

RESEARCH

Análisis de redes génicas asociadas al fenotipo HP:0002344 – Progressive neurologic deterioration mediante biología de sistemas

Patricia Rodríguez Lidueña*, Achraf Ousti El Moussati, Aissa Omar El Hammouti Chachoui and Hugo Salas Calderón

*Correspondence: capi13@uma.es
ETSI Informática, Universidad de Málaga, Málaga, España
Full list of author information is available at the end of the article

Abstract

Las enfermedades neurodegenerativas tienen en común el fenotipo HP:0002344 – Progressive neurologic deterioration, caracterizado por la pérdida progresiva de funciones neuronales. Su heterogeneidad genética dificulta la identificación de mecanismos comunes, por lo que la biología de sistemas ofrece un marco integrador para su estudio. El presente trabajo analiza cuatro patologías representativas: Spinocerebellar ataxia with epilepsy, Leigh syndrome, Frontotemporal lobar degeneration with TDP-43 inclusions y Rett syndrome, asociadas a los genes MECP2, IARS2 y GRN. Se propone integrar información de bases de datos biomédicas y construir redes de interacción proteína–proteína (PPI) para identificar rutas moleculares compartidas, genes candidatos y posibles dianas terapéuticas implicadas en la neurodegeneración progresiva.

Keywords: sample; article; author

1 Introducción

Las enfermedades neurodegenerativas constituyen uno de los mayores desafíos biomédicos contemporáneos debido a su alta prevalencia y a la ausencia de tratamientos curativos eficaces [1]. Aunque sus manifestaciones clínicas son diversas, comparten un rasgo común: la degeneración neurológica progresiva, entendida como un deterioro gradual e irreversible de las funciones neuronales que afecta al movimiento, la cognición y la función sensorial [2].

Este rasgo se encuentra descrito en la Ontología del Fenotipo Humano (HPO) bajo el identificador HP:0002344 – Progressive neurologic deterioration, que agrupa un conjunto heterogéneo de enfermedades hereditarias y adquiridas, entre ellas ataxias, encefalopatías metabólicas y demencias frontotemporales [3]. Aunque cada una de estas patologías es ultrarrara (frecuencia menor 1:1.000.000), su frecuencia combinada alcanza aproximadamente 1 por cada 100.000 personas, lo que convierte a este fenotipo en un problema relevante dentro de las enfermedades neurológicas raras [4].

La HPO constituye una herramienta clave para vincular fenotipos clínicos con genes causales mediante un vocabulario jerárquico estandarizado [5]. Esta ontología ha mejorado la interpretación de variantes genéticas y ha facilitado el diagnóstico computacional de enfermedades raras basadas en descripciones fenotípicas precisas. Sin embargo, la heterogeneidad genética y fenotípica de las enfermedades asociadas

a HP:0002344 dificulta la identificación de mecanismos comunes de patogénesis. Para acotar el análisis, nos centraremos en un conjunto representativo de cuatro patologías con este fenotipo: Spinocerebellar ataxia with epilepsy (ORPHA:778), Leigh syndrome (ORPHA:506), Frontotemporal lobar degeneration with TDP-43 inclusions (OMIM:607485) y Rett syndrome (ORPHA:778). Estas patologías ejemplifican distintos mecanismos de neurodegeneración —epigenético, metabólico y proteopático—, permitiendo explorar posibles rutas convergentes en el deterioro neurológico progresivo.

En la ataxia espinocerebelosa con epilepsia y el síndrome de Rett, las mutaciones en [6] [7] MECP2 alteran la regulación epigenética y la expresión génica neuronal. MECP2 codifica una proteína que se une al ADN metilado y modula la transcripción de genes esenciales para la plasticidad sináptica y la maduración neuronal; su disfunción produce desequilibrios en la excitabilidad cortical y pérdida progresiva de funciones cognitivas y motoras [6].

