

Genómica: donde la nanotecnología, la biología molecular y la informática convergen para romper récords

Por Andrés Iriarte y José Sotelo-Silveira*

La tecnología de la secuenciación de ADN ha avanzado en forma inimaginable para los científicos que iniciaron esa tarea. Lo que se hizo en 10 años, en cientos de metros cuadrados, con grandes equipos, requiriendo decenas de operarios, ahora puede hacerse en unas horas sobre una mesada de laboratorio. Lo que antes costaba millones de dólares, ahora cuesta unos pocos miles de dólares. Hoy es posible encontrar equipos para secuenciar genomas en varias instituciones públicas y privadas de Uruguay que tienen capacidad para determinar el ADN en un par de horas. La dificultad ya no es obtener la secuencia del genoma sino manejar, analizar y clasificar esa enorme cantidad de información.

En el ADN se encuentran las instrucciones para generar vida. Cuando José Sotelo, co-autor de esta nota, vio por primera vez ADN purificado, hace casi 25 años, estaba viendo millones de moléculas enredadas en un fino hilo blanco de pocos centímetros.

Esto ocurrió en una actividad de laboratorio en el primer año de la Licenciatura de Ciencias Biológicas en la Facultad de Ciencias. Increíblemente, se podía, usando una fina barra de vidrio, extraerlo de un líquido y mirar el famoso ADN con nuestros propios ojos. Allí se escondía, y todavía sigue parcialmente oculta, sin descifrar, la información necesaria para generar un individuo con todas sus características físicas y psíquicas, o sea, todas las instrucciones para desarrollar un organismo partiendo de una sola célula única original.

Información codificada

Toda la información mencionada está codificada en una larga sucesión de bloques básicos o subunidades moleculares, conocidos como nucleótidos, cada uno de ellos integrado, entre

otros componentes, por una de las siguientes cuatro bases: Adenina (A), Timina (T), Guanina (G) y Citosina (C), dispuestas en dos cadenas que se orientan en direcciones opuestas, anti-paralelas, y se mantienen apareadas de forma absolutamente específica, es decir cada base se combina con una de la otra cadena, una A con una T y viceversa y una C con una G y viceversa. Este fenómeno es conocido como complementariedad de las bases.

Cada una de estas 2 cadenas se forma por enlaces muy estables entre los nucleótidos que forman el sostén de la estructura. Ambas cadenas se mantienen unidas entre sí por enlaces de hidrógeno entre las bases, formando como escalones de una escalera que se enrolla en forma helicoidal. Si bien parece que fue ayer, la estructura del ADN fue descubierta por Watson y Crick en 1953. ¡Hace más de 60 años!

En la secuencia de bases de cada cadena, es decir, en el orden en que se colocan las bases, por ejemplo A, C, G, T, A, G, T, C, T, etc. y sus complementarias en la cadena opuesta (T, G, C, A, T, C, A, G, A), se codifica la información e incluso el mecanismo de herencia de los seres vivos. Esta información es vasta y los mecanismos que aseguran la interpretación de la información tienen un grado de complejidad extraordinario.

La escalera de un genoma humano tiene 3.200 millones de nucleótidos. Un nucleótido en el ADN tiene un largo de 0,34 nanómetros [un nanómetro (nm) es igual a la mil millonésima parte de un metro (m) o 0,000000001 m]. Es claro entonces que el código genético está escrito en caracteres extremadamente pequeños, pero al mismo tiempo la molécula tiene tanta información que si se desenrolla una de ellas, de una única célula, tendría un largo de aproximadamente un metro.

Secuenciación

Con ese tamaño de la molécula, conocer la secuencia de los nucleótidos de toda la cadena -identificados según la base que lo compone- parecía una tarea imposible.

Esta información es esencial para el ser vivo para producir decenas de miles de proteínas distintas, en forma precisa, tanto en cuanto a cantidad, como a lugar de ubicación y momento temporal de producción. Por si fuera poco, existe además un



Grandes cantidades de ADN se observan suspendidas desde una barra de vidrio



Diferentes dimensiones de lectores de ADN. Dos secuenciadores masivos en el IIBCE: el portable y de última generación (Oxford Nanopore), apoyado encima del equipo multipropósito pero de mayor tamaño, Ion Torrent PGM. Recuadro inferior izquierdo, la nueva generación de secuenciador por nanoporos se usa directamente conectado a una laptop (el mate no es necesario, solo se muestra con fines de escala ;). Recuadro inferior derecho, se muestra el chip donde se lee la secuencia del ADN dentro del secuenciador Ion Torrent (IIBCE).

número todavía no conocido de moléculas de ARN (1) funcionalmente fundamentales y otro importante número de regiones en la cadena de ADN que aseguran la regulación de la expresión génica. Con el término regulación nos referimos a todos los aspectos que permiten la interacción de la célula con su medio ambiente y las respuestas a sus cambios internos.

