

CARACTERIZACIÓN DEL ESTADO DE LA CITOGENÉTICA HUMANA MEDIANTE UN ESTUDIO BIBLIOMÉTRICO UTILIZANDO PUBMED.

CHARACTERIZATION OF THE STATE OF ART IN HUMAN CYTOGENETICS BY MEANS OF A BIBLIOMETRIC STUDY UTILIZING PUBMED

Autores: Michel Soriano, Ricardo Arencibia, Enny Morales, Arlay Castelví

Institución: Centro Nacional de Genética Médica (CNGM), Centro Nacional de Investigaciones Científicas (CNIC).

Teléfono: 2089991 extensiones: 1059, 1061

E-mail: michel.soriano@cngen.sld.cu

País: Cuba

Introducción

Apoyada en los trabajos de Lotka, Bradford y Zipf, la bibliometría surgió durante la primera mitad del siglo XX con el objetivo de medir la calidad de los trabajos recogidos en las colecciones de literatura científica (Pritchard, 1969). Esta disciplina permitió medir la producción científica de un investigador, una institución o un país dentro de un campo o especialidad. Actualmente constituye un elemento de apoyo estadístico que permite el mapeo y generación de diferentes indicadores para el manejo de la información y el conocimiento (Ugolini et al., 2007; Poletto y Medeiros, 2009). Los indicadores bibliométricos basados en publicaciones han resultado ser instrumentos útiles para estudiar los resultados de la actividad científica, lo cual resulta una necesidad ineludible en la actualidad ya que permite optimizar los recursos (siempre limitados) que se destinan a la investigación (Bordons y Zulueta, 1999).

A partir de una colección de referencias bibliográficas relacionadas con la Citogenética, ciencia dedicada al estudio de los aspectos celulares de la herencia y especialmente a la descripción de la estructura cromosómica y la identificación de las aberraciones genómicas que causan enfermedades, se puede mapear el desarrollo de esta disciplina científica. El análisis cromosómico convencional basado en el análisis de los patrones de bandas fue desarrollado en 1970 y es ampliamente utilizado en la actualidad. El descubrimiento en 1956 del número correcto de cromosomas en los humanos inició una nueva era para la Citogenética humana. El primer diagnóstico exitoso de una aberración cromosómica ocurrió en 1959 según fue reportado por Lejeune con la detección de una trisomía 21 en pacientes síndrome Down, provocando la emergencia de la genética médica como una disciplina clínica (Speicher y Carter, 2005, Ferguson-Smith, 2008).

La introducción de la técnica de reacción en cadena de la polimerasa permitió el desarrollo de la Citogenética molecular y su empleo extensivo en la producción y marcaje de sondas de pintado cromosómico que en 1992 fue utilizado para el análisis de cariotipos. Esta tecnología es usada con mayor frecuencia para estudiar la genética y la biología de las células normales (Ferguson-Smith, 2008; Spicer y Carter, 2005).

Objetivos

- Determinar el núcleo de revistas especializadas en Citogenética humana.
- Identificar los autores más productivos.
- Identificar las instituciones líderes a nivel mundial.
- Determinar las principales líneas investigativas.

Materiales y Métodos

a) Colección de los datos originales

La base de datos empleada fue Pubmed (www.ncbi.nlm.nih.gov/pubmed/), escogida como fuente de información por especializarse en literatura biomédica y por encontrarse disponible gratuitamente en Internet. Es publicada por el National Library of Medicine de los Estados Unidos de América (EUA) y es actualizada semanalmente. Contiene más de 20 millones de registros bibliográficos actualizados desde 1949 (consultado el 10 de diciembre de 2010).

Se creó una librería para la colección de datos utilizando el programa EndNote 10.0 (www.endnote.com). Para la estrategia de búsqueda se recopilaron todos los registros bibliográficos que incluyeran en el título o en el resumen los términos: *cytogenetic*, *chromosome*, *chromosomal*, *distal*, *centromere* y “*cell cycle*”.

