

Sección: ¿Cómo funciona?

Explorando datos abiertos en biomedicina: Una introducción práctica al uso de GEO y GEO2R

Exploring open data in biomedicine: A practical introduction to using GEO and GEO2R

Camila Fernanda Cunha-Brandao^{1,2}

Yan Leo Melo-Vieira^{1,2}

José de Jesús Martínez-González^{3*}

¹Internal Medicine Department, Ribeirão Preto Medical School, São Paulo State University, Ribeirão Preto, Brasil.

²Research Laboratory on Metabolism, Physiology and Physical Exercise (LAPEMFE), Minas Gerais State University, Divinópolis, Brasil.

³Universidad Autónoma del Estado de Morelos, CIICAp, Av. Universidad 1001, Col. Chamilpa 62209, Cuernavaca, Morelos, México.

*Autor para la correspondencia: jose_mtz09@hotmail.es

RESUMEN

El Gene Expression Omnibus (GEO) del NCBI es un repositorio público con miles de estudios de expresión génica y otros perfiles moleculares. Reutilizar estos datos abiertos permite generar hipótesis, validar hallazgos y realizar análisis exploratorios *in silico* sin costo de laboratorio. En este artículo explicamos, con lenguaje sencillo, qué es GEO, qué información contiene y cómo usar herramientas accesibles como GEO2R para análisis básicos reproducibles, incluso en instituciones con recursos limitados. Además, presentamos un ejemplo ilustrativo del flujo de trabajo y posibles resultados, sin establecer conclusiones biológicas o clínicas definitivas.

Palabras clave: GEO, bioinformática accesible, datos biomédicos abiertos.

SUMMARY

The NCBI Gene Expression Omnibus (GEO) is a public repository containing thousands of gene expression studies and other molecular profiles. Reusing these open data makes it possible to generate hypotheses, validate findings, and perform exploratory *in silico* analyses without laboratory costs. In this article, we explain in simple language what GEO is, what information it contains, and how to use accessible tools such as GEO2R to perform basic reproducible analyses, even in institutions with limited resources. We also present an illustrative example of the workflow and potential results, without establishing definitive biological or clinical conclusions.

Keywords: GEO, accessible bioinformatics, open biomedical data.

Introducción

En los últimos años, la disponibilidad de datos biomédicos abiertos ha crecido de forma sostenida. Entre los repositorios más usados se encuentra el Gene Expression Omnibus (GEO), administrado por el NCBI, que almacena estudios de expresión génica y otros perfiles moleculares provenientes de múltiples especies, tejidos y condiciones. Aparte de almacenar datos, no todos saben que los datos públicos pueden ser reanalizados y que los nuevos análisis in silico pueden publicarse; el reanálisis de resultados ya publicados es una práctica consistente con la ciencia abierta y la reproducibilidad [1].

En este contexto, el objetivo de este artículo es presentar, de manera accesible, el potencial de GEO como herramienta para el aprendizaje y la exploración inicial de datos biomédicos abiertos, así como ilustrar el uso básico de GEO2R mediante un ejemplo práctico. El caso mostrado tiene una finalidad exclusivamente didáctica y no pretende sustituir análisis bioinformáticos más rigurosos ni sustentar conclusiones biológicas o clínicas definitivas.

¿Qué es GEO y por qué importa?

GEO es, en esencia, una gran biblioteca digital de datos de expresión génica (actividad de los genes). Creada en el año 2000 por el Centro Nacional de Información Biotecnológica (NCBI) de Estados Unidos, fue el primer repositorio público dedicado a recopilar y compartir datos de este tipo. Surgió en paralelo al Proyecto Genoma Humano, cuando la comunidad científica necesitaba un lugar común para almacenar y consultar grandes volúmenes de información omica [2].

