



ENTREGA FASE II

Analisis de expresión diferencial en Fibrosis Quística.

Análisis de Datos Ómicos

Máster en Bioinformática

Autores:

Eva María López Fernández

Noelia Moreno González

Curso 2023/2024

Índice

1. Introducción	2
2. Metodología	3
2.1. Obtención de datos	3
2.2. Análisis de la calidad de los datos FastQ.	4
2.3. Alinear los readsRNA-Seq con el genoma humano	4
2.4. Calidad de alineamiento	4
2.5. Ensamblar los alineamientos	4
2.6. Conteo de lecturas	4
2.7. Obtención de transcritos expresados significativamente.	5
2.8. Análisis Funcional de genes diferencialmente expresados	5
3. Resultados	6
3.1. Analisis de la calidad de los datos FastQ	6
3.2. Comprobar la calidad del alineamiento	8
3.3. Obtención de genes expresados diferencialmente	8
3.3.1. Método Fold Change	9
3.3.2. Método NoiSEQ	9
3.4. Análisis funcional de genes expresados diferencialmente	11
3.4.1. Análisis de enriquecimiento GO	11
3.4.2. Análisis en Reactome	11
3.5. Conclusión	13
4. Bibliografía.	14

1. Introducción

La **Fibrosis Quística** es una enfermedad hereditaria monogénica, que se hereda de manera autosómica recesiva, está causada por una mutación en el gen CFTR. Ésta afecta al menos a 100.000 personas en todo el mundo.

El **gen CFTR** codifica una proteína que actúa como canal iónico epitelial facilitando el transporte de cloro y bicarbonato. Esta mutación provoca una alteración de la hidratación y eliminación de la mucosidad lo que lleva a complicaciones respiratorias y digestivas características de la enfermedad.

La Fibrosis Quística clásica se presenta con una serie de características distintivas, como infecciones pulmonares crónicas, problemas de digestión debido a la insuficiencia pancreática y posiblemente otras condiciones adicionales como la diabetes relacionada (CFRD) o enfermedades hepáticas específicas de la Fibrosis Quística [1].

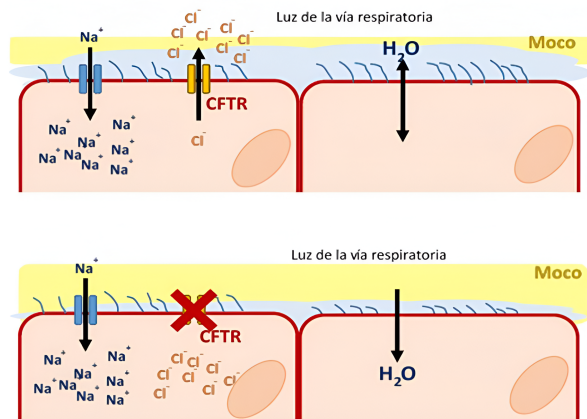


Figura 1: Representación esquemática de los canales iónicos de sodio y cloruro en la Fibrosis Quística, comparados con los mismos en células normales [2].

2. Metodología

Los ficheros y scripts generados en la terminal o de R se encuentran en el repositorio de GitHub (<https://github.com/noeliamorenog/ADO>), a excepción de los FASTQ por motivos de espacio. Todo lo generado a través de Galaxy se encuentra en el siguiente historial: (<https://usegalaxy.eu/u/eva.lopezf00/h/fibrosisquistica>).

2.1. Obtención de datos

Con el propósito de observar diferencias en la expresión en pacientes con Fibrosis Quística, hemos comparado los perfiles transcriptómicos de fibroblastos pulmonares en humanos con Fibrosis Quística y sanos, basándonos en un conjunto de datos obtenido de EMBL-EBI [3]. (<https://www.ebi.ac.uk/ena/browser/view/PRJNA593833>).

Este conjunto incluye los resultados de la secuenciación de ARN mediante el método RNA-seq de fibroblastos pulmonares afectados y sanos, cultivados en un entorno 2D hasta alcanzar aproximadamente el 100 % de confluencia. Este trabajo explota un novedoso modelo 3D *in vitro* del estroma pulmonar para demostrar las alteraciones morfológicas y moleculares en el tejido conectivo de las vías respiratorias con Fibrosis Quística. El modelo se obtuvo mediante el uso de fibroblastos pulmonares de Fibrosis Quística primaria y un enfoque ascendente de ingeniería de tejidos.

En cuanto al diseño general del estudio del artículo, se realizó un perfilado de expresión utilizando secuenciación de alto rendimiento para comparar los perfiles transcriptómicos, la plataforma de secuenciación utilizada fue Illumina novaSeq 6000 (*Homo sapiens*) El análisis se realizó mediante el método de secuenciación de ARN (RNA-seq). [4]

Así obtuvimos de este estudio 6 archivos fastq, correspondientes a tres repeticiones de una muestra de fibroblastos pulmonares humanos de individuos sanos, y tres repeticiones de una muestra de fibroblastos pulmonares de humanos afectados por Fibrosis Quística. [**Tabla 1**]

Nombre	Sample Accesion	Run Accesion	Condición
con_11	SAMN13492310	SRR10597211	Normal Human Lung Fibroblasts
con_12	SAMN13492309	SRR10597212	Normal Human Lung Fibroblasts
con_09	SAMN13492312	SRR10597209	Normal Human Lung Fibroblasts
mut_13	SAMN13492308	SRR10597213	Disease Human Lung Fibroblasts
mut_15	SAMN13492305	SRR10597215	Disease Human Lung Fibroblasts
mut_16	SAMN13492304	SRR10597216	Disease Human Lung Fibroblasts

Tabla 1: Muestras obtenidas para el estudio de expresión diferencial.

Para obtener los fastq se utiliza el comando *wget* en Mobaxterm aportando un fichero *wget.txt* con las

URLs ftp correspondientes a cada uno.

2.2. Análisis de la calidad de los datos FastQ.

Para evaluar la calidad de los archivos fastq, descargamos la herramienta **FastQC**, la cual genera un informe en formato HTML que detalla el análisis de calidad de estos archivos. El shell script utilizado para obtener los resultados se encuentra en *analisis_calidad_fastq.sh*

2.3. Alinear los readsRNA-Seq con el genoma humano

Para el alineamiento de las secuencias se ha utilizado la herramienta **Hisat2**, mediante **Galaxy**. La entrada para este programa serán nuestros archivos fastq y el genoma de referencia elegido. En este caso se utilizo el genoma de referencia GRCh38/hg38, obtenido en la base de datos **UCSC Genome Browser**. Todos los ficheros generados en este proceso se encuentran en la carpeta *Stringtie*. Tras este proceso se obtienen 6 archivos *.bam* los cuales se exportaron al servidor de *dayhoff*, con los que se obtuvieron los archivos *.bai* índices de alineamiento (BAI) mediante la herramienta **index** de **samtools**.

2.4. Calidad de alineamiento

Para comprobar la calidad de los alineamientos se utilizo ShellScript *flagstat.sh* donde se empleo la herramienta **Flagstat** de **samtools** , obteniendo unos ficheros (*flagstat_result_nombre_muestra.txt*) on las estadísticas de calidad del alineamiento de cada muestra.

Además, se utiliza **SeqMonk** para visualizar los resultados de los alineamientos y también comprobar la calidad de los mismos.

2.5. Ensamblar los alineamientos

Para el ensamblaje, se ha utilizado **StringTie** sobre cada archivo BAM obteniendo un fichero GTF para cada uno de ellos. Stringtie realiza un ensamblaje guiado por un fichero de anotación de referencia, en este caso se utilizó el genoma de referencia GRCh38/hg38 y se obtuvo las anotaciones del mismo (*encode.v38.annotation.gtf*), a través de la base de datos **UCSC Genome Browser** [5]. Esta herramienta se utilizó en la terminal del servidor, con el comando *stringtie -p 8 -G gencode.v38.annotation.gtf -o .gtf .bam*, indicando en *.gtf* el nombre del fichero que se genera y *.bam* el fichero que se toma como input.

2.6. Conteo de lecturas

Para asignar y contar la cantidad de lecturas de secuencias de ARN a elementos genómicos se utilizó la herramienta **featureCounts** a través de **Galaxy**. De ésta se obtiene un archivo tabular para cada muestra

(*count_n^oseq.tabular*), una de sus columnas contiene el Gene Id y la otra la expresión correspondiente en esa muestra a cada gen. También, se genera un resumen estadístico de la calidad del proceso.

2.7. Obtención de transcritos expresados significativamente.

Para llevar a cabo este proceso, se han seguido diferentes estrategias. En primer lugar, se ha utilizado el criterio del **Fold Change**. Para ello, se han utilizado diferentes herramientas de **Galaxy**, como:

1. **featureCounts**: herramienta bioinformática que se utiliza para medir la expresión génica de experimentos RNA-Seq a partir de archivos SAM o BAM. Para ello, se elige el archivo de anotaciones génicas del genoma de interés, en este caso hg38, y se seleccionan los BAM de los que se quiere medir la expresión génica.
2. **DESeq2**: herramienta bioinformática ampliamente utilizada para determinar genes diferencialmente expresados a partir de tablas de recuentos generadas en experimentos de RNA-Seq. Se deben especificar las diferentes condiciones del análisis, en este caso control y mutante, y qué muestras corresponden a cada una. Esto genera una tabla [Figura 2] con las siguientes columnas:

GeneID	Base mean	log2(FC)	StdErr	Wald-Stats	P-value	P-adj
387066	447.429365781581	3.50905081183423	0.264435531748315	13.2699671206594	3.45732409822154e-40	4.07964243590142e-36
1191	1618.65198839133	2.83077768530223	0.220727467579148	12.82476402393	1.19148776904047e-37	7.02977783733875e-34
6319	118.973290176683	3.68446981382968	0.334152845376806	11.0263008823848	2.85362417802874e-28	1.12242551002464e-24
715	299.116051420832	2.45866270437501	0.232701788237202	10.5657232933199	4.2964369461298e-26	1.26744889910829e-22
5673	85.986178571361	-3.78227533810836	0.402850813951324	-9.3887742239125	6.07025602612166e-21	1.43258042216471e-17

Figura 2: Tabla generada por DESeq2

Por último, se ha filtrado la tabla obtenida con DESeq2 para eliminar los genes cuya expresión era baja, estableciendo el límite en 200 lecturas. Por otro lado, se filtra el valor del log2(FC) para obtener los mayores de 0.5, siendo éstos los sobreexpresados, y menores de -0.5 para subexpresados. Todo esto se ha hecho a través de un script de R.

