

RESEARCH

Estudio fenotípico del gen HP:0030880 asociado al fenómeno de Raynaud

Jesús Aldana^{*} and Juan Carlos Ruiz

^{*}Correspondence:

jaldanam21@uma.es

ETSI Informática, Universidad de Málaga, Málaga, España

Full list of author information is available at the end of the article

Abstract

Esta memoria de investigación se realizó siguiendo la estructura convencional de una revista científica, con el objetivo de divulgar una enfermedad rara que afecta a numerosas personas en nuestro país. Hicimos un estudio acerca de aquellos genes a los que afecta y el fenotipo que producen. Toda la información investigada se organizó en forma de grafo y redes biológicas para su posterior análisis.

This research report was carried out following the conventional structure of a scientific journal, with the aim of disseminating a rare disease that affects many people in our country. We did a study about the genes it affects and the phenotype they produce. All the information investigated was organized in the form of a graph and biological networks for further analysis.

Keywords: Raynaud; gene; phenotype; ontology

1 Introducción

1.1 Ontologías

Para poder estudiar los sistemas influidos por el síndrome de Raynaud será necesario utilizar diferentes ontologías y bases de datos que podemos encontrar en Internet. Para ello en primer lugar debemos conocer qué es una ontología. Se podría definir como una manera formal de representar el conocimiento colectivo sobre un área en la que sus conceptos se describen por su significado y su relación entre el resto de componentes [1][2].

En nuestro caso, usaremos la ontología de GO (*Gene Ontology* 5) y la ontología de HPO (*Human Phenotype Ontology database* 5) [3]. Dicha ontología se encarga de mostrar las relaciones entre genes de diversas especies y HPO almacena ontologías asociadas a fenotipos de enfermedades del ser humano.

1.2 Fenómeno Raynaud

En particular el término HP:0030880 hace referencia en la ontología de HPO al fenómeno de Raynaud, esta enfermedad afecta al flujo de sangre en vasos arteriales cuando baja la temperatura corporal, esto suele afectar en su mayoría a las extremidades superiores e inferiores.

Más concretamente, el fenómeno de Raynaud causa una vasoconstricción de los vasos sanguíneos en situaciones de bajas temperaturas o estrés. Cuando esto ocurre, la sangre no puede llegar a la superficie de la piel y las áreas afectadas se vuelven blancas y azules. Cuando el flujo sanguíneo regresa, la piel se enrojece y se tiene una sensación de palpitación o de hormigueo. En casos severos, la pérdida del flujo sanguíneo puede causar llagas o muerte de los tejidos.[4]

1.2.1 Tipos de Raynaud

Es importante conocer que el fenómeno se divide en dos tipos de expresión en función del origen del mismo.

- **Raynaud primario o enfermedad de Raynaud:** Es la forma más frecuente y a su vez la que provoca síntomas más leves, típicamente afectan a mujeres menores de 30 años (normalmente en la adolescencia) de momento no se conoce su origen, ya que el fenómeno de Raynaud primario ocurre en pacientes sin otra enfermedad reumática. [5]
- **Raynaud secundario:** Es menos común y se denomina así ya que es provocado por otra condición. Estas son diversas y entre las diferentes enfermedades reumáticas que lo provocan se incluyen, entre otras, la esclerodermia, el lupus o el síndrome de Sjögren. Además puede deberse también a largas exposiciones al frío o ciertas sustancias químicas. Los síntomas provocados por el Raynaud secundario suelen ser más graves que en el primario. A menudo el fenómeno se presenta en mujeres mayores de los 30 años. [6]

1.3 Niveles de asociación

Para buscar información de genes y enfermedades en las que está asociado el fenómeno de raynaud encontramos en internet bases de datos de uso abierto como OMIM y HPO. Buscando en HPO el gen encontramos como la enfermedad de Raynaud está a su vez relacionado con otras 41 enfermedades distintas y unos 21 genes.

