



ESTUDIOS / RESEARCH STUDIES

Análisis bibliométrico de la influencia de la Genética en enfermedades raras, a partir de las bases de datos Pubmed y SCI (2000-2009)

Antonio Eleazar Serrano-López *, Carmen Martín-Moreno*, Elias Sanz-Casado*

* Universidad Carlos III de Madrid, Laboratorio de Estudios Métricos de Información, Unidad asociada CSIC. Madrid.
Correos-e: aeserran@bib.uc3m.es, cmartin@bib.uc3m.es, elias@bib.uc3m.es

Recibido: 30-04-2012; 2ª version: 13-09-2012; Aceptado: 30-011-2012.

Cómo citar este artículo/Citation: Serrano-López, A. E.; Martín-Moreno, C.; Sanz-Casado, E. (2013). Análisis bibliométrico de la influencia de la Genética en enfermedades raras, a partir de las bases de datos Pubmed y SCI (2000-2009). *Revista Española de Documentación Científica*, 36(4):e024. doi: <http://dx.doi.org/10.3989/redc.2013.4.970>

Resumen: Este trabajo se centra en el análisis bibliométrico de los artículos de revista publicados sobre dos enfermedades raras que generan trastornos mentales y del comportamiento: CADASIL y Síndrome de Rett, durante el periodo 2000-2009. Además, aunque el Rett afecta fundamentalmente al género femenino y las causas de ambas enfermedades son muy diferentes, tienen en común que en ambos casos se trata de trastornos con un origen genético. Para llevar a cabo el análisis se utilizaron dos bases de datos, Medline y SCI y una metodología de análisis multidimensional que permitiera determinar la relación existente entre la producción en estas enfermedades raras y los avances en el campo de la Genética. Los resultados encontrados muestran una clara diferenciación en los patrones de investigación en ambas enfermedades, aunque confluyen en un factor común, la influencia genética, a pesar de que esta está más acentuada en el Síndrome de Rett. Finalmente se concluye que los logros obtenidos en el campo de la genética, tanto específicos (mutaciones en los genes responsables) y generales (secuenciación del genoma humano), afectan significativamente a la actividad científica alrededor de estas enfermedades raras.

Palabras clave: Bibliometría; síndrome de Rett; CADASIL; genética; enfermedades raras.

Bibliometric analysis of the influence of Genetics on rare diseases from the Pubmed and SCI databases (2000-2009)

Abstract: This work focuses on the bibliometric analysis of journal articles published about two rare diseases that produce mental and behavioural disorders: Rett Syndrome and CADASIL, during the period 2000-2009. Furthermore, although Rett mainly affects females and the causes of both diseases are very different, they have in common that both are genetic disorders. To carry out the analysis we used two databases, Medline and SCI --each with its advantages and biases but complementary to each other-- together with a multidimensional analysis methodology in order to determine the relationship between output about these rare diseases and the advances in the field of Genetics. The results show a clear difference in research patterns for the two diseases, but a convergence on a common factor, genetic influence, although this is more pronounced in the Rett syndrome. Finally it concludes that the great milestones in the field of genetics, both specific (mutations in genes that cause disease) and general (the human genome sequencing), significantly affect the scientific activity surrounding these rare diseases.

Keywords: Bibliometrics; Rett syndrome; CADASIL; genetics; rare diseases.

Copyright: © 2013 CSIC. Este es un artículo de acceso abierto distribuido bajo los términos de la licencia Creative Commons Attribution-Non Commercial (by-nc) Spain 3.0.

INTRODUCCIÓN

Este trabajo surge a partir de la reflexión sobre los resultados de un estudio anterior (Serrano-López y Martín-Moreno, 2011), centrado en analizar descriptivamente la documentación científica existente sobre dos enfermedades raras, el Síndrome de Rett y la Arteriopatía cerebral autosómica dominante con infartos subcorticales y leucoencefalopatía (CADASIL). Ambos producen trastornos mentales y del comportamiento y se consideran enfermedades raras debido a su baja prevalencia puesto que el Síndrome de Rett se da entre 1 y 9 casos por cada 100.000 habitantes y el CADASIL se presenta entre 1 y 9 casos por cada millón (Orphanet, 2011, "International Rett Syndrome Foundation (IRSF)", 2012).

Las causas que generan ambos trastornos son muy diferentes. Mientras el CADASIL tiene causas fisiológicas que afectan a las arterias cerebrales, el Síndrome de Rett es una enfermedad neurológica que afecta fundamentalmente al sexo femenino. Esta enfermedad, descrita por primera vez en 1966 por el científico que le dio nombre (Rett, 1966), se ha relacionado tradicionalmente con trastornos del espectro autista. Por otro lado, tanto el CADASIL como el Síndrome de Rett poseen una causa genética claramente identificada. En el caso del CADASIL el gen *Notch3* en el cromosoma 19 (Joutel et al., 1996) es el responsable de las mutaciones que producen la enfermedad, mientras que en el caso del Síndrome de Rett el principal gen afectado es el *MECP2* (Mehtyl CpG Binding Protein 2), en el cromosoma X (Chahrour et al., 2008) y, aunque otros genes también ejercen cierta influencia sobre el desarrollo de la enfermedad, solo el 10% de los pacientes con Síndrome de Rett presentan mutaciones en genes diferentes del *MECP2* (Muse y Moore, 2012 y Merwick, O'Brien y Delanty, 2012).

Los resultados obtenidos por Serrano-López y Martín-Moreno (2011) ya encontraron evidencias de que el descubrimiento de estos genes (*NOTCH3* y *MECP2*) así como el desarrollo del mapa del genoma humano, han tenido una gran influencia sobre la investigación en estas enfermedades, haciendo que aumente significativamente, por lo que consideramos de gran interés el desarrollo de un estudio bibliométrico que nos permita determinar si estas causas de origen genético se reflejan también en la producción de información científica.

Del mismo modo, otros trabajos realizados sobre diferentes conjuntos de enfermedades raras (Heemstra y Van Weely, 2009; Luengo, 2001), también han encontrado evidencias de que los descubrimientos genéticos relacionados con las diferentes enfermedades raras y los trabajos relacionados con la secuenciación del genoma humano, ejercen cierta influencia sobre el volumen de la investigación en dichas enfermedades.

Los objetivos fundamentales de este trabajo han sido, comprobar si la investigación en genética que se realiza sobre el Síndrome de Rett y CADASIL, ha

influido en la producción científica en ambas enfermedades. Así mismo se ha pretendido determinar los países con mayor producción científica, así como las áreas de investigación de mayor interés. Finalmente otro de los objetivos del trabajo ha sido analizar la presencia de la investigación sobre estas enfermedades en las dos bases de datos consideradas.

