RESEARCH

Estudio del fenotipo Disgrafía

Nerea Martín Serrano*, Carlos Beltrán López
, Carlos Beltrán López
and Javier Mendez Parrilla

*Correspondence:
nmartins@uma.es
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 escritura es una habilidad que se desarrolla en la infancia, estamos rodeados de textos que leer y que implican nuestro día a día. La disgrafia es un trastorno de aprendizaje que surge en esta etapa del desarrollo que afecta a las habilidades de escritura [1]. Puede manifestarse mediante problemas en la memoria ortográfica a largo plazo, el proceso de conversión de sonido a escritura. Esto puede involucrar dificultades en diversos niveles, como la caligrafía, la escritura lenta y la ortografía.

La disgrafia puede tener un impacto negativo en el rendimiento escolar de los niños. Muchos niños que la sufren no pueden organizar coherentemente sus pensamientos en papel o escribir de manera legible. Esta discapacidad debe ser reconocida y tratada antes de que genere consecuencias negativas duraderas para el niño. [2]. Como tratamiento para el manejo de la disgrafia en la etapa escolar, el maestro debe tener en cuenta el contexto anamnésico (evolución de las funciones físicas, psíquicas...), sociopedagógico y datos sobre el lenguaje (vocabulario, lectura, escritura...)[3]. Para ello, se llevan a cabo intervenciones organizadas en tres categorías: acomodación, modificación y revalorización[1]. Las acomodaciones incluyen estrategias como proporcionar instrumentos de escritura especiales y permitir el uso de grabadoras y correctores ortográficos. Las modificaciones implican ajustar las expectativas académicas, dividiendo tareas extensas o permitiendo alternativas como informes orales. La revalorización se basa en un enfoque de respuesta a la intervención, es decir, un cálculo continuo del estado de su disgrafia para evaluar y proporcionar apoyo específico según las dificultades del individuo.

La causa más comúnmente propuesta es un déficit en el procesamiento fonológico, lo que dificultaría la comprensión de las relaciones entre sonidos y grafías en la escritura. Otras posibles causas distantes incluyen déficits en el dominio visual y problemas de control motor [4]. A menudo se asocia con otras dificultades específicas del aprendizaje (SLD), como la dislexia. A nivel neurofisiológico, estos trastornos parecen compartir áreas cerebrales similares [5, 6]. Aunque en la literatura actual se han llevado a cabo estudios sobre los genes que afectan a los SLD, no ha habido un consenso sobre qué genes afectan a cada SLD de manera específica [7, 8]. En este

estudio, se intentará identificar qué genes afectan específicamente a la aparición de la disgrafia, con el objetivo de diagnosticar este déficit y aplicar un tratamiento adecuado antes de que se desarrollen los síntomas.

2 Materiales y métodos

En esta sección, describiremos la metodología utilizada en el estudio de la Disgrafia, junto con los materiales empleados. La metodología se dividió en varias etapas, las cuales se detallarán a lo largo de esta sección y se pueden observar en la imagen 1.

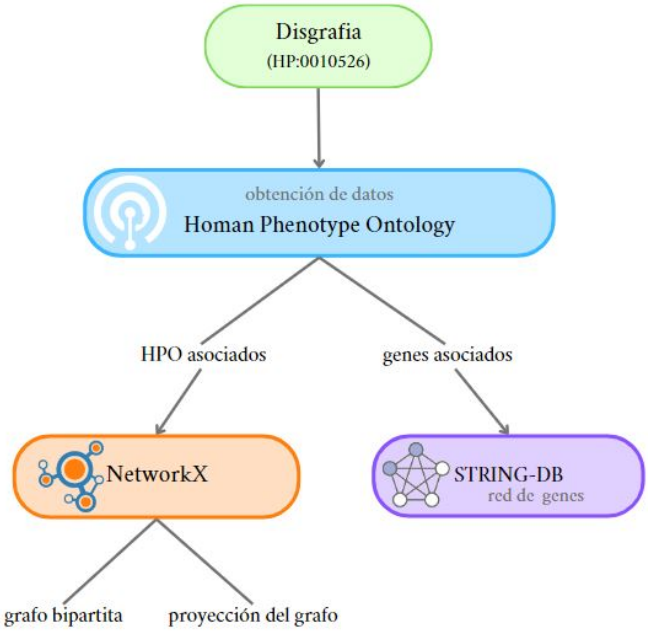

Figure 1 Flujo de trabajo

2.1 Datos biológicos

Lo primero que se realizó fue descargar los dataset necesarios para hacer el estudio. Se usaron dos base de datos: Human Phenotype Ontology (HPO) [9] y STRING-DB [13].

