

Pregunta 2.1

En cuanto a las ventajas y desventajas de los microarrays y el RNA-seq en estudios de expresión génica, hay varios puntos importantes a considerar.

Para los microarrays, sus ventajas incluyen un alto rendimiento, lo que permite analizar la expresión de miles de genes en una sola muestra. Además, inicialmente son más económicos en términos de costos de equipos y reactivos. Otra ventaja es la disponibilidad de numerosos conjuntos de datos públicos, lo que facilita su análisis y comparación.

Por otro lado, los microarrays presentan algunas limitaciones. No pueden detectar nuevos transcritos, ya que dependen de sondas de ADN diseñadas previamente. También tienen una menor sensibilidad y rango dinámico en comparación con el RNA-seq, lo que puede dificultar la detección de niveles bajos de expresión. Además, no proporcionan información precisa sobre la ubicación de las secuencias transcritas dentro del genoma.

Ahora, respecto al RNA-seq, sus ventajas son su mayor sensibilidad y rango dinámico. Puede detectar niveles de expresión más bajos y tiene la capacidad de identificar tanto transcritos conocidos como nuevos, lo que permite descubrir variantes y eventos de empalme alternativo. Además, proporciona información sobre la ubicación exacta de las secuencias transcritas y permite la detección de isoformas específicas.

Sin embargo, el RNA-seq también tiene sus desventajas. Es más costoso en términos de equipos, reactivos y análisis de datos. Requiere una mayor cantidad de ARN de alta calidad para su análisis, lo cual puede ser un desafío en muestras con cantidades limitadas de ARN. Además, genera grandes cantidades de datos, lo que implica el uso de capacidades de almacenamiento y análisis computacional más avanzadas.

Los microarrays pueden ser una opción adecuada si se busca analizar un gran número de genes de manera simultánea y se dispone de un presupuesto inicial limitado. Además, si se cuenta con datos públicos previos relevantes y se buscan resultados comparables con estudios anteriores, los microarrays podrían ser una elección viable.

Por otro lado, el RNA-seq ofrece una mayor sensibilidad, un rango dinámico más amplio y la capacidad de detectar tanto transcritos conocidos como nuevos. Además, proporciona información sobre la localización precisa de las secuencias transcritas y permite la detección de isoformas específicas. Si el estudio requiere una mayor profundidad de análisis, la identificación de variantes de transcritos o el descubrimiento de nuevos eventos de empalme alternativo, el RNA-seq podría ser la opción preferida.

Shi L, Reid LH, Jones WD, et al. The MicroArray Quality Control (MAQC)-II study of common practices for the development and validation of microarray-based predictive models. *Nat Biotechnol.* 2010;28(8):827-838. doi:10.1038/nbt.1665

Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet. 2009;10(1):57-63. doi:10.1038/nrg2484

Pregunta 2.2

Concepto de normalización: Tanto en el análisis de RNA-seq como en los microarrays, la normalización es un proceso esencial para corregir las diferencias técnicas entre las muestras y permitir la comparación de las expresiones génicas de manera precisa.

RNA-seq: En el análisis de RNA-seq, los datos de expresión génica se obtienen contando el número de fragmentos secuenciados que se alinean con cada gen. Estos fragmentos pueden tener longitudes diferentes y la profundidad de secuenciación puede variar entre las muestras.

Microarrays: En los microarrays, la expresión génica se mide indirectamente utilizando sondas de ARN que se hibridan con los ácidos nucleicos presentes en la muestra. La intensidad de la señal fluorescente capturada por las sondas se utiliza para inferir los niveles de expresión génica.

Diferencias en la normalización: Debido a las características intrínsecas de las tecnologías, la normalización en RNA-seq y en los microarrays difiere en los métodos utilizados.

Microarrays: En los microarrays, la normalización generalmente se realiza utilizando métodos como la suma global o métodos basados en la intensidad de la señal de los genes de control. Estos métodos asumen que la mayoría de los genes no cambian su expresión entre las muestras y se utilizan para corregir las diferencias técnicas entre los experimentos.

RNA-seq: En RNA-seq, se requiere una normalización más sofisticada debido a las diferencias en la longitud de los genes y la profundidad de secuenciación. Un método común de normalización en RNA-seq es el FPKM (fragments per kilobase of transcript per million mapped reads). Este método tiene en cuenta tanto la longitud del gen como la cantidad total de fragmentos mapeados en la muestra, permitiendo así la comparación de la expresión génica entre diferentes genes y muestras.