La enfermedad de Leigh, asociada a mutaciones en IARS2 [8], afecta la función de la isoleucil-ARNt sintetasa mitocondrial, lo que compromete la traducción proteica dentro de la mitocondria y conduce a un déficit energético severo, estrés oxidativo y degeneración neuronal en regiones con alta demanda metabólica [8]. Por su parte, la degeneración lobar frontotemporal con inclusiones de TDP-43 [9], causada por mutaciones en GRN, se caracteriza por la pérdida de programulina, una proteína implicada en la homeostasis lisosomal, la inflamación y la supervivencia neuronal. La deficiencia de programulina conduce a la acumulación de inclusiones citoplasmáticas de TDP-43, disfunción sináptica y neuroinflamación crónica [9].

A partir de esta base, se plantea la hipótesis de que la degeneración neurológica progresiva surge como resultado de alteraciones en redes moleculares interconectadas, compartidas por distintas enfermedades raras con orígenes genéticos diversos.

La integración de información procedente de bases de datos como HPO, OMIM, STRINGdb y GeneCards permitirá construir una visión sistémica del fenotipo HP:0002344, identificando redes funcionales compartidas, genes candidatos no descritos y nodos terapéuticos potenciales. Este enfoque de biología de sistemas busca arrojar luz a los mecanismos moleculares de convergencia que subyacen al deterioro neurológico progresivo y aportar una base para el desarrollo de estrategias terapéuticas dirigidas [10, 11, 12].

2 Objetivos

El presente trabajo tiene como finalidad estudiar los mecanismos moleculares implicados en el fenotipo **HP:0002344 – Progressive neurologic deterioration** desde un enfoque de biología de sistemas. Para ello, se plantean los siguientes objetivos específicos:

- 1 **Identificar** los genes y las enfermedades asociadas al fenotipo HP:0002344 a partir de bases de datos como *HPO*, *OMIM* y *GeneCards*.
- 2 **Describir** los mecanismos moleculares implicados en las enfermedades representativas —Spinocerebellar ataxia with epilepsy, Leigh syndrome, Frontotemporal lobar degeneration y Rett syndrome— y sus genes causales (*MECP2*, *IARS2* y *GRN*).

- 3 **Construir y analizar** redes de interacción proteína-proteína (*PPI*) a partir de los genes identificados mediante la base de datos *STRINGdb*, evaluando métricas topológicas como el grado, la centralidad y la modularidad.
- 4 **Detectar** grupos funcionales de genes y realizar un análisis de enriquecimiento para identificar procesos biológicos y rutas moleculares comunes.
- 5 **Desarrollar** y documentar el código necesario para reproducir el análisis completo y garantizar la trazabilidad de los resultados en el repositorio de *Github*.

3 Materiales y Métodos

3.1 Materiales

El análisis se basó en información procedente de bases de datos biomédicas de referencia. Se seleccionó el término **HP:0002344 – Progressive neurologic deterioration** de la Human Phenotype Ontology (HPO) [10] como punto de partida, recopilando los genes asociados a este fenotipo. La relación gen–enfermedad se contrastó con OMIM [13] y GeneCards [14], lo que permitió confirmar y ampliar la evidencia disponible sobre cada gen.

Las interacciones entre las proteínas codificadas por dichos genes se obtuvieron mediante STRINGdb (versión 12.0) [15], considerando únicamente interacciones de alta confianza (combined score ≥ 0.7) en *Homo sapiens*.

El procesamiento y análisis de redes se realizaron en R (versión 4.3.2), empleando los paquetes *igraph*, *tidygraph*, *ggraph* y *linkcomm* [16, 17, 18, 19]. Adicionalmente, se utilizaron los paquetes *clusterProfiler*, *enrichplot* y *org.Hs.eg.db* para el análisis de enriquecimiento funcional. Todos los scripts, tablas y visualizaciones se encuentran disponibles en el repositorio de GitHub del proyecto, garantizando la reproducibilidad del estudio.

