



Análisis de expresión diferencial en el cromosoma 1 de *Saccharomyces cerevisiae* en fermentación y quimiostato

Nasim Salim*, Nikola Tesla†, Rosalind Elsie Franklin*, Charles Robert Darwin†, Margaret Oakley Dayhoff†, Carl Edward Sagan*

*Department of Computer Science and Artificial Intelligence Massachusetts Institute of Technology 32 Vassar Street, Cambridge, MA 02139, USA

†Department of Genetics and Genomic Sciences Icahn School of Medicine at Mount Sinai New York, NY 10029, USA

Autor de correspondencia: Carl Edward Sagan cesagan@ingenieria.uner.edu.ar

Resumen

En este trabajo se analizaron cambios en la expresión de genes en el cromosoma 1 de *Saccharomyces cerevisiae* en un estado de recursos escasos, batch, y en quimiostato. Se utilizaron diversas herramientas bioinformáticas para su análisis, partiendo desde la obtención de los datos en una base de datos, preprocesamiento, control de calidad, alineamiento, análisis de la expresión diferencial y enriquecimiento de los genes resultantes. Se encontraron cambios en múltiples vías de señalización compleja, más allá de las que regulan la producción energética, como fueron rutas de síntesis de aminoácidos, modificaciones de proteínas y homeostasis de lípidos, lo cual es consistente con trabajos anteriores, y resalta la importancia de los genes localizados en el cromosoma 1 durante la adaptación de *Saccharomyces cerevisiae* a diferentes condiciones de entorno.

Palabras clave

transcriptoma, levadura, expresión, vías

I. INTRODUCCIÓN

S. cerevisiae ha sido utilizada durante más de un siglo en la industria de alimentos, como en la producción de pan y cerveza, debido a sus propiedades fermentativas[1]. Además, este microorganismo se ha convertido en un modelo de estudio para dilucidar los mecanismos moleculares que ocurren durante el ciclo celular en células eucariotas[2], [3]. En la producción de bioetanol de segunda generación, *S. cerevisiae* se ha mejorado mediante ingeniería evolutiva para aumentar su eficiencia en la producción de compuestos de interés[4]. También se ha empleado en la biotecnología para la producción de insulina, anticuerpos y albúmina, entre otros productos.[5]

Ha habido muchos estudios transcriptómicos de *S. cerevisiae* en diferentes condiciones de entorno[6], donde se resalta la gran importancia de respuestas al estrés como mecanismos de adaptación a cambios nutricionales en el entorno mediante vías de señalizaciones[7], [8], donde muchos genes como SNF1, GAL, entre otros, se han analizado en detalle por su implicación en muchas de estas vías[9], [10].

El objetivo de este trabajo es estudiar la expresión de genes del primer cromosoma de *Saccharomyces cerevisiae*, en particular de la cepa CEN.PK 113-7D[11] bajo dos condiciones metabólicas diferentes: metabolismo respiro-fermentativo (batch), donde la levadura consume azúcares para producir dióxido de carbono y alcohol, este metabolismo es útil para producir productos de fermentación como cerveza o vino, o totalmente respiratorio (quimiostato), donde la levadura emplea exclusivamente el oxígeno presente para producir energía a través de la respiración

celular, lo que resulta en un crecimiento más lento, pero una mayor producción de biomasa, este metabolismo es útil para producir levadura, para su uso en la fabricación de pan o para la producción de proteínas recombinantes en la industria farmacéutica.

II. MÉTODOS

Especificaciones

Se usó como sistema operativo Ubuntu 22.0.2 mediante WSL2 5.15.90.1 de Windows 10 Pro, en una computadora de escritorio con un procesador Intel(R) Core(TM) i5-7400 de 3.00GHz y 16GB de RAM.

Fuente de datos

Los datos de secuenciación para el análisis fueron obtenidos de un estudio anterior[12], junto con el genoma de referencia y anotación.

Software

Para el análisis exploratorio inicial de los datos se utilizó comandos bash y también la librería FASTQC[13] para generar reportes que luego fueron agrupados por MultiQC[14].

Para el alineamiento se emplearon STAR[15], como herramienta principal, e HISAT2[16], para tener una validación de los resultados. Se empleó Qualimap[17] para hacer un control de calidad a los alineamientos, que fueron agrupados nuevamente por MultiQC, y también fue necesario Samtools[18] para convertir los archivos de salida .sam de HISAT2 a archivos .bam.



Se empleó un script de R con las librerías pacman[19], DeSeq2[20], biomaRt[21], tidyverse[22], ggplot2[23], pheatmap[24] y ggrepel[25] para el análisis de expresión diferencial.

También se empleó YMLA[26] y YeastEnrichr[27], [28] en sus versiones web para enriquecer los genes resultantes.

Metodología

El primer paso consistió en hacer un control de calidad de las muestras, compuestas por 3 réplicas de *Saccharomyces cerevisiae* en estado batch y en quimiostato, dos archivos por cada muestra, conteniendo la secuencia en la hebra plus y minus, ya que se trata de lecturas pair-end.

Una vez comprobada se continuó con el alineamiento, el cual consistió en una primera etapa de creación de índice, con los parámetros overhang de 100, genomeSAindexNbases en 7 y el resto por defecto, y luego el alineamiento con el índice y la anotación con parámetros por defecto. Para validar el proceso se empleó un segundo alineador, con un proceso similar y parámetros por defecto, se compararon los resultados y luego se continuó con los resultados de STAR.

Posteriormente, se realizó un análisis de expresión diferenciada con los archivos de salida del paso anterior, consistió en el preprocesamiento, anotación con la base de datos de Ensembl, ajuste del modelo estadístico “quimiostato vs batch” y visualización de los resultados.

Finalmente, se ejecutó un análisis de enriquecimiento con YeasEnrichr con los genes que se encontraron expresados diferencialmente del script anterior, con un criterio de considerar como regulados de manera positiva, “upregulated”, a genes que presenten un log2FoldChange mayor a 0,5 y tengan un p-value menor a 0,005 y como genes regulados de manera negativa, “downregulated”, a genes que presenten un log2FoldChange menor a -0,5 y tengan un p-value menor a 0,005. A su vez, se efectuó otro análisis de enriquecimiento con un set extendido de genes expresados de diferencialmente de todos los cromosomas.

