Métodos de secuenciación: tercera generación

por JOSÉ MIGUEL VALDERRAMA MARTÍN*, FRANCISCO ORTIGOSA* Y RAFAEL A. CAÑAS+

*Estudiante del Programa de Doctorado en Biología Celular y Molecular, 4ª Planta Edificio I+D, Facultad de Ciencias, Universidad de Málaga, 29071.

+Profesor Titular del Departamento de Biología Molecular y Bioquímica, 4ª Planta Edificio I+D, Facultad de Ciencias, Universidad de Málaga, 29071.

jmvalderrama@uma.es, fortigosa@uma.es, rcanas@uma.es

Las primeras técnicas de secuenciación masiva de ácidos nucleicos (2.ª generación) han permitido un extraordinario desarrollo de la Genómica y su «democratización». Sin embargo, presentan una serie de flaquezas, principalmente su incapacidad para generar lecturas mayores de 1 kb y la necesidad de amplificación previa de las muestras. En los últimos años se han desarrollado las denominadas como técnicas de secuenciación de tercera generación, las cuales han venido a paliar algunos de los defectos de las tecnologías precedentes. Por un lado, estas técnicas permiten obtener tamaños medios de lecturas de 30 kb, con máximos de 2,3 Mb. Por otro lado, no necesitan amplificación previa de las muestras de ácidos nucleicos, por lo que no se pierden las marcas epigenéticas que puedan presentar. Además, estas técnicas de tercera generación realizan la secuenciación directa del ARN, algo que no es posible con las técnicas precedentes. Así mismo, abren un nuevo camino en la secuenciación de ácidos nucleicos, sentando un nuevo precedente en el desarrollo de las Ciencias Genómicas hasta ahora desconocido.

The first techniques for massive sequencing of nucleic acids (2nd generation) have allowed an extraordinary development of Genomics and its «democratization». However, they present a series of weaknesses, mainly their inability to generate readings greater than 1 kb and the need for prior amplification of the samples. In recent years, so-called third-generation sequencing techniques have appeared, alleviating some of the shortcomings of previous technologies. On the one hand, they average read size reaches 30 kb, with maximums of 2.3 Mb in a read On the other hand, they do not need prior amplification of the nucleic acid samples, so the epigenetic marks that they may present are not lost. In addition, these third generation sequencing techniques can also perform direct sequencing of RNA, which is not possible with the preceding techniques. Thus, they point to a new path in nucleic acid sequencing, setting a new precedent in genomic science development unknown until now.

Palabras clave: secuenciación, tercera generación, SMRT, PacBio, Oxford Nanopore, MinION. Enviado: 05/05/2020
Keywords: sequencing, third generation, SMRT, PacBio, Oxford Nanopore, MinION. Aceptado: 31/10/2020

Las tecnologías de secuenciación de ácidos nucleicos de 2.ª generación han revolucionado la Biología durante los últimos 15 años. Su alto rendimiento y fidelidad de las lecturas han permitido el impulso definitivo de la Genómica tanto a nivel estructural (secuenciación de genomas) como funcional (por ejemplo, análisis transcriptómicos o epigenómicos) reduciendo los gastos al punto de poder secuenciar el genoma humano por menos de 1.000[1]. De todos los métodos de secuenciación de 2.ª generación es el de Illumina el que se sitúa como referente o estándar actual en cuanto a secuenciación masiva de ácidos nucleicos, representando mejor que ninguna otra plataforma las fortalezas y debilidades de estas tecnologías. Con respecto a sus principales debilidades está el reducido tamaño de las lecturas que producen (hasta 300 pb) lo que aumenta la probabilidad de producir errores de ensamblaje. Cuanto más pequeñas son las lecturas, más fácil es encontrar porciones de secuencia comunes a dos lecturas sin que estas pertenezcan a un mismo fragmento genómico o transcrito. En el caso de transcriptomas se pueden producir numerosas secuencias quiméricas (producto del ensamblaje de secuencias procedentes de dos o más genes diferentes). En este sentido los procesos de ayuste alternativo (*alternative splicing*) son una gran fuente de perturbación de los ensamblajes transcriptómicos provocando ensamblajes de secuencias que no se dan en la realidad. En el caso de los ensamblajes de genomas, las secuencias de pequeño tamaño, cuando se usa estrategias de secuenciación no basadas en mapas genéticos o físicos, pueden dar lugar a ensamblajes muy fragmentados del genoma. Esto sucede principalmente en el caso de genomas eucariotas de gran tamaño y con