1 Se buscó el fenotipo Disgrafía en la HPO, conociendo que su identificador es HP:0010526. De esta base de datos, obtuvimos dos archivos tabulados; uno contiene los genes asociados y el otro contiene términos HPO asociados a la Disgrafía.

2 La base de datos STRING se usó para descargar la red de interacción de proteínas humanas

A continuación, se van a describir los distintos conjuntos de datos mencionados anteriormente.

- **HPO asociados**

Table 1 Cabecera del archivo de HPO asociados

Gene id (ncbi)	Gene symbol	HPO id	HPO name	frequency	Disease id
10	NAT2	HP:0000007	Autosomal recessive inheritance	-	OMIM:243400
10	NAT2	HP:0001939	Abnormality of metabolism/homeostasis	-	OMIM:243400
16	AARS1	HP:0002460	Distal muscle weakness	15/15	OMIM:613287
16	AARS1	HP:0002451	Limb dystonia	3/3	OMIM:616339
16	AARS1	HP:0008619	Bilateral sensorineural hearing impairment	HP:0040283	ORPHA:33364

Uno de los archivos enumera para cada gen las clases HPO más específicas. Las primeras cinco filas se pueden visualizar en la tabla 1.

La tabla 1 proporciona el identificador de gen NCBI, el símbolo del gen, el identificador HPO y el nombre del término. Si está disponible, se muestra la frecuencia. Para este campo, hay tres opciones:

1 Un identificador de término dentro de la sub-ontología de la HPO que esté relacionado con la frecuencia del fenotipo en cuestión
2 Un recuento de pacientes afectados dentro de un individuo. "7/13" indicaría que 7 de los 13 pacientes con la enfermedad especificada tienen la anormalidad fenotípica mencionada por el término de la HPO en cuestión. Por ejemplo, la mutación en el gen AARS1 causa *leucoencefalopatía*. La frecuencia del término HPO Ataxia sensorial esta anotada como 1 de 2 debido a la información en Sundal C, et al. [10].
3 Un valor porcentual. Nuevamente, esto se refiere al porcentaje de pacientes que tienen la anormalidad fenotípica mencionada por el término de la HPO.

La última columna muestra anotaciones realizadas por el equipo HPO (utilizando identificadores de enfermedades de OMIM), así como anotaciones proporcionadas por el equipo de Orphanet [11] (utilizando identificadores de enfermedades de ORPHA).

El archivo se introdujo en Python como un data frame utilizando la librería Pandas [12]. A través de operaciones lógicas aplicadas al data frame, se intentó inferir la existencia de algún gen que estuviera exclusivamente relacionado con la Disgrafía, sin tener asociación con otro término HPO.

- **Genes asociados**

El segundo archivo consiste en un listado de genes asociados a la disgrafía. En la tabla 2, se presenta la cabecera del archivo, donde también se visualizan tres columnas: la primera contiene el identificador de Entrez de los genes, la segunda el símbolo de los genes y la tercera el identificador de las enfermedades. Al igual que en el archivo anterior, el identificador de las enfermedades puede provenir de dos fuentes, OMIM o Orphanet.

Table 2 Cabecera del archivo de genes asociados

Gene id (entrez)	Gene symbol	DISEASE_IDS
10347	ABCA7	ORPHA:1020,OMIM:608907
351	APP	OMIM:605714,ORPHA:100006,ORPHA:1020,ORPHA:3247...
9031	BAZ1B	ORPHA:904
9275	BCL7B	ORPHA:904
657	BMPR1A	OMIM:174900,ORPHA:329971,OMIM:610069,ORPHA:157...

- **Red proteínas**

Este archivo contiene una red de interacciones entre proteínas humanas, presentes en la base de datos STRING, junto con una puntuación de los enlaces entre las proteínas. La cabecera de este archivo se puede visualizar en la tabla 3.

Table 3 Datos de la red de proteínas con puntuaciones combinadas.

