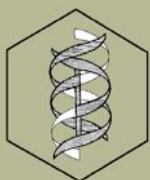


Revista de Educación Bioquímica

REB 2018



Órgano de información de la
Asociación Mexicana de
Profesores de Bioquímica, A.C.



Asociación Mexicana de
Bioquímicos, A.C.

Departamento de Bioquímica
Facultad de Medicina
UNAM

Facultad de Medicina



EDITADO DESDE 1982 COMO BOLETÍN DE EDUCACIÓN BIOQUÍMICA

COMITÉ EDITORIAL

EDITOR EN JEFE

JOSÉ VÍCTOR CALDERÓN SALINAS

Departamento de Bioquímica
Centro de Investigación y de Estudios Avanzados

EDITORES

RAFAEL CAMACHO CARRANZA

Instituto de Investigaciones Biomédicas
Departamento de Medicina, Genómica y Toxicología Ambiental
Universidad Nacional Autónoma de México

KARLA GUADALUPE CARVAJAL AGUILERA

Instituto Nacional de Pediatría

ALICIA GAMBOA DE BUEN

Instituto de Ecología
Universidad Nacional Autónoma de México

MARCO ANTONIO JUÁREZ OROPEZA

Facultad de Medicina
Universidad Nacional Autónoma de México

MARÍA ESTHER REVUELTA MIRANDA

Sección Bioquímica y Farmacología Humana
Facultad de Estudios Superiores Cuautitlán

ROCÍO SALCEDA SACANELLES

Instituto de Fisiología Celular
Universidad Nacional Autónoma de México

YOLANDA SALDAÑA BALMORI

Facultad de Medicina
Universidad Nacional Autónoma de México

VÍCTOR M. VALDES LÓPEZ

Facultad de Ciencias
Universidad Nacional Autónoma de México

ÁNGEL ZARAIN HERZBERG

Facultad de Medicina
Universidad Nacional Autónoma de México

EDITORES FUNDADORES

GUILLERMO CARVAJAL SANDOVAL

Instituto Nacional de Enfermedades Respiratorias
e Instituto Politécnico Nacional

JESÚS MANUEL LEÓN CÁZARES

Facultad de Ciencias Naturales
Universidad Autónoma de Querétaro

ENRIQUE PIÑA GARZA

Facultad de Medicina
Universidad Nacional Autónoma de México

YOLANDA SALDAÑA BALMORI

Facultad de Medicina
Universidad Nacional Autónoma de México

SERGIO SÁNCHEZ ESQUIVEL

Instituto de Investigaciones Biomédicas
Universidad Nacional Autónoma de México

CORRESPONSALES

ROCÍO SALCEDA SACANELLES

Coordinadora
Instituto de Fisiología Celular
Universidad Nacional Autónoma de México

ALEJANDRO MARTÍNEZ MARTÍNEZ

Instituto de Ciencias Biomédicas
Universidad Autónoma de Ciudad Juárez, Chihuahua

SERGIO SÁNCHEZ-ARMÁSS ACUÑA

Departamento de Fisiología y Farmacología
Facultad de Medicina, Universidad Autónoma de San Luis Potosí

MARTHA ANGÉLICA QUINTANAR ESCORZA

Departamento de Bioquímica
Facultad de Medicina Humana, Universidad Juárez del Estado de Durango

MARÍA MALDONADO VEGA

Departamento de Investigación. Centro de Innovación Aplicada en
Tecnologías Competitivas, AC. León, Gto., Mexico

Publicación incluida por el Centro de Información Científica y Humanística de la Universidad Nacional Autónoma de México en la base de datos Periódica, Iresie, Redalyc, Latindex y Scielo.

REVISTA DE EDUCACIÓN BIOQUÍMICA (REB), Volumen 37, Número 3, septiembre de 2018, publicación trimestral, editada por Asociación Mexicana de Profesores de Bioquímica, A.C., Avenida Universidad No. 3000, Colonia Universidad Nacional Autónoma de México, Delegación Coyoacán, C.P. 04510. Ciudad de México, México. Correspondencia: Apartado Postal 70-281, Coyoacán, C.P. 04510, Ciudad de México, México. Correo E: reb@bq.unam.mx

<http://bq.unam.mx/wikidep/pmwiki.php/Reb/HomePage>

http://www.facmed.unam.mx/marco/index.php?dir_ver=101

Editor responsable Dr. José Víctor Calderón Salinas. ISSN: 1870-3690. Reserva de derechos al Uso Exclusivo No. 04-2015-113014523300-203, ambos otorgados por el Instituto Nacional del Derecho de Autor. Diseño: Celia Virginia Sánchez Meza, Departamento de Bioquímica, Facultad de Medicina, UNAM. Disponible en septiembre del 2018.

El contenido de los artículos y las opiniones expresadas en ellos son responsabilidad de los autores y no reflejan necesariamente las del Comité Editorial. Se autoriza la reproducción parcial o total de los textos aquí publicados siempre y cuando se cite la fuente completa y la dirección electrónica de la publicación.

CONTENIDO

COMITÉ EDITORIAL

EDITORIAL

ACCESO LIBRE A LA INFORMACIÓN
CIENTÍFICA, ¿EXISTEN BARRERAS PARA LA
CIENCIA?
Jesús Martínez Sámano y
Marco Antonio Juárez Oropeza.....65

ARTÍCULOS

IMPLICACIONES CLÍNICAS DE LA
EXPRESIÓN DE GENES QUE CODIFICAN
PARA PROTEÍNAS DE LAS UNIONES
ESTRECHAS: EL CASO DE *TJP1*
Eric Ramírez-Bohórquez,
Diana García-Cruz,
Diego Ortega-Pacheco,
José I. Siliceo-Murrieta,
Lorena Guadalupe Ramón-Canul,
Sergio Alberto Ramirez-Garcia.....67

ESTUDIO DEL TRANSCRIPTOMA MEDIANTE
RNA-SEQ CON ÉNFASIS EN LAS ESPECIES
VEGETALES NO MODELO
Gustavo Rodríguez-Alonso,
Svetlana Shishkova.....75

OTRAS COMUNICACIONES

CRUCIBIOQ
CONCEPTOS DE GENÉTICA
Yolanda Saldaña Balmori.....89

AMINO-ESTRUCTURA: UNA PROPUESTA
PEDAGÓGICA PARA EL APRENDIZAJE
DE LOS VEINTE AMINOÁCIDOS QUE
COMÚNMENTE SE ENCUENTRAN EN LAS
PROTEÍNAS
Juan Carlos Vega Garzón.....91

SOLUCIÓN AL CRUCIBIOQ
CONCEPTOS DE GENÉTICA
Yolanda Saldaña Balmori.....95

INSTRUCCIONES PARA LOS
COLABORADORES DE LA REVISTA DE
EDUCACIÓN BIOQUÍMICA.....96

EDITORIAL

ACCESO LIBRE A LA INFORMACIÓN CIENTÍFICA, ¿EXISTEN BARRERAS PARA LA CIENCIA?

Todos aquellos que nos dedicamos a hacer ciencia y su divulgación tenemos la gran necesidad de realizar consultas en revistas especializadas en nuestras disciplinas, es un hecho que en muchas ocasiones no es posible el acceso debido a que se debe de pagar cierto costo por el acceso a la revista y/o al artículo. México es considerado un país de ingresos medios-altos en donde, de acuerdo a datos del Banco Mundial para el 2015, se destina aproximadamente 0.55% del Producto Interno Bruto a la ciencia (investigación y desarrollo); lejos de ser un tema de interés, es más alarmante que para el año 2018, se asignaron al Consejo Nacional de Ciencia y Tecnología 31 mil 091 millones 273 mil pesos, destinándose 18.6% del recurso para investigación, desarrollo y publicaciones, siendo 2.5% menor que en el 2017, todo esto como reflejo de los recortes presupuestales del Gobierno Federal, generando descontento entre la comunidad científica mexicana.

Para los investigadores, una forma de demostrar su productividad es a través de la publicación en revistas indizadas, arbitradas y con factor de impacto. Desafortunadamente, el publicar en revistas de estas características y el tener acceso a su contenido científico tiene un costo (que puede ser de hasta algunos miles de dólares) lo cual crea la necesidad de gastar presupuesto en adquirir publicaciones y en publicar nuestra investigación original o bien, buscar alternativas para la publicación y recuperación de artículos. Una de estas alternativas es la publicación en revistas de acceso abierto (open access), una postura impulsada desde sus inicios por la fundación Bill y Melinda Gates. Las publicaciones de acceso abierto están disponibles de forma gratuita para todos los lectores en cualquier lugar, generalmente mediante acceso a internet, lo cual significa que pueden ser descargadas, distribuidas o impresas por el usuario, el costo de la publicación es cubierto, en general, por el investigador o institución de trabajo o financiado-

ra. Actualmente impera el modelo tradicional en donde es necesario realizar el pago de suscripción a la revista para hacer uso de los servicios editoriales, entre ellos el de las publicaciones académicas. De acuerdo con la revista Nature en su publicación del 04 de septiembre de 2018 y llamada "Radical open-access plan could spell end to journal subscriptions" existen tres modelos de publicaciones académicas, 15% de las revistas publican los trabajos en el modelo de acceso abierto, financiados por costos de publicación o mediante convenios institucionales; un 38% de las revistas publican los trabajos en el modelo tradicional, que significa que para la recuperación del artículo se requiere el pago de cuotas de suscripción y, un 45% de las revistas trabaja en un modelo híbrido. Lo anterior indica la prevalencia del modelo en el cual es necesario cubrir un costo para tener acceso a la información.

Es aquí donde se crean diversos puntos de discusión: ¿Debe de existir una barrera económica para el acceso a la información científica?, ¿la publicación de la ciencia debe de ser capitalizable?, ¿los investigadores deberíamos de pagar por el acceso a la información científica de calidad? Si bien es cierto que las editoriales requieren de cubrir ciertos gastos para realizar la publicación de artículos (ahora en su mayoría de forma electrónica) y esto se convierte en un modelo de negocio. En este año la Comunidad Europea, en conjunto con diversas organizaciones dedicadas al financiamiento de la ciencia, establecieron su postura con respecto a las publicaciones de acceso abierto a través del Plan S (science, speed, solution): se pretende que para el año 2020 todas las publicaciones derivadas del financiamiento público deben de ser publicadas en revistas de acceso abierto. Sin duda, este planteamiento significa una gran razón de "preocupación" para las casas editoriales, como lo expresan los voceros de la Asociación Internacional de Editores Médicos, Científicos y

Técnicos, ya que la propuesta podría significar limitaciones a la libertad académica.

Sin duda es una propuesta controversial, desde la visión de los que hacen ciencia significaría el acceso a la información inmediatamente. Pero es un hecho que mientras no avancen iniciativas como

estas, seguiremos dependiendo del acceso limitado que nos pueden brindar nuestras instituciones.

Jesús Martínez Sámano y
Marco Antonio Juárez Oropeza
Departamento de Bioquímica, Facultad de Medicina
Universidad Nacional Autónoma de México

IMPLICACIONES CLÍNICAS DE LA EXPRESIÓN DE GENES QUE CODIFICAN PARA PROTEÍNAS DE LAS UNIONES ESTRECHAS: EL CASO DE *TJP1**

Eric Ramírez-Bohórquez¹, Diana García-Cruz², Diego Ortega-Pacheco¹, José I. Siliceo-Murrieta³, Lorena Guadalupe Ramón-Canul³, Sergio Alberto Ramírez-García^{3**}

¹Programa de Maestría en Salud Pública, División de Estudios de Posgrado, Universidad de la Sierra Sur, Miahuatlán de Porfirio Díaz, Oaxaca, México.

²Departamento de Biología Molecular y Genómica, Centro Universitario de Ciencias de la Salud, Universidad de Guadalajara.

³Instituto de Investigaciones sobre la Salud Pública de la Universidad de la Sierra Sur, Miahuatlán de Porfirio Díaz, Oaxaca, México.

**Autor de correspondencia correo E: sergio7genetica@hotmail.com

RESUMEN

El gen *TJP1* codifica para la proteína ZO-1, una de las proteínas de las uniones estrechas (UE) intercelulares, las cuales realizan diferentes funciones de acuerdo a su expresión. ZO-1 participa en el contacto célula-célula, proporciona estructura mecánica y regula el transporte de sustancias a través de las células. Su mayor expresión es en las células epiteliales así como endoteliales. Se ha propuesto que la alteración de la estructura, o la disrupción de las *TJP1* a causa de mutaciones o polimorfismos que conducen a cambios en la estructura de ZO-1 o en su expresión, podrían ser un factor patogénico en diferentes enfermedades, ya que modifica la arquitectura y función de los tejidos, como en síndrome de intestino irritable, infecciones por virus del hepatitis C, el virus de inmunodeficiencia humana (VIH), el virus del papiloma humano (VPH), y en las cardiomiopatías. Considerando que en el humano las UE tienen un rol crucial en la fisiología de los tejidos, realizamos una breve revisión de las bases moleculares y las implicaciones del efecto de la expresión del gen *TJP1* y de las UE relacionadas en el desarrollo de diferentes enfermedades del campo de las sub-especialidades de la medicina interna.

ABSTRACT

The *TJP1* gene encodes for the ZO-1 protein, from the intercellular tight junctions (TJP), which perform different functions according to their expression, such as in the cell-cell contact, provides mechanical and structural support, regulates the transport of nutrients through the cells. Its highest expression is in epithelial and endothelial cells. It has been proposed that the alteration of the structure or disruption of the *TJP1* due to mutations or polymorphisms that lead to change in the structure of ZO-1 or its expression, could be a pathogenic factor in different diseases, since it modifies the architecture and tissue function, such as irritable bowel syndrome, infections by hepatitis C virus, HIV, papilloma virus (HPV), and cardiomyopathies. Considering that in humans the *TJP1* gene plays a crucial role in the physiology of tissues; we carry out a brief review of the molecular bases and the implications of the effect of the expression of *TJP1* and related TJP genes, during the development of different diseases.

PALABRAS CLAVE:

Claudina, ocludinas, polimorfismos, uniones estrechas, Zona Occludens.

KEY WORDS:

Claudins, polymorphism, Occludins, Tight junctions, Zona Occludens.

INTRODUCCIÓN

El gen *TJP1* (del inglés, *Tight Junction Protein 1*) codifica para la fosfoproteína de membrana periférica ZO-1 (Zona Occludens-1), el gen tiene su *locus* en el cromosoma 15q13.1 (Fig. 1). Por corte y empalme alternativo se producen cinco isoformas que se expresan en diferentes tejidos, tal como se muestra en la figura 2 (1). ZO-1 se expresa principalmente en uniones estrechas de células epiteliales y endoteliales, preferentemente en los sitios de contacto célula-célula, formando barreras intercelulares que sirven para separar los compartimentos tisulares laterales al limitar el transporte de los solutos, agua, nutrientes entre otras moléculas, así como a las células del sistema inmune y agentes infecciosos para atravesar el espacio paracelular (2).

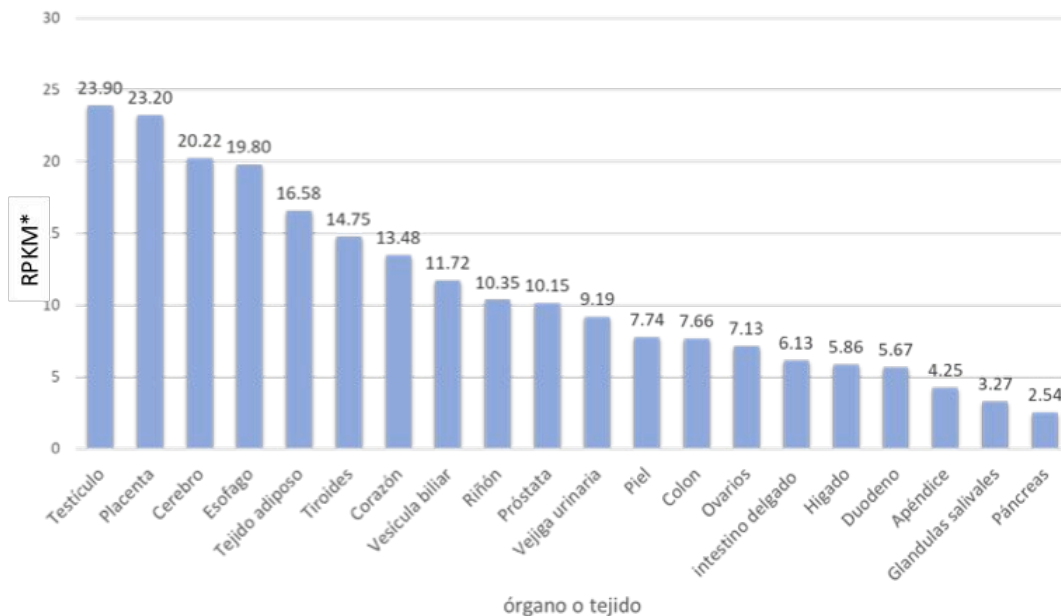
La isoforma más larga de la proteína ZO-1 tiene 7 dominios; cabe señalar que ZO-1 forma parte del complejo de proteínas transmembranales de las uniones estrechas (UE) formado por tres grupos

(2); 1) por proteínas de anclaje periférico (zonulina, zonula ocludens, ZO-1, ZO-2, ZO-3), 2) ocludinas claudinas y 3) moléculas de adhesión (JAM), las cuales brindan resistencia mecánica a la unión célula-célula (2), participan en la señalización intracelular, están vinculadas con la regulación de la expresión genética, diferenciación celular, (2-7). ZO-1, ocludinas y claudinas (3, 6), además se asocian a otras proteínas transmembranales, las cadherinas, las cuales pueden ligar a las UE con los elementos del citoesqueleto y las uniones adherentes (UA) (6), regulando la citocinesis, tráfico de las uniones GAP, quimiotaxis, apoptosis, así como el transporte mediado por vesículas. Se ha reportado que ZO-1 y las claudinas son determinantes de la absorción selectiva de nutrientes en las células epiteliales intestinales (Fig. 3) (1, 8).

La alteración en la estructura o expresión de las UE puede tener diferentes implicaciones clínicas, el presente trabajo se enfoca principalmente a *TJP1*/ZO-1, sin embargo se analizan otras UE, con las que interacciona. Por ejemplo, se han reportado



Figura 1. Locus del gen *TJP1* en el cromosoma 15. Fuente: elaboración propia.



*Datos de kilobases leídas por millón de lecturas mapeadas, bloque de secuencias de RNA

Figura 2. Abundancia de expresión del gen *TJP1* en diferentes órganos y tejidos del cuerpo humano. Fuente: adaptado de Gene, NCBI.

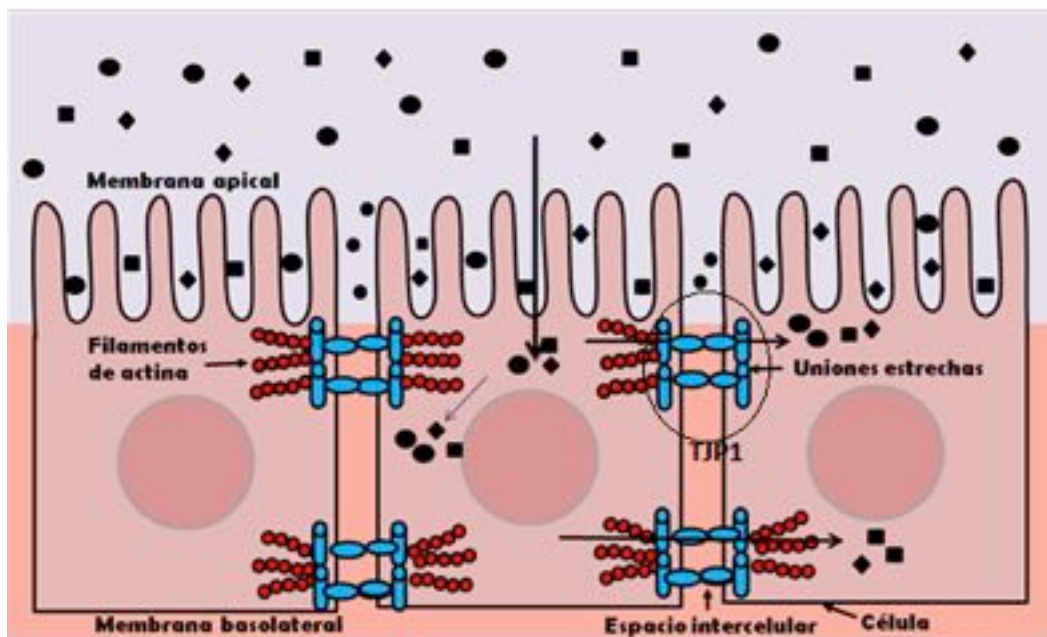


Figura 3. Localización de *TJP1/ZO-1* en los enterocitos, se aprecian en la parte apicolateral y basolateral de las células, de color azul, mientras que los filamentos de actina a los cuales se anclan, están de color rojo. El flujo de nutrientes a través de estas se observa con las flechas negras de la luz al enterocito.
Fuente: elaboración propia.

polimorfismos de un solo nucleótido (SNP, por sus siglas en inglés) que modifican la estructura de las *TJP1/ZO-1*, como el caso del polimorfismo rs2291166 el cual consiste en una transversión T>G en el exón 23 codón 1334, que se traduce en un cambio de aspartato por valina, se asocia con una alteración de la estructura secundaria de *ZO-1*; este polimorfismo se asocia con la albuminuria (3). En la Tabla 1 se muestran algunos de los SNP's reportados con *locus* en sitios crípticos intrónicos, los cuales se relacionan con la expresión

del gen. También se presentan algunos SNP's que conducen cambios de aminoácido que se traducen en modificaciones de la estructura secundaria y terciaria *ZO-1* (1). Así, las patologías de las UE se deben a dos causas, por alteración en la estructura (lo cual se traduce en la reducción o aumento de su función), o por modificación en su expresión (aumento o disminución). Las principales patologías se agrupan en:

- 1) Enfermedades del epitelio respiratorio
- 2) Enfermedades del tracto digestivo

Tabla 1. Polimorfismos de *TJP1* con implicaciones funcionales

Número de SNP	Locus cromosómico	Localización de la variante	Cambio de nucleótido	Consecuencia Funcional
rs785422	15:29881682	Variante Intrónica	C > T	Modificación en la expresión
rs1719013	15:29859586	Variante Intrónica	C > T	Modificación en la expresión
rs2229515	15:29726423	Variante exónica, cambio con sentido	T > C	Modificación de la estructura
rs2291166	15:29716773	Variante exónica, cambio con sentido	T > G	Modificación de la estructura
rs4779674	15:29916646	Variante Intrónica	G > A	Modificación en la expresión
rs17570821	15:29716773	Variante exónica, cambio con sentido	T > G	Modificación de la estructura

- 3) Enfermedades dentales
- 4) Enfermedades neurodegenerativas
- 5) Neoplasias
- 6) Nefropatías
- 7) Enfermedades cardiovasculares
- 8) Enfermedades infecciosas

Enfermedades del epitelio respiratorio. La rinitis alérgica es una enfermedad inflamatoria nasal crónica mediada por un proceso mediado por inmunoglobulina E a alérgenos ambientales. Aunque la atopia es un potente factor de riesgo predisponente para la rinitis alérgica, la susceptibilidad del tejido local es inevitable para la expresión de la enfermedad. El epitelio nasal mantiene la homeostasis tisular proporcionando una barrera física controlada por proteínas de unión epitelial, en este sentido Lee y cols.,(9) analizaron la expresión de proteínas de unión epitelial que no se ha estudiado en pacientes con rinitis alérgica, mediante la reacción en cadena de la polimerasa en tiempo real, Western blot y ensayos de inmunohistoquímica, se analizó la expresión de las proteínas E-cadherina y de ZO-1 en el epitelio del cornete. Además, los niveles de expresión de E-cadherina y ZO-1 se determinaron en células epiteliales cultivadas tratadas con interleucina-4 (IL-4), IL-5, factor de necrosis tumoral-alfa (TNF- α) e interferón gamma (citocinas incrementadas en este padecimiento). Ciertamente la expresión y la inmunorreactividad de E-cadherina y ZO-1 disminuyeron en el epitelio nasal de pacientes con rinitis alérgica. Cabe resaltar que la estimulación de células epiteliales cultivadas con IL-4, IL-5 y TNF- α dio como resultado la regulación negativa de la expresión de E-cadherina solo en células epiteliales cultivadas de pacientes con rinitis alérgica, mientras que la expresión de E-cadherina en células epiteliales cultivadas de los controles no se vieron afectados por la estimulación con el mismo panel de citocinas. Estos resultados son concordantes con la disminución de la expresión de las proteínas de unión epitelial que se encontró en pacientes con rinitis alérgica. La interrupción de la integridad epitelial se relacionó con la regulación corriente debajo de los genes para E-cadherina y ZO-1 (9).

Enfermedades del tracto digestivo. Las UE expresadas en la parte apical de las vellosidades del intestino delgado, tienen un rol crucial en la permeabilidad selectiva, debido a que en la superficie apical se absorben los nutrientes; mientras que a nivel lateral y basolateral, *TJP1*/ZO-1 actúa como regulador de la permeabilidad entre un enterocito y otro (6). Los mecanismos de interacción entre las UE a este nivel, son un campo de estudio vasto, por ejemplo se ha reportado que JAM-A (por su

siglas en inglés, Junctional adhesion molecule-A) no interacciona directamente con ZO-1, sino más bien lo hace a través de ZO-2 (10). Una disminución en la expresión o alteración estructural de estas proteínas, pueden causar un síndrome de mala absorción. Por ejemplo se ha asociado la reducción de la expresión de *TJP1* y otras UE en el síndrome de colon irritable (11-12). En la patogénesis de la enfermedad celiaca, se ha relacionado más con la reducción de la zonulina y con ZO-1 (13). En diabetes mellitus tipo 1 autoinmune (13) se ha reportado un incremento de la permeabilidad de las vellosidades intestinales y diarrea secretora relacionada con la pérdida de la arquitectura de las UE (zonulina y ZO-1) (14). Esto sugiere que los genes de las UE pueden ser factores pronósticos de la severidad de estas dos enfermedades autoinmunes (3, 15-16), así como para otros patógenos que pueden afectar el tracto digestivo o sus glándulas anexas, tal es el caso de *Helicobacter pylori*; en éste, la proteína CagA se trasloca con ZO-1 y JAM-A en las células epiteliales gástricas causando un ensamblaje ectópico de componentes de la unión estrecha en sitios de unión bacteriana y alterando la composición así como función del complejo apical-basolateral, su administración prolongada produce epitelios displásicos (3, 15-16). En estudios *in vitro*, ZO-1 y otras ocludinas se incrementan en el borde de las células Huh7 en presencia de hepatitis C, aumentan aún más en presencia de enterovirus, así ZO-1 y otras UE del tipo JAM-A, regulan la homeostasis, y son necesarias para la estabilización de las integrinas tipo B y para la proliferación de las células epiteliales intestinales (14).

Enfermedades dentales. Recientemente, Xu y cols. (17), han demostrado la alta expresión de ZO-1 en odontoblastos de la pulpa dental y del periodoncio; la alta expresión de ZO-1 incrementa gradualmente con la diferenciación. La diferenciación se inhibía en gran medida en caso de silenciamiento del gen *TJP1*, con lo que queda claro que ZO-1 es necesario para la diferenciación de estas células del tejido conectivo (17). Las implicaciones clínicas de estos hallazgos son importantes, porque la expresión genética anormal de *TJP1* o sus variantes génicas son factores de predisposición para pulpitis, necrosis pulpar y anomalías del desarrollo embrionario (agenesias dentales), un nueva frontera de investigación en la odontología.

Enfermedades neurodegenerativas. Se conoce que la presencia de las UE en la barrera hematoencefálica, destacando la familia de las claudinas, ocludinas y ZO-1 (18). Adicional a lo antes mencionado se ha encontrado que se expresan en células madre que dan origen a estirpes neuronales (19). Cabe señalar que, cuando se afecta la estructura

de las UE o la expresión de sus proteínas, como algún polimorfismo o mutación, las funciones que realizan se verán afectadas; por ejemplo la arquitectura y permeabilidad de las células endoteliales de la barrera hematoencefálica se mantiene por *TJP1*(ZO-1) (18); los cambios estructurales en esta proteína se han reportado en enfermedades neurodegenerativas como en la enfermedad de Alzheimer, la enfermedad de Parkinson, el edema vasogénico, la esclerosis múltiple, lo cual es una de las fronteras de investigación (9). Algunos polimorfismos como el rs1038306187 que conduce al cambio del aminoácido Gln791Glu en el dominio guanilato cinasa (residuo de localización 682-861), o que afectan a la expresión génica, mediante cambios de nucleótidos en los sitios crípticos intrónicos del corte y empalme alternativo, tal es el caso de los SNP's rs781148827, rs5478300017, rs78014403 y rs1020739943 en la región aceptora, y rs1029122894 en la región donante (9). La región del gen *TJP1* que codifica para el dominio ZU5 (residuos 1701-1799), es un "hotspot" con 33 polimorfismos que conducen a cambios conformacionales (23). En la Tabla 1 se puede ver otros polimorfismos del gen *TJP1* con implicaciones funcionales.

Neoplasias. *TJP1*/ZO-1 es uno de los principales blancos en el cáncer (2-4), primero porque forma parte de los contactos célula-célula que participan en la diferenciación celular. Segundo, porque la alteración en la arquitectura de ZO-1 tiene un efecto importante en la tumorigénesis, se presentan dos efectos: a) por disminución de la expresión y, b) por incremento de la expresión; en el primer caso están los colangiocarcinomas, tumores de la vesícula biliar y tumores de mama en fases avanzadas (3). En el segundo grupo, están el adenocarcinoma ductal pancreático y en largas metaplasias ductales (20), cáncer de colon estadio 1-2 y con metástasis al hígado (3). En casos de neoplasias gastrointestinales se ha reportado que el incremento de la expresión de *TJP1* es proporcional con el diámetro del tumor, grado histológico, así como mala supervivencia de los pacientes (3). De la misma manera, se ha asociado la disminución de los niveles de *TJP1* con la consecuente reducción en los niveles de ZO-1 al pronóstico poco favorable de supervivencia de pacientes con carcinoma hepatocelular después de una hepatectomía (21). Mientras que en el cáncer pulmonar de células no pequeñas, la alta expresión de este gen, es un indicador de buen pronóstico (3). Por otra parte, *TJP1* se ha postulado como un posible marcador para cáncer medular de tiroides, esto debido a que la pérdida del contacto célula-célula, es un factor sobresaliente en la transformación maligna (20). En tumores metastásicos se pierde la adhesión celular de UE, incluyendo ZO-1, de este

modo se pueden invadir los elementos estromales circundantes y continuar con los eventos posteriores de la metástasis. Se ha encontrado evidencia de la mayor permeabilidad por UE asociado a mutaciones oncogénicas (4); las mutaciones en oncogenes RAS de la familia de GTPasas, ya que se asocian con la reducción en los niveles de claudinas 1 y 2 (22), no así para ZO-1, y esto como consecuencia influye en la progresión de la tumorigénesis (4); una frontera de investigación en cáncer es la estabilidad de las UE, si se pudiera mantener la adhesión celular de las UE, podría significar un mejor pronóstico ante los tumores en etapas avanzadas, aunque esto aplicaría solo para tumores de origen epitelial.

Nefropatías. *TJP1* también es un candidato en nefropatía diabética ya que ZO-1 desempeña una función importante en la barrera de filtración glomerular al mantener los diafragmas entre los podocitos, lo que regula el paso de diferentes moléculas como la albumina. Cuando es alterada la arquitectura de la barrera de filtración se produce albuminuria. En este sentido, Lehman y cols. (24) reportaron que la albuminuria en mexico-americanos se asoció con polimorfismos de *TJP1* como el rs2291166, rs2035326, rs17671556, SNP e28.3, los cuales conducen a cambios estructurales de *TJP1*/ZO-1 (24).

Enfermedades cardiovasculares. Las células endoteliales se encuentran en la superficie interna de los vasos sanguíneos y linfáticos desempeñando un papel importante en su formación, así como en la función de los mismos. La proteína ZO-1 de las UE, regula la tensión sobre las uniones adherentes (UA) basadas en VE-cadherinas, que realizan la tensión célula-célula y así también regulan la migración celular, y la angiogénesis *in vitro* e *in vivo*, además son esenciales en la permeabilidad de la barrera endotelial (7). Se ha reportado que en células endoteliales, los ácidos grasos lisados por la lipoproteína lipasa, incrementan la permeabilidad de ZO-1, y que la hipercolesterolemia produce alteración en la vía de la PI3K (Fosfatidilinositol-3-cinasa). Recientemente se reportó que las proteínas ZO-1, claudina-1, ocludina-1 y las de unión comunicante (GJP) (Cx43 Cx45 y Cx46) juegan un papel importante en la morfogénesis del corazón y de los vasos sanguíneos. En la insuficiencia cardíaca por cardiomiopatía isquémica o dilatada, también existe una marcada disminución de ZO-1 y de la conexina 43 (3, 25-26). También se ha reportado que en ratas expuestas a una dieta alta en grasas, un aumento significativo de la formación de ateromas mostraron que la pared de la arteria estaba infiltrada con tejido adiposo, el tejido coronario y el tejido endotelial arterial mostraron que Cx43/45/46 y claudina-1 estaban

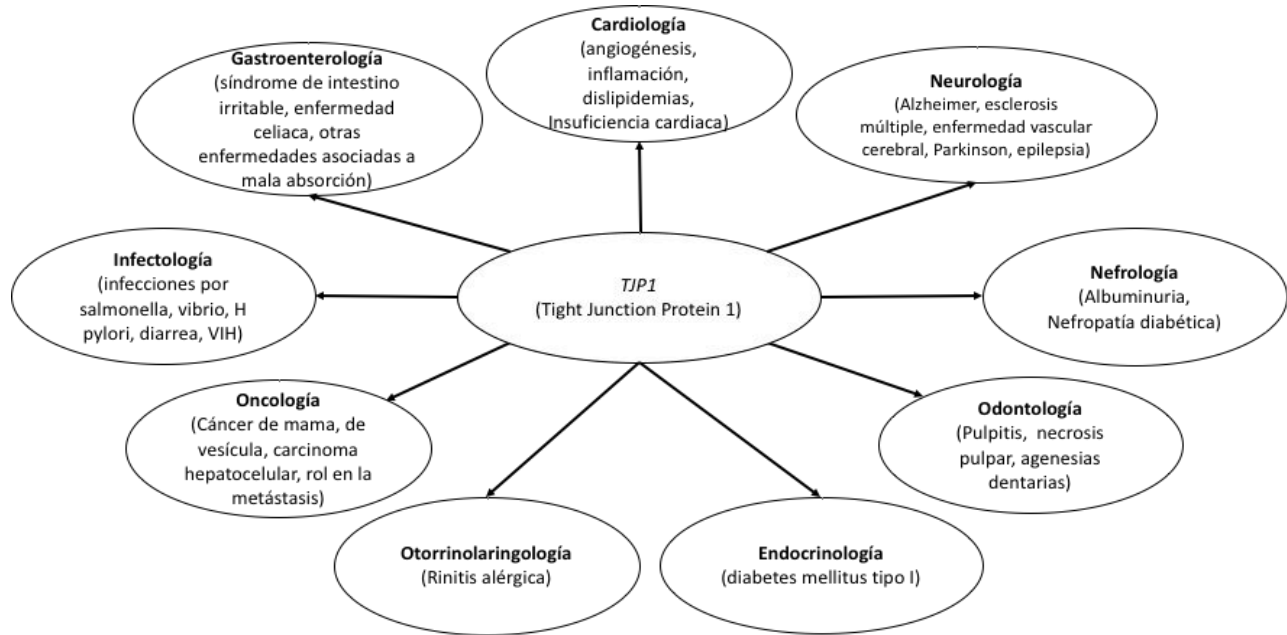


Figura 4. Patologías causadas por defectos en TJP1 y sus proteínas relacionadas. Fuente: elaboración propia.

regulados negativamente. En la arteria coronaria las células endoteliales mostraron una mayor permeabilidad, con menor expresión de Cx43 y ZO-1 y más penetración de monocitos CD14 en corazón y aorta, estos hallazgos sugieren que la formación de ateroma podría deberse a la pérdida de las proteínas TJP1 y GJP (3,25-26). Por ello recientemente se ha postulado que los polimorfismos funcionales de TJP1 podrían influenciar la permeabilidad vascular y la respuesta al tratamiento con estatinas. Estas uniones intercelulares también pueden ser útiles en el pronóstico y la estimación del riesgo cardiovascular en pacientes con enfermedad vascular (3, 25-26).

Enfermedades infecciosas. Dentro de las funciones de la UE se encuentran el regular el transporte de iones, nutrientes y patógenos, teniendo en cuenta esto último las UE pueden estar asociadas a algunas patologías infecciosas, destacan el VIH ya que interactúa con las proteínas específicas para éste (1); las infecciones por virus coxsackie y adenovirus, en las cuales mientras se mantenga normal la estructura de las UE la infección podrá contenerse. A nivel del tejido adiposo es importante porque las ZO-1 pueden facilitar adhesión del adenovirus 36, un factor que se asocia con el desarrollo de obesidad. También en la infección por virus del papiloma humano (VPH) se ha asociado la expresión de algunas proteínas de dicho virus con la disrupción de las proteínas ZO-1 ocasionando mayor

expresión del pseudovirón del virus para infectar las células basales y parabasales, que son en las que inicia el ciclo de vida del VPH (3). La infección por *Salmonella* y *Vibrio cholerae* en las células epiteliales intestinales provoca la disrupción de ZO-1 (1,3), favoreciendo la diarrea secretora.