METODOLOGÍA

Con el objetivo de maximizar la cobertura del estudio, se decidió consultar dos bases de datos bibliográficas de gran relevancia. Por un lado se utilizó Pubmed/Medline, una base de datos gratuita y de reconocido prestigio, que recoge la mayor parte de la producción científica en medicina (aproximadamente 5400 títulos de revista y 2,4 millones de artículos publicados desde 1948 en el área de medicina y biomedicina) y permite recuperar el mayor número de trabajos posible. A pesar de lo cual, presenta algunas limitaciones que deben tenerse en cuenta. La más importante se debe al hecho de que, en los datos de filiación institucional, solo recoge la dirección del primer autor firmante. Por tanto, esta limitación impide calcular indicadores relativos a la colaboración entre países, instituciones y autores.

Para minimizar estas limitaciones se decidió utilizar una segunda base de datos, el Science Citation Index (SCI) de Thomson Reuters. Esta base de datos, aunque cuenta con importantes sesgos relativos al idioma, concretamente a favor del inglés, debido a que recoge fundamentalmente revistas anglosajonas (Ortiz-Rivera, Suárez-Balseiro y Sanz-Casado, 2002; Van Leeuwen, Moed, Tijssen, Visser y Van Raan, 2001; Fernández, Gómez y Sebastián, 1998; Spinak, 1996); a la nacionalidad de las revistas e incluso a las direcciones de los autores, ofrece los datos de filiación institucional de los autores de forma prácticamente completa, especialmente desde finales de los años 90 (García-Zorita, Martín-Moreno, Lascrain-Sánchez, y Sanz-Casado, 2006). Por este motivo consideramos que se trata de una base de datos apta para el estudio de la colaboración entre países. Es importante tener en cuenta que ambas bases de datos tan solo hemos utilizado los registros tipificados como artículos de revista (limitando los registros a aquellos tipificados como "journal article" en Pubmed y "article" en SCI. Además, la base de datos SCI ha sido utilizada como complemento de Medline y, debido a su volumen, si incluyéramos todos los tipos documentales presentes en el SCI, se desvirtuarían los resultados del análisis.

Con el fin de recuperar el conjunto de documentos que traten sobre las dos enfermedades estudiadas, hemos elaborado la siguiente estrategia de búsqueda, utilizada para interrogar ambas bases de datos.

- CADASIL OR "CEREBRAL ARTERIOPATHY AUTOSOMAL DOMINANT WITH SUBCORTICAL

INFARCTS AND LEUKOENCEPHALOPATHY" OR CASIL OR "HEREDITARY MULTI-INFARCT DEMENTIA"

- "RETT SYNDROME" OR "AUTISM DEMENTIA ATAXIA AND LOSS OF PURPOSEFUL HAND USE"

Esta estrategia se ha ejecutado contra todos los campos de ambas bases de datos, a texto completo, en lugar de utilizar el tesoro MeSH, debido a que en muchos casos las enfermedades raras como estas se encuentran englobadas en términos más amplios que no las representan de forma unívoca, y además, de este modo se asegura la recuperación de todos los documentos que traten sobre ellas, aunque sea de forma tangencial.

En cuanto al ámbito temporal, limitamos el estudio al periodo 2000-2009. El periodo escogido se debe a que en los años 1999 y 2000 se aprobaron los reglamentos de la Unión Europea que establecían el marco regulador en el ámbito de las enfermedades raras y los medicamentos huérfanos (Unión Europea, 1999a, 1999b), por lo que consideramos esta fecha como el punto de partida ideal para el inicio de nuestro trabajo.

De las bases de datos se obtuvo un total de 1.608 registros en Medline y 1.144 en SCI. Una vez obtenidos, se procedió a su exportación a una base de datos en formato SQL con el fin de tratar y normalizar los datos descargados. La normalización se realizó a través de scripts en lenguaje Perl y consultas SQL que nos permitieron asignar a los diferentes registros su nacionalidad, idioma, etc.

Además, con el fin de emplear una terminología temática adecuada, se utilizaron los términos MeSH para etiquetar los registros descargados del SCI, a partir de las keywords asignadas por los autores, de forma que contáramos con un lenguaje normalizado para poder analizar la distribución temática. Sin embargo, a la hora de utilizar estos términos para los análisis multidimensionales, comprobamos que la asignación de términos por parte de Pubmed es insuficientemente descriptiva para el objetivo de este trabajo, por lo que optamos por utilizar, de forma mayoritaria, los términos asignados a los trabajos del SCI. De este modo, la utilización de ambas bases de datos ha permitido conformar una visión bastante completa de la producción científica publicada en este ámbito concreto.

Para el tratamiento estadístico y obtención de los diferentes indicadores se ha utilizado el paquete estadístico R (R Development Core Team, 2008). Se trata de un software estadístico, publicado bajo licencia GPL, y por tanto se trata de software libre, con toda una comunidad de desarrolladores que lo soporta. Este software nos ha permitido realizar análisis estadísticos mediante la conexión directa a la base de datos, en formato SQL, que contiene los datos obtenidos de Medline y SCI.

En cuanto a la elección de los indicadores multidimensionales como el método para determinar la influencia de la investigación sobre genética en la producción científica del Síndrome de Rett y el CADASIL, se debe a que nos permiten poner en

relación variables aparentemente no relacionadas y detectar patrones de comportamiento en la actividad científica. Para ello vamos a utilizar cuatro tipos de indicadores. Los de tipo "mosaico", que nos permiten determinar en que observaciones concretas de cada variable se producen cambios significativos. Análisis de correspondencias, con el fin de examinar el comportamiento de los países productores frente a las enfermedades estudiadas. Análisis de componentes principales, que nos permite estudiar en que medida afectan las temáticas más relacionadas con la genética a la producción científica estudiada.