III. RESULTADOS

Control de calidad

Se puede observar (fig1) que el número de reads varía entre las réplicas entre unos 35.000 y 60.000 reads (A), también se observa que las lecturas son paired end. El nivel de calidad de la secuencia en las distintas réplicas es muy bueno basado en sus valores Phred(B, C) y la baja cantidad de bases N (E). Se puede apreciar una diferencia en el porcentaje de GC entre las muestras batch y quimiostato (D). Los largos de las reads estan alrededor de las 100 bases (F), con algunos menores a tal cantidad que se debe a que fueron sometidas a un proceso de trimming. En cuanto a los niveles de duplicación (G), es esperable encontrar un pico de secuencias duplicadas más de 10 veces por la naturaleza del estudio. Finalmente, se observa un enriquecimiento de algunos nucleótidos en las primeras bases secuenciadas por un artefacto inherente a la técnica utilizada en la secuenciación (H).

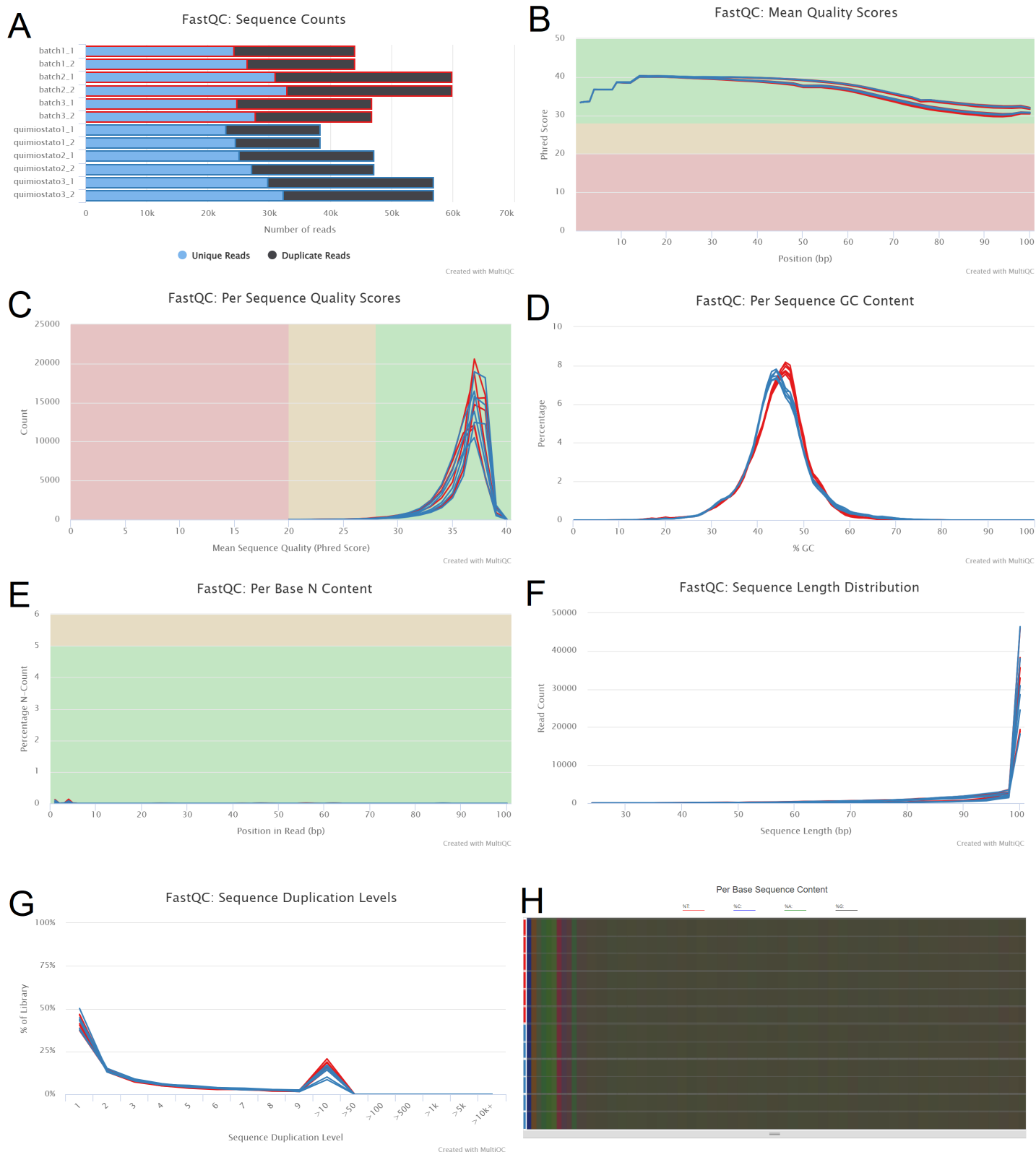
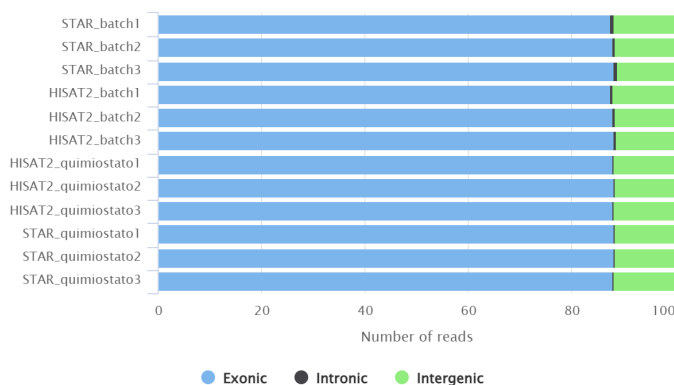


Figura 1. Resultados de FastQC de los datos crudos, visualizados en MultiQC. En rojo las muestras pertenecientes al estado batch y en azul las de quimiostato.