1.3.1 Enfermedades asociadas

El fenómeno de Raynaud como ya hemos comentado es una enfermedad cardiovascular que afecta principalmente al flujo de la sangre, sin embargo como es común en las enfermedades genéticas los genes a los que afecta según el contexto pueden producir unos fenotipos u otros y en nuestro caso encontramos dos enfermedades con las que el fenómeno de Raynaud tiene especial relación: *Lupus eritematoso sistémico* [7] o LES es una enfermedad auto-inmunitaria con múltiples fenotipos que van desde una leve irritación cutánea hasta afectar a varios órganos o incluso sistema nervioso central. La *esclerodermia* [8] es la otra enfermedad común cuando hablamos del fenómeno de Raynaud. La esclerodermia afecta de forma multisistémica a todo el organismo causando encogimiento en la piel y fallos vasculares.

1.3.2 Genes asociados:

Entre los 21 genes asociados que encontramos en HPO vemos algunos como STING1 al que solo está asociado además de con Raynaud con una vasculopatía con inicio en el lactante o el gen RNF125 asociado a Síndrome de Tenorio, sin embargo tras buscar información y papers en pubmed nos hemos centrado en el gen **COL4A1**. [9]

Este gen afecta a un espectro muy amplio de trastornos entre los que destacamos: enfermedad cerebral de vasos pequeños, porencefalia, calambres musculares, aneurismas cerebrales, fenómeno de Raynaud; El COL4A1 se hereda de forma autosómica dominante y la mayoría de las personas estudiadas en todas ellas coincidían en tener un padre afectado. De esta forma cada hijo de un individuo con un trastorno relacionado con COL4A1 tiene un 50% de probabilidades de heredar la variante patógena.

1.3.3 Fisiopatología

Como ya hemos comentado en el apartado 1.2 el síndrome de Raynaud provoca una vasoconstricción de los vasos sanguíneos. En este proceso cobran mucha importancia las células endoteliales (ECs), estas células se encuentran ubicadas en el lado luminal de la pared del vaso sanguíneo y realizan funciones como la formación de vasos sanguíneos, la coagulación, la fibrinólisis, la regulación del tono muscular y hasta tiene un papel en la inflamación.

Aquí entran en juego los genes de la familia STAT (Signal Transducer and Activator of Transcription) esta familia de genes proporcionan instrucciones para producir proteínas que forman parte de las vías de señalización química esenciales dentro de las células. Concretamente el STAT4 está relacionado con el síndrome de Raynaud debido a la vasoconstricción de los vasos sanguíneos. Queremos destacar un estudio realizado por (??) en el cual se demuestra como el interleukin 12 que participa en la activación del gen STAT4 no funciona correctamente y demuestra como el **IFN- α** (Interferon alpha) lo activa en su lugar. Activando así a las células endoteliales y provocando una reacción no prevista en el organismo.

?? PAPER DE ECS ?? PAPER TOCHO de la info

1.4 Objetivos

En este proyecto vamos a investigar los mecanismos moleculares y rutas biológicas detrás de la enfermedad descrita por el término HP:0030880 y sus genes asociados, además de intentar buscar opciones de terapia útiles para el tratamiento de la enfermedad. Para ello usaremos, como ya hemos comentado, ontologías (GO) y bases de datos como OMIM (*Online Mendelian Inheritance in Man database* 5) o HPO.

2 Materiales y métodos

Se ha llevado a cabo un estudio fenotípico sobre la población afectada por el fenómeno de Raynaud. Se pretende averiguar la influencia de los diferentes genes asociados a la enfermedad y su grado de importancia en el desarrollo de la misma.

Para llevar a cabo la siguiente metodología hemos será necesario descargar la última versión de R, que, hasta la fecha de este proyecto es la 4.2.1. Así como será necesario instalar todas las librerías indicadas en el repositorio de github.