un alto porcentaje de secuencias repetitivas como retrotransposones[2]. Por otro lado, las lecturas de pequeño tamaño también pueden llevar a problemas en la identificación de especies filogenéticamente cercanas en análisis metagenómicos de microorganismos. Otra flaqueza de estos métodos es la necesidad de amplificar la muestra de ácidos nucleicos previamente a la secuenciación, por lo que siempre se secuencian copias de las moléculas originales. Esto hace que se pierdan las modificaciones químicas, marcas epigenéticas, de los nucleósidos y, por tanto, generan la necesidad de aproximaciones previas y complementarias a la secuenciación para poder determinar la posición de estas modificaciones[3].

Desde un punto de vista tecnológico, el impulso creado por la aparición de las tecnologías de 2.ª generación ha propiciado el desarrollo de nuevos métodos que han intentado solventar los dos principales problemas de estos métodos descritos anteriormente: el tamaño de las lecturas y el uso directo de la muestra de ácidos nucleicos sin amplificar. A estos nuevos métodos se les ha denominado secuenciación de tercera generación. Entre sus características generales está la generación de lecturas de gran tamaño, una secuenciación monitorizada a tiempo real y no necesitan la amplificación previa de los ácidos nucleicos, siendo capaces de detectar modificaciones químicas de las bases nitrogenadas tanto de ADN como de ARN. A pesar de estas ventajas, estas tecnologías de secuenciación de 3.ª generación presentan dos grandes inconvenientes. El primero es que la cantidad de lecturas (cobertura) que pueden generar sigue siendo muy reducida en comparación con las que producen las tecnologías de secuenciación de segunda generación como Illumina. El segundo y principal problema de estas tecnologías es su menor fidelidad, siendo las tasas de acierto de hasta un 99,8 % (Q27) para PacBio[4] y hasta un 95 % (Q13) para Oxford Nanopore[5], las cuales son menores en comparación con las de tecnologías de segunda generación, cuyas tasas de fidelidad son cercanas al 99.99 % (>Q35). Esto es un condicionante muy importante para su desarrollo y expansión comercial a campos de estudio diferentes a la Genómica estructural, donde por el momento presentan sus mejores resultados. Gracias a su capacidad de generar lecturas largas pueden establecer guías robustas para realizar ensamblajes de calidad en especies con genomas grandes y repetitivos. Estos borradores genómicos, que por los inconvenientes de las técnicas previamente descritos no presentan una elevada fidelidad, se pueden corregir mediante el uso complementario de métodos de segunda generación, esto es conocido como ensamblaje genómico híbrido[6]. Por lo tanto, actualmente las plataformas de 3.ª generación no pueden sustituir a las previas y resulta esencial la coexistencia de diferentes tecnologías de secuenciación para diferentes usos.

Secuenciación a tiempo real de una sola molécula (SMRT, PacBio)

Esta técnica fue desarrollada durante la primera década del siglo XXI por la empresa *Pacific Biosciences* (PacBio) la cual lanzó al mercado su primer secuenciador en 2011[7]. El sistema de secuenciación a tiempo real de una sola molécula (SMRT, siglas en inglés) usa células de flujo que se basan en la tecnología «guía de onda de modo cero» (ZMW), que es un dispositivo que focaliza toda la luz hacia un punto para su detección. Esto permite registrar la señal lumínica emitida por un fluoróforo unido a un único nucleótido. En el fondo de cada ZMW se inmoviliza una molécula de ADN polimerasa que realiza la síntesis de la hebra complementaria de la molécula de ADN que se quiere secuenciar (Figura 1A y B). Esta ADN polimerasa está modificada para admitir nucleótidos hexafosfato que presentan un fluoróforo (específico para cada base nitrogenada) unido al último grupo fosfato[8]. Al introducirse un nucleótido durante la fase de síntesis se libera su fluoróforo emitiendo fluorescencia durante un intervalo de tiempo determinado (Figura 1C). La señal lumínica emitida es detectada e interpretada como la incorporación de un nucleótido específico dependiendo de la longitud de onda de emisión del fluoróforo[9]. De este modo, el proceso no requiere de ciclos de incorporación, sino que se realiza en continuo (a tiempo real) y permite la obtención de lecturas medias de unas 30 kb[10]. Los tamaños de secuenciación son dependientes de la vida media de las polimerasas ancladas al fondo de los micropocillos. Por otra parte, el sistema SMRT da la posibilidad de secuenciar una muestra de ácidos nucleicos sin necesidad de amplificarla como ocurría en las tecnologías de 2.ª generación, en las que los sistemas de detección necesitaban una «masa crítica» de moléculas clonales que se secuenciasen al mismo tiempo[3]. Una de las características de este tipo de metodología es la capacidad de secuenciar las dos hebras de un fragmento de ADN bicatenario gracias a adaptadores monohebra que forman bucles de horquilla (*SMRTbell template*) y unen ambas hebras del ADN molde. Uno de ellos une un cebador para la ADN polimerasa que así puede comenzar la síntesis/secuenciación y, tras sintetizar la hebra complementaria a una de las hebras del ADN molde, la polimerasa puede continuar sobre el otro adaptador del *SMRTbell template* y finalmente realizar la síntesis de la hebra complementaria de la segunda