Para entender muchas de las características de los organismos y de sus respuestas al ambiente, es necesario entender el funcionamiento y la interacción de vastas regiones del genoma. Es decir, no alcanza con entender cómo se produce una proteína particular, sino varias de ellas y sus interacciones. Aquí es donde radica la complejidad y el apasionante misterio que subyace a ese universo que es el genoma. Este misterio ha llevado a los biólogos moleculares, a invertir décadas en el estudio del mismo.

Primer genoma humano

Durante muchos años solamente se obtenía información detallada de algunos genes (zonas de la cadena con información para producir ciertas sustancias) particulares y específicos en el genoma, ya que acceder al orden de cientos de millones de

bases era un esfuerzo colosal, desde el punto de vista económico y técnico.

De hecho la secuenciación del primer genoma humano requirió la unión de estados y países, organizaciones académicas y científicas de todo el planeta, con una inversión de más de 3.000 millones de dólares. El esfuerzo conjunto comenzó decisivamente en el año 1990 y culminó 10 años más tarde. Hoy, el adelanto en los procedimientos ha sido tan grande que se puede secuenciar un genoma humano en días con un costo aproximado de mil dólares.

Cuando Andrés Iriarte, co-autor de esta nota, ingresó en la Facultad de Ciencias en el año 2000, se estaba terminando el primer borrador del genoma humano, un hito histórico que ha cambiado y seguirá cambiando la forma de estudiar la vida humana y la vida en general. Con el tiempo se ha comprobado que las primeras versiones de los genomas tienen errores. Cuando Andrés terminó de recibirse de Doctor en Ciencias Biológicas en el Programa de Desarrollo de Ciencias Básicas de Uruguay (PEDECIBA), en 2013, y comenzó a trabajar en el Instituto Clemente Estable, la situación superaba ampliamente las expectativas y predicciones originales.

Obtener genomas completos se había transformado en un experimento relativamente común. En estas últimas dos décadas la secuenciación de genomas se ha aplicado a otras miles de especies de animales y plantas, a decenas de miles de bacterias y arqueas (2) incluso a miles de personas. Es que la secuenciación de genomas se está aproximando rápidamente a una técnica de rutina, a una prueba “común” de laboratorio (no hay más que ver CSI Las Vegas, por dar un ejemplo entre muchos, donde secuencian ADN para identificar delincuentes).

La tecnología de la secuenciación logró paralelizar y miniaturizar el proceso de secuenciación a un nivel de ciencia ficción. Lo que al inicio se hizo en 10 años en cientos de metros cuadrados de laboratorios, con grandes equipos, requiriendo decenas de operarios, ahora puede hacerse en unas horas, sobre una mesada de laboratorio. Lo que antes costaba millones de dólares, ahora cuesta unos pocos miles de dólares. Hoy es posible encontrar equipos para secuenciar genomas en varias instituciones públicas y privadas de Uruguay.

Genómica

Ahora que se utilizan equipos con capacidad para obtener la información contenida en la secuencia de millones de nucleótidos de ADN, en un par de horas, la dificultad ya no es obtener la secuencia del genoma sino manejar, analizar y clasificar esa enorme cantidad de información. Además, ubicar los genes o regiones reguladoras, predecir la función de los mismos y entender el efecto de los cambios que se producen a nivel de la secuencia del ADN.

Ocurre que obtener el orden de los nucleótidos, es decir, leer la secuencia de ADN, no es suficiente para entender cómo la información contenida en el ADN es manejada por la célula para producir el fenotipo, es decir todo aquello que determina el aspecto físico y químico de cualquiera de los organismos vivos, pero es una etapa clave. Tal es así que a más de 10 años del hito histórico en el que se da a conocer la secuencia del primer genoma humano, el más estudiado de los genomas, todavía existen muchos segmentos con función desconocida.