El análisis diferencial consiste, de forma sencilla, en comparar dos o más grupos de muestras para identificar cambios moleculares

asociados con una condición biológica o clínica. En estudios de expresión génica, este análisis permite reconocer qué genes están más o menos activos entre grupos; por ejemplo, puede utilizarse para comparar tejido hepático sano y tejido hepático enfermo, e identificar qué genes aumentan o disminuyen su actividad en asociación con la enfermedad. El análisis de expresión génica diferencial consiste, de forma sencilla, en medir qué genes están más encendidos o apagados en diferentes grupos de muestras [1,3]. Por ejemplo, este tipo de análisis puede emplearse para comparar tejido hepático sano y tejido hepático enfermo, con el fin de identificar qué genes aumentan o disminuyen su actividad en asociación con una enfermedad. La expresión génica puede entenderse como una fotografía de los procesos celulares que ocurren en el momento en que se toma la muestra. GEO nació para almacenar esas “fotografías” en forma de tablas y ponerlas al alcance de cualquier persona, en cualquier país, sin costo.

GEO no almacena únicamente estudios de expresión génica, sino también otros tipos de datos omicos, como metilación del ADN, microARN y otros perfiles moleculares. En estudios de metilación, como el presentado en este trabajo, el análisis se orienta a identificar sitios CpG con diferencias en su nivel de metilación entre grupos, en lugar de cambios en la actividad génica. Por ello, los resultados obtenidos con GEO2R dependen del tipo de estudio analizado y pueden consistir en genes diferencialmente expresados o en sondas y sitios CpG diferencialmente metilados [3].

Cómo empezar con GEO sin programar

En la ciencia tradicional, los datos solían quedarse en el laboratorio donde se generaban, a veces “olvidados” tras la publicación del artículo. GEO cambió esa dinámica: una vez que el estudio se publica, los datos quedan

Figura 1. Página principal de Gene Expression Omnibus (GEO), repositorio público del NCBI para consulta de estudios y muestras de datos ómicos.. Disponible en: <https://www.ncbi.nlm.nih.gov/geo>

disponibles para toda la comunidad (Figura 1).

Esta posibilidad de reanalizar datos es especialmente valiosa para grupos que inician su trayectoria, que trabajan con presupuestos ajustados y que no tienen experiencia en análisis en laboratorio de biología molecular, pues solo con una computadora y conexión a internet es factible realizar experimentos *in silico* a gran escala sin incurrir en costos de reactivos, infraestructura o animalarios. Además, el uso de métodos estadísticos y plataformas bioinformáticas amigables permite contrastar hipótesis de forma rápida, comparar estudios y aumentar la solidez de los hallazgos mediante validaciones cruzadas.

Advertencia metodológica. El análisis presentado es únicamente ilustrativo y no debe considerarse evidencia concluyente. En estudios ómicos, en particular de metilación del ADN, la interpretación requiere un diseño estadístico y experimental riguroso, incluyendo control de calidad, normalización, efectos

de lote, covariables y validación independiente.

Los datos públicos de GEO también pueden promover colaboraciones internacionales, por ejemplo, para validar un resultado encontrado en determinada población con otras poblaciones.

Caso ilustrativo: búsqueda y reanálisis exploratorio en obesidad

La obesidad, campo de estudio de nuestro grupo de investigación, es una enfermedad crónica de creciente prevalencia a nivel mundial; su origen es multifactorial [4] y el uso de herramientas ómicas puede aportar más información sobre la condición [5]. La Figura 2 muestra la cantidad de datos disponibles al buscar el término *obesity* en GEO; se recuperaron 23,813 conjuntos de datos (consulta realizada el 15 de enero de 2026).