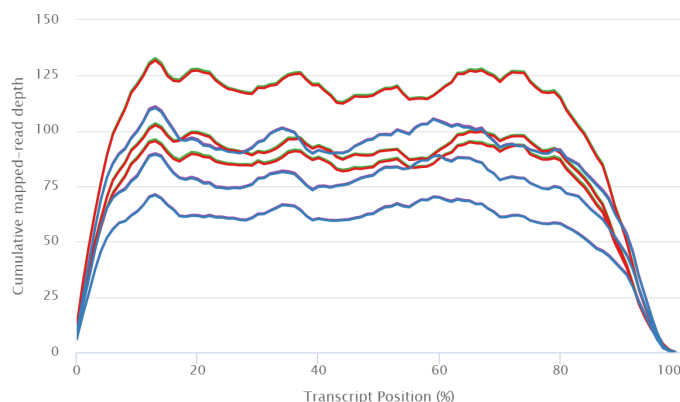


Qualimap RNAseq: Genomic Origin



Created with MultiQC

Qualimap RNAseq: Coverage Profile Along Genes (total)



Created with MultiQC

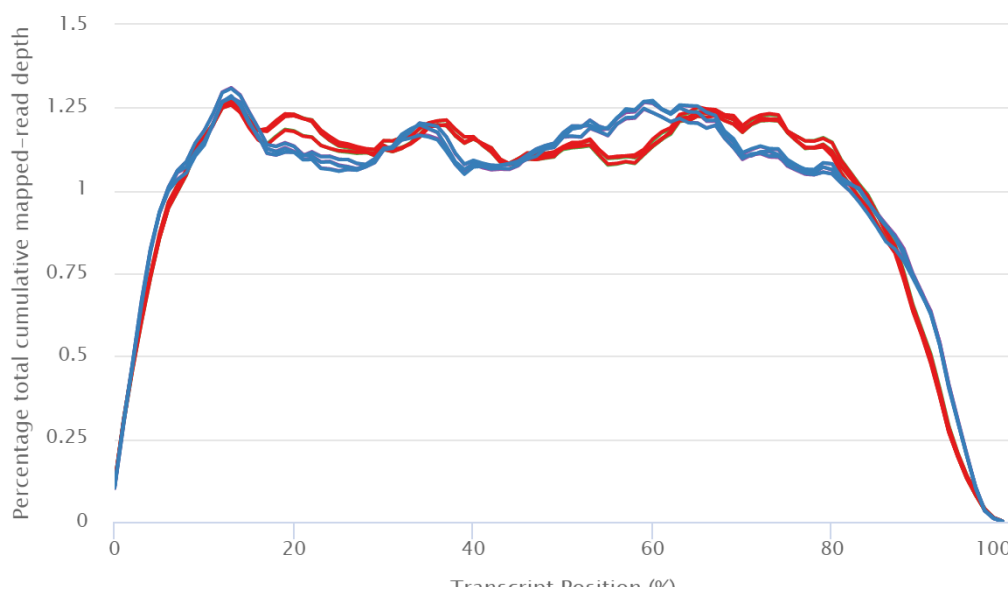


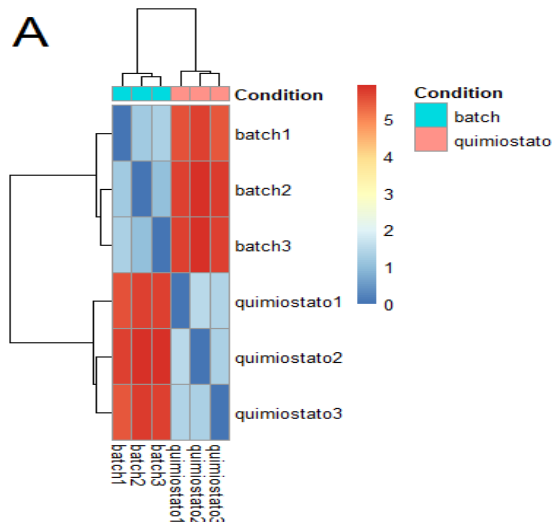
Figura 2. Resultados de Qualimap, visualizados en MultiQC. En rojo las muestras pertenecientes a batch analizadas con STAR, en azul las de quimiostato analizadas con STAR, en verde las batch analizadas con HISAT2 y en violeta las quimiostato analizadas con HISAT2.

Alineamiento

De las salidas de ambos alineadores (fig2) se observa cierta similitud, tanto en el origen genómico (A), como en la cobertura (B, C).

Análisis de expresión diferencial

En el análisis de expresión diferencial (fig3) se observó gran correlación entre las réplicas, diferenciadas por su condición (A), algo que también queda claro en el análisis de componentes (B), donde la condición explica el 93 % de la varianza. Finalmente, se observaron algunos genes regulados de manera positiva, como ACS1, GDH3, YAT1, y de manera negativa, PMT2, ADE1, CYS3, (C) con respecto al modelo considerado quimiostato vs batch.



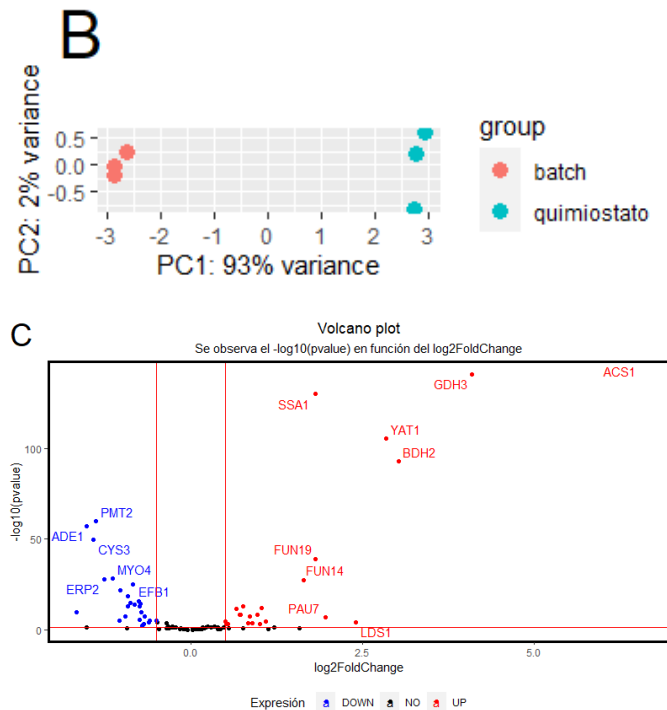


Figura 3. Gráficos generados en R para el análisis diferencial

Análisis de enriquecimiento

De los resultados devueltos por YeastEnrichR (fig4) se puede ver que en el caso de los genes que estaban regulados de manera negativa en quimiostato, principalmente, las funciones biológicas tienen que ver con transporte de proteínas (A), los componentes celular asociados a Golgi y vesículas (C), su función molecular a unión de GTP (E) y las rutas a factores de transcripción y modificación de proteínas (G).

