

Resumen en castellano de la tesis

La presente tesis muestra la utilización de las lecturas largas para resolver las limitaciones asociadas al ARN-Seq habitual, presentando innovaciones significativas en este campo. Las lecturas largas permiten capturar transcritos completos y detectar nuevas variantes de *splicing*, mejorando los resultados obtenidos con lecturas cortas en términos de precisión ya que no existe la necesidad de realizar un ensamblado de lecturas que podría dar lugar a isoformas quiméricas.

En el marco de este trabajo, se ha desarrollado la herramienta SQANTI3, diseñada para la evaluación y filtrado de transcriptomas. SQANTI3 clasifica modelos de transcripción de lecturas largas según categorías estructurales basadas en sus *splice junctions* (SJ) y anota diversas características de calidad, tales como la presencia de SJ no canónicas o la fiabilidad de las anotaciones de los sitios de inicio y término de transcripción (TSS y TTS, por sus siglas en inglés) utilizando datos ortogonales. También ofrece un módulo de filtrado de artefactos basado en aprendizaje automático y reglas definidas por el usuario, así como un módulo de “rescate” para evitar la pérdida de genes completos por un filtrado excesivo. Por último, SQANTI3 integra la anotación funcional de los transcriptomas con isoAnnot Lite, facilitando el análisis de cambios en la expresión de isoformas y sus implicaciones funcionales.

SQANTI3 se utilizó en los retos 1 y 3 del proyecto LRGASP (Long-read RNA-seq Genome Annotation Assessment Project), un esfuerzo internacional y multicéntrico para el *benchmarking* de herramientas bioinformáticas de lecturas largas en ARN-Seq. Ambos retos se centraron en la identificación correcta de transcritos en organismos altamente anotados (reto 1) y en organismos no modelo con limitaciones de información *a priori* (reto 3). LRGASP proporcionó datos de diferentes tecnologías y protocolos a los participantes para que presentaran los resultados obtenidos sus herramientas bioinformáticas. Estos resultados se evaluaron y compararon utilizando SQANTI3, dejando patente las diferencias de transcriptomas obtenidos para una misma muestra dependiendo de los datos y métodos empleados.

En resumen, esta tesis resalta la importancia que la utilización de lecturas largas para ARN-Seq puede tener en el futuro y como SQANTI3 es y será una herramienta clave para la evaluación y mejora de la calidad de los transcriptomas.