

## Proyecto de investigación

### Título

Desarrollo de algoritmos bioinformáticos basados en redes y Biología de Sistemas para la búsqueda de nuevos genes responsables de enfermedades raras de origen genético

### Descripción

La introducción de las técnicas de secuenciación masiva ha producido un aumento del porcentaje de casos diagnosticados en enfermedades hereditarias. En enfermedades raras, todavía estamos lejos de alcanzar un diagnóstico completo, principalmente porque no se conocen todos los mecanismos (genes y variantes) implicados. En la actualidad el descubrimiento de nuevos genes causantes o implicados (modificadores genéticos) en una patología se realiza a partir de sospechas experimentales en, normalmente, pocos casos que son recopilados por diversos grupos de investigación de todo el mundo.

Nuestro propósito es desarrollar métodos basados en la integración de datos, la Biología de Redes y de Sistemas y las capacidades predictivas de métodos estadísticos diseñados para el big data, para predecir probabilísticamente la relación entre genes y enfermedades. Proponemos por tanto un algoritmo capaz de asignar un valor predictivo a cada par gen/enfermedad que pueda ser utilizado en el mismo sentido que se hace con los muy extendidos predictores de patogenicidad en la priorización de variantes. Mediante este proyecto proponemos profundizar en el conocimiento de las bases moleculares de las enfermedades genéticas raras con especial foco en las que afectan a la retina aunque extensible a otras patologías (con cohortes de pacientes e incluso a otras). Nuestras herramientas y desarrollos serán además puestos a disposición de la práctica clínica contribuyendo de forma muy significativa a mejorar el diagnóstico y por tanto la calidad de vida de los pacientes afectados de estas enfermedades.

**OBJETIVO:** Desarrollar un método bioinformático para predecir nuevas asociaciones gen-enfermedad en enfermedades genéticas mendelianas.

**HIPÓTESIS:** Los genes causales de enfermedades genéticas están relacionados funcionalmente. A su vez, los genes con funciones cercanas se encuentran próximos en un interactoma que represente asociaciones funcionales. Una búsqueda exhaustiva en la vecindad de genes ligados a un mismo fenotipo puede predecir nuevos genes susceptibles de tener variantes que causen las mismas disfunciones.

**MÉTODOS:** Hemos recopilado ~40 fuentes de información que se han agrupado en 13 categorías no solapantes que incluyen desde información fenotípica, de expresión, de

interacciones físicas, de co-citaciones en la literatura científica, y otras. Cada una de estas fuentes de información es transformada en una red funcional (interactoma) que representa un tipo de conocimiento. Nuestro método calcula la capacidad de cada categoría de conocimiento para predecir nuevos genes en cada patología de forma específica para después priorizar genes provenientes de cada uno de los interactomas.

**RESULTADOS Y CONCLUSIONES:** Hemos realizado una validación de nuestro método realizando predicciones de genes ya conocidos en 82 patologías con genes asociados descritos en paneles virtuales por Genomics England (<https://panelapp.genomicsengland.co.uk>).

Nuestros resultados confirman que las diferentes patologías necesitan priorizar distintos tipos de información para aumentar la capacidad predictiva. Por último, también concluimos que la capacidad predictiva de las categorías de información de relaciones funcionales también varía en función del nivel de conocimiento que existe para cada enfermedad.

**Periodo:** 01/01/2019 - 31/12/2021

**Financiación:** Instituto de Salud Carlos III

**Referencia:** PI18/00579

**Investigadores principales:** Pablo Mínguez Paniagua