3.2 Metodología

3.2.1 Flujo de trabajo general

El procedimiento seguido en el estudio puede resumirse en las siguientes etapas:

- 1 Selección del fenotipo HP:0002344 en HPO.
- 2 Extracción de genes asociados al fenotipo mediante la API de HPO.
- 3 Validación de las asociaciones gen–enfermedad mediante OMIM y GeneCards.
- 4 Obtención de interacciones proteína–proteína (*PPI*) en STRINGdb.
- 5 Construcción y depuración de la red en R.
- 6 Cálculo de métricas estructurales de la red.
- 7 Identificación de comunidades mediante Louvain y Walktrap.
- 8 Análisis detallado de los genes de interés: *MECP2*, *IARS2* y *GRN*.
- 9 Enriquecimiento funcional de los módulos identificados (GO y KEGG).
- 10 Interpretación integrada de resultados.

3.2.2 Recopilación de genes y construcción de la red

A partir del término HP:0002344 se obtuvo programáticamente la lista de genes relacionados mediante la API REST de HPO [4]. La consulta se realizó mediante el endpoint <https://ontology.jax.org/api/network/annotation/HP:0002344>, utilizando el paquete *httr* de R para realizar las peticiones HTTP y el paquete *jsonlite* para parsear las respuestas JSON.

Tras eliminar duplicados y valores nulos, se verificó la validez de las asociaciones gen-enfermedad mediante OMIM [13] y GeneCards [14], obteniendo un conjunto depurado de genes candidatos.

Estos genes se introdujeron en STRINGdb (v12.0) [15] mediante el paquete de Bioconductor **STRINGdb**, estableciendo los siguientes parámetros:

- **Organismo:** *Homo sapiens* (NCBI Taxonomy ID: 9606)
- **Umbral de confianza:** 0.7 (combined score ≥ 700), correspondiente a interacciones de alta confianza
- **Versión de la base de datos:** 12.0

Para asegurar la calidad de las interacciones, únicamente se consideraron aquellas con evidencia experimental o provenientes de bases de datos curadas. El mapeo de los símbolos génicos a identificadores STRING se realizó mediante la función `map()` del paquete **STRINGdb**, eliminando los genes que no pudieron ser mapeados.

La red resultante se construyó obteniendo todas las interacciones entre los genes mapeados mediante la función `get_interactions()`. Los datos de interacción se transformaron en un objeto de clase `igraph` para su análisis estructural, estableciendo los nombres de genes como atributos de los vértices y los scores de confianza como pesos de las aristas.

3.2.3 Análisis estructural de la red

Se caracterizó la topología global de la red mediante métricas estándar implementadas en el paquete `igraph`.

El **grado** de cada nodo (número de conexiones) se calculó mediante la función `degree()`, permitiendo identificar genes altamente conectados (posibles hubs). Las centralidades de **intermediación** (*betweenness centrality*) y **cercanía** (*closeness centrality*) se computaron mediante las funciones `betweenness()` y `closeness()` respectivamente, ayudando a detectar nodos clave en la comunicación entre módulos.

Se calculó el **coeficiente de agrupamiento** (*clustering coefficient*) local para cada nodo mediante `transitivity(type = "local")` y el coeficiente de agrupamiento global mediante `transitivity(type = "global")`, evaluando la tendencia de los nodos a formar clústeres densos.

La **modularidad** global de la red se estimó mediante la función `modularity()`, cuantificando el grado de organización de la red en comunidades funcionales. Valores de modularidad superiores a 0.3 se consideran indicativos de estructura comunitaria significativa.

Todas las métricas calculadas se consolidaron en una tabla única que incluye, para cada gen: símbolo genético, grado, betweenness, closeness, clustering coefficient y membresía comunitaria.

3.2.4 Identificación de comunidades funcionales

La detección de comunidades se realizó mediante dos algoritmos complementarios implementados en `igraph`:

- **Algoritmo de Louvain** [20]: Método de optimización multinivel que maximiza la modularidad de la red mediante agrupación jerárquica. Se aplicó mediante la función `cluster_louvain()`.