Por último, realizamos diversos análisis de redes sociales, de forma similar a los estudios realizados en otros trabajos (González Alcaide, Castellano Gómez, Valderrama Zurián, y Aleixandre-Benavent, 2008) pero con un conjunto mucho mayor de datos. Para ello se ha utilizado el algoritmo diseñado por Fruchterman y Reingold (1991), a diferencia de otros estudios realizados con esta metodología (Pinto, Rodríguez Barquín, Moreiro González, y Kauric, 2009) que utilizan el método de Kamada y Kawai (1989). Esto se debe a varios motivos: En primer lugar el volumen de datos es lo suficientemente pequeño como para que puedan ser representados sin algoritmos de poda, como los que implementa el método de Kamada y Kawai, y en segundo lugar, el método de Fruchterman y Reingold, basado no solo en las conexiones entre nodos, si no también en las fuerzas de atracción y repulsión entre ellos, nos permite representar los diferentes aspectos de nuestro análisis de forma más completa. Estos análisis de redes sociales, sin ser indicadores multidimensionales propiamente dichos, aportan un valor añadido a los mismos, fundamentalmente en el análisis de colaboraciones y en las relaciones entre los términos que definen el contenido de los documentos.

Cabe destacar que, tanto para la carga de los datos como para la normalización y posterior análisis de los mismos, se ha utilizado exclusivamente software libre (MySQL, Perl, R) así como los módulos (tanto de tratamiento de datos como estadísticos) desarrollados por la comunidad para este software.

RESULTADOS Y DISCUSIÓN

El total de artículos fuente obtenidos en ambas bases de datos fue de 1.608 en Medline y 1.144 en el SCI, distribuidos tal y como puede observarse en la tabla 1 y figura 1. El solapamiento entre ambas bases de datos se sitúa en torno al 50%, pues de las 478 revistas que hemos analizado en el SCI y las 536 de Pubmed, ambas bases de datos poseen un total de 282 revistas en común y la presencia de los diferentes términos de búsqueda en las dos bases de datos es la que se muestra en la tabla 2. La tendencia en la producción anual es claramente creciente, con algunas rupturas en esta tendencia, concretamente en 2001 y 2005, que también aparecen en Bradsher (2006). El aumento en 2001 tiene como causa probable el hallazgo de los genes que más influyen en la presencia de trastornos del espectro autista (Ingram et al., 2000; Smith et al.,

2000) y de forma particular en el Síndrome de Rett (Amir, et al., 1999; Wan, Mimi et al., 1999).

El segundo aumento significativo, que se produce entre 2004 y 2007 puede estar relacionado con la influencia de los trabajos sobre genética en este tipo de enfermedades. Debemos recordar que la secuenciación del genoma humano, completada en 2003 (Collins, Morgan, y Patrinos, 2003; Frazier, Johnson, Thomassen, Oliver, y Patrinos, 2003; Jasnny, 2003), ha tenido un gran impacto en la literatura médica en general y ha despertado el interés por aquellas enfermedades que tienen una causa genética ya identificada.

Por otro lado, algunos trabajos bibliométricos sobre neuropatologías (Al-Shahi, Will, y Warlow, 2001; Bishop, 2010) han puesto de manifiesto que muchos los investigadores que participaron en los estudios, muestran una mayor preferencia por investigar sobre enfermedades con menor prevalencia que sobre aquellas más comunes.

En cuanto a las diferencias entre ambas bases de datos, se deben fundamentalmente a que la cobertura de Medline (especialmente en revistas biomédicas y revistas locales) es mucho mayor que la del SCI, más aún cuando para este trabajo solo hemos tenido en cuenta los artículos de revista.

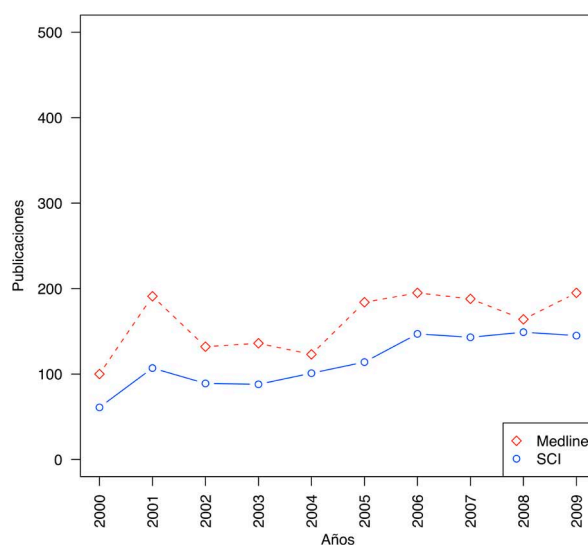
Tabla I. Producción por año y base de datos

Año	Artículos Pubmed	Artículos SCI
2000	100	61
2001	191	107
2002	132	89
2003	136	88
2004	123	101
2005	184	114
2006	195	147
2007	188	143
2008	164	149
2009	195	145

Tabla II. Documentos recuperados por términos de búsqueda

Términos	SCI	Pubmed
CADASIL	650	532
CASIL	1	1
CEREBRAL ARTERIOPATHY AUTOSOMAL DOMINANT WITH SUBCORTICAL INFARCTS AND LEUKOENCEPHALOPATHY	10	203
HEREDITARY MULTI-INFARCT DEMENTIA	1	25
RETT SYNDROME	1170	1106
AUTISM DEMENTIA ATAXIA AND LOSS OF PURPOSEFUL HAND USE	4	1

Figura 1. Evolución de la producción por base de datos



En cuanto a la distribución de trabajos por años en cada base de datos, agregadas por la enfermedad sobre la que tratan, obtenemos la figura 2, y nos indica que esta influencia es mayor en el Síndrome de Rett que en el CADASIL.

Este hecho probablemente se deba a los nuevos conocimientos que se tienen sobre el Síndrome de Rett, ya que tradicionalmente se consideraba como un trastorno del espectro autista, si bien a partir de 1999, cuando se descubrió del gen MECP2 como responsable de su aparición (Amir, et al., 1999; Wan, Mimi et al., 1999). También puede influir en estas diferencias la prevalencia de ambas enfermedades, pues la prevalencia del CADASIL es 10 veces menor que la del Síndrome de Rett.

Además, como señalan algunos autores (Frigotto y Riccaboni, 2011), el Síndrome de Rett es una especialidad con una comunidad de investigadores y una base de conocimientos muy definida y estable, especialmente tras el descubrimiento del gen MECP2, convirtiéndola en una disciplina mucho

más productiva de lo que cabría esperar por el número de personas afectadas.

Este comportamiento también se reproduce en la figura 3, que representa la distribución de documentos por fecha en función del gen sobre el que tratan. En ella se observa una mayor producción científica sobre el gen MECP2 que sobre el NOTCH3 (causante del CADASIL).