Como en muchos otros casos en la ciencia, la capacidad para acceder a un tipo de información genera muchas más preguntas que respuestas. Con la secuencia del genoma en nuestras manos es posible plantearnos preguntas sobre la verdadera complejidad del funcionamiento de los organismos vivos, es decir la forma como la información es decodificada en la célula a lo largo del tiempo, en los tejidos, en el organismo; la regulación de la forma y cómo es utilizada esa información.

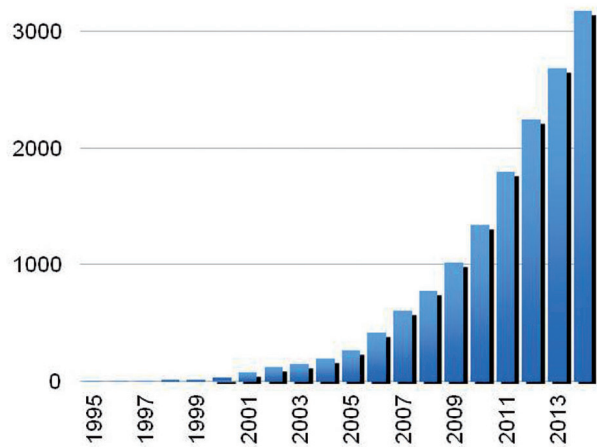
Por lo antedicho, queda claro que obtener la secuencia de nucleótidos es solo un primer paso, pero es un paso fundamental para entender en forma cabal todos los aspectos del funcionamiento de los seres vivos. Es aquí donde se oculta la verdadera importancia de la información genómica, como aporte fundamental al desarrollo de todas las disciplinas de las ciencias de la vida. Esto abarca desde la biología de la conservación a la medicina, pasando por la producción agrícola, el mejoramiento genético animal, la microbiología o la neurobiología.

Un caso de fundamental relevancia para la salud humana es el esfuerzo por catalogar todas las mutaciones o cambios nocivos que ocurren en el genoma humano y que caracterizan al cáncer. Proyectos multilaterales, es decir llevados a cabo por instituciones de varios países, han realizado, en poco tiempo, un atlas (<http://cancer-genome.nih.gov/>) que describe por primera vez todo el espectro de cambios que subyacen en diferentes tipos de cáncer. Conociendo el mapa con precisión será más fácil navegar.

Genómica de microorganismos

En 1995 se publicó la secuencia del primer genoma completo de un organismo de vida libre, la bacteria *Haemophilus influenzae*. Desde entonces el número de genomas completos ha aumentado enormemente de una forma que pocos podían prever.

Detrás de los proyectos de secuenciación de decenas o incluso centenas de genomas completos se encuentran objetivos tan importantes como el desarrollo de nuevos agentes antimicrobianos, la manipulación de cepas bacterianas especializadas en la biorremediación, o con potencial biotecnológico, la identificación de proteínas como biomarcadores de enfermedades, el desarrollo

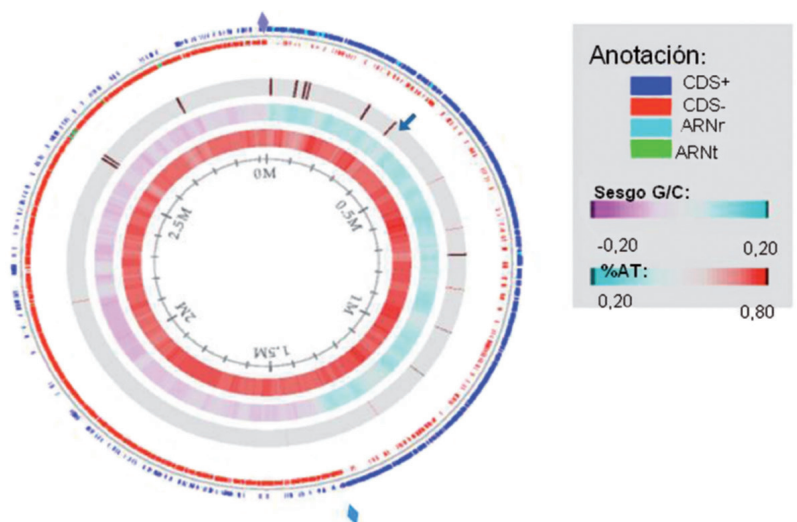


Genomas bacterianos secuenciados disponibles en la base de datos del genbank (base de datos donde se alberga la información genómica de todos los organismos secuenciados que se publican en revistas científicas).

“inteligente” de vacunas o identificación de genes de virulencia. Solo a modo de ejemplo la secuencia de *Clostridium perfringens*, publicada en el 2002, permitió localizar 23 genes responsables de la particular virulencia de este organismo. Con el tiempo se explicó el mecanismo de virulencia y se identificaron otros genes que también estaban involucrados.

Una de las características sobresalientes de los microorganismos procariontes (bacterias), y una de las claves de su potencial, es su diversidad ecológica. Podemos encontrar procariontes en prácticamente cualquier ambiente, incluso en los lugares más inhóspitos y extraños tales como: desiertos, el fondo del mar, ambientes contaminados, el hielo de la Antártida, cerca de volcanes, salares, etc. Ellos pueden poseer genes particulares con alguna capacidad inusual para, por ejemplo, el procesamiento de sustancias o nutrientes específicos, resistencia a antibióticos, capacidad de invadir células de un tejido u hospedero particular, entre otras. El descubrimiento, la manipulación y uso de estos genes es muy importante para el desarrollo de diferentes aplicaciones biotecnológicas.

Representación esquemática del genoma bacteriano de la bacteria *Clostridium perfringens*, secuenciado completamente. Cada círculo corresponde a propiedades particulares del mismo. En el más externo se muestra la posición de sus 2753 genes (en rojo y azul). Otras características pueden representarse en círculos sucesivos.



En el IIBCE existen, desde hace varios años, líneas de investigación que estudian e identifican capacidades metabólicas de interés biotecnológico.

Un caso particular es la búsqueda de enzimas capaces de trabajar a bajas temperaturas (como las de la Antártida), potencialmente utilizables en aplicaciones a nivel industrial. Una vez aislados los microorganismos productores de las enzimas de interés, es posible obtener la secuencia del genoma y analizarlo en búsqueda de los genes codificantes para la enzima. Estos genes son insertados en otros organismos seleccionados para producir enzimas en cantidades mayores en condiciones controladas. El proceso de producción de la enzima es luego escalado para que la enzima se transforme en un producto comerciable y aplicable a nivel industrial. Entre muchas otras cosas, también se estudian genes responsables de la fijación del nitrógeno (bacterias que proporcionan el nitrógeno a las plantas), procesos fundamentales como la nodulación, desde el punto de vista ecológico con aplicaciones en la producción agrícola.

La genómica permite una visión global de la información genética y esta es una diferencia fundamental con otras aproximaciones tradicionales. Integrando esta nueva perspectiva a las clásicas en fenómenos ampliamente estudiados, pensamos que se avanzará rápidamente hacia una mejor comprensión de las bases biológicas de la vida. El IIBCE ha incursionado en esta disciplina, formando un departamento nuevo de Genómica junto con un servicio de secuenciación y análisis de datos genómicos, disponible para la interacción con la comunidad científica que trabaje tanto en áreas de ciencia básica o aplicada, así como diferentes actores del sector privado.

Pensamos que en este campo, el avance tecnológico está generando nuevos caminos donde los científicos pueden transitar hacia metas concretas que mejorarían nuestra calidad de vida. Para lograr estas metas, es esencial seguir invirtiendo en tecnologías de punta, como éstas, manteniendo así un ritmo de desarrollo científico digno y competitivo.