En el caso de los genes regulados de manera positiva en quimiostato, principalmente, las funciones biológicas están relacionadas con organización mitocondrial, homeostasis de lípidos y reciclado de receptores (B), los componentes celulares en la membrana mitocondrial externa y el complejo ERMES (D), las funciones moleculares a uniones a estructuras flap de ADN, calcio y purinas (F) y las rutas a síntesis de glutamato (H).

Análisis de enriquecimiento extendido

En los resultados del análisis extendido (MS1) se encontraron resultados similares a los encontrados anteriormente analizando los genes por separado, gran parte involucrados en procesos de traducción y generación de ribosomas (A), en componentes citoplasmáticos y citosólicos (B), con actividad oxidoreductasa y de unión (C) y que afectan a rutas de respiración y síntesis de nucleótidos y lípidos(D).

IV. DISCUSIÓN

En este trabajo, se encontraron y analizaron las diferencias en los niveles de expresión de genes del cromosoma 1 de levadura en los estados de quimiostato y batch.

Quimiostato

En quimiostato, las rutas reguladas de manera positiva fueron la de síntesis de glutamato, debido al gen GDH3, el cual está relacionado directamente con la síntesis de glutamato a partir de amonio y 2-oxoglutarato, un producto del ciclo de Krebs[29], y la ruta de la glucólisis, particularmente el gen ACS1, el cual regula el paso de acetato a acetyl co-a dentro de la mitocondria para dar inicio al ciclo de Krebs[30].

También se observa mayor transcripción de genes con función biológica de homeostasis de lípidos, y organización mitocondrial, donde también se observó que uno de los componentes celulares principalmente afectados es la membrana mitocondrial y el complejo ERMES, lo que indica una mayor activación de mecanismos de interacciones entre mitocondrias y el retículo endoplasmático, además está relacionado con la mayor expresión de genes de unión de calcio[31]. La reorganización no solo ocurre en el exterior de la célula, sino también en el interior, ya que muchos de los genes están relacionados con una mayor unión a estructuras flap de ADN.

Batch

En el estado batch, las rutas mayoritariamente afectas de manera positiva son la de síntesis de cisteína, por el gen CYS3[32], síntesis de novo de nucleótidos de purina, ADE1[33], y modificaciones de proteína, en particular, glicosilación con manosa, mediada por el gen PMT2[34].

Hay una mayor transcripción de genes relacionados con el transporte y modificación de proteínas, lo que indicaría un gran reordenamiento del proteoma en levadura durante procesos de escasez de recursos energéticos, acompañado por una mayor producción de amino ácidos.

Conclusión

Mediante este análisis se observaron los genes del cromosoma 1 que se expresan de manera diferencial en los estados de quimiostato y batch. Muchos de estos genes, lógicamente, involucrados en procesos de respiración oxidativa y fermentación, pero muchos otros involucrados en rutas de síntesis de aminoácidos, modificaciones de proteínas y homeostasis de lípidos[10], [35].