La distribución de producción observada en ambos gráficos, nos hace pensar que existe una producción significativamente mayor en el Síndrome de Rett que en el CADASIL, por lo que efectuamos un test de hipótesis sobre la producción de ambas enfermedades tanto en SCI como en Medline, con un nivel de significación del 95%. En la prueba obtuvimos unos p-valores de 0,002 y 0,0005 respectivamente, por lo que podemos rechazar la hipótesis nula de que ambas distribuciones (producción en el Síndrome de Rett y producción en CADASIL) provienen de la misma población y, por tanto, concluir que existe diferencia significativa entre ambas.

Figura 2. Evolución de la producción por enfermedad en SCI y Pubmed

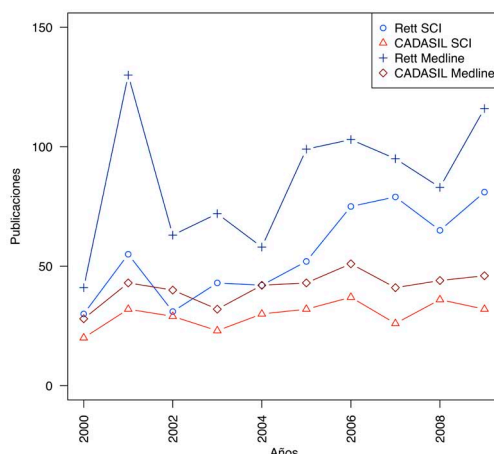
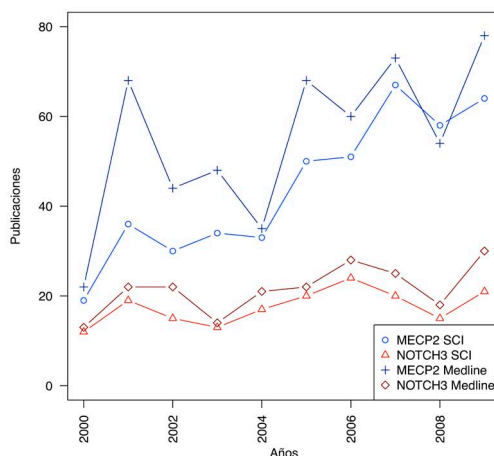


Figura 3. Evolución de la producción por genes en SCI y Medline



Para recoger la filiación institucional de los autores, se han utilizado las bases de datos de Thomson-Reuter, puesto que, al recoger las filiaciones institucionales completas, ofrecen una mayor cobertura para estudios de la producción por países, tal y como queda patente en la figura 4, donde se puede observar la distribución de trabajos por país en ambas bases de datos. Se recogen aquellos países que han publicado al menos 15 trabajos en todo el periodo de estudio y, quedan patentes las limitaciones de Medline, pues países como Bélgica, Dinamarca, Finlandia, Portugal, etc. que probablemente no aparecen como los primeros firmantes, se encuentran mucho mejor representados en el SCI que en Medline. Por otro lado merece ser destacada la presencia de países como China o Chile

(CHN y CHI), que cuentan con una mayor representación en Medline, motivada por la cobertura de esta base de datos sobre las revistas nacionales de dichos países, que sin embargo no recoge el SCI.

En cuanto a los idiomas de los documentos (figura 5), han sido extraídos del campo "language" de los registros bibliográficos (incluyendo los idiomas de las revistas multilingües), se produce el efecto contrario al de los países. Debido a la presencia en Medline de un gran número de revistas de ámbito nacional, los idiomas se encuentran mucho mejor representados en esta base de datos que en el SCI, incluyendo un gran número de idiomas que no tienen representación en el Science Citation Index.

Figura 4. Gráfico de mosaico entre bases de datos y países

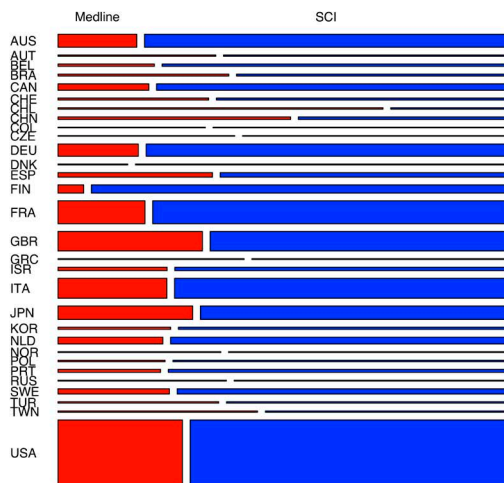
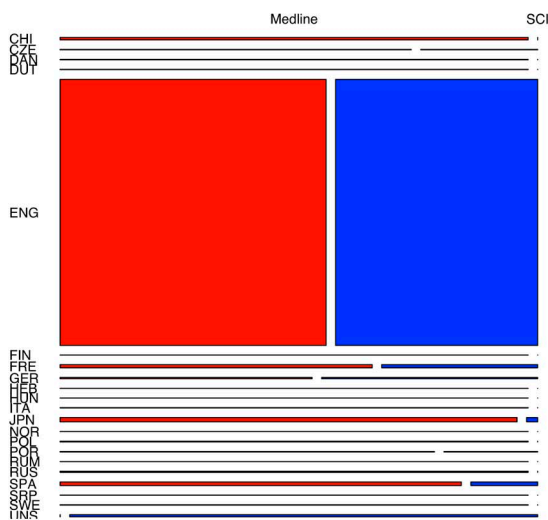


Figura 5. Gráfico de mosaico entre bases de datos e idiomas



Igualmente podemos observar claramente que, incluso con la presencia mayoritaria del inglés en ambas bases de datos, en Medline algunos idiomas sí que tienen cierta representación, como es el caso del francés (con 20 documentos), japonés (41), chino (22) o español (30), mientras que en SCI su presencia, cuando la hay, es meramente anecdótica y, por el contrario, cuenta con un número significativo (21) de trabajos en idiomas no identificados (UNS).

Sobre este grupo de idiomas "no identificados", cabe destacar que se trata de la etiqueta que se les asigna a aquellos idiomas que el sistema de Thomson-Reuters no ha sido capaz de identificar de forma automática, por lo que es probable que se encuentren incluidos en esta categoría gran parte de los idiomas regionales.

El análisis de la distribución de los idiomas por base de datos responde a la intención de mostrar las diferencias entre ambas bases de datos, y porque la información que ambas recogen es complementaria.

El análisis de correspondencias entre países y términos (figura 6) se ha realizado con aquellos términos más significativos en el SCI (aquellos con una frecuencia de aparición mayor o igual a 50). Se ha utilizado el SCI debido a las limitaciones de Medline que hemos mencionado anteriormente.