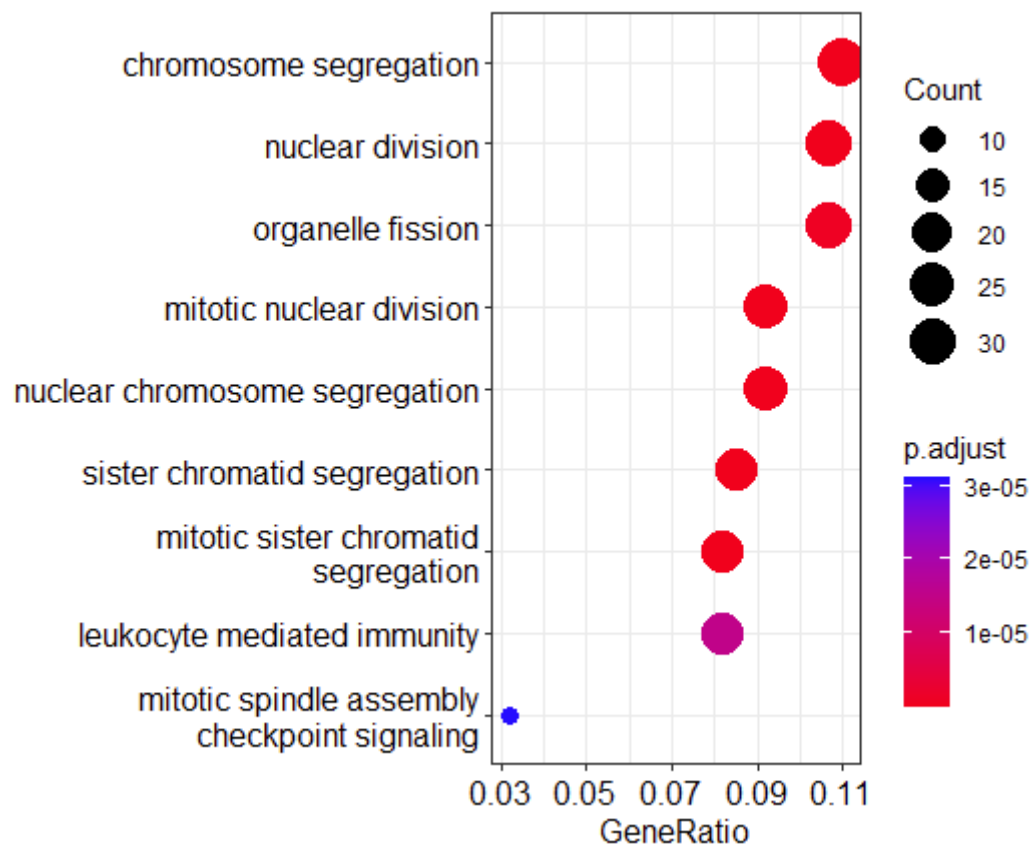
En resumen, la normalización en el análisis de RNA-seq y en los microarrays difiere en los métodos utilizados debido a las características intrínsecas de las tecnologías. Mientras que en los microarrays se utilizan métodos que asumen que la mayoría de los genes no cambian su expresión, en RNA-seq se requieren técnicas más avanzadas que tengan en cuenta las diferencias en la longitud de los genes y la profundidad de secuenciación para normalizar los datos de expresión génica.

Pregunta 2.3

Se ha logrado clonar el repositorio con éxito.

He realizado el filtraje de genes de una manera diferente a la mostrada en el repositorio de GitHub:

Con Deseq2:



Utilizando limma:

...1	ID	Description	GeneRatio	BgRatio
GO:0000070	GO:0000070	mitotic sister chromatid segregation	23/282	183/14423
GO:0140014	GO:0140014	mitotic nuclear division	26/282	260/14423
GO:0000819	GO:0000819	sister chromatid segregation	24/282	229/14423
GO:0007059	GO:0007059	chromosome segregation	31/282	385/14423
GO:0000280	GO:0000280	nuclear division	30/282	380/14423
GO:0098813	GO:0098813	nuclear chromosome segregation	26/282	298/14423
GO:0048285	GO:0048285	organelle fission	30/282	432/14423
GO:0002443	GO:0002443	leukocyte mediated immunity	23/282	306/14423
GO:0007094	GO:0007094	mitotic spindle assembly checkpoint signaling	9/282	42/14423
GO:0071173	GO:0071173	spindle assembly checkpoint signaling	9/282	42/14423
GO:0051784	GO:0051784	negative regulation of nuclear division	10/282	57/14423
GO:0033046	GO:0033046	negative regulation of sister chromatid segregation	9/282	44/14423
GO:0033048	GO:0033048	negative regulation of mitotic sister chromatid segregation	9/282	44/14423
GO:0045841	GO:0045841	negative regulation of mitotic metaphase/anaphase transition	9/282	44/14423