- **Algoritmo de Walktrap** [21]: Método basado en caminatas aleatorias que identifica comunidades mediante la estructura de caminos cortos en la red. Se aplicó mediante `cluster_walktrap()` con 4 pasos de caminata aleatoria.

Para cada partición resultante se calculó su modularidad y se analizó la distribución de tamaños de las comunidades. Los resultados de ambos métodos se compararon para evaluar la robustez de la estructura comunitaria detectada.

Cada comunidad se evaluó considerando su cohesión interna, su tamaño y la presencia de genes previamente relacionados con enfermedades neurológicas progresivas en la literatura.

3.2.5 Análisis de genes de interés

Se estudiaron en mayor profundidad los genes **MECP2**, **IARS2** y **GRN**, seleccionados por su relevancia clínica en patologías asociadas al deterioro neurológico progresivo.

El gen *MECP2* está implicado en el síndrome de Rett [11] y en el síndrome por duplicación *MECP2* [22]. El gen *IARS2* se ha asociado a síndrome de Leigh con deficiencia combinada en fosforilación oxidativa [12]. Por último, *GRN* está implicado en la degeneración lobar frontotemporal [23].

Para cada gen se verificó su presencia en la red y se extrajeron sus métricas topológicas (grado, centralidad de intermediaciación, coeficiente de agrupamiento).

Se generaron subredes de vecindad inmediata para cada gen de interés mediante la función `make_ego_graph()`, incluyendo el gen focal y todos sus vecinos directos (orden 1). Estas subredes locales se visualizaron mediante `ggraph`, destacando la conectividad del gen focal y su relación con nodos relevantes dentro de la red global. Los nodos se colorearon según su pertenencia a comunidades y se dimensionaron proporcionalmente a su grado.

3.2.6 Enriquecimiento funcional

El análisis de enriquecimiento funcional se realizó mediante el paquete `clusterProfiler` de Bioconductor.

Se evaluaron las categorías de **Gene Ontology** (GO) utilizando la función `enrichGO()`, con los siguientes parámetros:

- **Base de datos de anotación:** `org.Hs.eg.db`
- **Ontología:** Procesos biológicos (BP, *Biological Process*)
- **Tipo de identificador:** SYMBOL
- **Método de corrección:** Benjamini–Hochberg (BH)
- **Umbral de significancia:** p ajustado ≤ 0.05 , q-valor ≤ 0.2

El análisis se realizó tanto para la red completa como para cada comunidad individual que contenía al menos 5 genes.

Para el análisis de rutas metabólicas **KEGG**, primero se convirtieron los símbolos génicos a identificadores Entrez mediante la función `bitr()`. Posteriormente se aplicó `enrichKEGG()` con los siguientes parámetros:

- **Organismo:** `hsa (Homo sapiens)`
- **Umbral de significancia:** p-valor ≤ 0.05 , q-valor ≤ 0.2

Los resultados se visualizaron mediante gráficos de puntos (*dotplots*) y gráficos de barras, generados con las funciones `dotplot()` y `barplot()` del paquete `enrichplot`.

Este análisis permitió relacionar los módulos detectados con procesos neuronales y rutas metabólicas asociadas al deterioro neurológico progresivo, en consonancia con los mecanismos descritos en la literatura biomédica [1, 2].

3.2.7 Exportación y reproducibilidad

Todos los resultados, incluyendo métricas de nodos, resultados de enriquecimiento y la red en formatos GraphML y GML, se exportaron para su posterior análisis. El código completo del análisis se documentó en un notebook R Markdown y se encuentra disponible en el repositorio GitHub del proyecto, garantizando la total reproducibilidad del estudio.

4 Resultados

4.1 Red PPI asociada al fenotipo HP:0002344

La consulta al término HP:0002344 en la Human Phenotype Ontology produjo un conjunto inicial de **XX genes** asociados al fenotipo. Tras la depuración de duplicados y el mapeo en STRINGdb, **XX genes** fueron identificados correctamente y empleados para la construcción de la red PPI.