Otra de las estrategias seguidas para encontrar los genes expresados de manera diferencial es mediante el **método NOISeq**. Para esto, se ha adaptado el código de R proporcionado por el profesor en clase (*ExpresionDiferencialNoiseq2023.R*). Con esto, se obtienen los genes expresados diferencialmente en las 2 condiciones y se recupera información de BiomaRt.

2.8. Análisis Funcional de genes diferencialmente expresados

Con el fin de caracterizar funcionalmente los transcritos expresados diferencialmente, se utilizan las siguientes herramientas: análisis de enriquecimiento **GO** [6], utilizado para identificar las funciones biológi-

cas, los procesos celulares y las localizaciones subcelulares que están sobrerrepresentadas en el conjunto de genes en comparación con un fondo de referencia; y el análisis proporcionado por **Reactome**, para identificar y caracterizar los procesos biológicos y vías metabólicas en las que están involucrados los genes cuya expresión varía entre condiciones experimentales. [7].

3. Resultados

A continuación, se comentan los resultados obtenidos en el procedimiento de obtención de los datos para llevar a cabo el análisis de expresión diferencial.

3.1. Analisis de la calidad de los datos FastQ

En cuanto a los resultados obtenidos de los análisis de parámetros dados por FastQC, se muestran como ejemplo los gráficos generados de una muestra, siendo estos representativos de la tendencia general.

La calidad de las secuencias base, como se muestra en la **Figura 3**, es superior a 24 en todas las muestras. Esto se puede deber a que las muestras seleccionadas ya han sido sometidas a un preprocesamiento, por lo que no se ha realizado ningún filtrado para acortar las lecturas.

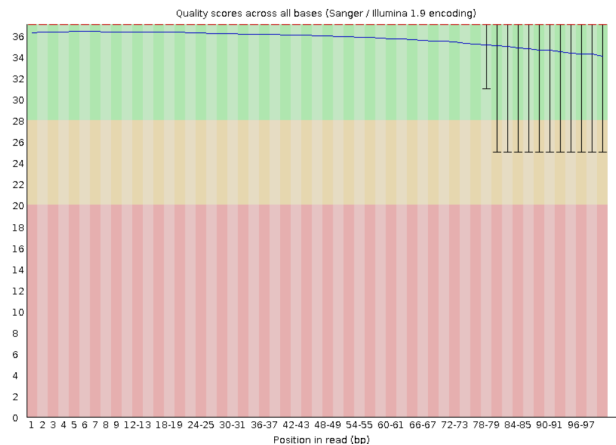


Figura 3: Gráfica de calidad por secuencia de base de SRR10597211.

Los demás parámetros, como las puntuaciones de calidad por secuencia [**Figura 4**], el contenido de N por base [**Figura 5**] y la distribución de longitud de secuencia [**Figura 6**] también muestran buenos resultados.

El contenido de bases por secuencia [**Figura 7**] es relativamente uniforme a lo largo de las secuencias en todas las muestras, esto indica que hay un equilibrio entre las bases nitrogenadas.

Las lecturas experimentales en el resultado de contenido de GC por base [**Figura 8**] presenta un buen alineamiento con las expectativas teóricas.

La presencia de adaptadores en la gráfica de contenido de adaptador [Figura 9], junto con la ausencia de secuencias sobreexpresadas, puede suponer que los adaptadores detectados no están ejerciendo un impacto significativo en la calidad de los datos. Esta observación puede deberse a la baja abundancia de adaptadores en las muestras o a su escasa influencia en el análisis global de secuencias. Por lo tanto, en este contexto, no se considera necesario realizar un filtrado de adaptadores, ya que su presencia no parece comprometer la integridad de los resultados obtenidos del análisis de los datos de secuenciación.

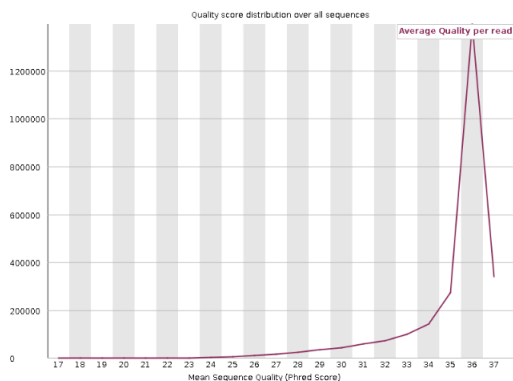


Figura 4: Gráfica de puntuaciones de calidad por secuencia de SRR10597211.

