

# Endo-Mining: herramienta web para la búsqueda automatizada de genes potencialmente relacionados con la endometriosis a través de minería de textos en PubMed

Jorge Vallejo Ortega

01/03/2021

## 1 Palabras clave

Endometriosis, minería de textos, Shiny.

## 2 Temática escogida

La endometriosis es una enfermedad del sistema reproductor femenino caracterizada por crecimiento de tejido endometrial ectópico fuera del útero. Afecta a alrededor del 10% de mujeres en edad reproductora (Malvezzi et al. 2020). Sus principales síntomas son infertilidad y dolor (Rolla 2019) por lo que, sin ser una enfermedad que amenace directamente la vida del paciente, sí supone una importante disminución en su calidad de vida.

El tratamiento medicamentoso de la endometriosis involucra el uso de analgésicos, hormonas y reguladores hormonales. De esta forma se tratan el dolor y la progresión de la enfermedad. La única forma de tratar las lesiones producidas por la enfermedad es mediante cirugía (Rolla 2019).

A día de hoy no existe ninguna prueba genética que identifique personas con mayor riesgo de desarrollar la enfermedad, aunque sí podemos acceder a una amplia literatura científica que explora la relación entre genes y endometriosis y desvela gran cantidad de genes y regiones génicas que podrían estar involucradas (Rolla 2019). Entre otros se han identificado loci relacionados en, o cercanos a, genes involucrados en rutas hormonales (*FN1*, *CCDC170*, *ESR1*, *SYNE1* y *FSHB*); así como otros genes de funciones diversas (*CDKN2BAS*, *WNT4*, *GREB1*, *ID4*, *CDKN2B-AS1*, *VEZT*, *KDR*, *TTC39B*, e *IL1A*) (Sapkota et al. 2017).

Los polimorfismos detectados incluyen tanto polimorfismos de nucleótido único (SNPs), como variaciones en el número de copias (CVNs). En el caso de los CVNs, los genes afectados más prometedores fueron *SRRM2*, *LGALS3BP*, *CABLES2* y *FCGBP* (Mafra et al. 2017).

## 3 Problemática a resolver

El origen y patogénesis de la endometriosis son desconocidos y, aunque existen varias teorías sobre su causa, ninguna ha sido probada de forma concluyente. El método de diagnóstico de referencia es la laparoscopia. Los métodos de diagnóstico por imagen (ultrasonidos, resonancia magnética) son menos invasivos pero no son capaces de alcanzar un diagnóstico en todos los casos. No se le conocen biomarcadores confiables, ni existe consenso sobre posibles factores de riesgo ambientales o genéticos (Rolla 2019).

Una de las causas que complican la investigación de nuevos tratamientos, biomarcadores y factores de riesgo es la dificultad para modelizar esta enfermedad. Los modelos animales in vivo no consiguen reproducir el desarrollo de la endometriosis (roedores, pollos conejos), o implican mayores dificultades logísticas y éticas

(macacos); con el uso de ratones transgénicos intentando salvar la distancia entre ambos tipos de modelos animales. Los modelos in vitro (cultivo de células y tejidos humanos), permiten el estudio rápido de gran cantidad de compuestos terapéuticos, pero sigue siendo necesario el modelo in vivo para estudiar el efecto de esos posibles fármacos en un organismo completo. El uso del análisis bioinformático de los datos aportados por las diferentes áreas “ómicas”, los estudios de asociación del genoma completo (GWAS, *genome wide association study*), y la minería de texto son las herramientas que se espera que aporten un mayor avance al estudio de esta enfermedad (Malvezzi et al. 2020).

## 4 Objetivos

### 4.1 Objetivo general

- Encontrar genes relacionados con la endometriosis aplicando técnicas de minería de textos.

### 4.2 Objetivos específicos

1. Desarrollar un script que permita realizar un procedimiento de minería de textos automáticamente, desde la recopilación de datos en bruto hasta la presentación de resultados.
2. Desarrollar una aplicación web implementando el script de minería de textos resultante del objetivo anterior.

## Referencias

- Mafrá, Fernanda, Diego Mazzotti, Renata Pellegrino, Bianca Bianco, Caio Parente Barbosa, Hakon Hakonarson, and Denise Christofolini. 2017. “Copy Number Variation Analysis Reveals Additional Variants Contributing to Endometriosis Development.” *Journal of Assisted Reproduction and Genetics* 34 (1): 117–24. doi:10.1007/s10815-016-0822-1.
- Malvezzi, Helena, Eliana Blini Marengo, Sérgio Podgaec, and Carla de Azevedo Piccinato. 2020. “Endometriosis: Current Challenges in Modeling a Multifactorial Disease of Unknown Etiology.” *Journal of Translational Medicine* 18 (1): 311. doi:10.1186/s12967-020-02471-0.
- Rolla, Edgardo. 2019. “Endometriosis: Advances and Controversies in Classification, Pathogenesis, Diagnosis, and Treatment.” *F1000Research* 8. doi:10.12688/f1000research.14817.1.
- Sapkota, Yadav, Valgerdur Steinthorsdottir, Andrew P. Morris, Amelie Fassbender, Nilufer Rahmioglu, Immaculata De Vivo, Julie E. Buring, et al. 2017. “Meta-Analysis Identifies Five Novel Loci Associated with Endometriosis Highlighting Key Genes Involved in Hormone Metabolism.” *Nature Communications* 8 (1): 15539. doi:10.1038/ncomms15539.