

## Bioinformática: Qué debe saber el personal del laboratorio clínico y para qué debe prepararse

Moderador: Gregory J. Tsongalis<sup>1\*</sup>

Expertos: Elizabeth Chao,<sup>2</sup> Jill M. Hagenkord,<sup>3</sup> Tina Hambuch,<sup>4</sup> y Jason H. Moore<sup>5</sup>

La iniciación en el laboratorio clínico de novedosas tecnologías de diagnóstico molecular que abarcan una amplia variedad de aplicaciones ha ocurrido a un ritmo sin precedentes y ha conducido a cambios revolucionarios en el área. Por ejemplo, las micromatrices se utilizan habitualmente en el laboratorio clínico como la prueba de elección para detectar anomalías generales relacionadas con el retraso mental y el autismo. Asimismo, las micromatrices que emplean millones de sondas para cada análisis se usan para la determinación de genotipos a gran escala y perfiles de expresión genética junto con algoritmos clínicos específicos. La secuenciación masiva en paralelo o de siguiente generación (NGS)<sup>6</sup> también se está volviendo habitual en el laboratorio clínico para la secuenciación holotómica, exómica y genómica dirigida. La cantidad de datos generados por estos análisis no tiene precedentes y requiere un sofisticado conocimiento de la bioinformática para el correcto almacenamiento, análisis y examen de estos conjuntos de datos. Si bien los laboratorios clínicos tienen experiencia en la informática y la manipulación de una gran cantidad de resultados, los sistemas usados para dichas tareas no resultan adecuados para el tratamiento de los datos de los estudios ómicos. En la presente publicación de Preguntas y respuestas, se invitó a investigadores académicos y de la industria, que habitualmente utilizan la bioinformática con fines de estudios ómicos, para analizar la importancia de la bioinformática y la manera en que el personal del laboratorio clínico puede prepararse mejor por su propia cuenta para poder manejar el creciente

volumen y la complejidad de los datos generados por sus laboratorios en la realización de estos estudios.

**La bioinformática y la bioestadística con frecuencia se utilizan indistintamente a pesar de sus considerables diferencias. ¿Puede definir esos términos e indicar cómo puede afectar la bioinformática en el laboratorio clínico?**



**Elizabeth Chao:** La bioestadística aplica la estadística a los temas biológicos. La atención se centra en el uso del análisis estadístico para diseñar experimentos que involucran grandes poblaciones. La bioestadística destaca los resultados significativos a través de la interferencia estadística

y así clasifica la señal del ruido. Por otra parte, la bioinformática es una ciencia interdisciplinaria moderna que agrupa la biología, la informática, las matemáticas aplicadas y la estadística en una disciplina. El principal objetivo de la bioinformática es el de desarrollar y usar programas informáticos para estudiar los procesos biológicos.

A pesar de sus diferencias, tanto la bioinformática como la bioestadística son fundamentales para el análisis de grandes conjuntos de datos. En los laboratorios

<sup>1</sup> Director of Molecular Pathology and Co-Director of the Translational Research Program, Geisel School of Medicine at Dartmouth, Dartmouth-Hitchcock Medical Center (Director de Patología Molecular y Codirector del Programa de Investigación de Transferencia, Facultad de Medicina de Geisel en Dartmouth), Lebanon, NH; <sup>2</sup> Director of Translational Medicine (Directora de Medicina de Transferencia), Ambry Genetics, Aliso Viejo, CA; <sup>3</sup> Chief Medical Officer and Senior Vice President (Directora General de Sanidad y Vicepresidenta Primera), Complete Genomics, Inc., Mountain View, CA; <sup>4</sup> Illumina Clinical Services Laboratory (Laboratorio de Servicios Clínicos Illumina), Illumina, Mountain View, CA; <sup>5</sup> Professor of Genetics and Community and Family Medicine, Director of the Institute for Quantitative Biomedical Sciences, Associate Director for Bioinformatics, Norris-Cotton Cancer Center, and Editor-in-Chief of BioData Mining, Geisel School of Medicine at Dartmouth, Dartmouth-Hitchcock Medical Center (Profesor de Genética y Medicina Comunitaria y Familiar, Director del

Instituto de Ciencias Biomédicas Cuantitativas, Director Asociado de Bioinformática, Centro para el Cáncer Norris-Cotton y Editor Jefe de BioData Mining, Facultad de Medicina de Geisel en Dartmouth, Centro Médico Dartmouth-Hitchcock), Lebanon, NH.