Se seleccionaron solo los artículos que fueron publicados en el periodo 1999-2008 y los datos recuperados fueron revisados manualmente por segmentos eliminando aquellas referencias no relacionadas con la Citogenética humana, entre estos se encontraban artículos asociados al estudio de las plantas, hongos, insectos, anfibios y aquellos publicados en revistas que se dedican fundamentalmente a la microbiología.

b) Selección de los elementos para análisis

Se emplearon en el análisis los datos recogidos de los campos: título, resumen, descriptores, revista científica, idioma y autor. Otros datos utilizados en el análisis fueron los países y las instituciones de los investigadores, los cuales fueron determinados manualmente. En el campo autor (por no hallarse normalizado) se decidió tener en cuenta para el análisis a los primeros 100 investigadores más productivos, los países e instituciones fueron hallados para estos de manera manual.

c) Procesamiento de los datos

Una vez seleccionados los elementos que serían analizados, la librería fue exportada en el formato *Show all*, escogido por ser el único que puede ser importado desde el programa Bibexcel (<http://www8.umu.se/inforsk/Bibexcel/>), una poderosa herramienta para el análisis de coocurrencia de palabras.

A continuación se realizó el cálculo de las frecuencias de coocurrencia de términos y de autores (número de veces que dos elementos ocurren a la vez) con el objetivo de elaborar las matrices de coocurrencia, las cuales constituyen tablas de contingencia que reflejan la frecuencia de aparición conjunta de dos indicadores que pueden ser o no de la misma naturaleza. Empleando el programa Microsoft Excel las matrices fueron construidas utilizando los términos presentes en el título, el resumen, en el campo descriptores y en el campo autor. Seguidamente se calculó la similitud entre los

elementos en base a las frecuencias de coocurrencia mediante el método de correlación de Pearson por ser el más empleado en estudios de esta naturaleza (McCain, 1990; White y McCain, 1998) y por ser más sencillo que el método que emplea la medida coseno, el índice de inclusión o el índice de Jaccard (Peters y van Raan, 1993).

Los programas Ucinet 6.0 y NetDraw 1.0 (<http://www.analytictech.com/ucinet/trial.htm>) fueron empleados para la visualización de las matrices de coocurrencia, seleccionando los algoritmos Jaccard y CONCOR (CONcordancia de CORrelaciones iterativas) para observar las agrupaciones de términos y de autores. El primero es usado frecuentemente para expresar el grado de intersección entre dos conjuntos de documentos y tiene como propósito brindar información sobre la "intensidad" de la cooperación (coocurrencia). Su valor se calcula a través de la siguiente fórmula: $J_{ij} = (C_{ij} / (C_i + C_j - C_{ij}))$. Este índice puede usarse para medir el grado relativo de solapamiento entre "áreas semánticas" de palabras dentro de una base de datos (Courtial, 1986). El segundo algoritmo empleado correlaciona cada par de actores, extrayendo cada columna de una matriz de correlación actor/actor y correlacionándola con la otra columna. Este proceso es repetido una y otra vez hasta que los elementos en esta "matriz de correlación iterativa" alcancen como valor +1 o -1. Luego, los datos son separados en dos grupos sobre la base de estas correlaciones y a continuación, el proceso es repetido dentro de cada grupo hasta que todos los actores sean separados. El método ha sido utilizado en el análisis bibliométrico de patentes y proyectos de investigación, lo que aproxima su aplicabilidad al contexto de la Ciencia de la Información. Recientemente, su empleo fue validado para la identificación de paradigmas en la estructura intelectual de la bibliometría (Vega Almeida, 2010).

d) Análisis de los elementos seleccionados

Se determinó la productividad anual dentro de la disciplina cuantificando el número de artículos publicados por año para conocer los años de mayor productividad dentro del periodo y si el campo de investigación se encuentra activo. A continuación, las revistas científicas fueron ordenadas en orden decreciente según la cantidad de artículos publicados. Se calculó el porcentaje de las publicaciones por cada revista con respecto al total y el porcentaje acumulativo de los mismos. A partir de este último se conformaron cuatro bloques con vistas a identificar el núcleo Bradford de las revistas más productivas. Para la selección de los 100 autores más productivos se decidió desestimar aquellos de origen asiático ya que fueron detectados un gran número de estos investigadores que trabajaban en distintas instituciones e incluso, en distintos países y compartían el mismo nombre, se comprobó que este fenómeno ocurría de igual manera para nombres de origen chino, japonés y coreano. De este modo se evitó la inclusión de un sesgo que podría alterar los resultados del presente estudio aunque debe tenerse en cuenta que de este modo se desestiman de igual manera las instituciones y los países asiáticos para ser identificados como altamente productivos. A continuación, se listaron todos los autores en orden decreciente según número de artículos publicados y se calculó el porcentaje y el porcentaje acumulativo de los artículos publicados por cada autor.