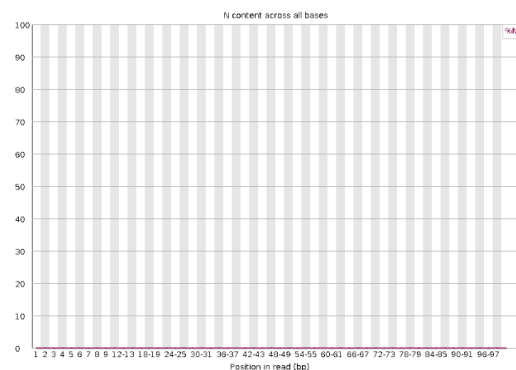


Figura 5: Gráfica de contenido de nitrógeno por base de SRR10597211.

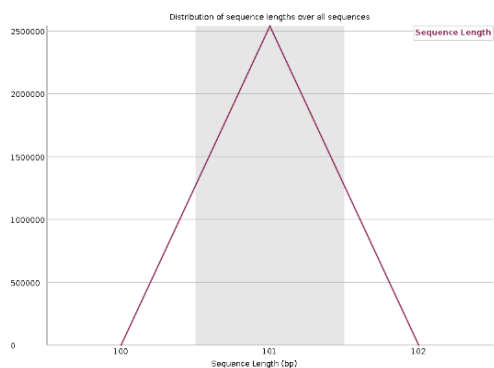


Figura 6: Gráfica distribución de longitud de secuencia de SRR10597211.

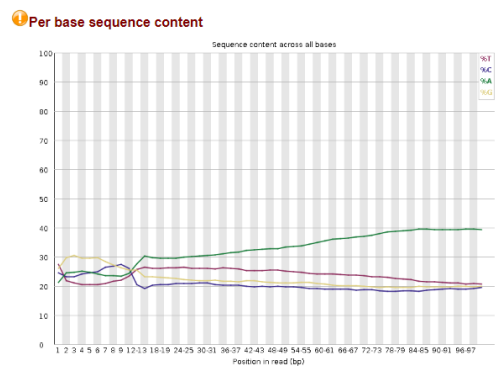


Figura 7: Gráfica de contenido de bases por secuencia de SRR10597211.

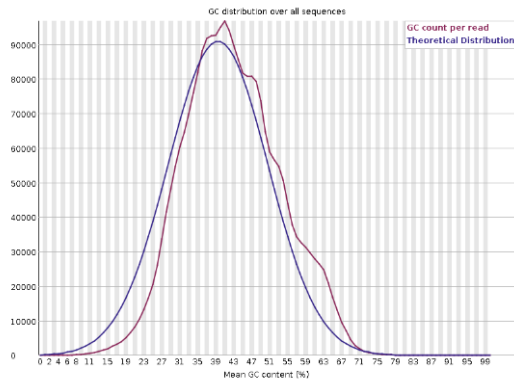


Figura 8: Gráfica de contenido de GC en SRR10597211.

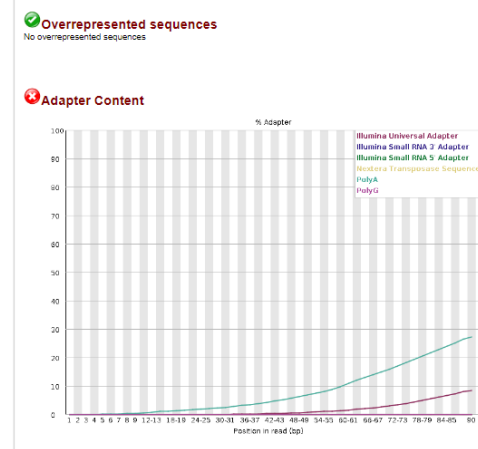


Figura 9: Contenido de adaptadores y secuencias sobreexpresadas en SRR10597211.

3.2. Comprobar la calidad del alineamiento

Los resultados que se obtuvieron con flagstat sugieren un exitoso alineamiento.[**Tabla 2**]

Muestra	Nombre archivo	Grupo	% de alineamiento	% de duplicados
SAMN13492308	estadisticas_13	Mutante	81.83 %	0 %
SAMN13492305	estadisticas_15	Mutante	82.61 %	0 %
SAMN13492304	estadisticas_16	Mutante	81.31 %	0 %
SAMN13492310	estadisticas_11	Control	78.82 %	0 %
SAMN13492312	estadisticas_09	Control	77.20 %	0 %
SAMN13492309	estadisticas_12	Control	82.55 %	0 %

Tabla 2: Resumen de los resultados de calidad del alineamiento

Se ha utilizado la herramienta **SeqMonk** para realizar una comparación entre las muestras control y mutantes frente al genoma de referencia. Además, se examinó la comparación entre las muestras control y mutantes entre sí. En la **Figura** se representa el alineamiento en el cromosoma 7 de una muestra control (**con.11.bam**) y una mutante (**mut.15.bam**) frente al genoma *Homo Sapiens* GRCh38. Se ha seleccionado ese cromosoma por ser donde se encuentra el gen CFTR en posición chr7:117465784-117715971, siendo este el causante de la Fibrosis Quística.

3.3. Obtención de genes expresados diferencialmente

Para ambas estrategias, se han preprocesado los datos obtenidos, imputando aquellos valores cuya expresión en ambas condiciones es nula. Por otro lado, también se han obviado los genes cuya expresión en ambas condiciones era pequeña.