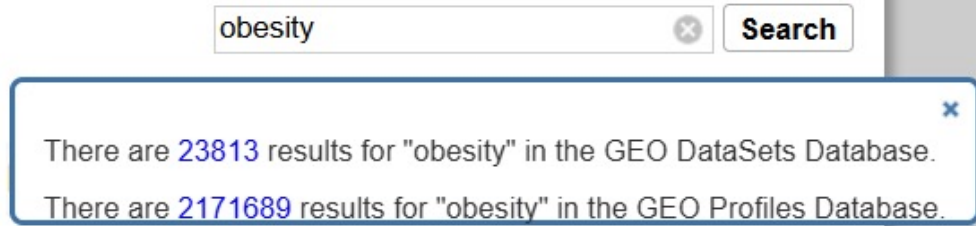
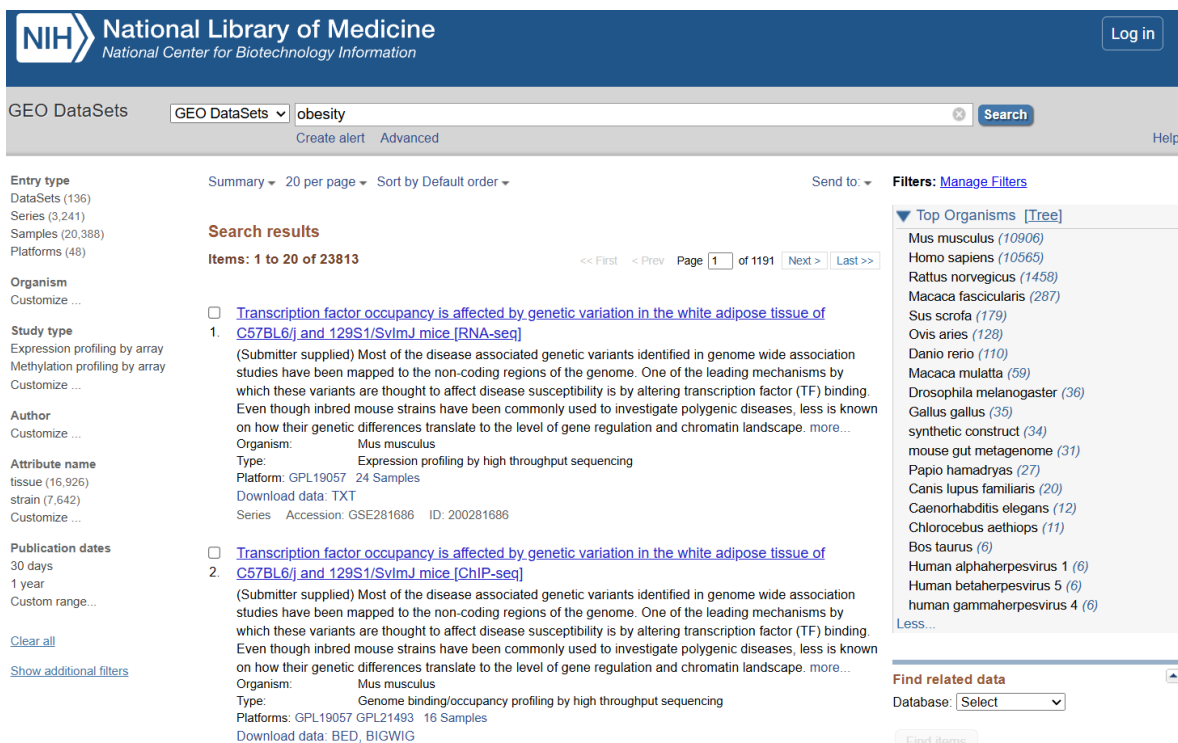


Figura 2. Resultados de búsqueda en GEO para el término “obesity”, que muestran el número de conjuntos de datos recuperados al momento de la consulta (15 de enero de 2026).



Search results
Items: 1 to 20 of 23813

1. [Transcription factor occupancy is affected by genetic variation in the white adipose tissue of C57BL/6j and 129S1/SvImJ mice \[RNA-seq\]](#)
(Submitter supplied) Most of the disease associated genetic variants identified in genome wide association studies have been mapped to the non-coding regions of the genome. One of the leading mechanisms by which these variants are thought to affect disease susceptibility is by altering transcription factor (TF) binding. Even though inbred mouse strains have been commonly used to investigate polygenic diseases, less is known on how their genetic differences translate to the level of gene regulation and chromatin landscape. more...
Organism: Mus musculus
Type: Expression profiling by high throughput sequencing
Platform: GPL19057 24 Samples
Download data: TXT
Series Accession: GSE281686 ID: 200281686

2. [Transcription factor occupancy is affected by genetic variation in the white adipose tissue of C57BL/6j and 129S1/SvImJ mice \[ChIP-seq\]](#)
(Submitter supplied) Most of the disease associated genetic variants identified in genome wide association studies have been mapped to the non-coding regions of the genome. One of the leading mechanisms by which these variants are thought to affect disease susceptibility is by altering transcription factor (TF) binding. Even though inbred mouse strains have been commonly used to investigate polygenic diseases, less is known on how their genetic differences translate to the level of gene regulation and chromatin landscape. more...
Organism: Mus musculus
Type: Genome binding/occupancy profiling by high throughput sequencing
Platforms: GPL19057 GPL21493 16 Samples
Download data: BED, BIGWIG

Top Organisms [Tree]

- Mus musculus (10906)
- Homo sapiens (10565)
- Rattus norvegicus (1458)
- Macaca fascicularis (287)
- Sus scrofa (179)
- Ovis aries (128)
- Danio rerio (110)
- Macaca mulatta (59)
- Drosophila melanogaster (36)
- Gallus gallus (35)
- synthetic construct (34)
- mouse gut metagenome (31)
- Papio hamadryas (27)
- Canis lupus familiaris (20)
- Caenorhabditis elegans (12)
- Chlorocebus aethiops (11)
- Bos taurus (6)
- Human alphaherpesvirus 1 (6)
- Human betaherpesvirus 5 (6)
- human gammaherpesvirus 4 (6)
- Less...

Figura 3. Resultados de GEO DataSets para el término obesity, mostrando los organismos más frecuentes asociados a los conjuntos de datos recuperados.

Salvo indicación en contrario, las búsquedas y capturas de pantalla incluidas en este trabajo corresponden a consultas realizadas en GEO en enero de 2026. En la búsqueda del término obesity, GEO recuperó 23,813 conjuntos de datos (15 de enero de 2026), como se muestra en la Figura 2.

De estos conjuntos de datos, encontramos para el término “obesity” estudios correspondientes a veinte grupos de organismos diferentes, como se muestra en la Figura 3.

La Figura 4 muestra qué tipos de experimentos se almacenan en la base de datos; en el menú de la izquierda, el usuario puede

Figura 4. Filtro de búsqueda para estudios con seres humanos y selección del tipo de análisis de expresión génica.

Figura 5. Un estudio sobre las firmas epigenéticas moduladas por la intervención del entrenamiento físico en mujeres obesas, desarrollado por el grupo.

seleccionar el tipo de experimento deseado eligiendo las técnicas en study design para algunos estudios.

En las siguientes imágenes, hemos se-

leccionado un estudio para ejemplificar las funcionalidades de la plataforma. La Figura 5 muestra el estudio seleccionado para los pasos siguientes. Para este ejemplo, seleccionamos un estudio de nuestro grupo donde evaluamos

NCBI > GEO > **Accession Display** [?](#) Not logged in | [Login](#) [?](#)

Scope: Format: Amount: GEO accession:

Series GSE296180 [Query DataSets for GSE296180](#)

| | |
|------------------|--|
| Status | Public on May 06, 2025 |
| Title | COMBINED EXERCISE AND ITS EPIGENETIC SIGNATURES: 8-WEEKS PROTOCOL MODULATED THERMOGENESIS, CARDIOVASCULAR AND LIPID BIOMARKERS RELATED PATHWAYS |
| Organism | Homo sapiens |
| Experiment type | Methylation profiling by genome tiling array |
| Summary | Background: Obesity, a health concern shaped by genetic, epigenetic, and environmental factors, has emerged as a global epidemic, particularly with a 60% increment in Brazil from 2006 to 2016. This study explored the effects of an 8-week integrated strength and aerobic training protocol on the DNA methylation patterns in 11 sedentary women with obesity |
| Overall design | MethylationBrazil Assay |
| Contributor(s) | Rodrigues G |
| Citation missing | <i>Has this study been published? Please login to update or notify GEO.</i> |

Analyze with GEO2R

| | |
|-------------------|--|
| Submission date | May 02, 2025 |
| Last update date | May 07, 2025 |
| Contact name | Guilherme Da Silva Rodrigues |
| E-mail(s) | guirodrigues@usp.br |
| Phone | +5516981757737 |
| Organization name | Universidade de São Paulo |
| Street address | Avenida Bandeirantes, 3900 |
| City | Ribeirão Preto |
| ZIP/Postal code | 14040907 |
| Country | Brazil |

Figura 6. Resumen del estudio seleccionado.

la metilación de ADN antes y después de la intervención de ejercicio físico durante ocho semanas en mujeres con y sin obesidad. Para utilizar las herramientas bioinformáticas incorporadas al sistema sin necesidad de programación, el usuario debe consultar los estudios que incluyen la opción “Analyze with GEO2R”.