\* Dirigir correspondencia para estos autores a: Geisel School of Medicine at Dartmouth and the Dartmouth-Hitchcock Medical Center, One Medical Center Drive, Lebanon, NH, 03756. Fax 603-650-6120; correo electrónico gregory.j.tsongalis@hitchcock.org.

Recibido para la publicación el 1 de mayo de 2013. Aceptado para la publicación el 9 de mayo de 2013.

<sup>6</sup> Abreviaturas no estándar: NGS, secuenciación de siguiente generación; HIPAA, Ley de Portabilidad y Responsabilidad de Seguros de Salud de 1996; IOM, Instituto de Medicina; ACMG, Colegio Estadounidense de Genética Médica.

clínicos, la bioinformática es indispensable para ajustar el registro de instrumentos de alto rendimiento y conjuntos de datos genómicos.



**Jill M. Hagenkord:** En mi opinión, la bioinformática es la utilización de herramientas informáticas para estudiar conjuntos de datos biológicos complejos y de gran tamaño. La bioinformática es de carácter interdisciplinario e incluye elementos de la biología, química, matemática, in-

formática y estadística. Los ejemplos incluyen algoritmos para la determinación de la secuencia de ADN del exoma de millones de cifras fragmentadas superpuestas o la previsión del impacto biológico de una variante de secuencia de ADN en la función proteínica. Considero que la bioestadística es la aplicación de la estadística en los ensayos clínicos; por ejemplo, cómo proporcionar suficiente impulso a un estudio o cómo representar la importancia estadística de los resultados.

Estamos observando la migración de cada vez más métodos de pruebas “masivas en paralelo” de los laboratorios de investigación a los laboratorios clínicos. Las matrices citogenómicas, que básicamente reconstruyen los cromosomas dentro de la computadora a partir del ADN interrumpido hibridado en una micromatriz, ya constituyen la prueba de elección para varias alteraciones genéticas generales y se están utilizando ampliamente para las aplicaciones del cáncer. La NGS se está utilizando en la secuenciación hologenómica (estirpe germinal, cáncer y microbios) total o parcial. No existe prácticamente ninguna área de la patología que no vaya a verse afectada por los análisis de NGS en el futuro cercano. El diseño, la validación, la implementación y la interpretación de los análisis genómicos clínicos recaerán en el ámbito de la patología y cada paso requiere la amplia colaboración de la bioinformática. La totalidad de los datos de NGS son digitales; básicamente convertimos las bases en bytes y luego las visualizamos como bases nuevamente en la computadora. En comparación con los datos estadísticos, esta información digital interactúa fácilmente con algoritmos o bases de datos adicionales para ayudar al personal del laboratorio clínico a interpretar los datos en el contexto de las cuestiones clínicas de cada paciente. Es esencial que cada patólogo responsable o científico de laboratorio comprenda las suposiciones detrás de cada algoritmo y las relativas fortalezas y limitaciones de las bases de datos externas. Es probable que

cada laboratorio precise emplear uno o más bioinformáticos clínicos.



