

RESEARCH

# Explorando los Misterios de la Megalencefalia: Avanzando en la Comprensión y Tratamiento de esta Anomalía Cerebral

Miguel Ángel Ruiz Villarrazo\*, Florín Babusca Voicu, Raúl Obrero Berlanga and Claudia Vega Rodriguez

\*Correspondence:  
mruizvillarrazo@gmail.com  
ETSI Informática, Universidad de  
Málaga, Málaga, España  
Full list of author information is  
available at the end of the article

## Abstract

**Keywords:** sample; article; author

## 1 Introducción

La megalencefalia se puede traducir en un trastorno del desarrollo caracterizado por un crecimiento excesivo de las estructuras cerebrales y un aumento del número de neuronas y células gliales como consecuencia de acontecimientos anormales posnatales. Se presenta como dos desviaciones estándar del perímetro cefálico por encima de la media correspondiente a la edad del paciente. Este trastorno suele provocar epilepsia, discapacidades en el desarrollo y problemas de conducta [?]. No se debe confundir la megalencefalia con la macrocefalia, pues bien pueden coexistir, presentan diferentes evaluaciones clínicas, pronóstico y tratamiento. A día de hoy, se conocen las bases moleculares de muchos de estos trastornos, lo cual permite acercarse un poco más a entender como la desregulación de ciertas vías puede conducir a la enfermedad, y con ello, un posible diagnóstico e intervención terapéutica [?].

Cuando se habla de las consecuencias de la patología, pueden llegar a surgir enfermedades como la de Canavan [?] y la enfermedad de Alexander [?], que conduce a problemas neurológicos [?], y tienen en común la acumulación de sustancias tóxicas en el cerebro.

Eventualmente, cuando se habla de los familiares de una persona que padece la patología, la calidad de vida puede ser extremadamente desafiante. El cuidado de un ser querido con necesidades médicas y neurológicas complejas puede ser emocionalmente agotador y financiera y físicamente demandante [?]. El apoyo continuo y los recursos médicos son esenciales para brindar la mejor atención posible a quienes padecen estas enfermedades y para mejorar la calidad de vida de sus familias.

La megalencefalia dispone de una prevalencia del 2% de la población infantil, lo que equivale a que 2 de cada 50 infantes presentan este fenotipo [?]. No obstante, aunque esta patología está relacionada con severas mutaciones génicas y moleculares [?], está también ligado al autismo, en el que aproximadamente el 15% de los niños presentan esta malformación [?]. Además, también se ha observado que afecta más a varones que a mujeres [?].

Este trastorno puede ser debido a tres causas principales: a un problema metabólico, del desarrollo o lesiones cerebrales. Por lo general, las megalencefalías metabólicas están causadas por anomalías genéticas en el funcionamiento celular. En cambio, se han constatado recientemente que las megalencefalías de origen del desarrollo están ocasionadas por desajustes en las vías de comunicación que supervisan la multiplicación, el desarrollo y el traslado de las neuronas [?].

En cuanto a los factores de riesgo podemos caracterizar 4 factores principalmente: Antecedentes familiares (La presencia de antecedentes familiares aumenta el riesgo), Trastornos genéticos (Trastornos genéticos que estén relacionados con la megalencefalia aumentan la probabilidad), Exposición prenatal a agentes teratogénicos (La exposición a infecciones virales o drogas aumenta el riesgo en el feto) y lesiones Cerebrales previas [?].

La principal motivación de la investigación es allanar el camino hacia la comprensión y eventual tratamiento de la megalencefalia, abordando esta condición desde múltiples perspectivas. De esta manera, se posibilitará a las personas afectadas con esta anomalía cerebral el acceso a diagnósticos más precisos y tratamientos efectivos, infundiéndole así esperanza y mejorando la calidad de sus vidas de manera significativa.

En cuanto a trabajos relacionados, centrándose más en el ámbito biológico del gen, se puede encontrar algunos como los siguientes, que estudian más la megalencefalia a fondo. Se tiene el primer estudio que relata acerca de una variación de la megalencefalia, como es la megalencefalia capilar [?] y [?], que se relata información más anatómica comparándolo con la microencefalia.

### 1.1 Hipótesis de trabajo

Las hipótesis que se plantean en este proyecto son las siguientes:

- H1: Se espera que con toda la información recolectada de diversas fuentes y bases de datos acerca de la patología, diseñar una red en la cual se muestre la relación entre la enfermedad y sus síntomas.
- H2: Cuando se superpone la información de la red que conecta genes con enfermedades y las interacciones físicas entre proteínas, se descubren conexiones inesperadas entre enfermedades y síntomas que previamente no se conocían.

### 1.2 Objetivos

- 1 Crear una red que relaciona enfermedades con síntomas a partir de la información recopilada de bases de datos de investigación y metadatos de diagnósticos.
- 2 Incorporar los datos de asociaciones entre genes y enfermedades, así como las interacciones entre proteínas, a la red que vincula enfermedades con síntomas.
- 3 Realizar un análisis topológico de las redes resultantes. Esto implica examinar la estructura y las propiedades de las redes que hemos construido para comprender mejor sus conexiones y patrones.

## 2 Materiales y métodos

A lo largo de la investigación, se ha hecho uso de diversos programas, herramientas y bases de datos para la adecuada puesta en marcha del estudio. Entre las bases de datos podemos encontrar **HPO (Human Phenotype Ontology)** [?] que es una ontología formal de fenotipos humanos y **StringDB** [?], base de datos la cual se ha usado poder realizar la interacción entre las proteínas del fenotipo de interés. Posteriormente, se han procesado los datos obtenidos para hacer uso de ellos en R y Python con el paquete de **iGraph** [?] y así, esbozar los grafos que se han considerado de interés en nuestro estudio.

### 2.1 Herramientas y materiales usados

Los elementos que han sido utilizados durante la realización de la investigación, los materiales, incluyendo marcas, modelos y proveedores, así como la especificación de cualquier otro instrumento utilizado en los experimentos son:

Equipo Usado:

- Portátil ROG STRIX G15:
  - Ubuntu 22.04.3 LTS
  - Procesador: Intel(R) Core(TM) i7-10750H CPU @ 2.60 GHz
  - RAM: 16,0 GB
  - Sistema: Sistema operativo de 64 bits, procesador basado en x86-64
  - Fabricante: ASUSTeK COMPUTER INC.

Bases de datos:

- **StringDB**: es una base de datos utilizada en biología de sistemas y genómica funcional. Su función principal ha sido proporcionar información sobre las interacciones entre proteínas del fenotipo de interés. La base de datos incluye datos experimentales y predicciones computacionales sobre estas interacciones, permitiendo así explorar y comprender las relaciones entre proteínas.
- **HPO**: Ontología de Fenotipos Humanos, es un sistema estandarizado para describir fenotipos humanos en el contexto de enfermedades genéticas. Organiza términos en una estructura jerárquica, establece relaciones entre ellos y se utiliza ampliamente en genómica clínica para analizar la asociación entre genotipos y fenotipos. Se ha usado para obtener los datos del fenotipo de Megalencefalia.

El conjunto de software utilizado en el proyecto incluye varias herramientas clave. En primer lugar, se emplea **Visual Studio Code (VS Code)** [?], un editor de código fuente gratuito y de código abierto desarrollado por Microsoft. Este editor es altamente personalizable y extensible, con un rico ecosistema de extensiones que permiten ampliar su funcionalidad. Adicionalmente, VS Code ofrece soporte para una amplia gama de lenguajes de programación, siendo Python uno de los principales enfoques.

Para la manipulación eficiente y el análisis de datos, se utiliza la biblioteca de Python llamada **Pandas** [?]. Esta biblioteca es conocida por proporcionar estructuras de datos flexibles y herramientas de análisis de datos eficaces.

En el ámbito de las solicitudes HTTP y la interacción con APIs web, se integra la biblioteca **Requests** [?] de Python. Esta biblioteca facilita la realización de solicitudes HTTP, siendo esencial para la comunicación con APIs web, como la API de StringDB.

En el contexto del lenguaje de programación R, se utiliza el entorno de desarrollo integrado (IDE) **RStudio**. Rstudio [?] ofrece herramientas avanzadas y una interfaz amigable, siendo ampliamente utilizado en el ámbito de estadísticas y análisis de datos.