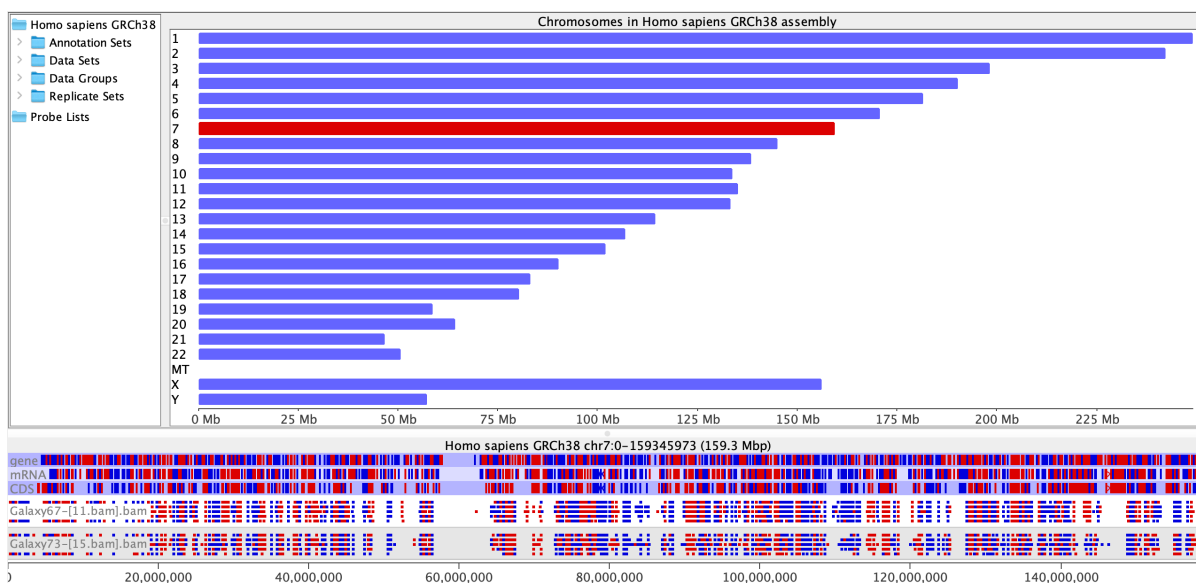


Figura 10: Alineamiento del cromosoma 7 en SeqMonk.

3.3.1. Método Fold Change

El Fold Change (FC) es una medida utilizada comúnmente en análisis de expresión génica para cuantificar la diferencia en la expresión de un gen entre dos condiciones o grupos experimentales. Esta medida proporciona información sobre la magnitud del cambio relativo en la expresión génica entre las condiciones comparadas. Se calcula como el cociente entre los niveles de expresión de un gen en las dos condiciones comparadas.

Una vez obtenidos los valores de $\log_2(\text{FC})$, se encuentra un total de 187 genes expresados diferencialmente, siendo 80 sobreexpresados y 107 subexpresados. A continuación, se muestran los 5 más sobreexpresados [Figura 11] y los 5 más subexpresados [Figura 12].

symbol	Gene_ID	base mean	$\log_2(\text{FC})$
MT1L	4500	286.9416	1.8519194
S100A4	6275	341.9826	1.6638186
TGM2	7052	219.7829	1.4975340
RND3	390	703.8204	1.4836489
FAM107B	83641	299.3715	1.4765669

Figura 11: Genes sobreexpresados con FC, orden descendente.

symbol	Gene_ID	base mean	$\log_2(\text{FC})$
SNHG5	387066	447.4294	-3.5090503
CLU	1191	1618.6520	-2.8307772
C1R	715	299.1161	-2.4586623
IGFBP5	3488	1255.4356	-2.0849963
SCARA3	51435	610.5197	-1.9588857

Figura 12: Genes subexpresados con FC, orden ascendente.

3.3.2. Método NoiSEQ

El método NOISeq es un enfoque utilizado en el análisis de datos de expresión génica para identificar genes diferencialmente expresados en experimentos de alto rendimiento. Se destaca por ser robusto frente

a la heterogeneidad en los datos y por ser útil cuando la replicación experimental es limitada.

Con esta estrategia, se obtienen un total de 249 genes expresados diferencialmente, siendo 57 sobreexpresados y 192 subexpresados. A continuación, se muestran los 5 más sobreexpresados [Figura 13] y los 5 más subexpresados [Figura 14].

	control_mean	mutante_mean	M
RND3	1103.0000	415.3333	1.4090912
MKI67	480.6667	223.0000	1.1079930
RAI14	887.3333	414.3333	1.0986843
BUB3	437.3333	208.3333	1.0698396
VIM	8156.0000	4015.3333	1.0223420

Figura 13: Genes sobreexpresados con NoiSEQ, orden descendente.

	control_mean	mutante_mean	M
CLU	405.3333	3108.3333	-2.9389605
IGFBP5	504.6667	2232.0000	-2.1449343
SCARA3	237.3333	1025.6667	-2.1115753
RPS28	621.3333	2538.6667	-2.0306316
PCOLCE	210.0000	855.0000	-2.0255351

Figura 14: Genes subexpresados con NoiSEQ, orden ascendente.