hebra del ADN molde[9]. Esto permite generar una secuencia consenso que aumenta la fidelidad de la secuenciación (Figura 1B). En el caso de tratarse de una secuenciación de ARN, en el sistema SMRT la ADN polimerasa es sustituida por una reversotranscriptasa, lo que permite obtener la secuencia del ARN de la muestra pero sin tratarse de una secuenciación directa de ARN debido a la reversotranscripción[11]. Teóricamente, esto permite analizar los transcritos completos (*full length*) de los ARN mensajeros y con ello la detección y cuantificación de las variantes de ayuste alternativo (*alternative splicing*).

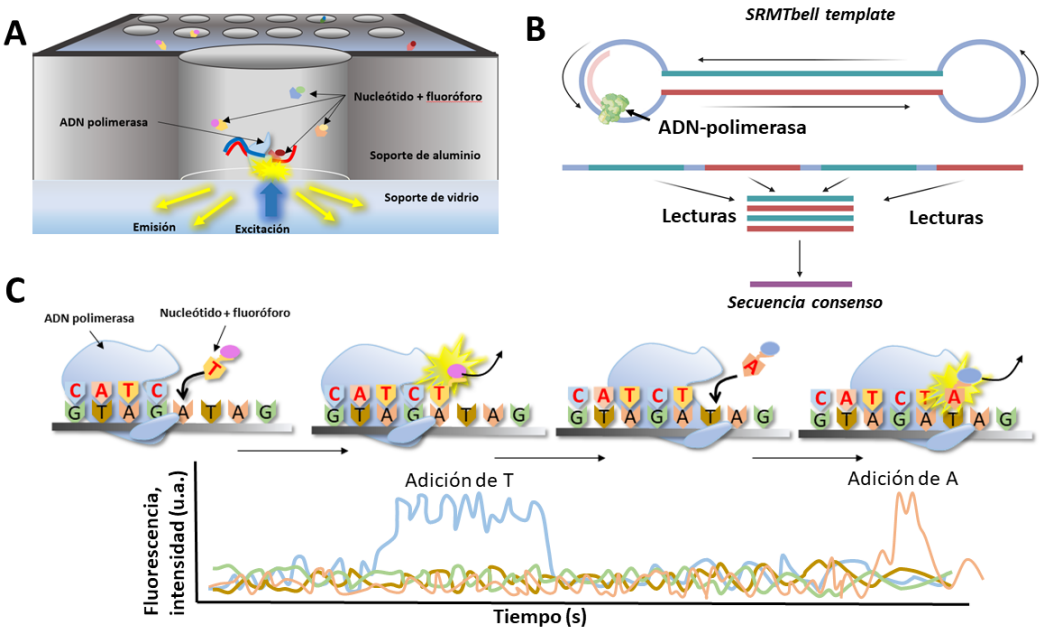

Figura 1. Representación esquemática de la tecnología de secuenciación SMRT de PacBio. A) ZMW contenido en una celda SMRT, en cuya parte inferior se encuentra una polimerasa inmovilizada. La unión del adaptador a la proteína anclada permite la síntesis de la hebra complementaria del ADN. La incorporación de un nucleótido marcado produce una liberación de fluorescencia que es detectada como la incorporación de un nucleótido específico atendiendo a la longitud de onda emitida. B) Proceso de síntesis de la hebra complementaria de ADN. Los adaptadores (lila) se encuentran ligados a los extremos de una doble cadena de ADN (azul y rojo) formando una estructura circular. La ADN polimerasa es la enzima encargada de llevar a cabo la síntesis de ADN. Posteriormente se obtienen un conjunto de fragmentos sintetizados (lecturas) que dan lugar a la obtención de una secuencia consenso. C) Cada nucleótido se encuentra marcado con un fluoróforo diferente. Cuando la ADN polimerasa incorpora un nucleótido marcado, se produce un pulso de fluorescencia que es detectado específicamente permitiendo la decodificación de la secuencia de ADN. El panel B fue diseñado con Biorender.

Una de las ventajas adicionales de la secuenciación *SMRT* es que se pueden determinar directamente modificaciones químicas de las bases nitrogenadas de los nucleótidos. En las técnicas de 2.ª generación se secuencian copias de la molécula de ADN o ARN original, lo que elimina las modificaciones químicas (marca epigenética) de los nucleótidos[3]. Por lo tanto, requieren procedimientos previos a la secuenciación para marcar los nucleótidos que presentan una determinada modificación química (por ejemplo, secuenciación por bisulfito para determinación de metilaciones en las citosinas[12]). Sin embargo, en el caso de las tecnologías de 3.ª generación es posible detectar la presencia de cualquier nucleótido modificado (casi cualquier modificación) en la secuencia sin necesidad de marcajes previos. La detección de los nucleótidos modificados se produce debido a que existe un retraso por parte de la ADN polimerasa en la introducción de los nucleótidos cuando se encuentra un nucleótido modificado[13] (Figura 2). Además, esta capacidad también permite la detección de marcas epitranscriptómicas en el ARN utilizando el mismo principio[11].