Proteína 1	Proteína 2	Puntuación Combinada
9606.ENSP00000000233	9606.ENSP00000356607	173
9606.ENSP00000000233	9606.ENSP00000427567	154
9606.ENSP00000000233	9606.ENSP00000253413	151
9606.ENSP00000000233	9606.ENSP00000493357	471

Cada fila contiene el StringID de las dos proteínas que interactúan, junto con el *combined score*. Esta puntuación se calcula al combinar las probabilidades de diferentes canales de evidencia y se corrige por la probabilidad de observar una interacción al azar.

Este archivo es demasiado grande para subirlo al repositorio, por lo que se ha filtrado por el *combined score* utilizando expresiones regulares:

```
awk -F" " '$3 > 800 {print $0}'
9606.protein.links.v12.0.txt > proteinas_filtrado.txt
```

De esta forma, nos quedamos con aquellas interacciones que tengan una puntuación mayor a 800.

2.2 Grafo bipartito

Al no encontrarse ningún gen que afecte solo a Disgrafía, se buscó aquellos términos HPO relacionados [...]

En un grafo bipartito, los vértices se organizan en dos conjuntos distintos, de modo que cada arista conecta un vértice de un conjunto con otro del segundo conjunto. En términos más simples, no existen aristas que conecten vértices dentro del mismo conjunto [14]. En nuestro contexto, los conjuntos de vértices representan genes y términos HPO. De esta manera, obtenemos un grafo bipartito que conecta distintos términos HPO al nuestro, a través de genes.

Para llevar a cabo esta representación y conexión entre genes y términos HPO, hemos utilizado la librería de Python NetworkX [15]. Usando las segunda y tercera columna de la tabla 1, es decir los símbolos de los genes y los identificadores de los términos HPO, se creó este grafo bipartito. [...]

De este grafo nos interesaba ver aquellos término HPO que se encuentran estrechamente relacionados con la Disgrafía, por lo que lo siguiente que hicimos fue hacer un subgrafo que con los nodos que se encuentre a dos pasos del término HPO Disgrafía y hacer una proyección de los términos HPO de ese subgrafo. De esta forma obtuvimos aquellos HPO que están relacionados con al Disgrafía a través de un gen. [...].

2.3 Red de genes

A continuación, se procedió a realizar un estudio de los genes relacionados con la disgrafía (usando el archivo descrito en 2.1). Lo primero fue obtener la red de genes utilizando la API de String-DB [13], haciendo uso de la biblioteca *strindb* para Python. Entre las funciones clave de esta biblioteca se encuentra get_network, la

cual requiere como parámetros la lista de nuestros genes y el identificador de la especie *Homo sapiens* (9606). Adicionalmente, se ha impuesto un score de 500, esta es una puntuación de corte para los bordes de la red, corresponde a la probabilidad de pertenecer a la misma vía funcional, lo que se traduce en que salgan más o menos genes en nuestra red.

La función devuelve la red de genes, donde se encuentran representados los genes y las relaciones entre ellos, la cual guardamos en un archivo .tsv.

2.4 Propagación de la red

Una vez que tenemos la red con los genes asociados, llevamos a cabo una propagación de la red con el objetivo de ampliar su tamaño y buscar otros genes relacionados con la Disgrafía. Este enfoque nos permite potencialmente descubrir otros genes implicados en la Disgrafía.

Para llevar a cabo este proceso, existen diferentes algoritmos, entre los cuales hemos seleccionado Diamond [16].

Este algoritmo toma como parámetros nuestra red de genes, la red de interacciones de proteínas (filtrada), el número de genes que queremos que tenga el grafo final, y el nombre del archivo donde se va a guardar la red ampliada.

```
python DIAMOnD.py proteinas_filtrado.txt grafo_51_genes.txt 200
propaged_genes.txt
```

2.5 Detección de comunidades

Para estudiar el grafo ampliado y la relación de los nuevos genes añadidos con el HPO de interés, se ha llevado a cabo un análisis de comunidades.