La red final incluyó un total de **XX nodos** y **XX interacciones** con un umbral de confianza de 0.7 (combined score ≥ 700). La densidad obtenida fue de **X.XXX**, con un grado medio de **X.X**. La Tabla ?? resume las métricas topológicas principales de la red.

4.2 Distribución de grados y nodos altamente conectados

La distribución de grados mostró valores comprendidos entre **X** y **X**. El nodo con mayor conectividad presentó un grado de **X**. El histograma de grados obtenido se presenta en la Figura ??.

4.3 Detección de comunidades

La aplicación del algoritmo Louvain detectó un total de **X comunidades**, con un valor de modularidad de **X.XXX**. Por su parte, el algoritmo Walktrap identificó **X comunidades**, alcanzando una modularidad de **X.XXX**. La Figura ?? muestra la partición obtenida mediante Louvain y la Figura ?? representa la partición correspondiente a Walktrap.

4.4 Métricas individuales y genes de interés

Las métricas de grado, centralidad de intermediación, cercanía y coeficiente de agrupamiento se calcularon para cada nodo de la red. La Tabla ?? resume estos valores.

Se generaron subredes de vecindad inmediata (orden 1) para los genes *MECP2*, *IARS2* y *GRN*. La Figura ?? muestra la subred de *MECP2* y las Figuras ?? y ?? las correspondientes a los otros genes.

5 Discusión

6 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.

Author details

ETSI Informática, Universidad de Málaga, Málaga, España.