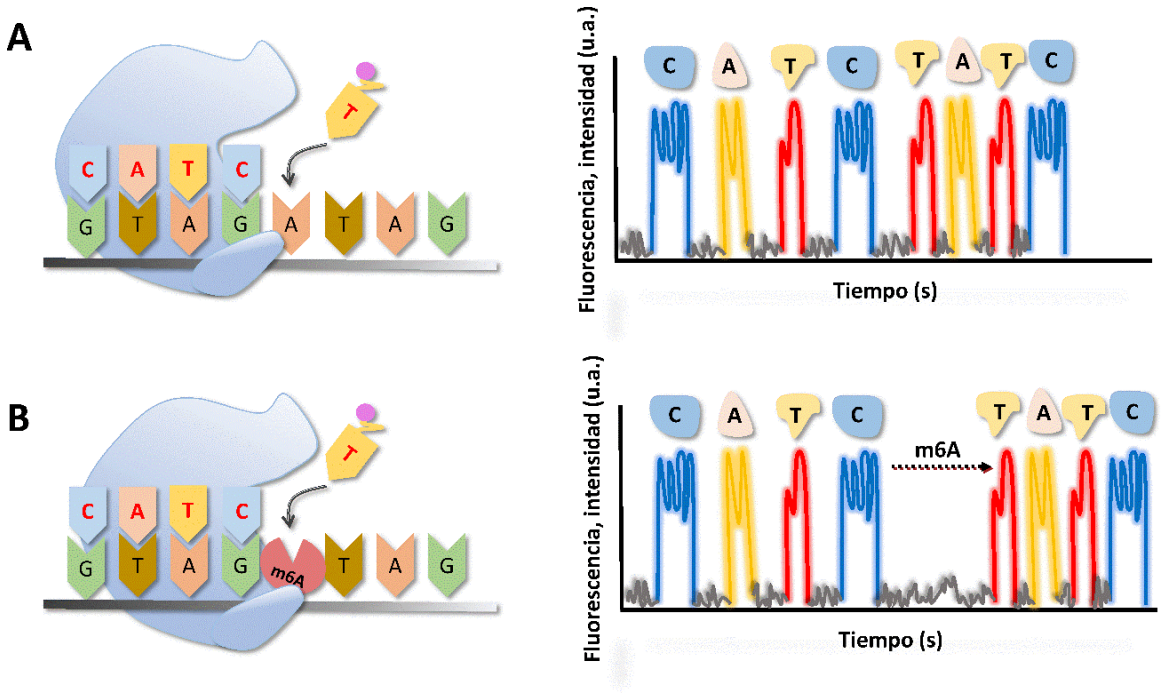

Figura 2. Detección de la modificación química del ADN N^6-metiladenosina (m^6A). Comparación esquemática del proceso de síntesis de ADN por la polimerasa en ausencia **(A)** y presencia **(B)** de m^6A. Cuando la ADN polimerasa encuentra en la secuencia molde una modificación química, se produce un retardo en el tiempo de incorporación del nucleótido lo que se traduce en una diferencia temporal en la detección del pulso de fluorescencia, indicando la presencia de m^6A.

Tecnología Oxford Nanopore (ONT)

De manera posterior a la aparición de la secuenciación SMRT de PacBio surge como alternativa *Oxford Nanopore Technology* (ONT). Esta empresa ha desarrollado su tecnología durante las dos primeras décadas del siglo XXI, incluso se puede considerar que todavía se encuentra en desarrollo. En abril de 2014, el MinION, un equipo de secuenciación miniaturizado y portátil que tiene un tamaño ligeramente mayor que el de una memoria USB, fue distribuido a más de 1.000 laboratorios para realizar pruebas beta a través del Programa de Acceso MinION (MAP). Un año después, comenzó su comercialización. Desde 2015 han aparecido dispositivos con diferentes capacidades incluso algunos que permiten la secuenciación sin la asistencia de un ordenador adicional para hacer secuenciación en el terreno[13].

Se trata de una tecnología de secuenciación totalmente diferente al resto de las presentadas y, a diferencia de ellas, no se basa en la complementariedad de los ácidos nucleicos. Su principio técnico se basa en los estudios iniciados por David Deamer durante la década de los 90. Se aprovecha de los perfiles de corriente eléctrica que genera cada molécula (en este caso nucleótidos) al pasar a través de un poro en una membrana a cuyos lados hay establecido un diferencial de voltaje[15,16]. De esta manera, es posible determinar la secuencia de una hebra de ácidos nucleicos de manera directa, sea ADN o ARN, algo que ni en el caso de la SMRT es posible puesto que realiza una reversotranscripción al secuenciar ARN. En el lado de entrada de las moléculas se establece una carga negativa y al otro extremo una carga positiva para facilitar el movimiento de los ácidos nucleicos. En el lado de la detección se incluye un electrodo y un circuito integrado (ASIC) (Figura 3A) para detectar la corriente eléctrica generada como consecuencia del paso de cada nucleótido a través de un poro y así registrar los datos en un nuevo formato de archivo de secuenciación FAST5 (los formatos generalmente obtenidos en secuenciación son los FASTQ). Estos archivos contienen la información bruta de los datos de corriente eléctrica de cada poro de la membrana que es traducido a nucleótidos cuando se realiza el proceso de *basecalling*[10]. Los poros incrustados en la membrana están formados por proteínas como la lipoproteína CsgG de *Escherichia coli* con modificaciones que permiten el paso específico a través de ella de ácidos nucleicos (Figura 3B)[17]. Además, los ácidos nucleicos a secuenciar deben unirse previamente a adaptadores de ADN unidos a una proteína motora, como la helicasa, que se unen al poro y promueven el paso de la hebra del ácido nucleico a través de és-

te[16] (Figura 3C). Gracias a ello, es posible generar lecturas extremadamente largas, llegando a obtenerse lecturas de hasta 2,3 Mb[10], tamaño superior que el de algunos genomas bacterianos.