Se calcularon los índices de centralidad (liderazgo dentro del grupo de colaboración) e intermediación (investigadores que colaboran con otros fuera de su propio grupo de colaboración) de los autores en las redes de coautoría y se ordenaron en tablas separadas de manera decreciente, seleccionando los autores que ocuparon las primeras 10 posiciones según cada índice. Las principales relaciones de coautoría

fueron representadas gráficamente, utilizando como valor de poda la coocurrencia en 15 o más artículos con el objetivo de que quedaran definidos grupos aislados e identificables.

La filiación de los autores más productivos de la colección fue hallada manualmente, posteriormente se sustituyó el nombre de los investigadores en la red de coautoría por el nombre de las instituciones a las cuales pertenecían y se pudo observar el tipo de colaboración dada entre estos (intrainstitucional o interinstitucional). Este dato es de mucha importancia para saber el tipo de relaciones que se establecen entre los autores dentro de la comunidad científica especializada en el tema. Las instituciones analizadas fueron clasificadas según cuatro categorías: Universidad, Instituto, Hospital y Academia de Medicina. Una vez clasificadas, se halló el porcentaje de contribución de cada una dentro de la muestra de autores más productivos.

Se realizó un análisis de las principales temáticas abordadas por las revistas científicas de mayor especialización en temas de Citogenética humana, las líneas investigativas en las que se desempeñan los autores más productivos y las temáticas sugeridas por las agrupaciones obtenidas a partir de las asociaciones de palabras en los campos título, resumen y descriptores. Se decidió comparar los resultados del presente estudio con las expectativas de los expertos dentro de la especialidad bajo el supuesto de que el cáncer y el uso de las técnicas de Citogenética molecular debían estar presentes. Se prestó especial atención a la posible aparición de nuevas líneas de investigación no identificadas por nuestro laboratorio con anterioridad.

Resultados

Se recuperaron un total de 10 388 artículos publicados indizados en Pubmed referentes a la rama de la genética humana analizada. La producción científica presentó un crecimiento sostenido durante todo el período, lo cual es indicativo del desarrollo experimentado por la especialidad. El crecimiento fue casi exponencial entre los años 1999 y 2004, con un promedio anual sobre los 1 200 artículos durante los últimos cinco años analizados.

Las revistas científicas, al ser divididas en bloques, presentaron una distribución donde cerca del 20 % de las mismas fueron responsables del 75 % de los artículos, mientras que el 25 % restante fue publicado en poco más del 80 % del total de publicaciones seriadas, en una clara adecuación al principio de Pareto. De igual modo, un núcleo de 17 revistas (1,3 %) conformaron el primer bloque de publicaciones seriadas más productivas, responsables del 25 % del total de artículos, similar a la tradicional distribución de Bradford descrita en 1934. En el primer bloque se ubicaron las revistas que publicaron un mayor número de artículos asociados a la especialidad (Tabla I).

Tabla I. Revistas pertenecientes al primer bloque ordenado de manera decreciente según la cantidad de artículos referentes a la Citogenética Humana.

Revista	País	Artículos	%	Art/Cum	% Cum
Cancer Genet Cytogenet	EUA	326	3,14	326	3,14
J Biol Chem	EUA	240	2,31	566	5,45
Am J Med Genet A	EUA	202	1,94	768	7,39
Oncogene	Inglaterra	163	1,57	931	8,96

Cancer Res	EUA	159	1,53	1090	10,49
Cytogenet Genome Res	Suiza	152	1,46	1242	11,96
Cell Cycle	EUA	152	1,46	1394	13,42
Proc Natl Acad Sci U S A	EUA	151	1,45	1545	14,87
Hum Reprod	Inglaterra	150	1,44	1695	16,32
Blood	EUA	146	1,41	1841	17,72
J Cell Biol	EUA	131	1,26	1972	18,98
Genes Chromosomes Cancer	EUA	125	1,20	2097	20,19
Mol Biol Cell	EUA	121	1,16	2218	21,35
Am J Med Genet	EUA	118	1,14	2336	22,49
J Cell Sci	Inglaterra	115	1,11	2451	23,59
Leukemia	Inglaterra	114	1,10	2565	24,69
Curr Biol	Inglaterra	111	1,07	2676	25,76