**Tina Hambuch:** La bioinformática es el procesamiento de conjuntos de datos biológicos y comprende procesos logísticos y analíticos. La bioestadística es específicamente un subconjunto de procesos analíticos centrado en la identificación de tipos específicos de patrones o

tendencias mediante el uso de herramientas estadísticas. A medida que el alcance de los datos se vuelve mayor, la bioinformática se vuelve más esencial. También permite que la ciencia de laboratorio clínico se vuelva más estandarizada y cuantitativa en términos del desempeño entre operadores y entre laboratorios. Sin embargo, dado que pocos laboratorios clínicos cuentan con un sólido grupo de bioinformática, este será un desafío importante para ellos en el corto plazo y el peligro de la aplicación incorrecta de la bioinformática es considerable. Por ejemplo, existen diferentes programas optimizados para la detección de diferentes tipos de variantes genéticas y su desempeño puede variar considerablemente; por lo tanto, es fundamental para realizar una aplicación adecuada comprender cómo se analizan los datos, así como las suposiciones y optimizaciones de las diferentes herramientas usadas para su análisis.



**Jason H. Moore:** La bioestadística es una disciplina formal que emplea las matemáticas para llevar a cabo la estimación puntual y el contraste de hipótesis con el objetivo de responder problemas biológicos y biomédicos. La estimación puntual se centra en la estimación precisa de

los parámetros demográficos, tales como la media y la varianza u otras medidas como la pendiente de una línea de regresión. El contraste de hipótesis se centra en el proceso formal de inferencia del contraste de una hipótesis nula acerca de uno o más parámetros. Comprender los conceptos y métodos básicos en bioestadística es esencial para el personal del laboratorio clínico así como los demás involucrados en las ciencias biomédicas.

La bioinformática es una disciplina relativamente nueva que combina la bioestadística con la informática para abordar problemas biomédicos complejos. La bioinformática se originó en la década de 1970 cuando se inició la necesidad de almacenar, administrar y analizar los datos de la secuencia del ADN. Luego prosperó en la década de 1990 con la propagación de Internet y las computadoras accesibles. Gran parte de la bioinformática se centra en el desarrollo, la evaluación y la aplicación de nuevas bases de datos y algoritmos informáticos para el análisis de datos de alta dimensión de tecnologías, tales como la espectrometría de masas y secuenciación de ADN. Este enfoque en la informática y sus diversas subdisciplinas, como el aprendizaje automático y la visualización, son lo que diferencia a la bioinformática de la bioestadística. Sin embargo, los buenos bioinformáticos pueden integrar métodos bioestadísticos sólidos a sus estrategias analíticas. Tanto la bioestadística como la bioinformática son esenciales para el ámbito del laboratorio clínico. Por ejemplo, la secuenciación de las muestras de ADN clínicas puede requerir métodos bioestadísticos formales para QC mientras que la bioinformática sería necesaria para almacenar los datos brutos seguido de la integración con las bases de datos de investigación o clínicas.

**En los laboratorios clínicos, las tecnologías de NGS y micromatriz representan herramientas de diagnóstico conocidas por generar grandes cantidades de datos. ¿Cuáles son algunas de las opciones para el análisis, el almacenamiento y la identificación de datos?**

**Elizabeth Chao:** Los laboratorios clínicos albergan infraestructuras informáticas de TI internas para interpretar y transmitir grandes cantidades de datos. Los proveedores pueden proporcionar soluciones que ofrecen una amplia variedad de soluciones de almacenamiento e informática de gran rendimiento para atender las necesidades particulares de los laboratorios.

**Jill M. Hagenkord:** Las micromatrices citogenómicas generan alrededor de un gigabyte de datos y la secuenciación hologenómica genera alrededor de un terabyte. Sin embargo, resulta poco frecuente que siquiera necesitemos regresar a los datos brutos y es suficiente, en mi opinión, con almacenar datos procesados. Los datos procesados para un genoma humano completo constituyen tienen unos pocos gigabytes, que resultan financieramente razonables de almacenar. Asimismo, la naturaleza ya ha encontrado la manera más económica de almacenar genomas: se denomina ADN. Por lo tanto, resulta más económico el almacenamiento de los archivos de datos procesados de mues-

tras clínicas en lugar de los datos brutos y luego el nuevo procesamiento del ADN en las ocasiones poco frecuentes en que sea necesario hacerlo. Es un modelo diferente del que estamos acostumbrados a usar en el laboratorio clínico pero conserva el espíritu de las exigencias del mantenimiento de los datos de las muestras clínicas.

Otra consideración es la de almacenar los datos genómicos en “la nube”, brutos o procesados, y acceder a estos de forma remota para redactar los informes según sea necesario. El informe final se incluye en el registro médico del paciente pero los datos genómicos se conservan en una base de datos que cumple con la ley Health Insurance Portability and Accountability Act (Ley de Portabilidad y Responsabilidad de Seguros de Salud, HIPAA) de 1996. Si se desean minimizar los costos, almacene el ADN en el congelador clínico y almacene en frío los datos procesados en la nube.

**Tina Hambuch:** Existen varias opciones disponibles y el desafío es identificar cuáles de estas son las más adecuadas. Esto dependerá de cómo intenta el laboratorio particular utilizar los datos y qué preguntas se realizan. Dicho esto, la estandarización de ciertos formatos, tales como BAM (alineación binaria/mapa) o VCF (formato de llamada variante), mejorará la situación en gran medida. Para el almacenamiento de datos, utilizamos Isilon (EMC) y para el análisis de datos, un Sun Grid Engine (clúster de cálculo). Si bien ciertos paquetes de software comerciales se están volviendo accesibles, utilizamos un conjunto de programas y secuencias de comandos personalizados.

**Jason M. Moore:** Dos de las herramientas más comúnmente utilizadas para el análisis de la secuencia de ADN son Galaxy y CLC Genomics Workbench. Galaxy es un paquete de software web gratuito que incluye docenas de herramientas de análisis para la manipulación de la secuencia de ADN, la alineación de QC, la detección de variantes, el análisis estadístico, etc. Ha surgido rápidamente como uno de los principales recursos para el análisis de la secuencia de ADN. CLC Genomics Workbench es un paquete de software comercial que desempeña una función muy similar con una interfaz gráfica de usuario mucho más intuitiva. Si bien CLC Genomics Workbench es costoso, resulta muy accesible para el personal ajeno a la bioinformática.

**¿Qué es la “nube”? ¿Es esta opción de almacenamiento de datos adecuada para los laboratorios que se rigen por restricciones financieras, HIPAA y otros asuntos normativos?**

**Elizabeth Chao:** La nube proporciona acceso sencillo a las capacidades de datos y procesamiento así como su

distribución. No presenta restricciones geográficas y permite la disponibilidad de sus servicios en forma permanente. Asimismo, es una solución comercial rentable destinada a las capacidades de procesamiento y almacenamiento a escala en respuesta a las necesidades de la empresa. Esta idea se analiza aún más al comparar las capacidades de red de la nube pública y privada. La principal diferencia es la infraestructura del host que predetermina la transmisión y el procesamiento de datos. La infraestructura pública utiliza la World Wide Web para comunicarse y es intrínsecamente un riesgo desde una perspectiva de la tecnología de la información. Por otra parte, un equipo interno de soporte de tecnología de la información responsable de salvaguardar los datos respalda la infraestructura privada. Este último requiere la adopción de rigurosos programas de seguridad de la información.