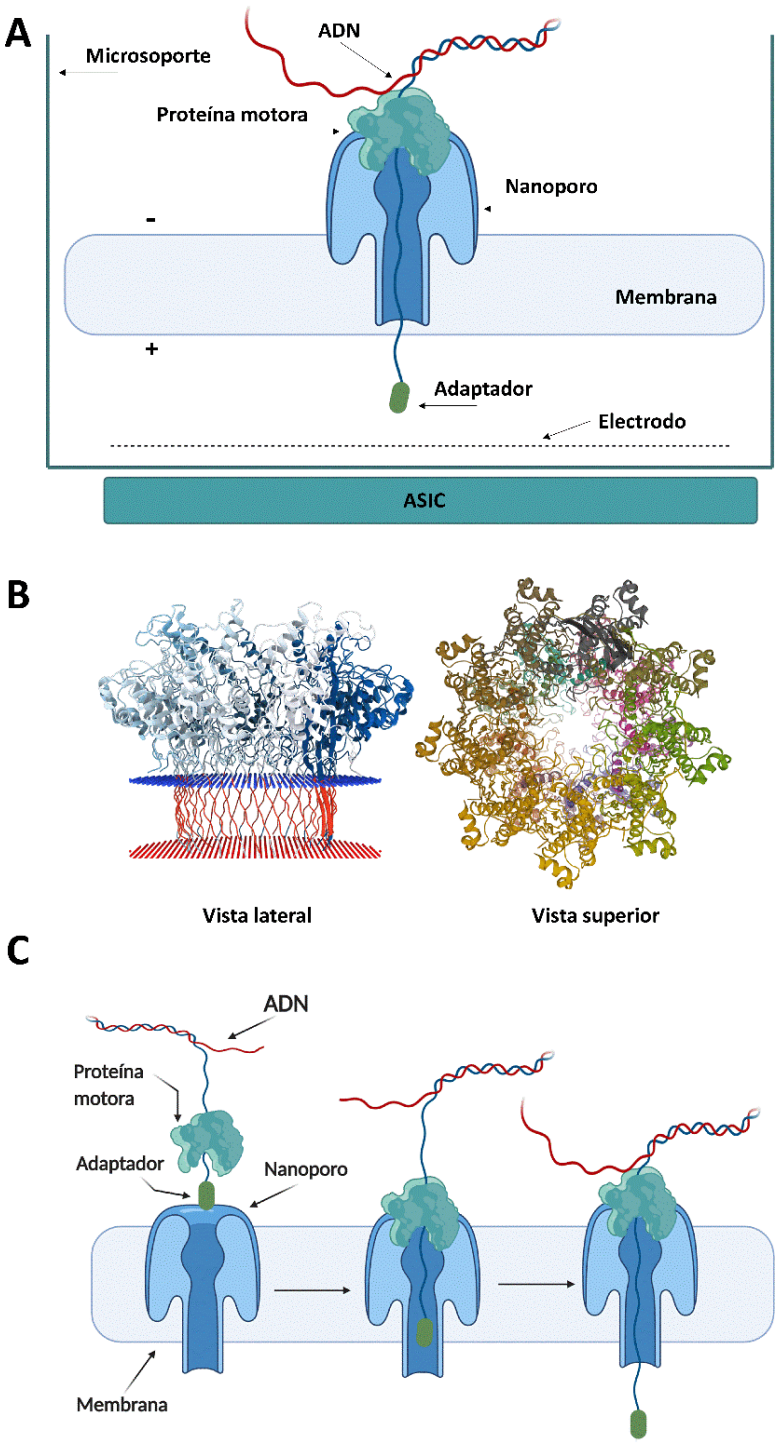

Figura 3. Representación de los principales componentes de la secuenciación de Oxford Nanopore. A) Ilustración esquemática de los componentes del soporte de secuenciación de la tecnología de Oxford Nanopore (ONT). B) Estructura de la proteína CsgG de *Escherichia coli*, una variante del poro proteico utilizado en el soporte MinION, PDB ID: 4UV3. C) Representación del complejo de secuenciación de ácidos nucleicos de la tecnología de Oxford Nanopore (ONT). Las figuras A y C fueron creadas usando Biorender.

En teoría, una de las ventajas de esta técnica sería que permitiría identificar cualquier molécula que atraviese el poro, incluyendo ácidos nucleicos, proteínas y metabolitos[17]. Por el momento, se usa solamente para la secuenciación de ácidos nucleicos, ADN y ARN. Esto convierte a ONT en la única plataforma de secuenciación distribuida en la actualidad que realiza la secuenciación directa de ARN sin necesidad de realizar una reversotranscripción. Además, como ocurre con la secuenciación SMRT de PacBio, al ser posible secuenciar directamente una muestra sin necesidad de realizar un proceso de amplificación, la tecnología ONT permite que en los análisis de RNA-Seq se necesiten menos lecturas para obtener un análisis estadístico robusto de las muestras[18], puesto que se evitan los sesgos derivados de los pasos de reversotranscripción y amplificación. La capacidad de detección de la huella eléctrica de cualquier ácido nucleico permite además la identificación de bases nitrogenadas con modificaciones químicas, pues éstas implican cambios en la corriente generada por la molécula al atravesar el poro con respecto al cambio que originaría el nucleótido en cuestión sin modificaciones químicas. Al igual que ocurre con el sistema SMRT de PacBio, estas características hacen idónea esta plataforma para su empleo en campos como la epigenómica, así como para la epitranscriptómica (estas marcas también se pueden detectar en el ARN[19]).