Se obtuvo una lista de los 100 autores más productivos dentro de la especialidad durante el periodo, estos pertenecen principalmente a Europa destacándose Alemania entre los 10 más productivos (Tabla II). El análisis de la red de coautoría, a partir de la representación visual de aquellos autores que colaboraron en más de 15 trabajos de investigación, permitió la identificación de 14 componentes fundamentales (Figura 4). La mayoría de los componentes (10, representando el 71,4 %) están formados por parejas. El componente principal lo integran 5 autores de la *Ludwig Maximilians University* estrechamente relacionados, que constituyen un grupo de investigación muy productivo liderado por Schoch, Hiddemann y Haferlach.

Tabla II. Se presentan los 10 autores más productivos, los cuales publicaron más de 35 artículos.

Autores	Institución	País	Artículos	%
Liehr, T.	Institute of Human Genetics and Anthropology, Jena	Alemania	66	0,64
Rocchi, M.	Universidad de Bari	Italia	46	0,44
Haferlach, T.	Ludwig Maximilians University	Alemania	45	0,43
Earnshaw, W. C.	University of Edinburgh	UK	45	0,43
Schoch, C.	Ludwig Maximilians University	Alemania	44	0,42
Starke, H.	Institute of Human Genetics and Anthropology, Jena	Alemania	40	0,39
Shaffer, L. G.	Signature Genomic Laboratories	EUA	39	0,38
Choo, K. H.	Murdoch Children's Research Institute	Australia	38	0,37
Egozcue, J.	Universitat Autònoma de Barcelona	España	38	0,37
Hiddemann, W.	Ludwig Maximilians University	Alemania	37	0,36

Otro componente compuesto por 5 autores, aunque con menor robustez estructural, está integrado por autores del *Institute of Human Genetics and Anthropology* y es liderado por Liehr y Starke. Un tercer componente intensamente relacionado está compuesto por los franceses Morel, Donet-Guilbert, Le Bris y De Braekeleer. Es importante señalar que Earnshaw, Shaffer y Egozcue no se encuentran ubicados en

ninguno de los componentes con más de 15 relaciones de colaboración por lo que la producción científica de estos autores es realizada mayormente sin colaboración.

Se obtuvo el índice de centralización para los autores más productivos y se determinaron aquellos autores que resultan líderes dentro de los grupos de colaboración que se establecen por participar en la publicación de un mayor número de artículos publicados. Estos autores suelen ser los de mayor reconocimiento dentro del grupo (Tabla III).

Tabla III. Primeros 10 autores con mayor índice de centralidad en sus relaciones. Nota: Network Centralization = 2.23%

Autores	Grado	Normalizado
Liehr, T.	157.000	2.683
Haferlach, T.	147.000	2.512
Schoch, C.	147.000	2.512
Kern, W.	131.000	2.239
Schnittger, S.	127.000	2.170
Starke, H.	124.000	2.119
Hiddemann, W.	117.000	1.999
Oliver-Bonet, M.	88.000	1.504
Morel, F.	86.000	1.470
De Braekeleer, M.	84.000	1.435
Media	28.361	0.485

La importancia de un nodo como mediador entre los otros fue medida con el indicador de intermediación (Freeman et al., 1991) revelando la existencia de autores proclives a colaborar con otros que no pertenecen al grupo de sus relaciones más estrechas (Tabla IV). Esto resulta de gran importancia en la identificación de investigadores altamente productivos con los cuales colaborar en la producción de publicaciones. Los investigadores Rochi, (Universidad de Bari), Earnshaw (Universidad de Edimburgo), Liehr (Institute of Human Genetics and Anthropology) y Schoch (Ludwig Maximilians University) se destacaron por su capacidad para intermediar relaciones de colaboración, lo que los convierte en personalidades de obligada referencia a la hora de definir los principales autores dedicados a la Citogenética humana.