Al comparar los genes obtenidos con ambos métodos, se observa que son bastante similares. Se decide confiar más en los datos proporcionados por NoiSEQ debido a varios motivos, siendo uno de ellos que tiene en cuenta la variabilidad en la expresión génica entre las réplicas biológicas, lo que puede proporcionar una evaluación más precisa de la significancia estadística de los cambios en la expresión.

Esta sobreexpresión se puede visualizar gráficamente, representando en un Volcano plot [Figura 15] la relación entre la magnitud del cambio en la expresión génica y la significancia estadística de ese cambio.

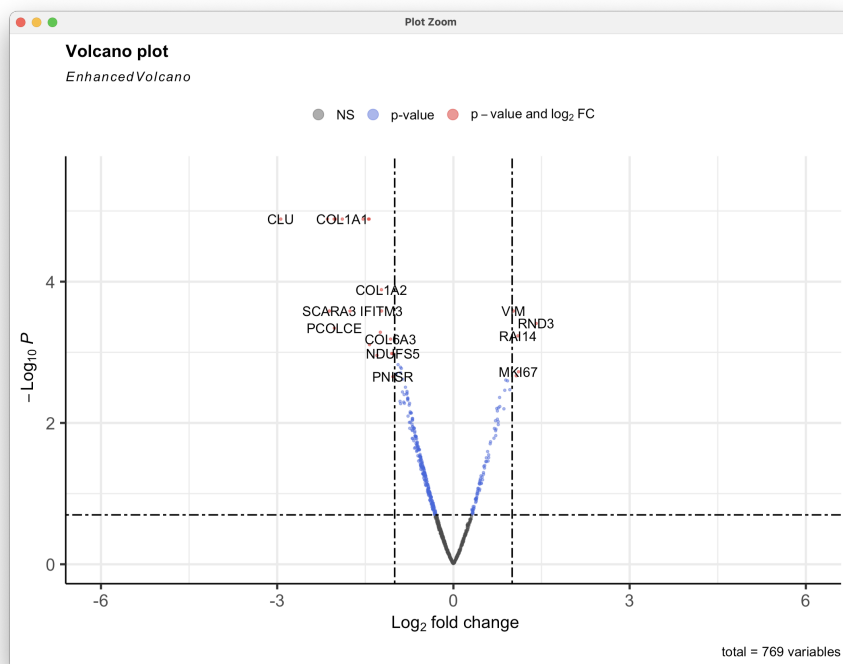


Figura 15: Volcano plot.

Cada punto representa un gen individual, cuya posición indica el tamaño del fold change y el valor de p

asociado con ese gen. Los que están más arriba en el eje Y y más a la izquierda serían los subexpresados, mientras que los de arriba a la derecha representan los genes sobreexpresados. Por otro lado, la línea horizontal indica el umbral de significancia estadística y las verticales los umbrales establecido para fold change.

3.4. Análisis funcional de genes expresados diferencialmente

El último paso de este análisis es caracterizar funcionalmente los genes obtenidos, proceso en el cual se asignan funciones biológicas o roles específicos a éstos.

Para ello, se han seleccionado dos herramientas de las propuestas para esta tarea: análisis de enriquecimiento **GO** y el análisis en **Reactome**.

3.4.1. Análisis de enriquecimiento GO

Al realizar el análisis filtrando por procesos biológicos, se observa un enriquecimiento significativo en procesos relacionados con la degradación de proteínas, como la dirección de proteínas al lisosoma involucradas en la autofagia mediada por chaperonas, así como en procesos metabólicos celulares generales. Además, se identifica un enriquecimiento en procesos relacionados con la localización celular y la regulación del transporte de proteínas a organelos específicos, como localización celular y localización de proteínas a organelo.

A continuación, se muestran en la [Figura 16] los procesos biológicos en los que se expresan diferencialmente los genes obtenidos y, en la [Figura 17], se detallan los diferentes procesos celulares, ya que esta es la categoría más abundante dentro de los procesos biológicos.

Se observa como, dentro de los procesos celulares, los más destacados son las actividades celulares relacionadas con el metabolismo y la organización y formación de estructuras celulares. En el contexto de la Fibrosis Quística, las disfunciones en el transporte de iones causadas por mutaciones en CFTR pueden alterar el equilibrio de iones en las células epiteliales, lo que afecta directamente a varios procesos metabólicos celulares. Por otro lado, las alteraciones en la función de CFTR pueden tener efectos significativos en la estructura y función de las células epiteliales afectadas, como puede ser la acumulación de moco espeso en los pulmones.

3.4.2. Análisis en Reactome

Se ha llevado a cabo el análisis funcional mediante Reactome, con el fin de identificar los diferentes pathways en los que participan el conjunto de datos de genes expresados diferencialmente.