Por todas las implicaciones que se mencionan anteriormente, se propone que el gen TJP1 el cual codifica para la proteína ZO-1 de las UE, puede ser usado como un biomarcador, lo cual es ilustrado en la Figura 4. Puede ser a nivel de polimorfismo, a nivel de expresión del gen, a nivel de proteína (ZO-1), en farmacología es una de las nuevas fronteras en la respuesta a los antipsicóticos (26).

Conclusión

El presente trabajo nos deja ver que las UE son importantes para la arquitectura tisular y transporte paracelular de los nutrientes, particularmente para ZO-1 y proteínas relacionadas con ella. Cuando se altera su estructura o expresión pueden ocurrir seis tipos de enfermedades: Enfermedades del epitelio respiratorio, del tracto digestivo, de los dientes, neurodegenerativas, cardiovasculares e infecciosas, pero es fundamental que se continúe explorando a fondo el papel que juegan el gen TJP1 y las UE en estas enfermedades, ciertamente las evidencias muestran que el gen TJP1 es un gen "housekeeping", el cual tiene un efecto pleiotrópico, que es caracterizado por los efectos fenotípicos distintos

y no relacionados en diferentes tejidos, por lo cual los polimorfismos de este gen son candidatos a biomarcadores para la práctica clínica

Declaración de conflicto de interés. Los autores declaran que no existe conflicto de interés. 

REFERENCIAS

- Gene [Internet]. Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information (NCBI); 2004 – [2017 02 15]. Disponible en: <https://www.ncbi.nlm.nih.gov/gene/7082>
- Jiménez-Salazar JE, González-Núñez L, Königsberg-Fainstein M, Gómez-Quiroz LE, Zentella-Dehesa A, Damián-Matsumura P (2012) Estructura y función de las uniones estrechas en la transición Epitelio-mesénquima (TEM) y la tumorigénesis del cáncer de mama humano. *REB* 31: 49-59.
- Ramírez-García SA, Flores-Alvarado LJ, Topete-González LR, Charles-Niño C, Mazariegos-Rubí M, Dávalos-Rodríguez NO (2015) Alta frecuencia de alelo ancestral del polimorfismo rs2291162 de *tjp1* en población mexicana, efecto conformacional, así como las aplicaciones en cirugía y medicina. *Cir Cir* 84: 28-36.
- Martin TA, Mason MD, Jiang WG (2011) Tight junctions in cancer metastasis. *Front in Biosci* 16: 898-936.
- Runkle EA, Mu D (2013) Tight Junction Proteins: From Barrier to Tumorigenesis. *Cancer Lett* 337:41-48.
- Hartsock A, Nelson WJ (2009) Adherens and Tight Junctions: Structure, Function and Connections to the Actin Cytoskeleton. *Biochim Biophys Acta* 1778: 660–669.
- Tornavaca O, Chia M, Dufton N, Osuna-Almagro L, Conway DE, Randi AM, Schwartz MA, Matter K, Balda MS (2015) ZO-1 controls endothelial adherens junctions, cell-cell tension, angiogenesis, and barrier formation. *J Cell Biol* 208: 821–838.
- Anderson JM, Van Itallie CM (2009) Physiology and Function of the Tight Junction. *Cold Spring Harb Perspect Biol* 1: 1-16.
- Lee H, Kim BSB, Im NR, Lee BSD, Kim HK, Lee SH, Lee HM, Lee SH, Baek SK, Kim TH (2016) Decreased expression of *E-cadherin* and *ZO-1* in the nasal mucosa of patients with allergic rhinitis: Altered regulation of E-cadherin by IL-4, IL-5, and TNF-alpha. *Am J Rhinol Allergy* 30: 173-178.
- Monteiro AC, Sumagina R, Rankina CR, Leonia G, Minab MJ, Reiterc DM, Stehlec T, Dermodyd TS, Schaeferg SA, Hallg RA, Nusrata A, Parkosa CA (2013) JAM-A associates with ZO-2, afadin, and PDZGEF1 to activate Rap2c and regulate epithelial barrier function. *Mol Biol Cell* 24: 2849-2860.
- Camilleri M, Carlson P, Valentin N, Acosta A, O'Neill J, Eckert D, Dyer R, Na J, Klee EW, Murray JA (2016) Pilot study of small bowel mucosal gene expression in patients with irritable bowel syndrome with diarrhea. *Am J Physiol Gastrointest Liver Physiol* 311: 365-376.
- Yu F, Huang S, Zhou F, Luo Q, Xie X, Zheng C (2014) Correlation between DQB1 genetic polymorphism and genetic susceptibility in patients diagnosed with irritable bowel syndrome with diarrhea. *Genetics and Mol Res* 13: 10285-10293.
- Fasano A (2011) Zonulin and Its Regulation of Intestinal Barrier Function: The Biological Door to Inflammation, Autoimmunity, and Cancer. *Physiol Rev* 91: 151–175.
- Vorobjova T, Raikkerus H, Kadaja L, Talja I, Uibo O, Heilman K, Uibo R (2017) Circulating Zonulin Correlates with Density of Enteroviruses and Tolerogenic Dendritic Cells in the Small Bowel Mucosa of Celiac Disease Patients. *Dig Dis Sci* 62: 358-371.
- Laukoetter MG, Nava P, Lee WY, Severson EA, Capaldo CT, Babbin BA, Williams IR, Koval M, Peatman E, Campbell JA, Dermody TS, Nusrat A, Parkos CA (2007) JAM-A regulates permeability and inflammation in the intestine in vivo. *J Exp Med* 204: 3067-3076.
- Nava P, Capaldo CT, Koch1 S, Kolegraff K, Rankin CR, Farkas AE, Feasel ME, Li L, Addis C, Parkos CA, Nusrat A (2011) JAM-A regulates epithelial proliferation through Akt/ β -catenin signalling. *EMBO reports* 12: 314-320.
- Xu J, Shao M, Pan H, Wang H, Cheng L, Yang H, Hu T (2016) Novel role of zonula occludens-1: A tight junction protein closely associated with the odontoblast differentiation of human dental pulp cells. *Cell Biol Int* 40: 787–795.

74. Ramírez-Bohórquez E, García-Cruz D, Ortega-Pacheco D, Siliceo-Murrieta JI, Ramón-Canul LG, Ramirez-Garcia SG
18. Mohammad MT (2016) Overproduction of nitric oxide intensifies brain infarction and cerebrovascular damage through reduction of *claudin-5* and *ZO-1* expression in striatum of ischemic brain. *Pathol Res Prac* 212(11): 959-964.
 19. Watters AK, Rom S, Hill JD, Dematatis MK, Zhou Y, Merkel SF, Andrews AM, Cena J, Potula R, Skuba A, Son Y, Persidsky Y, Ramirez SH (2015) Identification and Dynamic Regulation of Tight Junction Protein Expression in Human Neural Stem Cells. *Stem Cells Dev* 24: 1377-1389.
 20. Madrigal-Ruiz P, Dávalos-Rodríguez NO, Ramirez-Garcia SA, Topete-González LR, Mazariegos-Rubí M, Flores-Alvarado LJ, Ruiz-Mejia R, Aguilar-Aldrete ME, Magallanes-Ordoñez JJ, Ramón-Canul LG, Siliceo-Murrieta JI, Zurita-Vásquez GG, Cruz-Bastida JM (2015) Polimorfismo P.D1134A del gen *TJP1* en población con ancestría zapoteca; potencial marcador para cáncer de tiroides y otras neoplasias. *Revista Médica MD* 7: 21-26.
 21. Nagai T, Arao T, Nishio K, Matsumoto K, Hagiwara S, Sakurai T, Minami Y, Ida H, Ueshima K, Nishida N, Sakai K, Saijo N, Kudo K, Kaneda H, Tamura D, Aomatsu K, Kimura H, Fujita Y, Haji S, Kudo M (2016) Impact of *Tight Junction Protein ZO-1* and *TWIST* Expression on Postoperative Survival of Patients with Hepatocellular Carcinoma. *Dig Dis* 34: 702-707.
 22. Mullin JM, Leatherman JM, Valenzano MC, Rendon-Huerta E, Verrechio J, Smith DM, Snetselaar K, Liu M, Francis MK, Sell C (2005) Ras Mutation Impairs Epithelial Barrier Function to a Wide Range of Nonelectrolytes. *Mol Biol Cell* 16: 5538-5550.
 23. Aguilar- Aldrete ME, Madrigal-Ruíz PM, Flores-Alvarado LJ, Ruíz-Mejia R, Parada-Luna F, Ramirez-Garcia, SA (2017) Tight Junction Protein 1 Gene in Neurodegenerative Disease, *New Frontier. MOJ Gerontology & Geriatrics* 3: 84-85.
 24. Lehman DM, Leach RJ, Johnson-Pais T, Hamlington J, Fourler S, Almasy L, Duggirala R, Stern MP, Abboud HD (2006) Evaluation of *Tight Junction protein 1 Encoding Zona Occludens 1* as candidate Gene for Albuminuria in a Mexican American Population. *Exp Clin Endocrinol Diabetes* 114: 432-437.
 25. Madrigal PM, Aguilar ME, Flores LJ, Madrigal-Ruiz D, Ruiz-Mejía R, Madrigal D, Ramirez SA. (2018) TJP1 is a New Gene Target in Personalized Medicine. *Adv Biochem Biotechcnol: ABIO*-159.
 26. Sun B, Chen Z, Gu J, TseG, Jiang J, Huang F, Zhao C. (2016) Tight junction proteins and gap junction proteins play important roles in high fat dietary atherosclerosis pathogenesis. *Int J Clin Exp Pathol* 9: 7969-7976.

ESTUDIO DEL TRANSCRIPTOMA MEDIANTE RNA-SEQ CON ÉNFASIS EN LAS ESPECIES VEGETALES NO MODELO*

Gustavo Rodríguez-Alonso¹, Svetlana Shishkova^{1,2}

¹Departamento de Biología Molecular de Plantas, Instituto de Biotecnología, Universidad Nacional Autónoma de México. Avenida Universidad 2001, Colonia Chamilpa, Cuernavaca, Morelos 62210, México

²Autor de correspondencia correo E: sveta@ibt.unam.mx

RESUMEN

Históricamente el desarrollo de la biología se ha limitado al estudio de unas pocas especies, a las cuales conocemos como especies modelo, y cuyas particularidades representan ventajas prácticas para su mantenimiento y estudio en el laboratorio. Sin embargo, la viabilidad actual para secuenciar el genoma y los transcriptomas de virtualmente cualquier especie permite la inclusión de nuevos organismos como modelo de estudio. En esta revisión se presenta una descripción general de las principales plataformas de secuenciación de transcriptomas (RNA-seq), así como los pasos básicos para el ensamblaje de transcriptomas cuando no se cuenta con un genoma de referencia. Finalmente, se proveen algunos ejemplos de estudios de transcriptoma aplicados a organismos no modelo.

ABSTRACT

The advance of biology has largely relied on the detailed study of a small number of species collectively known as model organisms. These species were chosen as experimental models due to their peculiarities, which represent practical advantages for their study in laboratory conditions. However, the improvement of sequencing technologies allows nowadays the genome and transcriptome sequencing of virtually any organism without the need of a reference genome, and therefore, new species can be used now to study biological processes that were previously inaccessible from the study of classic model organisms. In this review, we present a general description of the main sequencing technologies that can be used for transcriptome sequencing (RNA-seq) and provide an overview of transcriptome assembly and analysis when no reference genome is available. Finally, some examples of transcriptome analysis applied to non-model organisms are provided.

INTRODUCCIÓN

El desarrollo de la biología como ciencia experimental se ha valido del uso de distintos organismos, los cuales, debido a diversas particularidades, se utilizan rutinariamente como objeto de investigación. En 1929 August Krogh pronosticó "para un gran número de problemas, habrá un animal de elección, o unos pocos animales, a los cuales convenga estudiar" (1). A los organismos o especies elegidos para su estudio detallado los conocemos como organismos modelo. Los organismos modelo, que pueden ser procarióticos o eucarióticos, se es-

tudian de manera extensa para generar inferencias y generalidades sobre diversos procesos biológicos. Krogh sugería que la adopción de una especie eucariótica como modelo de estudio debía obedecer a la idoneidad de ésta para el problema a estudiar. Más adelante, en 1975, Hans Krebs expandió la idea de Krogh al añadir que los organismos modelo deben tener un tamaño adecuado para su trabajo en el laboratorio, así como un arreglo anatómico tal que facilite el trabajo experimental (2). Actualmente, la selección de organismos modelo obedece a razones prácticas para su manejo o estudio (3); por ejemplo, las especies eucarióticas que se eligen

PALABRAS

CLAVE:

Transcriptoma, RNA-seq, Organismos no modelo, Illumina, Secuenciación masiva.

KEY WORDS:

Transcriptome, RNA-seq, Non-model organism, Illumina, Next Generation Sequencing.

como modelo de estudio se pueden mantener y propagar fácilmente en el laboratorio, tienen ciclos generacionales cortos y alta fecundidad, son especies diploides, lo cual facilita la obtención y el análisis de mutantes; en muchos casos tienen genoma nuclear pequeño, y existen protocolos establecidos para su manipulación genética.

En el caso particular de las plantas, *Arabidopsis thaliana* fue utilizada como modelo de estudio por Friedrich Laibach desde inicios de 1900 (4), sin embargo, su adopción masiva como organismo modelo sucedió hasta principios de 1980, cuando se generalizó su uso al reconocer las ventajas de esta especie para los estudios genéticos (4-6). El estudio de otras especies vegetales, tales como el arroz (*Oryza sativa*), el maíz (*Zea mays*), o la soya (*Glycine max*) ha recibido atención y financiamiento debido a la relevancia económica de su cultivo, pese a que las características genómicas de estas especies distan de las características genómicas ideales de un organismo modelo; por ejemplo, las especies mencionadas son recalcitrantes a la transformación genética; es decir, la obtención de plantas transgénicas de estas especies resulta complicada.

Aunque la adopción de organismos modelo para la investigación ha proporcionado información valiosa acerca de procesos celulares y moleculares fundamentales, las preguntas que pueden ser abordadas con organismos modelo se encuentran inherentemente limitadas por la información que puede recabarse a partir de ellos (7, 8). Las especies modelo representan un porcentaje insignificante entre las 11 millones de especies, aproximadamente, que se estima que existen en el planeta (9). Adicionalmente, no es posible estudiar todos los procesos biológicos a partir del análisis de las especies modelo; por ejemplo, *A. thaliana* no establece interacciones simbióticas con hongos micorrízicos (10) ni con bacterias fijadoras de nitrógeno (11), las cuales son de relevancia agro-económica y biogeoquímica (12, 13). Debido a esta limitante es necesaria la inclusión de especies no modelo en la actividad científica, es decir, de especies poco estudiadas y que no necesariamente cumplen con todas las características de los organismos modelo, pero que representan sistemas adecuados para el estudio de procesos biológicos inaccesibles mediante el estudio de los organismos modelo actuales (14,15). La adopción de organismos no modelo es posible, en gran medida, gracias a que los costos de secuenciación han disminuido considerablemente, por lo cual es viable secuenciar el genoma y los transcriptomas de virtualmente cualquier organismo pluricelular o de microorganismos cultivables *in vitro*. Adicionalmente, el desarrollo tecnológico y

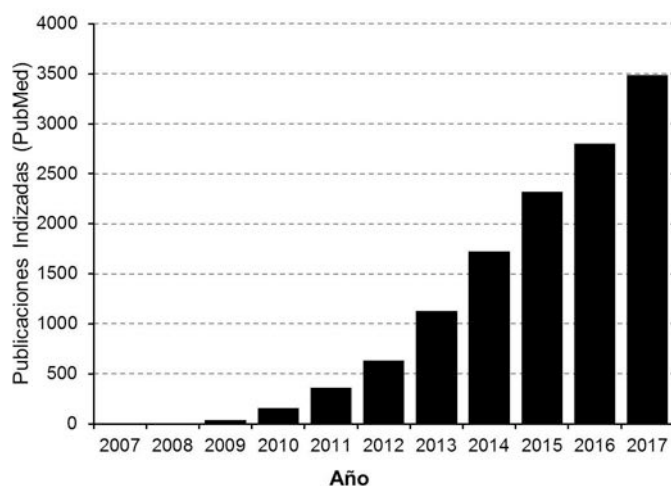


Figura 1. Aumento en el número de publicaciones que utilizan RNA-seq y que se encuentran indizadas por PubMed. La búsqueda se realizó al restringir a las publicaciones a estudios de expresión génica que incluyen "RNA-seq" en el título o el resumen.

el aumento de poder de cómputo hacen factible el análisis de transcriptomas en ordenadores personales. Una muestra clara de la adopción y relevancia de la secuenciación como herramienta de estudio es el número creciente de genomas secuenciados depositados en el NCBI (16), cifra que hasta febrero de 2018 asciende a más de 5,000 genomas de eucariontes y aproximadamente 127,000 genomas de procariontes; así como el número de publicaciones indizadas por PubMed sobre secuenciación masiva de cDNA sintetizado a partir de RNA (RNA-seq; Fig. 1). El RNA-seq es una herramienta que permite evaluar el estado transcripcional de un organismo, órgano o tejido, típicamente para comparar distintos tratamientos, condiciones, o estadios de desarrollo (17, 18). A diferencia de otros métodos de transcriptómica que se basan en la hibridación de un conjunto de moléculas de RNA marcadas, tales como los microarreglos, el RNA-seq es un método cuantitativo que no requiere del conocimiento *ab initio* de las secuencias de los RNA mensajeros. Adicionalmente, el RNA-seq permite detectar la transcripción de genes que se expresan a niveles bajos, además de polimorfismos de una sola base e isoformas, es decir, variantes de transcritos que se obtienen mediante el procesamiento diferencial de pre mensajeros provenientes del mismo gen (19).