La Figura 6 muestra la interfaz del estudio localizado. Para analizar utilizando las herramientas básicas, se debe hacer clic en el botón “Analyze with GEO2R”.

En este caso ilustrativo, la pregunta ex-

ploratoria planteada fue si existen sitios CpG con metilación diferencial entre participantes con circunferencia de cintura elevada y participantes con circunferencia de cintura normal dentro del conjunto de datos seleccionado. Esta comparación se utiliza únicamente para mostrar cómo GEO2R permite organizar grupos, ejecutar análisis básicos y visualizar resultados en un estudio de metilación del ADN.

Los datos públicos de GEO pueden reutilizarse con distintos fines, incluyendo la reproducción de análisis originales con fines de aprendizaje o verificación, así como la explo-

Use GEO2R to compare two or more groups of Samples in order to identify genes that are differentially expressed across experimental conditions. Results are presented as a table of genes ordered by significance. [Full instructions](#) [YouTube](#)

GEO accession: COMBINED EXERCISE AND ITS EPIGENETIC SIGNATURES: 8-WEEKS PROTOCOL MODULATED THERMOGENESIS, CARDIOVASCULAR AND LIPID BIOMARKERS RELATED PATHWAYS

| Group | Accession | Source name | Age | Characteristics | Vo2 max |
|-----------|------------|-------------------------|-----|--|---------|
| Higher WC | GSM8985518 | exercise [1_pre] | 40 | bm (kg): 83.5 wc (cm): 92.5 bmi (kg/m ²): 31.8 | 18.0 |
| - | GSM8985519 | POST Exercise [1_post] | 40 | bm (kg): 81.7 wc (cm): 88.0 bmi (kg/m ²): 31.0 | 25.0 |
| Lower WC | GSM8985520 | PRE Exercise [11_pre] | 31 | bm (kg): 74.3 wc (cm): 87.5 bmi (kg/m ²): 31.3 | 21.0 |
| - | GSM8985521 | POST Exercise [11_post] | 31 | bm (kg): 71.2 wc (cm): 86.0 bmi (kg/m ²): 30.0 | 27.4 |
| Higher WC | GSM8985522 | PRE Exercise [13_pre] | 38 | bm (kg): 78.4 wc (cm): 90.5 bmi (kg/m ²): 30.2 | 15.0 |
| - | GSM8985523 | POST Exercise [13_post] | 38 | bm (kg): 79.3 wc (cm): 86.0 bmi (kg/m ²): 30.6 | 20.0 |

Figura 7. Filtrado del estudio seleccionado y separación de las muestras según la circunferencia de cintura para realizar una comparación exploratoria de metilación del ADN entre grupos.

GEO2R Options Profile graph R script

Quick start

- Specify a GEO Series accession and a Platform if prompted.
- Click 'Define groups' and enter names for the groups of Samples you plan to compare, e.g., test and control.
- Assign Samples to each group. Highlight Sample rows then click the group name to assign those Samples to the group. Use the Sample metadata (title, source and characteristics) columns to help determine which Samples belong to which group.
- Click 'Analyze' to perform the calculation with default settings.
- You may change settings in the Options tab.

How to use

Run the differential expression analysis. Distribution plots can be viewed without specifying Sample groups.

Figura 8. Información sobre cómo iniciar el análisis después de la búsqueda, con una guía de "Quick start" con consejos sobre cómo analizar. Referencia.

GEO2R **Options** Profile graph R script

Apply adjustment to the P-values. [More...](#)

Benjamini & Hochberg (False discovery rate)

Benjamini & Yekutieli

Bonferroni

Holm

Apply log transformation to the data. [More...](#)

Auto-detect

Yes

No

Apply limma precision weights (vooma). [More...](#)

Yes

No

Force normalization. [More...](#)

Yes

No

Category of Platform annotation to display on results.

Submitter supplied

NCBI generated

Plot displays. [More...](#)

Significance level cut-off (enter number between 0 and 1)

Log 2 fold change threshold

Volcano and Mean-difference plot contrasts (select up to 5)
0 selected (clear)

Higher WC vs Lower WC

If you edit Options after performing an analysis, click Reanalyze to apply the edits.