Tabla IV. Primeros 10 autores con mayor índice de intermediación en sus relaciones. Nota: Network Centralization Index = 7.14%

Autores	Intermediación	Normalizada
Rocchi, M.	999.587	8.485
Earnshaw, W. C.	913.187	7.751
Liehr, T.	895.260	7.599
Schwartz, S.	847.758	7.196
Stankiewicz, P.	800.468	6.795

Harrison, C. J.	743.878	6.314
Haas, O. A.	732.723	6.220
Dewald, G. W.	709.388	6.021
Hagemeijer, A.	463.986	3.938
Schoch, C.	438.781	3.724
Media	164.284	1.394

Después de tomar en cuenta el análisis de las asociaciones entre palabras en los campos título y resumen, las temáticas de los artículos que publican las revistas de mayor especialización dentro de la colección y las líneas de trabajo en las que laboran los investigadores de mayor productividad, se determinó que las principales líneas investigativas abordadas en el periodo estudiado y relacionadas con la disciplina son: los procesos malignos (su diagnóstico y pronóstico), los elementos que influyen en el correcto progreso del ciclo celular y los síndromes cromosómicos que conllevan retraso mental.

Conclusiones

- Se identificó una selección de 17 revistas altamente productivas que recogen, cada una, más de 100 artículos relacionados con la temática y que centran su interés en la presentación de casos clínicos, los estudios relacionados con el cáncer y los elementos que interfieren con el progreso del ciclo celular. Coincidiendo con la expectativa de los expertos, un gran número de las revistas pertenecientes al primer bloque se dedican a la publicación de artículos relacionados con los procesos malignos.
- La identificación de los primeros 100 autores más productivos identificó investigadores que constituyen frentes de desarrollo en temáticas de mucho interés para la comunidad científica como son: los estudios de los marcadores supernumerarios y la evolución a nivel cromosómico. La inclusión de un gran número de los autores más productivos entre los de mayor grado de centralidad e intermediación los convierten en autores de referencia obligada y proclives a participar en la publicación de trabajos conjuntos.
- Las instituciones identificadas como líderes se encuentran en Europa, principalmente en Alemania y Francia.
- Las principales líneas de investigación estuvieron en consonancia con las expectativas creadas desde un inicio donde destaca el cáncer como elemento de gran interés por las potencialidades de las técnicas dentro de la especialidad por aportar valiosos datos al diagnóstico, pronóstico, selección del tratamiento y evaluación posterior de los pacientes con enfermedades neoplásicas.

Referencias bibliográficas

- Bordons, A. y M. A. Zulueta (1999): Evaluación de la actividad científica a través de indicadores bibliométricos. Revista española de cardiología. 52(10):790-800.
- Courtial, J. P. (1986): Technical issues and developments in methodology. En M. Callon, J. Law, & A. Rip (eds.), Mapping the dynamics of science and technology: Sociology of science in the real world. pp. 189-210. London: The Macmillan Press Ltd.

- Ferguson-Smith, M. A. (2008): Cytogenetics and the evolution of medical genetics. *Genet Med.* 10(8): 553-9.
- Freeman, L. C., S. P. Borgatti y D. R. White (1991): Centrality in Valued Graphs: A Measure of Betweenness Based on Network Flow. *Social Networks.* 13(1): 141-154.
- McCain, K. W. (1990): Mapping authors in intellectual space: a technical overview, *J. Am. Soc. Inf. Sci.* 41(6): 433-443.
- Peters, H. P. F. y A. F. J. van Raan (1993): Co-word-based science maps of chemical engineering. Part I: representations by direct multidimensional scaling, *Research Policy.* 22(5-6): 23-45.
- Poletto, V. C. e I. Medeiros (2010): Bibliometric study of articles published in a Brazilian journal of pediatric dentistry. *Braz Oral Res.* 24(1):83-8.
- Pritchard, A. (1969): Statistical Bibliography or Bibliometrics? *Journal of Documentation.* 25(4): 348-349.
- Speicher, M. y N. Carter (2005): The new cytogenetics: blurring the boundaries with molecular biology. *Nature Rev Genet.* 6(10): 782-792.
- Ugolini, D., R. Puntoni, F. P. Perera, P. A. Schulte y S. Bonassi (2007): A bibliometric analysis of scientific production in cancer molecular epidemiology. *Carcinogenesis.* 28(8): 1774-9.
- Vega Almeida, R. L. (2010): Ciencia de la Información y Paradigma Social: Enfoques Histórico, Epistemológico y Bibliométrico para un Análisis de Dominio [Tesis doctoral] España. Universidad de la Habana.
- White, H. D. y K. W. McCain (1998): Visualizing a discipline: an author co-citation analysis of information science, 1972-1995, *J. Am. Soc. Inf. Sci. Tec.* 49(4): 327-355.