Generalidades y conceptos básicos del RNA-seq

En general, los estudios que utilizan al RNA-seq siguen una serie de pasos comunes (Fig. 2), los cuales inician con la extracción de RNA total a

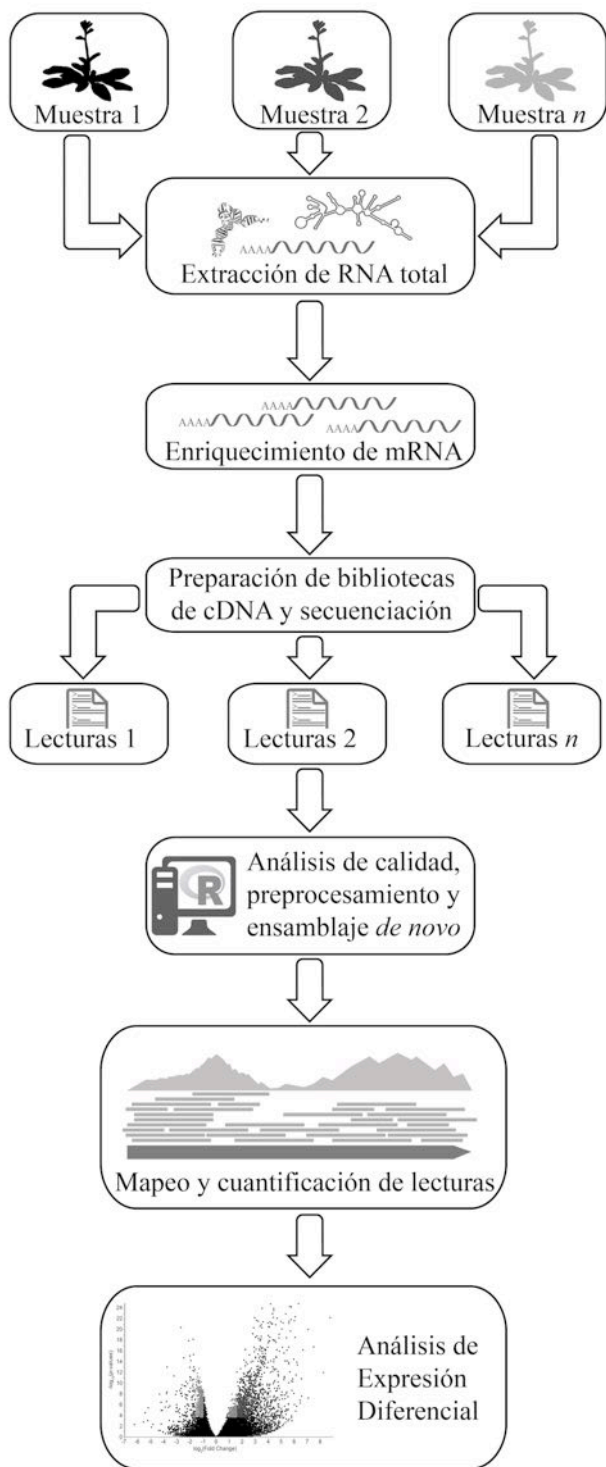


Figura 1. Diagrama de flujo de un experimento que emplea la secuenciación del transcriptoma mediante RNA-seq. Las muestras se procesan para extraer el mRNA, a partir del cual se prepararán y secuenciarán las bibliotecas de cDNA en alguna de las plataformas de secuenciación disponibles. Las lecturas que se obtienen se procesan, y en el caso de las especies sin genoma de referencia, se ensamblan *de novo* para reconstruir la secuencia completa de los transcritos y estimar sus niveles de abundancia en cada condición. A partir de los niveles de abundancia es posible realizar análisis de expresión diferencial y proceder hacia los objetivos particulares de cada estudio.

partir de las muestras biológicas de interés. La extracción se realiza mediante el método de conveniencia para cada especie. En el caso de las plantas, por ejemplo, el consorcio OneKP pone a disposición una serie de protocolos estandarizados para una gran variedad de familias y géneros (20). Dado que en cualquier transcriptoma el RNA ribosomal constituye a la mayor parte del RNA total, existen dos métodos principales de enriquecimiento de RNA mensajero (mRNA), estos son la selección positiva de los mRNA mediante la captura de RNA poliadenilado ("polyA-capture") y el enriquecimiento de mRNA mediante la eliminación de RNA ribosomal ("rRNA-depletion") (21). En general, la técnica más utilizada es la selección de RNAs poliadenilados pues los mRNAs eucarióticos, con muy pocas excepciones, son de este tipo. Sin embargo, el enriquecimiento de mRNA mediante la eliminación de RNA ribosomal es útil en procariontes, cuyos mRNAs no se poliadenilan; y en sistemas eucarióticos en casos muy particulares.

Una vez que el mRNA se enriqueció en la muestra, se procede a la preparación de bibliotecas ("libraries") para secuenciación. El mRNA se utiliza para la síntesis de ambas cadenas de cDNA, cuya secuencia permite deducir la secuencia de nucleótidos del mRNA original. Según la plataforma de secuenciación, las cuales se discutirán más adelante, se puede secuenciar a las moléculas completas de cDNA, o bien, se pueden fragmentar mediante métodos mecánicos y realizar una selección de fragmentos por tamaño, usualmente de entre de 200-1,000 nucleótidos. La secuenciación de los fragmentos cortos es mucho más barata, y debido a ello es la técnica más utilizada. A las secuencias cortas que se obtienen a partir de estos fragmentos de cDNA se les conoce como lecturas ("reads"), las cuales se pueden generar en los formatos de lecturas simples ("single-end") cuando la secuenciación se realiza sólo en alguno de los extremos de cada fragmento, o bien como lecturas apareadas ("paired-ends") cuando a partir del mismo fragmento se secuencian los dos extremos. Ambos formatos de lectura permiten cuantificar los niveles de abundancia de transcritos. Sin embargo, cuando no se cuenta con un genoma de referencia, las lecturas apareadas y más largas conducen a un ensamblaje *de novo* de mejor calidad, esto es, al traslape de lecturas en un mayor número de transcritos probables ("contigs") y de mayor tamaño. Cuando existe un genoma de referencia, las lecturas apareadas permiten

además detectar isoformas o eventos de empalme ("splicing") alternativo de RNAs mensajeros. Una vez que se ensamblaron los contigs, las lecturas se mapean sobre ellos y se cuantifican; al normalizar el número de lecturas mapeadas sobre cada transcrito con base en la longitud de cada contig y el tamaño de la biblioteca de secuenciación, se puede calcular la abundancia relativa de cada transcrito en las muestras biológicas que se secuenciaron. Los análisis posteriores al ensamblaje del transcriptoma estarán determinados por los objetivos particulares de cada estudio, aunque el uso más común del RNA-seq es la detección de expresión diferencial entre muestras biológicas.

Tecnologías de secuenciación más comunes

Actualmente las tres tecnologías principales para la secuenciación de transcriptomas son: la secuenciación por síntesis, la secuenciación por conducción y la secuenciación en tiempo real de molécula única. De éstas, la secuenciación por síntesis, desarrollada por la compañía Solexa y comercializada por la compañía Illumina, es la más utilizada (22). Las bibliotecas para la secuenciación por síntesis se preparan a partir de fragmentos de cDNA, cada uno de los cuales se liga a secuencias adaptadoras (Fig. 3). Los adaptadores fijan a los fragmentos en un arreglo espaciado y ordenado

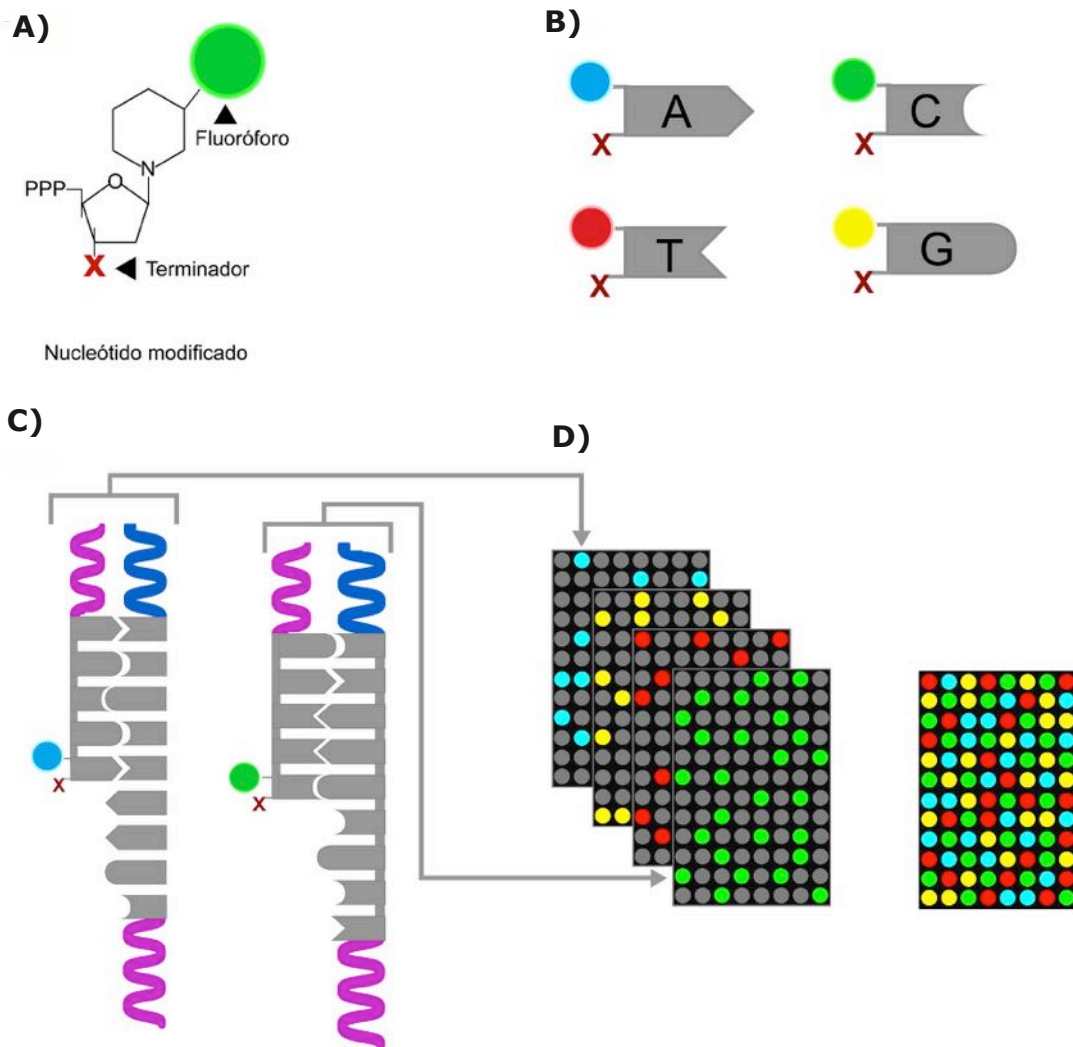


Figura 3. Secuenciación por síntesis. La tecnología de secuenciación por síntesis, desarrollada por la compañía Solexa, que actualmente pertenece a Illumina, utiliza nucleótidos modificados (A) que contienen una marca fluorescente distintiva para cada base (B) y un terminador en la posición 3' del nucleótido. El terminador limita cada ciclo de síntesis a la adición de un solo nucleótido por fragmento a secuenciar (C). Después de que el nucleótido se agregó a la cadena naciente, se capturan cuatro imágenes (D) para identificar al nucleótido que se agregó en cada cluster. Una vez que se resuelven los nucleótidos adicionados en un ciclo de secuenciación en cada cluster, el fluoróforo, así como el bloqueo del sitio 3' se escinden químicamente y se procede a un nuevo ciclo de secuenciación.

sobre una superficie de secuenciación, o chip. Cada fragmento está separado de otros espacialmente y se amplifica mediante una reacción de PCR de modo que se generan grupos discretos ("clusters") de copias de la misma molécula, lo cual permite detectar más fácilmente a las señales provenientes de cada cluster. La secuenciación se realiza mediante la adición consecutiva de 3'-O-azidometil-dNTPs, los cuales son nucleótidos modificados que incluyen un fluoróforo específico para cada base nitrogenada y una modificación química en la posición 3', de modo que la reacción de síntesis se detiene con la adición de cada nucleótido (23). En cada paso de adición de nucleótidos los fluoróforos se excitan mediante un láser y se capturan cuatro imágenes, cada una en la longitud de emisión de un fluoróforo particular; así en cada ciclo de secuenciación se detecta la señal de cada cluster en sólo una de las cuatro imágenes. De este modo se puede identificar, o resolver, al nucleótido que se incorpora en cada cluster durante cada ciclo de secuenciación. Una vez que las imágenes se capturaron, tanto el fluoróforo como el bloqueo del nucleótido modificado se remueven y se puede realizar un nuevo ciclo de síntesis. En teoría, el número de ciclos de secuenciación puede ser ilimitado, sin embargo, la adición de nucleótidos, la remoción de fluoróforos o del bloqueo no siempre es homogénea dentro de un cluster. Como resultado, la certeza de lectura de bases decrece de manera inversamente proporcional al número de ciclos de secuenciación, por lo cual los valores de calidad de las lecturas decaen después de varios ciclos de secuenciación. Actualmente se pueden obtener datos confiables hasta de 300 ciclos de síntesis, es decir, en el formato de lecturas apareadas se pueden secuenciar hasta 600 nucleótidos del mismo fragmento.

Las bibliotecas de secuenciación por semiconducción también se preparan a partir de cDNA, ligado mediante adaptadores a una esfera sobre la cual se realizará la amplificación clonal para generar un cluster (Fig. 4). Sin embargo, esta tecnología de secuenciación tiene como base un principio químico distinto al de la secuenciación por síntesis. Durante la síntesis de ácidos nucleicos, la reacción de ataque nucleofílico para la adición de un nucleótido a la cadena nascente libera a un protón como producto. Este principio se aprovecha en las plataformas de secuenciación por semiconducción (24), desarrolladas por IonTorrent, las cuales consisten en un transistor de efecto de campo sensible a iones que está acoplado a un circuito integrado (25). El chip de secuenciación es un arreglo de micropozos que contienen a las esferas con los clusters de cDNA; el transistor asocia un cambio de voltaje (ΔV) al cambio de pH (ΔpH) generado por la liberación

de protones debido a la adición de nucleótidos. Dado que la liberación de protones es inherente a la reacción de adición de los nucleótidos, éstos no están marcados con fluorescencia ni incluyen un terminador de la reacción de síntesis. La secuencia de los fragmentos se resuelve mediante la inyección de cada tipo de nucleótido de forma secuencial y al observar la presencia o ausencia de cambios de voltaje en cada cluster. En regiones en las cuales existen repeticiones de una sola base se puede estimar el número de nucleótidos repetidos a partir de la magnitud del cambio de voltaje, pues éste es proporcional al número de protones liberados. Actualmente el intervalo dinámico para la detección de repeticiones de una sola base es una limitante técnica de este tipo de secuenciación, sin embargo, al no requerir nucleótidos modificados, los costos de secuenciación son más bajos que en la plataforma Illumina, además, se pueden obtener lecturas de hasta 700 nucleótidos (26). Una ventaja adicional de la secuenciación por semiconducción es que dado que el sistema forma parte de un circuito integrado, las dimensiones de los secuenciadores pueden reducirse hasta dispositivos portátiles, proporcionando así la posibilidad de realizar secuenciación *in situ*.

Por último, la secuenciación en tiempo real de molécula única o SMRT-seq, por sus siglas en inglés (Single Molecule Real Time sequencing; Fig.5), desarrollada por Pacific Biosciences (PacBio), es la tecnología que permite obtener secuencias más largas, con un intervalo de tamaño entre 40 y 60,000 nucleótidos (27). A diferencia de la secuenciación por síntesis o por semiconducción, la secuenciación de molécula única en tiempo real no requiere generar clusters de secuenciación. El fragmento a secuenciar se liga a un adaptador con estructura secundaria de tallo-asa en cada extremo del cDNA (Fig. 5 A) con lo cual, al desnaturalizarse, el fragmento a secuenciar queda circularizado convirtiéndose en una molécula de cadena sencilla. La molécula circular es reconocida por una DNA polimerasa que la utilizará en rondas de síntesis consecutivas e ininterrumpidas. La DNA polimerasa se encuentra fija en el fondo de un nanopozo, formalmente conocido como guía de onda de modo cero, de 70 nm de diámetro y 100 nm de altura. La solución de reacción incluye a los nucleótidos ligados a un fluoróforo en el extremo trifosfatado (28), por lo cual el fluoróforo se escinde tras la formación del enlace fosfodiéster durante la síntesis del DNA. Los nucleótidos marcados, así como los fluoróforos liberados, difunden libremente en la guía de onda. La secuencia se obtiene al excitar a los fluoróforos desde el fondo del pozo con un láser y registrar la emisión de fluorescencia del nucleótido que se está

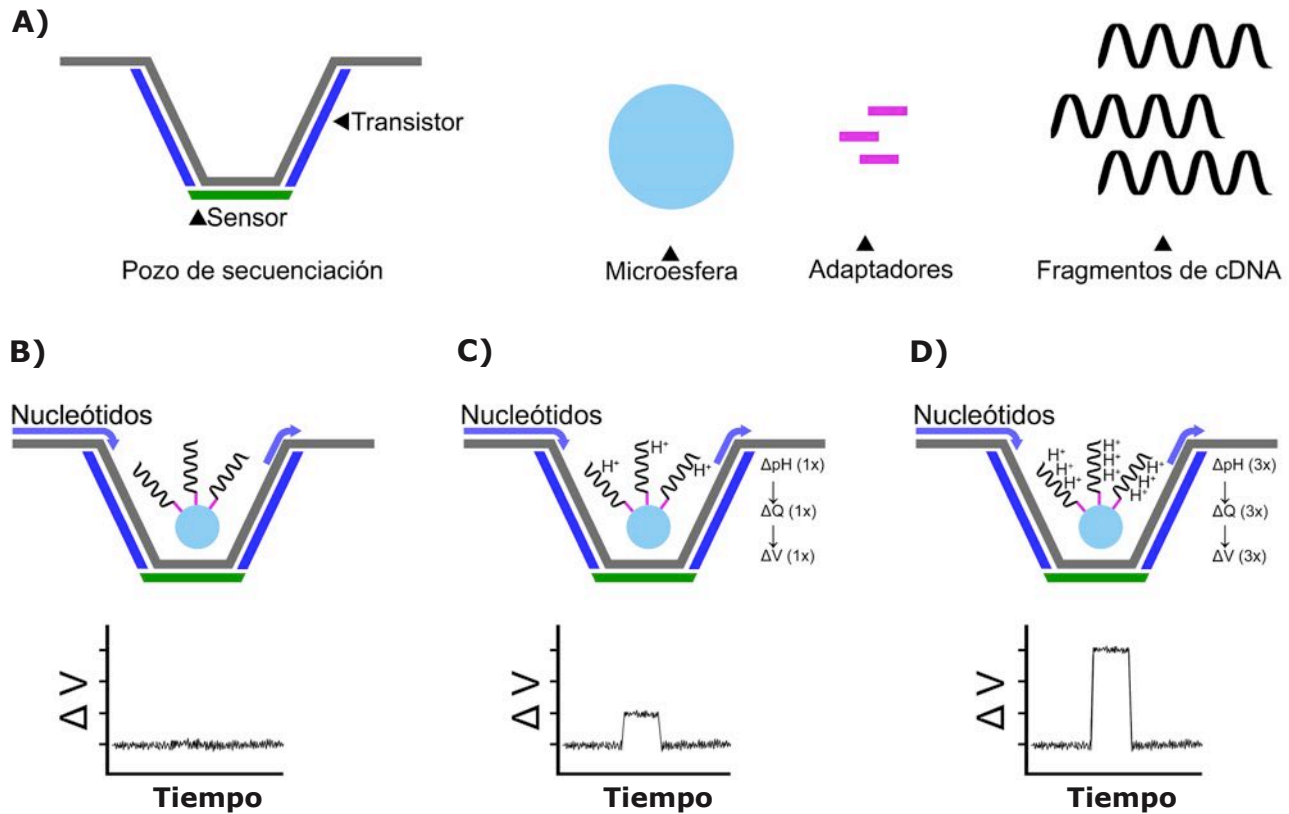


Figura 4. Secuenciación por semiconducción. Esta secuenciación de la compañía IonTorrent, utiliza transistores de campo sensible a iones en los nanopozos de secuenciación (A). La secuenciación ocurre mediante la inyección controlada de cada nucleótido de forma secuencial a los micropozos y al registrar el cambio de voltaje (ΔV) asociado al cambio de pH (ΔpH) debido a la liberación de protones. Si en un pozo durante un ciclo no se agregan nucleótidos a la cadena naciente, no se registra un ΔV (B), a diferencia de aquellos pozos en los cuales un nucleótido es adicionado (C). Cuando en un micropozo a la cadena naciente del DNA se agrega más de un nucleótido debido a repeticiones continuas de la misma base en la secuencia, el ΔV será proporcional al número de nucleótidos adicionados (D). Después de haber registrado la presencia o ausencia del ΔV en un ciclo, se lava a las moléculas del nucleótido que no fueron adicionados y el ciclo puede repetirse mediante la inyección de un nucleótido diferente.

adicionando. Debido a las dimensiones de la guía de onda, el volumen observacional está restringido a 20 zeptolitros (20×10^{-21} L), por lo cual la fluorescencia registrada corresponde al nucleótido contenido en el sitio activo de la polimerasa (29). En general, los resultados que se obtienen mediante SMRT-seq son de buena calidad debido a que cada fragmento se secuencia de manera continua múltiples veces al estar circularizado; lo cual permite deducir la secuencia del mRNA original a partir del consenso que se obtiene al eliminar la secuencia de los adaptadores y alinear a las sublecturas resultantes (Fig. 5D).