Algunas claves para poder leer el ADN en forma ultrarrápida

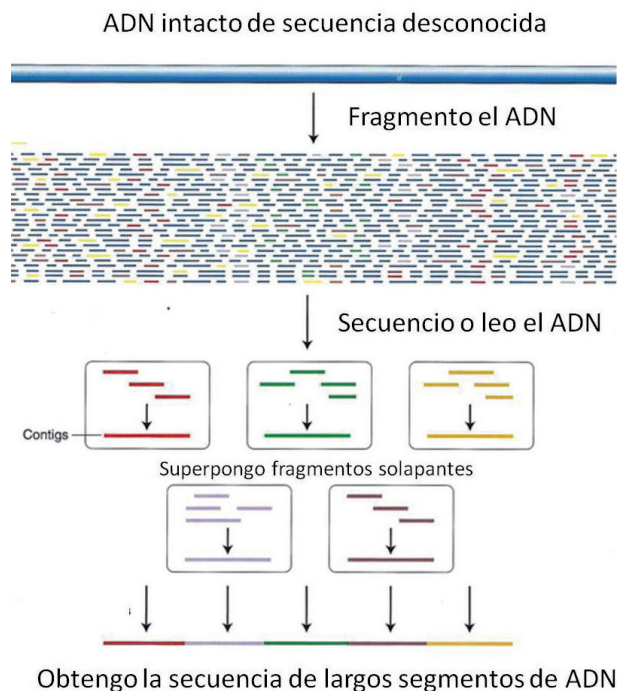
Para poder leer el ADN en forma rápida la ciencia ha dado enormes pasos basados en una premisa, esta es apoyarse en todas las ideas previas obtenidas en diferentes campos de la ciencia para innovar en estrategias, alternativas y metodologías. Dicho de otra manera, trabajar paso a paso, hasta que un paso tenga la capacidad de transformar la caminata en una carrera hacia la concreción del objetivo a largo plazo planteado inicialmente.

Primer Paso Divide y vencerás Fragmenta y leerás

Desde la invención de los primeros métodos de secuenciación o lectura de ADN, inventados por Fred Sanger, Walter Gilbert y Allan Maxam hace ya 40 años, el desafío primario era poder leer la secuencia del ADN, pero dentro de este estaba el desafío de leer moléculas extremadamente largas, ya que se conocía que los genomas de los seres vivos tenían longitudes que van desde fracciones de centímetros hasta metros. Esa barrera recién empieza a ser quebrada en estos momentos (ver más adelante el secuenciado a través de poros proteicos). Hace veinte años sólo podíamos leer pocos segmentos con un máximo de 800 bases en meses. Hoy seguimos leyendo segmentos de ADN, pero en "cantidades industriales" (como se dice en el lenguaje común) en cuestión de horas. La clave es que podemos, en un espacio físico muy pequeño, leer millones y millones de fragmentos cortos de ADN, de ahí el término utilizado para describir estas metodologías: "secuenciación masiva".

Segundo Paso Miniaturiza el proceso y genera miles de lecturas de ADN en paralelo

Para leer el primer genoma se utilizaron secuenciadores que ocupaban pisos enteros de los centros de investigación. Hoy, cada "lector" se ha reducido tanto en sus dimensiones que el lugar físico donde se lee el ADN puede caber en la palma de la



mano. Sin embargo estos lectores están rodeados por un sistema de captura y detección de la lectura de ADN, que termina siendo finalmente lo que ocupa más espacio hoy día.

Tercer Paso

Colabora con los mejores profesionales de la lectura de ADN o inventa nuevos lectores

Cuarenta años de desarrollo de la biología molecular nos han dejado una caja de herramientas moleculares para manejar las moléculas de ADN. En dicha caja tenemos tijeras, conectores, ligadores, marcadores, copadoras, reparadores y otras herramientas moleculares que nos permiten trabajar y operar en este universo ultra microscópico. Estas herramientas son nada más ni nada menos que proteínas (enzimas) capaces de realizar tareas con precisión. La polimerasa de ADN es capaz de copiar ADN en forma muy precisa. Esta capacidad de copiar es utilizada ingeniosamente, en varios de los métodos de secuenciación moderna, para obtener la información copiada a cada instante, es decir la secuencia del ADN. Los científicos hemos logrado observar esta enzima al copiar el ADN, registrando el proceso a medida que avanza. Ella lee por nosotros, nosotros seguimos su rastro e interpretamos su trabajo.

