



Valencia, lunes 10 de junio de 2024

El CSIC colidera un estudio pionero sobre la secuenciación larga del transcriptoma, clave en la enfermedad y el envejecimiento

- El Instituto de Biología Integrativa de Sistemas (CSIC-UV) impulsa un estudio internacional sobre los métodos para analizar datos obtenidos por secuenciación de lectura larga del transcriptoma
- Este es el sistema más utilizado hoy día para estudiar la complejidad de la vida. Los resultados, publicados en 'Nature Methods', revelan una diversidad de moléculas de ARN mucho mayor de la esperada



Molécula de ARNm, ilustración 3D. / iStock

Un grupo internacional de investigación coliderado por el Instituto de Biología Integrativa de Sistemas (I2SysBio), centro mixto del Consejo Superior de Investigaciones Científicas (CSIC) y la Universitat de València (UV), publica en *Nature Methods* el mayor estudio comparativo realizado hasta la fecha sobre los métodos que analizan datos obtenidos por secuenciación de lectura larga del transcriptoma humano. Se analizaron diferentes tecnologías y varias herramientas informáticas disponibles para la secuenciación de lectura larga de las moléculas del ARN, moléculas fundamentales para

que los genes cumplan su función. El estudio halló una mayor diversidad de ARN de la esperada, lo que podría tener grandes repercusiones en el estudio de enfermedades, envejecimiento y la misma complejidad de la vida en la Tierra.

Durante años, un consorcio internacional conocido como Proyecto de Evaluación de Anotación del Genoma de Lectura Larga de ARN-Seq (LRGASP, por sus siglas en inglés), evaluó los métodos y tecnologías en experimentos de secuenciación de lectura larga de ARN. Ahora, este consorcio global donde CSIC tiene un papel fundamental ha publicado los resultados de este esfuerzo, ofreciendo orientación para el futuro de la experimentación y el análisis de secuenciación de ARN. El trabajo, publicado en la revista *Nature Methods*, evalúa las fortalezas y debilidades de las dos principales plataformas de secuenciación de lectura larga de ARN, Oxford Nanopore Technologies y Pacific Biosciences, así como los métodos computacionales utilizados para evaluar los datos.

El ARN es el compuesto molecular de las células que transmite la información del ADN a las proteínas mediante los procesos de transcripción y traducción, que son universales para todos los seres vivos. La secuenciación de lectura larga del ARN permite observar moléculas completas de ARN e identificar pequeños cambios en la forma en la que los genes dan lugar a proteínas. Esos pequeños cambios son críticos para la constitución de organismos complejos como los seres humanos, y fallos en su síntesis se asocian a diversas enfermedades. La secuenciación de lectura larga del ARN se usa para identificar estos cambios y asociarlos con procesos biológicos diversos.

“Aunque el genoma humano ha sido secuenciado de punta a punta, aún nos enfrentamos a grandes desafíos para definir con exactitud cómo los genes dan lugar a la enorme diversidad de moléculas de ARN y proteínas que forman un ser vivo. Este conocimiento es muy importante, porque pequeños cambios en el paso de ADN a ARN pueden dar lugar a patologías”, explica **Ana Conesa**, profesora de investigación del CSIC en el I2SysBio y una de las investigadoras que han liderado este consorcio. Su equipo ha evaluado las predicciones de ARN propuestas por 14 laboratorios bioinformáticos de todo el mundo, utilizando el software SQANTI3 desarrollado por este grupo en el I2SysBio, una de las herramientas bioinformáticas de referencia en el campo.

Diversidad de ARN mayor de lo esperado

Se generaron más de 427 millones de secuencias de lectura larga que fueron analizados en el estudio. Los datos procedían de humanos, ratones y manatíes. El uso de datos de manatíes permitió probar los métodos en una especie sin un genoma de referencia. “Era importante probar las técnicas en una especie no modelo, ya que cada vez es más común ver estudios con secuenciación de lectura larga de ARN en estos organismos no tan bien estudiados. Esta falta de información previa debe ser tenida en cuenta durante los análisis porque puede afectar directamente a nuestros resultados”, asegura **Francisco J. Pardo Palacios**, investigador predoctoral del I2SysBio y primer firmante de este trabajo.

Tras una extensa recopilación y análisis de datos, el consorcio produjo un conjunto de recomendaciones para la secuenciación de ARN. En general, los enfoques de secuenciación de lectura larga funcionan mucho mejor que la secuenciación de lectura corta, siendo la calidad de las lecturas, más que su abundancia, el factor clave de

precisión. Además, encontraron una cantidad sorprendente de transcritos no documentados en los genomas humanos y de ratones. “Hemos visto que existe una diversidad de ARNs mucho más grande de lo pensábamos. Estamos viendo que cada individuo, incluso cada célula, tiene un transcriptoma propio y personal. El siguiente paso es averiguar la relevancia que tiene esto en la enfermedad, el envejecimiento y la diversidad de especies”, resume Ana Conesa.

El trabajo concluye que no existe un mejor enfoque único para la secuenciación de lectura larga de ARN. El artículo describe las mejores prácticas en función de los diferentes objetivos que puedan tener los estudios individuales. Las diferentes tecnologías existentes presentan diferencias en las tasas de error, rendimiento de secuenciación y la longitud de lectura, por lo que los investigadores deben priorizar cuál es más importante para su área de estudio. “Creo que esto ayudará a muchas personas que quieren desarrollar aún más la tecnología, todavía hay margen de mejora en muchos de estos métodos”, finaliza **Angela Brooks**, investigadora de la Universidad de California Santa Cruz (EE.UU.) y coautora del estudio.

Systematic assessment of long-read RNA-seq methods for transcript identification and quantification.

Nature Methods. DOI: 10.1038/s41592-024-02298-3

Isidoro García / CSIC Comunicación – Comunidad Valenciana

comunicacion@csic.es