References

1. Kelser, B.M., Teichner, E.M., Subtirelu, R.C., Hoss, K.N.: A review of proposed mechanisms for neurodegenerative disease. *Frontiers in Aging Neuroscience* **16**, 1370580 (2024)
2. Gao, H.-M.: Why neurodegenerative diseases are progressive (2008)
3. Wakap, S.N., et al.: Estimating cumulative point prevalence of rare diseases. *European Journal of Human Genetics* **28**, 165 (2020)
4. Human Phenotype Ontology Consortium: HPO Term HP:0002344. <https://hpo.jax.org/browse/term/HP:0002344>. Accessed: YYYY-MM-DD
5. Wakap, S.N., et al.: Estimating cumulative point prevalence of rare diseases. *Orphanet Journal of Rare Diseases* **14**, 261 (2019)
6. de Palma, L., Boniver, C., Cassina, M., Toldo, I., Nosadini, M., Clementi, M., Sartori, S.: Eating-induced epileptic spasms in a boy with MECP2 duplication syndrome: insights into pathogenesis of genetic epilepsies. *Epileptic Disorders* **14**(4), 414–417 (2012). doi:10.1684/epd.2012.0546
7. Ehrhart, F., Coort, S.L.M., Cirillo, E., Smeets, E., Evelo, C.T., Curfs, L.M.G.: Rett syndrome – biological pathways leading from MECP2 to disorder phenotypes. *Orphanet Journal of Rare Diseases* **11**(1), 158 (2016). doi:10.1186/s13023-016-0545-5
8. Dong, Q., Yin, X., Fan, S., Zhong, S., Yang, W., Chen, K., Wang, Q., Ma, X., Mahlatsi, R.L., Yang, Y., Lyu, J., Fang, H., Wang, Y.: lars2 mutations lead to Leigh syndrome with a combined oxidative phosphorylation deficiency. *Orphanet Journal of Rare Diseases* **19**(1), 305 (2024). doi:10.1186/s13023-024-03310-x
9. Karamysheva, Z.N., Tikhonova, E.B., Karamyshev, A.L.: Granulin in frontotemporal lobar degeneration: Molecular mechanisms of the disease. *Frontiers in Neuroscience* **13**, 395 (2019). doi:10.3389/fnins.2019.00395
10. Human Phenotype Ontology Consortium: Human Phenotype Ontology (HPO) database. <https://hpo.jax.org/>. Accessed: YYYY-MM-DD
11. Ehrhart, F., Coort, S.L.M., Cirillo, E., Smeets, E., Evelo, C.T., Curfs, L.M.G.: Rett syndrome – biological pathways leading from mecp2 to disorder phenotypes. *Orphanet Journal of Rare Diseases* **11**(1), 158 (2016). doi:10.1186/s13023-016-0545-5
12. Dong, Q., Yin, X., Fan, S., Zhong, S., Yang, W., Chen, K., Wang, Q., Ma, X., Mahlatsi, R.L., Yang, Y., Lyu, J., Fang, H., Wang, Y.: lars2 mutations lead to leigh syndrome with a combined oxidative phosphorylation deficiency. *Orphanet Journal of Rare Diseases* **19**(1), 305 (2024). doi:10.1186/s13023-024-03310-x
13. Amberger, J., Bocchini, C., Hamosh, A.: A new face and new challenges for Online Mendelian Inheritance in Man (OMIM®). *Human Mutation* (2011). doi:10.1002/humu.21466
14. Stelzer, G., Rosen, R., Plaschkes, I., Zimmerman, S., Twik, M., Fishilevich, S., Iny Stein, T., Nudel, R., Lieder, I., Mazor, Y., Kaplan, S., Dahary, D., Warshawsky, D., Guan-Golan, Y., Kohn, A., Rappaport, N., Safran, M., Lancet, D.: The genecards suite: From gene data mining to disease genome sequence analyses. *Current Protocols in Bioinformatics* **54**, 1–30113033 (2016). doi:10.1002/cpb.5
15. Szklarczyk, D., Kirsch, R., Koutrouli, M., Nastou, K., Mehryary, F., Hachilif, R., Gable, A.L., Fang, T., Doncheva, N.T., Pyysalo, S., Bork, P., Jensen, L.J., von Mering, C.: The STRING database in 2023: protein–protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Research* **51**(D1), 638–646 (2023). doi:10.1093/nar/gkac1000
16. Csárdi, G., Nepusz, T.: The igraph software package for complex network research. *InterJournal Complex Systems*, 1695 (2006)
17. Pedersen, T.L.: Tidygraph: A Tidy API for Graph Manipulation. (2025). R package version 1.3.1.9000, <https://github.com/thomasp85/tidygraph>. <https://tidygraph.data-imaginest.com>
18. Pedersen, T.L.: Ggraph: An Implementation of Grammar of Graphics for Graphs and Networks. (2025). R package version 2.2.2.9000. <https://ggraph.data-imaginest.com>
19. Kalinka, A.T., Tomancak, P.: linkcomm: an r package for the generation, visualization, and analysis of link communities in networks of arbitrary size and type. *Bioinformatics* **27**(15), 2011–2012 (2011). doi:10.1093/bioinformatics/btr311
20. Traag, V.A., Waltman, L., van Eck, N.J.: From louvain to leiden: guaranteeing well-connected communities. *Scientific Reports* **9**, 5233 (2019). doi:10.1038/s41598-019-41695-z
21. Hieu, D.D., Ha Duong, P.T.: Detecting communities in large networks using the extended walktrap algorithm, 100–105 (2022). doi:10.1109/RIVF55975.2022.10013880
22. de Palma, L., Boniver, C., Cassina, M., Toldo, I., Nosadini, M., Clementi, M., Sartori, S.: Eating-induced epileptic spasms in a boy with mecp2 duplication syndrome: insights into pathogenesis of genetic epilepsies. *Epileptic Disorders* **14**(4), 414–417 (2012). doi:10.1684/epd.2012.0546
23. Karamysheva, Z.N., Tikhonova, E.B., Karamyshev, A.L.: Granulin in frontotemporal lobar degeneration: Molecular mechanisms of the disease. *Frontiers in Neuroscience* **13**, 395 (2019). doi:10.3389/fnins.2019.00395