Consideraciones para el diseño experimental de estudios que involucran al RNA-seq

Aunque las aplicaciones del RNA-seq son diversas, existen consideraciones básicas que deben

tomarse en cuenta durante la planeación de experimentos de RNA-seq para optimizar la robustez y fiabilidad de los resultados que se derivan a partir de ellos. El proyecto *ENCyclopedia Of DNA Elements* (ENCODE, 30) recomienda incluir al menos dos réplicas biológicas de cada condición, aunque algunos estudios sugieren que la robustez estadística de los resultados se alcanza con al menos seis réplicas biológicas (31). Por su parte, las réplicas técnicas de secuenciación no se requieren pues la variabilidad asociada a las plataformas de secuenciación es mínima (32).

A diferencia del genoma, el transcriptoma es dinámico y difiere entre una condición y otra, o entre distintos órganos. Por ello, y debido a que la calidad del ensamblaje depende de la cantidad y calidad de lecturas obtenidas, se recomienda estimar la profundidad de secuenciación según los objetivos de cada experimento (33). Obtener

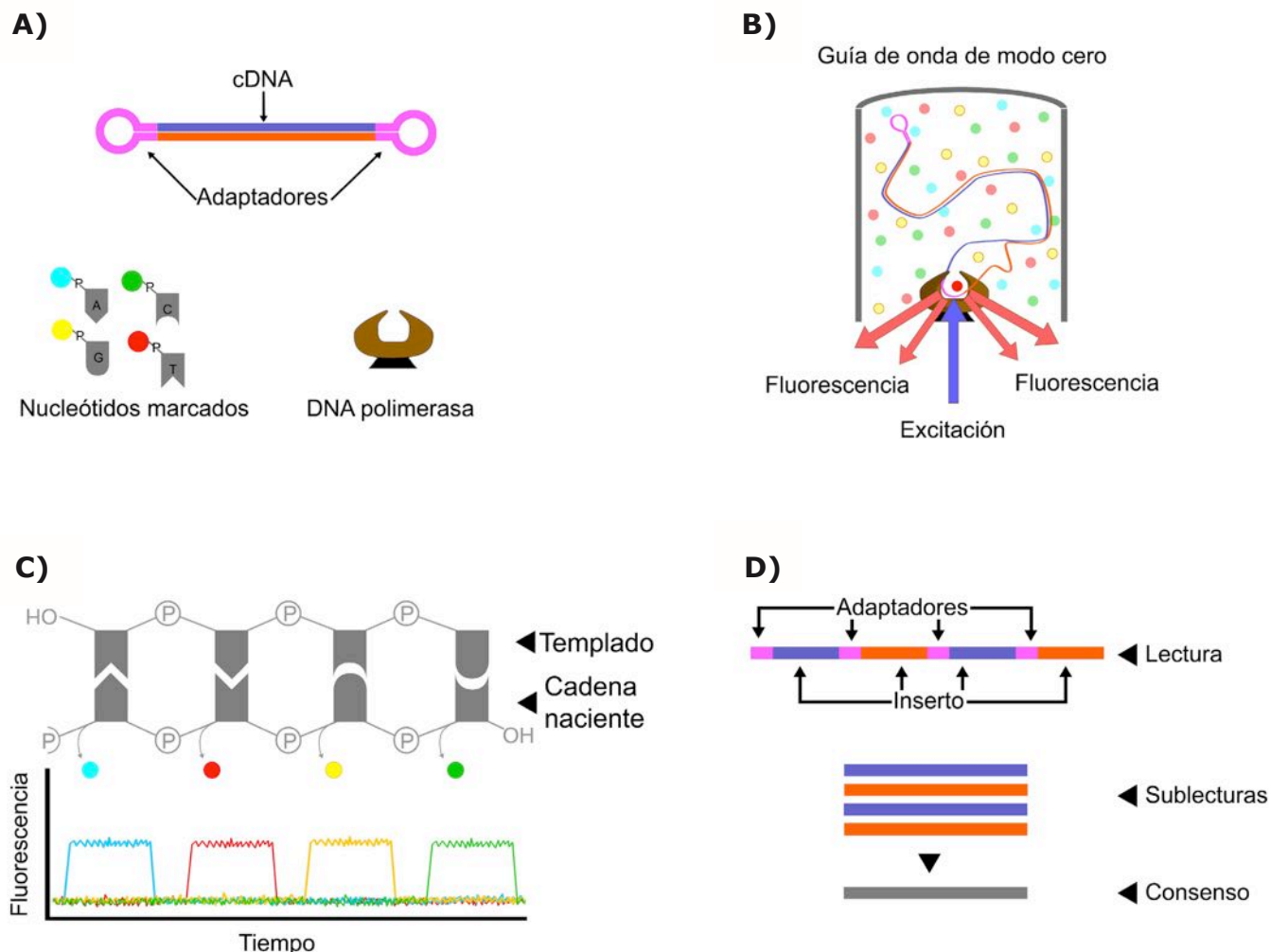


Figura 5. Secuenciación de molécula única en tiempo real (SMRT sequencing). La tecnología de secuenciación SMRT utiliza nucleótidos con marcas fluorescentes en el extremo trifosfatado y adaptadores que circularizan al fragmento de cDNA a secuenciar (A). La secuenciación ocurre en un nanopozo, o guía de onda de modo cero, en el fondo del cual está inmovilizada una DNA polimerasa. La secuencia se obtiene al excitar con láser, desde el fondo del pozo, a los fluoróforos que marcan a cada nucleótido. Debido a que la DNA polimerasa inmoviliza al nucleótido modificado mientras éste se incorpora a la cadena naciente, el fluoróforo de este nucleótido emitirá fluorescencia durante más tiempo que los nucleótidos que difunden libremente en la solución (B y C). Al ser una molécula circular, el mismo fragmento de cDNA se secuenciará varias veces de forma ininterrumpida dentro del nanopozo; dado que las secuencias de los adaptadores se conocen, son fácilmente identificables y pueden ser removidas. Las sublecturas generadas después de la remoción de adaptadores se pueden alinear para obtener una secuencia consenso, la cual corresponderá a la secuencia de cDNA derivado del mRNA original (D).

al menos 200 millones de lecturas mapeables permite detectar transcritos poco abundantes o eventos raros de empalme alternativo (34).

Finalmente, una práctica recomendable pero no del todo extendida es la inclusión de estándares cuantitativos ("spike-in") en las muestras de RNA. Estos estándares son moléculas de RNA exógeno, de secuencia conocida, que se agregan a las muestras en una concentración determinada durante cualquier paso experimental. El ERCC (External RNA Control Consortium) es un

consorcio internacional que provee de 92 RNAs sintéticos con un intervalo de tamaño de 250 a 2500 nucleótidos y del 5 al 51% de contenido de GC (35). El uso de estándares cuantitativos provee de una manera directa para estimar las tasas de error de lectura de bases y los sesgos de cobertura (36, 37). Adicionalmente, es posible contar con controles positivos y negativos en los análisis de expresión diferencial al agregar el spike-in en concentraciones significativamente distintas o iguales, respectivamente.

Ensamblaje *de novo* de transcriptomas de organismos sin genoma de referencia

Las lecturas que se obtienen mediante las plataformas de secuenciación deben ser preprocesadas para identificar y remover a los sitios no informativos. Estos incluyen a las secuencias adaptadoras de la plataforma de secuenciación, bases ambiguas y bases no resueltas. La remoción de sitios no informativos resultará en lecturas de diversos tamaños, a partir de las cuales es necesario descartar también a las lecturas muy cortas, pues éstas son más propensas a ser mapeadas en más de un sitio. Algunas herramientas bioinformáticas útiles y de acceso abierto para el preprocesamiento de las lecturas son FastQC, Trimmomatic y FastX (38-40). En el caso de las lecturas apareadas, es necesario fusionar a las lecturas que provienen del mismo fragmento cuando éstas se superponen en sus extremos (41).

Una vez que las lecturas han sido preprocesadas, se procede al ensamblaje del transcriptoma. Los algoritmos más comunes para el ensamblaje *de novo* se basan en gráficas de de *Bruijn*. Entre los algoritmos disponibles destaca Trinity (42) como el paquete más utilizado debido a su flexibilidad y el manejo eficiente de los recursos computacionales. De manera general, los ensambladores descomponen virtualmente a las lecturas de secuenciación en fragmentos pequeños, conocidos como palabras o k-meros, los cuales se utilizan en rondas de alineamiento para encontrar a los k-meros que se superponen en los extremos 5' o 3' (Fig. 6). De esta manera, a partir de un k-mero inicial ("seed") elegido aleatoriamente, se empieza el ensamblaje de los contigs y se continúa hasta que la secuencia ya no puede ser extendida. Cuando para alguno de los extremos se obtienen dos posibles k-meros que pueden ser agregados, se abre una burbuja ("bubble") y se continúa la extensión del contig en las dos posibles variantes, hasta que esta burbuja se cierra, o colapsa, debido a la existencia de un solo k-mero para ambas posibilidades. Cada burbuja duplica el número de variantes del contig, las cuales pueden representar isoformas, variantes alélicas o errores de secuenciación que resultan en artefactos del ensamblaje. Los tamaños de k-mero y burbuja pueden ser ajustados a conveniencia. Idealmente, deben realizarse varios ensamblajes al variar éstos dos parámetros hasta encontrar al "mejor" ensamblado del transcriptoma, el cual estará definido por la generación de contigs de mayor tamaño, con la mayor proporción de lecturas mapeadas en pares, y con un número bajo de lecturas que mapean a más de un sitio.

Anotación del transcriptoma

Una vez que las lecturas se han fusionado en contigs, es necesario asignar a cada uno de éstos una identidad y función putativa. A este proceso se le conoce como anotación y se basa principalmente en la búsqueda por homología de secuencias caracterizadas y depositadas en bases de datos tales como RefSeq (43). En general, la anotación requiere de la identificación de los marcos abiertos de lectura presentes en cada contig, una tarea que puede realizarse con programas tales como TransDecoder (44). Posteriormente, la secuencia de aminoácidos traducida a partir del contig se utiliza como sonda para interrogar a las bases de datos e identificar a sus secuencias homólogas. Paqueterías como Blast2GO (45) permiten asignar un conjunto de categorías jerárquicas, u ontologías génicas, a partir de la homología entre la proteína predicha para cada contig y las secuencias depositadas en las bases de datos. Otras herramientas disponibles para la anotación de secuencias son InterProScan (46), el cual identifica firmas proteicas; AnnotCrpt (47), el cual permite anotar probables RNAs largos no codificantes; y KAAS (48), el cual mapea a las secuencias proteicas en rutas metabólicas y de señalización.

Actualmente, la anotación de las secuencias de organismos no modelo depende de las secuencias depositadas en bases de datos de acceso público, las cuales a su vez fueron anotadas a partir de organismos modelo. Esto representa una limitante, pues en los organismos no modelo sólo se podrá anotar a aquellas secuencias que contengan a un ortólogo putativo anotado en el genoma de alguna especie modelo. Adicionalmente, las secuencias de RNA que carecen de marco abierto de lectura, tales como los RNAs largos no codificantes o los precursores de microRNAs, serán difíciles de anotar pues presentan tasas de evolución más altas y por lo tanto más divergencia entre especies (49-50). De esta forma, el porcentaje de secuencias no anotadas, así como el número de secuencias con anotaciones de baja calidad, será proporcional a la distancia evolutiva entre el organismo en estudio y el organismo modelo más cercano.

Mapeo de lecturas y evaluación de la expresión diferencial

Uno de los objetivos más comunes de los experimentos de RNA-seq es identificar a los transcritos que se regulan de manera diferencial entre dos condiciones. Existen distintas herramientas gratuitas que permiten mapear lecturas sobre los contigs cuando no se cuenta con un genoma de referen-

A) Descomposición del fragmento en palabras (*k*-meros; *k*=7)

```
GTCGTTATCATTAGTAGTCATCCTTCTTCTATTTCGTCCTCACCCTACTTGCTCACTCTTCTTCTTACATCAGCAGTACACAGATACAGTCCCTTCTTCTATGATGAACGTAGGAGACA
GTCGTTATCATTAGTAGTCATCCTTCTTCTATTTCGTCCTCACCCTACTTGCTCACTCTTCTTCTTACATCAGCAGTACACAGATACAGTCCCTTCTTCTATGATGAACGTAG
TCGTTATCATTAGTAGTCATCCTTCTTCTATTTCGTCCTCACCCTACTTGCTCACTCTTCTTCTTACATCAGCAGTACACAGATACAGTCCCTTCTTCTATGATGAACGTAGG
CGTTATCATTAGTAGTCATCCTTCTTCTATTTCGTCCTCACCCTACTTGCTCACTCTTCTTCTTACATCAGCAGTACACAGATACAGTCCCTTCTTCTATGATGAACGTAGG
GTTATCATTAGTAGTCATCCTTCTTCTATTTCGTCCTCACCCTACTTGCTCACTCTTCTTCTTACATCAGCAGTACACAGATACAGTCCCTTCTTCTATGATGAACGTAGG
TTATCATTAGTAGTCATCCTTCTTCTATTTCGTCCTCACCCTACTTGCTCACTCTTCTTCTTACATCAGCAGTACACAGATACAGTCCCTTCTTCTATGATGAACGTAGGAGA
TATCATTAGTAGTCATCCTTCTTCTATTTCGTCCTCACCCTACTTGCTCACTCTTCTTCTTACATCAGCAGTACACAGATACAGTCCCTTCTTCTATGATGAACGTAGGAGAC
ATCATTAGTAGTCATCCTTCTTCTATTTCGTCCTCACCCTACTTGCTCACTCTTCTTCTTACATCAGCAGTACACAGATACAGTCCCTTCTTCTATGATGAACGTAGGAGACA
```

B) Construcción de la gráfica de de Bruijn a partir de los *k*-meros de todos los fragmentos secuenciados



C) Gráfica de de Bruijn colapsada



D) Secuencias ensambladas (transcritos putativos)

```
GTCGTTATCATTAGTAGTCATCCTTCTTCTATTTCGTCCTCACCCTACTTGCTCACTCTTCTTCTTACATCAGCAGTACACAGATACAGTCCCTTCTTCTATGATGAACGTAGGAGACA
GTCGTTATCATTAGTAGTCATCCTTCTTCTATTTCGTCCTCACCCTACTTGCTCACTCTTCTTCTTACATCAGCAGTACAAAATCCTTCTTAGTATGAACGTAGGAGACA
GTCGTTATCATTAGTAGTCATGCTTCTGTATCGTCTCACCCTACTTGCTCACTCTTCTTCTTACATCAGCAGTACACAGATACAGTCCCTTCTTCTATGATGAACGTAGGAGACA
GTCGTTATCATTAGTAGTCATGCTTCTGTATCGTCTCACCCTACTTGCTCACTCTTCTTCTTACATCAGCAGTACAAAATCCTTCTTAGTATGAACGTAGGAGACA
```

Figura 6. Ensamblaje de novo del transcriptoma. Las lecturas obtenidas, en este caso a partir de la plataforma Illumina/Solexa, se descomponen virtualmente en palabras, o *k*-meros, de un tamaño particular definido por el usuario y ejemplificado aquí con un tamaño *k*=7 (A). Estos *k*-meros se utilizarán para construir alineamientos que permitan extender a la secuencia hacia los extremos 3' y 5'. Cuando existen dos *k*-meros que permiten extender el mismo extremo, se abre una burbuja, la cual colapsa mediante la adición de un *k*-mero común para ambas variantes (B). La gráfica se colapsa al fusionar las regiones sobrelapantes de los *k*-meros (C) y se leen las posibles variantes ensambladas del transcrito (D).

cia, entre ellas destaca RSEM (51). Los algoritmos para mapear lecturas sobre los contigs funcionan, básicamente, a partir de alineamientos globales entre las lecturas procesadas y las secuencias de referencia. Cada algoritmo presenta sus propias ventajas y desventajas respecto a los demás, par-

ticularmente en la calidad de los alineamientos y la versatilidad para evaluar los costos de las diferencias de un solo nucleótido, o *mismatches*, y la apertura y extensión de *gaps*. Un análisis comparativo de los distintos algoritmos para mapeo de lecturas cortas de RNA-seq (52), provee de la información

necesaria para decidir entre los diferentes algoritmos que existen según los criterios que resulten más relevantes en cada experimento. El mapeo de las lecturas sobre las secuencias de referencia, en este caso los transcritos ensamblados *de novo*, es el primer paso para estimar el nivel de expresión de cada uno de los contigs. Esto se logra al mapear y cuantificar, por separado, a las lecturas que se obtuvieron para cada condición y evaluar si existen diferencias significativas entre el número de lecturas que mapean sobre cada contig en las distintas condiciones. Existen diferentes métodos estadísticos para la evaluación de la expresión diferencial, los cuales pertenecen a los algoritmos paramétricos, cuando éste asume algún tipo de distribución estadística de los datos de expresión; o a los algoritmos no paramétricos, los cuales no ajustan ni asumen ningún tipo de distribución de los datos *a priori*. Los algoritmos paramétricos por lo general asumen una distribución binomial negativa de los datos, lo cual implica que al comparar dos muestras biológicas distintas se espera que la mayor parte de los transcritos no muestren diferencias significativas en su nivel de expresión. Entre los algoritmos paramétricos más utilizados se encuentran edgeR y DESEQ2 (53-54). Por su parte, NOISEQ es el algoritmo no paramétrico más ampliamente utilizado, el cual además permite simular réplicas cuando sólo se tiene una muestra biológica por condición (55). Los umbrales en el nivel de cambio ("*fold change*") de expresión y de valor p ("*p-value*") para definir a un transcrito como diferencialmente expresado se pueden ajustar de acuerdo con el rigor de los análisis; en general se recomienda ser más estricto cuando se dispone de pocas réplicas biológicas y se puede ser más laxo a medida que el número de réplicas biológicas por condición crece. Cabe mencionar que aunque el RNA-seq ha mostrado ser una herramienta muy útil y precisa, cuando se disponen de pocas muestras con pocas réplicas, es aconsejable realizar mediciones del nivel de abundancia de los transcritos por un método independiente, siendo los ensayos de PCR en tiempo real a partir de cDNA (RT-qPCR) el método de elección.