Figura 9. Opciones de análisis específicas relacionadas con el objetivo del estudio.

```
# Version info: R 4.2.2, Biobase 2.58.0, GEOquery 2.66.0, limma 3.54.0
#####
# Differential expression analysis with limma
library(GEOquery)
library(limma)
library(umap)

# load series and platform data from GEO
gset <- getGEO("GSE296180", GSEMatrix = TRUE, AnnotGPL = FALSE)
if (length(gset) > 1) idx <- grep("GPL21145", attr(gset, "names")) else idx <- 1
gset <- gset[[idx]]

# make proper column names to match toptable
fvarLabels(gset) <- make.names(fvarLabels(gset))

# group membership for all samples
gsms <- "0X1X0X0X0X0X0X0X1X0X0X"
sml <- strsplit(gsms, split="X")[[1]]

# filter out excluded samples (marked as "X")
sel <- which(sml != "X")
sml <- sml[sel]
gset <- gset[,sel]

# log2 transformation
ex <- exprs(gset)
qx <- as.numeric(quantile(ex, c(0., 0.25, 0.5, 0.75, 0.99, 1.0), na.rm=T))
LogC <- (qx[5] > 100) ||
  (qx[6] - qx[1] > 50 && qx[2] > 0)
if (LogC { ex[which(ex <= 0)] <- NaN
  exprs(gset) <- log2(ex) }
```

Figura 10. Opciones de análisis específicas relacionadas con el objetivo del estudio.

ración de nuevas preguntas de investigación. Dado que, en el contexto de la obesidad, una circunferencia de cintura superior a ochenta y ocho centímetros en mujeres constituye un factor de riesgo [6], en este ejercicio se reorganizaron las participantes del estudio de ejemplo de acuerdo con este criterio para ilustrar la identificación exploratoria de sitios CpG con metilación diferencial (Figura 7).

Consideramos High WC a las personas

Volcano plot
GSE296180: COMBINED EXERCISE AND ITS
EPIGENETIC SIGNATURES: 8-WEEKS...
 Higher WC vs Lower WC, Padj<0.05

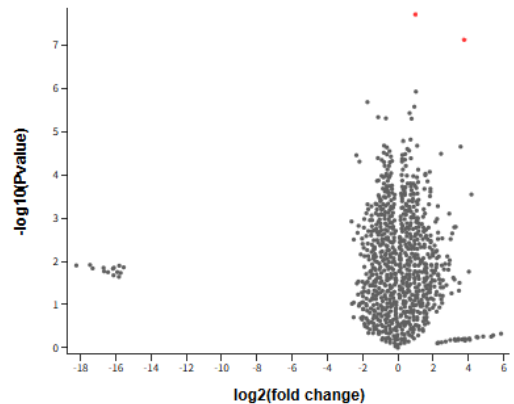


Figura 12. Resultados de la comparación entre los grupos como altos (grupo High WC) y normales (grupo Lower WC), mostrados en un gráfico de volcán generado con GEO, útil para visualizar diferencias estadísticas en estudios epigenéticos.

que, en el momento previo al entrenamiento físico, tenían una circunferencia de cintura mayor de ochenta y ocho centímetros (n = 9) y las comparamos con individuos que tenían una circunferencia de cintura menor de ochenta y ocho centímetros (n = 2).

Acceder a GEO no exige habilidades avanzadas de programación. Varias herramientas web permiten realizar análisis con pasos guiados. GEO2R es la interfaz del pro-

Reanalyze if you changed any options.