Como se ha mencionado en la descripción de esta técnica, un aspecto muy relevante es la longitud de las lecturas, actualmente las mayores de todas las plataformas de secuenciación. Además de en la Genómica estructural y del análisis de transcriptomas, esto tiene mucha relevancia en cuanto a los estudios metagenómicos, aquellos en los que se analiza el ADN/ARN de todos los organismos de una muestra biológica, suelo, medio marino, entre otros (generalmente microorganismos). Los tamaños de las lecturas hacen posible que la identificación de los organismos sea más precisa. Por norma general, las lecturas obtenidas en estos análisis se analizan frente a bases de datos de ácidos nucleicos, como las contenidas en GenBank, utilizando herramientas tipo BLAST[20] (*Basic Local Alignment Search Tool*, que realiza alineamientos de secuencias de ácidos nucleicos frente a bases de datos de ácidos nucleicos conocidos). La ventaja de tener grandes tamaños de lecturas es que permiten comparaciones más ajustadas con las secuencias de los organismos en las bases de datos. Esto se pudo comprobar en el seminario Málaga Nanopore celebrado en la Universidad de Málaga el 25 de abril de 2019[21], donde se presentó el trabajo de la empresa Xenogen. Su trabajo con esta tecnología demuestra que es muy superior para la identificación de microorganismos en estudios de metagenómica. Incluso puede ayudar en el caso del diagnóstico microbiológico de humanos en circunstancias en las que las técnicas de diagnóstico clásicas quedan rebasadas.

En contrapartida, también resulta importante hablar de los inconvenientes o factores limitantes que presenta esta técnica. La principal limitación es el desarrollo del software propio de análisis de los archivos de lecturas FAST5, que hasta el momento presenta una alta tasa de error. Sin embargo, gracias al continuo desarrollo de software de interpretación de datos se está consiguiendo aumentar de fiabilidad de los resultados obtenidos. Paralelamente a este inconveniente está el problema del almacenamiento de datos. De cada secuenciación se puede obtener más de 200 Gb de datos crudos, necesitando sistemas con un gran volumen de almacenaje y transferencia de datos. Por último, hay que tener en cuenta la naturaleza biológica (proteínas) de los nanoporos en los que se lleva a cabo la secuenciación, esto hace que las células de flujo que contienen los nanoporos tengan una vida media limitada. Todo ello marca la necesidad de una correcta organización para utilizar con éxito esta tecnología.

Conclusión

Desde las primeras aproximaciones de Frederick Sanger en los años 60 hasta la actualidad, los métodos de secuenciación de ácidos nucleicos han transformado las Ciencias Biológicas, dando lugar a que la Biología Molecular se encuentre presente en múltiples campos y abriendo el camino al desarrollo de nuevas disciplinas como la Genómica que comienza a ser ampliamente utilizada en el campo de la Medicina, Microbiología o Ciencias Ambientales, entre otros ejemplos. Estamos cerca de que estas técnicas nos permitan a un coste asequible la secuenciación del genoma de cualquier persona para su diagnóstico o para intentar implementar un tratamiento personalizado frente a las enfermedades. Gracias al desarrollo último de estas técnicas se ha podido secuenciar el virus SARS-CoV-2 en menos de un mes desde su detección[22], incluso realizar estudios sobre su propagación y analizar los diferentes subtipos o cepas que han aparecido. En retrospectiva, esto es algo increíble ya que el virus de la inmunodeficiencia humana (VIH) se aisló por primera vez en 1983[23] y hasta el año 1985[24] no se pudo secuenciar su genoma completo. Actualmente, se están intentando mejorar las técnicas de secuenciación y realizar nuevas aproximaciones que soslayen los problemas más frecuentes, como ensayar el uso de nanoporos compuestos de materiales robustos como el grafeno para sustituir