Complementando estas herramientas, se incorpora la biblioteca **iGraph**, especializada en el análisis de redes y grafos. iGraph proporciona funciones y herramientas para la creación, visualización y análisis de estructuras de red en diversos contextos. Esta biblioteca resulta especialmente útil en la representación y exploración de relaciones entre nodos en un grafo. En resumen, la combinación de Visual Studio Code, Pandas, Requests, RStudio e iGraph proporciona un conjunto integral de herramientas para el desarrollo, análisis y visualización de datos en el proyecto.

Para asegurar la reproducibilidad, se aconseja tener en cuenta que la versión de R utilizada es 4.3.1 y la versión de Python que se empleará es 3.10.

## 2.2 Diseño y procedimiento experimental

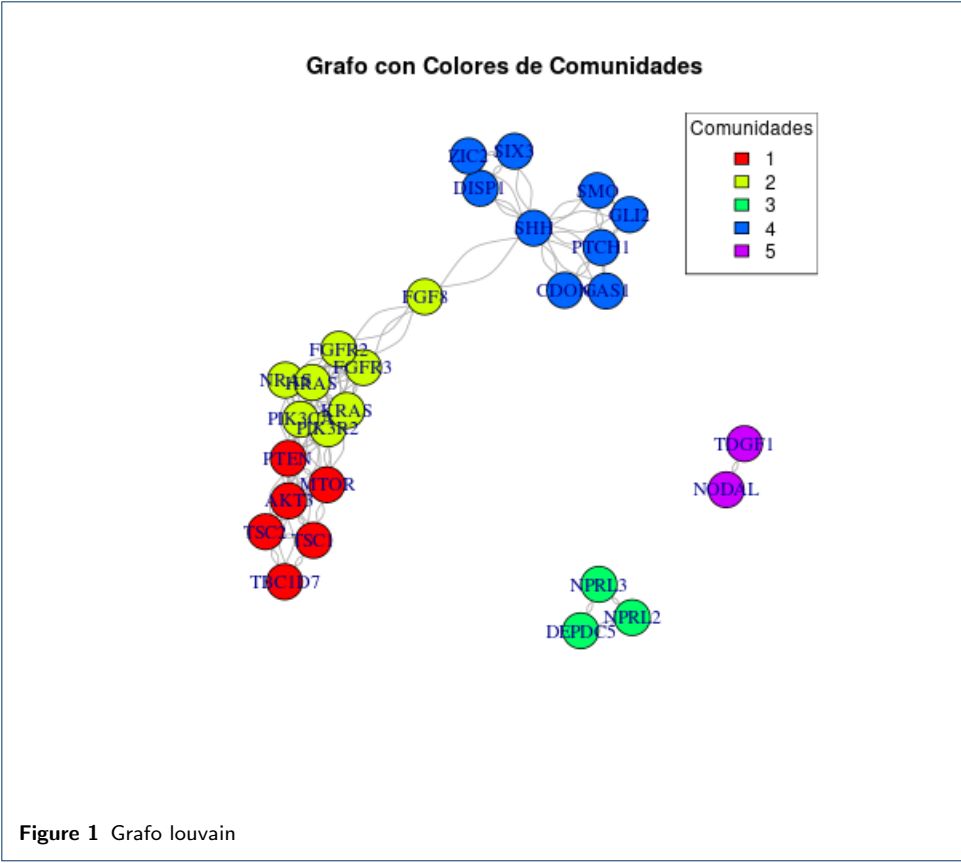
El presente estudio se basó en la recopilación de datos relacionados con megalencefalia desde la HPO. En concreto, se obtuvo el archivo de asociaciones de genes correspondiente al fenotipo HP:0001355 (megalencefalia). Posteriormente, se realizó el procesamiento de los datos mediante operaciones de limpieza y transformación para garantizar la calidad y consistencia de los datos.

Posteriormente, se empleó la API de StringDB para generar un modelo de interacción entre proteínas asociadas a megalencefalia. De la información devuelta, se eligió la más relevante para nuestro estudio, para luego guardarse y generar los grafos detallados de las relaciones entre las proteínas en Rstudio.