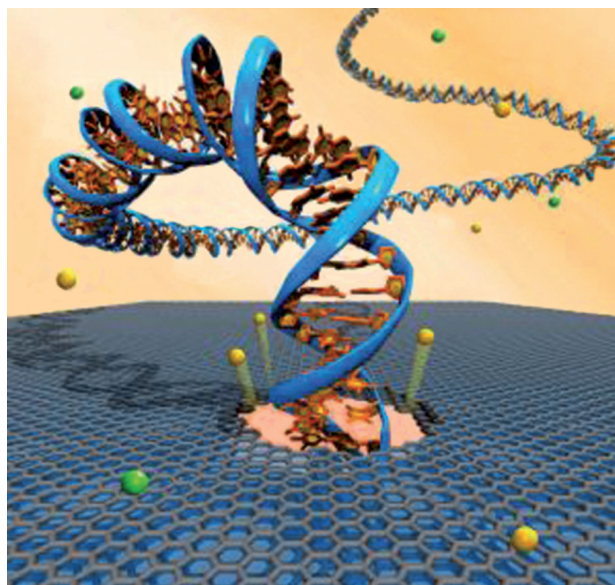
Recientemente otros métodos se han basado en un tipo de herramienta molecular diferente e innovadora: un complejo molecular compuesto por proteínas que forman lo que se puede describir como un poro, o tubo, o canal molecular. Los que lo desarrollaron han sido tan ingeniosos que hicieron posible descifrar la secuencia del ADN al hacerlo pasar por el mencionado canal. Este novedoso instrumento (de la compañía Oxford Nanopore), en uso hoy en el Departamento de Genómica del Instituto de Investigaciones Clemente Estable (IIBCE), es el que hoy puede leer los segmentos más largos de ADN, de unos diez mil nucleótidos o incluso más. Increíblemente el equipo entero cabe en la palma de la mano y además solo necesita una computadora portátil (*laptop*) para registrar y almacenar los datos de secuencia del ADN.

Cuarto Paso

Prepárate para analizar la gran cantidad de datos de secuencia producidos

Una vez terminada la lectura de un genoma humano realizado en el laboratorio, tenemos en nuestro poder el equivalente en datos a un disco duro de una computadora portátil (250 giga bytes o 250 mil millones de bytes). Para tener un punto de comparación, la colección completa de los libros de Harry Potter en versión electrónica que cargamos en un tablet tendría solamente cerca de 20 millones de bytes, 10.000 veces menos.

Para ordenar esta información es necesario hacer uso de modernos procedimientos informáticos. Si pensamos que la capacidad de generación de datos genómicos y sus aplicaciones se multiplican con el paso del tiempo, se necesitarán en breve nuevos y numerosos especialistas en análisis de datos. Algunos ya se están formando en el marco del programa de desarrollo de las ciencias básicas (PEDECIBA), a través de la maestría en bioinformática, que se creó en 2009.



Representación gráfica de secuenciación por nanoporos. Se observa una hebra de ADN que entra a través de un nanoporo de grafeno así como los iones que pasan por el mismo poro. Debido al tamaño reducido del poro, el ADN interfiere con la corriente de los iones y este cambio permite identificar los nucleótidos individuales. El instrumento de Oxford Nanopore utiliza poros proteicos. Imagen de Ayswaryak. Disponible bajo la licencia CC BY-SA 3.0 via Wikimedia Commons.

Agradecimientos

Agradecemos a Guillermo Eastman, Valeria Romero y a Joaquina Farías, estudiantes del Programa de Desarrollo de Ciencias Básicas (PEDECIBA) realizando sus estudios en el IIBCE, por su participación en la discusión y ensamblado de esta nota.

Notas del editor

1. El ácido ribonucleico (ARN) es un ácido nucleico formado por una cadena de ribonucleótidos. Está presente en las células, y es el único material genético de ciertos virus (virus ARN). El ARN celular es lineal y de hebra sencilla, pero en el genoma de algunos virus es de doble hebra.

En los organismos celulares desempeña diversas funciones. Dirige las etapas intermedias de la síntesis proteica; el ADN se vale del ARN para transferir la información durante la síntesis de proteínas (producción de las proteínas que necesita la célula para sus actividades y su desarrollo). Varios tipos de ARN regulan la expresión de los genes, mientras que otros tienen actividad catalítica.

2. Las archeas son un grupo de microorganismos unicelulares que, al igual que las bacterias, son de morfología procariota (sin núcleo ni, en general, orgánulos membranosos internos), pero que son fundamentalmente diferentes a estas, de tal manera que conforman su propio dominio o reino.

**Andrés Iriarte es becario posdoctoral del Departamento de Bioquímica y Genómica Microbianas y Profesor Adjunto del Departamento de Desarrollo Biotecnológico de la Facultad de Medicina, Universidad de la República y José Sotelo-Silveira es Profesor Adjunto del Departamento de Biología Celular y Molecular, Facultad de Ciencias, Universidad de la República y Profesor Titular de Investigación Profesional grado 5 del Departamento de Genómica del Instituto de Investigaciones Biológicas Clemente Estable.*