

1 Materiales y métodos

Para llevar a cabo el modelado de redes de coexpresión génica se ha utilizado el lenguaje R mediante el entorno RStudio. El análisis de coexpresión es un método de biología de sistemas utilizado para describir los patrones de correlación entre genes en muestras de microarrays. Las redes de correlación facilitan los métodos de cribado de genes basados en redes que se pueden utilizar para identificar posibles biomarcadores o dianas terapéuticas. A continuación, mostraremos y explicaremos los paquetes y funciones esenciales usados para llevar a cabo la realización de nuestro estudio.

WGCNA: Esta es la librería utilizada y en la que nos inspiraremos para el análisis de redes de coexpresión de genes ponderados. Este paquete podemos obtenerlo de Bioconductor instalando BiocManager en nuestro entorno. Hemos utilizado las siguientes funciones:

- `enableWGCNAThreads`: Estas funciones permiten y deshabilitan subprocesos múltiples para cálculos WGCNA que opcionalmente pueden ser multiproceso, lo que incluye todas las funciones que usan funciones `cor` o `bicor`.
- `pickSoftThreshold`: Análisis de topología libre de escala para múltiples poderes de umbral suave. El objetivo es ayudar al usuario a elegir una potencia de umbral suave adecuada para la construcción de la red.
- `adyacencia`: Calcula (mediante la correlación o distancia) la adyacencia de la red a partir de datos de expresión dados o de una similitud.
- `TOMsimilarity`: Cálculo de la matriz de superposición topológica, y la correspondiente disimilitud, a partir de una matriz de adyacencia dada.
- `label2colors`: Convierte un vector o matriz de etiquetas numéricas en un vector o matriz de colores correspondiente a las etiquetas.
- `moduleEigengenes`: Calcula los eigengenes del módulo (primer componente principal) de los módulos en un único conjunto de datos determinado.
- `mergeCloseModules`: Fusiona módulos en redes de expresión génica que están demasiado cerca según lo medido por la correlación de sus genes propios.
- `cor`: Estas funciones implementan un cálculo más rápido de la correlación de Pearson (ponderada).
- `TOMplot`: Representación gráfica de la matriz de superposición topológica utilizando un gráfico de mapa de calor combinado con el dendrograma de agrupamiento jerárquico correspondiente y los colores del módulo.
- `plotEigengeneNetworks`: Esta función traza representaciones de dendrogramas y genes propios de redes de genes propios (consenso). En el caso

de redes de genes propios de consenso, la función también traza medidas de preservación por pares entre redes de consenso en diferentes conjuntos.

- `exportNetworkToCytoscape`: Esta función exporta una red en archivos de lista de nodos y de borde en un formato adecuado para importar a Cytoscape.

cluster: Es el utilizado para hacer el agrupamiento de los datos. Hemos usado las siguientes funciones.

- `pam`: Esta función realiza un agrupamiento de los datos en k grupos "alrededor de medoides", una versión más robusta de K-means.
- `hclust`: Esta función realiza un análisis de agrupamiento jerárquico utilizando un conjunto de diferencias para los n objetos que se agrupan. Inicialmente, cada objeto se asigna a su propio grupo y luego el algoritmo procede de forma iterativa, en cada etapa uniendo los dos grupos más similares, continuando hasta que haya un solo grupo. En cada etapa, las distancias entre los conglomerados se vuelven a calcular mediante la fórmula de actualización de disimilitud de Lance-Williams de acuerdo con el método de conglomerado particular que se utilice.

DESeq2: Esta librería la usaremos para el análisis de datos de RNA-seq. Al igual que WGCNA, podemos obtenerla de Bioconductor instalando BiocManager en nuestro entorno. Las funciones usadas son:

- `DESeqDataSetFromMatrix`: Esta función es una subclase de `RangedSummarizedExperiment`, que se utiliza para almacenar los valores de entrada, cálculos intermedios y resultados de un análisis de expresión diferencial. La clase `DESeqDataSet` impone valores enteros no negativos en la matriz de "recuentos" almacenada como el primer elemento en la lista de análisis.
- `Counts`: La ranura de conteos contiene los datos de conteo como una matriz de valores de conteo de números enteros no negativos, una fila para cada unidad de observación (gen o similar) y una columna para cada muestra.
- `Deseq4`: Esta función realiza un análisis predeterminado a través de la estimación de factores de tamaño, a través de la estimación de dispersión y a través del ajuste GLM binomial negativo y estadísticas de Wald.

DCGL: El utilizado para el análisis de coexpresión diferencial y análisis de regulación diferencial de datos de microarrays de expresión genómica. Hemos utilizado las siguientes funciones.

- `qLinkfilter`: En esta función los enlaces genéticos con valores 'q' de pares de valores de coexpresión en cualquiera de las dos condiciones superiores al límite se retienen, mientras que los valores de coexpresión de otros enlaces se establecen en cero.

- WGCNA: El 'análisis de red de coexpresión de genes ponderados' pondera los vínculos con los coeficientes de correlación y compara las sumas de los coeficientes de correlación de un gen

coexnet: es el utilizado para la construcción de la red de coexpresión. Podremos obtenerla de BiocManager. Hemos utilizado las siguientes funciones.

- createNet: Esta, es una función que a partir de una secuencia biológica genera un grafo no direccionado teniendo como palabras vértices, pudiendo esto tener su parámetro de tamaño fijado por el parámetro 'palabra'. Las conexiones entre palabras dependen del parámetro 'paso' que indica la próxima conexión que se formará

También se han usado otras librerías de R que nos han facilitado el entendimiento de los resultados, así como la aplicación de estos para el uso de algunas funciones. Algunos son dplyr y base para la manipulación y el manejo de datos; grDevices para la manipulación de gráficos; Stats, para ciertas medidas estadísticas; entre otros.