Como siguiente paso, se cargaron en Rstudio los datos pertenecientes a la red de interacción de proteínas obtenidas mediante la herramienta de StringDB y se creó el grafo de interacciones de la siguiente manera:

Durante el análisis de interacciones de proteínas, se identificaron nodos únicos a partir de las interacciones, generando un grafo no dirigido que representa estas relaciones. Se utilizaron los algoritmos Louvain [?] y *edge betweenness* [?] para identificar comunidades en un grafo de interacciones de proteínas, siendo Louvain elegido por su eficiencia y capacidad para optimizar la modularidad en grandes conjuntos de datos. Este algoritmo destaca por su rapidez en el manejo de datos extensos y su flexibilidad para detectar comunidades superpuestas. Las comunidades fueron visualizadas mediante colores para mayor claridad. El segundo algoritmo, *edge betweenness*, se utilizará para una comparación adicional entre ambos enfoques.

Además, se desarrolló una función para realizar análisis de enriquecimiento funcional en cada comunidad y se aplicó la función a cada comunidad identificada, proporcionando información sobre las funciones biológicas asociadas. Sumado a esto, se aplicó el algoritmo de detección de comunidades por enlace al grafo y, de nuevo, se asignaron colores a las comunidades por enlace y se visualizó el grafo. Se compararon las dos comunidades a través de la modularidad [?]. También, se ha producido una comparación de las dos comunidades, observando qué nodos se encontraban en el mismo cluster y cuáles no. Por último, se ha generado un enriquecimiento funcional a las comunidades identificadas por enlace con el método ganador y se



obtuvo información sobre las funciones biológicas asociadas a estas comunidades con que conjunto de genes dentro de la comunidad lo provocaban.

**3 Resultados**

Los resultados finales que se han obtenido son el enriquecimiento funcional de las poblaciones que fueron generadas aplicando la metodología descrita anteriormente.

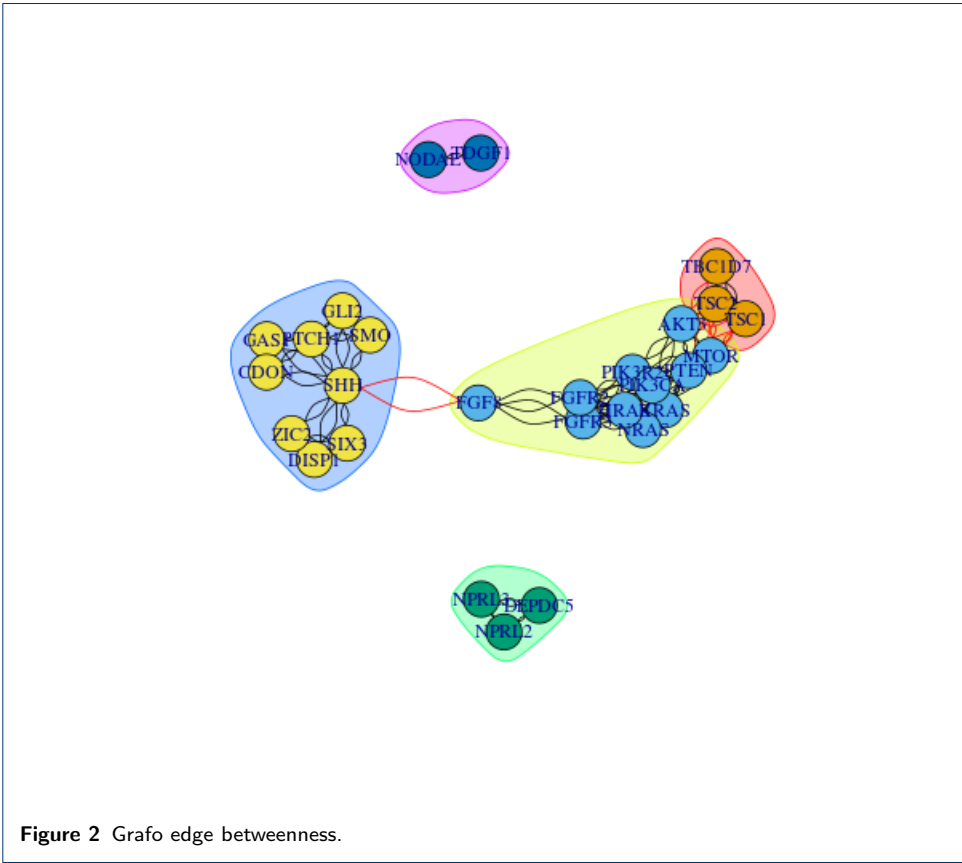
En primer lugar, se obtuvo la división de población a través del algoritmo de Louvain, que, como se ha observado en la Figura 1, existen 5 comunidades bastante pronunciadas.