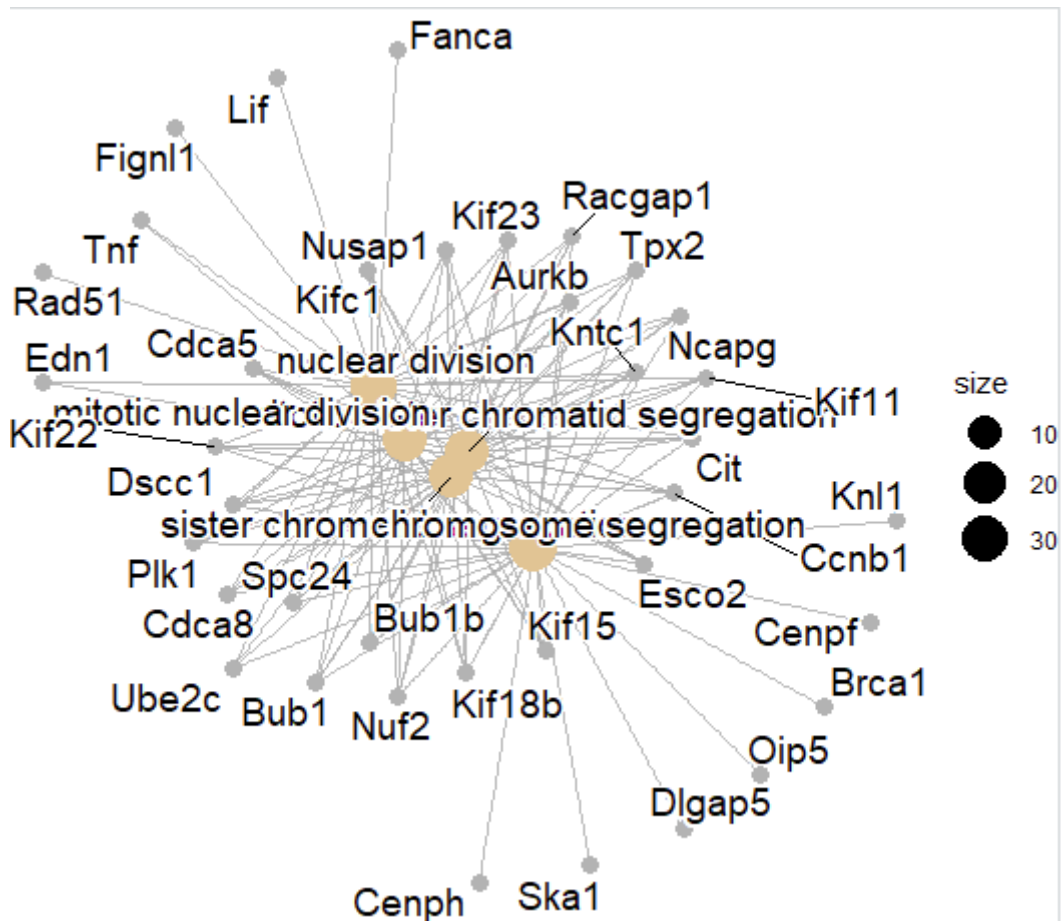
```
> head(ego_results)
```

```

      ID
GO:0000070 GO:0000070 mitotic sister chromatid segregation 23/282
GO:0140014 GO:0140014 mitotic nuclear division 26/282
GO:0000819 GO:0000819 sister chromatid segregation 24/282
GO:0007059 GO:0007059 chromosome segregation 31/282
GO:0000280 GO:0000280 nuclear division 30/282
GO:0098813 GO:0098813 nuclear chromosome segregation 26/282

```

No se han encontrado diferencias en los resultados obtenidos.



Se ha hecho una comparación entre el paquete limma y el paquete Deseq2 y se han obtenido los mismos resultados. En mi opinión Deseq2 es mucho más sencillo de utilizar porque una vez creas el objeto DeseqDataset, obtienes hasta las matrices de contraste. Haces todo en un solo paso.