Aplicaciones del RNA-seq

La rápida adopción del RNA-seq como herramienta para el estudio de procesos biológicos permea actualmente en una gran diversidad de disciplinas. En el área de la biología evolutiva, por ejemplo, el análisis del transcriptoma del tomate domesticado y de algunas especies silvestres se utilizó para comparar los patrones de expresión de distintos genes para identificar a aquellos que fueron seleccionados

artificialmente durante la domesticación de ésta solanácea, así como los tipos de presión evolutiva que actuaron sobre esta especie durante distintas etapas de su domesticación y diferentes eventos de hibridación con especies silvestres (56). En este ejemplo, el mapeo de las lecturas de RNA se realizó sobre un congénere con genoma secuenciado, es decir, sobre el genoma de referencia de una especie perteneciente al mismo género taxonómico. El conjunto de lecturas que no mapearon sobre el genoma de referencia se utilizaron para realizar un ensamblaje *de novo* e identificar transcritos putativos específicos de las especies silvestres. Por otro lado, proyectos colaborativos como el OneKP, en el cual se secuenciaron los transcriptomas de más de 1,300 especies pertenecientes a distintas familias de plantas, han permitido realizar análisis de transcriptómica comparativa para obtener filogenias moleculares con mejor resolución (57, 58); dilucidar el origen evolutivo, por ejemplo, de algunos receptores de hormonas vegetales (59) y de las desacetilasas de histonas (60) mediante la neofuncionalización de genes; analizar la divergencia de regiones regulatorias para la subfuncionalización y neofuncionalización de factores transcripcionales y su importancia en la remodelación de redes de regulación genética (61), entre otros.

Adicionalmente, el RNA-seq se ha empleado como herramienta para la exploración de procesos bioquímicos tales como la síntesis de metabolitos secundarios, por ejemplo, de la planta del té (*Camellia sinensis*; 62) o del peyote (*Lophophora williamsii*; 63); así como la caracterización del gametofito de los helechos (64), la caracterización de interacciones planta-patógeno (65), o la exploración de programas particulares del desarrollo, tales como el agotamiento del meristemo apical de la raíz en cactáceas (66).

El rápido desarrollo de las tecnologías de secuenciación, y de procesamiento de información masiva, permite actualmente que el RNA-seq ofrezca oportunidades diversas y atractivas en distintas áreas de la ciencia. Por ejemplo, la combinación del RNA-seq y de las técnicas histológicas permite realizar análisis de transcriptómica espacial para generar atlas de expresión con resolución a nivel de tejido (67). Por otro lado, la combinación con la técnica de separación de células asistida por dinámica de fluidos (FACS) permite secuenciar el transcriptoma de un solo tipo celular o de células únicas (68). La adopción de estas tecnologías permea ya en el estudio de diferentes especies modelo, y más lentamente, en especies cuyo genoma aún no está secuenciado, con lo cual se ampliará el estudio de especies no modelo y en consecuencia,

se fortalecerá nuestro conocimiento de las distintas especies, sus procesos metabólicos y moleculares.

Agradecimiento. El trabajo de los autores sobre el análisis de transcriptoma de *Pachycereus pringlei*,

una especie vegetal no modelo, está parcialmente financiado por proyectos CONACyT 240055 y PAPIIT-UNAM IN201318. A Gustavo Rodríguez-Alonso se le otorgó una beca doctoral del CONACyT (registro 290654).



REFERENCIAS

1. Krogh A (1929) The progress of physiology. *Science* 70:200-204.
2. Krebs HA (1975) The August Krogh principle: "for many problems there is an animal on which it can be most conveniently studied". *J Exp Zool* 194:221-226.
3. Russell JJ, Theriot JA, Sood P, Marshall WF, Landweber LF, Fritz-Laylin L, Polka JK, Oliferenko S, Gerbich T, Gladfelter A, Umen J, Bezanilla M, Lancaster MA, He S, Gibson MC, Goldstein B, Tanaka EM, Hu CK, Brunet A (2017) Non-model model organisms. *BMC Biol* 15:55.
4. Koornneef M, Meinke D (2010) The development of *Arabidopsis* as a model plant. *Plant J* 61:909-921.
5. Rédei GP (1975) *Arabidopsis* as a genetic tool. *Annu Rev Genet* 9:111-127.
6. Somerville C, Koornneef M (2002) A fortunate choice: the history of *Arabidopsis* as a model plant. *Nat Rev Genet* 3:883-889.
7. Bolker J (2012) There's more to life than rats and flies. *Nature* 491:31-33.
8. Alfred J, Baldwin IT (2015) New opportunities at the wild frontier. *eLife* 4:e06956.
9. Mora C, Tittlensor DP, Adl S, Simpson AGB, Worm B (2011) How many species are there on Earth and the ocean? *PLoS Biol* 9:e1001127.
10. Bonfante P, Genre A (2010) Mechanisms underlying beneficial plant-fungus interactions in mycorrhizal symbiosis. *Nat Commun* 1:48.
11. Mylona P, Pawlowski K, Bisseling T (1995) Symbiotic Nitrogen Fixation. *The Plant Cell* 7:869-885.
12. Read DJ (2003) Towards Ecological Relevance — Progress and pitfalls in the path towards an understanding of mycorrhizal functions in nature. En: *Mycorrhizal Ecology. Ecological Studies (Analysis and Synthesis)*. Editor: van der Heijden MGA, Sanders IR, vol. 157. Springer, Berlin, Heidelberg.
13. Vicente EJ, Dean DR (2017) Keeping the nitrogen-fixation dream alive. *Proc Natl Acad Sci USA*. 114:3009-3011.
14. Goldstein B, King N (2016) The Future of Cell Biology: Emerging model organisms. *Trends Cell Biol*. 26:818-824.
15. Gladfelter AS (2015) How nontraditional model systems can save us. *Mol. Biol. Cell* 26:3687-3689.
16. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM, Holko M, Yefanov A, Lee H, Zhang N, Robertson CL, Serova N, Davis S, Soboleva A. (2012) NCBI GEO: archive for functional genomics data sets - update. *Nuc Ac Res* 41:D991-D995.
17. Van Verk MC, Hickman R, Pieterse CM. J, Van Wees SCM (2013) RNA-seq: revelation of the messengers. *Trends Plant Sci* 18:175-179.
18. Wang Z, Gersten M, Snyder M (2009) RNA-seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* 10:57-63.
19. Zhao S, Fung-Leung WP, Bittner A, Ngo K, Liu X (2014) Comparison of RNA-seq and microarray in transcriptome profiling of activated T cells. *PLoS One* 9:e78644.
20. Johnson MT, Carpenter EJ, Tian Z, Bruskiwich R, Burris JN (2012) Evaluating methods for isolating total RNA and predicting the success of sequencing phylogenetically diverse plant transcriptomes. *PLoS One* 7:e50226.
21. Cui P, Lin Q, Ding F, Xin C, Gong W (2010) A comparison between ribo-minus RNA-sequencing and polyA-selected RNA-sequencing. *Genomics* 96:259-265.
22. Ari Ş, Arikan M (2016) Next-Generation Sequencing: Advantages, disadvantages, and future. En: *Plant Omics: Trends and Applications*. Editor: Hakeem K, Tombuloğlu H, Tombuloğlu G. Springer, Cham, Switzerland.
23. Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG, Hall KP, Evers DJ, Barnes CL, Bignell HR, Boutell JM, Bryant J, Carter RJ, Keira Cheatham R, Cox AJ, Ellis DJ, Flatbush MR, Gormley NA, Humphray SJ, Irving LJ, Karbelashvili MS, Kirk SM, Li H, Liu X, Maisinger KS, Murray

- LJ, Obradovic B, Ost T, Parkinson ML, Pratt MR, Rasolonjatovo IM, Reed MT, Rigatti R, Rodighiero C, Ross MT, Sabot A, Sankar SV, Scally A, Schroth GP, Smith ME, Smith VP, Spiridou A, Torrance PE, Tzonev SS, Vermaas EH, Walter K, Wu X, Zhang L, Alam MD, Anastasi C, Aniebo IC, Bailey DM, Bancarz IR, Banerjee S, Barbour SG, Baybayan PA, Benoit VA, Benson KF, Bevis C, Black PJ, Boodhun A, Brennan JS, Bridgham JA, Brown RC, Brown AA, Buermann DH, Bundu AA, Burrows JC, Carter NP, Castillo N, Chiara E, Catenazzi M, Chang S, Neil Cooley R, Crake NR, Dada OO, Diakoumakos KD, Dominguez-Fernandez B, Earnshaw DJ, Egbujor UC, Elmore DW, Etchin SS, Ewan MR, Fedurco M, Fraser LJ, Fuentes Fajardo KV, Scott Furey W, George D, Gietzen KJ, Goddard CP, Golda GS, Granieri PA, Green DE, Gustafson DL, Hansen NF, Harnish K, Haudenschild CD, Heyer NI, Hims MM, Ho JT, Horgan AM, Hoschler K, Hurwitz S, Ivanov DV, Johnson MQ, James T, Huw Jones TA, Kang GD, Kerelska TH, Kersey AD, Khrebtukova I, Kindwall AP, Kingsbury Z, Kokko-Gonzales PI, Kumar A, Laurent MA, Lawley CT, Lee SE, Lee X, Liao AK, Loch JA, Lok M, Luo S, Mammen RM, Martin JW, McCauley PG, McNitt P, Mehta P, Moon KW, Mullens JW, Newington T, Ning Z, Ling Ng B, Novo SM, O'Neill MJ, Osborne MA, Osnowski A, Ostadan O, Paraschos LL, Pickering L, Pike AC, Pike AC, Chris Pinkard D, Pliskin DP, Podhasky J, Quijano VJ, Raczky C, Rae VH, Rawlings SR, Chiva Rodriguez A, Roe PM, Rogers J, Rogert Bacigalupo MC, Romanov N, Romieu A, Roth RK, Rourke NJ, Ruediger ST, Rusman E, Sanches-Kuiper RM, Schenker MR, Seoane JM, Shaw RJ, Shiver MK, Short SW, Sizto NL, Sluis JP, Smith MA, Ernest Sohna Sohna J, Spence EJ, Stevens K, Sutton N, Szajkowski L, Tregidgo CL, Turcatti G, Vandevondele S, Verhovskiy Y, Virk SM, Wakelin S, Walcott GC, Wang J, Worsley GJ, Yan J, Yau L, Zuerlein M, Rogers J, Mullikin JC, Hurler ME, McCooke NJ, West JS, Oaks FL, Lundberg PL, Klenerman D, Durbin R, Smith AJ (2008) Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* 456:53-59.
24. Rothberg JM, Hinz W, Rearick TM, Schultz J, Mileski W, Davey M, Leamon JH, Johnson K, Milgrew MJ, Edwards M, Hoon J, Simons JF, Marran D, Myers JW, Davidson JF, Branting A, Nobile JR, Puc BP, Light D, Clark TA, Huber M, Branciforte JT, Stoner IB, Cawley SE, Lyons M, Fu Y, Homer N, Sedova M, Miao X, Reed B, Sabina J, Feierstein E, Schorn M, Alanjary M, Dimalanta E, Dressman D, Kasinskas R, Sokolsky T, Fidanza JA, Namsaraev E, McKernan KJ, Williams A, Roth GT, Bustillo J (2011) An integrated semiconductor device enabling non-optical genome sequencing. *Nature* 475:348-352.
 25. Lee CS, Kim SK, Kim M (2009) Ion-sensitive field-effect transistor for biological sensing. *Sensors (Basel)* 9:7111-71131.
 26. Salipante SJ, Kawashima T, Rosenthal C, Hoogestraat DR, Cummings LA (2014) Performance comparison of Illumina and ion torrent next-generation sequencing platforms for 16S rRNA-based bacterial community profiling. *Appl Environ Microbiol* 80:7583-7591.
 27. Travers KJ, Chin CS, Rank DR, Eid JS, Turner SW (2010) A flexible and efficient template format for circular consensus sequencing and SNP detection. *Nuc Ac Res* 38:e159.
 28. Korlach J, Bibillo A, Wegener J, Peluso P, Pham TT, Park I, Clark S, Otto GA, Turner SW (2008) Long, processive enzymatic DNA synthesis using 100% dye-labeled terminal phosphate-linked nucleotides. *Nucleos Nucleot Nucl* 27:1072-1083.
 29. Foquet M, Samiee KT, Kong X, Chaudhuri BP, Lundquist PM, Turner SW, Freudenthal J, Roitman DB (2008) Improved fabrication of zero-mode waveguides for single-molecule detection. *J Appl Phys* 103:034301.
 30. ENCODE project consortium (2004) The ENCODE (ENCyclopedia Of DNA Elements) project. *Science* 306:636-640.
 31. Schurch NJ, Schofield P, Gierliński M, Cole C, Sherstnev A, Singh V, Wrobel N, Gharbi K, Simpson GG, Owen-Hughes T, Blaxter M, Barton GJ (2016) How many biological replicates are needed in an RNA-seq experiment and which differential expression tool should you use? *RNA* 22:839-851.
 32. Marioni JC, Mason CE, Mane SM, Stephens M, Gilad Y (2008) RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. *Genome Res* 18:1509-1517.
 33. Hart SN, Therneau TM, Zhang Y, Poland GA, Kocher JP (2013) Calculating sample size estimates for RNA sequencing data. *J Comput Biol* 20:970-978.
 34. Sims D, Sudbery I, Ilott NE, Heger A, Ponting CP (2014) Sequencing depth and coverage: key considerations in genomic analysis. *Nat. Rev. Genet.* 15:121-132.
 35. Risso D, Ngai J, Speed TP, Dudoit S (2014) Normalization of RNA-seq data using factor

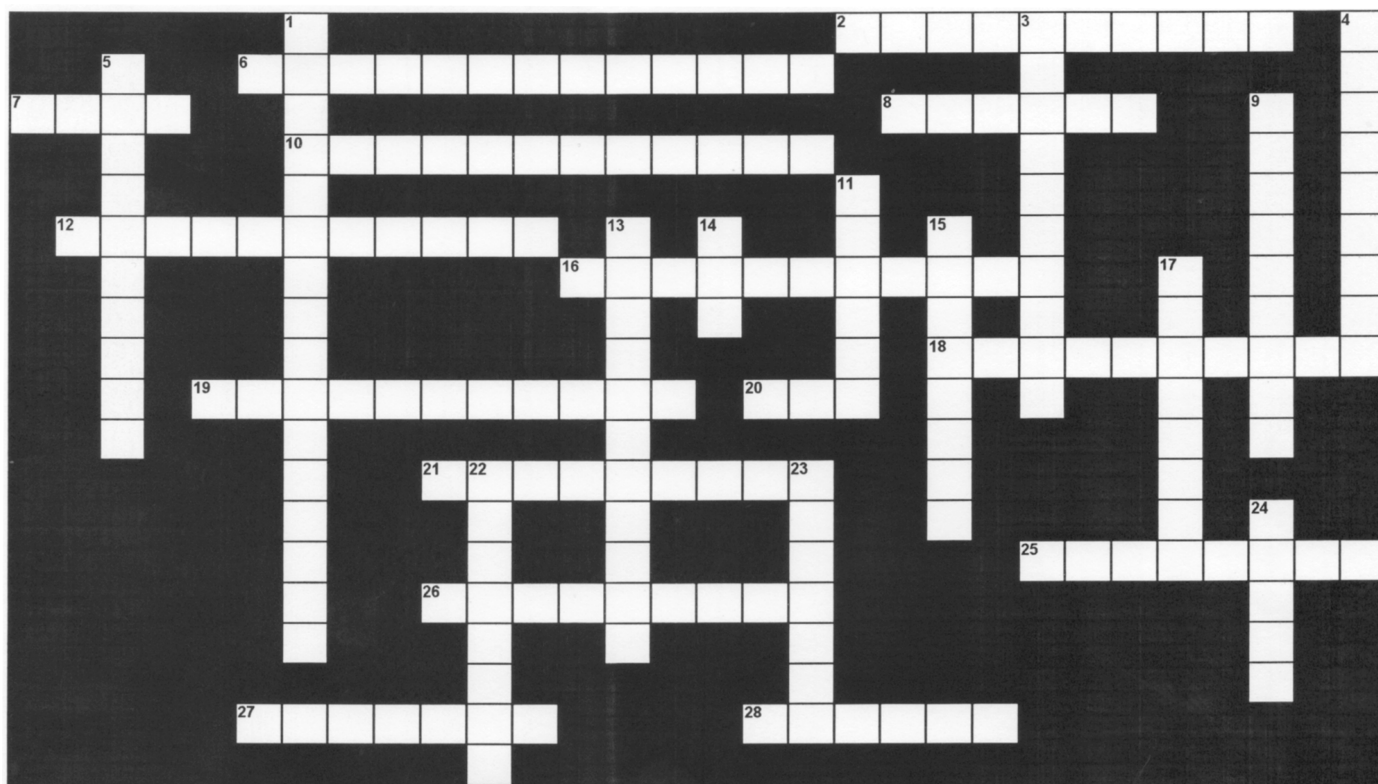
- analysis of control genes or samples. *Nat Biotechnol* 32:896-902.
36. Jiang L, Schlesinger F, Davis CA, Zhang Y, Li R, Salit M, Gingeras TR, Oliver B (2011) Synthetic spike-in standards for RNA-seq experiments. *Genome Res* 21:1543-1551.
 37. Munro SA, Lund SP, Pine PS, Binder H, Clevert DA (2014) Assessing technical performance in differential gene expression experiments with external spike-in RNA control ratio mixtures. *Nat Commun* 5:5125.
 38. Andrews S (2010) FastQC: a quality control tool for high throughput sequence data. <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
 39. Bolger AM, Lohse M, y Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114-2120.
 40. Hannon Lab. Cold Spring Harbor Laboratory. http://hannonlab.cshl.edu/fastx_toolkit/
 41. Zhang J, Kobert K, Flouri T, Stamatakis A (2014) PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics* 30:614-620.
 42. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A (2011) Full-length transcriptome assembly from RNA-seq data without a reference genome. *Nat Biotechnol* 29:644-652.
 43. Pruitt KD, Tatusova T, Brown GR, Maglott DR (2012) NCBI Reference Sequences (RefSeq): current status, new features and genome annotation policy. *Nuc Ac Res* 40:D130-D135.
 44. TransDecoder <https://github.com/TransDecoder>
 45. Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21:3674-3676.
 46. Zdobnov EM, y Apweiler R (2001) InterProScan-- an integration platform for the signature-recognition methods in InterPro. *Bioinformatics* 17:847-848.
 47. Musacchia F, Basu S, Petrosino G, Salvemini M, Sanges R (2015) Annocript: a flexible pipeline for the annotation of transcriptomes able to identify putative long noncoding RNAs. *Bioinformatics* 31:2199-2201.
 48. Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M (2007) KAAS: an automatic genome annotation and pathway reconstruction server. *Nuc Ac Res* 35:W182 - W185.
 49. Meunier J, Lemoine F, Soumillon M, Liechti A, Weier M, Guschanski K, Hu H, Khaitovich P, Kaessmann H (2013) Birth and expression evolution of mammalian microRNA genes. *Genome Res* 23:34-45.
 50. Johnsson P, Lipovich L, Grandér D, Morris KV (2014) Evolutionary conservation of long non-coding RNAs; sequence, structure, function. *Biochim Biophys Acta* 184:1063-1071.
 51. Li B, Dewey CN (2011) RSEM: accurate transcript quantification from RNA-seq data with or without a reference genome. *BMC Bioinformatics*. 12:323.
 52. Engström PG, Steijger T, Sipos B, Grant GR, Kahles A, Rättsch G, Goldman N, Hubbard TJ, Harrow J, Guigó R, Bertone P; RGASP Consortium. (2013) Systematic evaluation of spliced alignment programs for RNA-seq data. *Nat Meth* 10:1185-1191.
 53. Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26:139-140.
 54. Love MI, Huber W, Anders S (2014) Moderate estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15:550.
 55. Tarazona S, García-Alcalde F, Dopazo J, Ferrer A, Conesa A (2011) Differential expression in RNA-seq: a matter of depth. *Genome Res* 21:2213-2223.
 56. Koenig D, Jiménez-Gómez JM, Kimura S, Fulop D, Chitwood DH, Headland LR, Kumar R, Covington MF, Devisetty UK, Tat AV, Tohge T, Bolger A, Schneeberger K, Ossowski S, Lanz C, Xiong G, Taylor-Teeple M, Brady SM, Pauly M, Weigel D, Usadel B, Fernie AR, Peng J, Sinha NR, Maloof JN (2013) Comparative transcriptomics reveals patterns of selection in domesticated and wild tomato. *Proc Nat Acad Sci USA*, 110:E2655-E2662.
 57. Jiao Y, Leebens-Mack J, Ayyampalayam S, Bowers JE, McKain MR, McNeal J, Rolf M, Ruzicka DR, Wafula E, Wickett NJ, Wu X, Zhang Y, Wang J, Zhang Y, Carpenter EJ, Deyholos MK, Kutchan TM, Chanderbali AS, Soltis PS, Stevenson DW, McCombie R, Pires JC, Wong GK, Soltis DE, Depamphilis CW (2012) A genome triplication associated with early diversification of the core eudicots. *Genome Biol* 13:R3.
 58. Yang Y, Moore M. J, Brockington SF, Soltis DE, Wong GK, Carpenter EJ, Zhang Y, Chen L, Yan

