

# Bases de datos Entrez y NCBI

Este es un tutorial basado en el tutorial [Entrez del NCBI](#).

Entrez es una interfaz de búsqueda de bases de datos desarrollada por NCBI para acceder a bases de datos relacionadas, entre otras cosas, para:

- Bibliografía
- Secuencias de nucleótidos
- Secuencias de proteínas
- Genomas completos
- Estructuras de proteínas

Para mostrar cómo funciona Entrez vamos a buscar información sobre el gen MLH1, un gen humano involucrado en el cáncer de colon humano.

Queremos buscar:

- ARNm que representa mejor el gen
- Referencias bibliográficas más relevantes
- Secuencias de proteínas
- Dominios conservados
- Proteínas similares
- Mutaciones encontradas en el gen
- Estructura de la proteína
- Región del genoma en la que se encuentra el gen

Vaya al sitio [Entrez](#) y busque el nombre del gen: MLH1

Obtendrá resultados para la búsqueda en muchas de las bases de datos de NCBI.

Hay herramientas y filtros para reducir el número de resultados obtenidos:

- Puede buscar resultados que tienen más de un término de búsqueda. Por ejemplo, puede buscar MLH1 y Human usando *AND*. Busque "MLH1 Y human"
- También puede restringir la búsqueda a campos particulares; de lo contrario, los términos de búsqueda pueden aparecer en cualquier lugar de los registros encontrados. Para restringir la búsqueda, debe escribir la ID del campo entre corchetes. Pruebe con "MLH1 AND Human [ORGN]". También puede crear estas consultas complejas en la página de consulta avanzada de algunas de las bases de datos, por ejemplo, la página de [consulta avanzada](#) de la Base de datos de nucleótidos.

Principales campos de GenBank.

Campo	Descripción	Buscar en Entrez
Locus name	Nombre de secuencia único	[ACCN]
Sequence length	Longitud de secuencia	[SLEN]
Molecule Type	ADN, genómico, ARNm, etc.	[APUNTALAR]
Genbank Division	División para la secuencia	[APUNTALAR]
Modification Date	Fecha de la última edición	[MDAT]
Definition	Breve descripción	[TITL]
Accession	ID de acceso único. No cambia con modificaciones	[ACCN]

<b>Campo</b>	<b>Descripción</b>	<b>Buscar en Entrez</b>
Version	Número de versión de la secuencia	Todos los campos
Keywords	palabras clave que describen la secuencia	[KYWD]
Source	Nombre común de la especie fuente	[ORGN]
Organism	Nombre oficial de la especie fuente	[ORGN]
Reference	Publicaciones relacionadas	[TITL] [AUTH] [JOUR]
Features	Regiones de interés	[FKEY]
CDS	Secuencia de codificación	[FKEY]

Podemos ir a los resultados en una de las bases de datos, por ejemplo, Nucleótidos, y evaluar los resultados.

En la página de resultados de búsqueda de la base de datos, tiene filtros adicionales disponibles en la columna de la izquierda y la consulta de búsqueda utilizada en la columna de la derecha. También en la columna de búsqueda puede encontrar su historial de búsqueda reciente.

También puede buscar en la página de cualquiera de las bases de datos sin volver a la página de búsqueda general de Entrez. Por ejemplo, puede buscar solo en la base de datos de nucleótidos buscando en la página de la [base de datos de nucleótidos en](#) lugar de en la página general de Entrez.

## ARNm

Existen diferentes bases de datos de nucleótidos en Entrez, pero hagamos algunas búsquedas en la [Base de datos de nucleótidos](#).

- ¿Cuántos registros obtenemos si buscamos "cáncer de colon" en la Base de Datos de Nucleótidos?
- ¿Cuántos registros espera obtener si busca MLH1? ¿Cuántos obtienes? ¿Por qué?

Para buscar las secuencias que mejor representan el gen MLH1, es mejor buscar en la base de datos de RefSeq. Podemos filtrar los resultados que obtuvimos al seleccionar RefSeq en "Bases de datos de origen" en la columna de la izquierda.

Todavía hay muchos resultados para MLH1 en RefSeq. Podemos obtener menos resultados si restringimos la búsqueda al campo *Definition* haciendo la búsqueda "MLH1 [TITL]"

No hemos filtrado los resultados para obtener solo resultados humanos todavía. Podríamos hacerlo utilizando la búsqueda "MLH1 [TITL] Y Human [ORGN]", pero también podemos hacerlo yendo a la página de [consulta avanzada](#). En el formulario de consulta anticipada, podemos seleccionar los campos de los menús desplegables en lugar de utilizar la sintaxis del corchete cuadrado y también podemos usar una búsqueda previa y agregarle términos utilizando los números que se muestran en la tabla del historial. Por ejemplo, podríamos hacer algo similar a "Human [ORGN] Y # 1"

¿Cuántos resultados obtienes?

## Referencias Bibliográficas

Una vez que hemos localizado la secuencia de interés, es muy fácil encontrar información relacionada en la página de resultados de búsqueda o en la página de cualquiera de los registros yendo a la información relacionada de la sección. Por lo tanto, podemos buscar los documentos relacionados con nuestra búsqueda o con cualquiera de los registros de secuencia.

En el caso de Pubmed, podemos diferenciar las referencias que incluyen el texto completo mediante la búsqueda de PMC.

## Otra información relacionada

La mayoría de las bases de datos de NCBI están relacionadas de esta manera, así que podemos pasar de secuencias a documentos, desde documentos a genes, desde genes a estructuras, etc.

Busque las proteínas relacionadas con el gen humano MLH1 para sus genes homólogos.

En la página para cualquiera de sus proteínas, puede buscar dominios conservados.

- ¿Cuáles son sus dominios?
- ¿Cuáles son sus funciones?
- ¿Qué otras proteínas comparten alguno de sus dominios?

## Ejercicio 1

Busque la información más relevante para cualquiera de los genes humanos FXN, OCA2 o FOXP2.

- Función molecular
- Enfermedades relacionadas
- Estructura genómica
- Secuencia de proteínas
- Secuencias homólogas en otras especies

## Ejercicio 2

¿Hay algún gen descrito para la HIPERCOLESTEROLEMIA?

¿Cuál sería la base de datos más conveniente para buscar enfermedades humanas?

## Ejercicio 3

Busque información relacionada con los posibles efectos cancerígenos de: nicotina, leche, cannabis y plátano. Busque información tanto en Google como en Pubmed.

- ¿Cuántas referencias ha encontrado en ambas búsquedas? ¿Son consistentes las referencias? ¿Puedes obtener una conclusión sólida? ¿Qué evidencias valorarías más?
- Busque en Pubmed, pero solo en las referencias de "metanálisis". ¿A qué conclusiones puede llegar ahora?