**Jill M. Hagenkord:** Una forma sencilla de conceptualizar la nube es considerar a Yahoo Mail o Gmail. Estos son servicios de correo electrónico web que se ejecutan en servidores remotos mediante software remoto. Como usuario final, no necesita instalar ni mantener el hardware ni el software. La nube permite a los usuarios finales utilizar solo el ancho de banda que necesitan (y pagar por este solamente) mientras aprovechan las economías de escala proporcionadas por el modelo de “usuario compartido”. Los expertos en software, TI, seguridad y centros de datos se encargan de la tecnología en la nube de modo que, en el caso de los laboratorios de hospitales, los usuarios finales puedan centrarse en el uso de la información genómica para el cuidado de los pacientes. Las reglamentaciones de HIPAA presentan ciertos obstáculos a la informática de la nube pero no son insuperables. Por ejemplo, para mantener el cumplimiento de HIPAA, los datos no deben abandonar el territorio estadounidense; no siempre resulta claro para el usuario final *dónde* se encuentra la nube en cualquier momento dado. Esta es solo una de las diversas formas en que las pruebas ómicas definen las políticas de cumplimiento de ayer bajo su responsabilidad. ¡Intente interpretar las directrices de CLIA de 1988 en el contexto del funcionamiento de una fábrica de genoma clínico compatible! Sin embargo, existen diversas opciones que pueden aprovecharse al implementar soluciones de nube. Pueden aprovecharse los servicios tales como las nubes privadas virtuales, el cifrado y el control de acceso para crear soluciones compatibles con las normativas de HIPAA y otras normativas. Actualmente existen diversos servicios en ejecución en la producción de la nube que cumplen con dichos requisitos.

La informática de la nube y la genómica continuarán desarrollándose, ganarán aceptación y formarán parte del futuro panorama de la medicina genómica.

Como profesionales de laboratorio, debemos estar al tanto de las políticas normativas actuales y garantizar que nuestras soluciones ómicas cumplan con dichas directrices, incluso cuando formamos a los responsables de las políticas sobre lo que sería más adecuado para las pruebas genómicas.

**Tina Hambuch:** Estamos analizando esto pero existen inquietudes considerables respecto de HIPAA. Finalmente, considero que permitirá la exploración y utilidad optimizada de la información que se genera pero, en el corto plazo, existen diversos desafíos considerables en relación con la seguridad de los datos y la capacidad para compartirlos de forma apropiada.

**Jason H. Moore:** La informática de la nube es un servicio de análisis y almacenamiento de datos que se ofrece a través de Internet. Las entidades públicas o privadas venden acceso a su hardware de almacenamiento de datos e informática de alto rendimiento a los usuarios que se encuentran geográficamente distribuidos en Internet. La ventaja de este modelo es que no necesita mantener ningún recurso informático por su cuenta y solo paga por los servicios que utiliza. Para un laboratorio clínico, esto podría significar usar la nube para todas las necesidades de almacenamiento de datos. Usted pagaría por los gigabytes usados y por el tiempo que desea almacenar los datos. Lo que obtiene es un almacenamiento de datos económico con cierta redundancia que brinda protección contra la pérdida de datos ocasionada por fallas de hardware. Ciertamente, la desventaja es que los datos que abandonan sus instalaciones pueden no cumplir con sus normas de privacidad y seguridad. Una vez que los datos se encuentran en un servidor externo, usted ya no tiene control sobre su protección. Muchas personas en los laboratorios clínicos no desean correr ese riesgo en este momento.

*En marzo de 2012, el Institute of Medicine (Instituto de Medicina, IOM) emitió un informe sobre la validación de conjuntos de datos generado mediante pruebas de alta complejidad y el posterior análisis (Evolution of Translational Omics, Evolución de la Ómica de Transferencia). ¿Qué piensa acerca de la reproducibilidad de los análisis de bioinformática y la forma en que los laboratorios clínicos pueden evitar estas cuestiones problemáticas presentadas en el informe?*

**Elizabeth Chao:** La reproducibilidad de los análisis de informática ha sido una cuestión inminente con la rápida introducción de la genómica en las tareas clínicas. Resulta alentador que tanto el IOM como el American College of Medical Genetics (Colegio Estadounidense de Genética Médica, ACMG) hayan decidido abordar



estas prácticas con tanta rapidez. El ACMG presentó declaraciones de políticas y se encuentra trabajando activamente en directrices específicas para la secuenciación exómica y genómica total.

Como líderes y pioneros en la adopción en este campo, nuestra empresa se basó en referencias internas para la confiabilidad y reproducibilidad, que son más rigurosas que aquellas del actual informe del IOM. Dicho esto, nos entusiasma que se pongan en práctica estos estándares mínimos para garantizar una mayor estandarización en el futuro. No debería pasarse por alto la inversión necesaria para alcanzar esta etapa y es más que útil.