- Z, Xie Y, Xie Y, Sage RF, Covshoff S, Hibberd JM, Nelson MN, Smith SA (2016) Dissecting molecular evolution in the highly diverse plant clade Caryophyllales using transcriptome sequencing. *Mol Biol Evol* 32:2001-2014.
59. Bythell-Douglas R, Rothfels CJ, Stevenson DWD, Graham SW, Wong GK, Nelson DC, Bennett T (2017) Evolution of strigolactone receptors by gradual neo-functionalization of KAI2 paralogues. *BMC Biol.* 15:52.
 60. Bourque S, Jeandroz S, Grandperret V, Lehotai N, Aimé S, Soltis DE, Miles NW, Melkonian M, Deyholos MK, Leebens-Mack JH, Chase MW, Rothfels CJ, Stevenson DW, Graham SW, Wang X, Wu S, Pires JC, Edger PP, Yan Z, Xie Y, Carpenter EJ, Wong GKS, Wendehenne D, Nicolas-Francès V (2016) The Evolution of HD2 proteins in green plants. *Trends Plant Sci.* 21:1008-1016.
 61. Sayou C, Monniaux M, Nanao MH, Moyroud E, Brockington SF, Thévenon E, Chahtane H, Warthmann N, Melkonian M, Zhang Y, Wong GK, Weigel D, Parcy F, Dumas R (2014) A promiscuous intermediate underlies the evolution of LEAFY DNA binding specificity. *Science* 343:645-648.
 62. Chun-Fang L, Yan Z, Yao Y, Qiong-Yi Z, Sheng-Jun W, Xin-Chao W, Ming-Zhe Y, Da L, Xuan L (2015) Global transcriptome and gene regulation network for secondary metabolite biosynthesis of tea plant (*Camellia sinensis*) *BMC Genomics* 16:560.
 63. Ibarra-Laclette E, Zamudio-Hernández F, Pérez-Torres CA, Albert VA, Ramírez-Chávez E, Molina-Torres J, Fernández-Cortés A, Calderón-Vázquez C, Olivares-Romero JL, Herrera-Estrella A, Herrera-Estrella L (2015) De novo sequencing and analysis of *Lophophora williamsii* transcriptome, and searching for putative genes involved in mescaline biosynthesis. *BMC Genomics* 16:657.
 64. Der JP, Barker MS, Wickett NJ, Depamphilis CW, Wolf PG (2011) De novo characterization of the gametophyte transcriptome in bracken fern, *Pteridium aquilinum*. *BMC Genomics* 12:99.
 65. Nagano AJ, Honjo MN, Mihara M, Sato M, Kudoh H (2015) Detection of plant viruses in natural environments by using RNA-Seq. *Methods Mol Biol* 1236:89-98.
 66. Rodríguez-Alonso G, Matvienko M, López-Valle ML, Lázaro-Mixteco PE, Napsucialy-Mendivil S, Dubrovsky JG, Shishkova S (2018) Transcriptomics insights into the genetic regulation of root apical meristem exhaustion and determinate primary root growth in *Pachycereus pringlei* (Cactaceae). *Sci Rep* 8:8529.
 67. Lieben L (2017) Plant genetics: Spatial transcriptomics in plants. *Nat Rev Genet* 18:394.
 68. Tang F, Lao K, Surani MA (2011) Development and applications of single-cell transcriptome analysis. *Nat Methods* 8:S6-S11.

CRUCIBIOQ®

CONCEPTOS DE GENÉTICA

Yolanda Saldaña Balmori
 Correo E: balmori@bq.unam.mx



HORIZONTALES

- 2** Son el resultado de las modificaciones en el DNA las cuales pueden ser modificaciones de una forma evolutiva espontánea o bien ocasionadas entre otras causas, por radiaciones o por la acción de xenobióticos.
- 6** Mecanismo mediante el cual dos moléculas de DNA dan lugar a una tercera que combina segmentos procedentes de las dos progenitoras.
- 7** Grupo de secuencias de DNA, células u organismos que son genéticamente iguales debido a que proceden de un único antecesor.
- 8** Conjunto de la información genética total de un organismo, codificado en la secuencia de nucleótidos de su DNA.
- 10** Así se designa al mutágeno que se intercala entre dos bases nitrogenadas de un ácido nucleico ocasionando un desplazamiento del marco de lectura.
- 12** Constituidos por una base nitrogenada, un azúcar y un ácido fosfórico unidos entre sí por uniones éster, participan en tres tipos de funciones celulares al formar parte de los ácidos nucleicos, de las moléculas energéticas y de las coenzimáticas.
- 16** Organismo que tiene genes de otro, los cuales se han incorporado mediante el proceso de recombinación de DNA.
- 18** Es el proceso en el que la información genética del DNA y su asociación con el mRNA conducen a la síntesis de proteínas.
- 19** Grupo de nucleótidos formados a partir de carbamoil fosfato y aspartato, el anillo se cierra con la participación de la enzima dihidroorotasa.

- 20** Es una secuencia de DNA que codifica la estructura de una proteína o un RNA.
- 21** Es la adenosina monofosfato, se sintetiza a partir del inosinato por sustitución de un grupo amino por un carboxilo en el carbono 6.
- 25** Es el sitio donde se traduce la información transportada por el mRNA para dar lugar a una secuencia polipeptídica, es un orgánulo formado por dos subunidades una grande y una pequeña constituidas por rRNA y proteínas.
- 26** Molécula de RNA que traslada la información genética desde el DNA nuclear a los ribosomas donde actúa como molde para designar la secuencia de aminoácidos en una estructura polipeptídica.
- 27** Segmento inicial de una cadena nueva de DNA que es complementaria al molde; por su grupo 3'hidroxilo libre se le pueden añadir nucleótidos.
- 28** Así se llama a la secuencia de nucleótidos presente en un gen de célula eucariota que se transcribe, pero que se elimina del RNA antes de su traducción.
- 5** Nombre con el que se designa a la hebra del DNA que durante la replicación se sintetiza en el mismo sentido que la horquilla de replicación.
- 9** Formada por DNA y proteínas del tipo de las histonas en cantidades generalmente iguales y una pequeña cantidad de RNA; más otro tipo de proteínas que ayudan a mantener la estructura, y otras que regulan la expresión de algunos genes.
- 11** Unidad de expresión genética constituida por uno o varios genes, el promotor y las secuencias implicadas en la regulación.
- 13** Los ribosomas de _____ tienen 3 tipos de rRNA con los coeficientes de sedimentación de 23, 5 y 16S; la primera y 34 proteínas, más la segunda constituyen la subunidad de 50S; mientras que la de 16S más 21 proteínas forman la subunidad 30S, las dos subunidades unidas dan lugar a la estructura.
- 14** Siglas del ácido polinucleotídico constituido por residuos de los ácidos adenílico, guanílico, citidílico y uridílico.
- 15** Proteínas presentes en la cromatina de las células eucarióticas, formadas por un porcentaje muy alto de los aminoácidos básicos arginina y lisina que según sus proporciones se designan como H1, H2A, H2B, H3 y H4.
- 17** Región del núcleo de las células eucarióticas en la que se transcriben y procesan los RNA ribosómicos.

VERTICALES

- 1** Es la forma en la que el DNA se duplica ya que se conserva la mitad del material original en cada una de las dos copias.
- 3** Estructuras nucleares constituidas por DNA y proteínas, en el humano cada célula contiene 23 pares; en el último par en las mujeres la estructura es XX y en los varones es XY, es en este último par de los varones en donde se define el sexo del producto.
- 4** Secuencia de tres nucleótidos de una molécula de tRNA que es complementaria de un triplete de mRNA.
- 22** Este tipo de mutación es cuando se pierde uno o más nucleótidos de un gen o un cromosoma.
- 23** Así se llama al gen que muta y hace que las células tengan una proliferación rápida e incontrolada, lo que ocasiona la producción de un cáncer.
- 24** Conjunto de tres nucleótidos del mRNA que por su participación con una porción del tRNA seleccionará a un aminoácido.

AMINO-ESTRUCTURA: UNA PROPUESTA PEDAGÓGICA PARA EL APRENDIZAJE DE LOS VEINTE AMINOÁCIDOS QUE COMÚNMENTE SE ENCUENTRAN EN LAS PROTEÍNAS*

Juan Carlos Vega Garzón

Departamento de Bioquímica, Instituto de Biología. Universidade Estadual de Campinas Rua Monteiro Lobato, 255 Campinas, SP - 13083-862 - Brasil. Correo E: jcvegag@unal.edu.co

RESUMEN

Una de las principales tareas para un estudiante de bioquímica es el aprendizaje de los nombres, las estructuras y las abreviaturas de una y tres letras de los veinte aminoácidos que se encuentran comúnmente en las proteínas. Enseñar aminoácidos se ha convertido en un verdadero desafío pedagógico debido a que muchos estudiantes aprenden esta temática como un ejercicio de memorización y peor aun muchos profesores la enseñan de esta manera, pero la mera repetición no tiene ningún sentido ni mérito académico. Este enfoque no refuerza el aprendizaje ni le permite a los estudiantes reconocer las relaciones entre la estructura y función de los aminoácidos ni su transformación dentro de una vía metabólica. Con el objetivo de auxiliar la enseñanza y el aprendizaje de la estructura de los veinte aminoácidos que se encuentran comúnmente en las proteínas, fue desarrollado el aplicativo Amino Estructura. En este artículo se muestra como obtenerlo y usarlo, así como también la propuesta pedagógica para su utilización.

ABSTRACT

One of the main tasks for a biochemistry student is learning the names, structures, and abbreviations of one and three letters of the twenty amino acids commonly found in proteins. Teaching amino acids has become a real pedagogical challenge because many students learn this subject as an exercise in memorization, and worse yet many teachers teach it this way, but mere repetition has no meaning or academic merit. This approach neither reinforces learning nor allows students to recognize the relationships between the structure and function of amino acids nor their transformation within a metabolic pathway. To aid in teaching and to learn the structure of the twenty amino acids commonly found in proteins, the Amino Structure application was developed. This article shows how to obtain and use it, as well as the pedagogical proposal for its use.

INTRODUCCIÓN

Por décadas los profesores de bioquímica han explorado la manera de ayudar a los estudiantes a desarrollar un entendimiento conceptual de las representaciones químicas, las representaciones químicas se definen como varios tipos de fórmulas, estructuras y símbolos usados para representar

procesos químicos y entidades conceptuales (moléculas y átomos) y que pueden ser vistas como metáforas, modelos y constructos teóricos que los químicos utilizan para interpretar la naturaleza y la realidad. La principal función de estas representaciones es la de establecer un lenguaje común que los científicos usan para compartir sus descubrimientos (1). Una buena forma para conocer

PALABRAS

CLAVE:

Aminoácido, Estructura molecular, Modelo molecular, Representación molecular.

KEY WORDS:

Amino acid, Molecular structure, Molecular model, Molecular representation.

las estructuras de las moléculas es mediante la creación de modelos que las representen (2). Los químicos han representado la información molecular por medio de modelos moleculares, debido a que el cerebro humano comprende mejor una representación gráfica que datos numéricos de distancias y ángulos. La representación molecular más utilizada son los diagramas de estructura en 2D, sin embargo la representación de estructuras en 3D (Fig. 1) aporta información especialmente importante para el entendimiento de las propiedades químicas y biológicas de las moléculas y la relación entre características estructurales y funciones moleculares (3).

Una de las principales tareas para un estudiante de bioquímica es el aprendizaje de los nombres, las estructuras y las abreviaturas de una y tres letras de los veinte aminoácidos que se encuentran comúnmente en las proteínas (4, 5). El conocimiento de la estructura de los aminoácidos es esencial para entender su comportamiento físico (tamaño, forma y conformación espacial) y químico (reactividad) debido a que en gran parte su comportamiento está determinado por el tipo, número y uniones de sus átomos constituyentes (6). Este conocimiento es fundamental para abordar la función y transformación de los aminoácidos en el metabolismo.

Enseñar aminoácidos se ha convertido en un verdadero desafío pedagógico debido a que muchos estudiantes aprenden esta temática como un ejercicio de memorización y peor aun muchos profesores la enseñan de esta manera, pero la mera repetición no tiene ningún sentido ni mérito académico. Los estudiantes pueden aprender de memoria los nombres, estructuras y abreviaturas con el fin de pasar un examen, pero es poco probable que recuerden esta información después del período de evaluaciones (7). Este enfoque no refuerza el aprendizaje ni le permite a los estudiantes reconocer las relaciones entre la estructura y función de los aminoácidos ni su transformación dentro de una vía metabólica. Además los estudiantes van a necesitar de este conocimiento a lo largo de sus carreras académicas y profesionales debido a que en casi todas las disciplinas de las ciencias de la vida ya sea en el laboratorio o en estudios teóricos se deben poseer conocimientos básicos sobre los aminoácidos, también estos conocimientos son una poderosa herramienta para comprender artículos científicos (8, 9).

Con el objetivo de auxiliar la enseñanza y el aprendizaje de la estructura de los veinte aminoácidos que se encuentran comúnmente en las proteínas, fue desarrollado el aplicativo Amino Estructura.

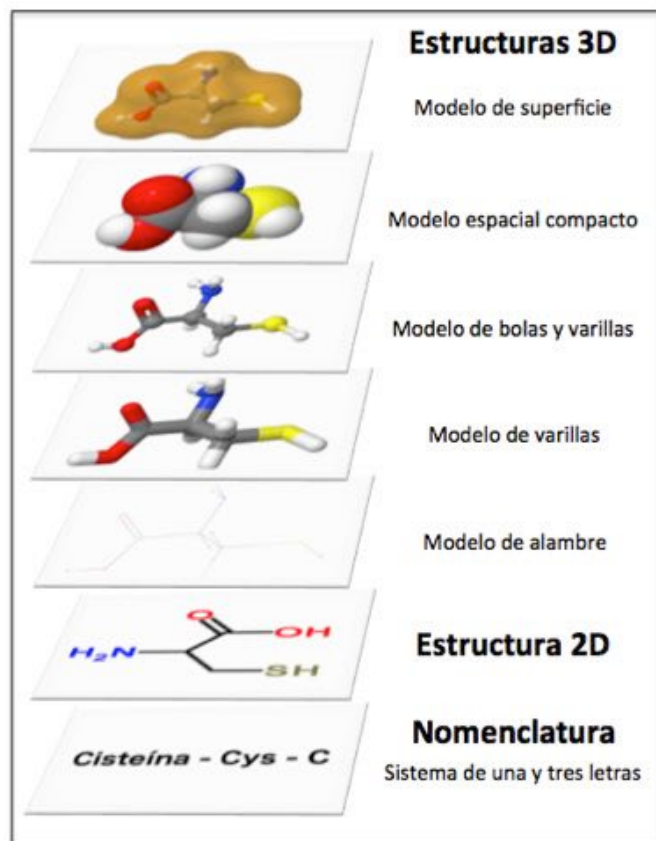


Figura 1. Diferentes representaciones gráficas del aminoácido Cisteína. Cada nivel aporta mayor información sobre sus propiedades químicas y biológicas.