Visualization

Top differentially expressed genes

| ID | adj.P.Val | P.Value | t | B | logFC | SPOT_ID |
|------------|-----------|----------|-------|--------|--------|------------|
| cg08977382 | 0.0145 | 1.98e-08 | 15.5 | 0.965 | 0.998 | cg08977382 |
| cg13920898 | 0.0278 | 7.58e-08 | 13.51 | 0.824 | 3.765 | cg13920898 |
| cg01015507 | 0.2021 | 1.20e-06 | 10.12 | 0.411 | 1.021 | cg01015507 |
| cg23817779 | 0.3824 | 2.09e-06 | -9.54 | 0.303 | -1.727 | cg23817779 |
| cg27027331 | 0.3917 | 2.97e-06 | 9.28 | 0.251 | 0.932 | cg27027331 |
| cg19255241 | 0.4114 | 3.78e-06 | 8.95 | 0.178 | 0.884 | cg19255241 |
| cg02041498 | 0.4114 | 4.87e-06 | -8.74 | 0.129 | -1.118 | cg02041498 |
| cg05075189 | 0.4114 | 4.97e-06 | -8.88 | 0.114 | -0.879 | cg05075189 |
| cg05545143 | 0.4114 | 5.05e-06 | 8.86 | 0.11 | 0.776 | cg05545143 |
| cg24949398 | 0.9704 | 1.54e-05 | 7.65 | -0.171 | 0.719 | cg24949398 |
| cg04878482 | 0.9704 | 1.67e-05 | 7.58 | -0.193 | 0.302 | cg04878482 |

Figura 11. Los resultados de los análisis con opciones gráficas generadas automáticamente por GEO eliminan la necesidad de software asociado o conocimientos profundos para seleccionar figuras específicas.

pio GEO para comparar grupos de muestras dentro de un estudio y obtener listas de genes diferencialmente expresados o metilados con métodos estadísticos consolidados (por ejemplo, limma). Presenta resultados tabulados y gráficas básicas, y ofrece el código empleado para fomentar la transparencia.

La Figura 8 muestra cómo iniciar el análisis después de seleccionar a las participantes.

Otras opciones de análisis (Figura 9) y también el código del script se pueden consultar en la interfaz (Figura 10). Aunque este estudio se centra en la metilación del ADN, se utiliza la función limma, que es el estándar de oro para analizar datos de expresión génica y también puede usarse para analizar datos de metilación de ADN [7].

La Figura 11 muestra los resultados de este análisis de ejemplo; se generan varios gráficos y análisis sin necesidad de conocimientos de programación.

En este ejercicio ilustrativo, el análisis mostró señales diferenciales en dos sitios CpG, cg08977382 (gen EDARADD) y cg13920856 (gen TANC1), los cuales pueden observarse en el gráfico de volcán (puntos rojos, Figura 12). Este tipo de visualización es ampliamente utilizado en estudios epigenéticos para mostrar, de manera exploratoria, la magnitud del cambio y su significancia estadística [8]. Los resultados aquí presentados se incluyen únicamente para ejemplificar el tipo de salida que GEO2R puede generar a partir de un conjunto de datos público. Debido al tamaño muestral reducido, al desbalance entre grupos y al carácter exploratorio del ejercicio, estos hallazgos no deben interpretarse como evidencia concluyente de asociación biológica o clínica, sino como una demostración del potencial de la herramienta para apoyar análisis iniciales y la generación de hipótesis [9].

Más allá de la obtención de tablas y gráficos, el análisis con GEO2R debe entender-

se como una aproximación exploratoria y no como una conclusión definitiva. Su interpretación requiere revisar metadatos, calidad muestral, balance entre grupos, covariables y posibles efectos de lote, además de contrastar los hallazgos con la literatura y, de ser posible, con otros conjuntos de datos o herramientas complementarias. En este sentido, GEO2R permite generar hipótesis iniciales a partir de datos públicos y familiarizar al usuario con la selección de muestras, la definición de grupos de comparación y la revisión de resultados estadísticos básicos en un entorno accesible, sin que ello implique establecer conclusiones biológicas o clínicas sin validación adicional.

Ciencia inclusiva: una oportunidad para los países con menos recursos

Plataformas como GEO ayudan a reducir la brecha entre centros ricos y pobres: con datos abiertos, un estudiante de una universidad rural puede analizar la misma información que un grupo de élite, sin invertir en laboratorios caros. Así, los países con menos recursos no solo consumen conocimiento, también generan y comparten evidencia sobre sus propios problemas de salud, incluyendo enfermedades desatendidas y variantes genéticas locales.

Impacto en formación e innovación en América Latina

En América Latina, GEO permite “aprender haciendo”: cursos cortos pueden incluir prácticas con GEO2R para que estudiantes desarrollen habilidades básicas en análisis de datos biomédicos usando estudios reales [10]. De este modo, GEO se convierte en un punto de encuentro global y un ejemplo de cómo los datos abiertos democratizan la ciencia y favorecen innovaciones en salud pública más allá de unos pocos centros privilegiados.