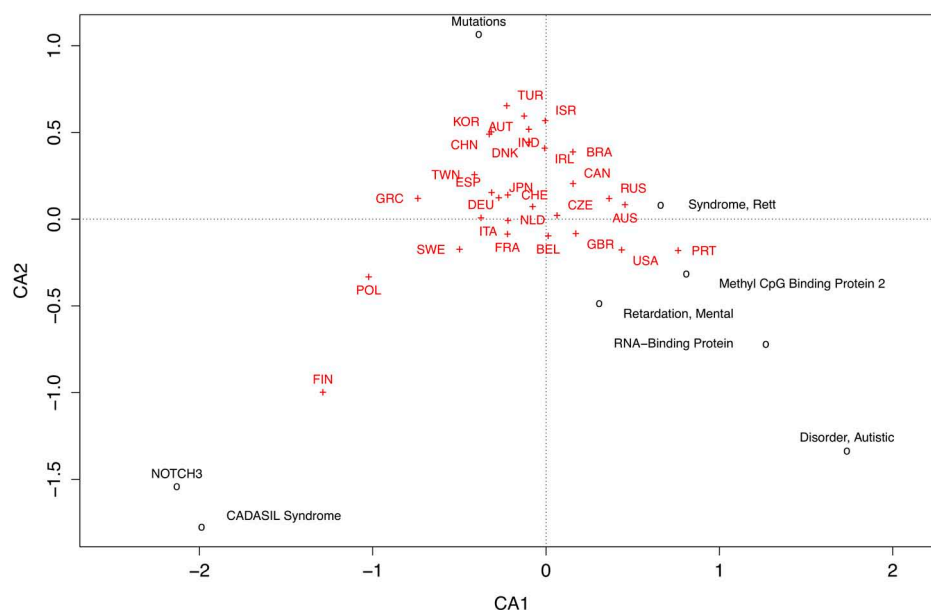
La matriz de doble entrada países-términos MeSH ha sido sometida a la prueba estadística del χ^2 , con el fin de determinar si existe asociación o no entre las variables temática y país. El p-valor

del estadístico ha sido inferior a $2,2e-16$, por lo que concluimos que existe relación de dependencia entre las variables y procedimos a realizar el análisis de correspondencias.

Los resultados que muestra el análisis de correspondencias nos indican que las investigaciones en Síndrome de Rett y en CADASIL son totalmente opuestas entre sí, situándose en dos puntos muy alejados y separadas en ambos ejes. La distribución de países en torno a los términos nos muestra un claro sesgo de tres países (Finlandia, Suecia y Polonia) hacia el CADASIL y por extensión hacia el gen Notch3, mientras que el resto de países se sitúan en torno al eje central, indicando que no se encuentran sesgados hacia ningún término concreto, aunque por volumen de producción se encuentran más cercanos al eje MECP2-Rett. Además, el acumulado de ambas dimensiones recoge una proporción de la variabilidad explicada del 74,81%, por lo que podemos afirmar que gran parte de la distribución de trabajos por país se debe a la temática de los mismos. El caso de los países se debe a que la prevalencia del CADASIL es mayor en la población caucásica y especialmente en Suecia y Finlandia (Kalimo y Ruchoux, 2002), donde además se ha detectado una variante de dicha enfermedad, con síntomas muy similares al CADASIL y causada también por una mutación presente en el gen NOTCH3 (Low y Junna, 2007).

Otro hecho a destacar a partir del análisis de correspondencias es que los trastornos del espectro autista se sitúan, respecto a la primera dimensión, completamente alejados del CADASIL, pero cercanos al Rett y al MECP2, debido, como menciona-

Figura 6. Análisis de correspondencias entre países y términos MeSH (SCI)



mos anteriormente, a la consideración que ha tenido dicha enfermedad como parte de los trastornos del espectro autista.

Siguiendo con la distribución de la temática, se han realizado dos análisis de componentes principales, el primero de los trabajos frente a categorías WOS (figura 7) y el segundo frente a los términos MeSH. (figura 8).

El primero, con un 48,68% de la variabilidad explicada, nos indica que los trabajos publicados en revistas de Genética muestran un perfil muy diferente de los publicados sobre estas enfermedades en el área de Neurología y Neurociencia, respecto del primer componente, mientras que el segundo componente nos indica una estrecha relación entre ellos, distanciándose sensiblemente del resto de categorías WOS, especialmente de la Pediatría, a pesar de que en el primer componente se encon-

trarían cerca de ambas áreas, indicando un carácter interdisciplinar.

Por otro lado, la figura 8, correspondiente al análisis de componentes principales de los términos MeSH, explica un 15,38% de la variabilidad y muestra un claro sesgo hacia la investigación genética sobre ambos genes responsables (MECP2 y Notch3), situándose los trabajos en torno a sus ejes de atracción. Del mismo modo, respecto a la primera componente los trabajos sobre Rett y CADASIL son diametralmente opuestos, a pesar de tener en común términos como "Mutations" o "RNA_Binding Protein", que sugieren una gran carga genética en ambas enfermedades, mientras que si situamos los términos sobre el plano de la segunda componente, observamos que ambas enfermedades se acercan entre si, por lo que podemos concluir que existe una gran influencia de la investigación genética sobre la producción científica en ambas enfermedades.

Figura 7. Análisis de componentes principales sobre categorías WOS

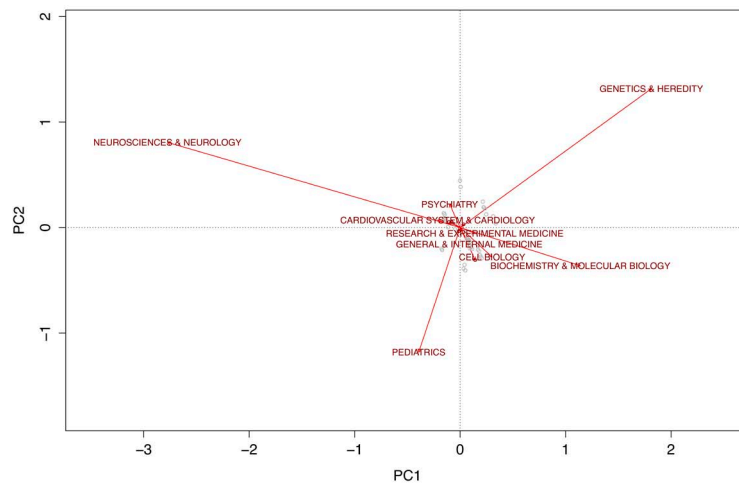
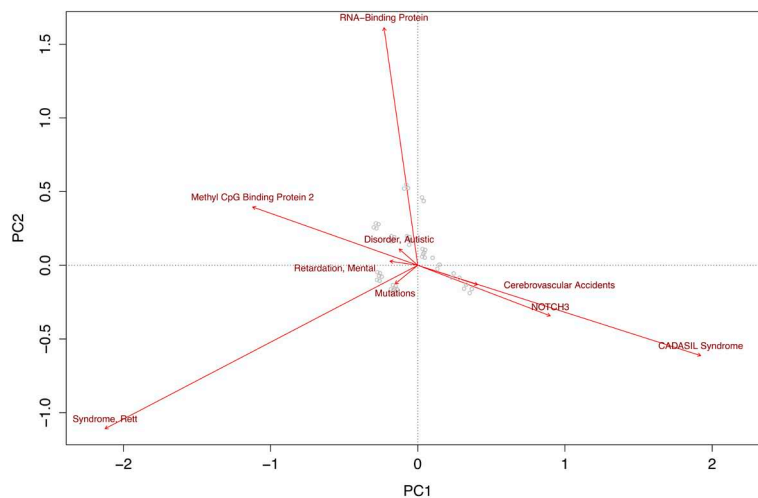