Diseño del aplicativo: Amino estructura fue desarrollado usando XCode (integrate development environment for iOS). Las moléculas 2D fueron obtenidas de ChemSpider database. Las moléculas 3D fueron obtenidas de Jmol (computer software for molecular modeling chemical structures in 3-dimensions). Blender fue usado para la creación de los modelos moleculares en 3D anáglifo.

Descargando el aplicativo: Amino Estructura se encuentra disponible de forma gratuita para dispositivos iOS en la App Store <https://itunes.apple.com/es/app/amino-estructura/id1074336561>

Instrucciones de uso: Amino Estructura esta conformado por cinco secciones (Info, 2D, 3D, Quiz y Créditos) para ingresar a cada sección solo es necesario tocar sobre el icono de cada sección en la parte inferior la pantalla del dispositivo móvil (Fig. 2). En la sección información (info, parte A de la figura 2) se muestra la representación gráfica de la estructura básica de los aminoácidos, la definición de aminoácido y una breve descripción de los cuatro

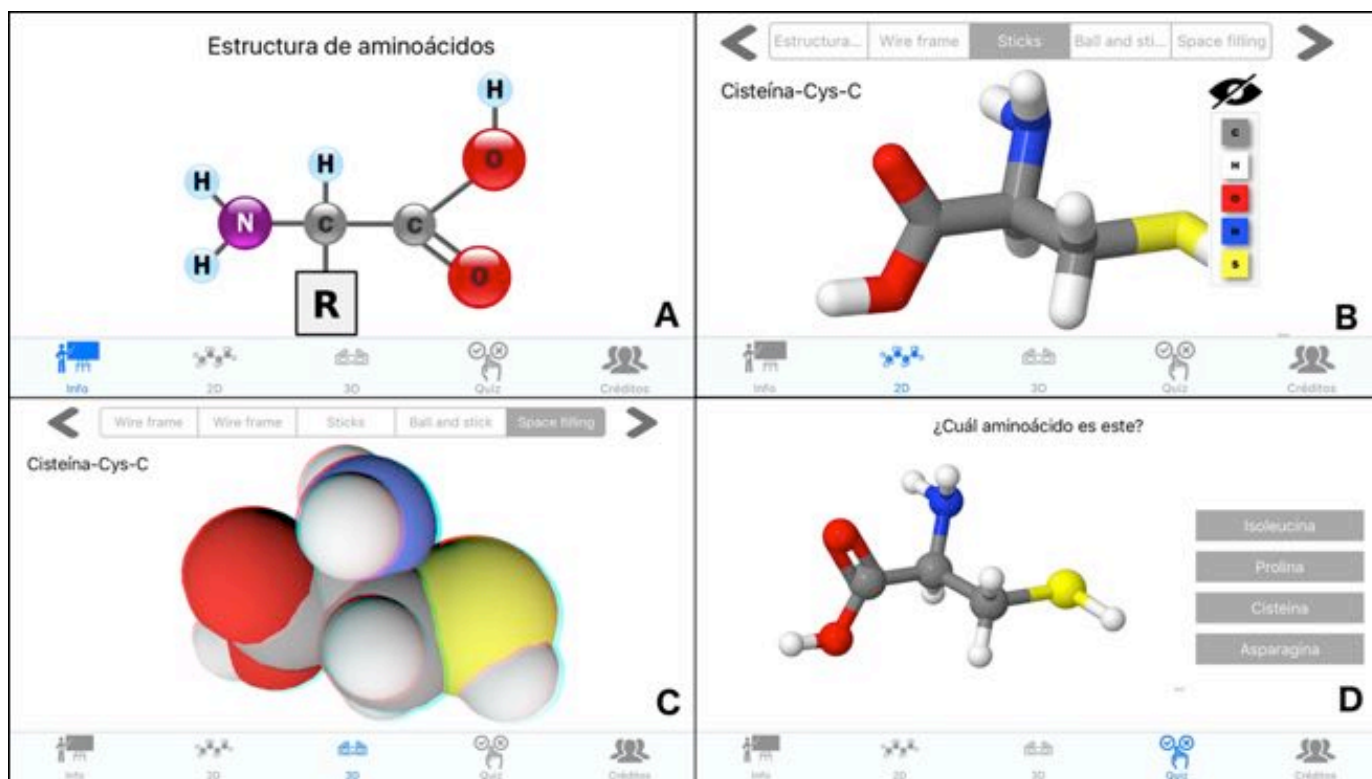


Figura 2. Cuatro principales secciones del aplicativo Amino Estructura. A Sección información, B sección 2D, C sección 3D y D sección Quiz.

principales modelos estructurales en 3D (estructura de alambre - Wire Frame Model, de barras -Sticks Model, barras y esferas -Ball and Stick Model y de espacio lleno - Space Filling Model).


En la sección 2D (parte B de la figura 2) se muestran cinco representaciones gráficas de cada uno de los veinte aminoácidos comúnmente presentes en las proteínas, junto con su nombre y su abreviatura de una y tres letras; para la mejor visualización de los diferentes tipos de átomos que conforman cada aminoácido se incluye el código de colores de la Unión Internacional de Química Pura y Aplicada (UIPAC). Los símbolos < y > son utilizados para ir al aminoácido anterior o siguiente respectivamente.

En la sección 3D (parte C de la figura 2) se muestran cinco representaciones gráficas de cada uno de los veinte aminoácidos comúnmente presentes en las proteínas, junto con su nombre y su abreviatura de una y tres letras. A diferencia de la sección 2D, la sección 3D muestra las mismas representaciones gráficas de los aminoácidos pero en 3D anaglifo (es necesaria la utilización de gafas adecuadas para visualizar este tipo de imágenes). Los símbolos < y > son utilizados

para ir para al aminoácido anterior o siguiente respectivamente.

La sección Quiz (parte D de la figura 2) corresponde al módulo de evaluación, este módulo está compuesto por veinte preguntas, pero solo diez (seleccionadas aleatoriamente) deben ser respondidas por el usuario cada vez que ingresa a esta sección del aplicativo.

PROPUESTA PEDAGÓGICA

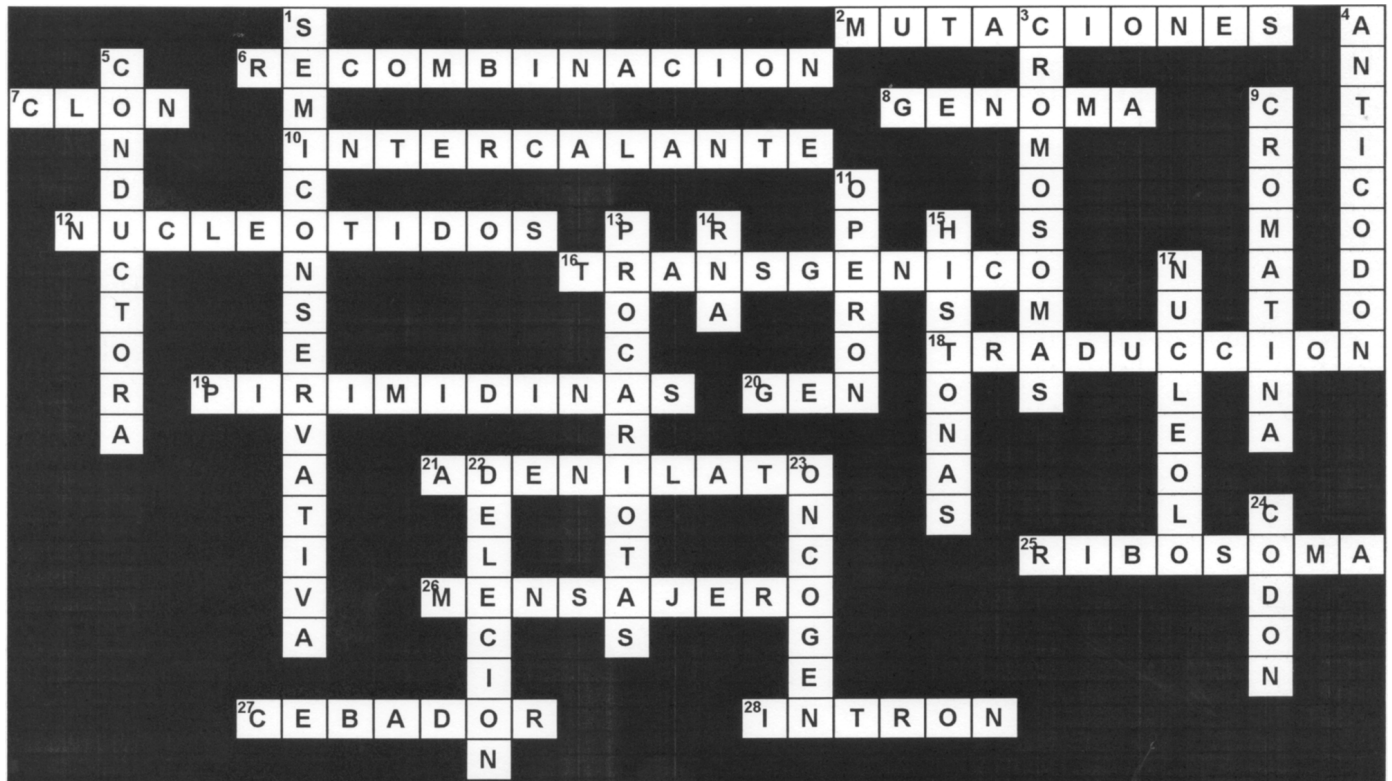
Amino estructura, esta diseñado para proporcionar a los estudiantes una herramienta de aprendizaje innovadora usando el poder de las tecnologías móviles de visualización en 3D. La idea es aportar la información más relevante para el aprendizaje de la estructura y nomenclatura de los aminoácidos. La sección Quiz es un juego de preguntas y respuesta, el aplicativo pone a disposición diez preguntas de un banco de veinte. El orden de las preguntas es seleccionado aleatoriamente, de modo que no existe ninguna secuencia lógica para responderlas. El desempeño del alumno es evaluado teniendo en cuenta el número de errores y el tiempo utilizado para responder todo el cuestionario. 

REFERENCIAS

1. Wu HK, Krajcik JS, Soloway E (2001) Promoting Understanding of Chemical Representations: Students' Use of a Visualization Tool in the Classroom. *J. Res. Sci. Teach* 38: 821-842.
2. Sanchez CV, Rosales LL (2008) Herramientas de visualización molecular para la enseñanza. *Mensaje Bioquímico* 32: 49-58.
3. Engel, T. (2004) Representation of Chemical Compounds. In *Chemoinformatics*. Editor: Gasteiger J, Engel T. Wiley-VCH Verlag GmbH & Co. KGaA, pp 15-168.
4. Moore DS (1987) Biochemically Informative Acronyms for the Twenty Common Amino Acids. *Biochem. Educ* 15: 74-76.
5. Saffran M (1998) Amino acid names and parlor games: from trivial names to a one-letter code, amino acid names have strained students' memories. Is a more rational nomenclature possible?. *Biochem. Educ* 26: 116-118.
6. Koolman J, Roehm KH (2005) *Color Atlas of Biochemistry*. Second edition. Georg Thieme Verlag. Stuttgart, Germany, p 467.
7. Winkle LJ (1985) A Summary of Amino Acid Metabolism Based on Amino Acid Structure. *Biochem. Educ* 13: 25-26.
8. Mezl VA (2001) The AAA amino acid list a mnemonic derivation of the structures and properties of the amino acids. *Biochem Mol Biol Educ* 29: 66-70
9. Pennington BO, Sears D, Clegg DO (2014) Interactive Hangman Teaches Amino Acid Structures and Abbreviations. *Biochem Mol Biol Educ* 42: 495 - 500.

SOLUCIÓN AL CRUCIBIOQ[®] CONCEPTOS DE GENÉTICA

Yolanda Saldaña Balmori
Correo E: balmori@bq.unam.mx



INSTRUCCIONES PARA LOS COLABORADORES DE LA REVISTA DE EDUCACIÓN BIOQUÍMICA (REB)

La REB es una revista dedicada a la divulgación, difusión, discusión, análisis y presentación de resultados derivados de investigaciones originales en temas relevantes en el campo de la bioquímica y áreas afines. La Revista está dirigida a investigadores, profesores y estudiantes de posgrado, licenciatura y educación media superior. Los trabajos que se someten a evaluación para su posible publicación no deben de haberse presentado total o parcialmente en otras publicaciones.

Se aceptan únicamente contribuciones originales con estricto contenido científico en forma de artículos de investigación, de revisión, de crítica, de análisis y otras comunicaciones que tengan que ver con diversas formas de estimular el aprendizaje de la bioquímica y sirvan de apoyo a investigadores, profesores y alumnos, tanto en aspectos de investigación, académicos y de actualización.

I. Los artículos se deben ajustar a los siguientes lineamientos editoriales:

1) Portada. En el primer párrafo incluir el título, el cual debe de ser claro, simple, atractivo y evitar las abreviaturas o en su caso, definir las al inicio del texto. En el siguiente párrafo se anotarán los nombres completos de los autores, iniciando por el nombre propio completo. La afiliación de los autores se escribirá en el siguiente párrafo, indicando departamento, institución, ciudad, estado y país y la dirección de correo electrónico del autor responsable. La afiliación de los autores se indicará con números entre paréntesis. Se proporcionará un título breve con un máximo de 60 caracteres.

2) Resumen. Se deberán incluir dos resúmenes, uno en idioma español y uno en inglés (Abstract) de no más de 350 caracteres.

3) Palabras clave. Se deberá proporcionar de tres a seis palabras clave en idioma español y en inglés.

4) Texto. El artículo deberá ser escrito en el procesador de textos "Word", con una extensión máxima recomendada de 15 cuartillas a doble espacio, en "Times New Roman 12" como fuente de la letra, sin formato de texto, tabuladores o pies de página. Las figuras y tablas se presentarán separadas del texto.

5) Referencias. Se indicarán en el texto con números entre paréntesis de acuerdo a su orden de aparición. Las referencias se enlistarán al final del trabajo y deben contener para los artículos: apellidos e iniciales de todos los autores, año

de publicación entre paréntesis, título completo del artículo y después de un punto, el nombre oficial de la revista abreviado como aparece en el Current Contents, número del volumen y antecedido por dos puntos, el número de la primera y última páginas, de acuerdo con el siguiente ejemplo: Martin GM, Austad SN, Johnson TE (1996) Generic analysis of ageing: role of oxidative damage and environmental stresses. Nature Gen 113:25-34. Los artículos en libros deberán citarse de la siguiente forma: Wood KJ (1992) Tolerance to alloantigens. En: The Molecular Biology of Immunosuppression. Editor: Thomson A W. John Wiley and Sons Ltd, Ann Arbor, Michigan, USA, pp 81-104. Los libros podrán incluir las páginas totales o las consultadas y se citarán de acuerdo con este ejemplo: Lehninger AL, Nelson DL, Cox MM (1993) Principles of Biochemistry. Worth Publishers, New York, NY, USA, p 1013.

6) Figuras y Tablas. Las figuras se pueden presentar en formato "jpg" o integradas en un archivo de "Power Point" o del mismo "Word" separadas del texto del artículo. Las figuras pueden presentarse en colores, con fondo y sombreado. Las tablas deben de estar en "Word" sin formatos especiales y separadas del texto del artículo. Las figuras y las tablas se deberán numerar con arábigos. Las leyendas y los pies de figuras se deberán presentar en una hoja aparte. Se deberá considerar que las figuras y las tablas se reducirán a la mitad o a un cuarto de las dimensiones de una hoja tamaño carta; las letras y números más pequeños no deben ser menores de dos milímetros después de la reducción. En caso de emplear figuras previamente publicadas, deberá darse el crédito correspondiente y en su caso

obtener el permiso para su publicación. Las figuras dentro del texto deberán mencionarse con minúsculas, la palabra entera y sin paréntesis; cuando se haga referencia a ellas deberá citarse con la abreviatura, la primera letra mayúscula y entre paréntesis (Fig. 2). Las tablas siempre llevarán la primera letra a mayúscula y entre paréntesis (Tabla 2).

7) Abreviaturas. Las abreviaturas seguirán las normas de la IUPAC, aquellas específicas o poco comunes deberán definirse entre paréntesis, la primera vez que se utilicen.

II. Otras comunicaciones incluyen: resúmenes de artículos científicos interesantes, relevantes o significativos, información científica o académica de interés general, avisos de reuniones académicas y cursos, problemas teóricos, ejercicios prácticos, juegos didácticos con orientación educativa, bolsa de trabajo o comentarios de artículos publicados previamente en la REB, cartas al editor, entre otras. Éstas deberán seguir los siguientes lineamientos:

- 1) El contenido deberá ser desarrollado en forma resumida y de una manera explícita.
- 2) Se aceptará un máximo de 10 referencias que se citarán entre paréntesis en el texto, como se indica en el inciso I-5. Se podrán incluir figuras o tablas, con las características que se indican en el inciso I-6.

Los manuscritos que no cumplan con las especificaciones de la Revista no serán aceptados de primera instancia para su revisión.

Los manuscritos serán evaluados por lo menos por tres revisores seleccionados por el Comité Editorial a quienes se les enviará el trabajo con los autores en anónimo y quienes en todo momento permanecerán anónimos para los autores y entre

ellos; los revisores opinarán sobre la relevancia del trabajo en un lapso no mayor a un mes. Las correcciones y sugerencias de los revisores serán enviadas con anonimato entre ellos al Editor en Jefe. El resultado de la evaluación puede ser: rechazado, enviado para correcciones o aceptado. Una vez obtenida una evaluación global, el Editor en Jefe enviará las evaluaciones al autor responsable para que corrija e incorpore en el manuscrito las diferentes observaciones o en su caso, indique su opinión a observaciones que considere discutibles en los dictámenes de los revisores. El manuscrito corregido por los autores deberá ser devuelto a la Revista, en un lapso no mayor a 30 días naturales; de otra forma se considerará como un manuscrito enviado por primera vez. De ser necesario el Comité Editorial podrá enviar nuevamente el manuscrito corregido a los revisores para tener una nueva ronda de revisión, todo ello continuando con el anonimato. Una vez aceptado el trabajo, las pruebas de galera, se enviarán al autor responsable.

Los archivos electrónicos se deberán enviar a la Revista de Educación Bioquímica como archivos adjuntos (reb@bq.unam.mx), con atención al Editor en Jefe y a partir de una dirección de correo electrónico que será considerada como la dirección oficial para la comunicación con los autores. El autor responsable deberá indicar plenamente su adscripción con teléfono, dirección electrónica y postal para comunicaciones posteriores. En el texto del mensaje enviado se deberá expresar la solicitud para considerar la posible publicación del artículo, el título del mismo, los nombres completos de los autores y su adscripción institucional, así como el número, tipo y nombre de los archivos electrónicos enviados. En el mismo texto se debe de aclarar que el trabajo no ha sido enviado a otra revista en su forma total o parcial, para su evaluación o que el mismo está en proceso de publicación en otra revista o en otro tipo de publicación. De igual manera se debe de aclarar que no existe conflicto de intereses entre los autores que envían el trabajo.