La vía más relevante descrita es la elongación de la cadena peptídica, que implica la formación de

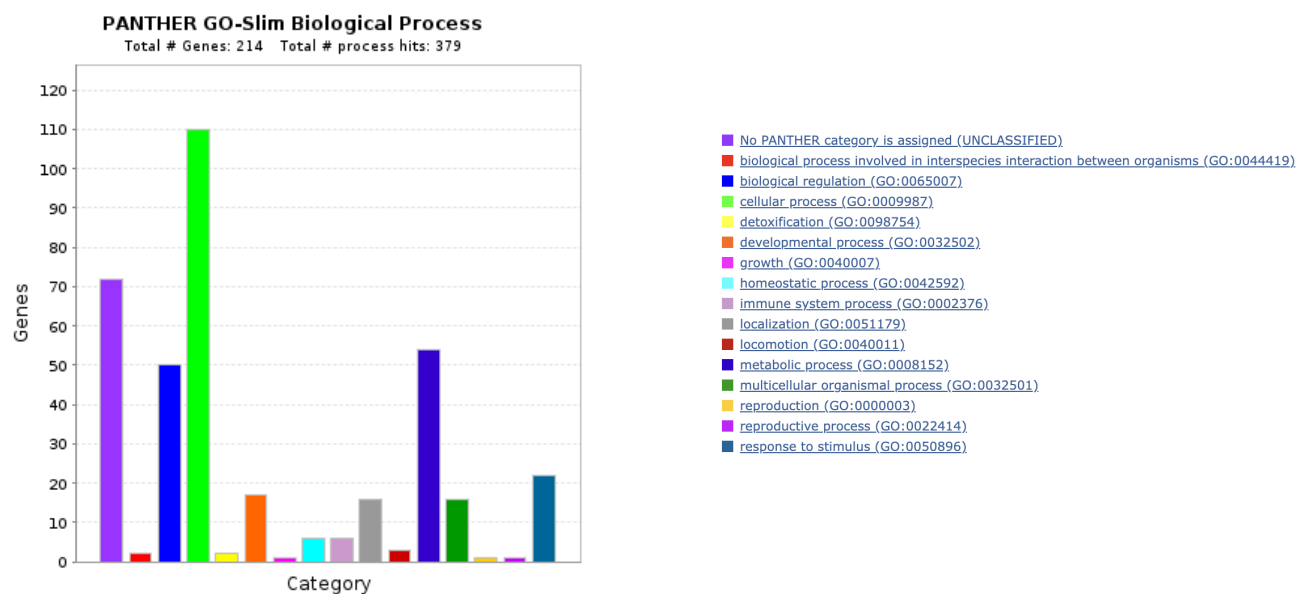


Figura 16: Representación gráfica de los procesos biológicos en los que participan los genes, junto con la leyenda correspondiente.

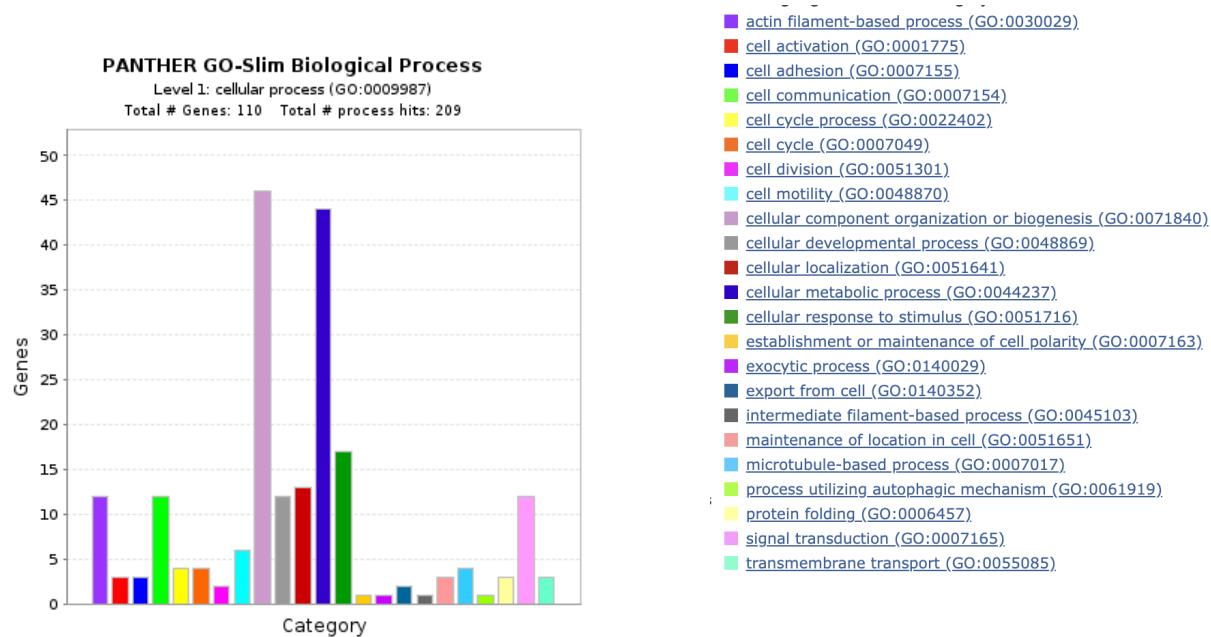


Figura 17: Procesos biológicos en los que los genes participan.