Figura 8. Análisis de componentes principales sobre términos MeSH (SCI)



Por otro lado, puede observarse una relación mucho más estrecha entre el CADASIL y el Notch3 que entre el Síndrome de Rett y el Mecp2, debido a que el Síndrome de Rett no es una enfermedad monogénica ("International Rett Syndrome Foundation (IRSF)", 2012), si no que otros genes también influyen en la aparición de este desorden (Cardoza et al., 2011; Mencarelli et al., 2010; Pecorelli et al., 2011), con lo cual el solapamiento entre los documentos que tratan del gen Mecp2 y el síndrome de Rett es menor que en el caso del CADASIL y el gen Notch3, tal y como indica el análisis de componentes principales de la Figura 8.

En las figuras 9 y 10 encontramos los grafos de asociación entre términos MeSH tanto en SCI como en Medline, realizados a través del algorit-

mo de Fruchterman y Reingold (1991). En el grafo del SCI podemos ver claramente como el Síndrome de Rett se encuentra estrechamente asociado al gen MECP2, a los trastornos del espectro autista y a las proteínas combinatorias del ARN, mientras que el CADASIL y el gen Notch3, que también se encuentran muy relacionados entre si, se alejan significativamente del resto de términos con un mayor volumen total de enlaces con otros términos (degree).

En el caso de Medline, el CADASIL y el Notch3 ni siquiera aparecen entre los términos con un degree más alto, mientras que si lo hacen el Síndrome de Rett y el gen MECP2, estrechamente unidos a otros términos (ADN, y secuenciación), muy relacionados con la Genética.

Figura 9. Análisis de redes sociales de términos MeSH (SCI)

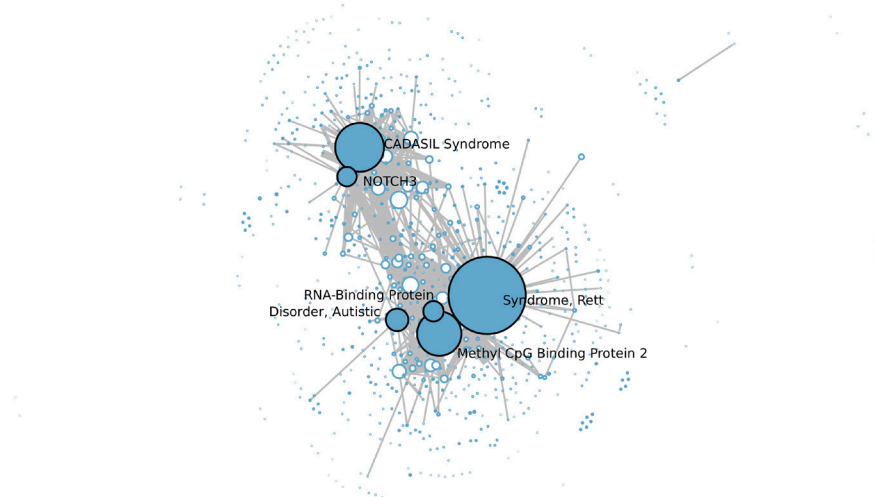
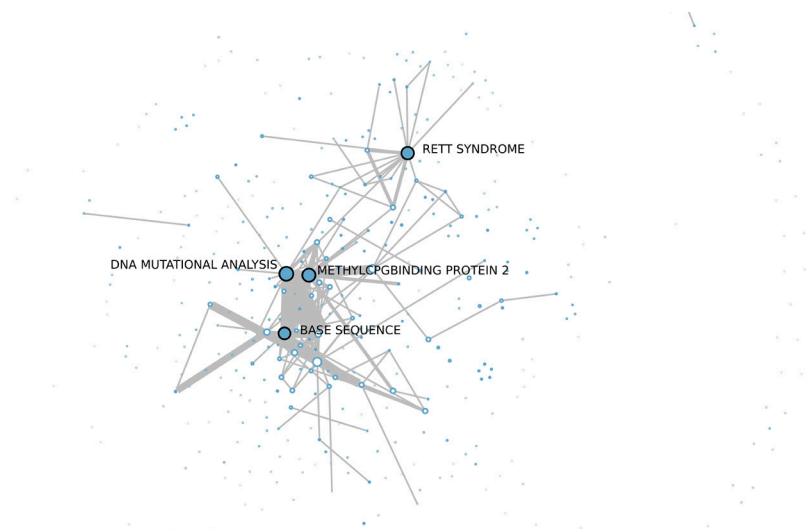


Figura 10. Análisis de redes sociales de términos MeSH (Medline)



Por último, para concluir con el análisis multidimensional de los datos, hemos elaborado una matriz de colaboraciones entre países, con el fin de determinar como colaboran entre si los diferentes países que desarrollan investigación científica sobre las enfermedades estudiadas. El resultado es la figura 11, en la que podemos ver como los países más colaborativos (con nodos de mayor tamaño) son en su mayor parte países occidentales, salvo Japón. Se encuentran muy relacionados entre si, con un alto grado de colaboración entre ellos y forman una red muy tupida, mientras que el resto de países son satélites de dicha red y se encuentra muy diseminados alrededor de ella, indicando que los centros de trabajo más importantes en estas enfermedades se encuentran muy concentrados en los países considerados como grandes productores, mientras en países menos desarrollados apenas se lleva a cabo este tipo de investigación.