La detección de comunidades consiste en fragmentar el grafo en conjuntos de nodos, denominados "comunidades", teniendo en cuenta la topología de la red. En resumen, puede conceptualizarse como un proceso de agrupamiento (clustering) aplicado a grafos. Aunque no hay una definición única de comunidad aceptada por la comunidad científica, generalmente se considera que una comunidad es de calidad cuando muestra más conexiones internas que externas. En otras palabras, los nodos dentro de la comunidad están más densamente conectados entre sí en comparación con los nodos fuera de la comunidad en el grafo [17].

Para hacer la detecciónn de comunidades del grafo se ha usado uno de los algoritmos proporcionados por el paquete NetworkX, *greedy_modularity_communities*. Este algoritmo utiliza la maximización de modularidad avariciosa de Clauset-Newman-Moore [18] para encontrar la partición de comunidades con la mayor modularidad.

2.6 Enriquecimiento funcional

El análisis funcional de genes implica registrar y analizar estadísticamente listas de genes con el objetivo de identificar anotaciones funcionales en relación con los genes analizados. El propósito principal es determinar si existe una asociación significativa entre los genes y las funciones específicas que desempeñan en procesos biológicos, rutas metabólicas u otras categorías funcionales.

El enriquecimiento funcional se ha llevado a cabo utilizando el paquete String de Python, específicamente la función *get_enrichment*. Este análisis se ha realizado en varias ocasiones: una vez para el grafo ampliado y otra para cada una de las comunidades detectadas en la sección anterior.

3 Resultados

En esta sección se van a presentar los resultados obtenidos tras realizar el estudio del fenotipo Disgrafía.

3.1 Red de genes asociados

Después de descargar la red de genes asociados a la Disgrafía de la HPO, observamos que contenía 51 genes. Posteriormente, obtuvimos una red de genes relacionados con la Disgrafía utilizando la base de datos STRING-DB, la cual se puede observar en la imagen 2.

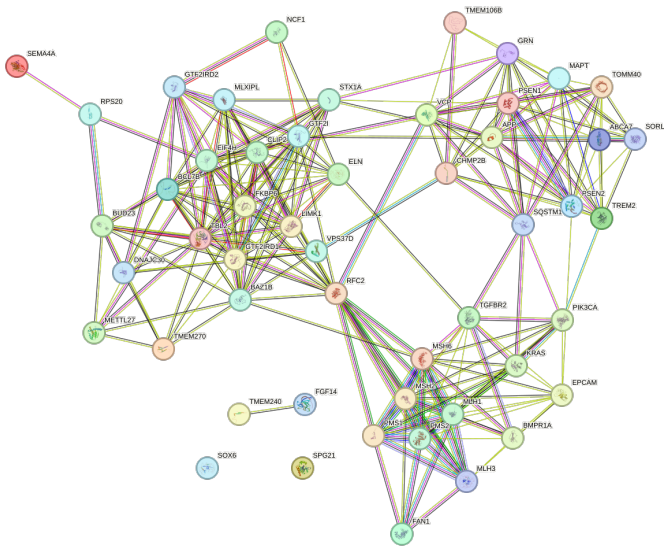

Figure 2 Red de genes asociados

4 Discusión

5 Conclusiones

Abreviaciones

SLD: dificultades específicas del aprendizaje HPO: Human Phenotype Ontology

Disponibilidad de datos y materiales

Enlace al repositorio de GithHub: https://github.com/nmartinser/HPO_Dysgraphia

Contribución de los autores

Usando las iniciales que habéis definido al comienzo del documento, debeis indicar la contribución al proyecto en el estilo:Debéis indicar aquí un enlace a vuestro repositorio de github. 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. Chung, P., R, D.P.: Dysgraphia. International Journal of Child and Adolescent Health **8**, 27–36 (2015)
2. Crouch, A.L., Jakubecy, J.J.: Dysgraphia: How It Affects A Studentâ•Žs Performance and What Can Be Done About It (2007). http://escholarship.bc.edu/education/tecplus/vol3/iss3/art5
3. del Sol, Y.S., Guerra, K.L., Medina, M.O.S., Gonzales-Sánchez, A., Oxolon, J.M.V.: Estudios sobre la corrección de la disgrafia caligráfica en escolares con discapacidad intelectual. Propósitos y Representaciones **9** (2021)
4. McCloskey, M., Rapp, B.: Developmental dysgraphia: An overview and framework for research. Cognitive Neuropsychology **34**, 65–82 (2017). doi:10.1080/02643294.2017.1369016