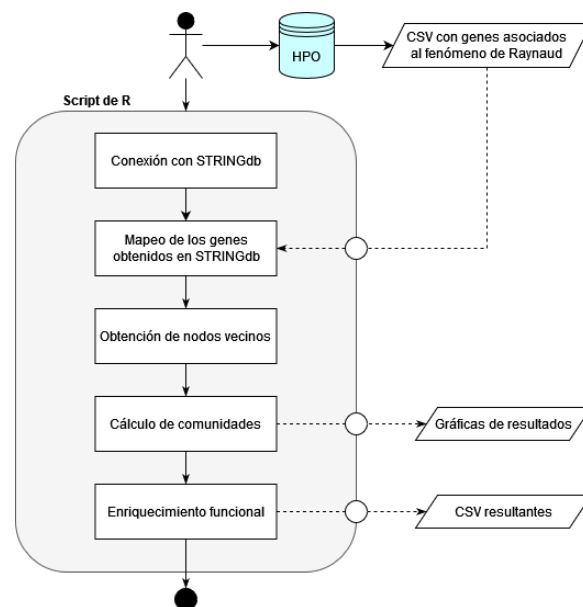


Figure 1: Flujo de trabajo

2.1 Obtención de la red

STRINGdb es una base de datos open-source que busca integrar todas las relaciones conocidas entre proteínas, así como sus interacciones físicas o funcionales. Esta herramienta también nos permite visualizar las redes que forman un conjunto de proteínas. Para la obtención automática de los datos es necesario utilizar la API de STRINGdb con R. Al tratarse de un estudio sobre una enfermedad que afecta a humanos, es necesario descargar de 'string' todas las proteínas, interacciones y peso de el enlace; del genoma humano al completo. El ID taxonómico con el que realizar la búsqueda es el **9606**. Como ya se ha comentado en el apartado 1.3.2 y como podemos ver en 1 es necesario descargar de HPO el nombre de los genes asociados, su rentrez.ID (que será útil para el mapeo de nuestros genes en el genoma) y el identificador del gen de la base de datos de OMIM; Todos estos datos relacionados con el fenómeno de Raynaud. Finalmente mediante haciendo uso de la API de STRINGdb realizamos un mapeo de los identificadores de rentrez de nuestros genes con la red completa del genoma humano de STRINGdb. Obteniendo así los genes con los que realizaremos el estudio. [10], [11]

2.2 Cálculo de comunidades

Tras conseguir los datos ahora debemos de realizar un estudio del grafo obtenido a partir de los datos del apartado 2.1. Para ellos realizamos un subgrafo a partir del mapeo de genes del apartado anterior para guardar nuestros datos en un objeto de tipo grafo. Como ya sabemos un grafo es un conjunto de nodos unidos por enlaces llamados aristas que nos permite representar las relaciones entre elementos de un conjunto. Haciendo uso de la teoría de grafos sabiendo que la vecindad de un vértice en un grafo G es el subgrafo inducido de G que está formado por todos los vértices adyacentes y todas las aristas que conecten dichos nodos. En otras palabras aplicado

a nuestros datos, vamos a obtener aquellos nodos que presenten un mayor número de relaciones para dividir todos nuestros genes en diferentes clusters. Cada cluster obtenido es una comunidad diferente, las ventajas que nos proporciona esto es el poder estudiar cada una por separado y realizar análisis de expresión solo a aquellos genes coexpresados. [12], [13]

2.3 Enriquecimiento funcional

El análisis de enriquecimiento de conjuntos (*Gene Sequence Enrichment Analysis* 5) es un método para identificar clases de genes o proteínas que están sobrerrepresentadas o empobrecidas en una comunidad de genes y pueden tener asociación con fenotipos de enfermedades. Los métodos de GSEA se basan en métodos estadísticos y en las últimas tecnologías de transcriptómica y proteómica.

2.3.1 GO

La base de datos de GO 5. Al realizar un enriquecimiento mediante GO conseguimos en formato tabla los términos GO compartidos significativos con los genes a los que aplicamos el enriquecimiento, así como su valor de sobrerrepresentación / subrepresentación y su valor p asociado. El valor p es la probabilidad de ver al menos x número de genes del total de genes en la lista anotados a un término GO en particular, dada la proporción de genes en todo el genoma.

2.3.2 KEGG

La base de datos de KEGG (*Kyoto Encyclopedia of Genes and Genomes* 5). KEGG es una base de datos que integra información genómica, química y de funciones del sistema. El núcleo de estos son las bases de datos KEGG pathway y KEGG orthology. En la base de datos KEGG pathway, las vías metabólicas biológicas se dividen en 6 categorías: procesos celulares, procesamiento de información ambiental, procesamiento de información genética, enfermedades humanas, metabolismo y sistemas orgánicos.[14].

Tras aplicar sendos enriquecimientos para cada comunidad que obtenemos en nuestro script de R. Obtenemos datos tabulador que nos serán útiles para realizar un estudio de funcionalidad de cada comunidad y sacar las conclusiones pertinentes.

3 Resultados

Todo lo comentado en esta sección se ha realizado haciendo uso de la librería linkcomm de R [15] y la ayuda de STRINGdb [10] . Tras la obtención de la red de genes al realizar el mapeo con STRINGdb hemos obtenido un grafo como el que podemos observar en la figura 2. En este grafo podemos observar todos los genes que pertenecen a la red asociada al fenómeno de Raynaud. Esto se ha obtenido a partir de STRINGdb como hemos explicado en el apartado 2.1.

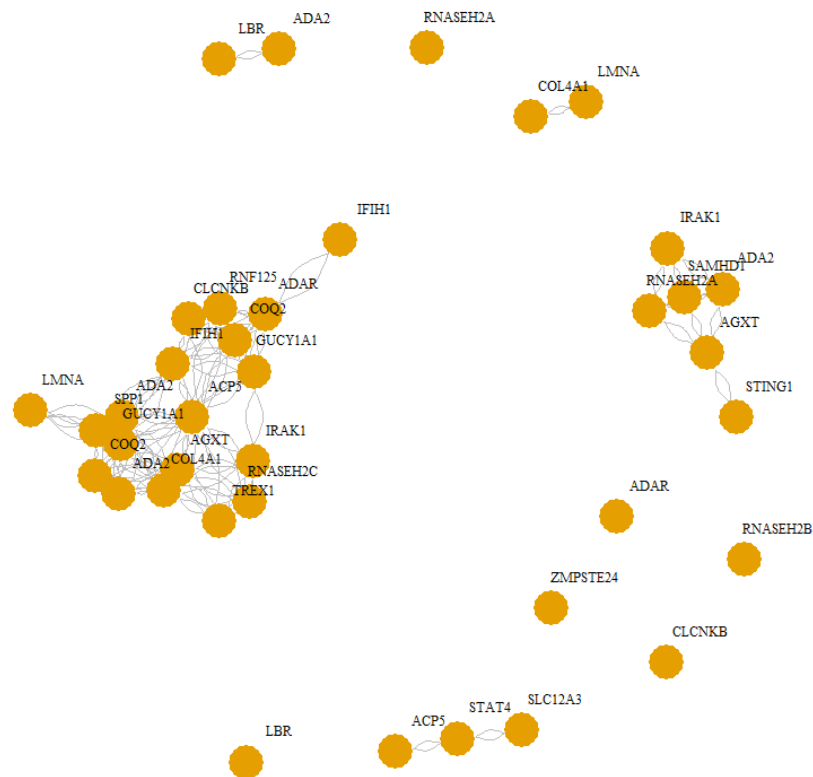


Figure 2: Grafo obtenido al realizar el mapeo de genes

Seguidamente hemos obtenido las comunidades que conforman nuestra red y a las que posteriormente se le harán los estudios de funcionalidad y podremos comprobar que genes conforman el fenotipo más grave de la enfermedad. Se han obtenido 3 gráficas en las que mostramos información acerca de estas comunidades.

En la figura 3, la primera obtenida del flujo de trabajo, se observa un dendrograma en la que, diferenciadas por colores, se pueden visualizar las diferentes comunidades obtenidas. Estas se han obtenido a una altura bastante alta (alrededor del 0.9) y nos indica que tenemos 6 comunidades cuyo agrupamiento más grande tiene 11 genes.

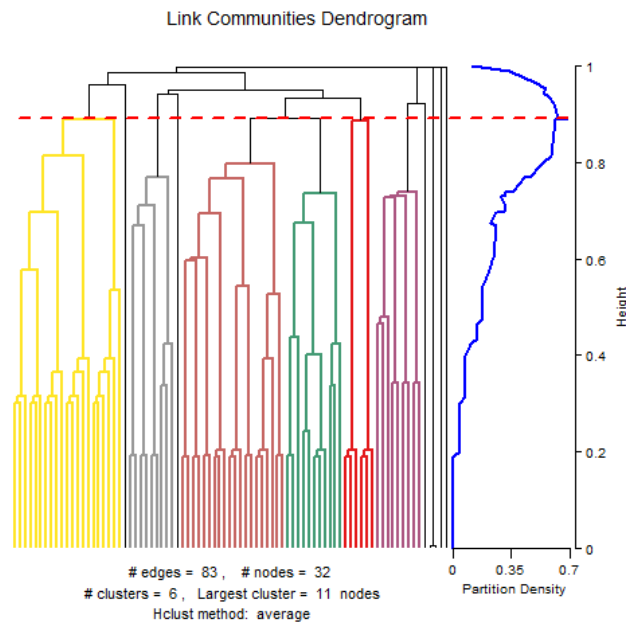


Figure 3: Dendrograma de la red mostrando las comunidades

La segunda figura que obtenemos del flujo de trabajo sería la número 4, en esta observamos una matriz de los miembros de cada comunidad. En este caso solo podemos ver los 10 primeros genes de nuestra red y en la matriz se nos muestra las comunidades a las que pertenece cada gen (señalado con un cuadrado en color para cada comunidad a la que pertenece). Además en los márgenes derecho e inferior de la matriz observamos los sumatorios del número de total de comunidades a las que pertenece cada gen y el número de genes que contiene cada comunidad, respectivamente.

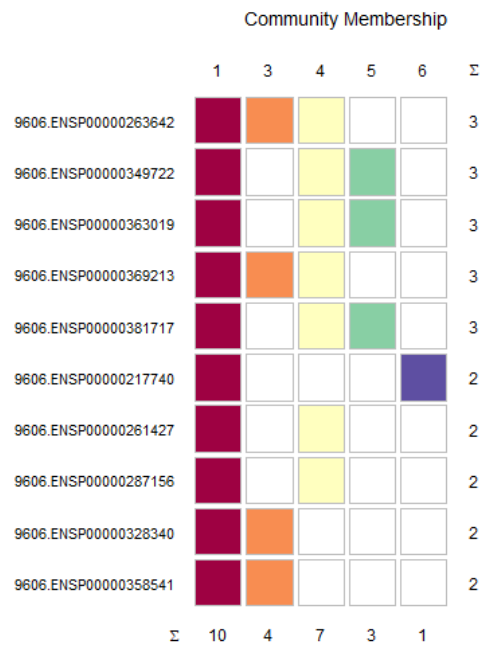


Figure 4: Matriz de los primeros genes para cada comunidad

Por último, al obtener las comunidades también hemos mostrado la figura 5 en la que se observa un grafo similar al de la figura 2, sin embargo, en este tenemos diferenciadas por colores las diferentes comunidades que hemos obtenido tras realizar el flujo de trabajo.

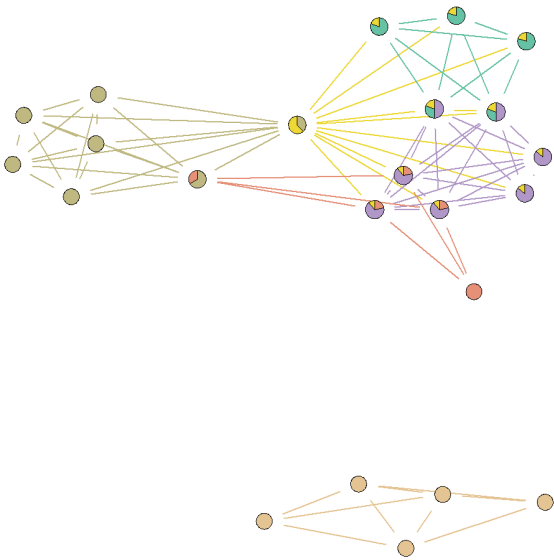


Figure 5: Grafo de los genes divididos por comunidad

Por último, hemos obtenido los CSV de cada una de las comunidades tras aplicarles el enriquecimiento. No todas las comunidades han devuelto una tabla con datos y los datos que sí hemos obtenido podemos observarlos en las tablas 1 y 2, correspondiendo a las comunidades 1 y 2, respectivamente.

En estas tablas podemos observar las columnas de *Ontología*, que nos indica la ontología de la que se ha obtenido esa referencia; *Término*, que indica el código de la ontología que relaciona a dicho gen; *N genes* y *Genes totales*, que indican el número de genes encontrados y el número de genes que tiene asociado dicho término; en la columna *Genes* vemos los nombres de los genes encontrados; y por último tenemos dos columnas para el *p valor* y el *fdr*, además de una descripción del término de la ontología.

Ontología	Categoría	Término	N genes	Genes totales	Taxon Id	genes	p valor	fdr	descripción
GO	Process	GO.0032480	2	43	9606	RNF125,IFIH1	5.17e-06	0.0011	negative regulation of type I interferon production
GO	Process	GO.0045088	2	361	9606	RNF125,IFIH1	0.00034	0.018	regulation of innate immune response
GO	Process	GO.0043900	2	653	9606	RNF125,IFIH1	0.0011	0.0347	regulation of multi-organism process
GO	Process	GO.0032446	2	690	9606	RNF125,IFIH1	0.0012	0.0347	protein modification by small protein conjugation

Table 1: Resultado del enriquecimiento de la comunidad 1

4 Discusión

5 Conclusiones

- Abreviaciones
- **GO**: Gene Ontology
 - **HPO**: Human Phenotype Ontology database
 - **OMIM**: Online Mendelian Inheritance in Man database

Ontología	Categoría	Término	N genes	Genes totales	Taxon Id	genes	p valor	fdi	descripción
GO	Process	GO.0090305	4	287	9606	RNA5EH2A,SAMHD1,TREX1,RNA5EH2B	2.37e-07	4.26e-05	nucleic acid phosphodiester bond hydrolysis
GO	Process	GO.0034655	4	394	9606	RNA5EH2A,SAMHD1,RNA5EH2C,RNA5EH2B	8.29e-07	7.46e-05	nucleobase-containing compound catabolic process
GO	Process	GO.0090501	3	137	9606	RNA5EH2A,SAMHD1,RNA5EH2B	3.35e-06	9.12e-05	RNA phosphodiester bond hydrolysis
GO	Process	GO.0006401	3	237	9606	RNA5EH2A,RNA5EH2C,RNA5EH2B	1.79e-05	4e-04	RNA catabolic process
GO	Process	GO.0006298	2	26	9606	RNA5EH2A,TREX1	1.97e-05	4e-04	mismatch repair
GO	Process	GO.0090502	2	70	9606	RNA5EH2A,RNA5EH2B	0.00013	0.0024	RNA phosphodiester bond hydrolysis, endonucleolytic
GO	Process	GO.0090304	5	3941	9606	RNA5EH2A,SAMHD1,TREX1,RNA5EH2C,RNA5EH2B	0.00033	0.0046	nucleic acid metabolic process
GO	Process	GO.0006260	2	203	9606	RNA5EH2A,TREX1	0.0011	0.01	DNA replication
GO	Process	GO.0016070	4	3430	9606	RNA5EH2A,SAMHD1,RNA5EH2C,RNA5EH2B	0.0041	0.0318	RNA metabolic process

Table 2: Resultado del enriquecimiento de la comunidad 2

- **GSEA:** Gene Sequence Enrichment Analysis
- **NASQAR:** Nucleic Acid Sequence Analysis Resource
- **KEGG:** Kyoto Encyclopedia of Genes and Genomes
- **ECs:** Endothelial Cells
- **STAT:** Signal Transducer and Activator of Transcription

Disponibilidad de datos y materiales

Repositorio de github: https://github.com/jesusaldanamartin/Gene_Phenotypical_Study

Contribución de los autores

J.C. : Encargado de la introducción, Resultados

J.A.M. : Encargado del abstract, parte de la introducción y Materiales y Métodos

Author details

ETSI Informática, Universidad de Málaga, Málaga, España.

References

1. Jepsen, T.C.: Just what is an ontology, anyway? IT Professional Magazine **11**, 22 (2009)
2. Bard, J.B.L., Rhee, S.Y.: Ontologies in biology: design, applications and future challenges. Nature Reviews Genetics 2004 5:3 5, 213–222 (2004). doi:[10.1038/nrg1295](https://doi.org/10.1038/nrg1295)
3. Human Phenotype Ontology. <https://hpo.jax.org/app/browse/term/HP:0030880>
4. Fenómeno de Raynaud — Temas de Salud — NIAMS. <https://www.niams.nih.gov/es/informacion-de-salud/fenomeno-de-raynaud>
5. Sufka, P.: El Fenómeno de Raynaud. <https://www.rheumatology.org/I-Am-A/Patient-Caregiver/Enfermedades-y-Condiciones/El-Fenomeno-de-Raynaud>
6. Pauling, J.D., Hughes, M., Pope, J.E.: Raynaud's phenomenon-an update on diagnosis, classification and management. Clinical rheumatology **38**, 3317–3330 (2019). doi:[10.1007/S10067-019-04745-5](https://doi.org/10.1007/S10067-019-04745-5)
7. Kuhn, A., Landmann, A.: Systemic lupus erythematosus. European Handbook of Dermatological Treatments, Third Edition, 547–560 (2022).
8. Mohameden, M., Vashisht, P., Sharman, T.: Scleroderma and primary myocardial disease. StatPearls (2022)
9. Plaisier, E., Ronco, P.: Col4a1-related disorders. GeneReviews® (1993)
10. Help - STRING Functional Protein Association Networks. <https://string-db.org/cgi/help?sessionId=bzJHBaJS0L92>
11. Igraph R Package. <https://igraph.org/r/#docs>
12. Larrión, F., Neumann-Lara, V., Pizañ, M.A.: Whitney triangulations, local girth and iterated clique graphs. Discrete Mathematics **258**, 123–135 (2002). doi:[10.1016/S0012-365X\(02\)00266-2](https://doi.org/10.1016/S0012-365X(02)00266-2)
13. Ákos Seress, Szabó, T.: Dense graphs with cycle neighborhoods. Journal of Combinatorial Theory, Series B **63**, 281–293 (1995). doi:[10.1006/jctb.1995.1020](https://doi.org/10.1006/jctb.1995.1020)
14. KEGG Pathway Enrichment Analysis — CD Genomics- Biomedical Bioinformatics. <https://bmb.cd-genomics.com/kegg-pathway-enrichment-analysis.html>
15. Tools for Generating, Visualizing, and Analysing Link Communities in Networks. <https://www.rdocumentation.org/packages/linkcomm/versions/1.0-14>