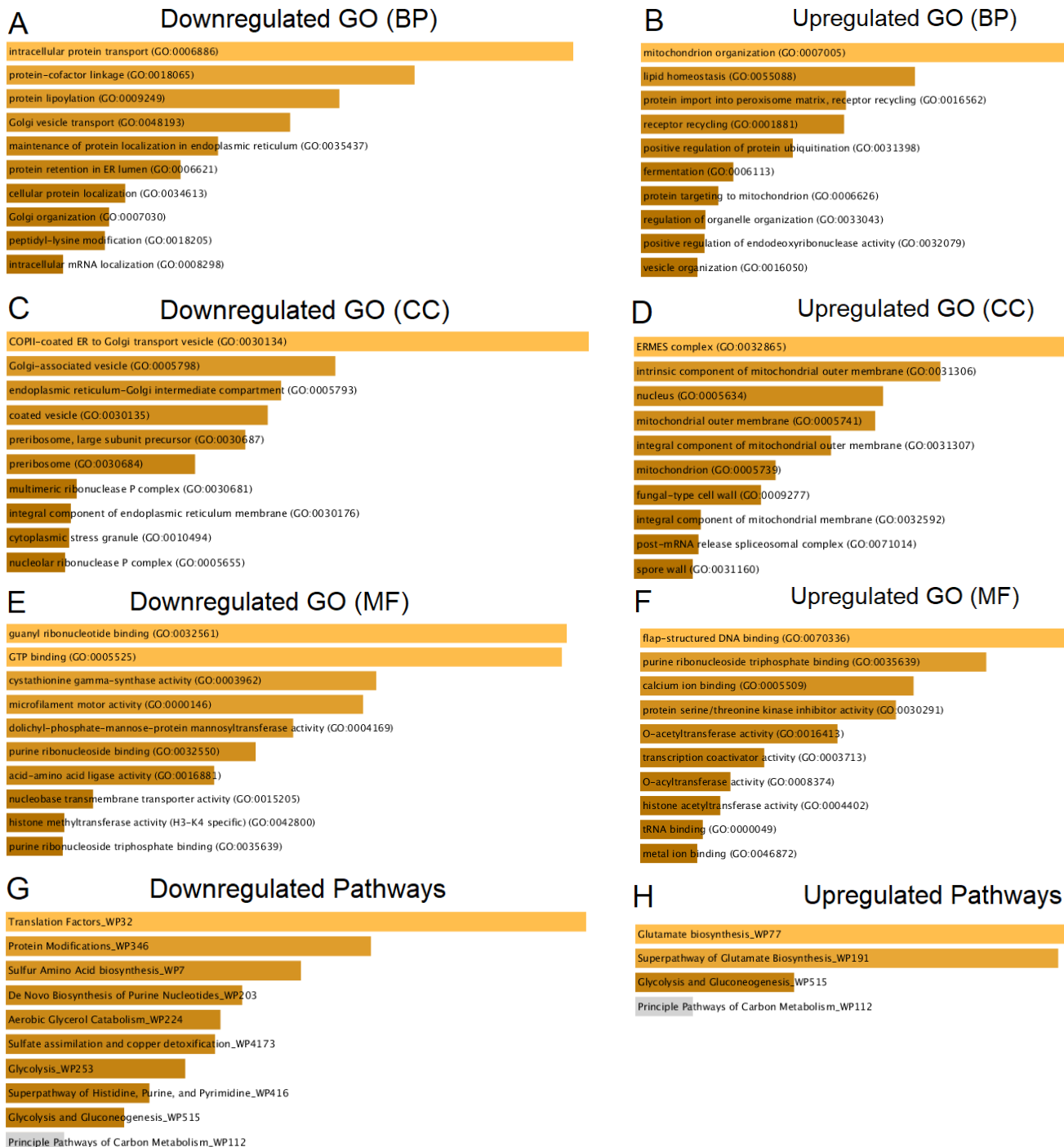


Figura 4. Resultados de YeastEnrichr: para proceso biológico(BP) en la lista de genes downregulados (A) y upregulados (B), para componente celular(CC) en la lista de genes downregulados (C) y upregulados (D), para función molecular(MF) en la lista de genes downregulados (E) y upregulados (F) y para los pathways afectados en la lista de genes downregulados (G) y upregulados (H)

DISPONIBILIDAD DE LOS DATOS

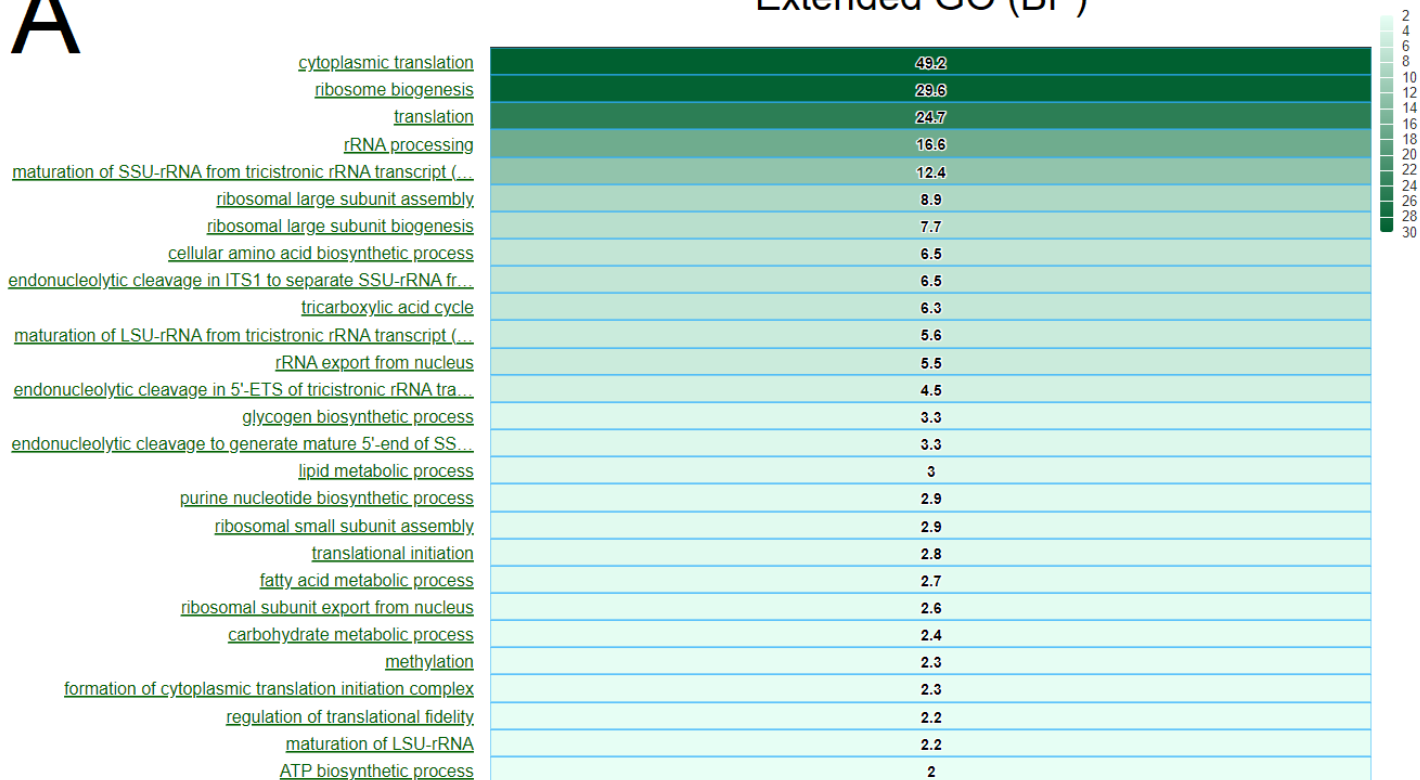
La carpeta con los datos y algoritmos utilizados en el análisis se encuentran en el siguiente repositorio de Github: <https://github.com/STNasim/University>. En el directorio /Nuevas_tecnologias_secuenciacion/TP_final.



