadas. Un método rápido y eficaz de preparar librerías para MPS de fragmentos cortos es la fragmentación del DNA por métodos enzimáticos.



Perfil de longitudes de fragmentación del ADN antes de añadir los adaptadores, visto con Bioanalyzer, TapeStation o Qiaxcel

todo el exoma o gran cantidad de genes. Esto tiene sus ventajas e inconvenientes. La secuenciación a partir de fragmentos de PCR es más barata porque es más específica y se secuencian sólo los fragmentos deseados, pero es más laborioso diseñar todos los cebadores. Por el contrario, los métodos de captura o enriquecimiento son más rápidos pero acaban secuenciando un 40% de secuencias fuera de la zona de interés, encareciendo la secuenciación.

En definitiva, las tres aproximaciones más comunes de selección de regiones para secuenciar a la hora de preparar la librería son:

- Estudio utilizando paneles de genes implicados en la enfermedad o relacionados con el fenotipo del paciente, que supone secuenciar una cantidad de genes determinados. Este tipo de estudio reduce la cantidad de genes a evaluar, facilitando el diagnóstico rápido y dirigido de la patología cuando hay una sospecha significativa de un síndrome concreto. Normalmente se hacen por PCR la selección de región para menos de 10 genes o diseñando sondas de captura si se va a amplificar más de 1 Mb.
- Estudio del exoma, que supone la secuenciación mayoritariamente de la parte codificante del ADN (aproximadamente un 1-2% de la secuencia genómica). Esto significa secuenciar entre 30 y 60 millones de bases 100 veces de media por individuo (cobertura 100x). Al cubrir la gran mayoría de las variantes interpretables, la secuenciación del exoma es útil cuando se desconoce la posible causa de la enfermedad del paciente en estudio o se han descartado los genes clásicos asociados con la enfermedad. La calidad de la secuenciación para cada gen está relacionada con el sistema de captura y sus coberturas finales dependerán de la eficiencia de las sondas diseñadas para ese gen.
- Estudio de genoma completo, que supone la secuenciación de todo el ADN del individuo. Esto significa secuenciar 3 mil millones de pares de

bases al menos 30 veces por individuo (normalmente indicado como una cobertura media mínima de 30x). Al secuenciar todo el genoma directamente, se obtiene una cobertura similar para todas las bases. Además no sólo da la información sobre las variantes de una sola base, sino que también permite identificar reordenamientos y variaciones en número de copias (CNVs).

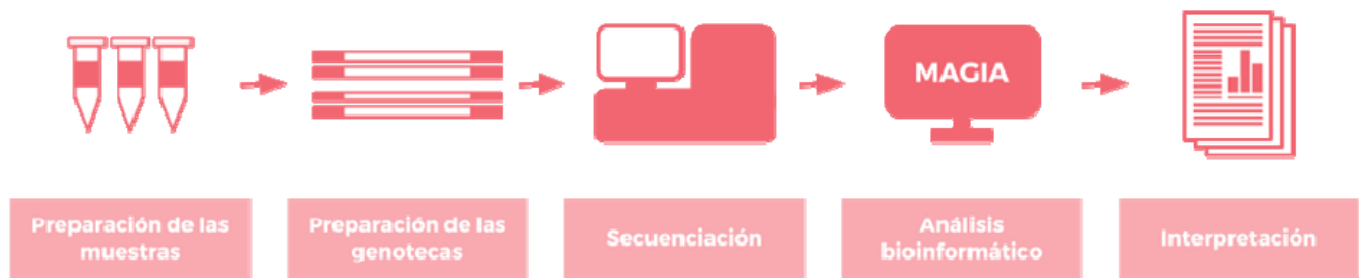
Cada uno de estos estudios tiene sus pros y sus contras. Sin embargo, debido a que todavía se desconoce la función de gran parte de las regiones del ADN y dado el alto coste de un genoma, para los estudios de diagnóstico clínico se utiliza mayoritariamente el análisis de exoma, o en caso de tener sospechas significativas que dirijan el diagnóstico hacia síndromes concretos de origen monogénico conocido, paneles de genes dirigidos.

## DISCUSIÓN

La democratización de la MPS en la clínica, ha dado lugar a un sinnúmero de posibilidades que permiten desarrollar nuevos métodos y procedimientos para facilitar y acelerar el diagnóstico genético y la ampliación de sus aplicaciones. La simplificación del procesado de muestras y la automatización del análisis bioinformático con las nuevas técnicas de MPS hace que este tipo de análisis esté llegando a los hospitales y sea una técnica más al alcance de los médicos para su uso en el diagnóstico diario (Figura 7).