Conclusiones

GEO y las plataformas complementarias de análisis fácil constituyen una palanca para acelerar la investigación biomédica con criterios de apertura, eficiencia y rigor. El reuso de datos hace posible que equipos con distintas realidades presupuestales participen en la generación de conocimiento, a la vez que mejora la transparencia y la reproducibilidad. Con preguntas bien planteadas, documentación cuidadosa y una combinación de herramientas accesibles, es factible transformar matrices públicas en hipótesis sólidas, listas para validación experimental y comunicación responsable.

Agradecimientos

Agradecemos a la Universidade do Estado de Minas Gerais (UEMG), a la Universidade de São Paulo (USP) y a la Universidad Autónoma del Estado de Morelos (UAEM) por el apoyo institucional brindado durante el desarrollo de este trabajo. Asimismo, expresamos nuestro agradecimiento a la Dra. Natalia Yumi Noronha por sus valiosas contribuciones académicas y apoyo científico durante la preparación del manuscrito. También extendemos nuestro agradecimiento a la Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG), cuyo respaldo fue fundamental para la realización de este estudio.

Referencias

- [1] Clough, E., Barrett, T., Wilhite, S. E., Ledoux, P., et al. (2024). NCBI GEO: archive for gene expression and epigenomics data sets: 23-year update. *Nucleic Acids Research*, 52(D1), D138–D144. <https://doi.org/10.1093/nar/gkad965>
- [2] Kesh, S., & Raghupathi, W. (2004). Critical issues in bioinformatics and computing. *Perspectives in health information management/AHIMA, American Health Infor-*

mation Management Association, 1, 9.

- [3] Wang, Z., Lachmann, A., & Ma'ayan, A. (2019). Mining data and metadata from the Gene Expression Omnibus. *Biophysical Reviews*, 11(1), 103–110. <https://doi.org/10.1007/s12551-018-0490-8>
- [4] Kułak, K., Sztybór, I., & Kamińska, K. (2024). Obesity—an epidemic of the 21st century—literature review. *Journal of Education, Health and Sport*, 70, 49557–49557.
- [5] Marchant, E., Singh, E., Kureel, S., Blair, B., Maroto, R., Sheetz, M., & Rasmussen, B. (2024). Abstract 2197 Whole-body ultrasound treatment increases insulin production and lowers fasting blood glucose in aged insulin-resistant mice. *Journal of Biological Chemistry*, 300(3). <https://doi.org/10.1016/j.jbc.2024.106486>
- [6] Alsulami S, Nyakotey DA, Dudek K, Bawah AM, Lovgrove JA, Annan RA, Ellahi B, Vimalaswaran KS. Interaction between metabolic genetic risk score and dietary fatty acid intake on central obesity in a Ghanaian population. *Nutrients* 12: 1906, 2020. <https://doi.org/10.3390/nu12071906>
- [7] Sahoo, K., & Sundararajan, V. (2024). Methods in DNA methylation array dataset analysis: A review. *Computational and Structural Biotechnology Journal*, 23, 2304–2325. <https://doi.org/10.1016/j.csbj.2024.05.015>
- [8] Chen, J. Q., Salas, L. A., Wiencke, J. K., Koestler, D. C., Molinaro, A. M., Andrew, A. S., & Christensen, B. C. (2022). Immune profiles and DNA methylation alterations related with non-muscle-invasive bladder cancer outcomes. *Clinical Epigenetics*, 14(1), 14. <https://doi.org/10.1186/s13148-022-01234-6>
- [9] Han, M. R., Jeong, J. H., Kim, Y. G., Yang, H. H., Seo, C. O., Kim, Y., ... & Choi, J. I. (2024). Epigenetic regulation on left atrial function and disease recurrence after catheter ablation in atrial fibrillation. *Clinical Epigenetics*, 16(1), 183. <https://doi.org/10.1186/s13148-024-01794-9>
- [10] Chen, E. Y., Tan, C. M., Kou, Y., Duan, Q., Wang, Z., et al. (2013). Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics*, 14, 128. <https://doi.org/10.1186/1471-2105-14-128>