MATERIAL SUPLEMENTARIO

A

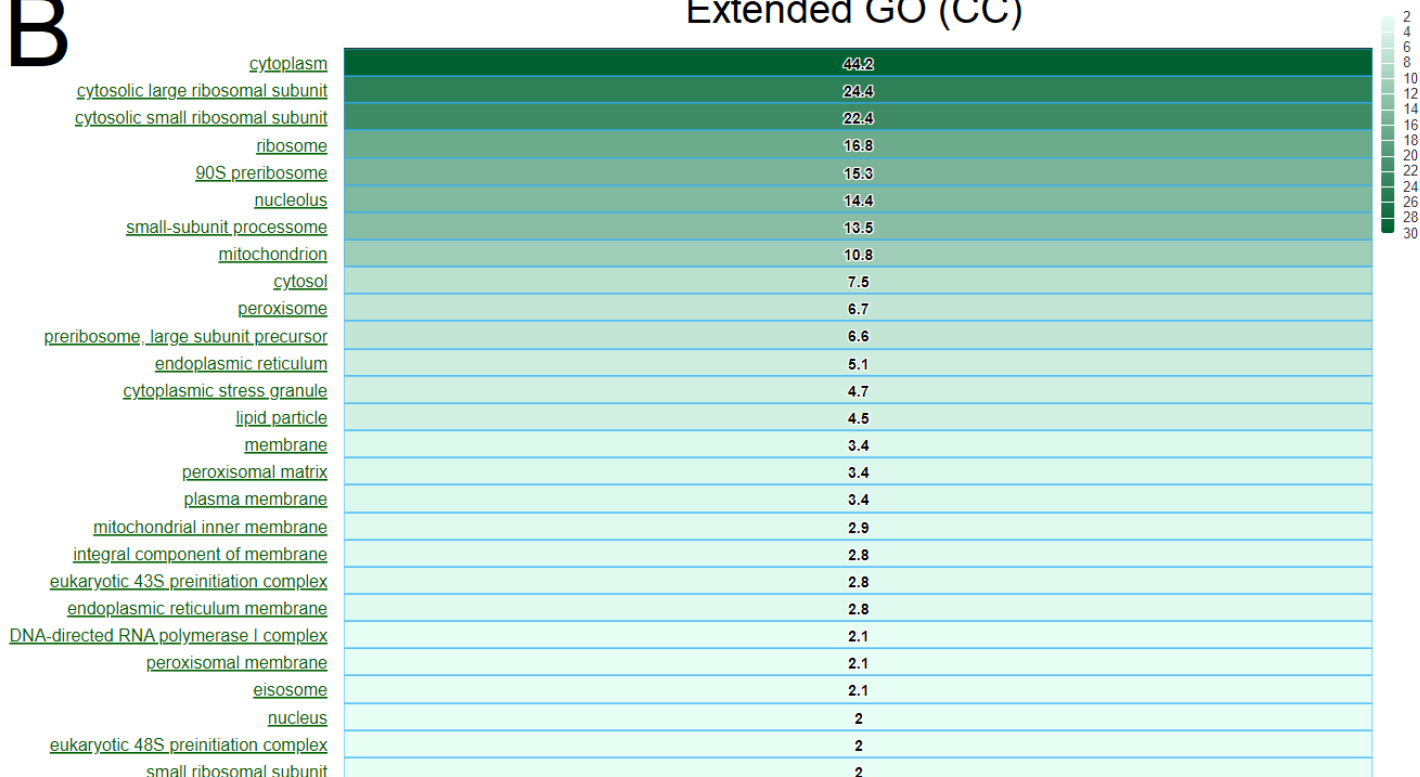
Extended GO (BP)



Highcharts.com

B

Extended GO (CC)