En cuanto a la presencia central de Estados Unidos se explica debido a que es el país más productivo y también el que presenta un mayor índice de colaboración, con el resto de países orbitando alrededor de él, especialmente aquellos con los que tiene una colaboración más intensa.

CONCLUSIONES

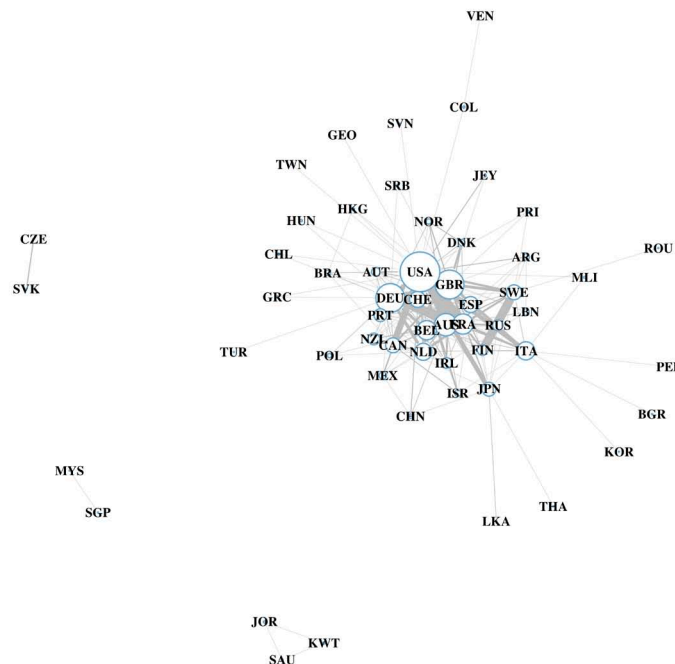
La primera conclusión que podemos obtener es que existe una gran relación entre la investigación

genética y las dos enfermedades estudiadas. Esta relación viene marcada por diferentes hitos en la historia de su investigación, como son el descubrimiento de los genes (MECP2 y NOTCH3) principales causantes de estos trastornos, o los recientes avances en la secuenciación del genoma humano, a pesar de lo cual la relación entre enfermedad y gen es mucho más estrecha en el caso del CADASIL, mientras que el Síndrome de Rett, al tratarse de una enfermedad poligénica, presenta una relación más difusa con el gen MECP2.

En cuanto a la distribución por idiomas y países, las publicaciones en ambas bases de datos siguen los patrones habituales en estas áreas, es decir, se observa una gran preeminencia del inglés, así como de los países occidentales. También se ha observado que existen algunas diferencias entre las bases de datos utilizadas, marcadas fundamentalmente por los sesgos propios de cada una de ellas. Si bien ambas bases de datos presentan información complementaria, con una mejor representación de los países productores en el Science Citation Index, mientras que Medline recoge mejor la producción en revistas de las áreas biomédicas, así como en idiomas locales, apenas presentes en el SCI.

Las diferencias entre las dos enfermedades estudiadas son patentes. Por un lado el volumen de la investigación sobre el Síndrome de Rett es mucho mayor que la que se realiza sobre el CADASIL,

Figura 11. Análisis de redes sociales de colaboración entre países



siendo también mucho menor el volumen de investigación genética realizada sobre esta enfermedad, pero además está el hecho de que la investigación de países como Estados Unidos está mucho más orientada al Síndrome de Rett que al CADASIL, lo que sesga enormemente el volumen total de publicaciones hacia el Síndrome de Rett. Sin embargo, a pesar de que el volumen de investigación genética sobre el CADASIL es menor que en el caso del Rett, el principal gen causante del CADASIL (Notch3) se encuentra presente en un mayor porcentaje de trabajos sobre el CADASIL, que su homólogo (el gen *Mecp2*) en el caso del Síndrome de Rett, debido a la presencia de los genes *CDKL5*, *FOXG1* y *NTNG1/BDNF*, que también influyen en la aparición de esta enfermedad.

Los indicadores multidimensionales muestra que, a pesar de tratarse de dos enfermedades claramente diferenciadas y de tener causas y comportamientos diferentes, existe una marcada relación entre ellas. Esta relación es especialmente visible en los trabajos con mayor volumen de investigación genética, así como por la influencia que los trabajos sobre el mapa del genoma humano han tenido sobre ambas enfermedades, pero de manera especial sobre el Síndrome de Rett.

Por último, cabe destacar el mayor interés en la investigación que presentan determinados países, especialmente Suecia, Polonia y sobre todo Finlandia, hacia la investigación sobre el CADASIL, debido a que la prevalencia de dicha enfermedad en estos países es notablemente mayor que en el resto y cuentan además con una variante propia de la enfermedad, por lo que despierta gran interés en la comunidad científica de dichos países.