Pathway name	Entities				Reactions	
	found	ratio	p-value	FDR*	found	ratio
Peptide chain elongation	46 / 105	0.005	1.11e-16	6.33e-15	5 / 5	3.96e-04
Nonsense Mediated Decay (NMD) independent of the Exon Junction Complex (EJC)	45 / 123	0.005	1.11e-16	6.33e-15	1 / 1	6.72e-05
Formation of a pool of free 40S subunits	47 / 111	0.005	1.11e-16	6.33e-15	2 / 2	1.34e-04
Response of EIF2AK4 (GCN2) to amino acid deficiency	47 / 304	0.013	1.11e-16	6.33e-15	16 / 16	0.001

Figura 18: Pathways más relevantes ordenadas por p-valor dado por reactome.

enlaces peptídicos en el compartimento celular del citosol, un proceso crucial en la síntesis de proteínas. En la Fibrosis Quística, esta actividad podría estar implicada en la sobreexpresión de proteínas asociadas con la regulación de la respuesta inflamatoria y la remodelación tisular, las cuales son características prominentes de la enfermedad.

3.5. Conclusión

En este análisis se esperaba encontrar una expresión diferencial destacada en el gen CFTR, el principal causante de la Fibrosis Quística. Sin embargo, ni las muestras de fibroblastos sanos ni las de pacientes con Fibrosis Quística expresaron CFTR, las diferencias observadas entre los sanos y afectados no están asociadas con la disfunción de CFTR *in vitro* sino con otros factores probablemente derivados de la condición pulmonar *in vivo* específica del paciente en el momento de la biopsia de la que se extrajeron las células primarias. A pesar de ello, sí que se han observado genes expresados diferencialmente en ambas condiciones que pueden estar relacionados con la patología.

Se encontró entre los genes sobreexpresados el gen TGM2 que codifica la proteína Transglutaminasa-2, enzima que juega un papel crucial en la formación de enlaces cruzados entre proteínas, lo que puede afectar la estructura y función de tejidos y órganos. En el contexto de la Fibrosis, se ha encontrado que la actividad de la Transglutaminasa-2 está implicada en la modulación de la respuesta inflamatoria, el proceso de remodelado tisular y la deposición de matriz extracelular [8].

Otro de los genes sobreexpresados fue el gen RND3, también conocido como RhoE, es parte de la familia de genes RND y desempeña un papel en la regulación del citoesqueleto y la señalización celular. Aunque su relación directa con la Fibrosis Quística no está completamente establecida, estudios sugieren que RND3 puede estar implicado en procesos clave asociados con la misma, como la migración celular y la remodelación del tejido [9].

4. Bibliografía.

1. Shteinberg, M., Haq, I. J., Polineni, D., & Davies, J. C. (2021). Cystic Fibrosis. *Lancet (London, England)*, **397**(10290), 2195–2211. [https://doi.org/10.1016/S0140-6736\(20\)32542-3](https://doi.org/10.1016/S0140-6736(20)32542-3)
2. Guerra-Morillo, M. O., Rabasco-Álvarez, A. M., González-Rodríguez, M. L., Guerra-Morillo, M. O., Rabasco-Álvarez, A. M., & González-Rodríguez, M. L. (2020). Fibrosis Quística: Tratamiento actual y avances con la nanotecnología. *Ars Pharmaceutica (Internet)*, **61**(2), 81-96. <https://doi.org/10.30827/ars.v61i2.11358>
3. European Bioinformatics Institute. [ENA Browser]. (s. f.). Recuperado 10 de abril de 2024, de <https://www.ebi.ac.uk/ena/browser/home>
4. Mazio, C., Scognamiglio, L. S., De Cegli, R., Galietta, L. J. V., Di Bernardo, D., Casale, C., Urciuolo, F., Imparato, G., & Netti, P. A. Intrinsic Abnormalities of Cystic Fibrosis Airway Connective Tissue Revealed by an In Vitro 3D Stromal Model. *Cells* 2020, **9**, 1371. <https://doi.org/10.3390/cells9061371>
5. UCSC Genome Browser Downloads. (s. f.). Recuperado 13 de abril de 2024, de <https://hgdownload.soe.ucsc.edu/downloads.html>
6. Gene Ontology Resource. (s. f.). Gene Ontology Resource. Recuperado 15 de abril de 2024, de <http://geneontology.org/>
7. Reactome — Analysis Tools. (s. f.). Recuperado 15 de abril de 2024, de <https://reactome.org/PathwayBrowser/#TOOL=AT>
8. Nguyen, T. N., Suzuki, H., Yoshida, Y., Ohkubo, J.-I., Wakasugi, T., & Kitamura, T. (2022). Decreased CFTR/PPAR γ and increased transglutaminase 2 in nasal polyps. *Auris, Nasus, Larynx*, **49**(6), 964-972. <https://doi.org/10.1016/j.anl.2021.10.006>
9. Scoffone, V. C., Trespidi, G., Barbieri, G., Irudal, S., Perrin, E., & Buroni, S. (2021). Role of RND Efflux Pumps in Drug Resistance of Cystic Fibrosis Pathogens. *Antibiotics*, **10**(7), 863. <https://doi.org/10.3390/antibiotics10070863>