



1. DATOS BÁSICOS DEL TFG:

Título: El impacto de la anotación de microRNAs sobre la inferencia biológica

Descripción general (resumen y metodología):

Los microRNAs son genes no-codificantes con importantes funciones en la regulación de la expresión génica (1). Para determinar su expresión mediante secuenciación masiva se requiere de una anotación, es decir conocer las secuencias de los genes de microRNAs. Existen varias bases de datos que anotan diferentes números de genes. Por ejemplo, miRBase (2) contiene 1917 genes en humano mientras MirGeneDB (3) anota solamente 514. Además, existe un alto número de publicaciones que reportan miles de microRNAs nuevos. En un principio podría parecer razonable anotar un alto número de microRNAs (aceptando la posibilidad de incluir falsos positivos) con la finalidad de no perder ninguno en el análisis.

No obstante, la presencia de falsos positivos impacta directamente en los niveles de expresión. Eso se debe a que no disponemos de métodos de alto rendimiento para cuantificar la abundancia absoluta de una molécula de ARN en una muestra. El nivel de expresión se suele estimar mediante una medida relativa llamado RPM (Reads per million, o lecturas por millón) que determina la frecuencia relativa de lecturas asignadas a un microRNA sobre el total de lecturas (4). Por lo tanto, la inclusión de falsos positivos (secuencias detectables que no son microRNAs) cambia los valores de expresión. Estos a su vez son el punto de partida para determinar los microRNAs que se expresan diferencialmente entre dos condiciones.

En este TFG se propone analizar sistemáticamente el impacto que tiene la anotación en la inferencia biológica, usando diferentes bases de datos, especies y métodos de normalización (5,6).

Tipología: Trabajo de investigación o desarrollo bioinformático

Objetivos planteados:

- Preparar las anotaciones de microRNAs obtenidas de diferentes bases de datos en 5 especies diferentes: Homo sapiens, Mus musculus, Drosophila melanogaster, Arabidopsis thaliana y Solanum lycopersicum
- Buscar en los repositorios públicos al menos 5 conjuntos de datos por especie cumpliendo los siguientes requisitos: existencia de al menos dos condiciones biológicas diferentes y 10 muestras por condición
- Determinar las matrices de conteo para los diferentes experimentos y anotaciones.
- Aplicar diferentes métodos de normalización y corrección de artefactos técnicos que darán lugar a las matrices de expresión.
- Determinar el número de microRNAs expresados diferencialmente, el estadístico asociado (magnitud de cambio, fold-change) y su significación estadística (valor P)

Bibliografía básica:

- 1) Bartel DP. (2018) Metazoan MicroRNAs. Cell. 2018 Mar 22;173(1):20-51. doi: 10.1016/j.cell.2018.03.006. PMID: 29570994; PMCID: PMC6091663.
- 2) Kozomara A, Birgaoanu M, Griffiths-Jones S. (2019) miRBase: from microRNA sequences to function. Nucleic Acids Res. 2019 Jan 8;47(D1):D155-D162. doi: 10.1093/nar/gky1141.
- 3) Clarke AW, Høye E, et. Al. (2025) MirGeneDB 3.0: improved taxonomic sampling, uniform nomenclature of novel conserved microRNA families and updated covariance models. Nucleic Acids Res. 2025 Jan 6;53(D1):D116-D128. doi: 10.1093/nar/gkae1094.

- 4) Cristina Gómez-Martín, Ernesto Aparicio-Puerta, Michael Hackenberg (2023) sRNAtoolbox: Dockerized Analysis of Small RNA Sequencing Data in Model and Non-model Species. Methods Mol Biol . doi: 10.1007/978-1-0716-2982-6_13.
- 5) Evans C., Hardin J., Stoebel D.M. Selecting between-sample RNA-Seq normalization methods from the perspective of their assumptions (2018) Briefings Bioinf. 2018; 19:776–792.
- 6) Chantal Scheepbouwer , Michael Hackenberg , Monique A J van Eijndhoven , Alan Gerber , Michiel Pegtel , Cristina Gómez-Martín (2023) NORMSEQ: a tool for evaluation, selection and visualization of RNA-Seq normalization methods Nucleic Acids Research, Volume 51, Issue W1, 5 July 2023, Pages W372–W378, <https://doi.org/10.1093/nar/gkad429>

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: