Aplicación Web interactiva para Minería de textos: Análisis del Síndrome de Williams-Beuren

María Pilar Natividad López
Máster universitario en Bioinformática y Bioestadística UOC-UB
Computación e inteligencia artificial en problemas biológicos y clínicos

Romina Rebrij
David Merino Arranz

01/2019

FICHA DEL TRABAJO FINAL

Título del trabajo:	*Aplicación Web interactiva para Minería de textos: Análisis del Síndrome de Williams-Beuren*
Nombre del autor:	*María Pilar Natividad López*
Nombre del consultor/a:	*Romina Rebrij*
Nombre del PRA:	*David Merino Arranz*
Fecha de entrega (mm/aaaa):	01/2019
Titulación:	*Máster Universitario en Bioinformática y Bioestadística*
Área del Trabajo Final:	*Computación e Inteligencia Artificial en problemas biológicos y clinicos*
Idioma del trabajo:	*Castellano*
Palabras clave	*Minería de textos; Pubmed; Síndrome Williams-Beuren; Web; R*

Resumen del Trabajo

La finalidad de este trabajo fin de máster ha sido el desarrollo de una aplicación web interactiva con Shiny Dashboard, para realizar minería de textos usando el paquete de R pubmed.mineR, sobre un corpus obtenido de Pubmed.

A partir de este corpus se puede navegar por los abstracts que lo componen, permitiendo posteriormente linkar con Pubmed para descargar el documento completo. Permite aplicar filtros, para encontrar los abstracts que contienen un término o un gen. Es posible también obtener las palabras y genes más frecuentes, y luego mostrar la nube de palabras y genes. También permite obtener la información existente en Uniprot para cada gen. Por último ofrece la posibilidad de obtener las enfermedades más comunes relacionadas con el corpus obtenido a través de Pubtator, y sobre ellas averiguar, si existen, las relaciones de similitud de dichas enfermedades con genes, a partir de un análisis semántico latente y la similitud del coseno.

Usando la aplicación se ha realizado un análisis sobre un corpus formado por 1580 abstracts, sobre el Síndrome de Williams-Beuren, trastorno del neurodesarrollo que se presenta desde el mismo momento del nacimiento del individuo, provocado por la pérdida de más de veinticinco genes, como GTF2I, GTF2IRD1, GTF2IRD2, ELN, LIMK1, STX1A, BAZIB, BCL7B y F2D9, en el cromosoma 7 en la posición 11.23. Este trastorno del neurodesarrollo puede provocar entre otras enfermedades, estenosis aórtica supravalvular, hipersociabilidad, deficits en cognición social, pérdida auditiva neurosensorial, hipercalcemia, retraso en el crecimiento y dismorfismo facial.

Abstract :

The purpose of this master's thesis has been the development of an interactive web application with Shiny Dashboard, to perform text mining using the R pubmed.mineR package, on a corpus obtained from Pubmed.

From this corpus you can browse the abstracts that compose it, allowing later to link with Pubmed to download the complete document. Allows to apply filters, to find the abstracts that contain a term or a gene. It is also possible to get the most frequent words and genes, and then display the cloud of words and genes. It also allows obtaining the existing information in Uniprot for each gene. Finally, it offers the possibility of obtaining the most common diseases related to the corpus obtained through Pubtator, and on them to find out, if they exist, the similarity relations of these diseases with genes, from a latent semantic analysis and the similarity of the cosine.

Using the application, an analysis was made of a corpus consisting of 1580 abstracts, about the Williams-Beuren syndrome, a neurodevelopmental disorder that occurs from the moment of the birth of the individual, caused by the deletion of more than twenty-five genes, such as GTF2I, GTF2IRD1, GTF2IRD2, ELN, LIMK1, STX1A, BAZIB, BCL7B and F2D9, on chromosome 7 at position 11.23. This neurodevelopmental disorder can cause among other diseases, supravalvular aortic stenosis, hyper sociability, deficits in social cognition, sensorineural hearing loss, hypercalcemia, growth retardation and facial dysmorphism.

Índice

Lista de figuras

1. Introducción

1.1 Contexto y justificación del Trabajo

En las últimas décadas, la gran cantidad de información, en formato de texto, existente en el campo de la Biología Molecular y de las Ciencias Ómicas a disposición de los investigadores, ha provocado que su tratamiento y análisis de forma manual sea una tarea inviable. Para vencer esta dificultad, las Ciencias de la Información y la Computación han facilitado técnicas automatizadas, como la minería de textos (*text mining*), que posibilitan gestionar esta información, filtrarla e interpretarla.

Cobra mucha importancia poder descubrir información a partir de los datos almacenados en las grandes bases de datos de literatura biomédica, tal como *MEDLINE*, ya que podrá ayudar a interpretar las relaciones existentes entre diversas entidades biológicas. Este es el papel de la minería de textos, y con este proyecto se ha intentado crear una herramienta que sirva de ayuda a los investigadores en tal labor, evitando que éstos tengan que procesar la información de forma manual, imposible hoy en día dada la cantidad de información de la que se dispone.

Los datos de los que se parte, son datos en lenguaje natural no estructurados, por lo que la minería de textos consigue obtener conocimiento a partir de la combinación de gran número de publicaciones que se encuentran almacenadas en estas grandes bases de datos. Es necesario, por tanto, descubrir patrones y clasificar toda esta información para poder observar relaciones existentes y confeccionar redes de conceptos.

La finalidad de este TFM ha sido crear una aplicación web interactiva que ayude al investigador a extraer información usando el sistema de búsqueda Pubmed, desarrollado por National Center for Biotechnology Information (NCBI), para acceder a bases de datos bibliográficas como Genbak, Complete Genoma, MEDLINE y PreMEDLINE, que será analizada aplicando técnicas de Minería de textos.

1.2 Objetivos del Trabajo

Objetivo general:

Desarrollar una aplicación web interactiva mediante Shiny, que use el sistema de búsqueda Pubmed, para extraer información que posteriormente se analizará aplicando técnicas de minería de textos.

Objetivos específicos

- Implementar pantallas en las que sea posible navegar por los abstracts descargados.
- Identificar las palabras más frecuentes.
- Identificar los genes más frecuentes.
- Identificar las relaciones entre genes y enfermedades.
- Identificar redes de asociación entre genes y enfermedades.
- Implementar gráficos que faciliten análisis de los datos.

1.3 Enfoque y método seguido

Existen numerosas plataformas en el mercado para realizar minería de textos. En el artículo (A.Kaur & Chopra, 2016) podemos encontrar una comparativa de todas ellas, clasificadas en free, open source o herramientas online, indicando las técnicas soportadas y sus usos más habituales.

Dado que R ha sido la herramienta de trabajo principal en este máster, dotada de multitud de librerías con distintos fines, su utilización es esencial en este TFM. Su paquete por excelencia para minería de textos es *tm* (Feinerer & Hornik, 2018), con él se pueden tratar formatos de texto heterogéneos, a partir de los cuales generar un corpus, que posteriormente se preprocesa realizando tareas de conversión con la variedad de funciones que contiene.

No obstante, hemos descartado su uso, ya que en este TFM se trabaja con artículos de investigación biomédica descargados de Pubmed. R dispone del paquete *pubmed.mineR* (Jyoti Rani, 2018), una librería específica para el tratamiento de un corpus confeccionado con un número ingente de artículos. Esta librería usa la clase de datos Abstracts, clase compuesta por los atributos propios de un artículo: Journal, Abstract y PMID, y contiene muchas funciones específicas para el tratamiento de los mismos.

Para la extracción del corpus desde Pubmed, se han evaluado las posibles alternativas a *pubmed.mineR*, ya que ésta no dispone de ninguna función para este propósito. La librería *easypubmed* nos proporciona una función que nos permite la descarga, en uno o varios ficheros, de los abstracts filtrados en la búsqueda realizada.

Se han seguido los pasos descritos en (J.RANI, SHAH, & RAMACHANDRAN., 2015), "*pubmed.mineR: An R package with text-mining algorithms to analyse PubMed abstracts*", para desarrollar una aplicación web interactiva creada con el paquete *Shiny Dashboard*.

Tras la extracción del corpus desde *Pubmed*, se ejecuta la tarea de clasificación documental, con el objetivo de filtrar de entre todos los informes encontrados, aquellos que nos interesan para posteriormente identificar entidades, tal como genes y enfermedades. A continuación, se procede a generar la matriz de términos, en la que se muestra la frecuencia con la que cada término aparece en cada informe y que nos sirve para obtener la asociación de términos, a través de un análisis semántico latente (LSA) y la similitud del coseno, permitiendo determinar y cuantificar las relaciones de significado entre términos que aparecen en el mismo corpus.

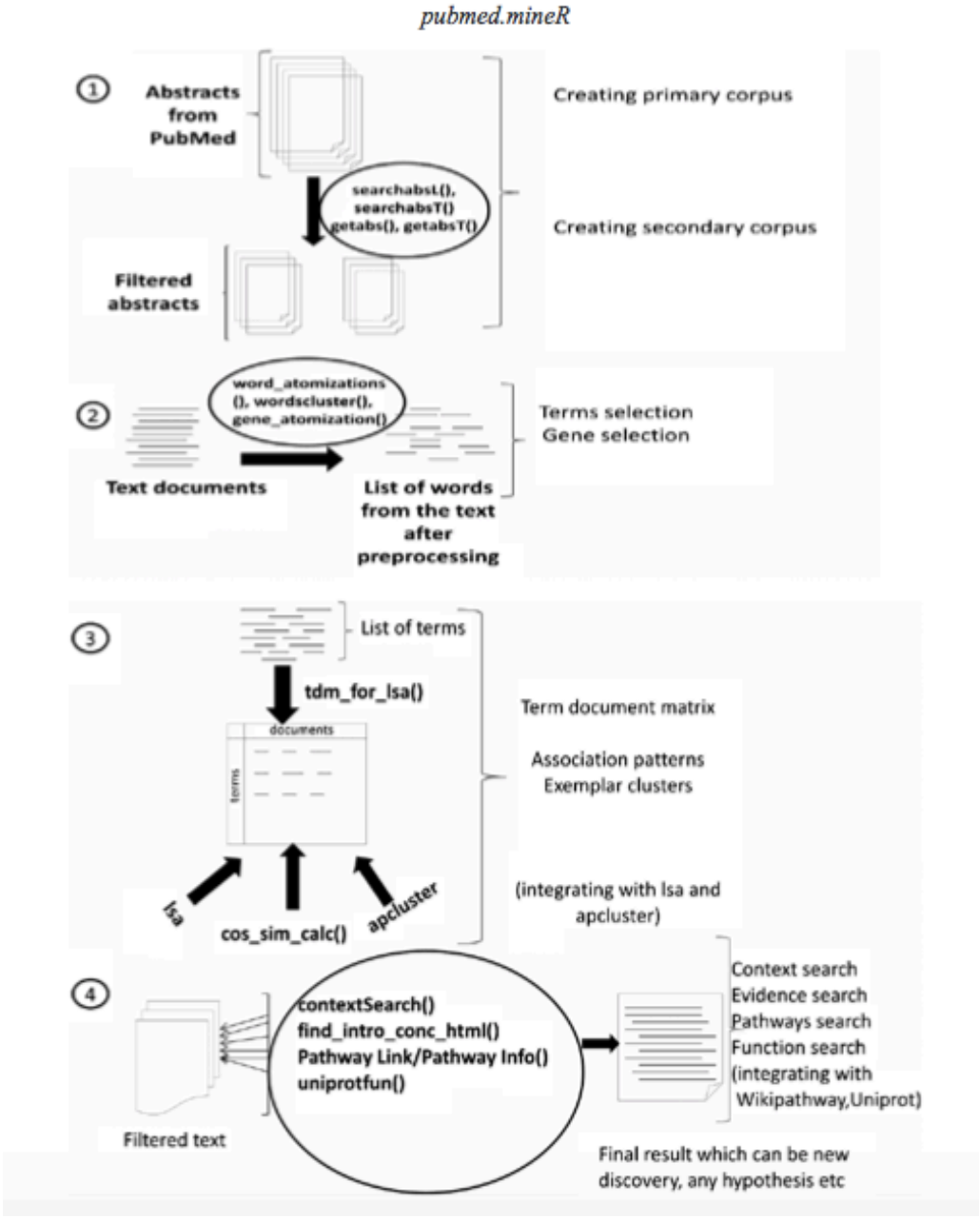


Fig. 1 Proceso mineria textos con pubmed-mineR (J.RANI, SHAH, & RAMACHANDRAN., 2015)

1.4 Planificación del Trabajo

1.4.1 Tareas:

1.4.1.1 Plan de Trabajo inicial

- Definir contenidos del TFM
- Definir las tareas y tiempo asignado a cada una de ellas.

1.4.1.2 Desarrollo del trabajo FASE 1

- Análisis de requerimientos y documentación:
    - Búsqueda de documentación necesaria para llevar a cabo el TFM.
    - Analizar funciones paquete pubmed.mineR, easyPubMed, lsa.
    - Definir sintaxis búsquedas Pubmed.
    - Documentación Síndrome Williams-Beuren.
    - Análisis herramienta Shiny Dashboard.
- Instalar y Configurar Shiny Server Open Source.
- Implementar documento markdown con pasos análisis del corpus fase 1:
    - Instalación librerías R necesarias.
    - Implementar descargar abstracts desde Pubmed.
    - Crear corpus: Concatenar los ficheros descargados en un solo archivo.
    - Explorar y realizar búsquedas en corpus:
        - Búsqueda en el corpus de los Abstracts que contiene un término.
        - Navegar por PMID y visualizar Abstract y Journal.
        - Navegar por PMID y visualizar información Pubtator: Genes, Enfermedades, Mutaciones, composición química, entidades.
    - Obtener frecuencia términos del corpus.
    - Obtener Wordcloud de términos.
    - Obtener genes más frecuentes.
- Diseñar aplicación web Shiny: estructura y aspecto de la interfaz.
- Definir interactividad de la aplicación web en fase 1 (user interface, ui.R): Determinar Inputs y Outputs
- Implementar cálculos de la aplicación fase 1(server.R), siguiendo documento markdown.
- Documentar memoria.

1.4.1.3 Desarrollo del trabajo FASE2

- Implementar documento markdown con pasos análisis del corpus fase 2:
    - Crear corpus secundarios con principales genes.
    - Obtener matriz de genes-enfermedades.
    - Obtener relaciones genes-enfermedades.
    - Obtener redes de asociación.
- Definir interactividad de la aplicación fase 2 (user interface, ui.R): Determinar Inputs y Outputs
- Implementar cálculos de la aplicación fase 2(server.R), siguiendo documento markdown.
- Documentar memoria

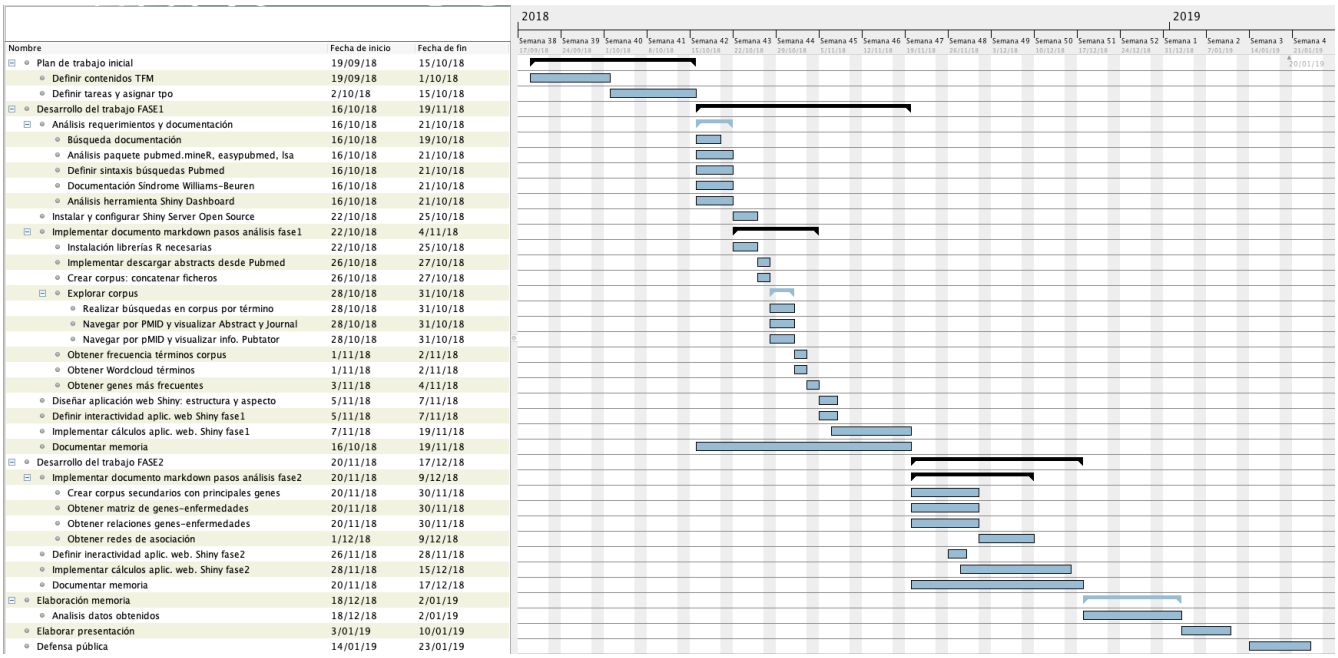
1.4.1.4 Elaboración de la memoria

La memoria se ha realizado durante todo el tiempo que ha durado la elaboración del TFM y se han analizado los datos obtenidos para el caso de estudio.

1.4.1.5 Elaboración de la presentación

1.4.1.6 Defensa pública

1.4.2 Calendario



Nombre	Fecha de inicio	Fecha de fin
Plan de trabajo inicial	19/09/18	15/10/18
Definir contenidos TFM	19/09/18	1/10/18
Definir tareas y asignar tpo	2/10/18	15/10/18
Desarrollo del trabajo FASE1	16/10/18	19/11/18
Análisis requerimientos y documentación	16/10/18	21/10/18
Búsqueda documentación	16/10/18	19/10/18
Análisis paquete pubmed.mineR, easypubmed, lsa	16/10/18	21/10/18
Definir sintaxis búsquedas Pubmed	16/10/18	21/10/18
Documentación Síndrome Williams-Beuren	16/10/18	21/10/18
Análisis herramienta Shiny Dashboard	16/10/18	21/10/18
Instalar y configurar Shiny Server Open Source	22/10/18	25/10/18
Implementar documento markdown pasos análisis fase1	22/10/18	4/11/18
Instalación librerías R necesarias	22/10/18	25/10/18
Implementar descargar abstracts desde Pubmed	26/10/18	27/10/18
Crear corpus: concatenar ficheros	26/10/18	27/10/18
Explorar corpus	28/10/18	31/10/18
Realizar búsquedas en corpus por término	28/10/18	31/10/18
Navegar por PMID y visualizar Abstract y Journal	28/10/18	31/10/18
Navegar por pMID y visualizar info. Pubtator	28/10/18	31/10/18
Obtener frecuencia términos corpus	1/11/18	2/11/18
Obtener Wordcloud términos	1/11/18	2/11/18
Obtener genes más frecuentes	3/11/18	4/11/18
Diseñar aplicación web Shiny: estructura y aspecto	5/11/18	7/11/18
Definir interactividad aplic. web Shiny fase1	5/11/18	7/11/18
Implementar cálculos aplic. web. Shiny fase1	7/11/18	19/11/18
Documentar memoria	16/10/18	19/11/18
Desarrollo del trabajo FASE2	20/11/18	17/12/18
Implementar documento markdown pasos análisis fase2	20/11/18	9/12/18
Crear corpus secundarios con principales genes	20/11/18	30/11/18
Obtener matriz de genes-enfermedades	20/11/18	30/11/18
Obtener relaciones genes-enfermedades	20/11/18	30/11/18
Obtener redes de asociación	1/12/18	9/12/18
Definir ineractividad aplic. web. Shiny fase2	26/11/18	28/11/18
Implementar cálculos aplic. web. Shiny fase2	28/11/18	15/12/18
Documentar memoria	20/11/18	17/12/18
Elaboración memoria	18/12/18	2/01/19
Analisis datos obtenidos	18/12/18	2/01/19
Elaborar presentación	3/01/19	10/01/19
Defensa pública	14/01/19	23/01/19

Fig. 2 Calendario TFM

1.5 Breve sumario de productos obtenidos

1. Plan de trabajo del TFM

2. Memoria

Documento que se ha generado a lo largo del proyecto, actualizándose en todas sus etapas.

3. Producto

Aplicación en plataforma web Shiny Dashboard útil para realizar análisis de minería de textos en tareas de investigación. Se incluye la URL del servidor en el que está alojada y el código fuente.

4. Presentación virtual

Exposición del trabajo usando una presentación de PowerPoint.

Autoevaluación del proyecto.

Una vez finalizado el TFM se ha realizado una autoevaluación, que sirve de reflexión acerca de todo lo que se ha aprendido en la elaboración del mismo.

1.6 Breve descripción de los otros capítulos de la memoria

En el capítulo 2 se presenta un breve estudio sobre Síndrome de Williams-Beuren, para dar una visión del fenotipo y enfermedades más características.

En el capítulo 3 se realiza una introducción a la minería de textos describiendo las funciones que en este TFM se han utilizado del paquete pubmed.mineR.

En el capítulo 4 se hace una introducción a las búsquedas desde Pubmed sobre MEDLINE, que se realizan usando la aplicación Shiny.

En el capítulo 5, se hace un breve resumen del paquete Shiny.

En el capítulo 6, se presenta el proceso seguido en este TFM para realizar minería de textos a partir de un corpus descargado en Pubmed, y como se aplica este proceso al caso de estudio.

Se incluye un Anexo con el código fuente de la aplicación desarrollada, así como un Glosario con términos y expresiones utilizadas en este texto.

2. Síndrome de Williams-Beuren

(JURADO, 1999) El Sindrome de Williams-Beuren, es un trastorno del neurodesarrollo, descrito por primera vez en 1961 por el cardiólogo John Williams y de forma paralela por el pediatra Alois Beuren, con una prevalencia de 1 en 7.500 (STRØMME, P.G.BIØMSTAD, & K.RAMSTAD, 2002), que se presenta desde el mismo momento del nacimiento del individuo, provocado por la pérdida de genes en el cromosoma 7 en la posición 11,23 (SO & KIM, 2015). Afecta a todas las razas, y tanto a hombres como a mujeres, y la pérdida de material genético supone aproximadamente más de millón y medio de nucleótidos, provocada por el alineamiento anormal de los dos cromosomas 7 durante la meiosis, bien en el cromosoma resultante que va al óvulo o al espermatozoide.

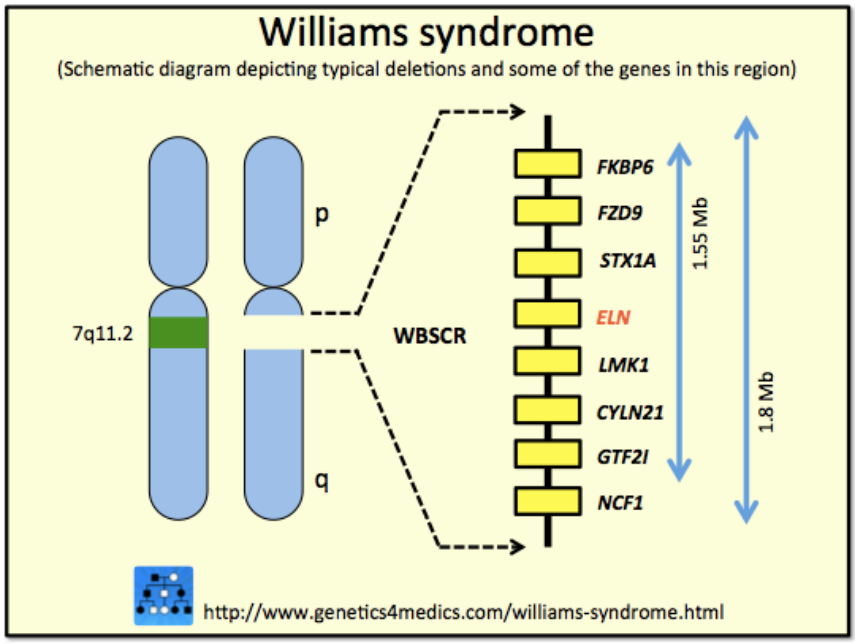


Fig. 3 Genes que han sido borrados del cromosoma 7 (SO & KIM, 2015)

Supone la pérdida de más de 25 genes, aunque no todos los genes tienen que tener un mal funcionamiento, debido a la existencia de la otra copia del cromosoma 7, aunque en la mayoría de los casos la delección puede ocurrir tanto en el cromosoma paterno como en el materno.

Aunque su herencia es autosómica dominante, no todos los casos los son por herencia, ya que no es habitual la existencia de otros casos en la familia.

No existe conocimiento detallado sobre la participación de cada gen perdido con las características clínicas del síndrome, aunque si que es conocida la relación de la pérdida del gen ELN con los problemas cardiovasculares. Es objetivo de este TFM será relacionar cada gen con las enfermedades en las que podría contribuir.

Para su diagnóstico clínico puede hacerse un análisis de sangre, aunque mediante la técnica de hibridación in situ fluorescente (FISH), puede detectarse la deleción de la elastina en el cromosoma 7, en más del 98% de las personas con SW. También se utilizan microsatélites para del diagnóstico, lo que requiere una muestra de tanto la madre como del padre.

En los pacientes pueden estar afectados diferentes órganos del cuerpo, y no siempre los mismos, aunque si que suponen problemas significativos.

Los síntomas más comunes del SW son (Casanelles & Jurado, 2010):

- *Retraso en el crecimiento*, ya que nacen con peso y talla reducidos, y al crecer no alcanzan la talla media familiar.

- *Anomalías craneofaciales,* presentándose un estrechamiento de la frente, pliegues epicánticos, región periorbital prominente, desviación de uno o ambos ojos, a veces de forma latente, otras de forma manifiesta, iris con un patrón estrellado en niños con ojos verdes o azules, nariz corta y respingada, zona alargada entre base de la nariz y labio superior, mofletes prominentes y algo caídos, boca ancha y labios gruesos y el mentón pequeño.

- *Defectos estructurales del corazón y de los vasos sanguíneos* son detectados en aproximadamente el 80% de los pacientes, pudiendo presentar estrechamiento aórtico supravalvular, estrechamiento en las arterias pulmonares y a veces también en las cerebrales, lo que puede provocar hipertensión arterial.

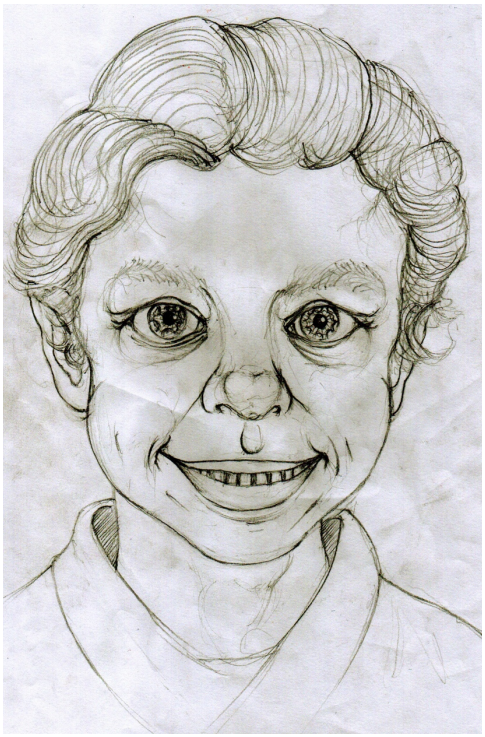


Fig. 4 Rostro característico SW.
(Pintura original R.Vivo 2018)

- *Laxitud articular e hipotonicidad,* con disminución del tono y fuerza muscular, provocada por la escasez de elastina, pudiendo sufrir a la larga escoliosis, cifosis y lordosis. Es típica la postura con hombros caídos y rodillas semiflexionadas.

- *Hiperacusia* o disminución de la tolerancia a los sonidos habituales y naturales del ambiente, que en un determinado umbral se presentan como molestos. En algunos casos presentan un talento musical similar al de los músicos con oído musical perfecto.

- *Signos de envejecimiento precoz*, con el cutis fino y laxo y pudiendo presentar canas prematuras.

- *Nefrocalcinosis*, debida a problemas renales y de vejiga urinaria, pudiendo presentar infecciones urinarias y en algunos casos urgencia en la micción o falta de control nocturno.

- *Hipercalcemia* o presencia de altos niveles de calcio en la sangre, en un porcentaje reducido de pacientes, siendo habitual que desaparezca a partir de los dos años, aunque debe controlarse durante toda la vida. En recién nacidos puede provocar cólicos que persisten hasta los 10 meses de edad.

- *Problemas en el aparato digestivo*, como complicaciones gastrointestinales y reflujo gastro-edofágico, que pueden interferir en el crecimiento.

- *Hernias inguinales y umbilicales*, pueden presentarse en una proporción superior a la de la población general.

- *Problemas de equilibrio y de la coordinación del movimiento*, así como de la orientación espacial.

- *Personalidad afectuosa y gran sociabilidad* con adultos, teniendo un comportamiento muy extrovertido, aunque a veces presentan problemas de relación con otros niños. No sienten temor hacia las personas extrañas, generalmente.

- *Retraso del lenguaje inicial*, consiguiendo con la edad buena expresividad, pero con deficiencias en los contenidos.

- *Coeficiente intelectual medio* en torno a los 60, siendo superior el coeficiente verbal que el manipulativo.

- *Buena memoria*, sobre todo la auditiva.

- *Hiperactividad* en edades tempranas, mejorando con la edad.

- *Capacidad de concentración escasa*, lo que requiere un apoyo continuo.

- *Anomalías dentales*, teniendo los dientes pequeños y muy separados, con problemas en la mordida, que pueden corregirse con ortodoncia.

En este TFM se ha realizado un análisis de las enfermedades y genes relacionadaos con el SW, buscado las relaciones de simulitud existentes.

3. Minería de textos con pubmed.mineR

Pubmen.mineR es una librería de funciones para hacer minería de textos con los documentos descargados de MEDLINE, base de datos bibliográfica producida por la National Library of Medicine de los Estados Unidos, donde se pueden encontrar millones de artículos sobre ciencia biomédica. Para la descarga de los documentos se usa Pubmed, que es un motor de búsqueda basado en tecnología word wide web.

Dado que *pubmed.mineR* no contiene ninguna función para realizar la descarga de los abstracts desde Pubmed, se ha hecho uso de la función *batch_pubmed_download* del paquete *easyPubMed*, que guarda en un disco local en uno o varios ficheros de texto, dependiendo la consulta realizada.

```
out.A <- batch_pubmed_download(pubmed_query_string = new_query,
                               format = "abstract",
                               batch_size = 1500,
                               dest_file_prefix = "pubmed_")
```

Mediante el parámetro *pubmed_query_string* se le pasa la consulta a realizar en Pubmed, y se le indica que el tipo de datos a descargar son abstracts con el parámetro *format*. Aunque el máximo número de registros que se pueden descargar en un fichero de texto es 5.000, se usará 1.500, indicándoselo con el parámetro *batch_size.* El nombre de los ficheros descargados tendrá el prefijo indicado en *dest_file_prefix***.**

Para realizar minería de textos con el fichero descargado desde Pubmed, éste se carga en un objeto de R de tipo S4, que es una clase usada en R, en este caso formada por 3 slots:

- Journal: vector que contiene, entre otros, los datos de autor, título, referencia bibliográfica, idioma (en caso de que no esté redactado en inglés), fecha de publicación.
- Abstract: vector con los resúmenes de los informes seleccionados.
- PMID: vector con los PMID, identificador formado por 8 dígitos que Pubmed asigna a cada informe.

Se usa la función *readabs()*, para realizar la carga del fichero de texto en el objeto S4 y a su vez genera un fichero de texto delimitado por tabuladores llamado "newabs.txt" en el disco local. Así, se consigue crear el corpus que posteriormente se analizará.

```
corpus <- readabs("pubmed_result.txt")
```

Se accede a cada unos de los slots del objeto S4 de la siguiente manera:

```
corpus@Journal
corpus@Abstracts
corpus@PMID
```

Si se quiere explorar la información del corpus se usa la función *pubtator_function(),* pasándole como parámetro el PMID del informe o un vector con varios PMID, generándose una lista de 6 vectores, con el PMID, información sobre qué genes, mutaciones, enfermedades, elementos químicos y entidades aparecen en el informe:

```
pubtator_output<- pubtator_function(corpus@PMID[1])
pubtator_output$Genes
pubtator_output$Mutations
pubtator_output$Diseases
pubtator_output$Chemicals
pubtator_output$Species
pubtator_output$PMID
```

Para realizar la tokenización del texto de los informes en palabras se usa la función *word_atomizations(),* cuyo argumento será el objeto S4 generado al ejecutar *readabs(),* y su salida un dataframe con dos campos: las palabras y su frecuencia.

```
corpus_words<- word_atomizations(corpus)
corpus_words$words
corpus_words$Freq
```

De igual manera se puede obtener la frecuencia de los genes que aparecen en los informes usando la función *gene_atomization(),* pasándole como argumento el objeto S4 y obteniendo como salida el fichero *table.txt* guardado en el disco local y una tabla con tres campos: el identificador HGNC del gen, su descripción y su frecuencia.

```
objetoS4gene<- gene_atomization(corpus)
objetoS4gene
```

	Gene_symbol	Genes	Freq
[1,]	"T"	"T, brachyury homolog (mouse)"	"93"
[2,]	"AR"	"androgen receptor"	"8"
[3,]	"CP"	"ceruloplasmin (ferroxidase)"	"6"
[4,]	"KL"	"klotho"	"6"
[5,]	"CS"	"citrate synthase"	"5"
[6,]	"MB"	"myoglobin"	"5"
[7,]	"GTF2I"	"general transcription factor IIi"	"3"
[8,]	"PC"	"pyruvate carboxylase"	"3"
[9,]	"GC"	"group-specific component (vitamin D binding protein)"	"2"
[10,]	"GTF2IRD1"	"GTF2I repeat domain containing 1"	"2"

Fig. 5 Salida función gene_atomization()

Para crear un corpus secundario se usa la función *searchabsL()*, de forma que se filtrarán aquellos abstracts del corpus inicial que contienen la palabra o palabras que se le pase en el vector *include*, y que no contienen la palabra o palabras que se le pase en el vector *exclude*. Dado que las palabras pasadas en el vector include usan el operador booleano OR, se puede restringir que en todos los informes aparezca una palabra o palabras usando el vector *restrict*, que usará el operador booleano AND. También se pueden filtrar los años de búsqueda usando *yr*.

```
corpus_secundario <- searchabsL(corpus,
                                yr = vector_year,
                                include = vector_include,
                                restrict = vector_restrict,
                                exclude = vector_exclude)
```

El método usado para analizar el texto y extraer relaciones de similitud entre palabras ha sido el Análisis Semántico Latente(LSA), que realiza una serie de cálculos algebraicos para obtener una representación de las palabras en un espacio multidimensional, usando la técnica de Descomposición en Valores Singulares(SVD) y obteniendo los componentes principales de la representación del texto (SLOMOVITZ, 2016). De esta forma se logra reducir la dimensionalidad de la matriz de términos y consecuentemente extraer su espacio semántico.

Para ello, el primer paso ha sido ejecutar la función *tdm_for_lsa(),* que crea la "*Term Document Matrix*", una matriz cuyas columnas son los informes de que consta el corpus secundario, que se le pasa como primer argumento, y cuyas filas son los términos pasados en un vector, como segundo argumento. Las frecuencias de cada término en cada informe son los valores de esta matriz.

```
terminos <- c("chromosome","social","deleted”,"phenotype")
tdm_corp_sec<- tdm_for_lsa(corpus_secundario, terminos)
```

El algoritmo LSA se ejecuta con la función *lsa()* de la librería del mismo nombre, que a partir de la matriz de términos de documentos y el número de valores singulares óptimo, generada mediante la función *dimcalc_share()*, obtiene un nuevo espacio vectorial LSA, cuyas dimensiones son el número de valores singulares y el número de palabras analizado. De esta forma se ha conseguido reducir la dimensionalidad de la matriz original.

```
library(lsa)
lsa_corp_sec <- lsa(tdm_corp_sec, dims=dimcalc_share())
GTF2IRD1_corpus<- searchabsL(tdm_corp_sec, include="GTF2IRD1")
terminos<-c("anxiety", "autism", "schizophrenia", "Lorna Wing","psychiatric disorders", "genetic disorder", "cardiovascular malformations", "mental retardation", "facial dysmorphism")
tdm_corp_sec<- tdm_for_lsa(GTF2IRD1_corpus, terminos)
lsa_corp <- lsa(tdm_corp_sec, dims=dimcalc_share())
```

```
$tk
                                 [,1]       [,2]
anxiety                    0.1170143 -0.5538127
autism                     0.5635980  0.6167244
schizophrenia              0.0000000  0.0000000
Lorna Wing                 0.0000000  0.0000000
psychiatric disorders      0.0000000  0.0000000
genetic disorder           0.7001681 -0.5232495
cardiovascular malformations 0.0000000  0.0000000
mental retardation         0.4224092  0.1978697
facial dysmorphism         0.0000000  0.0000000

$dk
             [,1]          [,2]
 [1,] 0.28921404 -6.276419e-01
 [2,] 0.19946641  3.593869e-01
 [3,] 0.00000000  1.110223e-16
 [4,] 0.04141326 -3.227261e-01
 [5,] 0.00000000  0.000000e+00
 [6,] 0.84456540 -1.351391e-01
 [7,] 0.19946641  3.593869e-01
 [8,] 0.00000000  0.000000e+00
 [9,] 0.00000000  0.000000e+00
[10,] 0.34896384  4.746925e-01
[11,] 0.00000000  0.000000e+00
[12,] 0.00000000  0.000000e+00

$sk
[1] 2.825528 1.716046

attr(,"class")
[1] "LSAspace"
```

Fig. 6 Salida función lsa()

Se puede observar que la matriz de términos se ha descompuesto en tres (Deerwester, 1990):

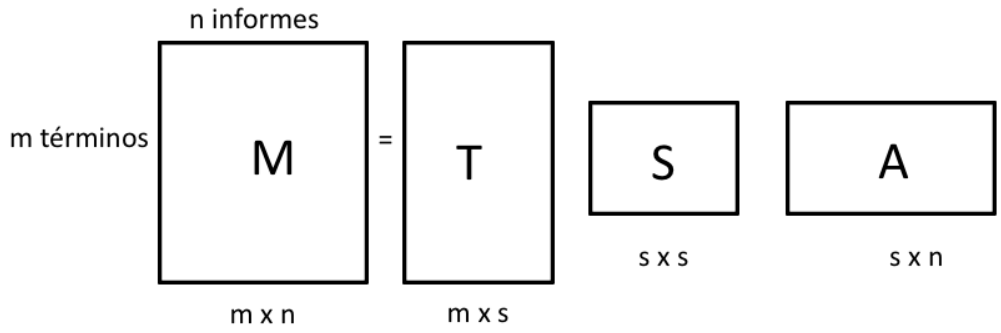


Fig. 7 Descomposición de la matriz de términos

Siendo S la matriz de valores singulares, T y A las matrices de términos y de abstracts respectivamente. Todas ellas son reducidas a un número de dimensiones k, obteniendo T_k S_k y A_k, que son las matrices resultado de ejecutar la función *lsa()*, y que juntas forman el espacio semántico multidimensional: (T_k , S_k , A_k).

Para poder obtener la asociación es necesario convertir el espacio latente semántico a una nueva matriz de términos clásica. Para ello se ha usado la función *as.textmatrix()*.

```
espacio_semantico <- as.textmatrix(lsa_corp)
```

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]	[,8]	[,9]	[,10]	[,11]	[,12]
anxiety	0.6921128	-0.2756008	-1.055120e-16	0.32040089	0	0.4076682	-0.2756008	0	0	-0.3357556	0	0
autism	-0.2036882	0.6979917	1.174979e-16	-0.27560080	0	1.2019170	0.6979917	0	0	1.0580918	0	0
schizophrenia	0.0000000	0.0000000	0.000000e+00	0.00000000	0	0.0000000	0.0000000	0	0	0.0000000	0	0
Lorna Wing	0.0000000	0.0000000	0.000000e+00	0.00000000	0	0.0000000	0.0000000	0	0	0.0000000	0	0
psychiatric disorders	0.0000000	0.0000000	0.000000e+00	0.00000000	0	0.0000000	0.0000000	0	0	0.0000000	0	0
genetic disorder	1.1357373	0.0719126	-9.968915e-17	0.37171193	0	1.7921857	0.0719126	0	0	0.2641349	0	0
cardiovascular malformations	0.0000000	0.0000000	0.000000e+00	0.00000000	0	0.0000000	0.0000000	0	0	0.0000000	0	0
mental retardation	0.1320674	0.3601001	3.769801e-17	-0.06015483	0	0.9621266	0.3601001	0	0	0.5776821	0	0
facial dysmorphism	0.0000000	0.0000000	0.000000e+00	0.00000000	0	0.0000000	0.0000000	0	0	0.0000000	0	0

Fig. 8 Nueva matriz de términos

A continuación se aplica la función *associate()* del paquete *lsa*, para encontrar la similitud entre términos usando el coseno como medida de asociación o cercanía. Para ello, internamente se calcula una tabla de similitud entre cada uno de los términos entre si, y se muestran aquellos términos entre los que realmente existe una relación superior a un umbral escogido mediante el argumento *threshold*.

```
terminos_asociados <- lapply(terminos, function(x){
associate(espacio_semantico, x , measure = "cosine", threshold = "0.7")})
names(terminos_asociados)<-terminos
```

```
$anxiety
NULL

$autism
mental retardation
           0.95252

$schizophrenia
NULL

$`Lorna Wing`
NULL

$`psychiatric disorders`
NULL

$`genetic disorder`
mental retardation
         0.7627486

$`cardiovascular malformations`
NULL

$`mental retardation`
          autism genetic disorder
       0.9525200         0.7627486

$`facial dysmorphism`
NULL
```

Fig. 9 Asociación de términos

También se ha utilizado la función *cos_sim_calc()* del paquete *pubmed.mineR*, para obtener un fichero de texto guardado en el disco local, con las distancias entre los términos:

```
cos_sim_calc(espacio_semantico)
```

```
anxiety facial dysmorphism      NaN
autism  schizophrenia   NaN
autism  Lorna Wing      NaN
autism  psychiatric disorders   NaN
autism  genetic disorder        0.447213595499958
autism  cardiovascular malformations    NaN
autism  mental retardation      0.707106781186547
autism  facial dysmorphism      NaN
schizophrenia   Lorna Wing      NaN
schizophrenia   psychiatric disorders   NaN
schizophrenia   genetic disorder        NaN
schizophrenia   cardiovascular malformations    NaN
schizophrenia   mental retardation      NaN
schizophrenia   facial dysmorphism      NaN
Lorna Wing      psychiatric disorders   NaN
Lorna Wing      genetic disorder        NaN
Lorna Wing      cardiovascular malformations    NaN
Lorna Wing      mental retardation      NaN
Lorna Wing      facial dysmorphism      NaN
psychiatric disorders   genetic disorder        NaN
psychiatric disorders   cardiovascular malformations    NaN
psychiatric disorders   mental retardation      NaN
psychiatric disorders   facial dysmorphism      NaN
genetic disorder        cardiovascular malformations    NaN
genetic disorder        mental retardation      0.632455532033676
genetic disorder        facial dysmorphism      NaN
cardiovascular malformations    mental retardation      NaN
cardiovascular malformations    facial dysmorphism      NaN
mental retardation      facial dysmorphism      NaN
```

Fig. 10 Fichero cossimdata.txt, generado al ejecutar cos_sim_calc()

Para tokenizar los documentos, se ha usado la función *SentenceToken*, que devuelve un vector con los párrafos del documento.

```
frases<-SentenceToken(documento@Abstract)
```

Y para tokenizar los párrafos en frases se ha usado la función *tokenize_sentences()* del paquete *tm*, que también devuelve un vector con cada una de las frases que integran los párrafos del abstract.

```
frasesTitulo<-tokenize_sentences(frases[1], lowercase= FALSE)
```

Para filtrar uno o más documentos del corpus, se crea un corpus secundario utilizando la función *subabs()*, cuyos parámetros son el corpus del que se extraen los documentos, y *desde_pos-hasta_pos* la posición en la que se encuentran los que se quieren extraer. Si sólo se quiere filtrar un documento, el parámetro *desde_pos* coincide con el parámetro *hasta_pos*.

```
doc<- subabs(corpus, desde_pos,hasta_pos)
```

4. Búsquedas en Medline con Pubmed

En 1996, la National Center for Biotechnology Information (NCBI) lanzó Pubmed (http://www.ncbi.nlm.nih.gov/PubMed/), el motor de búsqueda de libre acceso a la base de datos MEDLINE, PreMEDLINE, GenBank y Complete Genoma.

Desde la aplicación que ha sido desarrollada en este TFM se accede a Pubmed, mediante el uso de una función de la librería *easyPubMed*, por lo que es necesario para su ejecución que se construya una consulta usando la nomenclatura seguida por esta plataforma.

Pubmed permite realizar *búsquedas sencillas*, introduciendo tan sólo un término o frase entrecomillada en el cuadro de búsqueda, y *búsquedas avanzadas* usando operadores lógicos escritos en mayúsculas (AND, OR, NOT) y usando etiquetas entre corchetes. Permite recuperar todos los términos que tengan la misma raíz usando el comodín *.

La funcionalidad de los operadores booleanos se describe a continuación:

- AND o intersección, recuperará sólo los artículos que contengan todos los términos usados en el cuadro de búsqueda.
- OR o unión, recuperará los artículos que contengan todos los términos o al menos uno de ellos.
- NOT, excluye los artículos que contengan el término.

Por ejemplo, si se busca *Williams syndrome*, se encuentran 7001 resultados, entre los que aparecerá la palabra *Williams* y la palabra *syndrome*, pero no tienen porque ir juntas.

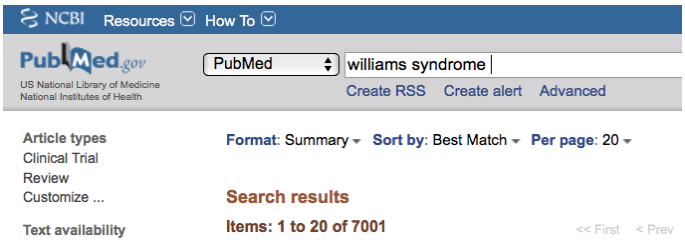


Fig. 11 Búsqueda en Pubmed

Esto no es correcto, así que se será más restrictivo, indicándole que deben ir juntas ambas palabras, usando doble comilla al principio y fin de la frase, siendo el número de artículos encontrados menor.

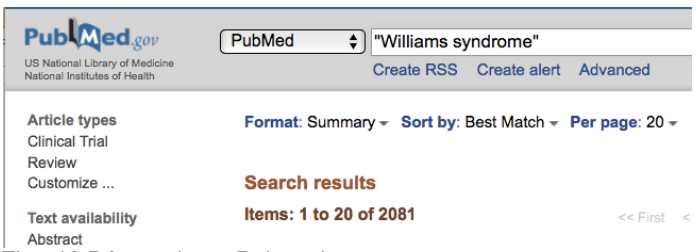


Fig. 12 Búsqueda en Pubmed

El uso de las etiquetas entre corchetes ayudará a definir de forma más clara los criterios de búsqueda:

- La etiqueta **[MH]** o *Medical Subject Headings*, asegura que el término que se ha usado forma parte del vocabulario controlado por MEDLINE, que consta de más de

33.000 términos ordenados en árboles, estructuras jerárquicas que se revisan anualmente para así garantizar que forma parte de la terminología médica, y de esta forma conseguir realizar búsquedas de alta calidad. (M.A.Fernández-Altunaa, y otros, 2016)

Por ejemplo, desde Pubmed se indica que se quiere realizar una búsqueda en MESH, usando el término en español "síndrome Williams" y avisa que ese término no existe, y muestra otras alternativas más genéricas.

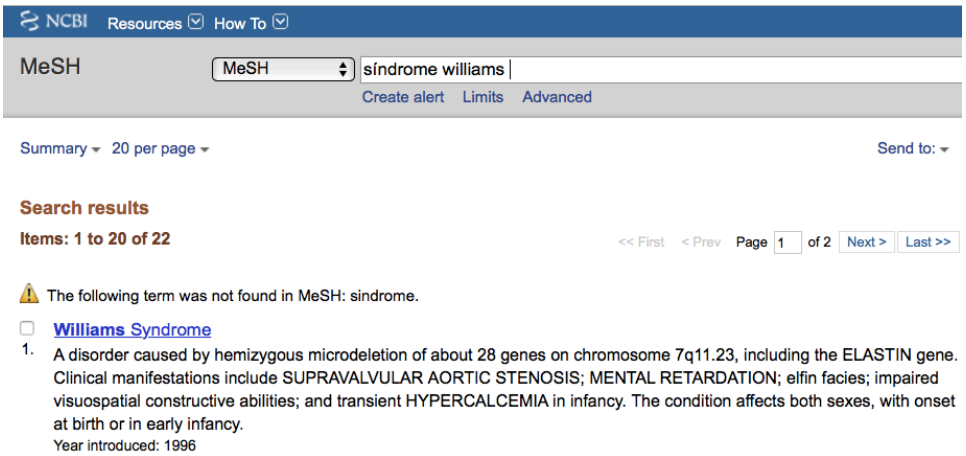


Fig. 13 Búsqueda desde Mesh

Ahora la búsqueda será más restrictiva:

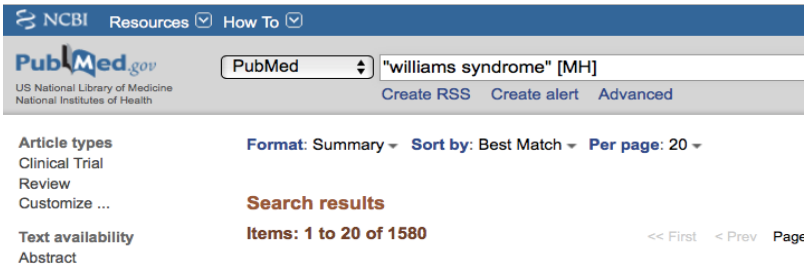


Fig. 14 Búsqueda con etiqueta Mesh

- La etiqueta [DP] o Publication Date, fecha de publicación con formato *aaaa/mm/dd*, ayuda a acotar la búsqueda.

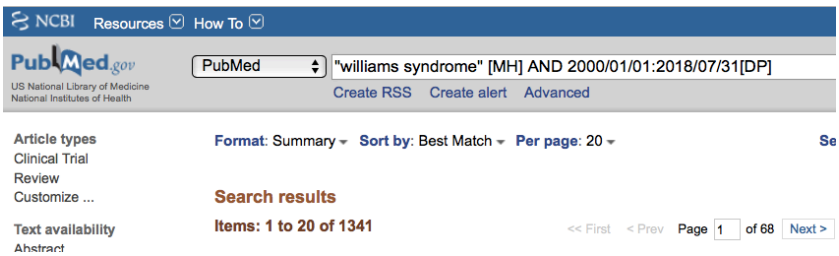


Fig. 15 Búsqueda en Pubmed con Mesh y filtro de fechas

5. Shiny y las aplicaciones interactivas

Shiny es un paquete de R desarrollado por RStudio, usado para crear aplicaciones web interactivas, usando los scripts de R, y que permite a los usuarios interactuar con los datos, sin que sea necesario manipular el código. Se fundamenta en la programación reactiva, creando un vínculo entre los datos de entrada y los datos de salida. Emplea una sintaxis amigable, que integra código HTML, CSS y JavaScript en el código fuente de sus componentes, sin que exista la necesidad de conocerlos, permitiendo crear una aplicación web de forma sencilla.

Para poder usarlo, se debe instalar y luego cargar en R el paquete *shiny*:

```
install.packages("shiny")
library("shiny")
```

Usando la librería *shinydashboard*, se podrá construir un cuadro de mando:

```
install.packages("shinydashboard")
library("shinydashboard")
```

Una aplicación Shiny consta al menos de dos archivos, uno de ellos que recibe los inputs o entradas y los outputs o salidas, y representa el diseño e interfaz gráfica de la aplicación, cuyo nombre es *ui.R*, y el otro en el que se realizan los cálculos que se ejecutarán para cada uno de los usuarios que ejecute la app, cuyo nombre es *server.R.*

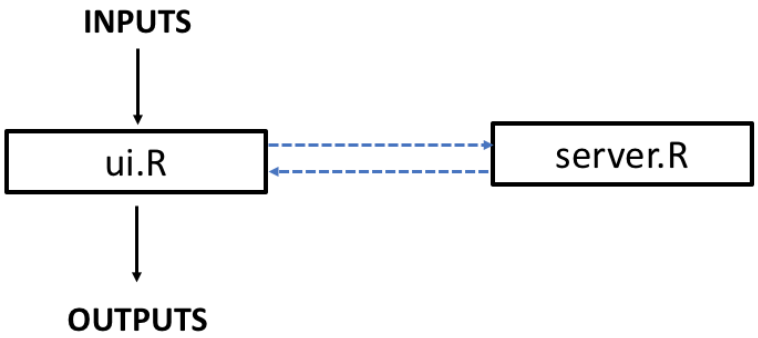


Fig. 16 Estructura aplicación Shiny

O bien, como es el caso adoptado en la aplicación desarrollada en este trabajo, puede constar de un solo fichero:

```
library("shiny")
library("shinydashboard")
ui<- dashboardPage(header, skin, sidebar, body)
server<- function(input, output){…..}
shinyApp(ui = ui, server = server)
```

En Shiny se trabaja con tres tipos de objetos de programación reactiva:

- Las fuentes reactivas, formadas por los inputs que aparecen en *ui* y se envían al *server*.
- Los conductores reactivos, que transforman los inputs que usa el *server*.
- Los puntos finales reactivos o outputs obtenidos en server y que se envían al ui para ser visualizados.

Los **inputs** permitidos en las aplicaciones Shiny se introducen mediante widgets o elementos web, y para su uso es necesario instalar el paquete que los contiene:

```
install.packages("shinyWidgets")
library(shinyWidgets)
```

Los **elementos web** que se pueden usar son:

- *textInput(inputId, label, value)* : Crea un campo para introducir texto.
- *numericInput(inputId, label, value, min, max, step, width)*: Crea un campo para introducir un valor numérico.
- *dateInput(inputId, label, value, min. max, format, startview, weekstart, language, width)*: Crea un campo para introducir una fecha, pudiendo introducirla mediante un calendario.
- *dateRangeInput(inputId, label, start, end, min. max, format, startview, weekstart, language, separator, width)*: Crea un par de campos para introducer fechas o bien seleccionarlas desde un calendario.
- *checkboxInput(inputId, label, value, width)*: Crea un campo checkbox para usar valores lógicos (TRUE o FALSE).
- *checkboxIGroupInput(inputId, label, choices, selected, inline, width)*: Crea un grupo de campos checkbox para usar múltiples valores lógicos (TRUE o FALSE).
- *actionButton(inputId, label, icon, width)*: Crea un botón para producir una acción tras ser pulsado.
- *radioButtons(inputId, label, choices, selected, inline, width):* Crea un conjunto de botones asociados a una lista, para seleccionar una opción de la misma.
- *fileInput(inputId, label, multiple, accept, width)*: Crea un control para cargar uno o más ficheros.
- *selectInput(inputId, label, choices, selected, multiple, selectize, width, size)*: Crea una lista desplegable en la que se puede elegir uno o más ítems.
- *sliderInput(inputId, label, min, max, value, step, round, format, locale, ticks, animate, width, sep, pre, post, timeFormat, timezone, dragRange)*: Crea un barra deslizante en la que se puede seleccionar un valor numérico de un rango.
- *helpText("")*: Crea una nota con ayuda al usuario.

Todos estos controles se introducen en ui, y cuando son referenciados en server se usa la nomenclatura **input$inputId,** y son los datos de entrada que se usarán para transformar y obtener las salidas de la aplicación.

Las operaciones necesarias para generar las salidas de la aplicación están incluidas en objetos de tipo reactivo, por lo que su nombre comienza por render y son generadas en *server*. A continuación se muestran los tipos de **objetos *render*** que se pueden utilizar:

- *renderImage*: para generar una imagen.
- *renderPlot*: para generar un gráfico.
- *renderPrint*: para generar cualquier tipo de salida.
- *renderTable*: para generar una tabla.
- *renderText*: para generar texto.
- *renderUI*: para generar código HTML o cualquier objeto shiny dinámico, tal como una lista desplegable.

Una vez generadas las salidas desde *server* pueden ser visualizadas usando las siguientes **salidas en *ui***:

- *htmlOutput(outputId, inline, container)*: se muestra la salida con tags div de HTML.
- *imageOutput(outputId, width, height, inline)*: la salida mostrará una imagen.
- *plotOutput(outputId, width, height, click, dblclick, hover, hoverDelay, hoverDelayType, brush, clickId, hoverId, inline)*: la salida mostrará un gráfico.
- *tableOutput(outputId)*: la salida mostrará una tabla.
- *dateTableOutput(outputId):* la salida mostrará una tabla interactiva.
- *textOutput(outputId)*: la salida mostrará texto.
- *uiOutput*: genera input de entrada de forma interactiva dependiendo de otras variables.

- *verbatimTextOutput(outputId, placeholder)*: la salida será mostrada con tags pre de HTML.

Es posible incluir código HTML en la aplicación, usando *tags$* y el código html que quiera utilizarse: tags$div, tags$img,…

Para la creación del menú en el lateral izquierdo del panel de mando, se usa el comando *menuItem*:

```
sidebarMenu(
  menuItem(“Op 1 del menú”, tabName=”op1”, icon=icon("nom_icono",lib="font-awesome"))
  menuItem(“Op 2 del menú”, tabName=”op1”, icon=icon("nom_icono",lib="font-awesome"))
   . . . )
```

Una vez elaborada la aplicación Shiny ha sido publicada, para que pueda ser ejecutada por cualquier persona que no tenga instalado R en su ordenador. RStudio ofrece a sus usuarios la oportunidad de alojar de forma gratuita sus aplicaciones en el dominio shinyapps.io o bien instalar Shiny Server en un servidor.

Gen-IO ha sido alojada en shinyapps.io y la ruta es :

https://pilarnatividad.shinyapps.io/genio/

Fig. 17 Pantalla inicio GEN-IO

6. Desarrollo del trabajo

6.1 Desarrollo del trabajo Fase1

Gen-IO es la aplicación web interactiva desarrollada en este TFM, usando Shiny Dashboard. Como ya se explicó en el capítulo 5, la aplicación consta de 2 componentes: la interfaz de usuario (*ui*) y el servidor (*server)*.

La interfaz de usuario se ha confeccionado con Shiny Dashboard:

```
ui <- dashboardPage(title="gen-IO", header, skin = "blue", sidebar, body)
```

Y se compone de:

- El título de la web que la aloja, una cabecera en la que aparece el logo de la aplicación:

```
header <- dashboardHeader(title =tags$img(src="gen-io-Logo.png", width="165px"),
                          titleWidth = 230)
```

- Un menú de opciones que el usuario podrá ejecutar:

```
sidebar<-  dashboardSidebar(
  sidebarMenu(id="sidebarmenu",
              menuItem("Busqueda en Pubmed",
                       tabName = "a",
                       icon= icon("download", lib="font-awesome")),
              menuItem("Explorar publicaciones",
                          tabName = "b",
                          icon= icon("globe", lib="font-awesome")),
                . . .
  ))
```

- Y el cuerpo, en el que se parametrizan los componentes de entrada y salida con las que el usuario podrá interactuar.

```
body <- dashboardBody(
        tabItem(
                tabName="a",
                h2("Búsqueda publicaciones en Pubmed"),
                br(),
                useShinyjs(),
                extendShinyjs(text=jsResetCode),
                textInput("busqueda", "Términos:", value="williams syndrome"),
                dateInput("date1", "Desde fecha publicación:",
                           value="2016-01-01", format="yyyy-mm-dd"),
                dateInput("date2", "Hasta fecha:", value="2018-12-31", format="yyyy-mm-dd"),
                actionBttn(inputId ="createPub",label="Visualizar consulta Pubmed",
                        color = "warning",style= "gradient",icon=icon("tv"),
                        block = FALSE, size = "md"),
                br(),br(),
                verbatimTextOutput("nText"),
                br(),
                actionBttn(inputId ="downPub", label="Buscar",color = "warning",
                         style= "gradient", icon=icon("download"),block = FALSE, size = "md"),
                hidden(actionBttn(inputId ="nuevaBusq",label="Nueva búsqueda", color = "warning",
                       style= "gradient", icon=icon("search"), block = FALSE, size = "md")),
                br(), br(),
                dataTableOutput("pmidResultado"),
                br(),
                wellPanel(htmlOutput("panelPublicacion"))
                ),
        . . .
        . . .
        )
```

Con estos datos se ejecuta la búsqueda, que devuelve uno o más ficheros dependiendo del número de publicaciones obtenidas. El código que aparece en el componente *server* para la búsqueda y descarga en Pubmed es el siguiente:

```
pubmedResult<-eventReactive( input$downPub,
            { . . .
             out.A<-batch_pubmed_download(pubmed_query_string =
             paste(c(input$busqueda, "[MH] AND " , as.character(input$date1,"%Y/%m/%d"),":",
             as.character(input$date2,"%Y/%m/%d"),"[DP]"), collapse=""),
             format = "abstract", batch_size = 1500, dest_file_prefix = "pubmed_")
              . . . }
```

Dependiendo del resultado de la búsqueda se pueden descargar uno o más ficheros, que deberán ser leídos para generar el corpus en la variable *pubmed_result.* Para ello, después de que el usuario haya pulsado el botón *Buscar*, se activará el objeto *eventReactive* y se concatenarán los ficheros descargados en un solo fichero, para luego ser leído con la función *readabs(),* creando el corpus en un objeto S4 de la clase Abstract:

```
pubmedResult<-eventReactive( input$downPub,
            {
            . . .
            file.create("pubmed_result.txt")
            ficheros<-out.A
            for(i in 1:length(ficheros)){ file.append("pubmed_result.txt",ficheros[i])}
            pmr<- readabs("pubmed_result.txt")
            . . .                                    })
            return(pmr) })
```

El corpus se muestra en forma de tabla conteniendo en cada una de sus filas las descripciones de las publicaciones filtradas desde Pubmed y el PMID de cada una de ellas. Para ello, el código se incluye en el objeto *output$pmidResultado* de tipo *renderDataTable*:

```
output$pmidResultado <- renderDataTable({
     corpus<- pubmedResult()
     pmidRes<- data.frame(corpus@PMID,corpus@Journal)
     colnames(pmidRes)<-c("PMID", "PUBLICACIONES")
     datatable(pmidRes,
               selection=list(mode='single', selected=1),
               options = list(language = list(url = 'Spanish.json'),
               pageLength = 5))
   })
```

Y para que sea visualizado en la aplicación en la interfaz de usuario *ui* se ha incluido el código:

```
tabItem(tabName="a",
        …
           dataTableOutput("pubmedResultado"),
           …
         )
```

Para mostrar el contenido del documento seleccionado en la tabla, ha sido necesario crear el objeto *output$panel_public*, a partir del corpus secundario generado con la función *subabs(),* pasándole como argumentos el corpus y la fila que ocupa la publicación seleccionada en la tabla.

```
  output$panelPublicacion = renderPrint({
     s = input$pmidResultado_rows_selected
     pmr<- pubmedResult()
     mostrar_info(s,pmr)
   })
```

Para darle formato al texto se ha creado la función *mostrar_info()*, que trocea la publicación por párrafos con la función *SentenceToken()*, y al primero de ellos se le aplica la función *tokenice_sentences()* para encontrar las frases que lo componen, de manera que se puede formatear el título.

```
mostrar_info = function (fila, corpus)
{
  s = fila
  pmr<- corpus
  corpusSecund<- subabs(pmr, s,s)
```

```
  frasesAbstract<-SentenceToken(corpusSecund@Abstract)
  frasesTitulo<-tokenize_sentences(frasesAbstract[1], lowercase= FALSE)
  cat(paste('<p>','<h4>','<font color=\"#4B04F6\"><b>',corpusSecund@Journal,
'</b></font>','</h4></p>'))
  for (i in (1:length(frasesTitulo))){
    if (i==1 || i==2)
        cat(paste('<p><h4>','<font color=\"#4B04F6\"><b>', frasesTitulo[i],'</b></font>','</h4>',
'\n','</p>'),fill = TRUE)
    else cat(paste('<p><i>',frasesTitulo[i],'</i></p>'), fill = TRUE)
  }
  cat('\n')
  for (i in (2:length(frasesAbstract))){
    cat(paste(' <p>',frasesAbstract[i], '</p>'), fill = TRUE)
  }
...
```

En la parte inferior del resumen de la publicación se incluye un enlace para navegar hacia la web de Pubmed, desde donde es posible acceder al documento completo.

```
...
  cat(paste("<a  href='https://www.ncbi.nlm.nih.gov/pubmed/",corpusSecund@PMID,"'  target=_blank
>", "Visualizar publicación en Pubmed","</a>"))
}
```

La minería de textos se usa para poder extraer información útil y relevante a partir de un corpus no estructurado. Y esto es lo que se pretende en las siguientes opciones de menú que muestra la aplicación. En un primer paso, se pueden *Explorar publicaciones*, para lo que se selecciona aquella que se quiere analizar desde una lista desplegable, y se muestra información referente a genes, enfermedades, mutaciones, elementos químicos y entidades, que ha sido posible encontrar en ella.

En la interfaz gráfica del usuario se incluye la lista desplegable con un control del tipo *uiOutput*, y las diferentes tablas que se activan, siempre y cuando exista información, usando *dataTableOutput*:

```
tabItem(tabName="b",
           h2("Explorar publicaciones Pubmed "),
           htmlOutput("tituloEnfBus"),
            br(),
            helpText("Seleccione una publicación para obtener la información sobre genes,
enfermedades, mutaciones, elementos químicos y entidades que aparece en la misma"),
            uiOutput("abs"),
            br(),br(),br(),
            tags$div(style='cursor:pointer',dataTableOutput("viewPubtatorGen")),
            tags$div(style='cursor:pointer',dataTableOutput("viewPubtatorDis")),
            tags$div(style='cursor:pointer',dataTableOutput("viewPubtatorMut")),
            tags$div(style='cursor:pointer',dataTableOutput("viewPubtatorChe")),
      tags$div(style='cursor:pointer',dataTableOutput("viewPubtatorEnt")),
      br(),br()
     ),
```

La lista desplegable dinámica *selectInput*, que muestra las publicaciones en la interfaz del usuario, se genera en *server* a través de un objeto reactivo de tipo *renderUI*.

```
output$abs<- renderUI({
     corpus<- pubmedResult()
     vectorAbs<- as.data.frame(corpus@Journal)
     colnames(vectorAbs)<-"Elija una Publicación..."
     selectInput("listaAbstracts","Publicación:", choices = vectorAbs, width='800px')})
```

Al elegir una publicación, de la lista desplegable, el objeto *eventReactive* ejecuta la función *pubtator_new_function()*, que busca toda la información relativa a genes, enfermedades, mutaciones, elementos químicos y entidades, que es posible encontrar en ella. En principio se utilizó la función *pubtator_function()* del paquete *pubmed.mineR*, pero la url que incluía dejo de funcionar mientras se desarrollaba la aplicación, por lo que partiendo de la misma función se cambió dicha url por otra que si que está operativa para acceder al servidor de Pubtator y recuperar la información. (Ramachandran)

```
pubtator_new_function = function (x)
{test =
getURL(paste("https://www.ncbi.nlm.nih.gov/research/bionlp/pubtator2/api/v1/publications/export/
pubtator?pmids=", x,sep = ""))
  testa = unlist(strsplit(test, "\n", fixed = T))
  table1 = NULL
  for (i in 3:length(testa)) {
    temps = unlist(strsplit(testa[i], "\t", fixed = T))
    if (length(temps) == 5) {
      temps = c(temps, "No Data")
    }
    table1 = rbind(table1, temps)
  }
  if (ncol(table1) == 6) {
    table2 = table1
    colnames(table2) = c("PMID", "Start", "End", "Term", "TermType", "TermID")
    gene = NULL
    disease = NULL
    mutation = NULL
    chemical = NULL
    species = NULL
    for (i in 1:length(table2[, 5])) {
      if (table2[i, 5] == "Gene")
        gene = c(gene, table2[i, 4])
      else if (table2[i, 5] == "Disease")
        disease = c(disease, table2[i, 4])
      else if (table2[i, 5] == "Mutation")
        mutation = c(mutation, table2[i, 4])
      else if (table2[i, 5] == "Chemical")
        chemical = c(chemical, table2[i, 4])
      else if (table2[i, 5] == "Species")
        species = c(species, table2[i, 4])
    }
    gene = union(gene, gene)
    disease = union(disease, disease)
    mutation = union(mutation, mutation)
    chemical = union(chemical, chemical)
    species = union(species, species)
    return(list(Genes = gene, Diseases = disease, Mutations = mutation,
                Chemicals = chemical, Species = species, PMID = x))
  }
  else return(" No Data ")
}

viewInfoPubtator<- eventReactive( c(input$listaAbstracts),{
     pmr <- pubmedResult()
     pmidPubtator<- pubtator_new_function(pmr@PMID[which(pmr@Journal==input$listaAbstracts)])
     return(pmidPubtator)
    })
```

Para generar las salidas por medio de tablas se usan objetos *renderDataTable* con la salida de la función *pubtator_new_function()*, que ha generado una lista con 6 componentes. Por ejemplo, al elegir el PMID=28751568 se puede obtener la siguiente información:

```
List of 6
 $ Genes    : chr [1:5] "mTOR" "Elastin" "ELN" "Eln" ...
 $ Diseases : chr [1:3] "Williams syndrome" "fibrosis" "arterial"
 $ Mutations: NULL
 $ Chemicals: chr [1:2] "Rapamycin" "rapamycin"
 $ Species  : chr [1:3] "Mice" "mice" "human"
 $ PMID     : num 28751568
```

A través del objeto *renderDataTable* se genera una tabla filtrando la primera componente de la lista *wPubtator*$*Genes* si se desea visualizar los genes que aparecen en la publicación elegida.

```
output$viewPubtatorGen<-renderDataTable({
     wPubtator<-viewopcion2()
     tabla<-data.frame(Genes=wPubtator$Genes)
     cap=tags$h3("Genes")
     datatable(tabla,
               caption=cap,
```

```
              selection=list(mode='single', selected=1),
              options = list(language = list(url = 'Spanish.json'),
                             pageLength = 5)
    })
```

Para mostrar las enfermedades se usa *wPubtator$Diseases*:

```
output$viewPubtatorDis< renderDataTable({
      wPubtator<-viewopcion2()
      tabla<- data.frame(Enfermedades=wPubtator$Disease)
      cap=tags$h3("Enfermedades")
      datatable(tabla,
                caption=cap,
                selection=list(mode='single', selected=1),
                options = list(language = list(url = 'Spanish.json'),
                               pageLength = 5))
    })
```

Y de forma similar se hace para las mutaciones (*wPubtator$Mutations*), los elementos químicos (*wPubtator$Chemicals*) y las entidades (*wPubtator$Species*).

Para implementar la búsqueda de la frecuencia de las palabras que aparecen en los textos, ha sido necesario programar en el server de la aplicación un objeto *reactive* que realiza el cálculo de frecuencias usando la función *word_atomizations():*

```
pubmedResultWords<- reactive({
                          pmr<- pubmedResult()
                          pmrWord<- word_atomizations(pmr)
                          row.names(pmrWord)<-NULL
                          return(pmrWord)})
```

Y generar la salida con *renderDataTable*:

```
    output$frecWords<-renderDataTable({
      tableWords<- pubmedResultWords ()
      colnames(tableWords)<-c("Palabras","Frecuencia")
      datatable(tableWords,
                selection=list(mode='single', selected=1),
                options = list(language = list(url = 'Spanish.json'),
                               pageLength = 20))
       })
```

El código en la interfaz de usuario incluye el *dataTableOutput*:

```
  tabItem(tabName="c",
            h2("Frecuencia de palabras"),
            htmlOutput("tituloFecPalabras"),
            br(),
            tags$div(style='cursor:pointer',dataTableOutput("frecWords")),
            br(),br()),
```

La función *gene_atomization* permite calcular la frecuencia de aparición de los genes en todas las publicaciones filtradas en la búsqueda realizada, a través de un objeto *reactive* en el *server*:

```
getFrecGene<- reactive({
      pmr<- pubmedResult()
      freqGenes <- gene_atomization(pmr)
      return(freqGenes)
    })
```

Y se genera una salida *renderDataTable*:

```
output$frecGene<-renderDataTable({
                     tabGene <-getFrecGene()
                     datatable(tabGene,
                         selection=list(mode='single', selected=1),
                         options = list(language = list(url = 'Spanish.json'),
                                        pageLength = 20))
    })
```

Que es mostrada en la interfaz de usuario:

```
tabItem(tabName="d",
            h2("Frecuencia de genes"),
            htmlOutput("tituloFecGenes"),
            br(),
            tags$div(style='cursor:pointer',dataTableOutput("frecGene")),
            br(),
            wellPanel(htmlOutput("urlGenSelec")),
            br(),br()),
```

Se ha incluido un enlace a un fichero de texto con información sobre el gen seleccionado en la tabla, que se obtiene del servidor Uniprot a través de la función *local_uniprotfun():*

```
output$urlGenSelec <- renderPrint({
      s = input$frecGene_rows_selected
      tabGene<-getFrecGene()
      gen<-tabGene[s,1]
      local_uniprotfun(gen)
      file.rename("x.txt","www/x.csv")
      cat(paste("<a href='../x.csv' target=_blank >
                <img src='Logo_uniprot.png' alt='Información de Uniprot'> ",gen,"</a>"))
```

Esta función descarga el fichero x.txt del servidor de Uniprot, y para que el fichero se pueda abrir en cualquier navegador es necesario renombrarlo con la extensión csv.

Para mostrar de forma visual qué palabras y qué genes se repiten con más frecuencia se ha implementado la opción de menú *Nube de Palabras y Genes*. Se crea un objeto *reactive* que responde ante la selección del usuario sobre qué gráfico mostrar, y devuelve una tabla con los datos que aparecen en el gráfico:

```
terms <- reactive({
      c(input$update,input$selection)
      isolate({
          if (input$selection=="Palabras") {twordcloud<-as.data.frame(pubmedResultWords ())}
          if (input$selection=="Genes") {twordcloud<-as.data.frame(getFrecGene())}
      })
      return(twordcloud)
      })
```

Para representar el gráfico en *server* se genera un *renderPlot*:

```
    output$plot <- renderPlot({
      v <- terms()
      if (input$selection=="Palabras") {wordcloud_rep(v$words, v$Freq, scale=c(4,0.5),
                                                      min.freq = input$freq, max.words=input$maxw,
                                                      colors=brewer.pal(8, "Dark2"))}
      if (input$selection=="Genes") {
                                       wordcloud_rep(v$Gene_symbol,as.numeric(v$Freq),
                                                     scale=c(4,0.5),
                                                     min.freq = input$freq, max.words=input$maxw,
                                                     colors=brewer.pal(8, "Dark2"))}
    })
```

Se muestra en la interfaz del usuario, en la que aparecen una lista desplegable con las opciones Palabras y Genes, y dos controles para indicar la frecuencia y el máximo número de palabras o genes que se presentarán:

```
tabItem(tabName="e",
          h2("Nube de Palabras y Genes"),
          br(),
          sidebarLayout(
            sidebarPanel(
              selectInput("selection", "Elije:", choices = c("Palabras","Genes")),
              actionButton("update", "Dibuja",class = "btn-primary"),
              hr(),
              sliderInput("freq","Mínima frecuencia:", min = 1, max = 50, value = 3),
              sliderInput("maxw","Máximo núm. Palabras/genes:", min = 1, max = 300, value = 100)),
              mainPanel(plotOutput("plot")))),
```

En la sección *Filtrar public. por genes* se permite seleccionar un gen y se buscan aquellos artículos en los que se habla de ese gen, creándose un corpus secundario sobre el que se analizan, los genes, enfermedades, mutaciones, elementos químicos y entidades que comparten las mismas publicaciones. No significa que estén relacionados con el gen seleccionado, tan sólo se presenta la información relevante que se expone en las publicaciones en las que aparece éste. Para ello, será necesario generar de forma dinámica una lista desplegable con los genes que aparecen en las publicaciones filtradas en la búsqueda en Pubmed, usando el objeto reactivo *getFrecGene()* que se usa para generar la frecuencia de los genes y filtrando tan sólo la descripción del gen:

```
getFrecGene<- reactive({
      pmr<- pubmedResult()
      freqGenes<- gene_atomization(pmr)
      return(freqGenes)
    })

getTablaGenes <- reactive({
                     freqGenes <- getFrecGene()
                     tablaGenes <- reqGenes[,2]
                     return(tablaGenes)
                      })

output$dataframeGenes <- renderUI({
      vectorGenes<- getTablaGenes ()
      selectInput("genes","Genes:", choices = vectorGenes, label="Genes:", width='350px')

    })
```

Con el gen seleccionado se crea un corpus secundario, que comprende las publicaciones en las que se habla del mismo, y para ello se ha implementado un objeto *eventReactive*, que se activa al seleccionar un gen de la lista desplegable de genes:

```
corpusDeUnGen<- eventReactive (c(input$genes), {
        pmr<- pubmedResult()
        genElegido<-input$genes
        tablaGenes<-getFrecGene()
        v<-(which(tablaGenes[,2]==genElegido))
        vecGen <-c()
        vecGen <- c(vecGen,tablaGenes[v,])
        vecGen <-unlist(vecGen[1:2], use.names=FALSE)
        vecGen<-as.vector(vecGen)
        corpusGen<- searchabsL(pmr, include=vecGen)
        return(corpusGen)
    })
```

Se muestra la tabla de documentos en los que se cita al gen seleccionado usando el objeto *renderDataTable*:

```
output$pmidCorpSecGen <- renderDataTable({
      corpusSecundario<- corpusDeUnGen()
      pmidCorpusSec<- data.frame(corpusSecundario@PMID,corpusSecundario@Journal)
      colnames(pmidCorpusSec)<-c("PMID", "PUBLICACIONES")
      datatable(pmidCorpusSec,
                selection=list(mode='single', selected=1),
                options = list(language = list(url = 'Spanish.json'),
                               pageLength = 5))
    })
```

El documento seleccionado se visualiza mediante el objeto renderPrint:

```
output$panelPublicacionGen = renderPrint({

      s = input$pmidCorpSecGen_rows_selected
      pmr<- corpusDeUnGen()
      mostrar_info(s,pmr)
    })
```

Y a partir de esta información se realiza una búsqueda similar a la que se hizo en *Explorar publicaciones*, ejecutándose la *pubtator_new_function()* en un objeto *eventReactive*, a partir del gen que el usuario haya seleccionado.

```
infoDeUnGen<- eventReactive (c(input$genes), {
      genFiltro<- corpusDeUnGen()
      genPmid<-genFiltro@PMID
      genSelInfo<- pubtator_new_function(genPmid)
      return(genSelInfo)
    })
```

Con la información obtenida en *infoDeUnGen(),* se filtran los datos de salida que tienen forma de tabla con objetos *renderDataTable*. Por ejemplo, para mostrar los genes que aparecen en el corpus secundario:

```
output$datosGenes<-renderDataTable({
      wPubtator<- infoDeUnGen()
      tabla<-data.frame(Genes=wPubtator$Genes)
      cap=tags$h3("Genes")
      datatable(tabla,
                caption=cap,
                #selection=list(mode='single', selected=1),
                options = list(language = list(url = 'Spanish.json'), pageLength = 10))
    })
```

De igual modo se procede para generar las salidas de enfermedades, mutaciones, elementos químicos y entidades que aparecen en el corpus secundario:

```
   output$datosDisease<-renderDataTable({
      wPubtator<- infoDeUnGen()
      tabla<- data.frame(Enfermedades=wPubtator$Disease)
      cap=tags$h3("Enfermedades")
      datatable(tabla,
                caption=cap,
                #selection=list(mode='single', selected=1),
                options = list(language = list(url = 'Spanish.json'), pageLength = 20))
    })

    output$datosMutation<-renderDataTable({
      wPubtator<- infoDeUnGen()
      tabla<-  data.frame(Mutaciones=wPubtator$Mutations)
      cap=tags$h3("Mutaciones")
      datatable(tabla,
                caption=cap,
                #selection=list(mode='single', selected=1),
                options = list(language = list(url = 'Spanish.json'),pageLength = 20))
    })

    output$datosChemicals<-renderDataTable({
      wPubtator<- infoDeUnGen()
      tabla<- data.frame(Elementos=wPubtator$Chemicals)
      cap=tags$h3("Elementos químicos")
      datatable(tabla,
                caption=cap,
                #selection=list(mode='single', selected=1),
                options = list(language = list(url = 'Spanish.json'),pageLength = 20))
    })

    output$datosEnti<-renderDataTable({
      wPubtator<- infoDeUnGen()
      tabla<- data.frame(Entidades=wPubtator$Species)
      cap=tags$h3("Entidades")
      datatable(tabla,
                caption=cap,
                #selection=list(mode='single', selected=1),
                options = list(language = list(url = 'Spanish.json'), pageLength = 20))
    })
```

Para mostrar toda esta información en la interfaz de salida se incluye el siguiente código:

```
tabItem(tabName="g",
        h2("Filtrar public. por genes"),
        htmlOutput("tituloFiltroGenes"),
        br(),
        uiOutput("dataframeGenes"),
```

```
        br(),
        wellPanel(htmlOutput("mostrarIdGen")),
        br(),
        tags$div(style='cursor:pointer',dataTableOutput("pmidCorpSecGen")),
        br(),
        wellPanel(htmlOutput("panelPublicacionGen")),
        tags$div(style='cursor:pointer',dataTableOutput("datosGenes")),
        tags$div(style='cursor:pointer',dataTableOutput("datosDisease")),
        tags$div(style='cursor:pointer',dataTableOutput("datosMutation")),
        tags$div(style='cursor:pointer',dataTableOutput("datosChemicals")),
        tags$div(style='cursor:pointer',dataTableOutput("datosEnti")),
        br(),br()
        ),
```

6.2 Desarrollo del trabajo Fase2

En la segunda fase se ha desarrollado una nueva opción de menú *Filtrar publicaciones por término*, para encontrar el corpus secundario de todas publicaciones que contienen los términos que indiquemos.

Ha sido necesario crear un objeto *eventReactive* que ejecuta, a partir de los términos introducidos en el objeto *input$nuevoCorpus*, la función *searchabsL() y* encuentra aquellos documentos que los contienen.

```
pmedCorSec <-eventReactive( input$nuevoCorpus,
            { pmr<- pubmedResult()
              corpusSecundario<- searchabsL(pmr, include=input$busquedaEnCorpus)
              return(corpusSecundario)})
```

A continuación, se crea el objeto de salida *renderDataTable*, que presenta la información en formato tabla, por la que el usuario podrá posteriormente navegar para obtener el contenido de la publicación en un objeto *renderPrint*, de la misma forma que se presentó en la búsqueda del corpus inicial.

```
output$pmidCorpSec <- renderDataTable({
     corpusSecundario<-pmedCorSec()
     pmidCorpusSec<- data.frame(corpusSecundario@PMID,corpusSecundario@Journal)
     colnames(pmidCorpusSec)<-c("PMID", "PUBLICACIONES")
     datatable(pmidCorpusSec,
               selection=list(mode='single', selected=1),
               options = list(language = list(url = 'Spanish.json'), pageLength = 5))
   })

output$panelPublicacion2 = renderPrint({
     s = input$pmidCorpSec_rows_selected
     pmr<- pmedCorSec()
     mostrar_info(s,pmr)
   })
```

La visualización de esta información en la interfaz de salida se realiza incluyendo el siguiente código:

```
tabItem(tabName="f",
        h2("Filtrar publicaciones que contienen un término"),
        htmlOutput("tituloFiltroTerm"),
        br(),
        useShinyjs(),
        extendShinyjs(text=jsResetCode),
        textInput("busquedaEnCorpus", "Términos:", value="mental retardation"),
        actionBttn(inputId ="btnNuevoCorpus",
                       label="Filtrar publicaciones",
                       color = "primary",
                       style= "gradient",
                       icon=icon("download"),
                       block = FALSE,
                       size = "md"),
        br(),br(),
        tags$div(style='cursor:pointer',dataTableOutput("pmidCorpSec")),
        br(),
```

```
        wellPanel(htmlOutput("panelPublicacion2")),
        br(),br()
    ),
```

En esta segunda fase del proyecto se pretende usar la aplicación web Gen-IO para encontrar las relaciones de similitud entre los genes relacionados con el síndrome de Williams-Beuren y las enfermedades diagnosticadas en el mismo. Para ello, se ha implementado el apartado *Relaciones de Similitud*, en el que se extraen todas las enfermedades relacionadas con la búsqueda que se realizó en Pubmed, y se muestran en una tabla, en la que el usuario puede navegar, incluso filtrar desde el control de *Buscar* escribiendo un término.

El proceso seguido, para encontrar las enfermedades relacionadas con la enfermedad analizada, ha sido crear un corpus formado por los artículos relacionados con ella, de los dos últimos años, desde la fecha en la que se ejecute el proceso. Se activará al pulsar el botón *Buscar Enfermedades*:

```
pubmedEnfer  <-eventReactive( input$downPub2,{
        out.A<-batch_pubmed_download(pubmed_query_string =
                paste(c(input$busqueda,"[Mesh] AND ",as.character(Sys.Date()-730,"%Y/%m/%d"),
                        ":", as.character(Sys.Date(),"%Y/%m/%d"),"[DP]"), collapse=""),
                 format = "abstract", batch_size = 1500, dest_file_prefix = "pubmed_enfer_")
        file.create("corpusEnf.txt")
        ficheros<-out.A
        for(i in 1:length(ficheros)){file.append("corpusEnf.txt",ficheros[i])}
        pmr<- readabs("corpusEnf.txt")
        return(pmr)})
```

Y luego de forma reactiva se ejecuta la función *pubtator_disease(),* que se ha implementado a partir de la función *pubtator_function(),* pero cambiando la url de consulta y eliminando parte de código no necesario:

```
pubtator_disease = function (x)
{
  test = getURL(
paste("https://www.ncbi.nlm.nih.gov/research/bionlp/pubtator2/api/v1/publications/export/pubtato
r?pmids=", x,sep = ""))
  testa = unlist(strsplit(test, "\n", fixed = T))
  table1 = NULL
  for (i in 3:length(testa)) {
    temps = unlist(strsplit(testa[i], "\t", fixed = T))
    if (length(temps) == 5) {
      temps = c(temps, "No Data")
    }
    table1 = rbind(table1, temps)
  }
  if (ncol(table1) == 6) {
    table2 = table1
    colnames(table2) = c("PMID", "Start", "End", "Term", "TermType", "TermID")
    disease = NULL
    for (i in 1:length(table2[, 5])) {
      if (table2[i, 5] == "Disease") disease = c(disease, table2[i, 4])
    }
    disease = union(disease, disease)
    return(list( Diseases = disease))
  }
  else return(" No Data ")
}
```

Esta función se ejecuta tantas veces como documentos existen en el corpus descargado, y consulta el servidor de Pubtator para descargar la información requerida, de ahí que se hayan restringido el número de documentos, para no saturar el servidor. Las enfermedades encontradas en cada documento se almacenan en un dataframe usando la función *rbind.fill()* del paquete *plyr*, que permite concatenar filas de dos dataframes. Seguidamente se convierten los textos a minúsculas, se eliminan los valores NA y duplicidades existentes:

```
tableDiseases <- reactive({
      pmr<- pubmedEnfer()
```

```
    pubEnf<-data.frame(Diseases=character())
    for (i in 1:length(pmr@PMID)){
      pub<-pubtator_disease(pmr@PMID[i][1])
      if (pub!='NULL') {
        pubEnf<-rbind.fill(pubEnf,data.frame(pub[1]))
      }
    }
    pubEnf$Diseases<-tolower(pubEnf$Diseases)
    pubEnf<-pubEnf[!is.na(pubEnf$Diseases),]
    pubEnf<-unique(pubEnf)
    tablaEnfer<-data.frame(pubEnf)
    colnames(tablaEnfer)<-c("Enfermedades")
    return(tablaEnfer)
  })
```

Por último, se usa un objeto *renderDataTable*, para mostrar la tabla de enfermedades al usuario:

```
output$listadoDisease2<-renderDataTable({
   tableenf<-tableDiseases()
   datatable(tableenf,
             selection=list(mode='single', selected=1),
             options = list(language = list(url = 'Spanish.json'),
                            pageLength =5))
  })
```

Como ya se discutió en el capítulo 3, el método usado para analizar el texto y extraer relaciones de similitud entre palabras será el Análisis Semántico Latente(LSA), que realiza una serie de cálculos algebraicos para obtener una representación de las palabras en un espacio multidimensional, usando la técnica de Descomposición en Valores Singulares (SVD) y obteniendo los componentes principales de la representación del texto (SLOMOVITZ, 2016). De esta forma se logra reducir la dimensionalidad de la matriz de términos y consecuentemente extraer su espacio semántico.

Para ello, el primer paso es crear el corpus secundario a partir de la enfermedad que se quiere analizar:

```
getRelLSA<- eventReactive (c(input$btnRelEnfGen), {
      pmr<- pubmedResult()
      tableEnf<-tableDiseases()
      filaElegida<-input$listadoDisease2_rows_selected
      enfElegida<- as.vector(tableEnf[filaElegida,])
      corpusSecundario<- searchabsL(pmr, include=enfElegida)
...
```

Por ejemplo, para la enfermedad hipercalcemia el corpus secundario está formado por los siguientes abstracts:

```
Slot "PMID":
 [1] 30300432 29226900 29193944 28259930 28085672 27938597 27574996 26271350 25691196 26958139 25332293 25029575 24985599 24953591
[15] 24885071 24331936 24057591 23581265 22608712 22570974 22566418 21980823 21528818 20437059 20146355 19764675 19097873 18662176
[29] 18350724 17471494 17009066 16817378 16506136 16406078 16380272 16108410 15774842 15730896 15466114 14751286 13677282 12792585
[43] 12687671 12432430 12041723 11928452 11314625 11180224 11075265 10872191 10502124 10440826 10326175 10232742  9926515  9860302
[57]  9850771  9782077  9637430  9558923  9071842  9003495  8812460  8862624  8620979
```

Fig. 18 Listado de PMIDs que contienen enfermedad hypercalcemia

El segundo paso es obtener los genes que aparecen en este corpus secundario:

```
...
      tablaGenes<-as.data.frame(gene_atomization(corpusSecundario))
...
```

Se controla si no aparece ningún gen, para mostrar un mensaje

```
...
      if (nrow(tablaGenes)==0){
```

```
          tabla<-data.frame(Genes=c( de aviso:"No se han encontrado genes asociados en
las publicaciones analizadas"))
        }
...
```

Gene_symbol <fctr>	Genes <fctr>	Freq <fctr>
ELN	elastin	12
LIMK1	LIM domain kinase 1	7
TRPC3	transient receptor potential cation channel, subfamily C, member 3	6
CLIP2	CAP-GLY domain containing linker protein 2	4
FKBP6	FK506 binding protein 6, 36kDa	4
BCL7B	B-cell CLL/lymphoma 7B	2
GTF2I	general transcription factor IIi	2
GTF2IRD1	GTF2I repeat domain containing 1	2
RFC2	replication factor C (activator 1) 2, 40kDa	2
YWHAG	tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, gamma polypeptide	2
AUTS2	autism susceptibility candidate 2	1
C3	complement component 3	1
CALN1	calneuron 1	1
HIP1	huntingtin interacting protein 1	1
HR	hair growth associated	1
SI	sucrase-isomaltase (alpha-glucosidase)	1

Fig. 19 Listado de genes que aparecen en el corpus secundario creado a partir de hypercalcemia

El tercer paso es crear la matriz de términos de documentos, usando la función *tdm_for_lsa(),* matriz cuyas columnas son los documentos de que consta el corpus secundario, obtenido a partir de la enfermedad elegida en la tabla mostrada, y los genes obtenidos con la función *gene_atomization()* sobre el corpus secundario obtenido de los dos últimos años.

```
...
       genesYenfer<-as.vector(c(as.vector(tablaGenes[,1]),enfElegida))
       tdmCorp<- tdm_for_lsa( corpusSecundario,genesYenfer)
...
```

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]	[,8]	[,9]	[,10]	[,11]	[,12]	[,13]	[,14]	[,15]	[,16]	[,17]	[,18]	[,19]	[,20]
ELN	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0
LIMK1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
TRPC3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CLIP2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0
FKBP6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
BCL7B	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
GTF2I	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0
GTF2IRD1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0
RFC2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
YWHAG	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
AUTS2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
CALN1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
HIP1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
HR	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SI	0	0	0	0	1	0	2	0	0	0	1	1	0	0	1	0	0	0	0	0
hypercalcemia	5	1	1	1	1	1	10	1	1	1	7	1	1	1	1	1	1	5	1	4

Fig. 20 Una muestra de la Matriz de términos de documentos obtenida para hypercalcemia

Si no se encuentran genes con los que buscar relaciones de similitud se muestra un aviso:

```
...
if (ncol(tdmCorp )<=1){
          tabla<-data.frame(Genes=c("No se han encontrado genes asociados en las
publicaciones analizadas"))
        }else{
...
```

El cuarto paso es aplicar el algoritmo LSA, que se ejecuta si la matriz de términos de documentos tiene más de una fila, con la función *lsa()* pasando como argumento la matriz de términos de documentos y el número de valores singulares óptimo, generado

mediante la función *dimcalc_share()*, de forma que se obtiene un nuevo espacio vectorial LSA, cuyas dimensiones son el número de valores singulares y el número de palabras analizadas, consiguiendo reducir la dimensionalidad de la matriz original.

```
...
        }else{
          lsaCorpus <- lsa( tdmCorp, dims=dimcalc_share())
...
$tk
                    [,1]         [,2]         [,3]
ELN          -0.056243746 -0.036910209 -0.853272881
LIMK1        -0.025111983 -0.018133726 -0.502756395
TRPC3        -0.092726628  0.653221412 -0.018480412
CLIP2        -0.015407400  0.031705688 -0.010270158
FKBP6        -0.023282535 -0.014328997  0.062681143
BCL7B        -0.004943114 -0.002009414  0.001836286
GTF2I        -0.012325920  0.025364550 -0.008216126
GTF2IRD1     -0.006162960  0.012682275 -0.004108063
RFC2         -0.008999442  0.002640753 -0.094941777
YWHAG        -0.007566774 -0.003201546  0.003121906
AUTS2        -0.002522258 -0.001067182  0.001040635
C3           -0.107398936  0.741215226 -0.022844495
CALN1        -0.002522258 -0.001067182  0.001040635
HIP1         -0.002522258 -0.001067182  0.001040635
HR           -0.004918397 -0.001980430  0.001783794
SI           -0.085756767 -0.036284184  0.035381597
hypercalcemia -0.983638001 -0.137297734  0.062400891

$dk
              [,1]         [,2]         [,3]
 [1,] -0.24530217 -0.05788374  0.036779215
 [2,] -0.04906043 -0.01157675  0.007355843
 [3,] -0.04930575 -0.01174373  0.007566117
 [4,] -0.04906043 -0.01157675  0.007355843
 [5,] -0.05333768 -0.01463618  0.011526640
 [6,] -0.04906043 -0.01157675  0.007355843
 [7,] -0.49915884 -0.12188634  0.081900024
 [8,] -0.04906043 -0.01157675  0.007355843
 [9,] -0.04906043 -0.01157675  0.007355843
[10,] -0.05467092 -0.01780118 -0.193812479
[11,] -0.34770029 -0.08409667  0.055661698
[12,] -0.05333768 -0.01463618  0.011526640
[13,] -0.04906043 -0.01157675  0.007355843
[14,] -0.04906043 -0.01157675  0.007355843
[15,] -0.05333768 -0.01463618  0.011526640
[16,] -0.04906043 -0.01157675  0.007355843
[17,] -0.04906043 -0.01157675  0.007355843
[18,] -0.24530217 -0.05788374  0.036779215
[19,] -0.06178219  0.07520452 -0.017424706
[20,] -0.19624174 -0.04630699  0.029423372
[21,] -0.23239051  0.96838313 -0.019596568
[22,] -0.04906043 -0.01157675  0.007355843
[23,] -0.34342304 -0.08103724  0.051490900
[24,] -0.04906043 -0.01157675  0.007355843
[25,] -0.05057005 -0.01265655  0.008827889
[26,] -0.05186568 -0.01468896 -0.093228318
[27,] -0.04906043 -0.01157675  0.007355843
[28,] -0.14718130 -0.03473024  0.022067529
[29,] -0.09812087 -0.02315350  0.014711686
[30,] -0.04906043 -0.01157675  0.007355843
[31,] -0.05333768 -0.01463618  0.011526640
[32,] -0.04906043 -0.01157675  0.007355843
[33,] -0.05333768 -0.01463618  0.011526640
[34,] -0.04906043 -0.01157675  0.007355843
[35,] -0.04906043 -0.01157675  0.007355843
[36,] -0.04906043 -0.01157675  0.007355843
[37,] -0.04906043 -0.01157675  0.007355843
[38,] -0.04906043 -0.01157675  0.007355843
[39,] -0.24530217 -0.05788374  0.036779215
[40,] -0.04906043 -0.01157675  0.007355843
[41,] -0.09812087 -0.02315350  0.014711686
[42,] -0.04930575 -0.01174373  0.007566117
[43,] -0.05747616 -0.02091340 -0.294396640
[44,] -0.14718130 -0.03473024  0.022067529
[45,] -0.04906043 -0.01157675  0.007355843
[46,] -0.04906043 -0.01157675  0.007355843
[47,] -0.04906043 -0.01157675  0.007355843
[48,] -0.04906043 -0.01157675  0.007355843
[49,] -0.05333768 -0.01463618  0.011526640
[50,] -0.04906043 -0.01157675  0.007355843
[51,] -0.19624174 -0.04630699  0.029423372
[52,] -0.05311817 -0.01621797 -0.152493450
[53,] -0.04906043 -0.01157675  0.007355843
[54,] -0.04906043 -0.01157675  0.007355843
[55,] -0.04906043 -0.01157675  0.007355843
[56,] -0.04955352 -0.01191561  0.007788767
[57,] -0.04906043 -0.01157675  0.007355843
[58,] -0.05835045 -0.02124235  0.066466876
[59,] -0.06809664 -0.03325387 -0.732625490
[60,] -0.04906043 -0.01157675  0.007355843
[61,] -0.05467092 -0.01780118 -0.193812479
[62,] -0.05747616 -0.02091340 -0.294396640
[63,] -0.05932613 -0.02194288 -0.393991428
[64,] -0.05333768 -0.01463618  0.011526640
[65,] -0.04906043 -0.01157675  0.007355843

$sk
[1] 20.049517 11.859784  8.483173

attr(,"class")
[1] "LSAspace"
```

Fig. 21 Ejecución lsa() para hypercalcemia.

Con el algoritmop *lsa()* se han obtenido tres matrices: la matriz de vectores singulares izquierda (*lsaCorpus$tk*), la matriz de vectores singulares derecha(*lsaCorpus$dk*) y la matriz diagonal (*lsaCorpus$sk*). Y el quinto paso, consiste

en multiplicar estas matrices, para obtener la matriz de términos de documentos reducida:

```
...
        }else{
          lsaCorpus <- lsa( tdmCorp, dims=dimcalc_share())
          matrizGenEnf<-as.textmatrix(lsaCorpus)
...
```

	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D32	D33	D34	D35	D36	D37	D38	D39	D40	D41	D56	D57
1. ELN	0.04	0.01	0.01	0.01	-0.02	0.01	0.02	0.01	0.01	1.47	0.01	-0.02	0.01	0.01	0.01	0.01	0.01	0.04	0.01	0.01	0.00	0.01
2. LIMK1	-0.02	0.00	0.00	0.00	-0.02	0.00	-0.07	0.00	0.00	0.86	0.00	-0.02	0.00	0.00	0.00	0.00	0.00	-0.02	0.00	-0.01	-0.01	0.00
3. TRPC3	0.00	0.00	0.00	0.00	-0.02	0.00	-0.03	0.00	0.00	-0.01	0.00	-0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
4. CLIP2	0.05	0.01	0.01	0.01	0.01	0.01	0.10	0.01	0.01	0.03	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.05	0.01	0.02	0.01	0.01
5. FKBP6	0.14	0.03	0.03	0.03	0.03	0.03	0.30	0.03	0.03	-0.07	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.14	0.03	0.06	0.03	0.03
6. BCL7B	0.03	0.01	0.01	0.01	0.01	0.01	0.05	0.01	0.01	0.00	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.03	0.01	0.01	0.01	0.01
7. GTF2I	0.04	0.01	0.01	0.01	0.01	0.01	0.08	0.01	0.01	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.04	0.01	0.02	0.01	0.01
8. GTF2IRD1	0.02	0.00	0.00	0.00	0.00	0.00	0.04	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.01	0.00	0.00
9. RFC2	0.01	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.17	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.01	0.00	0.00
10. YWHAG	0.04	0.01	0.01	0.01	0.01	0.01	0.08	0.01	0.01	0.00	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.04	0.01	0.02	0.01	0.01
11. AUTS2	0.01	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.01	0.00	0.00
12. C3	0.01	0.00	0.00	0.00	-0.02	0.00	-0.01	0.00	0.00	0.00	0.00	-0.02	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00
13. CALN1	0.01	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.01	0.00	0.00
14. HIP1	0.01	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.01	0.00	0.00
15. HR	0.03	0.01	0.01	0.01	0.01	0.01	0.05	0.01	0.01	0.00	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.03	0.01	0.01	0.01	0.01
16. SI	0.46	0.09	0.09	0.09	0.10	0.09	0.94	0.09	0.09	0.04	0.09	0.10	0.09	0.09	0.09	0.09	0.09	0.46	0.09	0.18	0.09	0.09
17. hypercalcemia	4.95	0.99	1.00	0.99	1.08	0.99	10.09	0.99	0.99	1.00	0.99	1.08	0.99	0.99	0.99	0.99	0.99	4.95	0.99	1.98	1.00	0.99

	D58	D59	D60	D61	D62	D63	D64	D65
1. ELN	-0.41	5.39	0.01	1.47	2.20	2.93	-0.02	0.01
2. LIMK1	-0.25	3.17	0.00	0.86	1.29	1.71	-0.02	0.00
3. TRPC3	-0.07	-0.02	0.00	-0.01	-0.01	0.00	-0.02	0.00
4. CLIP2	0.00	0.07	0.01	0.03	0.04	0.04	0.01	0.01
5. FKBP6	0.07	-0.35	0.03	-0.07	-0.13	-0.18	0.03	0.03
6. BCL7B	0.01	0.00	0.01	0.00	0.00	0.00	0.01	0.01
7. GTF2I	0.00	0.06	0.01	0.02	0.03	0.04	0.01	0.01
8. GTF2IRD1	0.00	0.03	0.00	0.01	0.01	0.02	0.00	0.00
9. RFC2	-0.04	0.60	0.00	0.17	0.25	0.33	0.00	0.00
10. YWHAG	0.01	-0.01	0.01	0.00	0.00	0.00	0.01	0.01
11. AUTS2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
12. C3	-0.07	0.00	0.00	0.00	0.00	0.01	-0.02	0.00
13. CALN1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
14. HIP1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
15. HR	0.01	0.00	0.01	0.00	0.00	0.00	0.01	0.01
16. SI	0.13	-0.09	0.09	0.04	0.02	-0.01	0.10	0.09
17. hypercalcemia	1.22	1.01	0.99	1.00	1.01	1.00	1.08	0.99

Fig. 22 Matriz de términos reducida

Sexto paso, mediante la función *associate()* se consiguen las asociaciones existentes entre los términos a partir de la matriz de términos reducida. No hace falta ejecutar la función para todos los términos, enfermedad seleccionada y genes obtenidos, ya que sólo nos interesa conocer la asociación existente entre la enfermedad y los genes:

```
...
        associatedWords <- associate(matrizGenEnf, enfElegida ,
                              measure = "cosine", threshold = "0.7")
...
```

HR	BCL7B	YWHAG	SI	AUTS2	CALN1	HIP1
0.9813188	0.9805754	0.9770327	0.9770327	0.9770327	0.9770327	0.9770327

Para algunas de las enfermedades obtenidas no se encuentran genes asociados, por lo que es necesario filtrar estos casos y devolver un mensaje que indique esta situación. En caso contrario se genera una tabla con tres columnas: la enfermedad, los genes relacionados con la misma, y el peso de asociación obtenido.

```
...
        if (is.null(associatedWords)){
            tabla<-data.frame(Genes=c("No se han encontrado genes asociados en las
publicaciones analizadas"))
          }else{
            tablaRelLSA<- data.frame(peso=associatedWords)
            col1<-rownames(tablaRelLSA)
            col0<-rep(enfElegida,length(col1))
            col2<-tablaRelLSA[,1]
            tabla<- data.frame(Enfermedad=character(),Gene=character(),Peso=double())
            tabla<-data.frame(Enfermedad=col0,Gene=col1,Peso=col2)
            rownames(tabla)<- NULL
```

```
            colnames(tabla)<-c("Enfermedad","Gene","Peso")
            tabla[,3]<-round(tabla$Peso,6)
            tablaSinCeros = tabla[tabla$Peso!=0,]
            tablaSinCeros[,3]<-round(tablaSinCeros[,3],6)
            tabla<-tablaSinCeros}}}
        return(tabla)
      })      })
```

Para mostrar la tabla obtenida se usará un objeto *renderDataTable*:

```
output$relacionesLSA <- renderDataTable({
          tabla<-getRelLSA()
          tabla
      })
```

En la interfaz gráfica del usuario se incluye la tabla de enfermedades con un control del tipo *dataTableOutput*, un botón con el control *actionBttn* para ejecutar la búsqueda de los genes relacionados, que se muestran con otro objeto *dataTableOutput*:

```
tabItem(tabName="h",
            h2("Relación de Similitud aplicando Análisis de la Semántica Latente"),
            htmlOutput("tituloRelSim"),
            br(),
            h4("Seleccione un enfermedad de la tabla y pulse el botón para
                obtener información de los genes relacionados:"),
            br(),
            actionBttn(inputId ="downPub2",
                       label="Buscar Enfermedades",
                       color = "primary",
                       style= "gradient",
                       icon=icon("download"),
                       block = FALSE,
                       size = "md"),
            tags$div(style='cursor:pointer',dataTableOutput("listadoDisease2")),
            hidden( actionBttn(inputId ="btnRelEnfGen",
                       label="Obtener Genes relacionados",
                       color = "primary",
                       style= "gradient",
                       icon=icon("sliders"),
                       block = FALSE,
                       size = "md")),
            br(),br(),
           tags$div(style='cursor:pointer',dataTableOutput("relLSA")),
           br(),br()
    ),
```

Por último, la aplicación calcula las relaciones de similitud usando la medida coseno, de forma que los datos se consideran vectores, cuyas componentes son los documentos, y el valor de la componente las ocurrencias del término en los documentos, tal que si el ángulo que los separa es de cero grados, su similaridad será de 1, y si el ángulo es de noventa grados, la similaridad será de 0 (Guzmán).

$$SimCoseno(gen, enfermedad) = \frac{\sum_{i=1}^{m} g_i \cdot e_i}{\sqrt{\sum_{i=1}^{m} g_i^2 \sum_{i=1}^{m} e_i^2}}$$

Para su cálculo se usa la función *cos_sim_calc(),* que crea un fichero de texto llamado *cossimdata.txt*, con la distancia ya calculada.

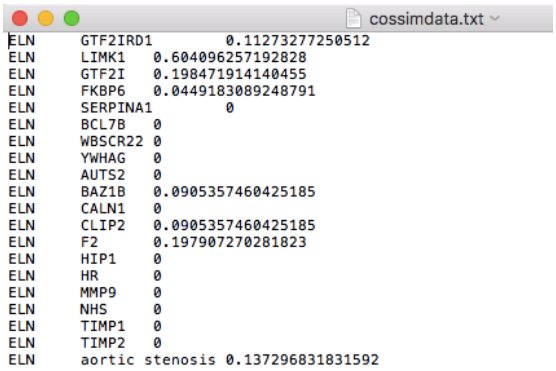


Fig. 23 Fichero cossimdata.txt, resultado de ejecutar la función *cos_sim_calc()*

Este fichero es leído por la aplicación GEN-IO para cargarlo en un dataframe:

```
getRelLSAGraf<- eventReactive (c(input$btnGraf), {
        pmr<- pubmedResult()
        tableEnf<-tableDiseases()
        filaElegida<-input$listadoDisease2_rows_selected
        enfElegida<- as.vector(tableEnf[filaElegida,])
        corpusSecundario<- searchabsL(pmr, include=enfElegida)
        tablaGenes<-as.data.frame(gene_atomization(corpusSecundario))
        if (nrow(tablaGenes)==0){
          tabla<-data.frame(Genes=c("No gráfico"))
        }else{
          genesYenfer<-as.vector(c(as.vector(tablaGenes[,1]),enfElegida))
          tdmCorp<- tdm_for_lsa( corpusSecundario,genesYenfer)
          if (ncol(tdmCorp )<=1){
            tabla<-data.frame(Genes=c("No gráfico"))
          }else{
            file.remove("cossimdata.txt")
            cos_sim_calc(tdmCorp)
            cos=read.delim("cossimdata.txt", header=FALSE, sep="\t")
            cos=na.omit(cos)
            relations <- data.frame(Desde=cos[,1], Hasta=cos[,2],
                                    Peso=abs(cos[,3]))
            relations2=relations[-row(relations)[relations == 0],]
            relations2
          }
        }
      })
```

Posteriormente, se configura el gráfico creado a partir de las relaciones encontradas:

```
output$plotRelaciones <- renderPlot({
     relations2<-getRelLSAGraf()
     g.1a <- graph.data.frame(relations2, directed=FALSE)
     V(g.1a)$size<-6
     min<-.1 #threshold
     layout1 <- layout.auto(g.1a)
     plot(g.1a, layout=layout1,edge.width=ifelse(E(g.1a)$Peso>=min,E(g.1a)$Peso, NA))
      })

output$tablaCos <- renderDataTable({
      tabla<-getRelLSAGraf()
      datatable(tabla, caption="Similitud del Coseno",
                  selection=list(mode='single', selected=1),
                  options = list(language = list(url = 'Spanish.json'),
                                 pageLength =20))
      })
```

En la interfaz gráfica del usuario se incluye la tabla de relaciones de similitud con un control del tipo *dataTableOutput*, un botón con el control *actionBttn* para generar el gráfico anterior.

```
tabItem(tabName="i",
          h2("Gráfico de Similitud del Coseno"),
          htmlOutput("tituloGrafRelSim"),
          br(),
          actionBttn(inputId ="btnGraf",
                           label="Obtener Gráfico",
                           color = "primary",
                           style= "gradient",
                           icon=icon("download"),
                           block = FALSE,
                           size = "md"),
          wellPanel(htmlOutput("tituloGraf")),
          plotOutput("plotRelaciones",width='1000px', height = '800px' ),
          br(),br(),
          tags$div(style='cursor:pointer',dataTableOutput("tablaCos")),
          br(),br()
          )
```

6.3 Resultados

La pantalla de inicio, *Búsqueda en Pubmed*, solicita al usuario el término que quiere buscarse en Medline, y el intervalo de fechas de publicación de los documentos seleccionados, permitiendo visualizar la consulta que se lanzará a Pubmed. La búsqueda que se va a realizar será obtener todas las publicaciones existentes en Medline sobre el Síndrome de Williams entre 1980 y 2018 y constituye la primera etapa del proceso de minería de textos, que nos permitirá obtener el corpus con toda la información que posteriormente se analizará.

Fig. 24 Pantalla inicio GEN-IO.

Fig. 25 Resultado de la búsqueda en Pubmed.

En el caso que se está analizando se han obtenido 1.580 documentos. Y al seleccionar cada una de las filas se muestra su contenido, para que pueda ser leído por el usuario, incluyéndose además un enlace al documento publicado en Pubmed:

Fig. 26 Resumen de la publicación seleccionada

Como se puede observar, en la parte inferior del resumen de la publicación se incluye un enlace para navegar hacia la web de Pubmed, desde donde será posible acceder al documento completo.

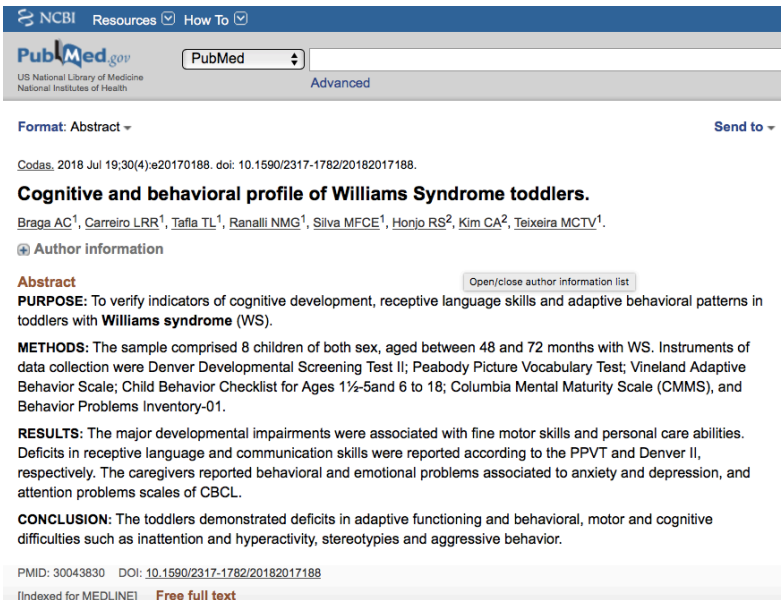


Fig. 27 Publicación en Pubmed

Usando el cuadro *Buscar*, podemos realizar búsquedas para filtrar, de entre todas las publicaciones, las que en su campo *Journal* contienen el término buscado. Por ejemplo, se han encontrado veinte registros que contienen 2018:

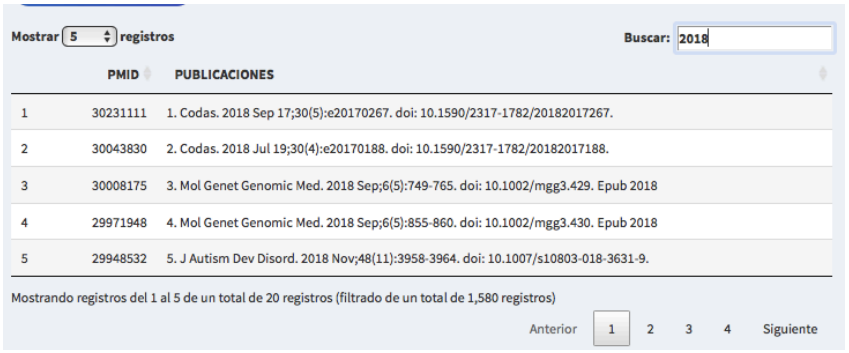


Fig. 28 Publicaciones que en su campo Journal contienen 2018

Desde la opción de menú Explorar publicaciones podemos navegar por todos los abstracts que componen el corpus, y extraer la información relevante de Pubtator:

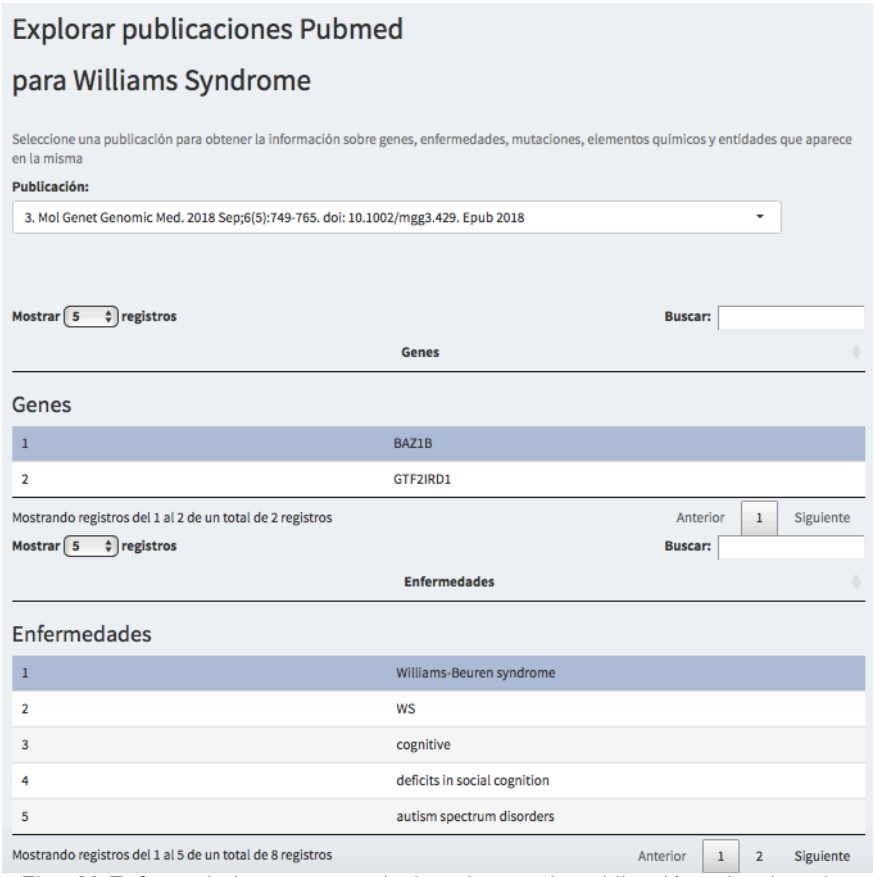


Fig. 29 Enfermedades y genes relacionados con la publicación seleccionada.

Es importante, a partir del corpus seleccionado, poder extraer la lista de palabras junto con el número de veces que cada una de ellas aparece, y esto se puede realizar desde la opción de menú *Frecuencia de palabras*. Como se puede observar, las palabras que con mayor frecuencia aparecen en este corpus son: *ws, syndrome, williams, children, patients, individuals, information, results,…*

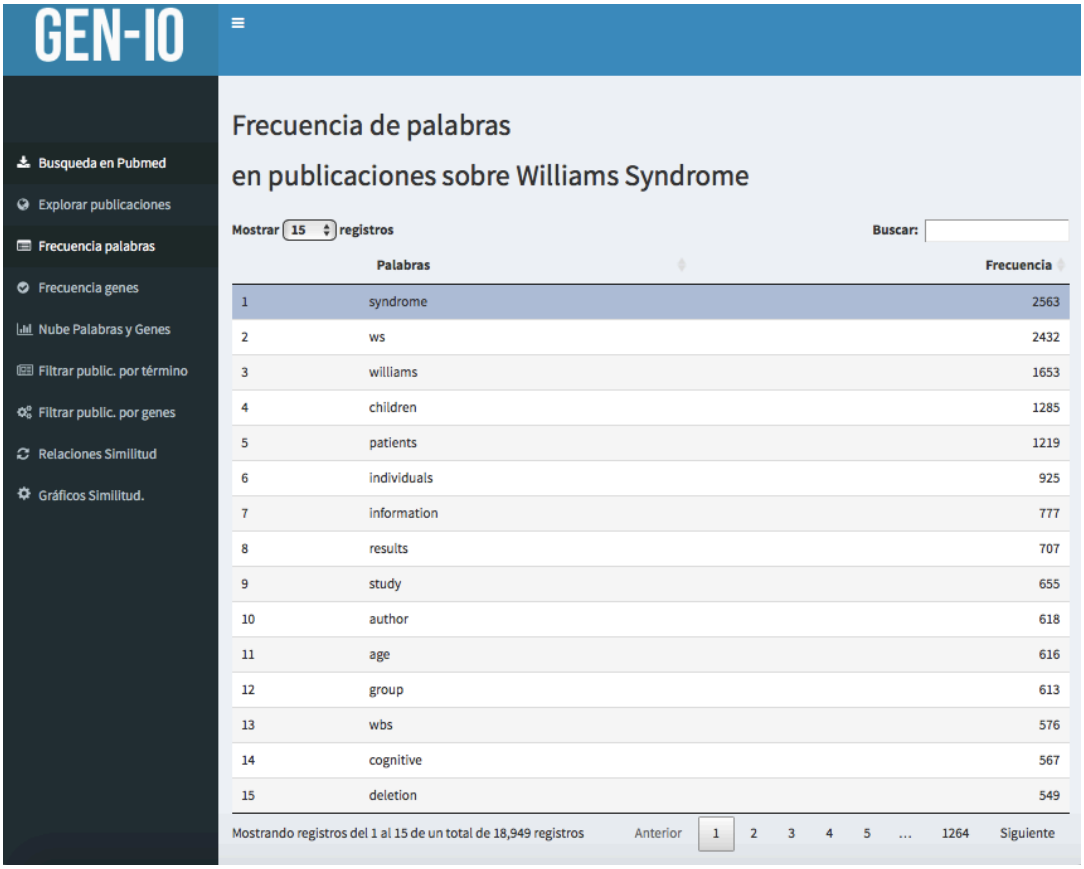


Fig. 30 Frecuencia palabras del corpus seleccionado.

En la tabla, usando el control *Buscar*, se pueden realizar búsquedas por un término, por ejemplo “denta”:

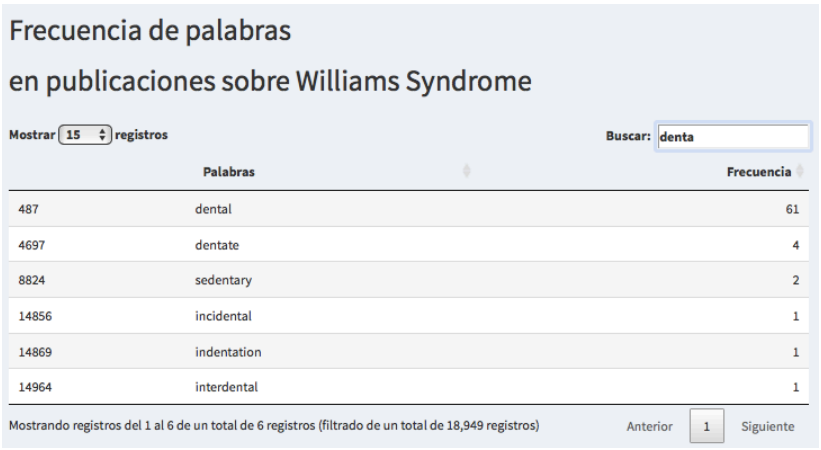
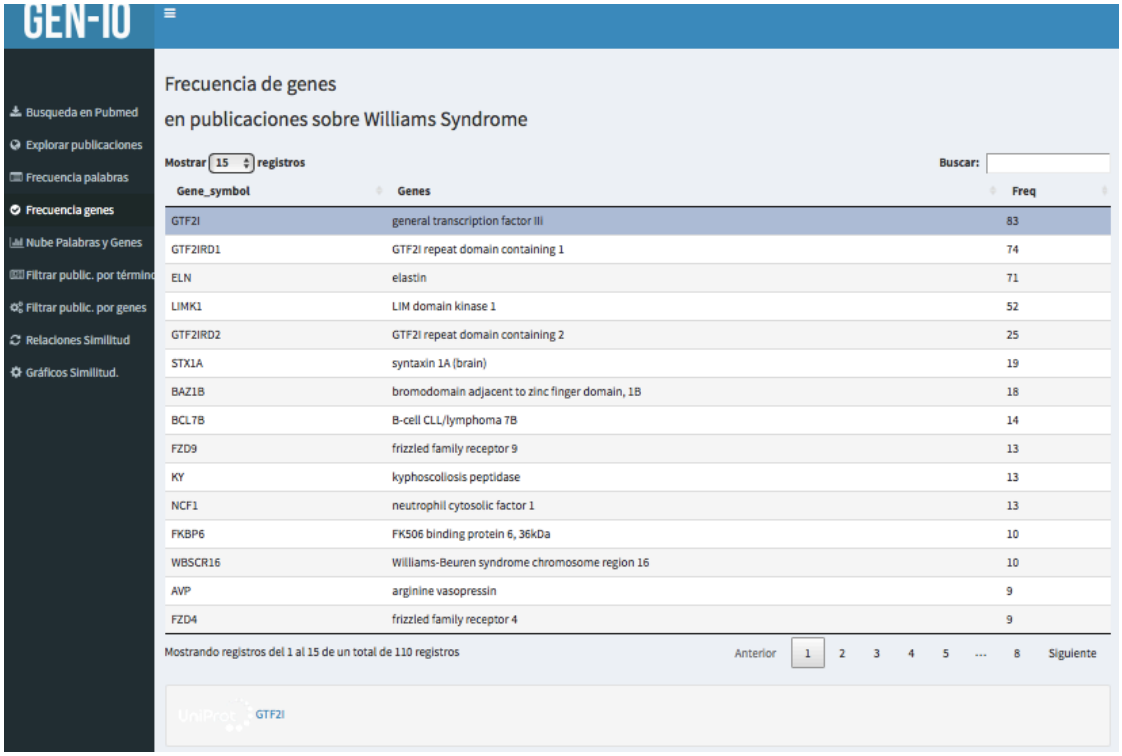


Fig. 31 Palabras que contienen “denta”

De igual modo, se puede obtener la *Frecuencia genes*, que ayuda a conocer qué genes han sido afectados en el síndrome de Williams, que como ya se comentó en el capítulo 2, éste se produce por la pérdida de información genética en la posición 11.23 del cromosoma 7. Y mediante la minería de textos se ha podido averiguar que los principales genes afectados son:



Gene_symbol	Genes	Freq
GTF2I	general transcription factor III	83
GTF2IRD1	GTF2I repeat domain containing 1	74
ELN	elastin	71
LIMK1	LIM domain kinase 1	52
GTF2IRD2	GTF2I repeat domain containing 2	25
STX1A	syntaxin 1A (brain)	19
BAZ1B	bromodomain adjacent to zinc finger domain, 1B	18
BCL7B	B-cell CLL/lymphoma 7B	14
FZD9	frizzled family receptor 9	13
KY	kyphoscoliosis peptidase	13
NCF1	neutrophil cytosolic factor 1	13
FKBP6	FK506 binding protein 6, 36kDa	10
WBSCR16	Williams-Beuren syndrome chromosome region 16	10
AVP	arginine vasopressin	9
FZD4	frizzled family receptor 4	9

Gene_symbol	Genes	Freq
T	T, brachyury homolog (mouse)	9
CLIP2	CAP-GLY domain containing linker protein 2	8
FZD3	frizzled family receptor 3	8
RFC2	replication factor C (activator 1) 2, 40kDa	8
VDR	vitamin D (1,25- dihydroxyvitamin D3) receptor	8
TBL2	transducin (beta)-like 2	7
MAGI2	membrane associated guanylate kinase, WW and PDZ domain containing 2	6
NF1	neurofibromin 1	6
NGF	nerve growth factor (beta polypeptide)	6
NHS	Nance-Horan syndrome (congenital cataracts and dental anomalies)	6
TRPC3	transient receptor potential cation channel, subfamily C, member 3	6
ATR	ataxia telangiectasia and Rad3 related	5
FAS	Fas (TNF receptor superfamily, member 6)	5
FMR1	fragile X mental retardation 1	5
STAG3	stromal antigen 3	5

Fig. 32 Treinta genes más frecuentes en el corpus analizado sobre WS

Para mostrar de forma visual qué palabras y qué genes se repiten con más frecuencia se utiliza la *Nube de Palabras y Genes*, que es el método más común para dicho propósito dentro de la minería de textos.

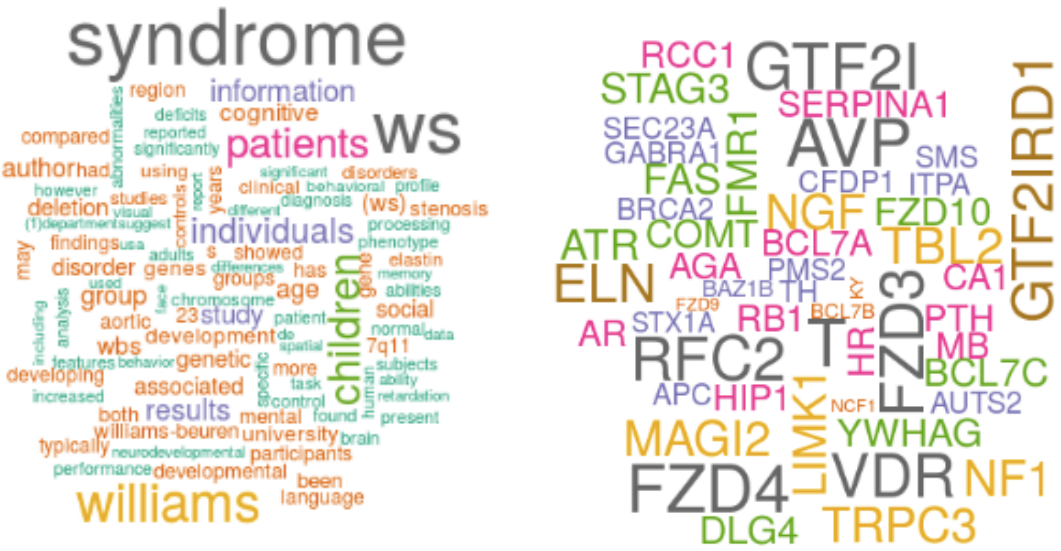


Fig. 33 Nube de palabras y genes

En la sección *Filtrar public. por genes* se permite seleccionar un gen y buscar aquellos artículos en los que se habla de ese gen. Por ejemplo, podemos ver que existen 58 publicaciones en las que se menciona al gen GTF2I. Y seleccionando cada una de ellas accedemos a la información relevante de esa publicación:

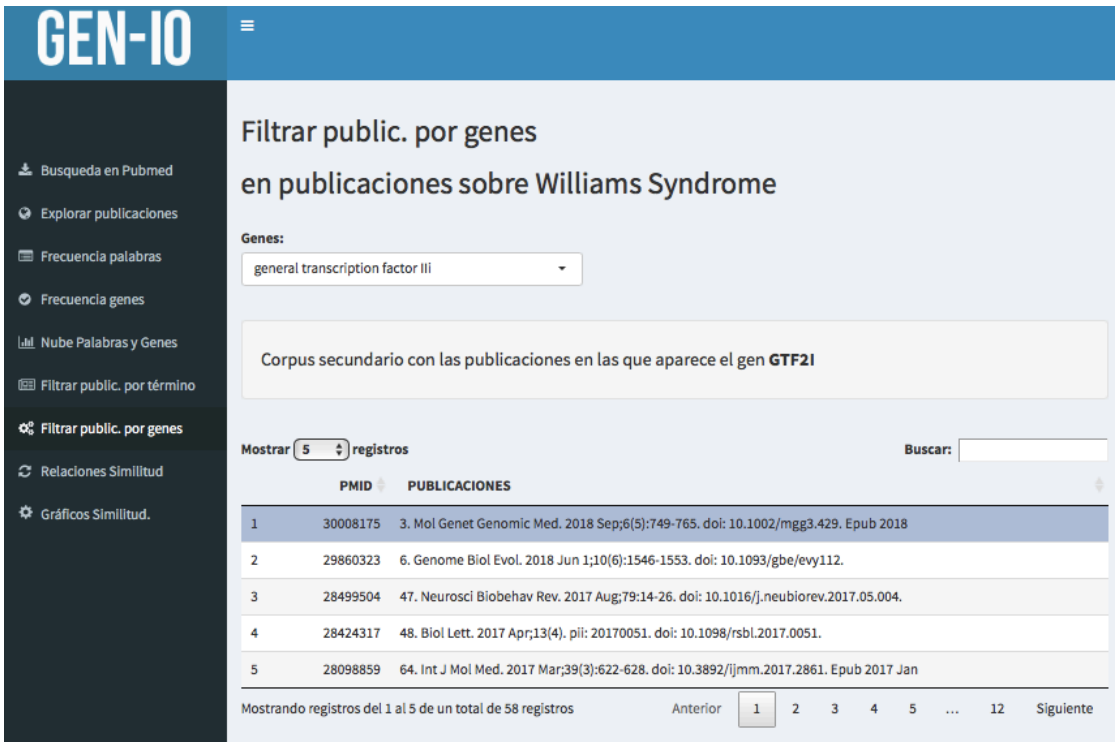
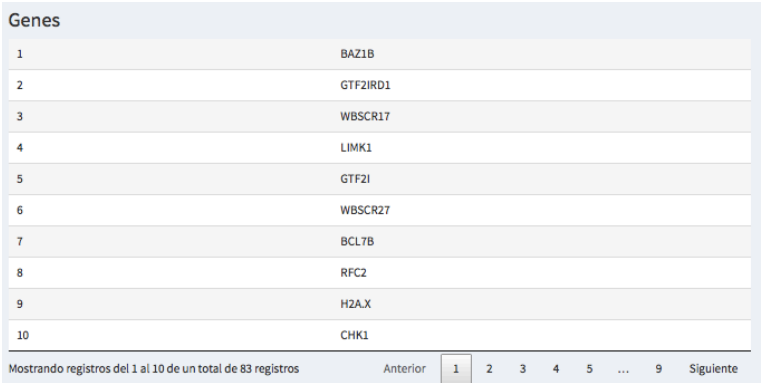


Fig. 34 Filtrar aquellas publicaciones en las que se menciona el gen GTF2I

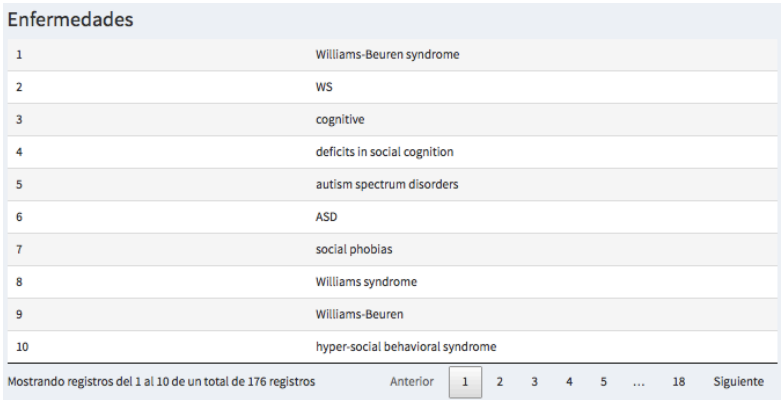


Genes

1	BAZ1B
2	GTF2IRD1
3	WBSCR17
4	LIMK1
5	GTF2I
6	WBSCR27
7	BCL7B
8	RFC2
9	H2A.X
10	CHK1

Mostrando registros del 1 al 10 de un total de 83 registros Anterior 1 2 3 4 5 ... 9 Siguiente

Fig. 35 Genes que aparecen en las mismas publicaciones que el gen GTF2I

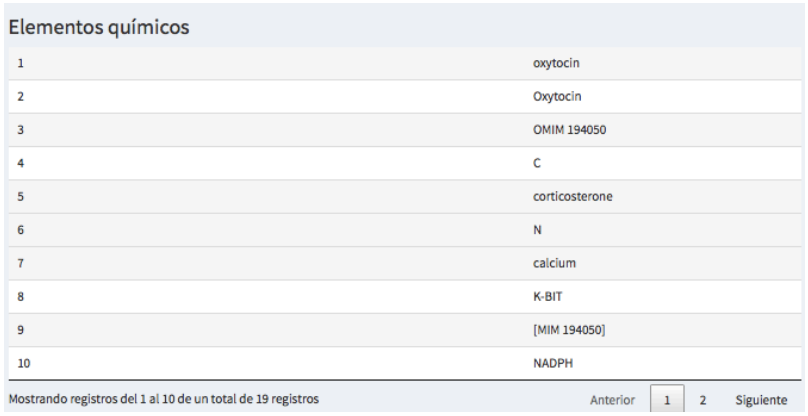


Enfermedades

1	Williams-Beuren syndrome
2	WS
3	cognitive
4	deficits in social cognition
5	autism spectrum disorders
6	ASD
7	social phobias
8	Williams syndrome
9	Williams-Beuren
10	hyper-social behavioral syndrome

Mostrando registros del 1 al 10 de un total de 176 registros Anterior 1 2 3 4 5 ... 18 Siguiente

Fig. 36 Enfermedades que aparecen en las mismas publicaciones que el gen GTF2I

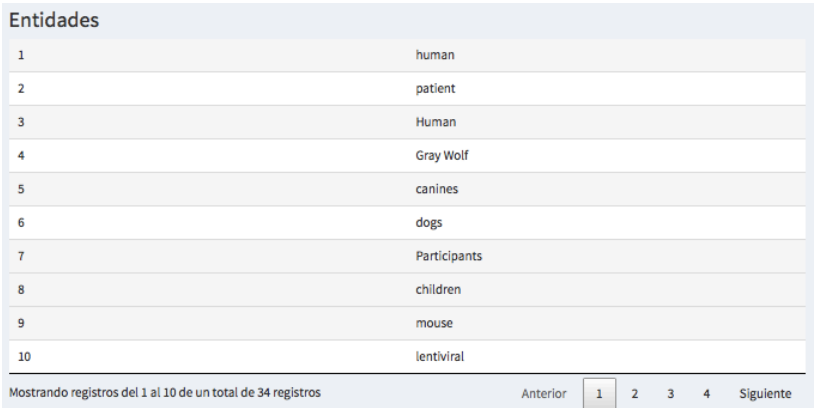


Elementos químicos

1	oxytocin
2	Oxytocin
3	OMIM 194050
4	C
5	corticosterone
6	N
7	calcium
8	K-BIT
9	[MIM 194050]
10	NADPH

Mostrando registros del 1 al 10 de un total de 19 registros Anterior 1 2 Siguiente

Fig. 37 Elementos químicos que aparecen en las mismas publicaciones que el gen GTF2I



Entidades

1	human
2	patient
3	Human
4	Gray Wolf
5	canines
6	dogs
7	Participants
8	children
9	mouse
10	lentiviral

Mostrando registros del 1 al 10 de un total de 34 registros Anterior 1 2 3 4 Siguiente

Fig. 38 Entidades que aparecen en las mismas publicacions que el gen GTF2I

Desde *Filtrar publicaciones por término*, se puede encontrar el corpus secundario de todas publicaciones que contienen los términos que indiquemos. Por ejemplo, se han encontrado 150 publicaciones en las que se menciona la expresión *mental retardation*:

Fig. 39 Filtrar publicacions en las que se menciona el termino mental retardation

Para encontrar las relaciones de similitud entre enfermedades y genes, es necesario obtener aquellas enfermedades relacionadas con el WS, pulsando *Buscar enfermedades*:

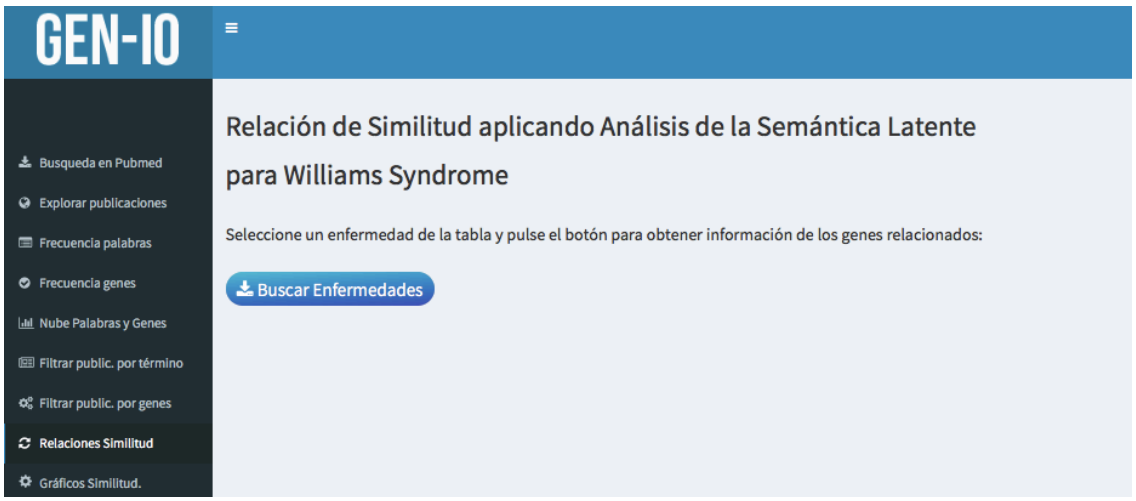


Fig. 40 Búsqueda de Relaciones de Similitud

Se han encontrado un total de 257 enfermedades relacionadas con el WS, en un corpus generado con los documentos de los últimos dos años. Para algunas de ellas no se ha podido encontrar ninguna relación de similitud con ningún gen.

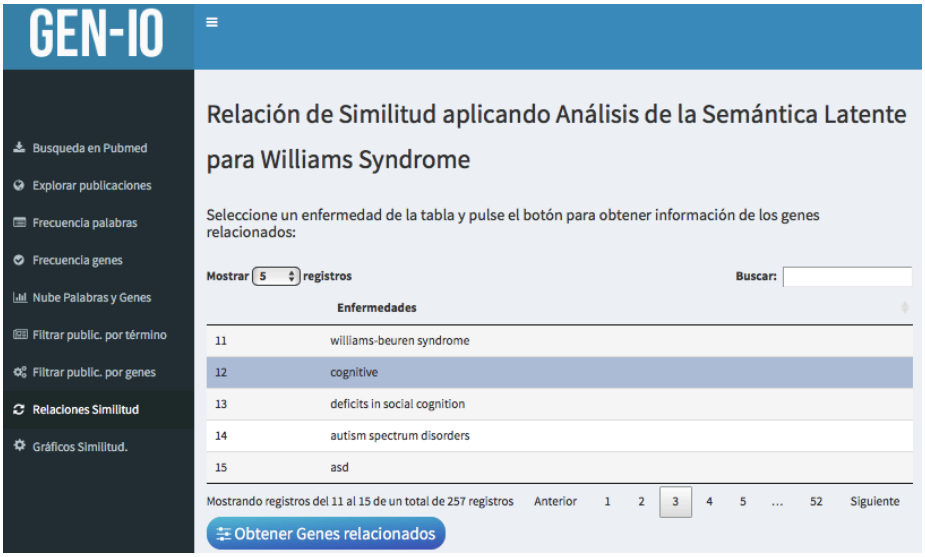


Fig. 41 Listado de enfermedades relacionades con el Síndrome de Williams

Se han calculado dos medidas para ver la relación de similitud entre enfermedad-gen, una con el análisis de la semántica latente y otra que se calcula con el coseno de similitud. Se ha podido comprobar que el coseno de similitud es mucho más restrictivo que LSA. Realmente lo que se obtienen son coincidencias en párrafos, y a continuación se muestran evidencias de los datos obtenidos.

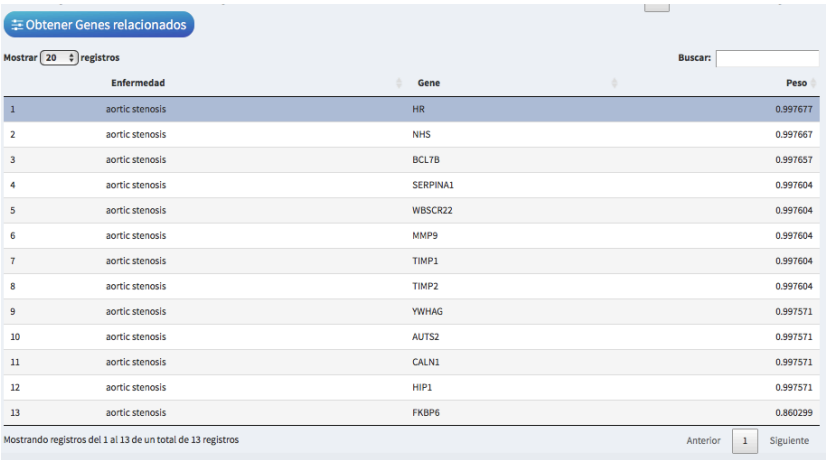


Fig. 42 Relaciones de similitud entre aortic stenosis y genes

	Desde	Hasta	Peso
1	ELN	GTF2IRD1	0.11273277250512
2	ELN	LIMK1	0.604096257192828
3	ELN	GTF2I	0.198471914140455
4	ELN	FKBP6	0.0449183089248791
10	ELN	BAZ1B	0.0905357460425185
12	ELN	CLIP2	0.0905357460425185
13	ELN	F2	0.197907270281823
20	ELN	aortic stenosis	0.137296831831592
21	GTF2IRD1	LIMK1	0.0599161759894059
22	GTF2IRD1	GTF2I	0.926782226247676
29	GTF2IRD1	BAZ1B	0.287347885566345
31	GTF2IRD1	CLIP2	0.287347885566345
32	GTF2IRD1	F2	0.924145571612652
39	GTF2IRD1	aortic stenosis	0.0439139071172663
40	LIMK1	GTF2I	0.189146497497384
41	LIMK1	FKBP6	0.0517260600111872
47	LIMK1	BAZ1B	0.208514414057075
49	LIMK1	CLIP2	0.208514414057075
50	LIMK1	F2	0.188608384038579
57	LIMK1	aortic stenosis	0.066183627509317

Mostrando registros del 1 al 20 de un total de 54 registros Anterior 1 2 3 Siguiente

Fig. 43 Relaciones de similitud del coseno para aortic stenosis

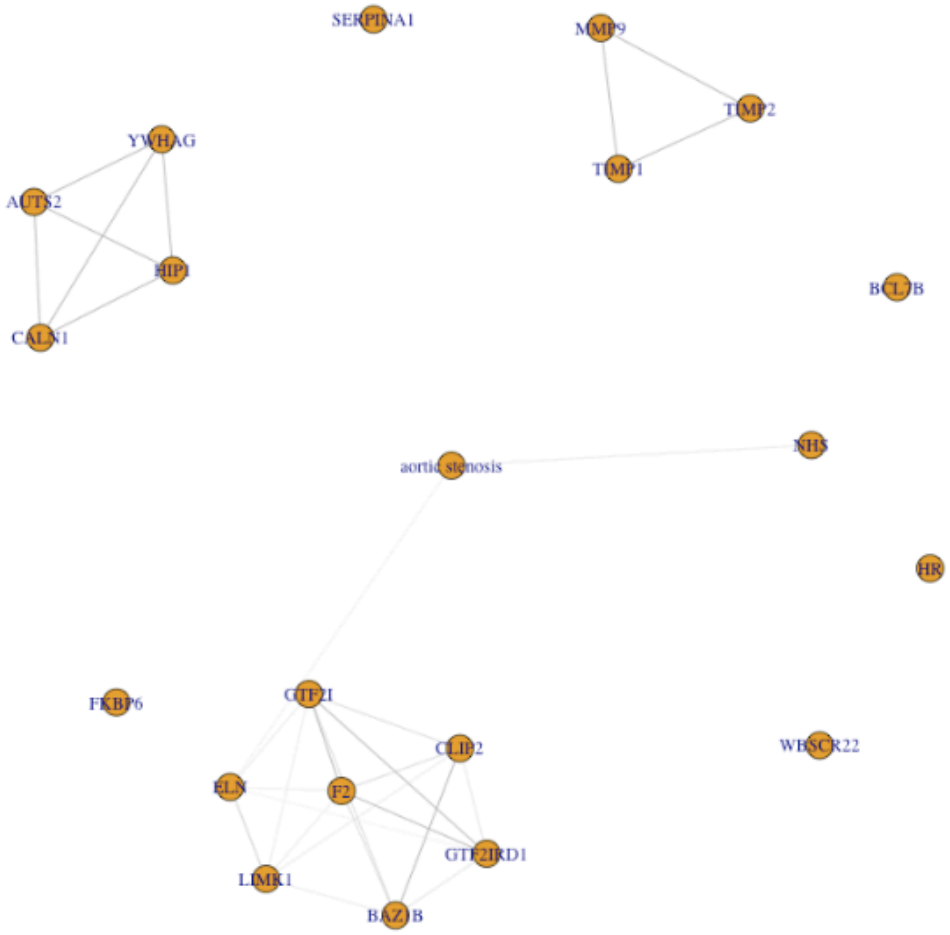


Fig. 44 Gráfico Relaciones de similitud del coseno para aortic stenosis

A continuación se detallan las enfermedades sobre las que sí que se han hallado relaciones con genes encontrados en párrafos del corpus. Se adjuntan pruebas de evidencia, mostrando los párrafos en los que aparecen los genes junto a las enfermedades, en los documentos que forman el corpus.

SUPRAVALVULAR AORTIC STENOSIS (estenosis aórtica supravalvular)

	Enfermedad	Gene	Peso
1	supravalvular aortic stenosis	HR	0.99647
2	supravalvular aortic stenosis	BCL7B	0.996437
3	supravalvular aortic stenosis	WBSCR22	0.996353
4	supravalvular aortic stenosis	MMP9	0.996353
5	supravalvular aortic stenosis	TIMP1	0.996353
6	supravalvular aortic stenosis	TIMP2	0.996353
7	supravalvular aortic stenosis	YWHAG	0.996301
8	supravalvular aortic stenosis	AUTS2	0.996301
9	supravalvular aortic stenosis	CALN1	0.996301
10	supravalvular aortic stenosis	HIP1	0.996301
11	supravalvular aortic stenosis	FKBP6	0.995194
12	supravalvular aortic stenosis	ELN	0.759203
13	supravalvular aortic stenosis	LIMK1	0.702793

- Characterization of two novel genes, WBSCR20 and **WBSCR22**, deleted in Williams-Beuren syndrome. Doll A Grzeschik KH Williams-Beuren syndrome (WBS), due to a contiguous gene deletion of approximately 1.5 Mb at 7q11.23, is a complex developmental disorder with multisystemic manifestations including **supravalvular aortic stenosis** (SVAS) and a specific cognitive phenotype. Large repeats containing genes and pseudogenes flank the deletion breakpoints, and the mutation mechanism commonly appears to be unequal meiotic crossover. Except for elastin, hemizygosity of which is associated with **supravalvular aortic stenosis**, it is unknown which of the 18 genes in the deletion area contributes to the phenotype. Here, we report the identification and characterization of two novel genes, WBSCR20 and **WBSCR22**, which map to the common WBS deletion region. **WBSCR22** encodes a putative methyltransferase protein strongly expressed in heart, skeletal muscle and kidney. WBSCR20 encodes a novel protein expressed in skeletal muscle with similarity to p120 (NOL1), a 120-kDa proliferation-associated nucleolar antigen, a member of an evolutionarily conserved protein family. A highly similar putative gene, WBSCR20B, flanks the WBS deletion at the telomeric side. Hemizygous deletion of either of the novel genes might contribute to the growth retardation, the myopathy or the premature aging effects in the pathogenesis of WBS.Copyright 2002 S. Karger AG, Basel https://www.ncbi.nlm.nih.gov/pubmed/?term=11978965
- Vascular wall remodeling in patients with **supravalvular aortic stenosis** and Williams Beuren syndrome. Dridi SM Foucault Bertaud A Igondjo Tchen S Senni K Ejeil AL Pellat B Lyonnet S Bonnet D Charpiot P Godeau G **Supravalvular aortic stenosis** (SVAS) and Williams Beuren syndrome (WBS) can be considered as inherited diseases affecting the whole arterial tree and causing narrowing of the vessels. It has been reported that abnormal deposition of elastin in arterial walls of patients with SVAS and WBS leads to increased proliferation of arterial smooth muscle cells (SMC), which result in the formation of hyperplastic intimal lesions. In this work, we conducted morphological and morphometrical analysis with stenotic aortas from patients suffering from **SVAS** and WBS and from healthy control subjects and demonstrated that the amount of elastic fibers and the loss of integrity of vascular elastic fibers in the aortas reflect similar changes in the skin of patients with SVAS or WBS, as reported in our previous work conducted on skin in these pathological states. On the other hand, we conducted investigations on metalloproteinases (MMP2, MMP9, MMP7) and their specific tissue inhibitors **TIMP1** and **TIMP2** to verify their possible involvement in the etiopathogeny of **SVAS** and WBS. We particularly evidenced an altered MMP9/**TIMP1** balance in favor of matrix degradation which could facilitate SMC migration and neointimal hyperplasia. Our findings suggest that elastinolytic enzymes secreted by arterial SMC, possibly including matrilysin 1, are critical for the development of arterial lesions in SVAS and WBS and contribute to perpetuate arterial stenosis in either SVAS or WBS.Copyright (c) 2005 S. Karger AG, Basel. https://www.ncbi.nlm.nih.gov/pubmed/?term=15832055
- Zebrafish gene knockdowns imply roles for human YWHAG in infantile spasms and cardiomegaly. Komoike Y Fujii K Nishimura A Hiraki Y Hayashidani M Shimojima K Nishizawa T Higashi K Yasukawa K Saitsu H Miyake N Mizuguchi T Matsumoto N Osawa M Kohno Y Higashinakagawa T Yamamoto T Williams-Beuren syndrome (WBS) is a neurodevelopmental disorder presenting with an elfin-like face, **supravalvular aortic stenosis**, a specific cognitive-behavioral profile, and infantile hypercalcemia. We encountered two WBS patients presenting with infantile spasms, which is extremely rare in WBS. Array comparative genomic hybridization (aCGH) and fluorescent in situ hybridization (FISH) analyses revealed atypical 5.7-Mb and 4.1-Mb deletions at 7q11.23 in the two patients, including the WBS critical region and expanding into the proximal side and the telomeric side, respectively. On the proximal side, **AUTS2** and **CALN1** may contribute to the phenotype. On the telomeric side, there are two candidate genes **HIP1** and **YWHAG**. Because detailed information of them was unavailable, we investigated their functions using gene knockdowns of zebrafish. When zebrafish **YWHAG** was knocked down, reduced brain size and increased diameter of the heart tube were observed, indicating that the infantile spasms and cardiomegaly seen in the patient with the telomeric deletion may be derived from haploinsufficiency of **YWHAG**.2010 Wiley-Liss, Inc. https://www.ncbi.nlm.nih.gov/pubmed/?term=20146355

SUPRAVALVAR AORTIC STENOSIS (estenosis aórtica supravalvular)

	Enfermedad	Gene	Peso
1	supravalvar aortic stenosis	NHS	0.994775
2	supravalvar aortic stenosis	SERPINA1	0.991637

- Alpha 1 antitrypsin deficiency alleles are associated with joint dislocation and scoliosis in Williams syndrome. Morris CA Pani AM Mervis CB Rios CM Kistler DJ Gregg RG Elastin haploinsufficiency is responsible for a significant portion of the Williams syndrome (WS) phenotype including hoarse voice, **supravalvar aortic stenosis** (SVAS), hernias, diverticuli of bowel and bladder, soft skin, and joint abnormalities. All of the connective tissue signs and symptoms are variable in the WS population, but few factors other than age and gender are known to influence the phenotype. We examined a cohort of 205 individuals with WS for mutations in **SERPINA1**, the gene that encodes alpha-1-antitrypsin (AAT), the inhibitor of elastase. Individuals with classic WS deletions and **SERPINA1** genotypes PiMS or PiMZ were more likely than those with a **SERPINA1** PiMM genotype to have joint dislocation or scoliosis. However, carrier status for AAT deficiency was not correlated with presence of inguinal hernia or with presence or severity of SVAS. These findings suggest that genes important in elastin metabolism are candidates for variability in the connective tissue abnormalities in WS. https://www.ncbi.nlm.nih.gov/pubmed/?term=20425789

SENSORINEURAL HEARING LOSS (Pérdida auditiva neurosensorial)

	Enfermedad	Gene	Peso
1	sensorineural hearing loss	ELN	0.942679

- **Sensorineural hearing loss** in children and adults with Williams syndrome. Marler JA Elfenbein JL Ryals BM Urban Z Netzloff ML Williams syndrome (WS) is a genetic neurodevelopmental disorder, most often accompanied by mild-to-moderate mental retardation. Individuals with WS show unique communication strengths and impairments that are challenging to treat in community, educational, and vocational settings. Many issues regarding characteristics of auditory sensitivity in WS remain to be resolved. Our purpose was to obtain behavioral (screening and pure-tone audiometry) and objective (distortion product otoacoustic emission-DPOAE) measures of auditory system function from a group of 27 individuals with WS, 6-48 years of age. These measures were gathered both at an international professional conference (n = 19) and in a clinic setting (n = 8). ========= ========= Post hoc analyses revealed a significant effect for age, suggesting a pattern of progressive hearing loss. An effect size analysis indicated a clinically meaningful difference in the hearing sensitivity between school-aged children and adults in the high frequencies (4,000 and 8,000 Hz). Similar hearing loss phenotype was observed in patients with familial nonsyndromic supravalvular aortic stenosis (SVAS), suggesting that molecular defects in the **elastin gene** in the pathogenesis of SNHL in WS. This study highlights the importance of early and regular hearing testing for WS patients and suggests that **elastin** may have a previously unappreciated function in maintaining hearing sensitivity.Copyright 2005 Wiley-Liss, Inc https://www.ncbi.nlm.nih.gov/pubmed/?term=16222677

ANXIETY (ansiedad):

	Enfermedad	Gene	Peso
1	anxiety	KY	0.957845
2	anxiety	ELN	0.918281
3	anxiety	DLG4	0.877285
4	anxiety	AVP	0.877285
5	anxiety	CA1	0.747924

- Williams syndrome and related disorders. Morris CA Mervis CB Three clinical conditions displaying phenotypic overlap have been linked to mutation or deletion of the elastin gene at 7q11.23. Supravalvar aortic stenosis, an autosomal dominant disorder characterized by elastin arteriopathy, is caused by mutation or intragenic deletions of **ELN**resulting in loss of function. Autosomal dominant cutis laxa, a primarily cutaneous condition, is the result of frameshift mutations at **ELN** that cause a dominant-negative effect on elastic fiber structure. Williams syndrome, a neurodevelopmental disorder is due to a 1.5 Mb deletion that includes **ELN** and at least 15 contiguous genes. The disorder is characterized by dysmorphic facies, mental retardation or learning difficulties, elastin arteriopathy, a unique cognitive profile of relative strength in auditory rote memory and language and extreme weakness in visuospatial constructive cognition, and a typical personality that includes overfriendliness, **anxiety**, and attention problems. The understanding of these disorders has progressed from phenotypic description to identification of causative mutations and insight into pathogenetic mechanisms for some aspects of the phenotype. https://www.ncbi.nlm.nih.gov/pubmed/?term=11701637
- Association of mouse **DLG4** (PSD-95) gene deletion and human **DLG4** gene variation with phenotypes relevant to autism spectrum disorders and Williams' syndrome. Feyder M Karlsson RM Mathur P Lyman M Bock R Momenan R Munasinghe J Scattoni ML Ihne J Camp M Graybeal C Strathdee D Begg A Alvarez VA Kirsch P Rietschel M Cichon S Walter H Meyer-Lindenberg A Grant SG Holmes A Research is increasingly linking autism spectrum disorders and other neurodevelopmental disorders to synaptic abnormalities ("synaptopathies"). PSD-95 (postsynaptic density-95, **DLG4**) orchestrates protein-protein interactions at excitatory synapses and is a major functional bridge interconnecting a neurexinneuroligin-SHANK pathway implicated in autism spectrum disorders.The authors characterized behavioral, dendritic, and molecular phenotypic abnormalities relevant to autism spectrum disorders in mice with PSD-95 deletion (**DLG4**â»(/)â»). The data from mice led to the identification of single-nucleotide polymorphisms (SNPs) in human **DLG4** and the examination of associations between these variants and neural signatures of Williams' syndrome in a normal population, using functional and structural neuroimaging.**DLG4**â»(/)â» showed increased repetitive behaviors, abnormal communication and social behaviors, impaired motor coordination, and increased stress reactivity and **anxiety**-related responses. **DLG4**â»(/)â» had subtle dysmorphology of amygdala dendritic spines and altered forebrain expression of various synaptic genes, including Cyln2, which regulates cytoskeletal dynamics and is a candidate gene for Williams' syndrome. A signifi-cant association was observed between variations in two human **DLG4** SNPs and reduced intraparietal sulcus volume and abnormal cortico-amygdala coupling, both of which characterize Williams' syndrome.These findings demonstrate that **DLG4** gene disruption in mice produces a complex range of behavioral and molecular abnormalities relevant to autism spectrum disorders and Williams' syndrome. The study provides an initial link between human **DLG4** gene variation and key neural endophenotypes of Williams' syndrome and perhaps corticoamygdala regulation of emotional and social processes more generally. https://www.ncbi.nlm.nih.gov/pubmed/?term=20952458
- Oxytocin and vasopressin are dysregulated in Williams Syndrome, a genetic disorder affecting social behavior. Dai L Carter CS Ying J Bellugi U Pournajafi-Nazarloo H Korenberg JR The molecular and neural mechanisms regulating human social-emotional behaviors are fundamentally important but largely unknown; unraveling these requires a genetic systems neuroscience analysis of human models. Williams Syndrome (WS), a condition caused by deletion of ~28 genes, is associated with a gregarious personality, strong drive to approach strangers, difficult peer interactions, and attraction to music. WS provides a unique opportunity to identify endogenous human gene-behavior mechanisms. Social neuropeptides including oxytocin (OT) and arginine vasopressin (**AVP**) regulate reproductive and social behaviors in mammals, and we reasoned that these might mediate the features of WS. Here we established blood levels of OT and **AVP** in WS and controls at baseline, and at multiple timepoints following a positive emotional intervention (music), and a negative physical stressor (cold). ========= ========= In WS, baseline OT but not **AVP**, was correlated positively with approach, but negatively with adaptive social behaviors. These results indicate that WS deleted genes perturb hypothalamic-pituitary release not only of OT but also of **AVP**, implicating more complex neuropeptide circuitry for WS features and providing evidence for their roles in endogenous regulation of human social behavior. The data suggest a possible biological basis for amygdalar involvement, for increased **anxiety**, and for the paradox of increased approach but poor social relationships in WS. They also offer insight for translating genetic and neuroendocrine knowledge into treatments for disorders of social behavior. https://www.ncbi.nlm.nih.gov/pubmed/?term=22719898

DEFICITS IN SOCIAL COGNITION (deficits en cognición social)

	Enfermedad	Gene	Peso
1	deficits in social cognition	GTF2IRD1	0.707107
2	deficits in social cognition	BAZ1B	0.707107

- Exome sequencing of 85 Williams-Beuren syndrome cases rules out coding variation as a major contributor to remaining variance in social behavior. Kopp ND Parrish PCR Lugo M Dougherty JD Kozel BA Large, multigenic deletions at chromosome 7q11.23 result in a highly penetrant constellation of physical and behavioral symptoms known as Williams-Beuren syndrome (WS). Of particular interest is the unusual social-cognitive profile evidenced by **deficits in social cognition** and communication reminiscent of autism spectrum disorders (ASD) that are juxtaposed with normal or even relatively enhanced social motivation. Interestingly, duplications in the same region also result in ASD-like phenotypes as well as social phobias. Thus, the region clearly regulates human social motivation and behavior, yet the relevant gene(s) have not been definitively identified.Here, we deeply phenotyped 85 individuals with WS and used exome sequencing to analyze common and rare variation for association with the remaining variance in social behavior as assessed by the Social Responsiveness Scale.We replicated the previously reported unusual juxtaposition of behavioral symptoms in this new patient collection, but we did not find any new alleles of large effect in the targeted analysis of the remaining copy of genes in the Williams syndrome critical region. However, we report on two nominally significant SNPs in two genes that have been implicated in the cognitive and social phenotypes of Williams syndrome, **BAZ1B** and **GTF2IRD1**. Secondary discovery driven explorations focusing on known ASD genes and an exome wide scan do not highlight any variants of a large effect.Whole exome sequencing of 85 individuals with WS did not support the hypothesis that there are variants of large effect within the remaining Williams syndrome critical region that contribute to the social phenotype. This deeply phenotyped and genotyped patient cohort with a defined mutation provides the opportunity for similar analyses focusing on noncoding variation and/or other phenotypic domains.Â© 2018 The Authors. Molecular Genetics & Genomic Medicine published by Wiley Periodicals, Inc. https://www.ncbi.nlm.nih.gov/pubmed/?term=30008175

COGNITIVE DEFICITS (déficits cognitivos)

	Enfermedad	Gene	Peso
1	cognitive deficits	ELN	0.88358
2	cognitive deficits	BAZ1B	0.833632
3	cognitive deficits	CLIP2	0.833632
4	cognitive deficits	GTF2I	0.82256
5	cognitive deficits	GTF2IRD1	0.819812
6	cognitive deficits	STX1A	0.815239

- An atypical 7q11.23 deletion in a normal IQ Williams-Beuren syndrome patient. Ferrero GB Howald C Micale L Biamino E Augello B Fusco C Turturo MG Forzano S Reymond A Merla G Williams-Beuren syndrome (WBS; OMIM no. 194050) is a multisystemic neurodevelopmental disorder caused by a hemizygous deletion of 1.55 Mb on chromosome 7q11.23 spanning 28 genes. Haploinsufficiency of the **ELN** gene was shown to be responsible for supravalvular aortic stenosis and generalized arteriopathy, whereas LIMK1, **CLIP2**, **GTF2IRD1** and **GTF2I** genes were suggested to be linked to the specific cognitive profile and craniofacial features. These insights for genotype-phenotype correlations came from the molecular and clinical analysis of patients with atypical deletions and mice models. Here we report a patient showing mild WBS physical phenotype and normal IQ, who carries a shorter 1 Mb atypical deletion. This rearrangement does not include the **GTF2IRD1** and **GTF2I** genes and only partially the **BAZ1B** gene. Our results are consistent with the hypothesis that hemizygosity of the **GTF2IRD1** and **GTF2I** genes might be involved in the facial dysmorphisms and in the specific motor and **cognitive deficits** observed in WBS patients. https://www.ncbi.nlm.nih.gov/pubmed/?term=19568270

- Contribution of CYLN2 and **GTF2IRD1** to neurological and cognitive symptoms in Williams Syndrome. van Hagen JM van der Geest JN van der Giessen RS Lagers-van Haselen GC Eussen HJ Gille JJ Govaerts LC Wouters CH de Coo IF Hoogenraad CC Koekkoek SK Frens MA van Camp N van der Linden A Jansweijer MC Thorgeirsson SS De Zeeuw CI Williams Syndrome (WS, [MIM 194050]) is a disorder caused by a hemizygous deletion of 25-30 genes on chromosome 7q11.23. Several of these genes including those encoding cytoplasmic linker protein-115 (CYLN2) and general transcription factors (**GTF2I** and **GTF2IRD1**) are expressed in the brain and may contribute to the distinct neurological and **cognitive deficits** in WS patients. Recent studies of patients with partial deletions indicate that hemizygosity of **GTF2I** probably contributes to mental retardation in WS. Here we investigate whether CYLN2 and **GTF2IRD1** contribute to the motoric and **cognitive deficits** in WS. Behavioral assessment of a new patient in which **STX1A** and LIMK1, but not CYLN2 and **GTF2IRD1** , are deleted showed that his cognitive and motor coordination functions were significantly better than in typical WS patients. Comparative analyses of gene specific CYLN2 and **GTF2IRD1** knockout mice showed that a reduced size of the corpus callosum as well as deficits in motor coordination and hippocampal memory formation may be attributed to a deletion of CYLN2, while increased ventricle volume can be attributed to both CYLN2 and **GTF2IRD1** . We conclude that the motor and **cognitive deficits** in Williams Syndrome are caused by a variety of genes and that heterozygous deletion of CYLN2 is one of the major causes responsible for such dysfunctions. https://www.ncbi.nlm.nih.gov/pubmed/?term=17270452

AUTISM SPECTRUM DISORDER (trastorno del espectro autista)

	Enfermedad	Gene	Peso
1	autism spectrum disorder	DLG4	1
2	autism spectrum disorder	GTF2IRD1	1
3	autism spectrum disorder	BAZ1B	1
4	autism spectrum disorder	LIMK1	1
5	autism spectrum disorder	USP7	1
6	autism spectrum disorder	GTF2I	1
7	autism spectrum disorder	GTF2IRD2	1
8	autism spectrum disorder	CTCF	1

- Exome sequencing of 85 Williams-Beuren syndrome cases rules out coding variation as a major contributor to remaining variance in social behavior. Kopp ND Parrish PCR Lugo M Dougherty JD Kozel BA Large, multigenic deletions at chromosome 7q11.23 result in a highly penetrant constellation of physical and behavioral symptoms known as Williams-Beuren syndrome (WS). Of particular interest is the unusual social-cognitive profile evidenced by deficits in social cognition and communication reminiscent of **autism spectrum disorders** (ASD) that are juxtaposed with normal or even relatively enhanced social motivation. Interestingly, duplications in the same region also result in ASD-like phenotypes as well as social phobias. Thus, the region clearly regulates human social motivation and behavior, yet the relevant gene(s) have not been definitively identified.Here, we deeply phenotyped 85 individuals with WS and used exome sequencing to analyze common and rare variation for association with the remaining variance in social behavior as assessed by the Social Responsiveness Scale.We replicated the previously reported unusual juxtaposition of behavioral symptoms in this new patient collection, but we did not find any new alleles of large effect in the targeted analysis of the remaining copy of genes in the Williams syndrome critical region. However, we report on two nominally significant SNPs in two genes that have been implicated in the cognitive and social phenotypes of Williams syndrome, **BAZ1B** and **GTF2IRD1**. Secondary discovery driven explorations focusing on known ASD genes and an exome wide scan do not highlight any variants of a large effect.Whole exome sequencing of 85 individuals with WS did not support the hypothesis that there are variants of large effect within the remaining Williams syndrome critical region that contribute to the social phenotype. This deeply phenotyped and genotyped patient cohort with a defined mutation provides the opportunity for similar analyses focusing on noncoding variation and/or other phenotypic domains.Â© 2018 The Authors. Molecular Genetics & Genomic Medicine published by Wiley Periodicals, Inc. https://www.ncbi.nlm.nih.gov/pubmed/?term=30008175

- Williams syndrome deletions and duplications: Genetic windows to understanding anxiety, sociality, **autism**, and schizophrenia. Crespi BJ Procyshyn TL We describe and evaluate an integrative hypothesis for helping to explain the major neurocognitive features of individuals with Williams syndrome region deletions and duplications. First, we demonstrate how the cognitive differences between Williams syndrome individuals, individuals with duplications of this region, and healthy individuals parallel the differences between individuals subject to effects of increased or decreased oxytocin. Second, we synthesize evidence showing that variation in expression of the gene **GTF2I** (General Transcription Factor II-I) underlies the primary social phenotypes of Williams syndrome and that common genetic variation in **GTF2I** mediates oxytocin reactivity, and its correlates, in healthy populations. Third, we describe findings relevant to the hypothesis that the **GTF2I** gene is subject to parent of origin effects whose behavioral expression fits with predictions from the kinship theory of genomic imprinting. Fourth, we describe how Williams syndrome can be considered, in part, as an autistic syndrome of Lorna Wing's 'active-but-odd' **autism** subtype, in contrast to associations of duplications with both schizophrenia and **autism**.Copyright Â© 2017 Elsevier Ltd. All rights reserved. https://www.ncbi.nlm.nih.gov/pubmed/?term=28499504

- A Novel Recurrent Breakpoint Responsible for Rearrangements in the Williams-Beuren Region. Plaja A Castells N Cueto-GonzÃ¡lez AM del Campo M Vendrell T Lloveras E Izquierdo L Borregan M RodrÃguez-Santiago B CarriÃ³ A MirÃ³ R Tizzano E Copy number variants (CNVs) of the Williams-Beuren syndrome (WBS) 7q11.23 region are responsible for neurodevelopmental disorders with multisystem involvement and variable expressivity. We found 2 patients with a deletion and 1 patient with a duplication in this region sharing a common breakpoint located between the **LIMK1** and EIF4H(WBSCR1) genes. One patient had a WBS phenotype, although testing with a commercially available FISH assay was negative for the deletion. A further test using array CGH showed an atypical WBS region deletion. The second patient showed global developmental delay, speech delay and poor motor skills with a deletion outside the WBS region. The third patient had manifestations compatible with an **autism spectrum disorder** showing a duplication in the WBS region. Our findings point to the existence of a previously unrecognized recurrent breakpoint responsible for rearrangements in the WBS region. Given that most commercial FISH assays include probes flanking this novel breakpoint, further testing with array CGH should be performed in patients with WBS and negative FISH results. Â© 2015 S. Karger AG, Basel. https://www.ncbi.nlm.nih.gov/pubmed/?term=26382598

- Cognitive-behavioral phenotypes of Williams syndrome are associated with genetic variation in the **GTF2I** gene, in a healthy population. Crespi BJ Hurd PL Individuals with Williams syndrome, a neurogenetic condition caused by deletion of a set of genes at chromosomal location 7q11.23, exhibit a remarkable suite of traits including hypersociality with high, nonselective friendliness and low social anxiety, expressive language relatively well-developed but under-developed social-communication skills overall, and reduced visual-spatial abilities. Deletions and duplications of the Williams-syndrome region have also been associated with **autism**, and with schizophrenia, two disorders centrally involving social cognition. Several lines of evidence have linked the gene **GTF2I** (General Transcription Factor IIi) with the social phenotypes of Williams syndrome, but a role for this gene in sociality within healthy populations has yet to be investigated.We genotyped a large set of healthy individuals for two single-nucleotide polymorphisms in the **GTF2I** gene that have recently been significantly associated with **autism**, and thus apparently exhibit functional effects on **autism**-related social phenotypes. **GTF2I** genotypes for these SNPs showed highly significant association with low social anxiety combined with reduced social-communication abilities, which represents a metric of the Williams-syndrome cognitive profile as described from previous studies.These findings implicate the **GTF2I** gene in the neurogenetic basis of social communication and social anxiety, both in Williams syndrome and among individuals in healthy populations. https://www.ncbi.nlm.nih.gov/pubmed/?term=25429715

- Association of **GTF2I** in the Williams-Beuren syndrome critical region with a**autism spectrum disorders**. Malenfant P Liu X Hudson ML Qiao Y Hrynchak M Riendeau N Hildebrand MJ Cohen IL Chudley AE Forster-Gibson C Mickelson EC Rajcan-Separovic E Lewis ME Holden JJ Duplications of 7q11.23, deleted in Williams-Beuren Syndrome, have been implicated in a**autism spectrum disorders** (ASDs). A 1.5 Mb duplication was identified in one girl with severe expressive language deficits and anxiety among 1,142 ASD individuals screened for this duplication. Family-based association studies of Tag-SNPs in three genes (STX1A , CYLN2 and **GTF2I**) in two multiplex **autism** family cohorts revealed strong association of two **GTF2I** SNPs and their haplotype in Cohort 1 and the combined families. The risk alleles and haplotype were associated with severe problems in social interaction and excessive repetitive behaviors. Our findings suggest the **GTF2I** gene is important in the etiology of **autism** in individuals with this duplication and in non-duplication cases with severe social interaction problems and repetitive behaviors. https://www.ncbi.nlm.nih.gov/pubmed/?term=22048961

- Structural variation-associated expression changes are paralleled by chromatin architecture modifications. Gheldof N Witwicki RM Migliavacca E Leleu M Didelot G Harewood L Rougemont J Reymond A Copy number variants (CNVs) influence the expression of genes that map not only within the rearrangement, but also to its flanks. To assess the possible mechanism(s) underlying this "neighboring effect", we compared intrachromosomal interactions and histone modifications in cell lines of patients affected by genomic disorders and control individuals. Using chromosome conformation capture (4C-seq), we observed that a set of genes flanking the Williams-Beuren Syndrome critical region (WBSCR) were often looping together. The newly identified interacting genes include **AUTS2**, mutations of which are associated with **autism** and intellectual disabilities. Deletion of the WBSCR disrupts the expression of this group of flanking genes, as well as long-range interactions between them and the rearranged interval. We also pinpointed concomitant changes in histone modifications between samples. We conclude that large genomic rearrangements can lead to chromatin conformation changes that extend far away from the structural variant, thereby possibly modulating expression globally and modifying the phenotype.GSE33784, GSE33867. https://www.ncbi.nlm.nih.gov/pubmed/?term=24265791

- Smaller and larger deletions of the Williams Beuren syndrome region implicate genes involved in mild facial phenotype, epilepsy and autistic traits. Fusco C Micale L Augello B Teresa Pellico M Menghini D Alfieri P Cristina Digilio M Mandriani B Carella M Palumbo O Vicari S Merla G Williams Beuren syndrome (WBS) is a multisystemic disorder caused by a hemizygous deletion of 1.5 Mb on chromosome 7q11.23 spanning 28 genes. A few patients with larger and smaller WBS deletion have been reported. They show clinical features that vary between isolated SVAS to the full spectrum of WBS phenotype, associated with epilepsy or **autism spectrum** behavior. Here we describe four patients with atypical WBS 7q11.23 deletions. Two carry ~3.5 Mb larger deletion towards the telomere that includes Huntingtin-interacting protein 1 (**HIP1**) and tyrosine 3-monooxygenase/tryptophan 5-monooxigenase activation protein gamma (**YWHAG**) genes. Other two carry a shorter deletion of ~1.2 Mb at centromeric side that excludes the distal WBS genes **BAZ1B** and **FZD9**. Along with previously reported cases, genotype-phenotype correlation in the patients described here further suggests that haploinsufficiency of **HIP1** and **YWHAG** might cause the severe neurological and neuropsychological deficits including epilepsy and autistic traits, and that the preservation of **BAZ1B** and **FZD9** genes may be related to mild facial features and moderate neuropsychological deficits. This report highlights the importance to characterize additional patients with 7q11.23 atypical deletions comparing neuropsychological and clinical features between these individuals to shed light on the pathogenic role of genes within and flanking the WBS region. https://www.ncbi.nlm.nih.gov/pubmed/?term=23756441

- Multiple recurrent de novo CNVs, including duplications of the 7q11.23 Williams syndrome region, are strongly associated with **autism**. Sanders SJ Ercan-Sencicek AG Hus V Luo R Murtha MT Moreno-De-Luca D Chu SH Moreau MP Gupta AR Thomson SA Mason CE Bilguvar K Celestino-Soper PB Choi M Crawford EL Davis L Wright NR Dhodapkar RM DiCola M DiLullo NM Fernandez TV Fielding-Singh V Fishman DO Frahm S Garagaloyan R Goh GS Kammela S Klei L Lowe JK Lund SC McGrew AD Meyer KA Moffat WJ Murdoch JD O'Roak BJ Ober GT Pottenger RS Raubeson MJ Song Y Wang Q Yaspan BL Yu TW Yurkiewicz IR Beaudet AL Cantor RM Curland M Grice DE GÃ¼nel M Lifton RP Mane SM Martin DM Shaw CA Sheldon M Tischfield JA Walsh CA Morrow EM Ledbetter DH Fombonne E Lord C Martin CL Brooks AI Sutcliffe JS Cook EH Geschwind D Roeder K Devlin B State MW We have undertaken a genome-wide analysis of rare copy-number variation (CNV) in 1124 **autism spectrum disorder** (ASD) families, each comprised of a single proband, unaffected parents, and, in most kindreds, an unaffected sibling. We find significant association of ASD with de novo duplications of 7q11.23, where the reciprocal deletion causes Williams-Beuren syndrome, characterized by a highly social personality. We identify rare recurrent de novo CNVs at five additional regions, including 16p13.2 (encompassing genes **USP7** and C16orf72) and Cadherin 13, and implement a rigorous approach to evaluating the statistical significance of these observations. Overall, large de novo CNVs, particularly those encompassing multiple genes, confer substantial risks (OR = 5.6; CI = 2.6-12.0, p = 2.4 Ã— 10(-7)). We estimate there are 130-234 ASD-related CNV regions in the human genome and present compelling evidence, based on cumulative data, for association of rare de novo events at 7q11.23, 15q11.2-13.1, 16p11.2, and Neurexin 1.Copyright Â© 2011 Elsevier Inc. All rights reserved. https://www.ncbi.nlm.nih.gov/pubmed/?term=21658581

- Haploinsufficiency of **GTF2I**, a gene deleted in Williams Syndrome, leads to increases in social interactions. Sakurai T Dorr NP Takahashi N McInnes LA Elder GA Buxbaum JD Identifying genes involved in social behavior is important for **autism** research. Williams-Beuren syndrome (WBS) is a developmental syndrome with unique neurocognitive features, including low IQ, deficits in visuospatial and visual-motor abilities, hypersensitivity to sounds, hypersociability, and increased general anxiety. The syndrome is caused by a recurrent hemizygous deletion of the 7q11.23 region, containing about 28 genes. One of genes in the region, **GTF2I**, has been implicated in the hypersociability and visuospatial deficits of WBS based on genotype-phenotype correlation studies of patients with atypical deletions. In order to clarify the involvement of **GTF2I** in neurocognitive function, especially social behavior, we have developed and characterized **GTF2I**-deficient mice. ========= ========= Furthermore, heterozygous animals show no alterations in learning and memory, including spatial memory as assessed by the Morris water maze, but show alterations in the recognition of novel objects. Interestingly, they show increased social interaction with unfamiliar mice and do not show typical social habituation processes, reminiscent of the hypersociability observed in WBS patients. The mice do not appear to show increased anxiety, supporting a specific effect of **GTF2I** on defined domains of the WBS phenotype. These data indicate that **GTF2I** is involved in several aspects of embryonic development and the development of social neurocircuitry and that **GTF2I** haploinsufficiency could be a contributor to the hypersociability in WBS patients.Copyright Â© 2010, International Society for Autism Research, Wiley Periodicals, Inc. https://www.ncbi.nlm.nih.gov/pubmed/?term=21328569

- Association of mouse **Dlg4** (PSD-95) gene deletion and human **DLG4** gene variation with phenotypes relevant to a**autism spectrum disorders** and Williams' syndrome. Feyder M Karlsson RM Mathur P Lyman M Bock R Momenan R Munasinghe J Scattoni ML Ihne J Camp M Graybeal C Strathdee D Begg A Alvarez VA Kirsch P Rietschel M Cichon S Walter H Meyer-Lindenberg A Grant SG Holmes A Research is increasingly linking a**autism spectrum disorders** and other neurodevelopmental disorders to synaptic abnormalities ("synaptopathies"). PSD-95 (postsynaptic density-95, **DLG4**) orchestrates protein-protein interactions at excitatory synapses and is a major functional bridge interconnecting a neurexinneuroligin-SHANK pathway implicated in a**autism spectrum disorders**.The authors characterized behavioral, dendritic, and molecular phenotypic abnormalities relevant to a**autism spectrum disorders** in mice with PSD-95 deletion (**DLG4**â»/â»). The data from mice led to the identification of single-nucleotide polymorphisms (SNPs) in human**DLG4** and the examination of associations between these variants and neural signatures of Williams' syndrome in a normal population, using functional and structural neuroimaging.**DLG4**â»/â» showed increased repetitive behaviors, abnormal communication and social behaviors, impaired motor coordination, and increased stress reactivity and anxiety-related responses. **DLG4**â»/â» had subtle dysmorphology of amygdala dendritic spines and altered forebrain expression of various synaptic genes, including Cyln2, which regulates cytoskeletal dynamics and is a candidate gene for Williams' syndrome. A signifi-cant association was observed between variations in two human **DLG4**SNPs and reduced intraparietal sulcus volume and abnormal cortico-amygdala coupling, both of which characterize Williams' syndrome.These findings demonstrate that **DLG4** gene disruption in mice produces a complex range of behavioral and molecular abnormalities relevant to a**autism spectrum disorders** and Williams' syndrome. The study provides an initial link between human **DLG4** gene variation and key neural endophenotypes of Williams' syndrome and perhaps corticoamygdala regulation of emotional and social processes more generally. https://www.ncbi.nlm.nih.gov/pubmed/?term=20952458

- An atypical deletion of the Williams-Beuren syndrome interval implicates genes associated with defective visuospatial processing and **autism**. Edelmann L Prosnitz A Pardo S Bhatt J Cohen N Lauriat T Ouchanov L GonzÃ¡lez PJ Manghi ER Bondy P Esquivel M Monge S Delgado MF Splendore A Francke U Burton BK McInnes LA During a genetic study of **autism**, a female child who met diagnostic criteria for **autism spectrum disorder**, but also exhibited the cognitive-behavioural profile (CBP) associated with Williams-Beuren syndrome (WBS) was examined. The WBS CBP includes impaired visuospatial ability, an overly friendly personality, excessive non-social anxiety and language delay.Using array-based comparative genomic hybridisation (aCGH), a deletion corresponding to BAC RP11-89A20 in the distal end of the WBS deletion interval was detected. Hemizygosity was confirmed using fluorescence in situ hybridisation and fine mapping was performed by measuring the copy number of genomic DNA using quantitative polymerase chain reaction.The proximal breakpoint was mapped to intron 1 of **GTF2IRD1** and the distal breakpoint lies 2.4-3.1 Mb towards the telomere. The subject was completely hemizygous for **GTF2I**, commonly deleted in carriers of the classic approximately 1.5 Mb WBS deletion, and **GTF2IRD2**, deleted in carriers of the rare approximately 1.84 Mb WBS deletion.Hemizygosity of the GTF2 family of transcription factors is sufficient to produce many aspects of the WBS CBP, and particularly implicate the GTF2 transcription factors in the visuospatial construction deficit. Symptoms of **autism** in this case may be due to deletion of additional genes outside the typical WBS interval or remote effects on gene expression at other loci. https://www.ncbi.nlm.nih.gov/pubmed/?term=16971481

HYPER-SOCIABILITY (hiper sociabilidad)

	Enfermedad	Gene	Peso
1	hyper-sociability	BAZ1B	0.707107
2	hyper-sociability	LIMK1	0.707107
3	hyper-sociability	WBSCR27	0.707107
4	hyper-sociability	GTF2I	0.707107

- Activity of Genes with Functions in Human Williams-Beuren Syndrome Is Impacted by Mobile Element Insertions in the Gray Wolf Genome. vonHoldt BM Ji SS Aardema ML Stahler DR Udell MAR Sinsheimer JS In canines, transposon dynamics have been associated with a **hyper-social** behavioral syndrome, although the functional mechanism has yet to be described. We investigate the epigenetic and transcriptional consequences of these behavior-associated mobile element insertions (MEIs) in dogs and Yellowstone gray wolves. We posit that the transposons themselves may not be the causative feature; rather, their transcriptional regulation may exert the functional impact. We survey four outlier transposons associated with **hyper-sociability**, with the expectation that they are targeted for epigenetic silencing. We predict hyper-methylation of MEIs, suggestive that the epigenetic silencing of and not the MEIs themselves may be driving dysregulation of nearby genes. We found that transposon-derived sequences are significantly hyper-methylated, regardless of their copy number or species. Further, we have assessed transcriptome sequence data and found evidence that MEIs impact the expression levels of six genes (**WBSCR17, LIMK1, GTF2I, WBSCR27, BAZ1B, and BCL7B**), all of which have known roles in human Williams-Beuren syndrome due to changes in copy number, typically hemizygosity. Although further evidence is needed, our results suggest that a few insertions alter local expression at multiple genes, likely through a cis-regulatory mechanism that excludes proximal methylation. https://www.ncbi.nlm.nih.gov/pubmed/?term=29860323

AORTIC STENOSIS (estenosis aórtica)

	Enfermedad	Gene	Peso
1	aortic stenosis	HR	0.997677
2	aortic stenosis	NHS	0.997667
3	aortic stenosis	BCL7B	0.997657
4	aortic stenosis	SERPINA1	0.997604
5	aortic stenosis	WBSCR22	0.997604
6	aortic stenosis	MMP9	0.997604
7	aortic stenosis	TIMP1	0.997604
8	aortic stenosis	TIMP2	0.997604
9	aortic stenosis	YWHAG	0.997571
10	aortic stenosis	AUTS2	0.997571
11	aortic stenosis	CALN1	0.997571
12	aortic stenosis	HIP1	0.997571
13	aortic stenosis	FKBP6	0.860299

- Alpha 1 antitrypsin deficiency alleles are associated with joint dislocation and scoliosis in Williams syndrome. Morris CA Pani AM Mervis CB Rios CM Kistler DJ Gregg RG Elastin haploinsufficiency is responsible for a significant portion of the Williams syndrome (WS) phenotype including hoarse voice, supravalvar **aortic stenosis** (SVAS), hernias, diverticuli of bowel and bladder, soft skin, and joint abnormalities. All of the connective tissue signs and symptoms are variable in the WS population, but few factors other than age and gender are known to influence the phenotype. We examined a cohort of 205 individuals with WS for mutations in **SERPINA1**, the gene that encodes alpha-1-antitrypsin (AAT), the inhibitor of elastase. Individuals with classic WS deletions and **SERPINA1** genotypes PiMS or PiMZ were more likely than those with a **SERPINA1** PiMM genotype to have joint dislocation or scoliosis. However, carrier status for AAT deficiency was not correlated with presence of inguinal hernia or with presence or severity of SVAS. These findings suggest that genes important in elastin metabolism are candidates for variability in the connective tissue abnormalities in WS. https://www.ncbi.nlm.nih.gov/pubmed/?term=20425789
- Zebrafish gene knockdowns imply roles for human **YWHAG** in infantile spasms and cardiomegaly. Komoike Y Fujii K Nishimura A Hiraki Y Hayashidani M Shimojima K Nishizawa T Higashi K Yasukawa K Saitsu H Miyake N Mizuguchi T Matsumoto N Osawa M Kohno Y Higashinakagawa T Yamamoto T Williams-Beuren syndrome (WBS) is a neurodevelopmental disorder presenting with an elfin-like face, supravalvular **aortic stenosis**, a specific cognitive-behavioral profile, and infantile hypercalcemia. We encountered two WBS patients presenting with infantile spasms, which is extremely rare in WBS. Array comparative genomic hybridization (aCGH) and fluorescent in situ hybridization (FISH) analyses revealed atypical 5.7-Mb and 4.1-Mb deletions at 7q11.23 in the two patients, including the WBS critical region and expanding into the proximal side and the telomeric side, respectively. On the proximal side, **AUTS2** and **CALN1** may contribute to the phenotype. On the telomeric side, there are two candidate genes **HIP1** and **YWHAG**. Because detailed information of them was unavailable, we investigated their functions using gene knockdowns of zebrafish. When zebrafish **YWHAG**1 was knocked down, reduced brain size and increased diameter of the heart tube were observed, indicating that the infantile spasms and cardiomegaly seen in the patient with the telomeric deletion may be derived from haploinsufficiency of **YWHAG**.2010 Wiley-Liss, Inc. https://www.ncbi.nlm.nih.gov/pubmed/?term=20146355

- Vascular wall remodeling in patients with supravalvular **aortic stenosis** and Williams Beuren syndrome. Dridi SM Foucault Bertaud A Igondjo Tchen S Senni K Ejeil AL Pellat B Lyonnet S Bonnet D Charpiot P Godeau G Supravalvular **aortic stenosis** (SVAS) and Williams Beuren syndrome (WBS) can be considered as inherited diseases affecting the whole arterial tree and causing narrowing of the vessels. It has been reported that abnormal deposition of elastin in arterial walls of patients with SVAS and WBS leads to increased proliferation of arterial smooth muscle cells (SMC), which result in the formation of hyperplastic intimal lesions. In this work, we conducted morphological and morphometrical analysis with stenotic aortas from patients suffering from SVAS and WBS and from healthy control subjects and demonstrated that the amount of elastic fibers and the loss of integrity of vascular elastic fibers in the aortas reflect similar changes in the skin of patients with SVAS or WBS, as reported in our previous work conducted on skin in these pathological states. On the other hand, we conducted investigations on metalloproteinases (MMP2, **MMP9**, MMP7) and their specific tissue inhibitors **TIMP1** and **TIMP2** to verify their possible involvement in the etiopathogeny of SVAS and WBS. We particularly evidenced an altered **MMP9/TIMP1** balance in favor of matrix degradation which could facilitate SMC migration and neointimal hyperplasia. Our findings suggest that elastinolytic enzymes secreted by arterial SMC, possibly including matrilysin 1, are critical for the development of arterial lesions in SVAS and WBS and contribute to perpetuate arterial stenosis in either SVAS or WBS.Copyright (c) 2005 S. Karger AG, Basel. https://www.ncbi.nlm.nih.gov/pubmed/?term=15832055

IMPAIRMENTS (discapacidad)

	Enfermedad	Gene	Peso
1	impairments	KY	0.97572
2	impairments	NGF	0.974937
3	impairments	COMT	0.958532
4	impairments	GTF2IRD2	0.811272
5	impairments	GTF2I	0.810696
6	impairments	GTF2IRD1	0.810408
7	impairments	ABR	0.807834

- The role of **GTF2IRD1** in the auditory pathology of Williams-Beuren Syndrome. Canales CP Wong AC Gunning PW Housley GD Hardeman EC Palmer SJ Williams-Beuren Syndrome (WBS) is a rare genetic condition caused by a hemizygous deletion involving up to 28 genes within chromosome 7q11.23. Among the spectrum of physical and neurological defects in WBS, it is common to find a distinctive response to sound stimuli that includes extreme adverse reactions to loud, or sudden sounds and a fascination with certain sounds that may manifest as strengths in musical ability. However, hearing tests indicate that sensorineural hearing loss (SNHL) is frequently found in WBS patients. The functional and genetic basis of this unusual auditory phenotype is currently unknown. Here, we investigated the potential involvement of **GTF2IRD1**, a transcription factor encoded by a gene located within the WBS deletion that has been implicated as a contributor to the WBS assorted neurocognitive profile and craniofacial abnormalities. Using **GTF2IRD1** knockout mice, we have analysed the expression of the gene in the inner ear and examined hearing capacity by evaluating the auditory brainstem response (**ABR**) and the distortion product of otoacoustic emissions (DPOAE). Our results show that **GTF2IRD1** is expressed in a number of cell types within the cochlea, and **GTF2IRD1** null mice showed higher auditory thresholds (hypoacusis) in both **ABR** and DPOAE hearing assessments. These data indicate that the principal hearing deficit in the mice can be traced to **impairments** in the amplification process mediated by the outer hair cells and suggests that similar mechanisms may underpin the SNHL experienced by WBS patients. https://www.ncbi.nlm.nih.gov/pubmed/?term=25248400
- [Williams syndrome]. Alleva E Cirulli F Calamandrei G Rondinini C Capirci O Aloe L Volterra V Williams syndrome (WS) is a rare (2-5/100,000) genetic human disorder characterised by a typical facies and mental retardation with a deficit in the visuospatial cognitive function and a relative preservation of linguistic abilities in general, and spoken language in particular. This syndrome also includes morphological anomalies, metabolic functional **impairments**, and likely deficits in the pattern of brain ontogenesis. The genetic basis of WS, recently identified, are presented. A cognitive profile of the WS individuals is defined and compared to Down syndrome (DS) and autism cognitive profiles. Neuroanatomical features of WS, including a reduction in brain volume, preservation of cerebellum and frontal lobes, and a reduction of posterior cortical systems, are described. The possible role of **NGF** (nerve growth factor)--a neurotrophin involved in the development of brain cholinergic systems and the associated behavioural functions--in the aetiology of the typical mental retardation of WS patients, is critically discussed. Future research avenues, including the identification of potential neurobiological markers in order to precociously diagnose this syndrome, are reviewed. https://www.ncbi.nlm.nih.gov/pubmed/?term=10645654
- Anxious, hypoactive phenotype combined with motor deficits in **GTF2IRD1** null mouse model relevant to Williams syndrome. Schneider T Skitt Z Liu Y Deacon RM Flint J Karmiloff-Smith A Rawlins JN Tassabehji M Williams-Beuren syndrome (WBS) is a rare genetic disorder caused by a hemizygous deletion of around 28 genes on the long arm of chromosome 7 (7q11.23), characterized by a unique spectrum of behavioral **impairments**, including mental retardation, deficits in visuospatial constructive cognition, hypersociability, anxiety and simple phobias. Physical characteristics include dysmorphic faces, short stature, oculomotor deficits, gross and fine coordination **impairments**, diminished control of balance and mild extrapyramidal signs as well as gait abnormalities resembling gait hypokinesia. Genes near the distal deletion breakpoint appear to contribute most to the WBS cognitive and behavioral profile and include the GTF family of transcription factors: **GTF2I**, **GTF2IRD1**, **GTF2IRD2**. We have previously shown that heterozygous deletions of **GTF2IRD1** in humans and homozygous deletion in mice contributes to craniofacial abnormalities. Here we show an important role of this gene in motor coordination and anxiety ascertained from extensive behavioral mouse phenotyping. **GTF2IRD1** null mice showed lower body weight, decreased spontaneous and circadian locomotor activity, diminished motor coordination and strength, gait abnormalities, increased anxiety and an elevated endocrinological response to stress. **GTF2IRD1** heterozygous mice displayed lower body weight and decreased circadian activity, but only minor motor coordination and anxiety-related behavioral dysfunctions. Our study strongly supports a role for **GTF2IRD1** in the motoric and anxiety-related abnormalities seen in Williams-Beuren syndrome, and suggests basal ganglia and potentially cerebellar abnormalities in **GTF2IRD1** mice.Copyright Â© 2012 Elsevier B.V. All rights reserved. https://www.ncbi.nlm.nih.gov/pubmed/?term=22652393

- A comparative study of hearing loss in two microdeletion syndromes: velocardiofacial (22q11.2 deletion) and Williams (7q11.23 deletion) syndromes. Zarchi O Attias J Raveh E Basel-Vanagaite L Saporta L Gothelf D To comprehensively assess auditory **impairments** in velocardiofacial syndrome (VCFS) and Williams syndrome (WS).Audiologic measurements were conducted with 62 subjects with VCFS and 44 subjects with WS, as well as two control groups consisting of 22 subjects with idiopathic developmental disability and 23 typically developing controls. An association between severity of hearing loss in VCFS and the (158)Val/Met polymorphism of the catechol-O-methyltransferase gene (**COMT**) was explored.Hearing was significantly more impaired in the VCFS and WS groups compared with the developmental disability and typically developing groups. Audiologic abnormalities identified in both the VCFS and WS groups included high-tone hearing loss (predominantly sensorineural or mixed type), loss of acoustic reflex, and middle ear pathologies. In both the VCFS and WS groups, hearing loss severity was positively correlated with age. In the VCFS group, hearing loss was more severe in the subgroup carrying the **COMT** Val allele compared with the subgroup carrying the **COMT** Met allele.Hearing **impairments**, including sensorineural hearing loss and acoustic reflex dysfunction, are very common in both VCFS and WS. Hearing loss is less severe in subjects with the **COMT** Met allele, possibly due to the protective effect of dopamine on the hearing system.Copyright Â© 2011 Mosby, Inc. All rights reserved. https://www.ncbi.nlm.nih.gov/pubmed/?term=20846670

GENETIC DISORDER (trastorno genético)

	Enfermedad	Gene	Peso
1	genetic disorder	ARC	0.97812
2	genetic disorder	CYP24A1	0.97812
3	genetic disorder	KY	0.977538
4	genetic disorder	RHD	0.976298
5	genetic disorder	HR	0.972683
6	genetic disorder	LIMK1	0.965942
7	genetic disorder	GCK	0.75981

- Importance of dietary calcium and vitamin D in the treatment of hypercalcaemia in Williams-Beuren syndrome. Lameris AL Geesing CL Hoenderop JG Schreuder MF Williams-Beuren syndrome (WBS) is a rare **genetic disorder** caused by the deletion of 26-28 genes on chromosome 7. Fifteen percent of WBS patients present with hypercalcaemia during infancy, which is generally mild and resolves spontaneously before the age of 4 years. The mechanisms underlying the transient hypercalcaemia in WBS are poorly understood.We report a case of severe symptomatic hypercalcaemia in a patient with WBS, in which treatment with mild calcium restriction, hyperhydration and repeated bisphosphonate administration only resulted in short-lasting effects. Long-term lowering of serum calcium was only achieved after reducing calcium and vitamin D intake to the bare minimum.This case illustrates the potential severity of hypercalcaemia in WBS, and demonstrates that both the cause as well as the solution of this problem may be found in the intestinal absorption of calcium. We hypothesise that the phenotypical resemblance between WBS and transient idiopathic infantile hypercalcaemia can be explained by similarities in the underlying genetic defect. Patients suffering from transient infantile hypercalcaemia were recently described to have mutations in **CYP24A1**, the key enzyme in 1,25-dihydroxyvitamin D3 degradation. In the light of this new development we discuss the role of one of the deleted genes in WBS, Williams syndrome transcription factor (WSTF), in the etiology of hypercalcaemia in WBS. https://www.ncbi.nlm.nih.gov/pubmed/?term=24572979

- Inefficient search of large-scale space in Williams syndrome: further insights on the role of **LIMK1** deletion in deficits of spatial cognition. Smith AD Gilchrist ID Hood B Tassabehji M Karmiloff-Smith A Williams syndrome (WS) is a **genetic disorder** associated with impairments of spatial cognition. This has primarily been studied in small-scale space, and rarely in large-scale environments. In order to fully characterise the spatial deficits in WS, and also to address claims that the deletion of LIM-kinase 1 (**LIMK1**) on chromosome 7 is responsible for those deficits, we report an automated large-scale search task for humans that places the participant egocentrically within the search space. Search locations were defined as lights and switches embedded in the floor, and participants attempted to locate a hidden target by pressing the switch at potential locations. We compared individuals with WS to patients with smaller deletions (including **LIMK1**) in the critical region on chromosome 7. Whilst partial-deletion participants performed efficiently on the task, participants with WS demonstrated inefficient search profiles: their search slopes were steeper and they made significantly more erroneous revisits to previously inspected locations. Our findings indicate that spatial deficits associated with WS also affect large-scale spatial processing and suggest that hemizygous deletion of **LIMK1** is not sufficient to account for any of the spatial deficits associated with WS. https://www.ncbi.nlm.nih.gov/pubmed/?term=19662944

INTELLECTUAL DISABILITY (discapacidad intelectual)

	Enfermedad	Gene	Peso
1	intellectual disability	KY	0.994617
2	intellectual disability	ELN	0.987881

- A new diagnosis of Williams-Beuren syndrome in a 49-year-old man with severe bullous emphysema. Wojcik MH Carmichael N Bieber FR Wiener DC Madan R Pober BR Raby BA Williams-Beuren syndrome (WBS) is a chromosomal microdeletion syndrome typically presenting with **intellectual disability**, a unique personality, a characteristic facial appearance, and cardiovascular disease. Several clinical features of WBS are thought to be due to haploinsufficiency of elastin (**ELN**), as the **ELN** locus is included within the WBS critical region at 7q11.23. Emphysema, a disease attributed to destruction of pulmonary elastic fibers, has been reported in patients without WBS who have pathogenic variants in **ELN** but only once (in one patient) in WBS. Here we report a second adult WBS patient with emphysema where the diagnosis of WBS was established subsequent to the discovery of severe bullous emphysema. Haploinsufficiency of **ELN** likely contributed to this pulmonary manifestation of WBS. This case emphasizes the contribution of rare genetic variation in cases of severe emphysema and provides further evidence that emphysema should be considered in patients with WBS who have respiratory symptoms, as it may be under-recognized in this patient population.Â© 2017 Wiley Periodicals, Inc. https://www.ncbi.nlm.nih.gov/pubmed/?term=28574231

NEURODEVELOPMENTAL DISORDERS (trastornos del neurodesarrollo)

	Enfermedad	Gene	Peso
1	neurodevelopmental disorder	KY	0.987702
2	neurodevelopmental disorder	BCL7B	0.974831
3	neurodevelopmental disorder	MLXIPL	0.974831
4	neurodevelopmental disorder	FZD9	0.974572
5	neurodevelopmental disorder	CTCF	0.973763
6	neurodevelopmental disorder	AUTS2	0.971668
7	neurodevelopmental disorder	CALN1	0.971668
8	neurodevelopmental disorder	HIP1	0.971668
9	neurodevelopmental disorder	YWHAG	0.971668
10	neurodevelopmental disorder	AVP	0.966753
11	neurodevelopmental disorder	BAZ1B	0.920764
12	neurodevelopmental disorder	STX1A	0.883707
13	neurodevelopmental disorder	GC	0.874703
14	neurodevelopmental disorder	DLG4	0.849064
15	neurodevelopmental disorder	NCF1	0.849064
16	neurodevelopmental disorder	CA1	0.803658

- Williams syndrome and mature B-Leukemia: A random association? Decimi V Fazio G Dell'Acqua F Maitz S Galbiati M Rizzari C Biondi A Cazzaniga G Selicorni A Williams syndrome (WBS) is a rare **neurodevelopmental disorder** with specific phenotypic characteristics and cardiac abnormalities, but is not considered as a cancer predisposing condition. However, in rare cases, malignancies have been described in patients with WBS, with hematologic cancer (mainly Burkitt Lymphoma and Acute Lymphoblastic Leukemia) as the most represented. We report here the case of a boy with WS and B-NHL. This is the unique case within the large cohort of patients (nÂ =Â 117) followed in our institution for long time (mean clinical follow-up, 13 years). We herewith propose that the **BCL7B** gene, located in the chromosomal region commonly deleted in Williams syndrome, could potentially have a role in this particular association.Copyright Ã‚Â© 2016 Elsevier Masson SAS. All rights reserved. https://www.ncbi.nlm.nih.gov/pubmed/?term=27771473

- Detection of an atypical 7q11.23 deletion in Williams syndrome patients which does not include the **STX1A** and **FZD3** genes. Botta A Novelli G Mari A Novelli A Sabani M Korenberg J Osborne LR Digilio MC Giannotti A Dallapiccola B We present two patients with the full Williams syndrome (WS) phenotype carrying a smaller deletion than typically observed. The deleted region spans from the elastin gene to marker D7S1870. This observation narrows the minimal region of deletion in WS and suggests that the syntaxin 1A and frizzled genes are not responsible for the major features of this **developmental disorder** and provides important insight into understanding the genotype-phenotype correlation in WS. https://www.ncbi.nlm.nih.gov/pubmed/?term=10874638

- A human neurodevelopmental model for Williams syndrome. Chailangkarn T Trujillo CA Freitas BC Hrvoj-Mihic B Herai RH Yu DX Brown TT Marchetto MC Bardy C McHenry L Stefanacci L JÃ¤rvinen A Searcy YM DeWitt M Wong W Lai P Ard MC Hanson KL Romero S Jacobs B Dale AM Dai L Korenberg JR Gage FH Bellugi U Halgren E Semendeferi K Muotri AR Williams syndrome is a genetic **neurodevelopmental disorder** characterized by an uncommon hypersociability and a mosaic of retained and compromised linguistic and cognitive abilities. Nearly all clinically diagnosed individuals with Williams syndrome lack precisely the same set of genes, with breakpoints in chromosome band 7q11.23 (refs 1-5). The contribution of specific genes to the neuroanatomical and functional alterations, leading to behavioural pathologies in humans, remains largely unexplored. Here we investigate neural progenitor cells and cortical neurons derived from Williams syndrome and typically developing induced pluripotent stem cells. Neural progenitor cells in Williams syndrome have an increased doubling time and apoptosis compared with typically developing neural progenitor cells. Using an individual with atypical Williams syndrome, we narrowed this cellular phenotype to a single gene candidate, frizzled 9 (**FZD9**). At the neuronal stage, layer V/VI cortical neurons derived from Williams syndrome were characterized by longer total dendrites, increased numbers of spines and synapses, aberrant calcium oscillation and altered network connectivity. Morphometric alterations observed in neurons from Williams syndrome were validated after Golgi staining of post-mortem layer V/VI cortical neurons. This model of human induced pluripotent stem cells fills the current knowledge gap in the cellular biology of Williams syndrome and could lead to further insights into the molecular mechanism underlying the disorder and the human social brain. https://www.ncbi.nlm.nih.gov/pubmed/?term=27509850

- Haploinsufficiency of **BAZ1B** contributes to Williams syndrome through transcriptional dysregulation of neurodevelopmental pathways. Lalli MA Jang J Park JH Wang Y Guzman E Zhou H Audouard M Bridges D Tovar KR Papuc SM Tutulan-Cunita AC Huang Y Budisteanu M Arghir A Kosik KS Williams syndrome (WS) is a **neurodevelopmental disorder** caused by a genomic deletion of â^¼28 genes that results in a cognitive and behavioral profile marked by overall intellectual impairment with relative strength in expressive language and hypersocial behavior. Advancements in protocols for neuron differentiation from induced pluripotent stem cells allowed us to elucidate the molecular circuitry underpinning the ontogeny of WS. In patient-derived stem cells and neurons, we determined the expression profile of the Williams-Beuren syndrome critical region-deleted genes and the genome-wide transcriptional consequences of the hemizygous genomic microdeletion at chromosome 7q11.23. Derived neurons displayed disease-relevant hallmarks and indicated novel aberrant pathways in WS neurons including over-activated Wnt signaling accompanying an incomplete neurogenic commitment. We show that haploinsufficiency of the ATP-dependent chromatin remodeler, **BAZ1B**, which is deleted in WS, significantly contributes to this differentiation defect. ========= ========= Altogether, these results reveal a pivotal role for **BAZ1B** in neurodevelopment and implicate its haploinsufficiency as a likely contributor to the neurological phenotypes in WS. Â© The Author 2016. Published by Oxford University Press. All rights reserved. For Permissions, please email: journals.permissions@oup.com. https://www.ncbi.nlm.nih.gov/pubmed/?term=26755828

- Metabolic abnormalities in Williams-Beuren syndrome. Palacios-VerdÃº MG Segura-Puimedon M Borralleras C Flores R Del Campo M Campuzano V PÃ©rez-Jurado LA Williams-Beuren syndrome (WBS, OMIM-194050) is a **neurodevelopmental disorder** with multisystemic manifestations caused by a 1.55-1.83 Mb deletion at 7q11.23 including 26-28 genes. Reported endocrine and metabolic abnormalities include transient hypercalcaemia of infancy, subclinical hypothyroidism in âˆ¼ 30% of children and impaired glucose tolerance in âˆ¼ 75% of adult individuals. The purpose of this study was to further study metabolic alterations in patients with WBS, as well as in several mouse models, to establish potential candidate genes.We analysed several metabolic parameters in a cohort of 154 individuals with WBS (data available from 69 to 151 cases per parameter), as well as in several mouse models with complete and partial deletions of the orthologous WBS locus, and searched for causative genes and potential modifiers.Triglyceride plasma levels were significantly decreased in individuals with WBS while cholesterol levels were slightly decreased compared with controls. Hyperbilirubinemia, mostly unconjugated, was found in 18.3% of WBS cases and correlated with subclinical hypothyroidism and hypotriglyceridemia, suggesting common pathogenic mechanisms. Haploinsufficiency at **MLXIPL** and increased penetrance for hypomorphic alleles at the UGT1A1 gene promoter might underlie the lipid and bilirubin alterations. Other disturbances included increased protein and iron levels, as well as the known subclinical hypothyroidism and glucose intolerance.Our results show that several unreported biochemical alterations, related to haploinsufficiency for specific genes at 7q11.23, are relatively common in WBS. The early diagnosis, follow-up and management of these metabolic disturbances could prevent long-term complications in this disorder.Published by the BMJ Publishing Group Limited. For permission to use (where not already granted under a licence) please go to http://group.bmj.com/group/rights-licensing/permissions. https://www.ncbi.nlm.nih.gov/pubmed/?term=25663682
- Oxytocin and vasopressin systems in genetic syndromes and **neurodevelopmental disorder**s. Francis SM Sagar A Levin-Decanini T Liu W Carter CS Jacob S Oxytocin (OT) and arginine vasopressin (**AVP**) are two small, related neuropeptide hormones found in many mammalian species, including humans. Dysregulation of these neuropeptides have been associated with changes in behavior, especially social interactions. We review how the OT and **AVP** systems have been investigated in Autism Spectrum Disorder (ASD), Prader-Willi Syndrome (PWS), Williams Syndrome (WS) and Fragile X syndrome (FXS). All of these **neurodevelopmental disorder**s (NDD) are marked by social deficits. While PWS, WS and FXS have identified genetic mutations, ASD stems from multiple genes with complex interactions. ========= ========= This has implications for ongoing studies of the therapeutic application of OT in NDD. This article is part of a Special Issue entitled Oxytocin and Social Behav. Copyright Â© 2014. Published by Elsevier B.V. https://www.ncbi.nlm.nih.gov/pubmed/?term=24462936
- Functional and genetic characterization of two extremely rare cases of Williams-Beuren syndrome associated with chronic granulomatous disease. Stasia MJ Mollin M Martel C Satre V Coutton C Amblard F Vieville G van Montfrans JM Boelens JJ Veenstra-Knol HE van Leeuwen K de Boer M Brion JP Roos D Williams-Beuren syndrome (WBS) is a **neurodevelopmental disorder** with multi-systemic manifestations, caused by a heterozygous segmental deletion of 1.55-1.83â€‰Mb at chromosomal band 7q11.23. The deletion can include the **NCF1** gene that encodes the p47(phox) protein, a component of the leukocyte NADPH oxidase enzyme, which is essential for the defense against microbial pathogens. It has been postulated that WBS patients with two functional **NCF1** genes are more susceptible to occurrence of hypertension than WBS patients with only one functional **NCF1** gene. We now describe two extremely rare WBS patients without any functional **NCF1** gene, because of a mutation in **NCF1** on the allele not carrying the **NCF1**-removing WBS deletion. These two patients suffer from chronic granulomatous disease with increased microbial infections in addition to WBS. Interestingly, one of these patients did suffer from hypertension, indicating that other factors than NADPH oxidase in vascular tissue may be involved in causing hypertension. https://www.ncbi.nlm.nih.gov/pubmed/?term=23340515
- Intelligence in Williams Syndrome is related to **STX1A**, which encodes a component of the presynaptic SNARE complex. Gao MC Bellugi U Dai L Mills DL Sobel EM Lange K Korenberg JR Although genetics is the most significant known determinant of human intelligence, specific gene contributions remain largely unknown. To accelerate understanding in this area, we have taken a new approach by studying the relationship between quantitative gene expression and intelligence in a cohort of 65 patients with Williams Syndrome (WS), a **neurodevelopmental disorder** caused by a 1.5 Mb deletion on chromosome 7q11.23. We find that variation in the transcript levels of the brain gene **STX1A** correlates significantly with intelligence in WS patients measured by principal component analysis (PCA) of standardized WAIS-R subtests, r = 0.40 (Pearson correlation, Bonferroni corrected p-value = 0.007), accounting for 15.6% of the cognitive variation. These results suggest that syntaxin 1A, a neuronal regulator of presynaptic vesicle release, may play a role in WS and be a component of the cellular pathway determining human intelligence. https://www.ncbi.nlm.nih.gov/pubmed/?term=20422020
- Association of mouse **DLG4** (PSD-95) gene deletion and human **DLG4** gene variation with phenotypes relevant to autism spectrum disorders and Williams' syndrome. Feyder M Karlsson RM Mathur P Lyman M Bock R Momenan R Munasinghe J Scattoni ML Ihne J Camp M Graybeal C Strathdee D Begg A Alvarez VA Kirsch P Rietschel M Cichon S Walter H Meyer-Lindenberg A Grant SG Holmes A Research is increasingly linking autism spectrum disorders and other **neurodevelopmental disorder**s to synaptic abnormalities ("synaptopathies"). PSD-95 (postsynaptic density-95, **DLG4**) orchestrates protein-protein interactions at excitatory synapses and is a major functional bridge interconnecting a neurexinneuroligin-SHANK pathway implicated in autism spectrum disorders.The authors characterized behavioral, dendritic, and molecular phenotypic abnormalities relevant to autism spectrum disorders in mice with PSD-95 deletion (**DLG4**â»(/)â»). The data from mice led to the identification of single-nucleotide polymorphisms (SNPs) in human **DLG4** and the examination of associations between these variants and neural signatures of Williams' syndrome in a normal population, using functional and structural neuroimaging.**DLG4**â»(/)â» showed increased repetitive behaviors, abnormal communication and social behaviors, impaired motor coordination, and increased stress reactivity and anxiety-related responses. **DLG4**â»(/)â» had subtle dysmorphology of amygdala dendritic spines and altered forebrain expression of various synaptic genes, including Cyln2, which regulates cytoskeletal dynamics and is a candidate gene for Williams' syndrome. A signifi-cant association was observed between variations in two human **DLG4** SNPs and reduced intraparietal sulcus volume and abnormal cortico-amygdala coupling, both of which characterize Williams' syndrome.These findings demonstrate that **DLG4** gene disruption in mice produces a complex range of behavioral and molecular abnormalities relevant to autism spectrum disorders and Williams' syndrome. The study provides an initial link between human **DLG4** gene variation and key neural endophenotypes of Williams' syndrome and perhaps corticoamygdala regulation of emotional and social processes more generally. https://www.ncbi.nlm.nih.gov/pubmed/?term=20952458

- Zebrafish gene knockdowns imply roles for human **YWHAG** in infantile spasms and cardiomegaly. Komoike Y Fujii K Nishimura A Hiraki Y Hayashidani M Shimojima K Nishizawa T Higashi K Yasukawa K Saitsu H Miyake N Mizuguchi T Matsumoto N Osawa M Kohno Y Higashinakagawa T Yamamoto T Williams-Beuren syndrome (WBS) is a **neurodevelopmental disorder** presenting with an elfin-like face, supravalvular aortic stenosis, a specific cognitive-behavioral profile, and infantile hypercalcemia. We encountered two WBS patients presenting with infantile spasms, which is extremely rare in WBS. Array comparative genomic hybridization (aCGH) and fluorescent in situ hybridization (FISH) analyses revealed atypical 5.7-Mb and 4.1-Mb deletions at 7q11.23 in the two patients, including the WBS critical region and expanding into the proximal side and the telomeric side, respectively. On the proximal side, **AUTS2** and **CALN1** may contribute to the phenotype. On the telomeric side, there are two candidate genes **HIP1** and **YWHAG**. Because detailed information of them was unavailable, we investigated their functions using gene knockdowns of zebrafish. When zebrafish **YWHAG**1 was knocked down, reduced brain size and increased diameter of the heart tube were observed, indicating that the infantile spasms and cardiomegaly seen in the patient with the telomeric deletion may be derived from haploinsufficiency of **YWHAG**.2010 Wiley-Liss, Inc. https://www.ncbi.nlm.nih.gov/pubmed/?term=20146355
- Frizzled 9 knock-out mice have abnormal B-cell development. Ranheim EA Kwan HC Reya T Wang YK Weissman IL Francke U The binding of frizzled (Fzd) receptors by their Wnt ligands results in the inhibition of beta-catenin degradation and subsequent transcription of beta-catenin/LEF-inducible genes. The beta-catenin pathway is known to be involved in development, tumorigenesis, and stem cell self-renewal. In humans, the **FZD9** gene lies in the region of chromosome 7q11.23 deleted in the **neurodevelopmental disorder**, Williams-Beuren syndrome (WBS). **FZD9**-/- mice show no obvious features of WBS, but reveal a role for **FZD9** in lymphoid development and maturation. **FZD9**-/- mice show pronounced splenomegaly, thymic atrophy, and lymphadenopathy with age, with accumulation of plasma cells in lymph nodes. There is a depletion of developing B cells in the bone marrow (BM), particularly in the pre-B stage where immunoglobulin heavy chains are expressed and the cells are undergoing clonal expansion prior to light chain rearrangement. The pre-B defect is partially intrinsic to the hematopoietic system; as in competitive BM reconstitution studies, **FZD9**-/- -derived BM exhibits defective B-cell development when implanted into a wild-type host. Mature B cells are present in normal numbers in lymph node and spleen. These findings suggest a role for **FZD9** signaling in lymphoid development, particularly at points where B cells undergo self-renewal prior to further differentiation. https://www.ncbi.nlm.nih.gov/pubmed/?term=15572594

DEATH (muerte)

	Enfermedad	Gene	Peso
1	death	ELN	0.997648
2	death	APC	0.997217
3	death	BRCA2	0.997217
4	death	BRCA1	0.997217
5	death	NCF1	0.997217

- mTOR (Mechanistic Target of Rapamycin) Inhibition Decreases Mechanosignaling, Collagen Accumulation, and Stiffening of the Thoracic Aorta in Elastin-Deficient Mice. Jiao Y Li G Li Q Ali R Qin L Li W Qyang Y Greif DM Geirsson A Humphrey JD Tellides G Elastin deficiency because of heterozygous loss of an **ELN** allele in Williams syndrome causes obstructive aortopathy characterized by medial thickening and fibrosis and consequent aortic stiffening. Previous work in **ELN**-null mice with a severe arterial phenotype showed that inhibition of mTOR (mechanistic target of rapamycin), a key regulator of cell growth, lessened the aortic obstruction but did not prevent early postnatal **death**. We investigated the effects of mTOR inhibition in **ELN**-null mice partially rescued by human **ELN** that manifest a less severe arterial phenotype and survive long term.Thoracic aortas of neonatal and juvenile mice with graded elastin deficiency exhibited increased signaling through both mTOR complex 1 and 2. Despite lower predicted wall stress, there was increased phosphorylation of focal adhesion kinase, suggestive of greater integrin activation, and increased transforming growth factor-Î²-signaling mediators, associated with increased collagen expression. Pharmacological blockade of mTOR by rapalogs did not improve luminal stenosis but reduced mechanosignaling (in delayed fashion after mTOR complex 1 inhibition), medial collagen accumulation, and stiffening of the aorta. Rapalog administration also retarded somatic growth, however, and precipitated neonatal **death**s. Complementary, less-toxic strategies to inhibit mTOR via altered growth factor and nutrient responses were not effective.In addition to previously demonstrated therapeutic benefits of rapalogs decreasing smooth muscle cell proliferation in the absence of elastin, we find that rapalogs also prevent aortic fibrosis and stiffening attributable to partial elastin deficiency. Our findings suggest that mTOR-sensitive perturbation of smooth muscle cell mechanosensing contributes to elastin aortopathy.Â© 2017 American Heart Association, Inc. https://www.ncbi.nlm.nih.gov/pubmed/?term=28751568

- Williams syndrome predisposes to vascular stiffness modified by antihypertensive use and copy number changes in **NCF1**. Kozel BA Danback JR Waxler JL Knutsen RH de Las Fuentes L Reusz GS Kis E Bhatt AB Pober BR Williams syndrome is caused by the deletion of 26 to 28 genes, including elastin, on human chromosome 7. Elastin insufficiency leads to the cardiovascular hallmarks of this condition, namely focal stenosis and hypertension. Extrapolation from the **ELN**(+/-) mouse suggests that affected people may also have stiff vasculature, a risk factor for stroke, myocardial infarction, and cardiac **death**. **NCF1**, one of the variably deleted Williams genes, is a component of the nicotinamide adenine dinucleotide phosphate (NADPH) oxidase complex and is involved in the generation of oxidative stress, making it an interesting candidate modifier for vascular stiffness. Using a case-control design, vascular stiffness was evaluated by pulse wave velocity in 77 Williams cases and matched controls. Cases had stiffer conducting vessels than controls (P<0.001), with increased stiffness observed in even the youngest children with Williams syndrome. Pulse wave velocity increased with age at comparable rates in cases and controls, and although the degree of vascular stiffness varied, it was seen in both hypertensive and normotensive Williams participants. Use of antihypertensive medication and extension of the Williams deletion to include **NCF1** were associated with protection from vascular stiffness. These findings demonstrate that vascular stiffness is a primary vascular phenotype in Williams syndrome and that treatment with antihypertensives or agents inhibiting oxidative stress may be important in managing patients with this condition, potentially even those who are not overtly hypertensive. https://www.ncbi.nlm.nih.gov/pubmed/?term=24126171

- Pancreatic adenocarcinoma: epidemiology and genetics. Flanders TY Foulkes WD Pancreatic adenocarcinoma is an important cause of **death** from cancer throughout the developed world. There are few established environmental risk factors, but a previous history of pancreatitis and exposure to tobacco and salted food appear to be the most important. A family history of pancreatic adenocarcinoma is not common in patients with this disease, but recent research has shown that pancreatic adenocarcinoma can be a feature of cancer susceptibility syndromes associated with germline mutations in p16, **BRCA1**, **BRCA2**, and **APC**. This highlights the need for a full family history in apparently sporadic cases. Somatic mutations in p16, **BRCA2**, and **APC** have also been reported in pancreatic cancer; however, K-RAS mutations appear to be the commonest oncogenic alteration. Recent advances in our understanding of the basis of hereditary cancer syndromes may be applicable to the diagnosis, treatment, and possibly prevention of pancreatic adenocarcinoma in the future. https://www.ncbi.nlm.nih.gov/pubmed/?term=8950667

TREMOR (temblor)

	Enfermedad	Gene	Peso
1	tremor	ELN	1

- [Clinical features of a senior patient with Williams syndrome]. Nakaji A Kawame Y Nagai C Iwata M We experienced a case of 62-year-old woman who was admitted for the evaluation of her trembling hands. She was diagnosed as Williams syndrome (WS) by fluorescent in situ hybridization (FISH) analysis. She was short in stature, had a characteristic face and moderate mental retardation, whereas she was talkative and gregarious. She also presented impaired visuospatial cognition, cerebellar ataxia and **TREMOR** like involuntary movement of the hands. No remarkable abnormality is noted in MRI of the brain. MRA study of the brain revealed the arteriosclerotic vascular change, such as elongation of basilar artery and dilatation of bilateral carotid arteries. Heterozygous microdeletion of chromosome 7q11.23 of this patient is typical for WS, the delction including elastin (**ELN**) and LIMK 1 gene. Although she was complicated by diabetes mellitus and hyperlipidemia, she had no cardiovascular abnormalities like supravalvular aortic stenosis (SVAS), and survived to her age in good condition. The **TREMOR**-like involuntary movement disappeared after her discharge and its mechanism remains to be elucidated. https://www.ncbi.nlm.nih.gov/pubmed/?term=11968743
- Biomechanical Description of Phonation in Children Affected by Williams Syndrome. Hidalgo I GÃ³mez Vilda P GarayzÃ¡bal E The voice of persons with Williams syndrome (WS) is described as hoarse with a deep and unstable fundamental frequency (f0). These observations may be justified by the deficit of elastin due to a haplo-insufficiency in the **ELN** gene characteristic of the syndrome. In view of the possible relationship between elastin deficit and dysphonia, a study of the dynamic function of WS phonation was conducted by means of biomechanical analysis. In order to assess the presence of dysphonic symptoms and their degree of severity, the biomechanical description of WS phonation has been evaluated in terms of dynamic mass and viscoelasticity estimates. Glottal biomechanical features such as vocal fold dynamic mass, stiffness, unbalances, and laryngeal **TREMOR** of 12 children with WS aged 3 to 8 years (five girls and seven boys) have been estimated and compared with the normative phonation of 97 children with typical development (53 girls and 44 boys). ========= ========= The conclusions may help to make a more complete view of the connection between WS and dysphonia based on objective assessments.Copyright Â© 2018 The Voice Foundation. Published by Elsevier Inc. All rights reserved. https://www.ncbi.nlm.nih.gov/pubmed/?term=28779989

OBSTRUCTIVE AORTOPATHY (aortopatía obstructiva)

	Enfermedad	Gene	Peso
1	obstructive aortopathy	ELN	0.980581

- mTOR (Mechanistic Target of Rapamycin) Inhibition Decreases Mechanosignaling, Collagen Accumulation, and Stiffening of the Thoracic Aorta in Elastin-Deficient Mice. Jiao Y Li G Li Q Ali R Qin L Li W Qyang Y Greif DM Geirsson A Humphrey JD Tellides G Elastin deficiency because of heterozygous loss of an **ELN** allele in Williams syndrome causes **obstructive aortopathy** characterized by medial thickening and fibrosis and consequent aortic stiffening. Previous work in **ELN**-null mice with a severe arterial phenotype showed that inhibition of mTOR (mechanistic target of rapamycin), a key regulator of cell growth, lessened the aortic obstruction but did not prevent early postnatal death. We investigated the effects of mTOR inhibition in **ELN**-null mice partially rescued by human **ELN** that manifest a less severe arterial phenotype and survive long term.Thoracic aortas of neonatal and juvenile mice with graded elastin deficiency exhibited increased signaling through both mTOR complex 1 and 2. Despite lower predicted wall stress, there was increased phosphorylation of focal adhesion kinase, suggestive of greater integrin activation, and increased transforming growth factor-Î²-signaling mediators, associated with increased collagen expression. Pharmacological blockade of mTOR by rapalogs did not improve luminal stenosis but reduced mechanosignaling (in delayed fashion after mTOR complex 1 inhibition), medial collagen accumulation, and stiffening of the aorta. Rapalog administration also retarded somatic growth, however, and precipitated neonatal deaths. Complementary, less-toxic strategies to inhibit mTOR via altered growth factor and nutrient responses were not effective.In addition to previously demonstrated therapeutic benefits of rapalogs decreasing smooth muscle cell proliferation in the absence of elastin, we find that rapalogs also prevent aortic fibrosis and stiffening attributable to partial elastin deficiency. Our findings suggest that mTOR-sensitive perturbation of smooth muscle cell mechanosensing contributes to elastin aortopathy.Â© 2017 American Heart Association, Inc. https://www.ncbi.nlm.nih.gov/pubmed/?term=28751568
- Deficient Circumferential Growth Is the Primary Determinant of Aortic Obstruction Attributable to Partial Elastin Deficiency. Jiao Y Li G Korneva A Caulk AW Qin L Bersi MR Li Q Li W Mecham RP Humphrey JD Tellides G Williams syndrome is characterized by **obstructive aortopathy** attributable to heterozygous loss of **ELN**, the gene encoding elastin. Lesions are thought to result primarily from excessive smooth muscle cell (SMC) proliferation and consequent medial expansion, although an initially smaller caliber and increased stiffness of the aorta may contribute to luminal narrowing. The relative contributions of such abnormalities to the obstructive phenotype had not been defined.We quantified determinants of luminal stenosis in thoracic aortas of **ELN**-/- mice incompletely rescued by human **ELN**. Moderate obstruction was largely because of deficient circumferential growth, most prominently of ascending segments, despite increased axial growth. Medial thickening was evident in these smaller diameter elastin-deficient aortas, with medial area similar to that of larger diameter control aortas. There was no difference in cross-sectional SMC number between mutant and wild-type genotypes at multiple stages of postnatal development. Decreased elastin content was associated with medial fibrosis and reduced aortic distensibility because of increased structural stiffness but preserved material stiffness. Elastin-deficient SMCs exhibited greater contractile-to-proliferative phenotypic modulation in vitro than in vivo. We confirmed increased medial collagen without evidence of increased medial area or SMC number in a small ascending aorta with thickened media of a Williams syndrome subject.Deficient circumferential growth is the predominant mechanism for moderate obstructive aortic disease resulting from partial elastin deficiency. Our findings suggest that diverse aortic manifestations in Williams syndrome result from graded elastin content, and SMC hyperplasia causing medial expansion requires additional elastin loss superimposed on **ELN** haploinsufficiency.Â© 2017 American Heart Association, Inc. https://www.ncbi.nlm.nih.gov/pubmed/?term=28254817

HYPERTENSION (hipertensión)

	Enfermedad	Gene	Peso
1	hypertension	ELN	0.994661
2	hypertension	F2	0.933101
3	hypertension	GTF2IRD2	0.913734
4	hypertension	NCF1	0.831787

- Functional and genetic characterization of two extremely rare cases of Williams-Beuren syndrome associated with chronic granulomatous disease. Stasia MJ Mollin M Martel C Satre V Coutton C Amblard F Vieville G van Montfrans JM Boelens JJ Veenstra-Knol HE van Leeuwen K de Boer M Brion JP Roos D Williams-Beuren syndrome (WBS) is a neurodevelopmental disorder with multi-systemic manifestations, caused by a heterozygous segmental deletion of 1.55-1.83â€‰Mb at chromosomal band 7q11.23. The deletion can include the **NCF1** gene that encodes the p47(phox) protein, a component of the leukocyte NADPH oxidase enzyme, which is essential for the defense against microbial pathogens. It has been postulated that WBS patients with two functional **NCF1** genes are more susceptible to occurrence of **hypertension** than WBS patients with only one functional **NCF1** gene. We now describe two extremely rare WBS patients without any functional **NCF1** gene, because of a mutation in **NCF1** on the allele not carrying the **NCF1**-removing WBS deletion. These two patients suffer from chronic granulomatous disease with increased microbial infections in addition to WBS. Interestingly, one of these patients did suffer from **hypertension**, indicating that other factors than NADPH oxidase in vascular tissue may be involved in causing **hypertension**. https://www.ncbi.nlm.nih.gov/pubmed/?term=23340515

- Williams syndrome predisposes to vascular stiffness modified by antihypertensive use and copy number changes in **NCF1**. Kozel BA Danback JR Waxler JL Knutsen RH de Las Fuentes L Reusz GS Kis E Bhatt AB Pober BR Williams syndrome is caused by the deletion of 26 to 28 genes, including elastin, on human chromosome 7. Elastin insufficiency leads to the cardiovascular hallmarks of this condition, namely focal stenosis and **hypertension**. Extrapolation from the **ELN**(+/-) mouse suggests that affected people may also have stiff vasculature, a risk factor for stroke, myocardial infarction, and cardiac death. **NCF1**, one of the variably deleted Williams genes, is a component of the nicotinamide adenine dinucleotide phosphate (NADPH) oxidase complex and is involved in the generation of oxidative stress, making it an interesting candidate modifier for vascular stiffness. Using a case-control design, vascular stiffness was evaluated by pulse wave velocity in 77 Williams cases and matched controls. Cases had stiffer conducting vessels than controls (P<0.001), with increased stiffness observed in even the youngest children with Williams syndrome. Pulse wave velocity increased with age at comparable rates in cases and controls, and although the degree of vascular stiffness varied, it was seen in both hypertensive and normotensive Williams participants. Use of antihypertensive medication and extension of the Williams deletion to include **NCF1** were associated with protection from vascular stiffness. These findings demonstrate that vascular stiffness is a primary vascular phenotype in Williams syndrome and that treatment with antihypertensives or agents inhibiting oxidative stress may be important in managing patients with this condition, potentially even those who are not overtly hypertensive. https://www.ncbi.nlm.nih.gov/pubmed/?term=24126171

- Reduction of NADPH-oxidase activity ameliorates the cardiovascular phenotype in a mouse model of Williams-Beuren Syndrome. Campuzano V Segura-Puimedon M Terrado V SÃ¡nchez-RodrÃguez C Coustets M Menacho-MÃ¡rquez M Nevado J Bustelo XR Francke U PÃ©rez-Jurado LA A hallmark feature of Williams-Beuren Syndrome (WBS) is a generalized arteriopathy due to elastin deficiency, presenting as stenoses of medium and large arteries and leading to **hypertension** and other cardiovascular complications. Deletion of a functional **NCF1** gene copy has been shown to protect a proportion of WBS patients against **hypertension**, likely through reduced NADPH-oxidase (NOX)-mediated oxidative stress. DD mice, carrying a 0.67 Mb heterozygous deletion including the **ELN** gene, presented with a generalized arteriopathy, **hypertension**, and cardiac hypertrophy, associated with elevated angiotensin II (angII), oxidative stress parameters, and **NCF1** expression. Genetic (by crossing with **NCF1** mutant) and/or pharmacological (with ang II type 1 receptor blocker, losartan, or NOX inhibitor apocynin) reduction of NOX activity controlled hormonal and biochemical parameters in DD mice, resulting in normalized blood pressure and improved cardiovascular histology. We provide strong evidence for implication of the redox system in the pathophysiology of the cardiovascular disease in a mouse model of WBS. The phenotype of these mice can be ameliorated by either genetic or pharmacological intervention reducing NOX activity, likely through reduced angII-mediated oxidative stress. Therefore, anti-NOX therapy merits evaluation to prevent the potentially serious cardiovascular complications of WBS, as well as in other cardiovascular disorders mediated by similar pathogenic mechanism. https://www.ncbi.nlm.nih.gov/pubmed/?term=22319452

- Genetic modifiers of cardiovascular phenotype caused by elastin haploinsufficiency act by extrinsic noncomplementation. Kozel BA Knutsen RH Ye L Ciliberto CH Broekelmann TJ Mecham RP Elastin haploinsufficiency causes the cardiovascular complications associated with Williams-Beuren syndrome and isolated supravalvular aortic stenosis. Significant variability exists in the vascular pathology in these individuals. Using the **ELN**(+/-) mouse, we sought to identify the source of this variability. Following outcrossing of C57Bl/6J **ELN**(+/-), two backgrounds were identified whose cardiovascular parameters deviated significantly from the parental strain. F1 progeny of the C57Bl/6J; **ELN**(+/-)x129X1/SvJ were more hypertensive and their arteries less compliant. ========= ========= Additional peaks were identified that affect only **ELN**(+/-), including a region upstream of **ELN** on chromosome 5 (LOD 4.5). Bioinformatic analysis of the quantitative trait locus peaks revealed several interesting candidates, including Ren1, **NCF1**, and Nos1; genes whose functions are unrelated to elastic fiber assembly, but whose effects may synergize with elastin insufficiency to predispose to **hypertension** and stiffer blood vessels. Real time RT-PCR studies show background-specific increased expression of **NCF1** (a subunit of the NOX2 NAPDH oxidase) that parallel the presence of increased oxidative stress in **ELN**(+/-) aortas. This finding raises the possibility that polymorphisms in genes affecting the generation of reactive oxygen species alter cardiovascular function in individuals with elastin haploinsufficiency through extrinsic noncomplementation. https://www.ncbi.nlm.nih.gov/pubmed/?term=22049077

- Induced chromosome deletion in a Williams-Beuren syndrome mouse model causes cardiovascular abnormalities. Goergen CJ Li HH Francke U Taylor CA The Williams-Beuren syndrome (WBS) is a genetic disorder caused by a heterozygous ~1.5-Mb deletion. The aim of this study was to determine how the genetic changes in a Wbs mouse model alter **ELN** expression, blood pressure, vessel structure, and abdominal aortic wall dynamics in vivo.Elastin (**ELN**) transcript levels were quantified by qRT-PCR and blood pressure was measured with a tail cuff system. M-mode ultrasound was used to track pulsatile abdominal aortic wall motion. Aortas were sectioned and stained to determine medial lamellar structure.**ELN** transcript levels were reduced by 38-41% in Wbs mice lacking one copy of the **ELN** gene. These mice also had a 10-20% increase in mean blood pressure and significantly reduced circumferential cyclic strain ($p < 0.001$). Finally, histological sections showed disorganized and fragmented elastin sheets in Wbs mice, but not the characteristic increase in lamellar units seen in **ELN**(+/-) mice.The deletion of **ELN** in this Wbs mouse model results in lower gene expression, **hypertension**, reduced cyclic strain, and fragmented elastin sheets. The observation that the number of medial lamellar units is normal in Wbs deletion mice, which is in contrast to **ELN**(+/-) mice, suggests other genes may be involved in vascular development.Copyright Â© 2010 S. Karger AG, Basel. https://www.ncbi.nlm.nih.gov/pubmed/?term=20926892

- Hemizygosity at the **NCF1** gene in patients with Williams-Beuren syndrome decreases their risk of **hypertension**. Del Campo M Antonell A Magano LF MuÃ±oz FJ Flores R BayÃ©s M PÃ©rez Jurado LA Williams-Beuren syndrome (WBS), caused by a heterozygous deletion at 7q11.23, represents a model for studying **hypertension**, the leading risk factor for mortality worldwide, in a genetically determined disorder. Haploinsufficiency at the elastin gene is known to lead to the vascular stenoses in WBS and is also thought to predispose to **hypertension**, present in approximately 50% of patients. Detailed clinical and molecular characterization of 96 patients with WBS was performed to explore clinical-molecular correlations. Deletion breakpoints were precisely defined and were found to result in variability at two genes, **NCF1** and **GTF2IRD2**. **hypertension** was significantly less prevalent in patients with WBS who had the deletion that included **NCF1** (P=.02), a gene coding for the p47(phox) subunit of the NADPH oxidase. Decreased p47(phox) protein levels, decreased superoxide anion production, and lower protein nitrotyrosination were all observed in cell lines from patients hemizygous at **NCF1**. Our results indicate that the loss of a functional copy of **NCF1** protects a proportion of patients with WBS against **hypertension**, likely through a lifelong reduced angiotensin II-mediated oxidative stress. Therefore, antioxidant therapy that reduces NADPH oxidase activity might have a potential benefit in identifiable patients with WBS in whom serious complications related to **hypertension** have been reported, as well as in forms of essential **hypertension** mediated by a similar pathogenic mechanism. https://www.ncbi.nlm.nih.gov/pubmed/?term=16532385

MENTAL RETARDATION (retraso mental)

	Enfermedad	Gene	Peso
1	mental retardation	NSD1	0.969706
2	mental retardation	KY	0.969179
3	mental retardation	NGF	0.969179
4	mental retardation	BCL7B	0.968078
5	mental retardation	CPD	0.968078
6	mental retardation	NF1	0.959826
7	mental retardation	AVP	0.950964
8	mental retardation	FZD9	0.764971

- Complete physical map of the common deletion region in Williams syndrome and identification and characterization of three novel genes. Meng X Lu X Li Z Green ED Massa H Trask BJ Morris CA Keating MT Williams syndrome (WS) is a contiguous gene deletion disorder caused by haploinsufficiency of genes at 7q11.23. We have shown that hemizygosity of elastin is responsible for one feature of WS, supravalvular aortic stenosis (SVAS). We have also implicated LIM-kinase 1 hemizygosity as a contributing factor to impaired visual-spatial constructive cognition in WS. However, the common WS deletion region has not been completely characterized, and genes for additional features of WS, including **mental retardation**, infantile hypercalcemia, and unique personality profile, are yet to be discovered. Here, we present a physical map encompassing 1.5 Mb DNA that is commonly deleted in individuals with WS. ========= ========= WS-betaTRP has four putative beta-transducin (WD40) repeats, and WS-bHLH is a novel basic helix-loop-helix leucine zipper (bHLHZip) gene. **BCL7B** belongs to a novel family of highly conserved genes. We describe the expression profile and genomic structure for each of these genes. Hemizygous deletion of one or more of these genes may contribute to developmental defects in WS. https://www.ncbi.nlm.nih.gov/pubmed/?term=9860302

- MLPA analysis for a panel of syndromes with **mental retardation** reveals imbalances in 5.8% of patients with **mental retardation** and dysmorphic features, including duplications of the Sotos syndrome and Williams-Beuren syndrome regions. Kirchhoff M Bisgaard AM Bryndorf T Gerdes T MLPA analysis for a panel of syndromes with **mental retardation** (MRS-MLPA) was used for investigation of 258 mentally retarded and dysmorphic patients with normal conventional karyotypes (P064 probe set, MRC-Holland, for detection of (micro)deletions associated with 1p36-deletion, Sotos, Williams-Beuren, Prader-Willi, Angelman, Miller-Dieker, Smith-Magenis, and 22q11-deletion syndromes). Patients were initially referred for HR-CGH analysis and MRS-MLPA was performed retrospectively. MRS-MLPA analysis revealed imbalances in 15/258 patients (5.8%). Ten deletions were identified, including deletions of 1p36, 5q35 (Sotos syndrome), 7q11 (Williams-Beuren syndrome), 17p11 (Smith-Magenis syndrome), 15q11 (Angelman syndrome) and 22q11. Duplications were detected in 5q35, 7q11, 17p13, 17p11 and 22q11. ========= ========= Clinical data regarding three patients investigated by MRS-MLPA are presented. The imbalances carried by these patients include a small interstitial 1p36 deletion, a small duplication of 5q35 (encompassing the **NSD1** gene, which is deleted/mutated in Sotos syndrome) and a duplication of 7q11 (reciprocal of the Williams-Beuren syndrome deletion), respectively. MRS-MLPA allows testing for a number of micro-deletions/-duplications in a single experiment, thereby filling a gap between array techniques and single locus techniques. MRS-MLPA combined with Subtelomeric MLPA represents an attractive first test in a clinical algorithm for **mental retardation**. https://www.ncbi.nlm.nih.gov/pubmed/?term=17090394

- VI. Genome structure and cognitive map of Williams syndrome. Korenberg JR Chen XN Hirota H Lai Z Bellugi U Burian D Roe B Matsuoka R Williams syndrome (WMS) is a most compelling model of human cognition, of human genome organization, and of evolution. Due to a deletion in chromosome band 7q11.23, subjects have cardiovascular, connective tissue, and neurodevelopmental deficits. Given the striking peaks and valleys in neurocognition including deficits in visual-spatial and global processing, preserved language and face processing, hypersociability, and heightened affect, the goal of this work has been to identify the genes that are responsible, the cause of the deletion, and its origin in primate evolution. To do this, we have generated an integrated physical, genetic, and transcriptional map of the WMS and flanking regions using multicolor metaphase and interphase fluorescence in situ hybridization (FISH) of bacterial artificial chromosomes (BACs) and P1 artificial chromosomes (PACs), BAC end sequencing, PCR gene marker and microsatellite, large-scale sequencing, cDNA library, and database analyses. The results indicate the genomic organization of the WMS region as two nested duplicated regions flanking a largely single-copy region. ========= ========= Primate studies indicate an evolutionary hot spot for chromosomal inversion in the WMS region. A cognitive phenotypic map of WMS is presented, which combines previous data with five further WMS subjects and three atypical WMS subjects with deletions; two larger (deleted for D7S489L) and one smaller, deleted for genes telomeric to **FZD9**, through LIMK1, but not WSCR1 or telomeric. The results establish regions and consequent gene candidates for WMS features including **mental retardation**, hypersociability, and facial features. The approach provides the basis for defining pathways linking genetic underpinnings with the neuroanatomical, functional, and behavioral consequences that result in human cognition. https://www.ncbi.nlm.nih.gov/pubmed/?term=10953236

HYPERCALCEMIA (hipercalcemia)

	Enfermedad	Gene	Peso
1	hypercalcemia	HR	0.980177
2	hypercalcemia	BCL7B	0.979389
3	hypercalcemia	YWHAG	0.975633
4	hypercalcemia	SI	0.975633
5	hypercalcemia	AUTS2	0.975633
6	hypercalcemia	CALN1	0.975633
7	hypercalcemia	HIP1	0.975633

- Zebrafish gene knockdowns imply roles for human **YWHAG** in infantile spasms and cardiomegaly. Komoike Y Fujii K Nishimura A Hiraki Y Hayashidani M Shimojima K Nishizawa T Higashi K Yasukawa K Saitsu H Miyake N Mizuguchi T Matsumoto N Osawa M Kohno Y Higashinakagawa T Yamamoto T Williams-Beuren syndrome (WBS) is a neurodevelopmental disorder presenting with an elfin-like face, supravalvular aortic stenosis, a specific cognitive-behavioral profile, and infantile **hypercalcemia**. We encountered two WBS patients presenting with infantile spasms, which is extremely rare in WBS. Array comparative genomic hybridization (aCGH) and fluorescent in situ hybridization (FISH) analyses revealed atypical 5.7-Mb and 4.1-Mb deletions at 7q11.23 in the two patients, including the WBS critical region and expanding into the proximal side and the telomeric side, respectively. On the proximal side, **AUTS2** and **CALN1** may contribute to the phenotype. On the telomeric side, there are two candidate genes **HIP1** and **YWHAG**. Because detailed information of them was unavailable, we investigated their functions using gene knockdowns of zebrafish. When zebrafish **YWHAG**1 was knocked down, reduced brain size and increased diameter of the heart tube were observed, indicating that the infantile spasms and cardiomegaly seen in the patient with the telomeric deletion may be derived from haploinsufficiency of **YWHAG**.2010 Wiley-Liss, Inc.
https://www.ncbi.nlm.nih.gov/pubmed/?term=20146355

- Complete physical map of the common deletion region in Williams syndrome and identification and characterization of three novel genes. Meng X Lu X Li Z Green ED Massa H Trask BJ Morris CA Keating MT Williams syndrome (WS) is a contiguous gene deletion disorder caused by haploinsufficiency of genes at 7q11.23. We have shown that hemizygosity of elastin is responsible for one feature of WS, supravalvular aortic stenosis (SVAS). We have also implicated LIM-kinase 1 hemizygosity as a contributing factor to impaired visual-spatial constructive cognition in WS. However, the common WS deletion region has not been completely characterized, and genes for additional features of WS, including mental retardation, infantile **hypercalcemia**, and unique personality profile, are yet to be discovered. Here, we present a physical map encompassing 1.5 Mb DNA that is commonly deleted in individuals with WS. ========= ========= WS-betaTRP has four putative beta-transducin (WD40) repeats, and WS-bHLH is a novel basic helix-loop-helix leucine zipper (bHLHZip) gene. **BCL7B** belongs to a novel family of highly conserved genes. We describe the expression profile and genomic structure for each of these genes. Hemizygous deletion of one or more of these genes may contribute to developmental defects in WS.
https://www.ncbi.nlm.nih.gov/pubmed/?term=9860302

MENTAL DISABILITY (discapacidad mental)

	Enfermedad	Gene	Peso
1	mental disability	COMT	0.722779

- A comparative study of hearing loss in two microdeletion syndromes: velocardiofacial (22q11.2 deletion) and Williams (7q11.23 deletion) syndromes. Zarchi O Attias J Raveh E Basel-Vanagaite L Saporta L Gothelf D To comprehensively assess auditory impairments in velocardiofacial syndrome (VCFS) and Williams syndrome (WS).Audiologic measurements were conducted with 62 subjects with VCFS and 44 subjects with WS, as well as two control groups consisting of 22 subjects with idiopathic develop**mental disability** and 23 typically developing controls. An association between severity of hearing loss in VCFS and the (158)Val/Met polymorphism of the catechol-O-methyltransferase gene (**COMT**) was explored.Hearing was significantly more impaired in the VCFS and WS groups compared with the develop**mental disability** and typically developing groups. Audiologic abnormalities identified in both the VCFS and WS groups included high-tone hearing loss (predominantly sensorineural or mixed type), loss of acoustic reflex, and middle ear pathologies. In both the VCFS and WS groups, hearing loss severity was positively correlated with age. In the VCFS group, hearing loss was more severe in the subgroup carrying the **COMT** Val allele compared with the subgroup carrying the **COMT** Met allele.Hearing impairments, including sensorineural hearing loss and acoustic reflex dysfunction, are very common in both VCFS and WS. Hearing loss is less severe in subjects with the **COMT** Met allele, possibly due to the protective effect of dopamine on the hearing system.Copyright Â© 2011 Mosby, Inc. All rights reserved.
https://www.ncbi.nlm.nih.gov/pubmed/?term=20846670

SCHIZOPHRENIA AND AUTISM (esquizofrenia y autismo)

	Enfermedad	Gene	Peso
1	schizophrenia and autism	GTF2I	0.707107

- Williams syndrome deletions and duplications: Genetic windows to understanding anxiety, sociality, **autism, and schizophrenia**. Crespi BJ Procyshyn TL We describe and evaluate an integrative hypothesis for helping to explain the major neurocognitive features of individuals with Williams syndrome region deletions and duplications. First, we demonstrate how the cognitive differences between Williams syndrome individuals, individuals with duplications of this region, and healthy individuals parallel the differences between individuals subject to effects of increased or decreased oxytocin. Second, we synthesize evidence showing that variation in expression of the gene **GTF2I** (General Transcription Factor II-I) underlies the primary social phenotypes of Williams syndrome and that common genetic variation in **GTF2I** mediates oxytocin reactivity, and its correlates, in healthy populations. Third, we describe findings relevant to the hypothesis that the **GTF2I** gene is subject to parent of origin effects whose behavioral expression fits with predictions from the kinship theory of genomic imprinting. Fourth, we describe how Williams syndrome can be considered, in part, as an autistic syndrome of Lorna Wing's 'active-but-odd' autism subtype, in contrast to associations of duplications with both schizophrenia and autism.Copyright Â© 2017 Elsevier Ltd. All rights reserved.
https://www.ncbi.nlm.nih.gov/pubmed/?term=28499504

ABNORMAL THYROID FUNCTION (función anormal tiroides)

	Enfermedad	Gene	Peso
1	abnormal thyroid function	TH	0.707107

- **thyroid** evaluation of children and adolescents with Williams syndrome in Zhejiang Province. Chen WJ Ji C Yao D Zhao ZY The objective of the study was to describe the prevalence of abnormal **thyroid** function and volume in children and adolescents with Williams syndrome (WS) in Zhejiang Province, China.**thyroid** function, including **thyroid**-stimulating hormone (TSH), free triiodothyronine (fT3), free thyroxine (fT4), and **thyroid** antibodies (**thyroid** peroxidase and thyroglobulin) were measured in 83 patients with WS, aged 0.2-16.5 years. Twenty-three patients were followed for an average of 1.7 years (0.4-4.1), and multiple TSH determinations were considered. **thyroid** ultrasonography was performed on 49 patients.One patient was diagnosed with overt hypo**thyroid**ism, and 23 patients (27%) had subclinical hypo**thyroid**ism (SH). **thyroid** antibodies were absent in all patients. In five age groups (0-1 years, 1-3 years, 3-6 years, 6-9 years, 9-18 years), the prevalence of patients with subclinical hypo**thyroid**ism was 25%, 28.5%, 44.4%, 16.7% and 4.7%, respectively. Through ultrasound examination, 21 patients (42%) were observed to have **thyroid** hypoplasia (TH), and there were no cases of **thyroid** haemiagenesis. The incidence rate of **TH** increased with age, rising from 20% in the youngest group to 66% in the oldest.SH and **TH** is common in children and adolescents with WS. Yearly evaluation of **thyroid** must be performed in all patients in this population, regardless of the result of the neonatal screening. Age under 6 years and existing **thyroid** abnormalities are risk factors for developing SH, and a shorter follow-up interval is needed for screening in these individuals, SH is often self-limiting, and clinicians should be alert to overt hypo**thyroid**ism. https://www.ncbi.nlm.nih.gov/pubmed/?term=29127763

SOCIAL COGNITION (cognición social)

	Enfermedad	Gene	Peso
1	social cognition	ADC	0.973785
2	social cognition	MOG	0.973785

- Altered microstructure within social-cognitive brain networks during childhood in Williams syndrome. Haas BW Barnea-Goraly N Sheau KE Yamagata B Ullas S Reiss AL Williams syndrome (WS) is a neurodevelopmental condition caused by a hemizygous deletion of â^¼26-28 genes on chromosome 7q11.23. WS is associated with a distinctive pattern of **social cognition**. Accordingly, neuroimaging studies show that WS is associated with structural alterations of key brain regions involved in **social cognition** during adulthood. However, very little is currently known regarding the neuroanatomical structure of social cognitive brain networks during childhood in WS. This study used diffusion tensor imaging to investigate the structural integrity of a specific set of white matter pathways (inferior fronto-occipital fasciculus [IFOF] and uncinate fasciculus [UF]) and associated brain regions [fusiform gyrus (FG), amygdala, hippocampus, medial orbitofrontal gyrus (**MOG**)] known to be involved in **social cognition** in children with WS and a typically developing (TD) control group. Children with WS exhibited higher fractional anisotropy (FA) and axial diffusivity values and lower radial diffusivity and apparent diffusion coefficient (**ADC**) values within the IFOF and UF, higher FA values within the FG, amygdala, and hippocampus and lower **ADC** values within the FG and **MOG** compared to controls. These findings provide evidence that the WS genetic deletion affects the development of key white matter pathways and brain regions important for **social cognition**. Â© The Author 2013. Published by Oxford University Press. All rights reserved. For Permissions, please e-mail: journals.permissions@oup.com. https://www.ncbi.nlm.nih.gov/pubmed/?term=23709644

CARDIOVASCULAR DISEASE (enfermedad cardiovascular)

	Enfermedad	Gene	Peso
1	cardiovascular disease	NCF1	0.979962
2	cardiovascular disease	HR	0.968386

- Reduction of NADPH-oxidase activity ameliorates the cardiovascular phenotype in a mouse model of Williams-Beuren Syndrome. Campuzano V Segura-Puimedon M Terrado V SÃ¡nchez-RodrÃguez C Coustets M Menacho-MÃ¡rquez M Nevado J Bustelo XR Francke U PÃ©rez-Jurado LA A hallmark feature of Williams-Beuren Syndrome (WBS) is a generalized arteriopathy due to elastin deficiency, presenting as stenoses of medium and large arteries and leading to hypertension and other cardiovascular complications. Deletion of a functional **NCF1** gene copy has been shown to protect a proportion of WBS patients against hypertension, likely through reduced NADPH-oxidase (NOX)-mediated oxidative stress. DD mice, carrying a 0.67 Mb heterozygous deletion including the Eln gene, presented with a generalized arteriopathy, hypertension, and cardiac hypertrophy, associated with elevated angiotensin II (angII), oxidative stress parameters, and **NCF1** expression. Genetic (by crossing with **NCF1** mutant) and/or pharmacological (with ang II type 1 receptor blocker, losartan, or NOX inhibitor apocynin) reduction of NOX activity controlled hormonal and biochemical parameters in DD mice, resulting in normalized blood pressure and improved cardiovascular histology. We provide strong evidence for implication of the redox system in the pathophysiology of the **cardiovascular disease** in a mouse model of WBS. The phenotype of these mice can be ameliorated by either genetic or pharmacological intervention reducing NOX activity, likely through reduced angII-mediated oxidative stress. Therefore, anti-NOX therapy merits evaluation to prevent the potentially serious cardiovascular complications of WBS, as well as in other cardiovascular disorders mediated by similar pathogenic mechanism. https://www.ncbi.nlm.nih.gov/pubmed/?term=22319452

EMPHYSEMA (enfisema)

	Enfermedad	Gene	Peso
1	emphysema	ELN	0.761243

- A new diagnosis of Williams-Beuren syndrome in a 49-year-old man with severe bullous **emphysema**. Wojcik MH Carmichael N Bieber FR Wiener DC Madan R Pober BR Raby BA Williams-Beuren syndrome (WBS) is a chromosomal microdeletion syndrome typically presenting with intellectual disability, a unique personality, a characteristic facial appearance, and cardiovascular disease. Several clinical features of WBS are thought to be due to haploinsufficiency of elastin (**ELN**), as the **ELN** locus is included within the WBS critical region at 7q11.23. **emphysema**, a disease attributed to destruction of pulmonary elastic fibers, has been reported in patients without WBS who have pathogenic variants in **ELN** but only once (in one patient) in WBS. Here we report a second adult WBS patient with **emphysema** where the diagnosis of WBS was established subsequent to the discovery of severe bullous **emphysema**. Haploinsufficiency of **ELN** likely contributed to this pulmonary manifestation of WBS. This case emphasizes the contribution of rare genetic variation in cases of severe **emphysema** and provides further evidence that **emphysema** should be considered in patients with WBS who have respiratory symptoms, as it may be under-recognized in this patient population.Â© 2017 Wiley Periodicals, Inc. https://www.ncbi.nlm.nih.gov/pubmed/?term=28574231
- Pulmonary function and **emphysema** in Williams-Beuren syndrome. Wan ES Pober BR Washko GR Raby BA Silverman EK Williams-Beuren syndrome (WBS) is caused by a submicroscopic deletion on chromosome 7q11.23 that encompasses the entire elastin (**ELN**) gene. Elastin, a key component of elastic fibers within the lung, is progressively destroyed in **emphysema**. Defects in the elastin gene have been associated with increased susceptibility towards developing chronic obstructive pulmonary disease (COPD) and **emphysema** in both humans and mice. We postulate that hemizygosity at the elastin gene locus may increase susceptibility towards the development of COPD and **emphysema** in subjects with WBS. We describe an adult subject with WBS who was a lifelong non-smoker and was found to have moderate **emphysema**. We also examined the pulmonary function of a separate cohort of adolescents and young adults with WBS. Although no significant spirometric abnormalities were identified, a significant proportion of subjects reported respiratory symptoms. Thus, while significant obstructive disease does not appear to be common in relatively young adults with WBS, subclinical **emphysema** and lung disease may exist which possibly could worsen with advancing age. Further investigation may elucidate the pathogenesis of non-smoking-related **emphysema**.(c) 2010 Wiley-Liss, Inc. https://www.ncbi.nlm.nih.gov/pubmed/?term=20186780

HAPLOINSUFFICIENCY (haploinsuficiencia)

	Enfermedad	Gene	Peso
1	haploinsufficiency	MLXIPL	0.932447
2	haploinsufficiency	WBSCR16	0.931242
3	haploinsufficiency	WBSCR17	0.931242
4	haploinsufficiency	WBSCR22	0.931242
5	haploinsufficiency	BCL7B	0.930621
6	haploinsufficiency	SERPINA1	0.927316
7	haploinsufficiency	FKBP6	0.908955
8	haploinsufficiency	AUTS2	0.898187
9	haploinsufficiency	CALN1	0.898187
10	haploinsufficiency	CA1	0.89165
11	haploinsufficiency	TBL2	0.879504
12	haploinsufficiency	YWHAG	0.870248
13	haploinsufficiency	HIP1	0.830472
14	haploinsufficiency	BAZ1B	0.817112
15	haploinsufficiency	FZD9	0.806702
16	haploinsufficiency	BMP4	0.761414
17	haploinsufficiency	EPHA4	0.761414
18	haploinsufficiency	PAX2	0.761414
19	haploinsufficiency	SHH	0.761414
20	haploinsufficiency	SOX2	0.761414

- **haploinsufficiency** of **BAZ1B** contributes to Williams syndrome through transcriptional dysregulation of neurodevelopmental pathways. Lalli MA Jang J Park JH Wang Y Guzman E Zhou H Audouard M Bridges D Tovar KR Papuc SM Tutulan-Cunita AC Huang Y Budisteanu M Arghir A Kosik KS Williams syndrome (WS) is a neurodevelopmental disorder caused by a genomic deletion of â¼28 genes that results in a cognitive and behavioral profile marked by overall intellectual impairment with relative strength in expressive language and hypersocial behavior. Advancements in protocols for neuron differentiation from induced pluripotent stem cells allowed us to elucidate the molecular circuitry underpinning the ontogeny of WS. In patient-derived stem cells and neurons, we determined the expression profile of the Williams-Beuren syndrome critical region-deleted genes and the genome-wide transcriptional consequences of the hemizygous genomic microdeletion at chromosome 7q11.23. Derived neurons displayed disease-relevant hallmarks and indicated novel aberrant pathways in WS neurons including over-activated Wnt signaling accompanying an incomplete neurogenic commitment. We show that **haploinsufficiency** of the ATP-dependent chromatin remodeler, **BAZ1B**, which is deleted in WS, significantly contributes to this differentiation defect. ========= ========= Altogether, these results reveal a pivotal role for **BAZ1B** in neurodevelopment and implicate its **haploinsufficiency** as a likely contributor to the neurological phenotypes in WS. Â© The Author 2016. Published by Oxford University Press. All rights reserved. For Permissions, please email: journals.permissions@oup.com. https://www.ncbi.nlm.nih.gov/pubmed/?term=26755828

- Metabolic abnormalities in Williams-Beuren syndrome. Palacios-VerdÃº MG Segura-Puimedon M Borralleras C Flores R Del Campo M Campuzano V PÃ©rez-Jurado LA Williams-Beuren syndrome (WBS, OMIM-194050) is a neurodevelopmental disorder with multisystemic manifestations caused by a 1.55-1.83 Mb deletion at 7q11.23 including 26-28 genes. Reported endocrine and metabolic abnormalities include transient hypercalcaemia of infancy, subclinical hypothyroidism in â¼ 30% of children and impaired glucose tolerance in â¼ 75% of adult individuals. The purpose of this study was to further study metabolic alterations in patients with WBS, as well as in several mouse models, to establish potential candidate genes.We analysed several metabolic parameters in a cohort of 154 individuals with WBS (data available from 69 to 151 cases per parameter), as well as in several mouse models with complete and partial deletions of the orthologous WBS locus, and searched for causative genes and potential modifiers.Triglyceride plasma levels were significantly decreased in individuals with WBS while cholesterol levels were slightly decreased compared with controls. Hyperbilirubinemia, mostly unconjugated, was found in 18.3% of WBS cases and correlated with subclinical hypothyroidism and hypotriglyceridemia, suggesting common pathogenic mechanisms. **haploinsufficiency** at **MLXIPL** and increased penetrance for hypomorphic alleles at the UGT1A1 gene promoter might underlie the lipid and bilirubin alterations. Other disturbances included increased protein and iron levels, as well as the known subclinical hypothyroidism and glucose intolerance.Our results show that several unreported biochemical alterations, related to **haploinsufficiency** for specific genes at 7q11.23, are relatively common in WBS. The early diagnosis, follow-up and management of these metabolic disturbances could prevent long-term complications in this disorder.Published by the BMJ Publishing Group Limited. For permission to use (where not already granted under a licence) please go to http://group.bmj.com/group/rights-licensing/permissions. https://www.ncbi.nlm.nih.gov/pubmed/?term=25663682

- Alpha 1 antitrypsin deficiency alleles are associated with joint dislocation and scoliosis in Williams syndrome. Morris CA Pani AM Mervis CB Rios CM Kistler DJ Gregg RG Elastin **haploinsufficiency** is responsible for a significant portion of the Williams syndrome (WS) phenotype including hoarse voice, supravalvar aortic stenosis (SVAS), hernias, diverticuli of bowel and bladder, soft skin, and joint abnormalities. All of the connective tissue signs and symptoms are variable in the WS population, but few factors other than age and gender are known to influence the phenotype. We examined a cohort of 205 individuals with WS for mutations in **SERPINA1**, the gene that encodes alpha-1-antitrypsin (AAT), the inhibitor of elastase. Individuals with classic WS deletions and **SERPINA1** genotypes PiMS or PiMZ were more likely than those with a **SERPINA1** PiMM genotype to have joint dislocation or scoliosis. However, carrier status for AAT deficiency was not correlated with presence of inguinal hernia or with presence or severity of SVAS. These findings suggest that genes important in elastin metabolism are candidates for variability in the connective tissue abnormalities in WS. https://www.ncbi.nlm.nih.gov/pubmed/?term=20425789

- Loss of WSTF results in spontaneous fluctuations of heterochromatin formation and resolution, combined with substantial changes to gene expression. Culver-Cochran AE Chadwick BP Williams syndrome transcription factor (WSTF) is a multifaceted protein that is involved in several nuclear processes, including replication, transcription, and the DNA damage response. WSTF participates in a chromatin-remodeling complex with the ISWI ATPase, SNF2H, and is thought to contribute to the maintenance of heterochromatin, including at the human inactive X chromosome (Xi). WSTF is encoded by **BAZ1B**, and is one of twenty-eight genes that are hemizygously deleted in the genetic disorder Williams-Beuren syndrome (WBS).To explore the function of WSTF, we performed zinc finger nuclease-assisted targeting of the **BAZ1B** gene and isolated several independent knockout clones in human cells. Our results show that, while heterochromatin at the Xi is unaltered, new inappropriate areas of heterochromatin spontaneously form and resolve throughout the nucleus, appearing as large DAPI-dense staining blocks, defined by histone H3 lysine-9 trimethylation and association of the proteins heterochromatin protein 1 and structural maintenance of chromosomes flexible hinge domain containing 1. In three independent mutants, the expression of a large number of genes were impacted, both up and down, by WSTF loss.Given the inappropriate appearance of regions of heterochromatin in **BAZ1B** knockout cells, it is evident that WSTF performs a critical role in maintaining chromatin and transcriptional states, a property that is likely compromised by WSTF **haploinsufficiency** in WBS patients. https://www.ncbi.nlm.nih.gov/pubmed/?term=24168170

- Smaller and larger deletions of the Williams Beuren syndrome region implicate genes involved in mild facial phenotype, epilepsy and autistic traits. Fusco C Micale L Augello B Teresa Pellico M Menghini D Alfieri P Cristina Digilio M Mandriani B Carella M Palumbo O Vicari S Merla G Williams Beuren syndrome (WBS) is a multisystemic disorder caused by a hemizygous deletion of 1.5 Mb on chromosome 7q11.23 spanning 28 genes. A few patients with larger and smaller WBS deletion have been reported. They show clinical features that vary between isolated SVAS to the full spectrum of WBS phenotype, associated with epilepsy or autism spectrum behavior. Here we describe four patients with atypical WBS 7q11.23 deletions. Two carry ~3.5 Mb larger deletion towards the telomere that includes Huntingtin-interacting protein 1 (**HIP1**) and tyrosine 3-monooxygenase/tryptophan 5-monooxigenase activation protein gamma (**YWHAG**) genes. Other two carry a shorter deletion of ~1.2 Mb at centromeric side that excludes the distal WBS genes **BAZ1B** and **FZD9**. Along with previously reported cases, genotype-phenotype correlation in the patients described here further suggests that **haploinsufficiency** of **HIP1** and **YWHAG** might cause the severe neurological and neuropsychological deficits including epilepsy and autistic traits, and that the preservation of **BAZ1B** and **FZD9** genes may be related to mild facial features and moderate neuropsychological deficits. This report highlights the importance to characterize additional patients with 7q11.23 atypical deletions comparing neuropsychological and clinical features between these individuals to shed light on the pathogenic role of genes within and flanking the WBS region. https://www.ncbi.nlm.nih.gov/pubmed/?term=23756441

- Zebrafish gene knockdowns imply roles for human **YWHAG** in infantile spasms and cardiomegaly. Komoike Y Fujii K Nishimura A Hiraki Y Hayashidani M Shimojima K Nishizawa T Higashi K Yasukawa K Saitsu H Miyake N Mizuguchi T Matsumoto N Osawa M Kohno Y Higashinakagawa T Yamamoto T Williams-Beuren syndrome (WBS) is a neurodevelopmental disorder presenting with an elfin-like face, supravalvular aortic stenosis, a specific cognitive-behavioral profile, and infantile hypercalcemia. We encountered two WBS patients presenting with infantile spasms, which is extremely rare in WBS. Array comparative genomic hybridization (aCGH) and fluorescent in situ hybridization (FISH) analyses revealed atypical 5.7-Mb and 4.1-Mb deletions at 7q11.23 in the two patients, including the WBS critical region and expanding into the proximal side and the telomeric side, respectively. On the proximal side, **AUTS2** and **CALN1** may contribute to the phenotype. On the telomeric side, there are two candidate genes **HIP1** and **YWHAG**. Because detailed information of them was unavailable, we investigated their functions using gene knockdowns of zebrafish. When zebrafish **YWHAG**1 was knocked down, reduced brain size and increased diameter of the heart tube were observed, indicating that the infantile spasms and cardiomegaly seen in the patient with the telomeric deletion may be derived from **haploinsufficiency** of **YWHAG**.2010 Wiley-Liss, Inc. https://www.ncbi.nlm.nih.gov/pubmed/?term=20146355

- Identification of additional transcripts in the Williams-Beuren syndrome critical region. Merla G Ucla C Guipponi M Reymond A Williams-Beuren syndrome (WBS) is a developmental disorder associated with **haploinsufficiency** of multiple genes at 7q11.23. Here, we report the characterization of **WBSCR16**, **WBSCR17**, WBSCR18, WBSCR20A, WBSCR20B, WBSCR20C, WBSCR21, **WBSCR22**, and WBSCR23, nine novel genes contained in the WBS commonly deleted region or its flanking sequences. They encode an RCC1-like G-exchanging factor, an N-acetylgalactosaminyltransferase, a DNAJ-like chaperone, NOL1/NOP2/sun domain-containing proteins, a methyltransferase, or proteins with no known homologies. **haploinsufficiency** of these newly identified WBSCR genes may contribute to certain of the WBS phenotypical features. https://www.ncbi.nlm.nih.gov/pubmed/?term=12073013

- An atypical 7q11.23 deletion in a normal IQ Williams-Beuren syndrome patient. Ferrero GB Howald C Micale L Biamino E Augello B Fusco C Turturo MG Forzano S Reymond A Merla G Williams-Beuren syndrome (WBS; OMIM no. 194050) is a multisystemic neurodevelopmental disorder caused by a hemizygous deletion of 1.55 Mb on chromosome 7q11.23 spanning 28 genes. **haploinsufficiency** of the ELN gene was shown to be responsible for supravalvular aortic stenosis and generalized arteriopathy, whereas LIMK1, CLIP2, GTF2IRD1 and GTF2I genes were suggested to be linked to the specific cognitive profile and craniofacial features. These insights for genotype-phenotype correlations came from the molecular and clinical analysis of patients with atypical deletions and mice models. Here we report a patient showing mild WBS physical phenotype and normal IQ, who carries a shorter 1 Mb atypical deletion. This rearrangement does not include the GTF2IRD1 and GTF2I genes and only partially the **BAZ1B** gene. Our results are consistent with the hypothesis that hemizygosity of the GTF2IRD1 and GTF2I genes might be involved in the facial dysmorphisms and in the specific motor and cognitive deficits observed in WBS patients. https://www.ncbi.nlm.nih.gov/pubmed/?term=19568270

- Autosomal dominant inheritance of Williams-Beuren syndrome in a father and son with **haploinsufficiency** for **FKBP6**. Metcalfe K Simeonov E Beckett W Donnai D Tassabehji M Williams-Beuren syndrome (WBS) is a neurodevelopmental microdeletion disorder that usually occurs sporadically due to its location within a highly repetitive genomic region that is unstable and prone to unequal cross-over during meiosis. The consequential loss of chromosomal material includes approximately 1.5 Mb of DNA at 7q11.23. Whilst cases of dominant inheritance have been described in the literature, there have been few reports of molecular confirmation and none have carried out detailed genotyping. We describe a Bulgarian father and son with WBS detected by fluorescent in situ hybridisation (with an elastin gene probe) and loss of heterozygosity mapping using microsatellite markers located in the critical region. These individuals appear to have a common WBS heterozygous deletion, confirming the expected dominant transmission and adding to the few familial cases reported. The deletion includes the gene **FKBP6** which has recently been shown to play a role in homologous chromosome pairing in meiosis and male fertility in mouse models. Homozygous **FKBP6** -/- male mice are infertile and our data suggests that **haploinsufficiency** for **FKBP6** does not appear to preclude male fertility in WBS, although male infertility involving this gene has the potential to follow the mouse model as a human autosomal recessive condition. https://www.ncbi.nlm.nih.gov/pubmed/?term=15770126

- Cloning, expression, and chromosomal mapping of the human 14-3-3gamma gene (**YWHAG**) to 7q11.23. Horie M Suzuki M Takahashi E Tanigami A The 14-3-3 family of proteins exerts diverse influences on the signal transduction pathways of cells. We have newly identified a human cDNA encoding the gamma subtype of the 14-3-3 family of genes. The deduced amino acid sequence of human 14-3-3gamma was identical to that of rat 14-3-3gamma. The human 14-3-3gamma gene (HGMW-approved symbol **YWHAG**) is highly expressed in brain, skeletal muscle, and heart. By fluorescence in situ hybridization analysis, the human 14-3-3gamma gene was mapped to chromosome 7q11.23. Radiation hybrid mapping has shown that this gene is localized 2.33 cR telomeric to D7S1870, a polymorphic marker located at the most telomeric end of the common deletion region of Williams-Beuren syndrome (WBS). This suggests that **haploinsufficiency** of 14-3-3gamma may not contribute to the WBS phenotype. However, information regarding the precise chromosomal location of a member of the 14-3-3 family of genes will aid in examining the relationship between this family of proteins and human disorders.Copyright 1999 Academic Press. https://www.ncbi.nlm.nih.gov/pubmed/?term=10486217

- GTF2I hemizygosity implicated in mental retardation in Williams syndrome: genotype-phenotype analysis of five families with deletions in the Williams syndrome region. Morris CA Mervis CB Hobart HH Gregg RG Bertrand J Ensing GJ Sommer A Moore CA Hopkin RJ Spallone PA Keating MT Osborne L Kimberley KW Stock AD Most individuals with Williams syndrome (WS) have a 1.6 Mb deletion in chromosome 7q11.23 that encompasses the elastin (ELN) gene, while most families with autosomal dominant supravalvar aortic stenosis (SVAS) have point mutations in ELN. The overlap of the clinical phenotypes of the two conditions (cardiovascular disease and connective tissue abnormalities such as hernias) is due to the effect of **haploinsufficiency** of ELN. SVAS families often have affected individuals with some WS facial features, most commonly in infancy, suggesting that ELN plays a role in WS facial gestalt as well. To find other genes contributing to the WS phenotype, we studied five families with SVAS who have small deletions in the WS region. None of the families had mental retardation, but affected family members had the Williams Syndrome Cognitive Profile (WSCP). All families shared a deletion of LIMK1, which encodes a protein strongly expressed in the brain, supporting the hypothesis that LIMK1 hemizygosity contributes to impairment in visuospatial constructive cognition. While the deletions from the families nearly spanned the WS region, none had a deletion of **FKBP6** or GTF2I, suggesting that the mental retardation seen in WS is associated with deletion of either the centromeric and/or telomeric portions of the region. Comparison of these five families with reports of other individuals with partial deletions of the WS region most strongly implicates GTF2I in the mental retardation of WS.Copyright 2003 Wiley-Liss, Inc. https://www.ncbi.nlm.nih.gov/pubmed/?term=14556246

- **TBL2**, a novel transducin family member in the WBS deletion: characterization of the complete sequence, genomic structure, transcriptional variants and the mouse ortholog. PÃ©rez Jurado LA Wang YK Francke U Cruces J Williams-Beuren syndrome (WBS) is a developmental disorder with multi-system manifestations caused by **haploinsufficiency** for contiguous genes deleted in chromosome region 7q11.23. The size of the deletion is similar in most patients due to a genomic duplication that predisposes to unequal meiotic crossover events. While hemizygosity at the elastin locus is responsible for the cardiovascular features, the contribution of other genes to the WBS phenotype remains to be demonstrated. We have identified a novel gene, **TBL2**, in the common WBS deletion. **TBL2** is expressed as a 2. 4-kb transcript predominantly in testis, skeletal muscle, heart and some endocrine tissues, with a larger approximately 5-kb transcript detected ubiquitously at lower levels. ========= ========= The mouse homolog, **TBL2**, shows 84% sequence identity at the nucleotide level and 92% similarity at the amino acid level. Comparison of the mouse and human sequences identifies a conserved region that extends upstream of the previously published sequence with an initiation codon common to both species that adds 21 amino acids at the N-terminus. The **TBL2** gene has been mapped to mouse chromosome 5 in a region of conserved synteny with human 7q11.23. Since **haploinsufficiency** has been shown for other WD-repeat containing proteins, hemizygosity of **TBL2** may contribute to some of the aspects of the complex WBS phenotype. https://www.ncbi.nlm.nih.gov/pubmed/?term=10575226

- Complete physical map of the common deletion region in Williams syndrome and identification and characterization of three novel genes. Meng X Lu X Li Z Green ED Massa H Trask BJ Morris CA Keating MT Williams syndrome (WS) is a contiguous gene deletion disorder caused by **haploinsufficiency** of genes at 7q11.23. We have shown that hemizygosity of elastin is responsible for one feature of WS, supravalvular aortic stenosis (SVAS). We have also implicated LIM-kinase 1 hemizygosity as a contributing factor to impaired visual-spatial constructive cognition in WS. However, the common WS deletion region has not been completely characterized, and genes for additional features of WS, including mental retardation, infantile hypercalcemia, and unique personality profile, are yet to be discovered. Here, we present a physical map encompassing 1.5 Mb DNA that is commonly deleted in individuals with WS. ========= ========= WS-betaTRP has four putative beta-transducin (WD40) repeats, and WS-bHLH is a novel basic helix-loop-helix leucine zipper (bHLHZip) gene. **BCL7B** belongs to a novel family of highly conserved genes. We describe the expression profile and genomic structure for each of these genes. Hemizygous deletion of one or more of these genes may contribute to developmental defects in WS. https://www.ncbi.nlm.nih.gov/pubmed/?term=9860302

- A novel human gene **FKBP6** is deleted in Williams syndrome. Meng X Lu X Morris CA Keating MT Williams syndrome (WS) is a developmental disorder caused by **haploinsufficiency** of genes at 7q11.23. We have shown that hemizygosity of elastin is responsible for one feature of WS, supravalvular aortic stenosis. We have also implicated LIM-kinase 1 hemizygosity as a contributing factor to impaired visual-spatial constructive cognition in WS. Here we identify and characterize a novel gene, **FKBP6**, within the common WS deletion region. **FKBP6** shows homology to the FK-506 binding protein (FKBP) class of immunophilins. **FKBP6** has a putative N-terminal FK-506 binding and peptidylproyl isomerase (rotamase) domain and, like known high-molecular-weight FKBPs, an imperfect C-terminal tetratricopeptide repeat domain. **FKBP6** is expressed in testis, heart, skeletal muscle, liver, and kidney. **FKBP6** consists of nine exons and is completely contained within a 35-kb cosmid clone. Fluorescence in situ hybridization experiments show that **FKBP6** gene is deleted in 40/40 WS individuals. Hemizygous deletion of **FKBP6** may contribute to certain defects such as hypercalcemia and growth delay in WS.Copyright 1998 Academic Press. https://www.ncbi.nlm.nih.gov/pubmed/?term=9782077

SOCIAL ANXIETY (ansiedad social)

	Enfermedad	Gene	Peso
1	social anxiety	GTF2I	0.733625

- The Williams syndrome prosociality gene **GTF2I** mediates oxytocin reactivity and **social anxiety** in a healthy population. Procyshyn TL Spence J Read S Watson NV Crespi BJ The neurohormone oxytocin plays a central role in human social behaviour and cognition, and oxytocin dysregulation may contribute to psychiatric disorders. However, genetic factors influencing individual variation in the oxytocinergic system remain poorly understood. We genotyped 169 healthy adults for a functional polymorphism in **GTF2I** (general transcription factor II-I), a gene associated with high prosociality and reduced **social anxiety** in Williams syndrome, a condition reported to involve high oxytocin levels and reactivity. Participants' salivary oxytocin levels were measured before and after watching a validated empathy-inducing video. Oxytocin reactivity, defined as pre- to post-video percentage change in salivary oxytocin, varied substantially and significantly between individuals with different **GTF2I** genotypes, with, additionally, a trend towards an interaction between genotype and sex. Individuals with more oxytocin-reactive genotypes also reported significantly lower **social anxiety**. These findings suggest a model whereby **GTF2I** has a continuum of effects on human sociality, from the extreme social phenotypes and oxytocin dysregulation associated with gene deletion in Williams syndrome, to individual differences in oxytocin reactivity and sociality associated with common polymorphisms in healthy populations.Â© 2017 The Author(s). https://www.ncbi.nlm.nih.gov/pubmed/?term=28424317

- A Common Polymorphism in a Williams Syndrome Gene Predicts Amygdala Reactivity and Extraversion in Healthy Adults. Swartz JR Waller R Bogdan R Knodt AR Sabhlok A Hyde LW Hariri AR Williams syndrome (WS), a genetic disorder resulting from hemizygous microdeletion of chromosome 7q11.23, has emerged as a model for identifying the genetic architecture of socioemotional behavior. Common polymorphisms in **GTF2I**, which is found within the WS microdeletion, have been associated with reduced **social anxiety** in the general population. Identifying neural phenotypes affected by these polymorphisms would help advance our understanding not only of this specific genetic association but also of the broader neurogenetic mechanisms of variability in socioemotional behavior.Through an ongoing parent protocol, the Duke Neurogenetics Study, we measured threat-related amygdala reactivity to fearful and angry facial expressions using functional magnetic resonance imaging, assessed trait personality using the Revised NEO Personality Inventory, and imputed **GTF2I** rs13227433 from saliva-derived DNA using custom Illumina arrays. Participants included 808 non-Hispanic Caucasian, African American, and Asian university students.The **GTF2I** rs13227433 AA genotype, previously associated with lower **social anxiety**, predicted decreased threat-related amygdala reactivity. An indirect effect of **GTF2I** genotype on the warmth facet of extraversion was mediated by decreased threat-related amygdala reactivity in women but not men.A common polymorphism in the WS gene **GTF2I** associated with reduced **social anxiety** predicts decreased threat-related amygdala reactivity, which mediates an association between genotype and increased warmth in women. These results are consistent with reduced threat-related amygdala reactivity in WS and suggest that common variation in **GTF2I** contributes to broader variability in socioemotional brain function and behavior, with implications for understanding the neurogenetic bases of WS as well as **social anxiety**.Copyright Â© 2016 Society of Biological Psychiatry. Published by Elsevier Inc. All rights reserved. https://www.ncbi.nlm.nih.gov/pubmed/?term=26853120

- Cognitive-behavioral phenotypes of Williams syndrome are associated with genetic variation in the **GTF2I** gene, in a healthy population. Crespi BJ Hurd PL Individuals with Williams syndrome, a neurogenetic condition caused by deletion of a set of genes at chromosomal location 7q11.23, exhibit a remarkable suite of traits including hypersociality with high, nonselective friendliness and low **social anxiety**, expressive language relatively well-developed but under-developed social-communication skills overall, and reduced visual-spatial abilities. Deletions and duplications of the Williams-syndrome region have also been associated with autism, and with schizophrenia, two disorders centrally involving social cognition. Several lines of evidence have linked the gene **GTF2I** (General Transcription Factor IIi) with the social phenotypes of Williams syndrome, but a role for this gene in sociality within healthy populations has yet to be investigated.We genotyped a large set of healthy individuals for two single-nucleotide polymorphisms in the **GTF2I** gene that have recently been significantly associated with autism, and thus apparently exhibit functional effects on autism-related social phenotypes. **GTF2I** genotypes for these SNPs showed highly significant association with low **social anxiety** combined with reduced social-communication abilities, which represents a metric of the Williams-syndrome cognitive profile as described from previous studies.These findings implicate the **GTF2I** gene in the neurogenetic basis of social communication and **social anxiety**, both in Williams syndrome and among individuals in healthy populations. https://www.ncbi.nlm.nih.gov/pubmed/?term=25429715

AUTISM SPECTRUM DISORDER (desorden del espectro autista)

	Enfermedad	Gene	Peso
1	autism spectrum disorder	DLG4	1
2	autism spectrum disorder	GTF2IRD1	1
3	autism spectrum disorder	BAZ1B	1
4	autism spectrum disorder	LIMK1	1
5	autism spectrum disorder	USP7	1
6	autism spectrum disorder	GTF2I	1
7	autism spectrum disorder	GTF2IRD2	1
8	autism spectrum disorder	CTCF	1

- Exome sequencing of 85 Williams-Beuren syndrome cases rules out coding variation as a major contributor to remaining variance in social behavior. Kopp ND Parrish PCR Lugo M Dougherty JD Kozel BA Large, multigenic deletions at chromosome 7q11.23 result in a highly penetrant constellation of physical and behavioral symptoms known as Williams-Beuren syndrome (WS). Of particular interest is the unusual social-cognitive profile evidenced by deficits in social cognition and communication reminiscent of **autism spectrum disorders** (ASD) that are juxtaposed with normal or even relatively enhanced social motivation. Interestingly, duplications in the same region also result in ASD-like phenotypes as well as social phobias. Thus, the region clearly regulates human social motivation and behavior, yet the relevant gene(s) have not been definitively identified.Here, we deeply phenotyped 85 individuals with WS and used exome sequencing to analyze common and rare variation for association with the remaining variance in social behavior as assessed by the Social Responsiveness Scale.We replicated the previously reported unusual juxtaposition of behavioral symptoms in this new patient collection, but we did not find any new alleles of large effect in the targeted analysis of the remaining copy of genes in the Williams syndrome critical region. However, we report on two nominally significant SNPs in two genes that have been implicated in the cognitive and social phenotypes of Williams syndrome, **BAZ1B** and **GTF2IRD1**. Secondary discovery driven explorations focusing on known ASD genes and an exome wide scan do not highlight any variants of a large effect.Whole exome sequencing of 85 individuals with WS did not support the hypothesis that there are variants of large effect within the remaining Williams syndrome critical region that contribute to the social phenotype. This deeply phenotyped and genotyped patient cohort with a defined mutation provides the opportunity for similar analyses focusing on noncoding variation and/or other phenotypic domains.Â© 2018 The Authors. Molecular Genetics & Genomic Medicine published by Wiley Periodicals, Inc. https://www.ncbi.nlm.nih.gov/pubmed/?term=30008175

- A Novel Recurrent Breakpoint Responsible for Rearrangements in the Williams-Beuren Region. Plaja A Castells N Cueto-GonzÃ¡lez AM del Campo M Vendrell T Lloveras E Izquierdo L Borregan M RodrÃguez-Santiago B CarriÃ³ A MirÃ³ R Tizzano E Copy number variants (CNVs) of the Williams-Beuren syndrome (WBS) 7q11.23 region are responsible for neurodevelopmental disorders with multisystem involvement and variable expressivity. We found 2 patients with a deletion and 1 patient with a duplication in this region sharing a common breakpoint located between the **LIMK1** and EIF4H(WBSCR1) genes. One patient had a WBS phenotype, although testing with a commercially available FISH assay was negative for the deletion. A further test using array CGH showed an atypical WBS region deletion. The second patient showed global developmental delay, speech delay and poor motor skills with a deletion outside the WBS region. The third patient had manifestations compatible with an **autism spectrum disorder** showing a duplication in the WBS region. Our findings point to the existence of a previously unrecognized recurrent breakpoint responsible for rearrangements in the WBS region. Given that most commercial FISH assays include probes flanking this novel breakpoint, further testing with array CGH should be performed in patients with WBS and negative FISH results. Â© 2015 S. Karger AG, Basel. https://www.ncbi.nlm.nih.gov/pubmed/?term=26382598

- Association of **GTF2I** in the Williams-Beuren syndrome critical region with **autism spectrum disorders**. Malenfant P Liu X Hudson ML Qiao Y Hrynchak M Riendeau N Hildebrand MJ Cohen IL Chudley AE Forster-Gibson C Mickelson EC Rajcan-Separovic E Lewis ME Holden JJ Duplications of 7q11.23, deleted in Williams-Beuren Syndrome, have been implicated in **autism spectrum disorders** (ASDs). A 1.5 Mb duplication was identified in one girl with severe expressive language deficits and anxiety among 1,142 ASD individuals screened for this duplication. Family-based association studies of Tag-SNPs in three genes (STX1A , CYLN2 and **GTF2I**) in two multiplex autism family cohorts revealed strong association of two **GTF2I** SNPs and their haplotype in Cohort 1 and the combined families. The risk alleles and haplotype were associated with severe problems in social interaction and excessive repetitive behaviors. Our findings suggest the **GTF2I** gene is important in the etiology of autism in individuals with this duplication and in non-duplication cases with severe social interaction problems and repetitive behaviors. https://www.ncbi.nlm.nih.gov/pubmed/?term=22048961

- Multiple recurrent de novo CNVs, including duplications of the 7q11.23 Williams syndrome region, are strongly associated with autism. Sanders SJ Ercan-Sencicek AG Hus V Luo R Murtha MT Moreno-De-Luca D Chu SH Moreau MP Gupta AR Thomson SA Mason CE Bilguvar K Celestino-Soper PB Choi M Crawford EL Davis L Wright NR Dhodapkar RM DiCola M DiLullo NM Fernandez TV Fielding-Singh V Fishman DO Frahm S Garagaloyan R Goh GS Kammela S Klei L Lowe JK Lund SC McGrew AD Meyer KA Moffat WJ Murdoch JD O'Roak BJ Ober GT Pottenger RS Raubeson MJ Song Y Wang Q Yaspan BL Yu TW Yurkiewicz IR Beaudet AL Cantor RM Curland M Grice DE GÃ¼nel M Lifton RP Mane SM Martin DM Shaw CA Sheldon M Tischfield JA Walsh CA Morrow EM Ledbetter DH Fombonne E Lord C Martin CL Brooks AI Sutcliffe JS Cook EH Geschwind D Roeder K Devlin B State MW We have undertaken a genome-wide analysis of rare copy-number variation (CNV) in 1124 **autism spectrum disorder** (ASD) families, each comprised of a single proband, unaffected parents, and, in most kindreds, an unaffected sibling. We find significant association of ASD with de novo duplications of 7q11.23, where the reciprocal deletion causes Williams-Beuren syndrome, characterized by a highly social personality. We identify rare recurrent de novo CNVs at five additional regions, including 16p13.2 (encompassing genes **USP7** and C16orf72) and Cadherin 13, and implement a rigorous approach to evaluating the statistical significance of these observations. Overall, large de novo CNVs, particularly those encompassing multiple genes, confer substantial risks (OR = 5.6; CI = 2.6-12.0, p = 2.4 Ã— 10(-7)). We estimate there are 130-234 ASD-related CNV regions in the human genome and present compelling evidence, based on cumulative data, for association of rare de novo events at 7q11.23, 15q11.2-13.1, 16p11.2, and Neurexin 1.Copyright Â© 2011 Elsevier Inc. All rights reserved. https://www.ncbi.nlm.nih.gov/pubmed/?term=21658581

- Association of mouse **DLG4** (PSD-95) gene deletion and human **DLG4** gene variation with phenotypes relevant to **autism spectrum disorders** and Williams' syndrome. Feyder M Karlsson RM Mathur P Lyman M Bock R Momenan R Munasinghe J Scattoni ML Ihne J Camp M Graybeal C Strathdee D Begg A Alvarez VA Kirsch P Rietschel M Cichon S Walter H Meyer-Lindenberg A Grant SG Holmes A Research is increasingly linking **autism spectrum disorders** and other neurodevelopmental disorders to synaptic abnormalities ("synaptopathies"). PSD-95 (postsynaptic density-95, **DLG4**) orchestrates protein-protein interactions at excitatory synapses and is a major functional bridge interconnecting a neurexinneuroligin-SHANK pathway implicated in **autism spectrum disorders**.The authors characterized behavioral, dendritic, and molecular phenotypic abnormalities relevant to **autism spectrum disorders** in mice with PSD-95 deletion (**DLG4**â»(/)â»). The data from mice led to the identification of single-nucleotide polymorphisms (SNPs) in human **DLG4** and the examination of associations between these variants and neural signatures of Williams' syndrome in a normal population, using functional and structural neuroimaging.**DLG4**â»(/)â» showed increased repetitive behaviors, abnormal communication and social behaviors, impaired motor coordination, and increased stress reactivity and anxiety-related responses. **DLG4**â»(/)â» had subtle dysmorphology of amygdala dendritic spines and altered forebrain expression of various synaptic genes, including Cyln2, which regulates cytoskeletal dynamics and is a candidate gene for Williams' syndrome. A signifi-cant association was observed between variations in two human **DLG4** SNPs and reduced intraparietal sulcus volume and abnormal cortico-amygdala coupling, both of which characterize Williams' syndrome.These findings demonstrate that **DLG4** gene disruption in mice produces a complex range of behavioral and molecular abnormalities relevant to **autism spectrum disorders** and Williams' syndrome. The study provides an initial link between human **DLG4** gene variation and key neural endophenotypes of Williams' syndrome and perhaps corticoamygdala regulation of emotional and social processes more generally. https://www.ncbi.nlm.nih.gov/pubmed/?term=20952458
- An atypical deletion of the Williams-Beuren syndrome interval implicates genes associated with defective visuospatial processing and autism. Edelmann L Prosnitz A Pardo S Bhatt J Cohen N Lauriat T Ouchanov L GonzÃ¡lez PJ Manghi ER Bondy P Esquivel M Monge S Delgado MF Splendore A Francke U Burton BK McInnes LA During a genetic study of autism, a female child who met diagnostic criteria for **autism spectrum disorder**, but also exhibited the cognitive-behavioural profile (CBP) associated with Williams-Beuren syndrome (WBS) was examined. The WBS CBP includes impaired visuospatial ability, an overly friendly personality, excessive non-social anxiety and language delay.Using array-based comparative genomic hybridisation (aCGH), a deletion corresponding to BAC RP11-89A20 in the distal end of the WBS deletion interval was detected. Hemizygosity was confirmed using fluorescence in situ hybridisation and fine mapping was performed by measuring the copy number of genomic DNA using quantitative polymerase chain reaction.The proximal breakpoint was mapped to intron 1 of **GTF2IRD1** and the distal breakpoint lies 2.4-3.1 Mb towards the telomere. The subject was completely hemizygous for **GTF2I**, commonly deleted in carriers of the classic approximately 1.5 Mb WBS deletion, and **GTF2IRD2**, deleted in carriers of the rare approximately 1.84 Mb WBS deletion.Hemizygosity of the GTF2 family of transcription factors is sufficient to produce many aspects of the WBS CBP, and particularly implicate the GTF2 transcription factors in the visuospatial construction deficit. Symptoms of autism in this case may be due to deletion of additional genes outside the typical WBS interval or remote effects on gene expression at other loci. https://www.ncbi.nlm.nih.gov/pubmed/?term=16971481

GROWTH RETARDATION (retraso en el crecimiento)

	Enfermedad	Gene	Peso
1	growth retardation	SDS	0.961815
2	growth retardation	WBSCR22	0.937461

- Characterization of two novel genes, WBSCR20 and **WBSCR22**, deleted in Williams-Beuren syndrome. Doll A Grzeschik KH Williams-Beuren syndrome (WBS), due to a contiguous gene deletion of approximately 1.5 Mb at 7q11.23, is a complex developmental disorder with multisystemic manifestations including supravalvular aortic stenosis (SVAS) and a specific cognitive phenotype. Large repeats containing genes and pseudogenes flank the deletion breakpoints, and the mutation mechanism commonly appears to be unequal meiotic crossover. Except for elastin, hemizygosity of which is associated with supravalvular aortic stenosis, it is unknown which of the 18 genes in the deletion area contributes to the phenotype. Here, we report the identification and characterization of two novel genes, WBSCR20 and **WBSCR22**, which map to the common WBS deletion region. **WBSCR22** encodes a putative methyltransferase protein strongly expressed in heart, skeletal muscle and kidney. WBSCR20 encodes a novel protein expressed in skeletal muscle with similarity to p120 (NOL1), a 120-kDa proliferation-associated nucleolar antigen, a member of an evolutionarily conserved protein family. A highly similar putative gene, WBSCR20B, flanks the WBS deletion at the telomeric side. Hemizygous deletion of either of the novel genes might contribute to the **growth retardation**, the myopathy or the premature aging effects in the pathogenesis of WBS.Copyright 2002 S. Karger AG, Basel https://www.ncbi.nlm.nih.gov/pubmed/?term=11978965

LUMINAL STENOSIS (estenosis luminal)

	Enfermedad	Gene	Peso
1	luminal stenosis	ELN	0.800641

- mTOR (Mechanistic Target of Rapamycin) Inhibition Decreases Mechanosignaling, Collagen Accumulation, and Stiffening of the Thoracic Aorta in Elastin-Deficient Mice. Jiao Y Li G Li Q Ali R Qin L Li W Qyang Y Greif DM Geirsson A Humphrey JD Tellides G Elastin deficiency because of heterozygous loss of an **ELN** allele in Williams syndrome causes obstructive aortopathy characterized by medial thickening and fibrosis and consequent aortic stiffening. Previous work in **ELN**-null mice with a severe arterial phenotype showed that inhibition of mTOR (mechanistic target of rapamycin), a key regulator of cell growth, lessened the aortic obstruction but did not prevent early postnatal death. We investigated the effects of mTOR inhibition in **ELN**-null mice partially rescued by human **ELN** that manifest a less severe arterial phenotype and survive long term.Thoracic aortas of neonatal and juvenile mice with graded elastin deficiency exhibited increased signaling through both mTOR complex 1 and 2. Despite lower predicted wall stress, there was increased phosphorylation of focal adhesion kinase, suggestive of greater integrin activation, and increased transforming growth factor-Î²-signaling mediators, associated with increased collagen expression. Pharmacological blockade of mTOR by rapalogs did not improve **luminal stenosis** but reduced mechanosignaling (in delayed fashion after mTOR complex 1 inhibition), medial collagen accumulation, and stiffening of the aorta. Rapalog administration also retarded somatic growth, however, and precipitated neonatal deaths. Complementary, less-toxic strategies to inhibit mTOR via altered growth factor and nutrient responses were not effective.In addition to previously demonstrated therapeutic benefits of rapalogs decreasing smooth muscle cell proliferation in the absence of elastin, we find that rapalogs also prevent aortic fibrosis and stiffening attributable to partial elastin deficiency. Our findings suggest that mTOR-sensitive perturbation of smooth muscle cell mechanosensing contributes to elastin aortopathy.Â© 2017 American Heart Association, Inc. https://www.ncbi.nlm.nih.gov/pubmed/?term=28751568

- Deficient Circumferential Growth Is the Primary Determinant of Aortic Obstruction Attributable to Partial Elastin Deficiency. Jiao Y Li G Korneva A Caulk AW Qin L Bersi MR Li Q Li W Mecham RP Humphrey JD Tellides G Williams syndrome is characterized by obstructive aortopathy attributable to heterozygous loss of **ELN**, the gene encoding elastin. Lesions are thought to result primarily from excessive smooth muscle cell (SMC) proliferation and consequent medial expansion, although an initially smaller caliber and increased stiffness of the aorta may contribute to luminal narrowing. The relative contributions of such abnormalities to the obstructive phenotype had not been defined.We quantified determinants of **luminal stenosis** in thoracic aortas of **ELN**-/- mice incompletely rescued by human **ELN**. Moderate obstruction was largely because of deficient circumferential growth, most prominently of ascending segments, despite increased axial growth. Medial thickening was evident in these smaller diameter elastin-deficient aortas, with medial area similar to that of larger diameter control aortas. There was no difference in cross-sectional SMC number between mutant and wild-type genotypes at multiple stages of postnatal development. Decreased elastin content was associated with medial fibrosis and reduced aortic distensibility because of increased structural stiffness but preserved material stiffness. Elastin-deficient SMCs exhibited greater contractile-to-proliferative phenotypic modulation in vitro than in vivo. We confirmed increased medial collagen without evidence of increased medial area or SMC number in a small ascending aorta with thickened media of a Williams syndrome subject.Deficient circumferential growth is the predominant mechanism for moderate obstructive aortic disease resulting from partial elastin deficiency. Our findings suggest that diverse aortic manifestations in Williams syndrome result from graded elastin content, and SMC hyperplasia causing medial expansion requires additional elastin loss superimposed on **ELN** haploinsufficiency.Â© 2017 American Heart Association, Inc. https://www.ncbi.nlm.nih.gov/pubmed/?term=28254817

FACIAL DYSMORPHISM (dismorfismo facial)

	Enfermedad	Gene	Peso
1	facial dysmorphism	SET	0.989201
2	facial dysmorphism	ELN	0.973484
3	facial dysmorphism	LIMK1	0.887658
4	facial dysmorphism	ATR	0.769166
5	facial dysmorphism	BAZ1B	0.74337

- DNA damage response defect in Williams-Beuren syndrome. Guenat D Merla G Deconinck E Borg C Rohrlich PS Williams-Beuren syndrome (WBS, no.Â OMIM 194050) is a rare multisystem genetic disorder caused by a microdeletion on chromosomeÂ 7q11.23 and characterized by cardiovascular malformations, mental retardation, and a specific **facial dysmorphism**. Recently, we reported that a series of nonâ€'Hodgkin's lymphoma occurs in children with WBS and thus hypothesized that a predisposition to cancer may be associated with this genetic disorder. The aim of the present study was to ascertain the role played by three genes hemizygously deleted in WBS (RFC2, GTF2I and **BAZ1B**) in DNA damage response pathways. Cell proliferation, cell cycle analysis, Î³â€'H2A.X induction, and expression of DNA damage response proteins were investigated upon exposure to genotoxic treatments in WBS patientâ€'derived primary fibroblasts and in the 293TÂ cell line treated with specific siRNAs targeting RFC2, GTF2I and **BAZ1B**. An impaired hydroxyureaâ€'induced phosphorylation of CHK1 was observed in the WBS cells. However, this defective DNA damage response was not associated with an increased sensitivity to genotoxic agents. In addition, depletion of RFC2, GTF2I and **BAZ1B** using specific siRNAs did not have a significant impact on the DNA damage response in 293TÂ cells. Our results highlight that the ATRâ€'dependent DNA damage response is impaired in WBS patient cells but is also dispensable for viability when these cells undergo a genotoxic stress. The mechanism by which the ATR pathway is impaired in WBS warrants elucidation through further investigation. https://www.ncbi.nlm.nih.gov/pubmed/?term=28098859
- Molecular and phenotypic characterization of atypical Williams-Beuren syndrome. Euteneuer J Carvalho CM Kulkarni S Vineyard M Grady RM Lupski JR Shinawi M Williams-Beuren syndrome (WBS) is a multisystemic genomic disorder typically caused by a recurrent Ëœ1.5-1.8â€‰Mb deletion on 7q11.23. Atypical deletions can provide important insight into the genotype-phenotype correlations. Here, we report the phenotypic and molecular characterization of a girl with a de novo 81.8â€‰kb deletion in the WBS critical region, which involves the **ELN** and **LIMK1** genes only. The patient presented at 2â€‰months of age with extensive vascular abnormalities, mild **facial dysmorphism** and delays in her fine motor skills. We discuss potential molecular mechanisms and the role of **ELN** and **LIMK1** in the different phenotypic features. We compare the findings in our patient with previously reported overlapping deletions. The phenotypic variability among these patients suggests that other factors are important in the phenotype and possibly include: position effects related to copy number variation size, variations in the non-deleted alleles, genetic modifiers elsewhere in the genome, or reduced penetrance for specific phenotypes. Â© 2013 John Wiley & Sons A/S. Published by John Wiley & Sons Ltd. https://www.ncbi.nlm.nih.gov/pubmed/?term=24246242
- An atypical 7q11.23 deletion in a normal IQ Williams-Beuren syndrome patient. Ferrero GB Howald C Micale L Biamino E Augello B Fusco C Turturo MG Forzano S Reymond A Merla G Williams-Beuren syndrome (WBS; OMIM no. 194050) is a multisystemic neurodevelopmental disorder caused by a hemizygous deletion of 1.55 Mb on chromosome 7q11.23 spanning 28 genes. Haploinsufficiency of the **ELN** gene was shown to be responsible for supravalvular aortic stenosis and generalized arteriopathy, whereas **LIMK1**, CLIP2, GTF2IRD1 and GTF2I genes were suggested to be linked to the specific cognitive profile and craniofacial features. These insights for genotype-phenotype correlations came from the molecular and clinical analysis of patients with atypical deletions and mice models. Here we report a patient showing mild WBS physical phenotype and normal IQ, who carries a shorter 1 Mb atypical deletion. This rearrangement does not include the GTF2IRD1 and GTF2I genes and only partially the **BAZ1B** gene. Our results are consistent with the hypothesis that hemizygosity of the GTF2IRD1 and GTF2I genes might be involved in the **facial dysmorphisms** and in the specific motor and cognitive deficits observed in WBS patients. https://www.ncbi.nlm.nih.gov/pubmed/?term=19568270

SPASMS (espasmos)

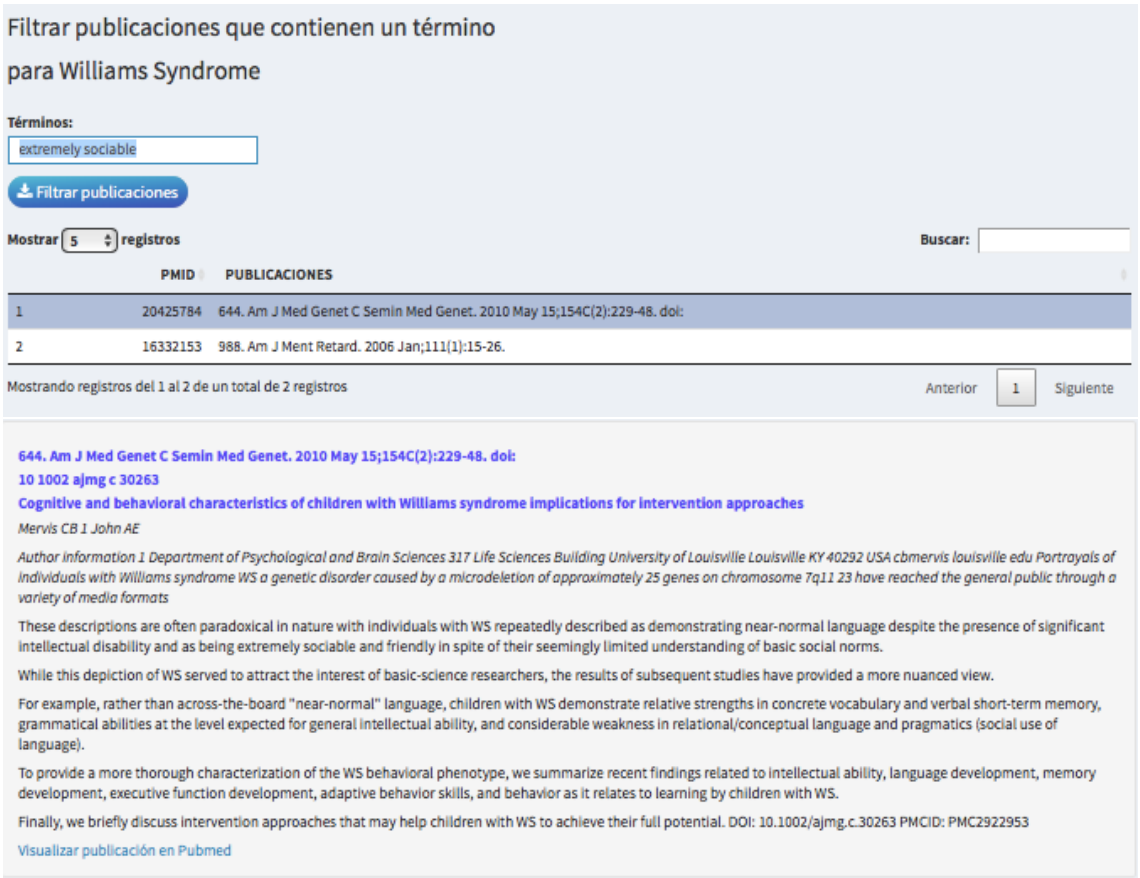
	Enfermedad	Gene	Peso
1	spasms	ELN	0.796079
2	spasms	HIP1	0.777212
3	spasms	GABRA1	0.705843

- Do the exome: A case of Williams-Beuren syndrome with severe epilepsy due to a truncating de novo variant in **GABRA1**. Popp B Trollmann R BÃ¼ttner C Caliebe A Thiel CT HÃ¼ffmeier U Reis A Zweier C Williams-Beuren syndrome (WBS) is a relatively common, clinically recognizable microdeletion syndrome. In most cases the typical heterozygous deletion of 1.5Â Mb on chromosome 7q11.23 spanning about 26 genes can be identified. Also some larger or smaller atypical deletions have been reported and associated with additional or atypical phenotypic aspects. We report on an individual with typical WBS due to the common deletion and with refractory infantile **spasms**. Using trio-exome sequencing, we identified a de novo truncating variant c.1200del, p (Lys401Serfs*25) in **GABRA1** as the likely cause of the early onset epilepsy. This unique case not only allows to further define the phenotypic spectrum of infantile epileptic encephalopathy associated with rare de novo **GABRA1** variants but exemplifies the need for a sensitive review of unclear associations in clinically defined syndromes and for extended diagnostic work-up in individuals with unusual presentations of a genetically confirmed diagnosis. Copyright Â© 2016 Elsevier Masson SAS. All rights reserved. https://www.ncbi.nlm.nih.gov/pubmed/?term=27613244

- Zebrafish gene knockdowns imply roles for human YWHAG in infantile **spasms** and cardiomegaly. Komoike Y Fujii K Nishimura A Hiraki Y Hayashidani M Shimojima K Nishizawa T Higashi K Yasukawa K Saitsu H Miyake N Mizuguchi T Matsumoto N Osawa M Kohno Y Higashinakagawa T Yamamoto T Williams-Beuren syndrome (WBS) is a neurodevelopmental disorder presenting with an elfin-like face, supravalvular aortic stenosis, a specific cognitive-behavioral profile, and infantile hypercalcemia. We encountered two WBS patients presenting with infantile **spasms**, which is extremely rare in WBS. Array comparative genomic hybridization (aCGH) and fluorescent in situ hybridization (FISH) analyses revealed atypical 5.7-Mb and 4.1-Mb deletions at 7q11.23 in the two patients, including the WBS critical region and expanding into the proximal side and the telomeric side, respectively. On the proximal side, AUTS2 and CALN1 may contribute to the phenotype. On the telomeric side, there are two candidate genes **HIP1** and YWHAG. Because detailed information of them was unavailable, we investigated their functions using gene knockdowns of zebrafish. When zebrafish ywhag1 was knocked down, reduced brain size and increased diameter of the heart tube were observed, indicating that the infantile **spasms** and cardiomegaly seen in the patient with the telomeric deletion may be derived from haploinsufficiency of YWHAG.2010 Wiley-Liss, Inc. https://www.ncbi.nlm.nih.gov/pubmed/?term=20146355

Es posible realizar búsquedas de cualquiera de estas enfermedades o cualquier otro término usando la opción de menú *Filtrar publicaciones por término*, como también buscar las publicaciones en las que se menciona a un gen concreto desde *Filtrar publicaciones por gen*.

Por ejemplo, se realiza la búsqueda del término “extremely sociable”, que es una de las características de los individuos que padecen este síndrome, y encontramos que en 2 documentos del corpus se menciona, pudiendo acceder al contenido de los mismos:



3. Conclusiones

Con este trabajo se ha usado la minería de textos en el campo de la genética para poder obtener relaciones de similitud entre genes y enfermedades, usando los algoritmos de Análisis Semántico Latente(LSA) y de la similitud del coseno, sobre los abstracts extraidos de PUBMED.

La herramienta web que se ha implementado es pública, y se encuentra disponible en la url http://pilarnatividad.shinyapps.io/genio/.

Se han encontrado todas los genes y enfermedades asociados al Síndrome de Williams-Beuren, y se ha conseguido averiguar qué genes aparecen en los mismos párrafos que las enfermedades buscadas, pero en algunos casos aunque aparecen en el mismo párrafo no significa que realmente estén relacionados.

El algoritmo LSA ha sido más permisivo, encontrando más coincidencias que las halladas por la similitud del coseno.

En algunos casos se ha confundido la aparición de agrupaciones de palabras como HR o KY, que son iniciales de nombres, considerándolas genes, ya que sus nombres coinciden.

Si evaluamos el alcance de los objetivos propuestos podemos concluir que han sido alcanzados, ya que sí que se ha conseguido crear una herramienta web que puede ser útil para el investigador, que permite extraer información valiosa de cantidades ingentes de datos.

Se propone como mejora investigar cómo solucionar problemas de confusión de genes, así como la forma de obtener las enfermedades relacionadas con los términos de búsqueda, ya que en este momento se usa el servidor de Pubtator, para extraer esta información de los abstracts analizados, y puede confundirse como un ataque a este servidor, debido al número tan alto de peticiones que se relizan en pocos minutos. Es por ello, que se ha restringido la búsqueda en Pubtator a los abstracts publicados en los últimos dos años.

Otro problema que se tuvo con Pubtator es que la función utilizada para extraer información usaba una url que dejó de funcionar en la realización de la segunda fase del proyecto, por lo que se decidió implementar una función que incluyese la nueva url permitida por Pubtator.

Shiny ha resultado ser una buena herramienta para poder crear una web interactiva en la que se puede incluir código R de una forma muy intuitiva, haciéndola muy interesante para todo bioinformático.

4. Glosario

Abstract, un articulo científico publicado en una revista medica.
Corpus, colección de textos
Delección, en genética consiste en la pérdida de un fragmento de ADN de un cromosoma.
FISH, técnica citogenética para marcar los cromosomas mediante hibridación con sondas, que emiten fluorescencia, permitiendo detectar anomalías en los mismos.
HGNC, comité de la Organización del Genoma Humano encargado asignar un nombre único y con sentido a cada uno de los genes humanos que se van conocieno.
HTML o HyperText Markup Language, es un lenguaje de programación dirigido al desarrollo de página web.
Inputs, entradas en una aplicación.
LSA, es un modelo computacional usado para encontrar similitudes semánticas en palabras que aparecen juntas o en contextos similares.
MEDLINE, es una base de datos bibliográfica producida por la National Library of Medicine de los Estados Unidos.
MESH, palabras clave usadas por los documentalistas, para normalizar las búsquedas en MEDLINE.
Microsatélite, son secuencias de ADN con n repeticiones en tándem, una a continuación de la otra, usados en genética para detectar marcadores moleculares.
minería de textos, técnica para extraer información útil en formatos heterogéneos, como artículos de revistas, libros, correos electrónicos, etc..
Ómica, término que en Biología molecular se utiliza como sufijo para referirse al estudio de la totalidad, y hace referencia a disciplinas como la genómica, la proteómica, la trascriptómica y la metabolómica.
Outputs, salidas o resultados en un aplicación.
PMID, identificador único de los abstracts publicados en PUBMED, compuesto por ocho dígitos.
PUBMED, es un sistema de búsqueda, desarrollado por National Center for Biotechnology Information (NCBI), para acceder a bases de datos bibliográficas como Genbak, Complete Genoma, MEDLINE y PreMEDLINE.
PUBTATOR, herramienta web que usa minería de textos sobre los abstracts publicados en PUBMED, para buscar genes, enfermedades, elementos químicos, entidades y mutaciones.
Shiny, es un paquete de R desarrollado por RStudio, usado para construir páginas web interactivas, basada en programación reactiva, que vincula valores de entrada con los de salida.
Síndrome, conjunto de síntomas que se presentan juntos y que son característicos de un cuadro patológico, causado normalmente por más de una enfermedad.
SVD, técnica de Descomposición en Valores Singulares
Tokenizar, dividir un texto en partes más pequeñas, como pueden ser párrafos o palabras
url, dirección específica que se asigna a cada uno de los recursos disponibles en la red para permitir su localización.

5. Bibliografía

A.Kaur, & Chopra, D. (2016). *Comparison of Text Mining Tools.* https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=7784950.

C.GALVEZ. (2008). *Minería de textos: la nueva generación de análisis de literatura científica en Biología molecular y Genómica.*

Casanelles, M. d., & Jurado, L. P. (2010). *PROTOCOLO DE SEGUIMIENTO EN EL SÍNDROME DE WILLIAMS.*

Deerwester, S. (1990). *Indexing by Latent Semantic Analysis.* http://lsa.colorado.edu/papers/JASIS.lsi.90.pdf.

F.WILD. (2008). *Latent Semantic Analysis.*

F.WILD. (s.f.). *An LSA Package for R.*

F.WILD, & STAHL., C. (2006). *Investigating Unstructured Texts with Latent Semantic Analysis.*

Feinerer, I., & Hornik, K. (2018). *Text Mining Package.*

Guzmán, E. (s.f.). *Minería de datos.* Universidad Nacional de Colombia: http://www.disi.unal.edu.co/profesores/eleonguz/cursos/md/presentaciones/Sesion11_Agrupacion.pdf.

J.Mulero. (s.f.). *Aplicaciones interactivas diseñadas con Shiny.* https://rua.ua.es/dspace/bitstream/10045/54325/1/shiny.pdf.

J.RANI, SHAH, A., & RAMACHANDRAN., S. (2015). *pubmed.mineR: An R package with text-mining algorithms to analyse PubMed abstracts.*

JURADO, L. P. (1999). *Guía clínica para el Síndrome de Williams o Williams-Beuren.* http://www.sindromewilliams.org/wp-content/uploads/2017/09/Guia-clinica-para-el-S%C3%ADndrome-de-Williams.pdf.

Jyoti Rani, S. A. (2018). *Text Mining of PubMed Abstracts.* https://cran.r-project.org/web/packages/pubmed.mineR/pubmed.mineR.pdf.

M.A.Fernández-Altunaa, A, M. d., Rodríguez, E., Rayón, D., Castillo, H. A., & Guinzberg, A. (2016). *Uso de los MeSH: una guía práctica.* https://www.sciencedirect.com/science/article/pii/S2007505716000302.

P.ZWEIGENBAUM, D.DEMNER-FUSHMAN, YU, H., & COHEN, K. B. (2007). *Frontiers of biomedical text mining:current progress.*

Ramachandran, S. (s.f.). *Código Fuente funciones pubmed.mineR.* https://github.com/cran/pubmed.mineR/tree/master/R.

SLOMOVITZ, G. (2016). *LSA (Latent Semantic Analisys) en registros médicos.* https://eva.fing.edu.uy/mod/resource/view.php?id=62478.

SO, E., & KIM, J. (2015). *WILLIAMS SYNDROME.* http://williams-disorder.weebly.com/causes-detection-and-diagnosis.html.

TORRE, M. J. (2008). *Nuevas técnicas de minería de textos: Aplicaciones. .*

6. Anexo

```
library(shiny)
library(shinydashboard)
library(shinyWidgets)
library(easyPubMed)
library(DT)
library(wordcloud)
library(pubmed.mineR)
library(lsa)
library(igraph)
library(stringr)
library(textreuse)
library(readr)
library(R2HTML)
library(htmltools)
library(shinyjs)
library(V8)
library(plyr)
library(RCurl)
utils::globalVariables(c("HGNC2UniprotID", "HGNCdata"))
jsResetCode <- "shinyjs.reset = function() {history.go(0)}"

header <- dashboardHeader(title =tags$img(src="gen-io-Logo.png", width="180px"),
                          tags$li(class = "dropdown",
                                  tags$style(".main-header {max-height: 80px }"),
                                  tags$style(".main-header .logo {height: 80px}")
                          ))

sidebar<-  dashboardSidebar(
  tags$style(".left-side, .main-sidebar {padding-top: 150px}"),
  sidebarMenu(id="sidebarmenu",
              menuItem("Busqueda en Pubmed", tabName = "a",  icon= icon("download", lib="font-awesome")),
              menuItem("Explorar publicaciones", tabName = "b", icon= icon("globe", lib="font-awesome")),
              menuItem("Frecuencia palabras", tabName = "c", icon= icon("list-alt", lib="font-awesome")),
              menuItem("Frecuencia genes", tabName = "d", icon= icon("check-circle", lib="font-awesome")),
              menuItem("Nube Palabras y Genes", tabName = "e", icon= icon("bar-chart-o", lib="font-awesome")),
              menuItem("Filtrar public. por término", tabName = "f", icon= icon("newspaper-o", lib="font-awesome")),
              menuItem("Filtrar public. por genes", tabName = "g", icon= icon("gears", lib="font-awesome")),
              menuItem("Relaciones Similitud", tabName = "h", icon= icon("refresh", lib="font-awesome")),
              menuItem("Gráficos Similitud.", tabName = "i", icon= icon("cog", lib="glyphicon"))
  ))

body <- dashboardBody(
  useShinyjs(),
  tags$style(type="text/css",
             ".shiny-output-error { visibility: hidden; }",
             ".shiny-output-error:before { visibility: hidden; }"
  ),
  tabItems(
    tabItem(tabName="a",
            h2("Búsqueda publicaciones en Pubmed"),
            br(),
            useShinyjs(),
            extendShinyjs(text=jsResetCode),
            textInput("busqueda", "Términos:", value="williams syndrome"),
            dateInput("date1", "Desde fecha publicación:", value="1980-01-01", format="yyyy-mm-dd"),
            dateInput("date2", "Hasta fecha:", value=Sys.Date(), format="yyyy-mm-dd"),
            actionLink(inputId ="createPub",
                       label="Visualizar consulta Pubmed",
                       color = "primary",
                       icon=icon("tv")),
            tags$div(id="espacio1",br()),
            verbatimTextOutput("nText"),
            tags$div(id="espacio2",br()),
            actionBttn(inputId ="downPub",
                       label="Buscar",
                       color = "primary",
                       style= "gradient",
                       icon=icon("download"),
                       block = FALSE,
                       size = "md"),
            hidden(actionBttn(inputId ="nuevaBusq",
                              label="Nueva búsqueda",
                              color = "primary",
                              style= "gradient",
                              icon=icon("search"),
                              block = FALSE,
                              size = "md")),
            br(),br(),
            tags$div(style='cursor:pointer',dataTableOutput("pmidResultado")),
            br(),
            wellPanel(htmlOutput("panelPublicacion")),
            br(),br()
    ),

    tabItem(tabName="b",
            h2("Explorar publicaciones Pubmed "),
```

```
            htmlOutput("tituloEnfBus"),
            br(),
            helpText("Seleccione una publicación para obtener la información sobre genes, enfermedades, mutaciones,
elementos químicos
                     y entidades que aparece en la misma"),
            uiOutput("abs"),
            br(),br(),
            tags$div(style='cursor:pointer',dataTableOutput("viewPubtatorGen")),
            tags$div(style='cursor:pointer',dataTableOutput("viewPubtatorDis")),
            tags$div(style='cursor:pointer',dataTableOutput("viewPubtatorMut")),
            tags$div(style='cursor:pointer',dataTableOutput("viewPubtatorChe")),
            tags$div(style='cursor:pointer',dataTableOutput("viewPubtatorEnt")),
            br(),br()
            ),

    tabItem(tabName="c",
            h2("Frecuencia de palabras"),
            htmlOutput("tituloFecPalabras"),
            br(),
            tags$div(style='cursor:pointer',dataTableOutput("frecWords")),
            br(),br()
    ),

    tabItem(tabName="d",
            h2("Frecuencia de genes"),
            htmlOutput("tituloFecGenes"),
            br(),
            tags$div(style='cursor:pointer',dataTableOutput("frecGene")),
            br(),
            wellPanel(htmlOutput("urlGenSelec")),
            br(),br()
    ),

    tabItem(tabName="e",
            h2("Nube de Palabras y Genes"),
            htmlOutput("tituloNube"),
            br(),
            sidebarLayout(
              sidebarPanel(
                selectInput("selection", "Elije:",
                            choices = c("Palabras","Genes")),
                actionButton("update", "Dibuja",class = "btn-primary"),
                hr(),
                sliderInput("freq",
                            "Mínima frecuencia:",
                            min = 1,  max = 50, value = 3),
                sliderInput("maxw",
                            "Máximo núm. palabras/genes:",
                            min = 1,  max = 300,  value = 100)
              ),
              mainPanel(
                plotOutput("plot")
              )
            ),
            br(),br()),

    tabItem(tabName="f",
            h2("Filtrar publicaciones que contienen un término"),
            htmlOutput("tituloFiltroTerm"),
            br(),
            useShinyjs(),
            extendShinyjs(text=jsResetCode),
            textInput("busquedaEnCorpus", "Términos:", value="mental retardation"),
            actionBttn(inputId ="btnNuevoCorpus",
                       label="Filtrar publicaciones",
                       color = "primary",
                       style= "gradient",
                       icon=icon("download"),
                       block = FALSE,
                       size = "md"),
            br(),br(),
            tags$div(style='cursor:pointer',dataTableOutput("pmidCorpSec")),
            br(),
            wellPanel(htmlOutput("panelPublicacion2")),
            br(),br()
    ),

    tabItem(tabName="g",
            h2("Filtrar public. por genes"),
            htmlOutput("tituloFiltroGenes"),
            br(),
            uiOutput("dataframeGenes"),
            br(),
            wellPanel(htmlOutput("mostrarIdGen")),
            br(),
            tags$div(style='cursor:pointer',dataTableOutput("pmidCorpSecGen")),
            br(),
            wellPanel(htmlOutput("panelPublicacionGen")),
            tags$div(style='cursor:pointer',dataTableOutput("datosGenes")),
            tags$div(style='cursor:pointer',dataTableOutput("datosDisease")),
            tags$div(style='cursor:pointer',dataTableOutput("datosMutation")),
```

```
            tags$div(style='cursor:pointer',dataTableOutput("datosChemicals")),
            tags$div(style='cursor:pointer',dataTableOutput("datosEnti")),
            br(),br()
    ),

    tabItem(tabName="h",
            h2("Relación de Similitud aplicando Análisis de la Semántica Latente"),
            htmlOutput("tituloRelSim"),
            br(),
            h4("Seleccione un enfermedad de la tabla y pulse el botón para obtener información de los genes
relacionados:"),
            br(),
            actionBttn(inputId ="downPub2",
                       label="Buscar Enfermedades",
                       color = "primary",
                       style= "gradient",
                       icon=icon("download"),
                       block = FALSE,
                       size = "md"),
            tags$div(style='cursor:pointer',dataTableOutput("listadoDisease2")),
            hidden( actionBttn(inputId ="btnRelEnfGen",
                               label="Obtener Genes relacionados",
                               color = "primary",
                               style= "gradient",
                               icon=icon("sliders"),
                               block = FALSE,
                               size = "md")),
            br(),br(),
            tags$div(style='cursor:pointer',dataTableOutput("relLSA")),
            br(),br()
    ),

    tabItem(tabName="i",
            h2("Gráfico de Similitud del Coseno"),
            htmlOutput("tituloGrafRelSim"),
            br(),
            actionBttn(inputId ="btnGraf",
                       label="Obtener Gráfico",
                       color = "primary",
                       style= "gradient",
                       icon=icon("download"),
                       block = FALSE,
                       size = "md"),
            wellPanel(htmlOutput("tituloGraf")),
            plotOutput("plotRelaciones",width='1000px', height = '800px' ),
            br(),br(),
            tags$div(style='cursor:pointer',dataTableOutput("tablaCos")),
            br(),br()
    )
  ),
  tags$footer("2018 GEN-IO (TFM MINERIA DE TEXTOS UOC)",
              tags$a(href='mailto:pilarnatividad@gmail.com' ,target='_top','Email: pilarnatividad@gmail.com'),
              style = "
              position:absolute;
              bottom:0;
              width:100%;
              height:50px;
              color: #999999;
              padding: 20px;
              z-index: 1000;")
  )

ui <- dashboardPage(title="gen-IO",header, skin = "blue",sidebar, body)

server <- function(input, output) {
  source("funciones.R", local = TRUE)
  nText<- eventReactive( input$createPub,{paste(c(input$busqueda,
                                                  "[MH] AND " ,
                                                  as.character(input$date1,"%Y/%m/%d"),
                                                  ":",
                                                  as.character(input$date2,"%Y/%m/%d"),
                                                  "[DP]"),
                                                collapse="")})

  output$nText<- renderText(nText())
  pubmedResult  <-eventReactive( input$downPub,
            {withProgress(message = 'Descargando Abstracts desde Pubmed...',
                          detail = 'Espere...', value = 0, {
                          incProgress(7/15)
                           out.A<-batch_pubmed_download(pubmed_query_string =
                                     paste(c(input$busqueda,
                                             "[Mesh] AND " ,
                                             as.character(input$date1,"%Y/%m/%d"),
                                             ":",
                                             as.character(input$date2,"%Y/%m/%d"),
                                             "[DP]"),
                                           collapse=""),
                                           format = "abstract",
                                           batch_size = 1500,
                                           dest_file_prefix = "pubmed_")
                           file.create("pubmed_result.txt")
```

```
                              ficheros<-out.A
                              for(i in 1:length(ficheros)){
                                file.append("pubmed_result.txt",ficheros[i])

                              }
                              file.rename("pubmed_result.txt","corpus.txt")
                              pmr<- readabs("corpus.txt")
                              incProgress(15/15)
                                          })
                                return(pmr)

                              }
)

output$pmidResultado <- renderDataTable({
  corpus<- pubmedResult()
  pmidRes<- data.frame(corpus@PMID,corpus@Journal)
  colnames(pmidRes)<-c("PMID", "PUBLICACIONES")
  datatable(pmidRes,
            selection=list(mode='single', selected=1),
            options = list(language = list(url = 'Spanish.json'),
                           pageLength = 5
            ))
})

observeEvent(input$downPub,{
  hide("busqueda")
  hide("date1")
  hide("date2")
  show("nuevaBusq")
  hide("downPub")
  hide("createPub")
  hide("espacio1")
  hide("espacio2")
})
observeEvent(input$nuevaBusq,{
  show("busqueda")
  show("date1")
  show("date2")
  show("downPub")
  show("createPub")
  hide("nuevaBusq")
  js$reset()
})

output$panelPublicacion = renderPrint({
  s = input$pmidResultado_rows_selected
  pmr<- pubmedResult()
  mostrar_info(s,pmr)
})

output$tituloFiltroTerm <- renderPrint({
  cat('<h2>para <font style="text-transform: capitalize;">',input$busqueda,'</font></h2>')
})

pmedCorSec  <-eventReactive( input$btnNuevoCorpus,
                             { pmr<- pubmedResult()
                             corpusSecundario<- searchabsL(pmr, include=input$busquedaEnCorpus)
                             return(corpusSecundario)})

output$pmidCorpSec <- renderDataTable({
  corpusSecundario<-pmedCorSec()
  pmidCorpusSec<- data.frame(corpusSecundario@PMID,corpusSecundario@Journal)
  colnames(pmidCorpusSec)<-c("PMID", "PUBLICACIONES")
  datatable(pmidCorpusSec,
            selection=list(mode='single', selected=1),
            options = list(language = list(url = 'Spanish.json'),
                           pageLength = 5
            ))
})

output$panelPublicacion2 = renderPrint({
  s = input$pmidCorpSec_rows_selected
  pmr<- pmedCorSec()
  mostrar_info(s,pmr)
})

output$tituloEnfBus <- renderPrint({
  cat('<h2>para <font style="text-transform: capitalize;">',input$busqueda,'</font></h2>')
})

output$abs<- renderUI({
  corpus<- pubmedResult()
  vectorAbs<- as.data.frame(corpus@Journal)
  colnames(vectorAbs)<-"Elija una Publicación..."
  selectInput("listaAbstracts","Publicación:", choices = vectorAbs, width='800px')
})

viewInfoPubtator<- eventReactive( c(input$listaAbstracts),{
  pmr <- pubmedResult()
  pmidPubtator<- pubtator_new_function(pmr@PMID[which(pmr@Journal==input$listaAbstracts)])
```

```
  return(pmidPubtator)
})

output$viewPubtatorGen<-renderDataTable({
  wPubtator<-viewInfoPubtator()
  tabla<-data.frame(Genes=wPubtator$Genes)
  cap=tags$h3("Genes")
  datatable(tabla,
            caption=cap,
            selection=list(mode='single', selected=1),
            options = list(language = list(url = 'Spanish.json'),
                           pageLength = 5
            ))
})

output$viewPubtatorDis<-renderDataTable({
  wPubtator<-viewInfoPubtator()
  tabla<- data.frame(Enfermedades=wPubtator$Disease)
  cap=tags$h3("Enfermedades")
  datatable(tabla,
            caption=cap,
            selection=list(mode='single', selected=1),
            options = list(language = list(url = 'Spanish.json'),
                           pageLength = 5
            ))
})

output$viewPubtatorMut<-renderDataTable({
  wPubtator<-viewInfoPubtator()
  tabla<-  data.frame(Mutaciones=wPubtator$Mutations)
  cap=tags$h3("Mutaciones")
  datatable(tabla,
            caption=cap,
            selection=list(mode='single', selected=1),
            options = list(language = list(url = 'Spanish.json'),
                           pageLength = 5
            ))
})

output$viewPubtatorChe<-renderDataTable({
  wPubtator<-viewInfoPubtator()
  tabla<- data.frame(Elementos=wPubtator$Chemicals)
  cap=tags$h3("Elementos químicos")
  datatable(tabla,
            caption=cap,
            selection=list(mode='single', selected=1),
            options = list(language = list(url = 'Spanish.json'),
                           pageLength = 5
            ))
})

output$viewPubtatorEnt<-renderDataTable({
  wPubtator<-viewInfoPubtator()
  tabla<- data.frame(Entidades=wPubtator$Species)
  cap=tags$h3("Entidades")
  datatable(tabla,
            caption=cap,
            selection=list(mode='single', selected=1),
            options = list(language = list(url = 'Spanish.json'),
                           pageLength = 5
            ))
})

pubmedResultWords<- reactive({
  pmr<- pubmedResult()
  pmrWord<- word_atomizations(pmr)
  row.names(pmrWord)<-NULL
  return(pmrWord)
})

output$tituloFecPalabras <- renderPrint({
  cat('<h2>en publicaciones sobre <font style="text-transform: capitalize;">',input$busqueda,'</font></h2>')
})

output$frecWords<-renderDataTable({
  withProgress(message = 'Obteniendo información palabras',
               detail = 'Espere...', value = 0, {
                 incProgress(2/15)
                 tableWords<-pubmedResultWords()
                 incProgress(15/15)
               })
  colnames(tableWords)<-c("Palabras","Frecuencia")
  datatable(tableWords,
            selection=list(mode='single', selected=1),
            options = list(language = list(url = 'Spanish.json'),
                           pageLength = 15
            ))
})

output$tituloFecGenes <- renderPrint({
  cat('<h2>en publicaciones sobre <font style="text-transform: capitalize;">',input$busqueda,'</font></h2>')
```

```
  })

  getFrecGene<- reactive({
    pmr<- pubmedResult()
    freqGenes<- gene_atomization(pmr)
    return(freqGenes)
  })

  getTablaGenes<- reactive({
    freqGenes<- getFrecGene()
    tablaGenes<-freqGenes[,2]
    return(tablaGenes)
  })

  output$tablaGenes<-renderDataTable({
    tab<-getTablaGenes()
    tab
  })

  output$frecGene<-renderDataTable({
    withProgress(message = 'Obteniendo información genes.',
                 detail = 'Puede tardar unos minutos. Espere...', value = 0, {
                   incProgress(7/15)
                   tabGene<-getFrecGene()
                   incProgress(15/15)
                 })
    datatable(tabGene,
              selection=list(mode='single', selected=1),
              options = list(language = list(url = 'Spanish.json'),
                             pageLength =15
              ))
  })

  output$urlGenSelec <- renderPrint({
    s = input$frecGene_rows_selected
    tabGene<-getFrecGene()
    gen<-tabGene[s,1]
    local_uniprotfun(gen)
    file.rename("x.txt","www/x.csv")
    cat(paste("<a   href='../x.csv'   target=_blank   ><img   src='Logo_uniprot.png'   alt='Información   de   Uniprot'>
",gen,"</a>"))
  })

  output$tituloFiltroGenes <- renderPrint({
    cat('<h2>en publicaciones sobre <font style="text-transform: capitalize;">',input$busqueda,'</font></h2>')
  })

  output$dataframeGenes<- renderUI({
    vectorGenes<-getTablaGenes()
    selectInput("genes","Genes:", choices = vectorGenes, label="Genes:",width='350px')
  })

  output$mostrarIdGen<- renderPrint({
    withProgress(message = 'Obteniendo información genes.',
                 detail = 'Puede tardar unos minutos. Espere...', value = 0, {
                   incProgress(7/15)
                   genElegido<-input$genes
                   freqGenes<- getFrecGene()
                   idGen<- freqGenes[(which(freqGenes[,2]==genElegido)),1]
                   incProgress(15/15)
                 })

    cat(paste('<h4>Corpus secundario con las publicaciones en las que aparece el gen <b>',idGen,'</b></h4>'))
  })

  corpusDeUnGen<- eventReactive (c(input$genes), {
    pmr<- pubmedResult()
    genElegido<-input$genes
    tablaGenes<-getFrecGene()
    v<-(which(tablaGenes[,2]==genElegido))
    vecGen <-c()
    vecGen <- c(vecGen,tablaGenes[v,])
    vecGen <-unlist(vecGen[1:2], use.names=FALSE)
    vecGen<-as.vector(vecGen)
    corpusGen<- searchabsL(pmr, include=vecGen)
    return(corpusGen)
  })

  output$pmidCorpSecGen <- renderDataTable({
    corpusSecundario<- corpusDeUnGen()
    pmidCorpusSec<- data.frame(corpusSecundario@PMID,corpusSecundario@Journal)
    colnames(pmidCorpusSec)<-c("PMID", "PUBLICACIONES")
    datatable(pmidCorpusSec,
              selection=list(mode='single', selected=1),
              options = list(language = list(url = 'Spanish.json'),
                             pageLength = 5
              ))
  })

  output$panelPublicacionGen = renderPrint({
    s = input$pmidCorpSecGen_rows_selected
```

```
  pmr<- corpusDeUnGen()
  mostrar_info(s,pmr)
})

infoDeUnGen<- eventReactive (c(input$genes), {

  genFiltro<- corpusDeUnGen()
  genPmid<-genFiltro@PMID
  genSelInfo<- pubtator_new_function(genPmid)
  return(genSelInfo)
})

output$datosGenes<-renderDataTable({
  wPubtator<- infoDeUnGen()
  tabla<-data.frame(Genes=wPubtator$Genes)
  cap=tags$h3("Genes")
  datatable(tabla,
            caption=cap,
            options = list(language = list(url = 'Spanish.json'),
                           pageLength = 10
            ))
})

output$datosDisease<-renderDataTable({
  wPubtator<- infoDeUnGen()
  tabla<- data.frame(Enfermedades=wPubtator$Disease)
  cap=tags$h3("Enfermedades")
  datatable(tabla,
            caption=cap,
            options = list(language = list(url = 'Spanish.json'), pageLength = 10
            ))
})

output$datosMutation<-renderDataTable({
  wPubtator<- infoDeUnGen()
  tabla<-  data.frame(Mutaciones=wPubtator$Mutations)
  cap=tags$h3("Mutaciones")
  datatable(tabla,
            caption=cap,
            options = list(language = list(url = 'Spanish.json'), pageLength = 10
            ))
})

output$datosChemicals<-renderDataTable({
  wPubtator<- infoDeUnGen()
  tabla<- data.frame(Elementos=wPubtator$Chemicals)
  cap=tags$h3("Elementos químicos")
  datatable(tabla,
            caption=cap,
            options = list(language = list(url = 'Spanish.json'), pageLength = 10
            ))
})

output$datosEnti<-renderDataTable({
  wPubtator<- infoDeUnGen()
  tabla<- data.frame(Entidades=wPubtator$Species)
  cap=tags$h3("Entidades")
  datatable(tabla,
            caption=cap,
            options = list(language = list(url = 'Spanish.json'), pageLength = 10
            ))
})

output$tituloNube <- renderPrint({
  cat('<h2>en publicaciones sobre <font style="text-transform: capitalize;">',input$busqueda,'</font></h2>')
})

terms <- reactive({
  c(input$update,input$selection)
  isolate({
    withProgress({
      setProgress(message = "Processing corpus...")
      if (input$selection=="Palabras") {twordcloud<-as.data.frame(pubmedResultWords())}
      if (input$selection=="Genes") {twordcloud<-as.data.frame(getFrecGene())}
    })
  })
  return(twordcloud)
})

wordcloudRep <- repeatable(wordcloud)

output$plot <- renderPlot({
  v <- terms()
  if (input$selection=="Palabras") {wordcloudRep(v$words, v$Freq, scale=c(4,0.5),
                                                 min.freq = input$freq, max.words=input$maxw,
                                                 colors=brewer.pal(8, "Dark2"))}
  if (input$selection=="Genes") {
    wordcloudRep(v$Gene_symbol, as.numeric(v$Freq), scale=c(4,0.5),
                 min.freq = input$freq, max.words=input$maxw,
                 colors=brewer.pal(8, "Dark2"))}
})
```

```
  output$tituloRelSim <- renderPrint({
    cat('<h2>para <font style="text-transform: capitalize;">',input$busqueda,'</font></h2>')
  })

  pubmedEnfer  <-eventReactive( input$downPub2,
        {out.A<-batch_pubmed_download(pubmed_query_string =
paste(c(input$busqueda,"[Mesh] AND ",as.character(Sys.Date()-730,"%Y/%m/%d"),":",as.character(Sys.Date(),"%Y/%m/%d"),
      "[DP]"), collapse=""), format = "abstract", batch_size = 1500, dest_file_prefix = "pubmed_enfer_")
                                file.create("corpusEnf.txt")
                                ficheros<-out.A
                                for(i in 1:length(ficheros)){ file.append("corpusEnf.txt",ficheros[i])}
                                pmr<- readabs("corpusEnf.txt")
                                return(pmr)
                                })

  tableDiseases <- reactive({
    pmr<- pubmedEnfer()
    pubEnf<-data.frame(Diseases=character())
    for (i in 1:length(pmr@PMID)){
      pub<-pubtator_disease(pmr@PMID[i][1])
      if (pub!='NULL') {
        pubEnf<-rbind.fill(pubEnf,data.frame(pub[1]))
      }
    }
    pubEnf$Diseases<-tolower(pubEnf$Diseases)
    pubEnf<-pubEnf[!is.na(pubEnf$Diseases),]
    pubEnf<-unique(pubEnf)
    tablaEnfer<-data.frame(pubEnf)
    colnames(tablaEnfer)<-c("Enfermedades")
    return(tablaEnfer)
  })

  output$listadoDisease2<-renderDataTable({
    withProgress(message = 'Obteniendo información enfermedades.',
                 detail = 'Puede tardar unos minutos. Espere...', value = 0, {
                   incProgress(7/15)
                   tableenf<-tableDiseases()
                   incProgress(15/15)
                 })
    datatable(tableenf,
              selection=list(mode='single', selected=1),
              options = list(language = list(url = 'Spanish.json'),
                             pageLength =5
              ))
  })

  observeEvent(input$downPub2,{
    show("btnRelEnfGen")
    hide("downPub2")
  })

  getRelLSA<- eventReactive (c(input$btnRelEnfGen), {
    pmr<- pubmedResult()
    tableEnf<-tableDiseases()
    filaElegida<-input$listadoDisease2_rows_selected
    enfElegida<- as.vector(tableEnf[filaElegida,])
    corpusSecundario<- searchabsL(pmr, include=enfElegida)
    tablaGenes<-as.data.frame(gene_atomization(corpusSecundario))
    if (nrow(tablaGenes)==0){
      tabla<-data.frame(Genes=c("No se han encontrado genes asociados en las publicaciones analizadas"))
    }else{
      genesYenfer<-as.vector(c(as.vector(tablaGenes[,1]),enfElegida))
      tdmCorp<- tdm_for_lsa( corpusSecundario,genesYenfer)
      if (ncol(tdmCorp )<=1){
        tabla<-data.frame(Genes=c("No se han encontrado genes asociados en las publicaciones analizadas"))
      }else{
        lsaCorpus <- lsa( tdmCorp, dims=dimcalc_share())
        matrizGenEnf<-as.textmatrix(lsaCorpus)
        associatedWords <- associate(matrizGenEnf, enfElegida , measure = "cosine", threshold = "0.7")
        if (is.null(associatedWords)){
          tabla<-data.frame(Genes=c("No se han encontrado genes asociados en las publicaciones analizadas"))
        }else{
          tablaRelLSA<- data.frame(peso=associatedWords)
          col1<-rownames(tablaRelLSA)
          col0<-rep(enfElegida,length(col1))
          col2<-tablaRelLSA[,1]
          tabla<- data.frame(Enfermedad=character(),Gene=character(),Peso=double())
          tabla<-data.frame(Enfermedad=col0,Gene=col1,Peso=col2)
          rownames(tabla)<- NULL
          colnames(tabla)<-c("Enfermedad","Gene","Peso")
          tabla[,3]<-round(tabla$Peso,6)
          tablaSinCeros = tabla[tabla$Peso!=0,]
          tablaSinCeros[,3]<-round(tablaSinCeros[,3],6)
          tabla<-tablaSinCeros}}}
    return(tabla)
  })
  output$relLSA <- renderDataTable({
    withProgress(message = 'Obteniendo información genes.',
                 detail = 'Puede tardar unos minutos. Espere...', value = 0, {
                   incProgress(7/15)
```

```
                  tabla<-getRelLSA()
                  incProgress(15/15)
                })

    datatable(tabla,
              selection=list(mode='single', selected=1),
              options = list(language = list(url = 'Spanish.json'), pageLength =20
              ))
  })

  output$tituloGrafRelSim <- renderPrint({
    cat('<h2>para <font style="text-transform: capitalize;">',input$busqueda,'</font></h2>')
  })

  output$plotRelaciones <- renderPlot({
    tablaSinCeros<-getRelLSA()
    tableEnf<-tableDiseases()
    filaElegida<-input$listadoDisease2_rows_selected
    enfElegida<- as.vector(tableEnf[filaElegida,])
    g.la<- graph.data.frame(tablaSinCeros,directed = FALSE)
    layout1<- layout.auto(g.la)
    wc=cluster_walktrap(g.la)
    modularity(wc)
    par(  cex=1.5, bg='#F4F6F6', font=2,  lwd=2, pch=3)
    plot(wc, g.la, layout=layout1, width='1024px', main=enfElegida )
  })

  getRelLSAGraf<- eventReactive (c(input$btnGraf), {
    pmr<- pubmedResult()
    tableEnf<-tableDiseases()
    filaElegida<-input$listadoDisease2_rows_selected
    enfElegida<- as.vector(tableEnf[filaElegida,])
    corpusSecundario<- searchabsL(pmr, include=enfElegida)
    tablaGenes<-as.data.frame(gene_atomization(corpusSecundario))
    if (nrow(tablaGenes)==0){
      tabla<-data.frame(Genes=c("No gráfico"))
    }else{
      genesYenfer<-as.vector(c(as.vector(tablaGenes[,1]),enfElegida))
      tdmCorp<- tdm_for_lsa( corpusSecundario,genesYenfer)
      if (ncol(tdmCorp )<=1){
        tabla<-data.frame(Genes=c("No gráfico"))
      }else{
        file.remove("cossimdata.txt")
        cos_sim_calc(tdmCorp)
        cos=read.delim("cossimdata.txt", header=FALSE, sep="\t")
        cos=na.omit(cos)
        relations <- data.frame(Desde=cos[,1], Hasta=cos[,2],   Peso=abs(cos[,3]))
        relations2=relations[-row(relations)[relations == 0],]
        relations2
      }
    }
  })

  output$tituloGraf<- renderPrint({
    tableEnf<-tableDiseases()
    filaElegida<-input$listadoDisease2_rows_selected
    enfElegida<- as.vector(tableEnf[filaElegida,])
    cat(paste('Relaciones Genes y ',enfElegida))
  })

  output$plotRelaciones <- renderPlot({
    withProgress(message = 'Obteniendo información genes.',
                 detail = 'Puede tardar unos minutos. Espere...', value = 0, {
                   incProgress(7/15)
                   relations2<-getRelLSAGraf()
                   incProgress(15/15)
                 })
    g.1a <- graph.data.frame(relations2, directed=FALSE)
    V(g.1a)$size<-16
    min<-.5 #threshold
    layout1 <- layout.auto(g.1a)
    plot(g.1a, layout=layout1, edge.width=ifelse(E(g.1a)$Peso>=min, E(g.1a)$Peso, NA))
  })

  output$tablaCos <- renderDataTable({
    tabla<-getRelLSAGraf()
    datatable(tabla, caption="Similitud del Coseno",
              selection=list(mode='single', selected=1),
              options = list(language = list(url = 'Spanish.json'),
                             pageLength =20
              ))
  })
}

# Run the application
shinyApp(ui = ui, server = server)
```

funciones.R

```
pubtator_new_function = function (x)
{
  test                                                                                                                                                =
getURL(paste("https://www.ncbi.nlm.nih.gov/research/bionlp/pubtator2/api/v1/publications/export/pubtator?pmids=", x,sep
= ""))
  testa = unlist(strsplit(test, "\n", fixed = T))
  table1 = NULL
  for (i in 3:length(testa)) {
    temps = unlist(strsplit(testa[i], "\t", fixed = T))
    if (length(temps) == 5) {
      temps = c(temps, "No Data")
    }
    table1 = rbind(table1, temps)
  }
  if (ncol(table1) == 6) {
    table2 = table1
    colnames(table2) = c("PMID", "Start", "End", "Term",
                         "TermType", "TermID")
    gene = NULL
    disease = NULL
    mutation = NULL
    chemical = NULL
    species = NULL
    for (i in 1:length(table2[, 5])) {
      if (table2[i, 5] == "Gene")
        gene = c(gene, table2[i, 4])
      else if (table2[i, 5] == "Disease")
        disease = c(disease, table2[i, 4])
      else if (table2[i, 5] == "Mutation")
        mutation = c(mutation, table2[i, 4])
      else if (table2[i, 5] == "Chemical")
        chemical = c(chemical, table2[i, 4])
      else if (table2[i, 5] == "Species")
        species = c(species, table2[i, 4])
    }
    gene = union(gene, gene)
    disease = union(disease, disease)
    mutation = union(mutation, mutation)
    chemical = union(chemical, chemical)
    species = union(species, species)
    return(list(Genes = gene, Diseases = disease, Mutations = mutation,
                Chemicals = chemical, Species = species, PMID = x))
  }
  else return(" No Data ")
}
```

```
pubtator_disease = function (x)
{
  test                                                                                                                                                =
getURL(paste("https://www.ncbi.nlm.nih.gov/research/bionlp/pubtator2/api/v1/publications/export/pubtator?pmids=", x,sep
= ""))
  testa = unlist(strsplit(test, "\n", fixed = T))
  table1 = NULL
  for (i in 3:length(testa)) {
    temps = unlist(strsplit(testa[i], "\t", fixed = T))
    if (length(temps) == 5) {
      temps = c(temps, "No Data")
    }
    table1 = rbind(table1, temps)
  }
  if (ncol(table1) == 6) {
    table2 = table1
    colnames(table2) = c("PMID", "Start", "End", "Term",
                         "TermType", "TermID")
    disease = NULL
    for (i in 1:length(table2[, 5])) {
      if (table2[i, 5] == "Disease")
        disease = c(disease, table2[i, 4])
    }

    disease = union(disease, disease)
    return(list( Diseases = disease))
  }
  else return(" No Data ")
}
```

```
mostrar_info = function (fila, corpus)
{
  s = fila
  pmr<- corpus
  corpusSecund<- subabs(pmr, s,s)
  frasesAbstract<-SentenceToken(corpusSecund@Abstract)
  frasesTitulo<-tokenize_sentences(frasesAbstract[1], lowercase= FALSE)
  cat(paste('<p>','<h4>','<font color=\"#4B04F6\"><b>', corpusSecund@Journal,'</b></font>','</h4></p>'))
  for (i in (1:length(frasesTitulo))){
    if  (i==1  ||  i==2)  cat(paste('<p><h4>','<font  color=\"#4B04F6\"><b>',  frasesTitulo[i],'</b></font>','</h4>',
'\n','</p>'),fill = TRUE)
    else cat(paste('<p><i>',frasesTitulo[i],'</i></p>'), fill = TRUE)
  }
  cat('\n')
  for (i in (2:length(frasesAbstract))){
    cat(paste(' <p>',frasesAbstract[i], '</p>'), fill = TRUE)
  }
  cat(paste("<a   href='https://www.ncbi.nlm.nih.gov/pubmed/",corpusSecund@PMID,"'   target=_blank   >",   "Visualizar
publicación en Pubmed","</a>"))
}
```