También, se ha obtenido la división por el algoritmo de edge betweenness, que como se ha observado en la Figura 2, son las mismas comunidades pero con una ligera diferencia entre dos comunidades.

Para facilitar la comprensión de los diferentes modelos, se muestra en la Tabla 3 una comparación por nodos. Las filas indican los nodos, La primera columna representa el algoritmo de louvain, y la segunda a la de edge betweenness. Los valores de las celdas indican en que comunidad se encuentra este nodo. Finalmente, en la última columna se muestra si el nodo es asociado a la misma comunidad ("iguales"), o bien, si pertenecen a diferentes comunidades ("diferentes").

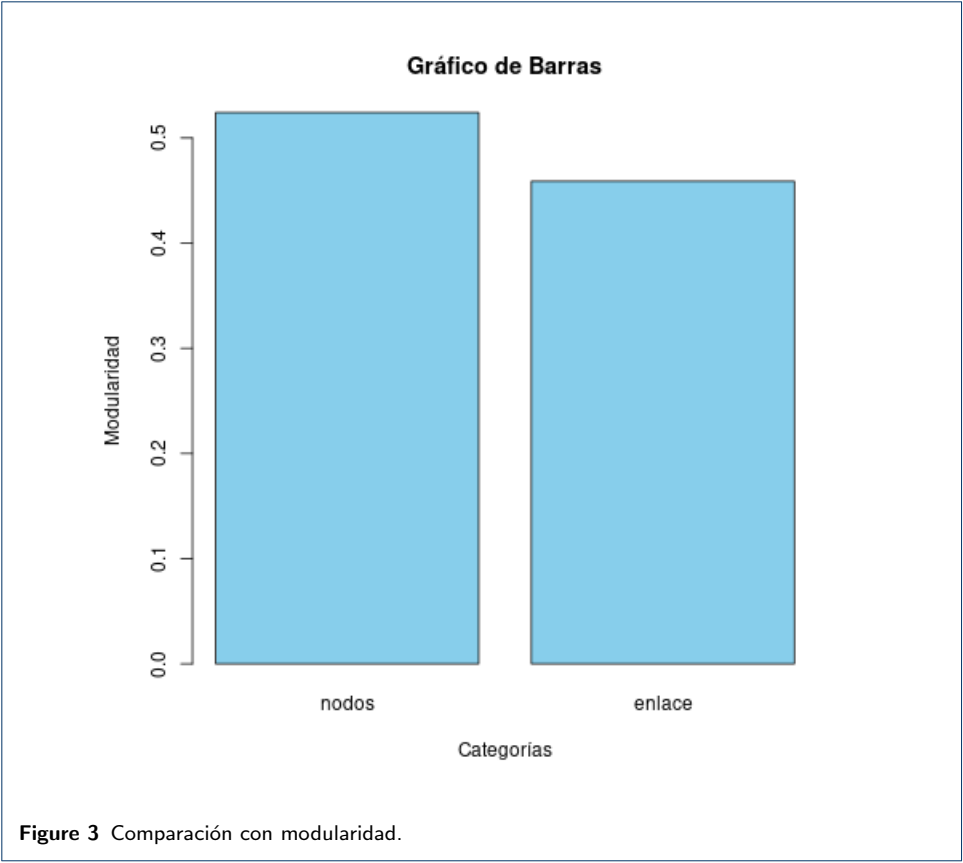
Posteriormente, se calculó la modularidad y se representó en la Figura 3.

Finalmente, se obtuvo el enriquecimiento funcional del mejor algoritmo basándonos en la modularidad. Para representar las funciones asociadas a cada comunidad, se han creado tantas tablas como comunidades (ver Figuras 3, 3, 3, 3 y 3), en las que,



**Table 1** Descripción de Nodos, Enlaces y Diferentes

Nodos	Enlaces	Diferentes
1	1	Iguals
2	2	Iguals
3	3	Iguals
4	4	Iguals
2	2	Iguals
4	4	Iguals
2	2	Iguals
4	4	Iguals
5	5	Iguals
4	4	Iguals
1	1	Iguals
4	4	Iguals
2	2	Iguals
4	4	Iguals
2	2	Iguals
1	2	Diferentes
2	2	Iguals
1	2	Diferentes
2	2	Iguals
1	1	Iguals
3	3	Iguals
1	2	Diferentes
2	2	Iguals
3	3	Iguals
4	4	Iguals
4	4	Iguals
5	5	Iguals
4	4	Iguals