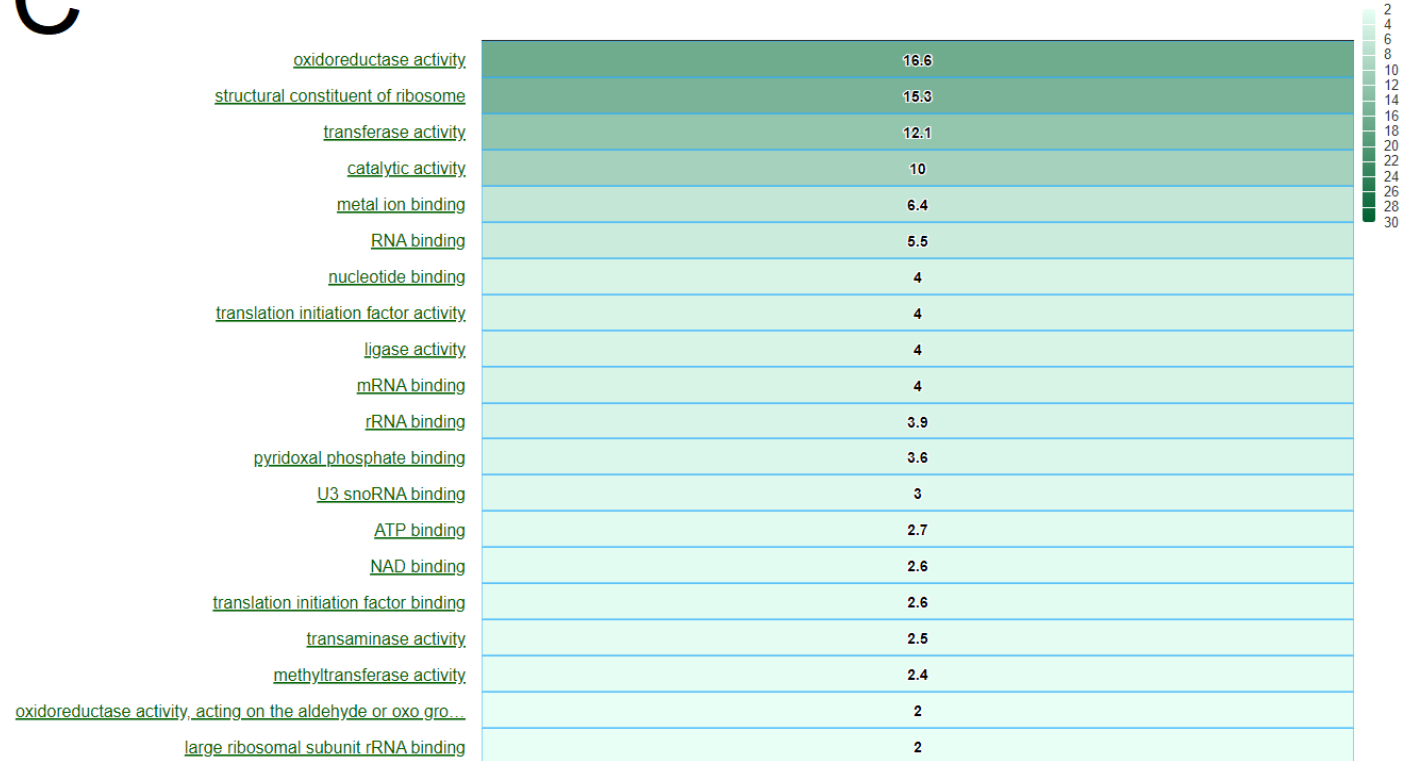


Highcharts.com



C

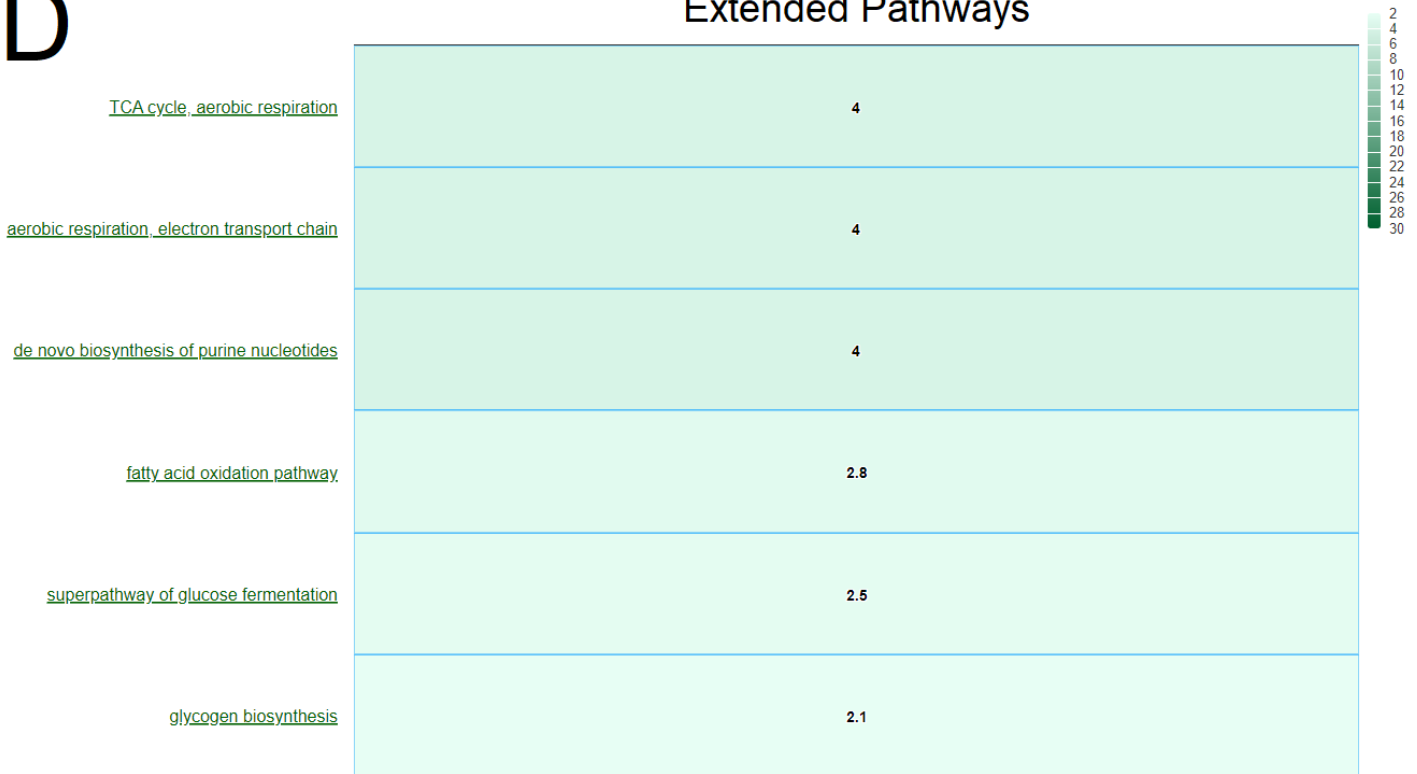
Extended GO (MF)



Highcharts.com

D

Extended Pathways



MSI. Resultados de YMLA con set de genes extendido.



