



1. DATOS BÁSICOS DEL TFG:

Título: El uso de técnicas paired-end en la cuantificación de miRNAs

Descripción general (resumen y metodología):

En la cuantificación de microRNAs (miRNAs) mediante secuenciación masiva de nueva generación (NGS) se emplea generalmente el método de lectura única (single-end, SE), es decir, se secuencia solamente la hebra original del ARN monocatenario [1]. Por el otro lado, existen las técnicas de lectura emparejada (paired-end, PE), que secuencian el inserto desde ambos extremos generando dos lecturas (Read 1 o R1, y Read 2 o R2) [2]. Esto tiene importantes ventajas para resolver la estructura de intrones/exones en genes (RNA-seq) o detectar reordenamientos cromosómicos (WGS).

Dado que los miRNAs son moléculas extremadamente cortas, en un principio la información en ambas lecturas es la misma (secuencias complementarias e inversas), por lo que no se suele emplear PE en la secuenciación del ARN pequeño (small RNA-seq). No obstante, cabe la posibilidad de usar esta información como filtro de concordancia al exigir la condición analítica de que $R1 == R2$. De esta forma se pueden, por ejemplo, identificar y eliminar gran parte de los errores de secuenciación introducidos por la propia plataforma de lectura. Aunque el impacto de este filtrado sobre la tasa de errores técnicos es evidente, apenas se ha explorado a fondo cómo afecta a aspectos críticos del análisis de miRNAs, tales como la detección de variantes de secuencia o isomiRs, la identificación de posibles eventos de edición del ARN, o la estimación precisa de la abundancia transcripcional [3].

Actualmente, existen numerosos conjuntos de datos públicos de miRNAs que contienen secuencias R1 y R2, debido en gran medida a que muchos proveedores de servicios de secuenciación ya no ofertan la opción single-end por cuestiones de optimización de costes y rendimiento. A pesar de la disponibilidad de estos datos, el análisis se suele basar únicamente en las lecturas R1, ya que el ecosistema de software bioinformático popular para small RNA —como **sRNAbench** [4], **miRDeep2** [5] o **Chimira** [6]— está diseñado computacionalmente para trabajar de forma exclusiva con lecturas únicas.

En este Trabajo Fin de Grado se propone investigar exhaustivamente el impacto metodológico y biológico que supone emplear de forma conjunta la información de R1 y R2 en el análisis de small RNA-seq. El estudio se centrará en evaluar tres aspectos clave:

1. La estimación de la abundancia cuantitativa de los miRNAs.
2. La eficiencia en la corrección de errores mediante el uso de Identificadores Moleculares Únicos (UMIs).
3. La resolución y fidelidad en la detección de isomiRs.

Para ello, se utilizarán conjuntos de datos transcriptómicos públicos y la nueva versión del programa sRNAbench que permite el uso adaptado de lecturas R2. Se evaluará cómo la inclusión activa de la lectura R2 altera o mejora la inferencia biológica final, específicamente en los resultados de los análisis de expresión diferencial y en la caracterización de los patrones de isomiRs.

Tipología: Estudio de casos, teóricos o prácticos, relacionados con la temática del Grado.

Objetivos planteados:

- Recopilar estudios en los repositorios públicos (como GEO o SRA) que cumplan con los siguientes requisitos: i) que se hayan secuenciado con la tecnología paired-end, y ii) que existan al menos dos grupos biológicos diferentes (como tratado/placebo o sano/enfermo) para poder calcular la expresión diferencial.
- Procesar los conjuntos de datos públicos de small RNA-seq de manera comparativa, ejecutando el análisis con y sin la opción R2 disponible en la nueva versión del programa sRNAbench.
- Comparar cuantitativamente la abundancia estimada de miRNAs y la diversidad de isomiRs detectados utilizando únicamente R1 frente al uso combinado y filtrado de R1/R2.
- Determinar el efecto del análisis paired-end sobre la tasa de falsos positivos en la identificación fina de eventos de edición de ARN.
- Estimar la tasa de errores intrínsecos de secuenciación frente a los errores introducidos por la PCR y la variación biológica real presente en las muestras.

Bibliografía básica:

1. Motameny S, Wolters S, Nürnberg P, Schumacher B (2010). **Next-generation sequencing of small RNAs—a review of bioinformatics methods**. Briefings in Bioinformatics, 11(5): 451-470. doi: 10.1093/bib/bbq008.
2. Bentley DR, Balasubramanian S, Swerdlow HP, et al. (2008). **Accurate whole human genome sequencing using reversible terminator chemistry**. Nature, 456(7218): 53-59. doi: 10.1038/nature07517.
3. Sánchez-Herrero JF, Pluvinet R, Luna de Haro A, Sumoy L (2021). **Paired-end small RNA sequencing reveals a possible overestimation in the isomiR sequence repertoire previously reported from conventional single read data analysis**. BMC Bioinformatics, 22(1): 215. doi: 10.1186/s12859-021-04128-1.
4. Gómez-Martín C, Aparicio-Puerta E, Hackenberg M (2023). **sRNAtoolbox: Dockerized Analysis of Small RNA Sequencing Data in Model and Non-model Species**. Methods Mol Biol, 2585: 201-222. doi: 10.1007/978-1-0716-2982-6_13.
5. Friedländer MR, Mackowiak SD, Li N, Chen W, Rajewsky N (2012). **miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades**. Nucleic Acids Res, 40(1): 37-52. doi: 10.1093/nar/gkr688.
6. Vitsios DM, Enright AJ (2015). **Chimira: analysis of small RNA sequencing data and microRNA modifications**. Bioinformatics, 31(20): 3365-3367. doi: 10.1093/bioinformatics/btv380.

Recomendaciones y orientaciones para el estudiante:

Plazas: 1

2. DATOS DEL TUTOR/A:

Nombre y apellidos: MICHAEL HACKENBERG

Ámbito de conocimiento/Departamento: GENÉTICA

Correo electrónico: hackenberg@ugr.es

3. COTUTOR/A DE LA UGR (en su caso):

Nombre y apellidos:

Ámbito de conocimiento/Departamento:

Correo electrónico:

4. COTUTOR/A EXTERNO/A (en su caso):

Nombre y apellidos:

Correo electrónico:

Nombre de la empresa o institución:

Dirección postal:

Puesto del tutor en la empresa o institución:

Centro de convenio Externo:

5. DATOS DEL ESTUDIANTE:

Nombre y apellidos:

Correo electrónico: