

RESEARCH

Aniridia y su relación con otras patologías

Alberto Montero Solera^{??}
, Teresa Vega Martínez
, Alexandra Rodríguez Serantes
and Xian Li Xia

Abstract

Keywords: aniridia; patologías; iris; ocular

1 Introducción

La aniridia es un trastorno congénito del desarrollo ocular caracterizado por la ausencia parcial o total del iris, la estructura que da color al ojo y que regula la cantidad de luz que entra en la pupila. Esta condición, además de afectar la estética del ojo, puede causar una serie de problemas visuales significativos, como fotofobia (sensibilidad a la luz), nistagmo (movimientos involuntarios de los ojos) y disminución de la agudeza visual. Los individuos con aniridia a menudo presentan malformaciones oculares asociadas, como cataratas, glaucoma (enfermedad ocular que daña el nervio óptico, esencial para la visión) y problemas en la córnea, lo que agrava su salud visual y afecta gravemente su calidad de vida [?].

No obstante, la aniridia no es solo un trastorno ocular, sino que también está vinculada a diversas patologías sistémicas, lo que sugiere que sus mecanismos genéticos y funcionales son complejos y multifactoriales[?]. Por ejemplo, estudios han mostrado que esta enfermedad puede estar asociada con el síndrome de WAGR, una condición genética que incluye anomalías renales, tumores y problemas de desarrollo, además de la aniridia [?]. Otros síndromes relacionados incluyen el síndrome de Gillespie, caracterizado por discapacidad intelectual, y el síndrome de Axenfeld-Rieger, que afecta tanto los ojos como otros órganos [?]. Estas asociaciones con condiciones sistémicas destacan la importancia de investigar en profundidad los mecanismos genéticos que subyacen a la aniridia, así como su relación con otras patologías.

Uno de los avances más significativos en la investigación de la aniridia ha sido la identificación del gen **PAX6** como regulador clave del desarrollo ocular. PAX6 es un factor de transcripción que juega un papel crucial en la diferenciación de diversas estructuras oculares, regulando la expresión de otros genes esenciales para este proceso [?]. Las mutaciones en PAX6 han sido vinculadas no solo con la aniridia, sino también con otras anomalías oculares como la displasia corneal (alteración en el desarrollo o estructura de la córnea) y el glaucoma congénito [?]. Sin embargo, para entender completamente la heterogeneidad clínica de la aniridia, es fundamental analizar cómo interactúa PAX6 con otros genes en redes genéticas más amplias.

Además de **PAX6**, varios genes han demostrado ser importantes para entender la complejidad de la aniridia y sus trastornos asociados, entre ellos **FOXC1**, **WT1**, **FOXE3** y **PITX2**.

Las mutaciones en **FOXC1** [?] están asociadas con el síndrome de Axenfeld-Rieger, un trastorno que incluye glaucoma y otras anomalías oculares [?]. Este gen codifica un factor de transcripción que, al igual que **PAX6**, regula el desarrollo del ojo, y su interacción con otros genes es clave para comprender las manifestaciones clínicas variadas que se presentan en estos pacientes. La integración de **FOXC1** en los análisis de redes génicas permite explorar cómo las vías de desarrollo ocular pueden estar interconectadas con otras rutas reguladoras sistémica.

Otro gen relevante es **WT1**, conocido por su papel en el desarrollo renal y la formación de tumores. Las mutaciones en **WT1** están vinculadas al síndrome de WAGR. La investigación sobre **WT1** desde una perspectiva de biología de sistemas permite investigar cómo los defectos en los programas de desarrollo que afectan tanto a los ojos como a otros órganos pueden estar mediadas por la interacción de redes génicas compartidas [?].

FOXE3 es un gen sin intrones que pertenece a la familia de factores de transcripción "Forkhead", que regulan la expresión de otros genes específicos de tejido cruciales para la diferenciación y desarrollo. La proteína que se codifica desempeña un importante papel en la formación del cristalino en vertebrados, siendo un factor de transcripción específico. Además, las mutaciones de este gen están relacionadas con disgenesia mesenquimal (desarrollo incompleto del tejido mesenquimático) del segmento anterior del ojo y afaquia primaria congénita (ausencia del cristalino).

Finalmente, el gen **PITX2** está involucrado en la morfogénesis craneofacial y cardiovascular. Las mutaciones en este gen están asociadas con el síndrome de Axenfeld-Rieger, afectando no solo a los ojos, sino también al desarrollo de órganos internos, debido a su rol en la simetría del desarrollo[?]. Desde la perspectiva de la biología de sistemas, **PITX2** representa un nodo crítico dentro de redes regulatorias que conectan el desarrollo de múltiples órganos, lo que resalta cómo las disfunciones en un solo gen pueden dar lugar a manifestaciones clínicas en varios sistemas.

El objetivo de este trabajo es examinar las patologías asociadas al fenotipo de la aniridia desde el punto de vista de la biología de sistemas, enfocándonos en los genes que tienen una relación funcional con **PAX6**. A través de la identificación de genes asociados y el análisis de redes génicas y moleculares mediante técnicas como la creación de clusters[?] y el análisis de redes funcionales [?], se tratará de avanzar en la comprensión de cómo las interacciones entre estos genes contribuyen a la presentación clínica de la aniridia y sus trastornos relacionados. Este enfoque no solo amplía nuestro conocimiento sobre la aniridia, sino que también ofrece información valiosa sobre los mecanismos patológicos que interconectan diversas condiciones sistémicas.

2 Materiales y métodos

En esta sección serán expuestos las diferentes herramientas y métodos utilizados para llevar a cabo el proyecto.

2.1 Materiales

2.1.1 *Human Phenotype Ontology*

La **Human Phenotype Ontology (HPO)** es una base de datos estructurada que organiza y describe las características clínicas (fenotipos) observadas en humanos, utilizando términos estandarizados y jerárquicos [?].

2.1.2 *STRING*

La **STRING Database** es una base de datos que integra información sobre interacciones entre proteínas, incluyendo interacciones físicas y funcionales, en una amplia variedad de organismos [?].

2.1.3 *Python*

Python es un lenguaje de programación versátil, fácil de aprender y ampliamente usado en desarrollo web, ciencia de datos, IA y más, gracias a su sintaxis clara y su gran ecosistema de herramientas [?].

2.2 Metodología

2.2.1 *Obtención de genes*

El primer paso que se llevó a cabo para comenzar el proyecto es la obtención de los genes relacionados con el fenotipo seleccionado. Para ello se visitó directamente la web de HPO (<https://hpo.jax.org/>) y se introdujo el identificador del fenotipo, HP:0000526 en este caso. Como resultado de la búsqueda se obtuvo el conjunto de genes relacionados con la aniridia. Finalmente descargamos el conjunto de genes resultante de HPO.

2.2.2 *Obtención de la red de interacciones inicial*

Tras descargar los genes, se utilizó la página web de STRING para obtener su red de interacciones (<https://string-db.org/>). Para ello es necesario seleccionar la opción para introducir múltiples proteínas y especificar la especie, en este caso *Homo Sapiens*. Se introdujo la lista de genes obtenida de HPO. Finalmente, dándole a exportar aparecen varias opciones. En el caso de este estudio se descargó la imagen en formato PNG de la red de interacciones. Además se descargaron los datos de la red como salida de texto en formato tsv.

2.2.3 *Análisis de la red de interacciones*

En este punto del proyecto se llevó a cabo un análisis exhaustivo de la red de interacciones de todos los genes implicados en el fenotipo. El análisis de dicha red implica examinar cómo los genes están conectados entre sí para identificar patrones, relaciones funcionales y estructuras que revelen información clave.

Se calcularon métricas clave como el grado de los nodos, la centralidad y la modularidad, con el objetivo de identificar los nodos más importantes dentro de la red. A través del análisis estructural y funcional, se detectaron los nodos clave, aquellos que desempeñan un papel central en la red.

El resto del proyecto gira en torno a los genes clave seleccionados tras el análisis de la red.

2.2.4 Propagación de red

Para analizar de manera exhaustiva los genes funcionalmente relevantes asociados con nuestro fenotipo de interés, aplicamos el algoritmo **DIAMOnD (Disease Module Detection)** a una red de interacción proteína-proteína (PPI). Este procedimiento nos permitió expandir un módulo génico inicial definido por el conjunto de genes semilla previamente seleccionado mediante la integración de información de interacción genética y funcional. Los genes semilla iniciales, seleccionados por su relevancia biológica y su asociación conocida con el fenotipo en estudio, fueron: **WNT10B, SEM1, WT1, PAX6, y NF1**. Estos genes representan el núcleo inicial desde el cual se propagaron las interacciones para identificar genes adicionales funcionalmente relacionados. Respecto a la red base, utilizamos un archivo que contiene todas las interacciones proteína-proteína humanas disponibles en la base de datos STRING. El algoritmo DIAMOnD se utilizó para expandir el módulo génico inicial. En este método, en cada iteración, el algoritmo evalúa los nodos vecinos de los genes ya presentes en el módulo y asigna un puntaje basado en el número de conexiones con los genes actuales del módulo. El nodo con el puntaje más alto se añade al módulo. Este proceso se repite hasta alcanzar el número máximo de nodos adicionales (en este caso, 200 nodos) o hasta que no queden candidatos relevantes. Una vez completada la expansión, se construyó un subgrafo que incluye todos los genes del módulo expandido junto con las aristas que los conectan.

2.2.5 Clustering

Para continuar con el análisis se desarrolló un algoritmo de clustering con el objetivo de identificar agrupaciones en la red de genes obtenida en el apartado anterior.

El **clustering** consiste en dividir un conjunto de datos en grupos o comunidades, de forma que los nodos de un mismo grupo tengan mayor relación entre sí que con los de los demás grupos.

Aplicado en el campo de la biología de sistemas, este método permite analizar redes complejas como las redes de interacción proteína-proteínas, rutas metabólicas y redes de regulación génica. A menudo, las comunidades identificadas con este tipo de clustering se corresponden a grupos de proteínas, genes (como es en este caso) o componentes moleculares que trabajan juntos para realizar una función específica. Para este proyecto se emplea un código para realizar el clustering de los genes de la red usando la **biblioteca igraph**. En este lo primero que se hace es cargar los datos de las interacciones entre genes desde el archivo de la red obtenida previamente en formato tabular.

A continuación, se crea un grafo no dirigido con los datos previos. Se estudian algunas propiedades de la red.

Después, se calculan las **métricas de centralidad**, es decir, el **grado de centralidad** (mide el número de conexiones de cada nodo), la **centralidad de cercanía** (mide la distancia media entre un nodo y los demás) y la **conectividad** (mide la densidad de conexiones en el grafo).

Una vez obtenidas las medidas se aplican varios métodos de clustering. Estos son **Edge Betweenness Clustering**, basado en la intermediación de aristas e ir eliminando iterativamente las aristas con mayor betweenness; **Clustering Basado en el Algoritmo Fast Greedy**, este consiste en agrupar nodos que tienen más conexiones entre ellos que con otros, optimizando de manera eficiente la modularidad; y

Clustering Basado en el Algoritmo Louvain, el cual encuentra comunidades en redes grandes optimizando su estructura jerárquica para maximizar la modularidad. Además, se realiza la identificación de comunidades con **Link Communities**, método que considera las relaciones entre aristas, es decir, entre nodos.

Estos resultados se guardan como imágenes en la carpeta del proyecto "results" y se añadirán en el apartado de resultados de este informe.

Por último, se buscan vecinos interesantes de los genes de interés en el estudio (**WNT10B**, **WT1**, **SEM1**, **PAX6**, **NF1**). Estos se guardan en un archivo plano que se empleará para el enriquecimiento funcional.

2.2.6 Enriquecimiento funcional de los clusters

Para profundizar en la caracterización funcional del cluster seleccionado en el apartado anterior, hemos llevado a cabo un análisis de enriquecimiento funcional utilizando un enfoque basado en la base de datos STRINGdb. En este análisis, evaluamos específicamente dos categorías: procesos biológicos (Process) y ontología del fenotipo humano (HPO, Human Phenotype Ontology). Además de poder evaluar los procesos biológicos en los que están envueltos estos genes, otro de los objetivos era determinar si entre los fenotipos resultantes del análisis de enriquecimiento se encuentran aquellos descritos en HPO como relacionados con el fenotipo de interés. Durante el proceso se ha llevado a cabo la conversión a identificadores UniProt de los genes del cluster mediante el uso de la API de MyGene.info, asegurando compatibilidad con la API de STRINGdb.

Posteriormente, dichos identificadores UniProt fueron sometidos al análisis de enriquecimiento utilizando la base de datos STRINGdb, con un enfoque en las categorías Process y HPO.

En este proceso nosotros consideramos oportuno establecer un umbral de significancia estadística de **FDR < 0.001**, para identificar términos significativamente enriquecidos. Se evaluaron las asociaciones entre los genes del cluster y términos fenotípicos relevantes descritos en HPO.

Como resultado de la ejecución del algoritmo desarrollado se obtienen dos archivos en formato csv. Uno recoge los resultados para la categoría HPO y los de la categoría Process.

3 Resultados

3.1 Red de interacciones inicial

4 Discusión

5 Conclusiones

Abreviaciones

Indicar lista de abreviaciones mostrando cada acrónimo a que corresponde

Disponibilidad de datos y materiales

https://github.com/monteero13/project_template.git

Contribución de los autores

Usando las iniciales que habéis definido al comienzo del documento, debeis 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.