REFERENCIAS

- [1] S. Maicas, «The Role of Yeasts in Fermentation Processes», *Microorganisms*, vol. 8, n.º 8, p. 1142, jul. 2020, doi: 10.3390/microorganisms8081142.
- [2] H. Karathia, E. Vilaprinyo, A. Sorribas, y R. Alves, «Saccharomyces cerevisiae as a model organism: a comparative study», *PloS One*, vol. 6, n.º 2, p. e16015, feb. 2011, doi: 10.1371/journal.pone.0016015.
- [3] C. Boone, «Yeast Systems Biology: Our Best Shot at Modeling a Cell», *Genetics*, vol. 198, n.º 2, pp. 435-437, oct. 2014, doi: 10.1534/genetics.114.169128.
- [4] N. Gurdo, «Mejoramiento de la levadura Saccharomyces cerevisiae Y138 por ingeniería evolutiva para la producción de bioetanol de segunda generación», 2016. doi: 10.13140/RG.2.1.4246.3602.
- [5] M. Parapouli, A. Vasileiadis, A.-S. Afendra, y E. Hatziloukas, «Saccharomyces cerevisiae and its industrial applications», *AIMS Microbiol.*, vol. 6, n.º 1, pp. 1-31, feb. 2020, doi: 10.3934/microbiol.2020001.
- [6] H. Taymaz-Nikerel, A. Cankorur-Cetinkaya, y B. Kirdar, «Genome-Wide Transcriptional Response of Saccharomyces cerevisiae to Stress-Induced Perturbations», *Front. Bioeng. Biotechnol.*, vol. 4, 2016, Accedido: 31 de mayo de 2023. [En línea]. Disponible en: <https://www.frontiersin.org/articles/10.3389/fbioe.2016.00017>
- [7] M. C. Teixeira, N. P. Mira, y I. Sá-Correia, «A genome-wide perspective on the response and tolerance to food-relevant stresses in Saccharomyces cerevisiae», *Curr. Opin. Biotechnol.*, vol. 22, n.º 2, pp. 150-156, abr. 2011, doi: 10.1016/j.copbio.2010.10.011.
- [8] S. Zaman, S. I. Lippman, L. Schneper, N. Slonim, y J. R. Broach, «Glucose regulates transcription in yeast through a network of signaling pathways», *Mol. Syst. Biol.*, vol. 5, n.º 1, p. 257, ene. 2009, doi: 10.1038/msb.2009.20.
- [9] M. Conrad, J. Schothorst, H. N. Kankipati, G. Van Zeebroeck, M. Rubio-Teixeira, y J. M. Thevelein, «Nutrient sensing and signaling in the yeast Saccharomyces cerevisiae», *FEMS Microbiol. Rev.*, vol. 38, n.º 2, pp. 254-299, mar. 2014, doi: 10.1111/1574-6976.12065.
- [10] V. M. Boer, J. H. de Winde, J. T. Pronk, y M. D. W. Piper, «The Genome-wide Transcriptional Responses of Saccharomyces cerevisiae Grown on Glucose in Aerobic Chemostat Cultures Limited for Carbon, Nitrogen, Phosphorus, or Sulfur *», *J. Biol. Chem.*, vol. 278, n.º 5, pp. 3265-3274, ene. 2003, doi: 10.1074/jbc.M209759200.
- [11] J. F. Nijkamp *et al.*, «De novo sequencing, assembly and analysis of the genome of the laboratory strain Saccharomyces cerevisiae CEN.PK113-7D, a model for modern industrial biotechnology», *Microb. Cell Factories*, vol. 11, n.º 1, p. 36, mar. 2012, doi: 10.1186/1475-2859-11-36.
- [12] I. Nookaew *et al.*, «A comprehensive comparison of RNA-Seq-based transcriptome analysis from reads to differential gene expression and cross-comparison with microarrays: a case study in Saccharomyces cerevisiae», *Nucleic Acids Res.*, vol. 40, n.º 20, pp. 10084-10097, nov. 2012, doi: 10.1093/nar/gks804.
- [13] S. Andrews, «FastQC A Quality Control tool for High Throughput Sequence Data», 2010. <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
- [14] P. Ewels, M. Magnusson, S. Lundin, y M. Käller, «MultiQC: summarize analysis results for multiple tools and samples in a single report», *Bioinformatics*, vol. 32, n.º 19, pp. 3047-3048, oct. 2016, doi: 10.1093/bioinformatics/btw354.
- [15] A. Dobin *et al.*, «STAR: ultrafast universal RNA-seq aligner», *Bioinformatics*, vol. 29, n.º 1, pp. 15-21, ene. 2013, doi: 10.1093/bioinformatics/bts635.
- [16] D. Kim, J. M. Paggi, C. Park, C. Bennett, y S. L. Salzberg, «Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype», *Nat. Biotechnol.*, vol. 37, n.º 8, Art. n.º 8, ago. 2019, doi: 10.1038/s41587-019-0201-4.
- [17] F. García-Alcalde *et al.*, «Qualimap: evaluating next-generation sequencing alignment data», *Bioinformatics*, vol. 28, n.º 20, pp. 2678-2679, oct. 2012, doi: 10.1093/bioinformatics/bts503.
- [18] P. Danecek *et al.*, «Twelve years of SAMtools and BCFtools», *GigaScience*, vol. 10, n.º 2, feb. 2021, doi: 10.1093/gigascience/giab008.
- [19] T. W. Rinker y D. Kurkiewicz, *pacman: Package Management for R*. Buffalo, New York, 2018. [En línea]. Disponible en: <http://github.com/trinker/pacman>
- [20] M. I. Love, W. Huber, y S. Anders, «Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2», *Genome Biol.*, vol. 15, n.º 12, p. 550, 2014, doi: 10.1186/s13059-014-0550-8.
- [21] D. Smedley *et al.*, «BioMart--biological queries made easy», *BMC Genomics*, vol. 10, p. 22, ene. 2009, doi: 10.1186/1471-2164-10-22.
- [22] H. Wickham *et al.*, «Welcome to the Tidyverse», *J. Open Source Softw.*, vol. 4, p. 1686, nov. 2019, doi: 10.21105/joss.01686.
- [23] H. Wickham, *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2016. [En línea]. Disponible en: <https://ggplot2.tidyverse.org>
- [24] R. Kolde, «pheatmap: Pretty Heatmaps». 4 de enero de 2019. [En línea]. Disponible en: <https://cran.r-project.org/web/packages/pheatmap/index.html>
- [25] K. Slowikowski, *ggrepel: Automatically Position Non-Overlapping Text Labels with «ggplot2»*. 2023. [En línea]. Disponible en: <https://github.com/slowkow/ggrepel>
- [26] T.-H. Yang *et al.*, «YMLA: A comparative platform to carry out functional enrichment analysis for multiple gene lists in yeast», *Comput. Biol. Med.*, vol. 151, p. 106314, dic. 2022, doi: 10.1016/j.compbio.2022.106314.



- [27] E. Y. Chen *et al.*, «Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool», *BMC Bioinformatics*, vol. 14, p. 128, abr. 2013, doi: 10.1186/1471-2105-14-128.
- [28] M. V. Kuleshov *et al.*, «Enrichr: a comprehensive gene set enrichment analysis web server 2016 update», *Nucleic Acids Res.*, vol. 44, n.º W1, pp. W90-97, jul. 2016, doi: 10.1093/nar/gkw377.
- [29] M. Braymer, W. M. Bot, D. Digles, E. Willighagen, M. Summer-Kutmon, y E. Weitz, «Glutamate biosynthesis superpathway», may 2021, Accedido: 30 de mayo de 2023. [En línea]. Disponible en: <https://www.wikipathways.org/instance/WP191>
- [30] Kdahlquist *et al.*, «Glycolysis and gluconeogenesis», may 2021, Accedido: 30 de mayo de 2023. [En línea]. Disponible en: <https://www.wikipathways.org/instance/WP515>
- [31] G. Hajnóczky *et al.*, «Reliance of ER-mitochondrial calcium signaling on mitochondrial EF-hand Ca²⁺ binding proteins: Miros, MICUs, LETM1 and solute carriers», *Curr. Opin. Cell Biol.*, vol. 29, pp. 133-141, ago. 2014, doi: 10.1016/j.ceb.2014.06.002.
- [32] J. Heckman *et al.*, «Sulfur amino acid biosynthesis», ene. 2023, Accedido: 30 de mayo de 2023. [En línea]. Disponible en: <https://www.wikipathways.org/instance/WP7>
- [33] M. Braymer, W. M. Bot, C. Chichester, E. Willighagen, M. Summer-Kutmon, y E. Weitz, «De novo biosynthesis of purine nucleotides», may 2021, Accedido: 30 de mayo de 2023. [En línea]. Disponible en: <https://www.wikipathways.org/instance/WP203>
- [34] M. Braymer, W. M. Bot, D. Digles, E. Willighagen, M. Summer-Kutmon, y E. Weitz, «Protein modifications», may 2021, Accedido: 30 de mayo de 2023. [En línea]. Disponible en: <https://www.wikipathways.org/instance/WP346>
- [35] X. Fu, P. Li, L. Zhang, y S. Li, «RNA-Seq-based transcriptomic analysis of *Saccharomyces cerevisiae* during solid-state fermentation of crushed sweet sorghum stalks», *Process Biochem.*, vol. 68, pp. 53-63, may 2018, doi: 10.1016/j.procbio.2018.02.024.