5. Drotár, P., Dobeš, M.: Dysgraphia detection through machine learning. Scientific Reports **10** (2020). doi:10.1038/s41598-020-78611-9
6. Nicolson, R.I., Fawcett, A.J.: Dyslexia, dysgraphia, procedural learning and the cerebellum. Cortex; a journal devoted to the study of the nervous system and behavior **47**, 117–27 (2011). doi:10.1016/j.cortex.2009.08.016
7. Abbott, R.D., Raskind, W.H., Matsushita, M., Price, N.D., Richards, T., Berninger, V.W.: Patterns of biomarkers for three phenotype profiles of persisting specific learning disabilities during middle childhood and early adolescence: A preliminary study. Biomarkers and genes **1** (2017)
8. Berninger, V., Richards, T.: Inter-relationships among behavioral markers, genes, brain and treatment in dyslexia and dysgraphia. Future neurology **5**, 597–617 (2010). doi:10.2217/fnl.10.22
9. Köhler, S., Gargano, M., Matentzoglu, N., Carmody, L.C., Lewis-Smith, D., Vasilevsky, N.A., Danis, D., Balagura, G., Baynam, G., Brower, A.M., Callahan, T.J., Chute, C.G., Est, J.L., Galer, P.D., Ganesan, S., Griese, M., Haimel, M., Pazmandi, J., Hanauer, M., Harris, N.L., Hartnett, M., Hastreiter, M., Hauck, F., He, Y., Jeske, T., Kearney, H., Kindle, G., Klein, C., Knoflach, K., Krause, R., Lagorce, D., McMurry, J.A., Miller, J.A., Munoz-Torres, M., Peters, R.L., Rapp, C.K., Rath, A.M., Rind, S.A., Rosenberg, A., Segal, M.M., Seidel, M.G., Smedley, D., Talmy, T., Thomas, Y., Wiafe, S.A., Xian, J., Yüksel, Z., Helbig, I., Mungall, C.J., Haendel, M.A., Robinson, P.N.: The human phenotype ontology in 2021. Nucleic Acids Research **49**, 1207–1217 (2021). doi:10.1093/nar/gkaa1043
10. Sundal, C., Carmona, S., Yhr, M., Almström, O., Ljungberg, M., Hardy, J., Hedberg-Oldfors, C., Åsa Fred, Brás, J., Oldfors, A., Andersen, O., Guerreiro, R.: An aars variant as the likely cause of swedish type hereditary diffuse leukoencephalopathy with spheroids. Acta neuropathologica communications **7**, 188 (2019). doi:10.1186/s40478-019-0843-y
11. Weinreich, S.S., Mangon, R., Sikkens, J.J., en Teeuw, M.E., Cornel, M.C.: [orphanet: a european database for rare diseases]. Nederlands tijdschrift voor geneeskunde **152**, 518–9 (2008)
12. McKinney, W.: pandas: powerful Python data analysis toolkit (2012)
13. Szklarczyk, D., Gable, A.L., Nastou, K.C., Lyon, D., Kirsch, R., Pyysalo, S., Doncheva, N.T., Legeay, M., Fang, T., Bork, P., Jensen, L.J., von Mering, C.: The string database in 2021: Customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. Nucleic Acids Research **49**, 605–612 (2021). doi:10.1093/nar/gkaa1074
14. He, X., Gao, M., Kan, M.-Y., Wang, D.: Birank: Towards ranking on bipartite graphs. IEEE Transactions on Knowledge and Data Engineering **29**, 57–71 (2017). doi:10.1109/TKDE.2016.2611584
15. Platt, E.L.: Network Science with Python and NetworkX Quick Start Guide: Explore and Visualize Network Data Effectively vol. 190 páginas. Packt Publishing Ltd., ??? (2019)
16. Ghiassian, S.D., Menche, J., Barabási, A.L.: A disease module detection (diamond) algorithm derived from a systematic analysis of connectivity patterns of disease proteins in the human interactome. PLoS Computational Biology **11** (2015). doi:10.1371/journal.pcbi.1004120
17. Lledot, A.P.: Algoritmos bio-inspirados para la detección de comunidades dinámicas en redes complejas
18. Clauset, A., Newman, M.E.J., Moore, C.: Finding community structure in very large networks. Phys. Rev. E **70** (2004)