a las proteínas[25]. Cabe recalcar el hecho de que el avance y desarrollo de estas tecnologías no solo pasan por la mejora y subsanación de sus debilidades. La empresa de Oxford Nanopore Technology se encuentra en el desarrollo de un mecanismo que permita, además de la secuenciación de ácidos nucleicos, la secuenciación y análisis de proteínas utilizando las mismas plataformas físicas existentes[26,26], lo que podría ayudar al incremento en la eficiencia y rapidez de procesos como la identificación y validación de nuevas proteínas, pasos clave en la detección de biomarcadores de enfermedades. Partiendo de una tecnología diseñada para el estudio de los ácidos nucleicos se podrían revolucionar las Ciencias Biológicas transformando completamente disciplinas como la Proteómica y sustituyendo sistemas indirectos de identificación y cuantificación de proteínas como la espectrometría de masas. Probablemente, dentro de poco tiempo haya que añadir un nuevo capítulo a la historia de los métodos de secuenciación sin incluir «de ácidos nucleicos».

Referencias

[1] NIH, Coste de la secuenciación de un genoma humano. https://www.genome.gov.

[2] De la Torre y otros. Insights into Conifer Giga-Genomes. *Plant Physiology* 166: 1724-1732, 2014.

[3] Valderrama Martín y otros. Métodos de secuenciación de ácidos nucleicos: Segunda generación. Encuentros en la Biología 174: vol 13, verano 2020.

[4] Wenger y otros. Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. *Nature Biotechnology* 37: 1155-1162, 2019.

[5] Jain y otros. Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nature Biotechnology* 36: 338-345, 2018.

[6] Sohn y Nam. The present and future of de novo whole-genome assembly. *Brief Bioinformatics* 19: 23-40, 2018.

[7] GenomeWeb. https://www.genomeweb.com.

[8] Ermert y otros. Phosphate-Modified Nucleotides for Monitoring Enzyme Activity. *Topics in Current Chemistry* (Cham) 375: 28, 2017.

[9] Rhoads y Au. PacBio Sequencing and Its Applications. *Genomics Proteomics Bioinformatics*. 13: 278-289, 2015.

[10] Amarasinghe y otros. Opportunities and challenges in long-read sequencing data analysis. *Genome Biology* 21: 30, 2020.

[11] Vilfan y otros. Analysis of RNA base modification and structural rearrangement by single-molecule real-time detection of reverse transcription. *Journal of Nanobiotechnology* 11: 8, 2013.

[12] Wreczycka y otros. Strategies for analyzing bisulfite sequencing data. *Journal of Biotechnology* 261: 105-115, 2017.

[13] Flusberg y otros. Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nature Methods* 7: 461, 2010.

[14] Oxford Nanopore history. https://nanoporetech.com.

[15] Clarke J y otros. Continuous base identification for single-molecule nanopore DNA sequencing. *Nature Nanotech.* 4: 265-270, 2009.

[16] Feng y otros. Nanopore-based fourth-generation DNA sequencing technology. *Genomics, Proteomics and Bioinformatics*, 13: 4-16, 2015.

[17] Oxford Nanopore. Nanopore sensing - how it Works. https://nanoporetech.com/how-it-works.

[18] Byrne y otros. Nanopore long-read RNAseq reveals widespread transcriptional variation among the surface receptors of individual B cells. *Nature Communications* 8: 16027, 2017.

[19] Garalde DR y otros. Highly parallel direct RNA sequencing on an array of nanopores. *Nature Methods*, 15: 201, 2018.

[20] Altschul, S.F. y otros. Basic local alignment search tool. *J. Mol. Biol.* 215:403-410, 1990.

[21] Málaga Nanopore, noticia. https://www.aulamagna.com.es.

[22] Zhou y otros. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* 579: 270-273, 2020.

[23] Barré-Sinoussi y otros. Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science* 220: 868-871, 1983.

[24] Ratner y otros. Complete nucleotide sequence of the AIDS virus, HTLV-III. *Nature* 313: 277-284, 1985.

[25] Mojtabavi y otros. Single-Molecule Sensing Using Nanopores in Two-Dimensional Transition Metal Carbide (MXene) *Membranes*. ACS Nano 13: 3042-3053, 2019.

[26] Oxford Nanopore Protein Analysis. https://nanoporetech.com/applications.

[27] Nanopore-based 5D fingerprinting of single proteins in real-time (London Calling Presentation). https://nanoporetech.com.