BIBLIOGRAFÍA

- Al-Shahi, R.; Will, R. G.; Warlow, C. P. (2001). Amount of research interest in rare and common neurological conditions: bibliometric study. *British Medical Journal*, 323.
- Amir, R. E.; Van Den Veyver, I. B.; Wan, M.; Tran, C. Q.; Francke, U.; Zoghbi, H. Y. (1999). Rett syndrome is caused by mutations in X-linked *MECP2*, encoding methyl-CpG-binding protein 2. *Nature Genetics*, 23(2), 185-188.
- Bishop, D. V. M. (2010). Which Neurodevelopmental Disorders Get Researched and Why? *PLoS One*, 5(11), e15112. Public Library of Science.
- Bradsher, E. (2006). An Analysis of the Publication Pattern of Pervasive Developmental Disorders, 2000-2005.
- Chahrour, M.; Jung, S. Y.; Shaw, C.; Zhou, X.; Wong, S. T. C.; Qin, J.; Zoghbi, H. Y. (2008). *MeCP2*, a key contributor to neurological disease, activates and represses transcription. *Science*, 320(5880), 1224. American Association for the Advancement of Science.
- Collins, F. S.; Morgan, M.; Patrinos, A. (2003). The Human Genome Project: lessons from large-scale biology. *Science*, 300(5617), 286. American Association for the Advancement of Science.
- Fernández, M. T.; Gómez, I.; Sebastián, J. (1998) La cooperación científica de los países de América Latina a través de indicadores bibliométricos. *Inter-ciencia*, vol. 23(6), 328-337.
- Frazier, M. E.; Johnson, G. M.; Thomassen, D. G.; Oliver, C. E.; Patrinos, A. (2003). Realizing the potential of the genome revolution: the genomes to life program. *Science*, 300(5617), 290. American Association for the Advancement of Science.
- Frigotto, M. L.; Riccaboni, M. (2011). A few special cases: scientific creativity and network dynamics in the field of rare diseases. *Scientometrics*, 89(1), 397-420. Akadémiai Kiadó, co-published with Springer Science+ Business Media BV, Formerly Kluwer Academic Publishers BV.
- Fruchterman, T. M. J.; Reingold, E. M. (1991). Graph drawing by force-directed placement. *Software: Practice and Experience*, 21(11), 1129-1164. Cite-seer. doi:10.1002/spe.4380211102
- García-Zorita, C.; Martín-Moreno, C.; Lascurain-Sánchez, M. L.; Sanz-Casado, E. (2006). Institutional addresses in the Web of Science: the effects on scientific evaluation. *Journal of Information Science*.
- González Alcaide, G.; Castellano Gómez, M.; Valderrama Zurián, J. C.; Aleixandre-Benavent, R. (2008). Literatura científica de autores españoles sobre análisis de citas y factor de impacto en Biomedicina (1981-2005). *Revista española de Documentación Científica*, 31(3). Consejo Superior de Investigaciones Científicas (CSIC). doi:10.3989/redc.2008.v31.i3.433
- Heemstra, H.; Weely, S. van. (2009). Translation of rare disease research into orphan drug development: disease matters. *Drug discovery Today*, 14, (23-24), 2009.
- Ingram, J. L.; Stodgell, C. J.; Hyman, S. L.; Figlewicz, D. A.; Weitkamp, L. R.; Rodier, P. M. (2000). Discovery of allelic variants of *HOXA1* and *HOXB1*: genetic susceptibility to autism spectrum disorders. *Teratology*, 62(6), 393-405.
- International Rett Syndrome Foundation (IRSF) (2012). Recuperado de <http://www.rettssyndrome.org/>
- Jasny, B. R. (2003). Building on the DNA Revolution. *Science*, 300(5617), 277-277. Science. doi:10.1126/science.300.5617.277
- Joutel, A.; Corpechot, C.; Ducros, A.; Vahedi, K.; Chabriat, H.; Mouton, P.; Alamowitch, S. (1996). Notch3 mutations in CADASIL, a hereditary adult-onset condition causing stroke and dementia. *Nature*, 383(6602), 707-710. Nature Publishing Group.
- Kalimo, H.; Ruchoux, M. (2002). CADASIL: a common form of hereditary arteriopathy causing brain infarcts and dementia. *Brain Pathol.*, 12(3), 371-384.
- Kamada, T.; Kawai, S. (1989). An algorithm for drawing general undirected graphs. *Information processing letters*, (31), 7-15.
- Low, W.; Junna, M. (2007). Hereditary multi-infarct dementia of the Swedish type is a novel disorder different from NOTCH3 causing CADASIL. *Brain*, 130(2), 357-367.

- Luengo, S.; Aranda, M. T.; la Fuente, M. (2001). *Enfermedades Raras: Situación y demandas sociosanitarias*. Madrid: IMSERSO.
- Merwick, A.; O'Brien, M.; Delanty, N. (2012). Complex single gene disorders and epilepsy. *Epilepsia*, 53, 81-91, 2012.
- Muse, M.; Moore, B. (Eds.) (2012). *Handbook of Clinical Psychopharmacology for Psychologists*. John Wiley & Sons, 2012.
- Orphanet. (2011). *Orphanet*. Recuperado de <http://www.orpha.net/consor/cgi-bin/index.php>
- Ortiz-Rivera, L. A.; Suárez-Balseiro, C.; Sanz-Casado, E. (2002). Enfoque bibliométrico de la producción científica en Ciencias de la Salud en Puerto Rico durante el período de 1990 a 1998. *Revista Española de Documentación Científica*, vol. 25 (1), 9-28.
- Pinto, A. L.; Rodríguez Barquín, B. A.; Moreiro González, J. A.; Kauric, A. (2009). Análisis de las redes sociales en las publicaciones seriadas: su representación en el Journal of Documentation. *Investigación bibliotecológica*, 23(48), 13-32.
- R Development Core Team. (2008). R: A Language and Environment for Statistical Computing. (R. D. C. Team, Ed.), *Vienna Austria R Foundation for Statistical Computing*. R Foundation for Statistical Computing. Recuperado de <http://www.r-project.org>
- Rett, A. (1966). On a unusual brain atrophy syndrome in hyperammonemia in childhood. *Wiener medizinische Wochenschrift (1946)*, 116(37), 723.
- Serrano-López, A. E.; Martín-Moreno, C. (2011). Cadasil e Síndrome de Rett: Estudo de caso de dois doenças raras neurológicas. *Ponto de Acesso*, 5(3), 130-148. Recuperado de <http://www.portalseer.ufba.br/index.php/revistaici/article/viewArticle/5506>
- Smith, M.; Filipek, P. A.; Wu, C.; Bocian, M.; Hakim, S.; Modahl, C.; Spence, M. A. (2000). Analysis of a 1-megabase deletion in 15q22-q23 in an autistic patient: identification of candidate genes for autism and of homologous DNA segments in 15q22-q23 and 15q11-q13. *American Journal of Medical Genetics*.
- Spinak, E. (1996). Los análisis cuantitativos de la literatura científica y su validez para juzgar la producción latinoamericana. *Boletín de la Oficina Sanitaria Panamericana*, vol. 120(2), 139-145
- Unión Europea. (1999a). Decisión nº 1295/1999/CE del Parlamento Europeo y del Consejo, de 29 de abril de 1999, por la que se aprueba un programa de acción comunitaria sobre las enfermedades poco comunes en el marco de la acción en el ámbito de la salud pública (1999-2003).
- Unión Europea. (1999b). Reglamento (CE) nº 141/2000 del Parlamento Europeo y del Consejo, de 16 de diciembre de 1999, sobre los medicamentos huérfanos.
- Van Leeuwen, T. N.; Moed, H. F.; Tijssen, R. J. W.; Visser, M. S.; Van Raan, A. F. J. (2001). Language biases in the coverage of the Science Citation Index and its consequences for international comparisons of national research performance. *Scientometrics*, vol. 51(1), 335-346.
- Wan, M.; Lee, S. S. J.; Zhang, X.; Houwink-Manville, I.; Song, H. R.; Amir, R. E.; Budden, S. (1999). Rett syndrome and beyond: recurrent spontaneous and familial MECP2 mutations at CpG hotspots. *The American Journal of Human Genetics*, 65(6), 1520-1529.