Hay que mentalizarse de que la MPS está aquí para quedarse. Son técnicas robustas de alta fiabilidad, reproducibilidad, automatización y capacidad de detección. En concreto, la secuenciación de exoma completo identifica las bases genéticas de las enfermedades en un 25-40% de los casos (dependiendo de si se hace un exoma completo de inicio o cuando no se ha encontrado nada en los genes candidatos).

Aunque la MPS es muy prometedora y útil aún tiene



Dada la importancia de la MPS en el área clínica, el procesado de los datos se ofrece ya muchas veces como un software cerrado al que se proporcionan unos datos de entrada y una configuración y se obtienen "mágicamente" unos resultados. Pero cuidado, esos resultados hay que interpretarlos teniendo en cuenta las medidas de calidad de cada parámetro en la interpretación y visualizar los alineamientos para descartar artefactos.

sus limitaciones. Por ejemplo, no se puede secuenciar regiones largas de homopolímeros (una misma base repetida muchas veces) si se utiliza una tecnología como la de IonTorrent. La saturación de H+ que producen, impide el conteo correcto. Con Illumina pasa algo parecido pero es menos dramática la pérdida de calidad. Tampoco se puede analizar genes que comparten zonas homólogas con otras regiones del genoma (pseudogenes, exones muy homólogos) si se utilizan técnicas de secuenciación que generan fragmentos cortos, ya que durante el alineamiento de estas secuencias se pueden producir mapeados ambiguos y no se podrá asignar a una posición concreta. Esto se puede evitar usando tecnologías de lecturas largas como PacBio o Nanopore.

Con todo, la implementación de la MPS en la rutina del sistema público de salud será una realidad en los próximos años y facilitará la personalización de la medicina y la adaptación del tratamiento del paciente, tanto para el manejo adecuado de sus signos y síntomas, como para la anticipación al progreso de la enfermedad.

## BIBLIOGRAFÍA

- Bentley DR, et al. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature*. 2008; 456:53–59. doi: 10.1038/nature07517
- Birney E y Soranzo N. Human genomics: The end of the start for population sequencing. *Nature*. 2015; 526:52–53. doi: 10.1038/526052a
- ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 2012; 489(7414):57-74. doi: 10.1038/nature11247
- Genomics England. The 100,000 Genomes Project Protocol. 2017; doi: 10.6084/mg.figshare.4530893.v2. 2017.
- Green ED, et al. Charting a course for genomic medicine from base pairs to bedside. *Nature*. 2011; 470:204-213. doi:10.1038/nature09764
- Goodwin S et al. Coming of age: ten years of next-generation sequencing technologies. *Nature Reviews*. 2016; 17:333-351. doi: 10.1038/nrg.2016.49
- Haque F, et al. Solid-state and biological nanopore

“LA IMPLEMENTACIÓN DE LA MPS EN LA RUTINA DEL SISTEMA PÚBLICO DE SALUD SERÁ UNA REALIDAD EN LOS PRÓXIMOS AÑOS Y FACILITARÁ LA PERSONALIZACIÓN DE LA MEDICINA Y LA ADAPTACIÓN DEL TRATAMIENTO DEL PACIENTE, TANTO PARA EL MANEJO ADECUADO DE SUS SIGNOS Y SÍNTOMAS, COMO PARA LA ANTICIPACIÓN AL PROGRESO DE LA ENFERMEDAD.”

for real-time sensing of single chemical and sequencing of DNA. *Nano Today*. 2013; 8:56–74. doi: 10.1016/j.nantod.2012.12.008

International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. *Nature*. 2004; 431:931–945. doi: 10.1038/nature03001

Ledford H. AstraZeneca launches project to sequence 2 million genomes. *Nature*. 2016; 427. doi: 10.1038/nature.2016.19797

Monkol L, et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature*. 2016; 536: 285–291. doi: 10.1038/nature19057

Nyrén PI y Lundin A. Enzymatic method for continuous monitoring of inorganic pyrophosphate synthesis. *Anal. Biochem*. 1985; 509:504–509

Psifidi A, et al. Comparison of Eleven Methods for Genomic DNA Extraction Suitable for Large-Scale Whole-Genome Genotyping and Long-Term DNA Banking Using Blood Samples. *Plos One*. 2015; DOI:10.1371/journal.pone.0115960

Rhoads A, Au KF. PacBio Sequencing and Its Applications. *Genomics Proteomics Bioinformatics*. 2015;

13: 278–289. doi: 10.1016/j.gpb.2015.08.002

Sankar PL y Parker LS. The Precision Medicine Initiative's All of Us Research Program: an agenda for research on its ethical, legal, and social issues. *Genet Med*. 2016; (Online advance). doi: 10.1038/gim.2016.183

Vermeesch JR, et al. Prenatal and pre-implantation genetic diagnosis. *Nature Reviews Genetics*, 2016; 17:643–656