**Jill M. Hagenkord:** Existen diferentes tipos de análisis ómicos, algunos más claros que otros. Los patrones de expresión pueden ser particularmente opacos para el usuario final y las variables deben controlarse rigurosamente para obtener resultados uniformes. Además, estos tipos de análisis requieren cuidadosos estudios de validación y utilidad clínica así como validación técnica. Otros tipos de análisis ómicos tienen rendimientos más familiares. Por ejemplo, una matriz citogenómica que produce un cariógrama, que presenta una eliminación del cromosoma 13q14 en el ADN de las células tumorales de un paciente con leucemia linfocítica crónica. Si bien existen algoritmos complejos que convierten el ADN interrumpido en el cariógrama informático, la representación de los datos es familiar (un cromosoma) y se corresponde con nuestra comprensión de la biología tumoral de la leucemia linfocítica crónica, que aporta cierta comprobación de validez. Dado que las matrices citogenómicas utilizadas de este modo son solo un método alternativo para detectar cambios de importancia clínica consolidada, los estudios de rendimiento diagnóstico o validación técnica pueden ser suficientes.

Los patólogos clínicos y científicos de laboratorio se capacitan durante años a fin de validar el uso clínico de las pruebas. Cumplimos con estándares muy altos en relación con este asunto. La bioinformática formará parte del diseño, la validación, la implementación y la interpretación del progreso de las pruebas desarrolladas en laboratorio. Los laboratorios clínicos que ofrecen pruebas genómicas deben tener bioinformáticos clínicos altamente involucrados en el proceso y en comunicación estrecha y permanente con el personal encargado de las pruebas de laboratorio y los directores médicos. Deben validarse los algoritmos y los resultados y se deben controlar sus versiones como parte de la prueba desarrollada en laboratorio.

**Tina Hambuch:** La reproducibilidad es fundamental para la precisión de los datos y nosotros, ciertamente evaluamos la reproducibilidad de las validaciones téc-

nicas de nuestro software bioinformático. Puede alcanzarse pero no darse por sentado.

**Jason H. Moore:** Los laboratorios clínicos tienen, de modo intencional, sumo cuidado de evitar los errores en sus mediciones. Se han puesto en práctica diversas garantías, controles y comprobaciones para garantizar la confiabilidad y precisión de los datos clínicos dado que la atención de los pacientes depende de estos. Los datos ómicos son intrínsecamente menos confiables debido a la naturaleza de la tecnología de alto rendimiento, como la secuenciación de ADN. Asimismo, los datos ómicos de gran escala requieren varias series de procesamiento para proporcionar información útil. La información útil luego se convierte en conocimiento mediante la aplicación de diferentes métodos de análisis de bioinformática y bioestadística. Cada paso de la tramitación del análisis desde QC al análisis final y la interpretación puede generar errores no intencionales. Por ejemplo, varios métodos de aprendizaje automático tienen numerosos ajustes que ejercen grandes efectos en los resultados. Resulta sencillo especificar e informar erróneamente un ajuste de forma que los resultados no serán válidos o posteriormente no podrán reproducirse.

Existe un movimiento en el campo de la bioinformática, y la informática más generalmente, para proporcionar el software y el ajuste preciso que se usaron en un conjunto de datos específico a fin de generar los resultados publicados. Cualquiera debería poder descargar los datos y el software para reproducir fácilmente un resultado. El cambio en la cultura del uso compartido de datos y metodología ha sido lento pero es necesario si deseamos confiar en los resultados publicados. Esta es otra buena razón por la cual es fundamental para el personal del laboratorio clínico contar con conocimientos sobre las tareas de la bioinformática y la bioestadística.