**Table 2** Descripción de Genes y Funciones de la comunidad 1

Genes	Funciones
TSC2, TSC1, MTOR, PTEN, TBC1D7, AKT3	negative regulation of intracellular signal transduction (GO:1902532)

**Table 3** Descripción de Genes y Funciones de la comunidad 2

Genes	Funciones
PIK3R2, KRAS, PIK3CA, FGF8, FGFR3, NRAS, HRAS, FGFR2	MAPK cascade (GO:0000165)
PIK3R2, KRAS, PIK3CA, FGF8, FGFR3, NRAS, HRAS, FGFR2	Transmembrane receptor protein tyrosine kinase signaling pathway (GO:0007169)
PIK3R2, KRAS, PIK3CA, FGF8, FGFR3, NRAS, HRAS, FGFR2	Positive regulation of intracellular signal transduction (GO:1902533)

en la primera columna, se muestra el conjunto de nodos dentro de la comunidad con una función en común, y en la segunda columna, la propia función.

**Table 4** Descripción de Genes y Funciones de la comunidad 3

Genes	Funciones
NPRL2, NPRL3, DEPDC5	Cellular response to amino acid starvation (GO:0034198)
NPRL2, NPRL3, DEPDC5	Negative regulation of TOR signaling (GO:0032007)
NPRL2, NPRL3, DEPDC5	Regulation of TOR signaling (GO:0032006)
NPRL2, NPRL3, DEPDC5	Cellular response to starvation (GO:0009267)
NPRL2, NPRL3, DEPDC5	Negative regulation of intracellular signal transduction (GO:1902532)

**Table 5** Descripción de Genes y Funciones de la comunidad 4

Genes	Funciones
SMO, SIX3, DISP1, SHH, GAS1, PTCH1, GLI2, ZIC2, CDON	Smoothened signaling pathway (GO:0007224)
SMO, SIX3, DISP1, SHH, GAS1, PTCH1, GLI2, ZIC2, CDON	Negative regulation of transcription, DNA-templated (GO:0045892)

**Table 6** Descripción de Genes y Funciones de la comunidad 5

Genes	Funciones
NODAL, TDGF1	BMP signaling pathway (GO:0030509)
NODAL, TDGF1	Cellular response to BMP stimulus (GO:0071773)
NODAL, TDGF1	Transmembrane receptor protein serine/threonine kinase signaling pathway (GO:0007178)
NODAL, TDGF1	Positive regulation of protein phosphorylation (GO:0001934)
NODAL, TDGF1	Positive regulation of cell proliferation (GO:0008284)
NODAL, TDGF1	Regulation of apoptotic process (GO:0042981)
NODAL, TDGF1	Regulation of transcription from RNA polymerase II promoter (GO:0006357)



## 4 Discusión

### 4.1 Limitaciones

Una limitación significativa de este proyecto reside en la falta de sujetos humanos reales para estudiar la megalencefalia. Dada la naturaleza de nuestro enfoque, nos basamos en datos y simulaciones, sin contar con la participación directa de individuos afectados. Aunque se ha empleado un enfoque riguroso utilizando modelos y datos disponibles, la ausencia de casos clínicos reales podría afectar la generalización de los resultados a situaciones específicas de pacientes con megalencefalia. Se recomienda cautela al extrapolar los hallazgos a poblaciones humanas y se sugiere la validación adicional mediante estudios clínicos con participantes reales en futuras investigaciones.

## 5 Conclusiones

### Abreviaciones

Indicar lista de abreviaciones mostrando cada acrónimo a que corresponde

### Disponibilidad de datos y materiales

Debéis indicar aquí un enlace a vuestro repositorio de github.

### Contribución de los autores

Usando las iniciales que habéis definido al comienzo del documento, debéis indicar la contribución al proyecto en el estilo: J.E : Encargado del análisis de coexpresión con R, escritura de resultados; J.R.S : modelado de red con python y automatizado del código, escritura de métodos; ... OJO: que sea realista con los registros que hay en vuestros repositorios de github.