**Si solo tuviera un recurso disponible para capacitarse sobre la bioinformática, ¿cuál sería y por qué?**

**Elizabeth Chao:** El campo de la bioinformática evoluciona más rápido que los recursos educativos actuales. Por lo tanto, el aporte global de los investigadores y líderes en el campo en referencia a los nuevos métodos de análisis de datos continúa siendo un principal recurso educativo. Esto se traduce directamente a la asistencia a conferencias locales y nacionales y la adhesión a oradores en diversas universidades, así como las reuniones personales con expertos en el campo.

**Jill M. Hagenkord:** Es una pregunta difícil. No creo que se pueda aprender lo que es necesario saber sobre la genómica clínica de una fuente. Pero, dicho eso, creo

que la mejor forma de aprender es haciendo. Las sociedades profesionales deberían realizar talleres prácticos para los miembros interesados y los alumnos deberían presentar casos genómicos de público conocimiento regularmente en las conferencias sobre casos. Varias empresas genómicas y proveedores de software cuentan con conjuntos de datos disponibles públicamente para fines educativos, en caso de que su institución aún no ofrezca análisis genómicos. He tenido la suerte de mantenerme en estrecho contacto con diversos bioinformáticos muy inteligentes a fin de obtener ayuda para aclarar cuestiones sobre algoritmos y aparentes inconsistencias en la representación de los datos. El personal del laboratorio clínico debe esforzarse por capacitarse y mantenerse en contacto con bioinformáticos, así como ser pacientes y mantener su sentido del humor mientras los profesionales médicos y los bioinformáticos aprenden a comunicarse unos con otros. Tengo numerosas historias graciosas sobre lo difícil que puede ser la comunicación entre un médico y un bioinformático.

**Tina Hambuch:** No existe realmente un único recurso adecuado, el campo es quizás muy diverso y además muy incipiente. También es lo suficientemente complejo y las personas necesitan una capacitación integral que va más allá de un simple libro o sitio web. Asimismo, los recursos y la infraestructura informática son necesarios para el aprendizaje en este campo.

**Jason H. Moore:** El paquete de software de programación de estadística R es un enfoque ideal para el aprendizaje de la bioinformática. R es un programa de código abierto, está disponible libremente y se ejecuta en sistemas operativos Linux, Mac y Windows. La ven-

taja de R es que se ha convertido rápidamente en el primer paquete de software de análisis bioinformático. Esto se debe en parte a que es gratuito pero también a que es extensible. Diversos informáticos lanzan sus nuevos métodos como paquetes dentro de R que cualquiera puede descargar y usar inmediatamente. Existe amplia documentación en línea y una colección de herramientas bioinformáticas en un paquete denominado bioconductor. La curva de aprendizaje inicial de R es algo acentuada. Sin embargo, el tiempo invertido bien vale las puertas que abre el conocimiento de R. R incluye ahora paquetes para casi todo lo que le gustaría conocer sobre bioinformática y bioestadística.

---

**Contribuciones de los autores:** *Todos los autores confirmaron que han contribuido al contenido intelectual de este documento y han cumplido con los siguientes 3 requerimientos: (a) contribuciones significativas a la concepción y el diseño, la adquisición de datos o el análisis e interpretación de estos; (b) redacción o revisión del artículo en relación con su contenido intelectual; y (c) aprobación final del artículo publicado.*

**Declaración de los autores o posibles conflictos de interés:** *Tras la presentación del manuscrito, todos los autores completaron el formulario de declaración del autor. Declaraciones o posibles conflictos de interés:*

**Empleo o liderazgo:** E. Chao, Ambry Genetics; J. M. Hagenkord, InVita Corporation; T. Hambuch, Illumina.

**Papel del consultor o asesor:** No se declara.

**Propiedad de acciones:** T. Hambuch, Illumina.

**Honorarios:** No se declara.

**Financiamiento de la investigación:** No se declara.

**Testimonio de expertos:** No se declara.

**Patentes:** No se declara.

---

Previously published online at DOI: 10.1373/clinchem.2012.198226

---