

BASES DE DATOS EJERCICIOS

Ejercicio 1

Imaginemos que estamos interesados en conocer más sobre los genes relacionados en el cáncer de mama.

- ¿Cuáles son estos genes?
- ¿Cuáles son los números de acceso para varios de estos genes?
 - PMIDs de artículos científicos relevantes
 - Entrez IDs para los genes
 - Uniprot IDs para sus productos
- ¿Alguna otra información relevante?
 - Localización en cromosomas
 - Diagnósticos y relevancia de cada gen involucrado
 - Mutaciones importantes
 - Familias de proteínas
 - ...

Para identificarlos, tenemos muchas bases de datos en las que buscar. Probad con todas estas aproximaciones

- PubMed para publicaciones relacionadas. Podemos también empezar con Wikipedia o una búsqueda en Google para una introducción y explicaciones más “sencillas” de la función y relación de los genes con el cáncer de mama.
- Entrez para genes, su localización en el genoma, anotaciones funcionales, etc.
- ExPASy para proteínas, proteínas homólogas, relación con el gen, etc.
- OMIM para desordenes de origen genético.
- ...

Encontrando un gen “nuevo”

La primera práctica a entregar consistirá en la búsqueda de un gen “nuevo” relacionado con algún gen ya conocido (por ejemplo, un gen en una planta parecido al gen XXX relacionado con el cáncer de mama), involucrando los contenidos de los primeros temas de la asignatura.

Con este fin, puede ser interesante repetir este ejercicio para un tema (enfermedad, desarrollo, tejido) que te interese realmente y para el que vayas a buscar un gen “nuevo”.

Ejercicio 2

¿Cuántas ubiquitinas hay en el genoma humano, y cuál es la secuencia de una ubiquitina prototipo?

- Busca bien en la literatura (PubMed), o en algún otro medio fiable (Wikipedia?) cuál es la ubiquitina prototipo y descarga su secuencia de 76 aminoácidos.
- Utiliza Entrez Gene y las búsquedas avanzadas para limitar la búsqueda a ubiquitinas en humano. Seguramente obtendrás muchas coincidencias, prueba a repetir la búsqueda descartando “pseudogene”, “-like”, “associated”, “peptidase”, “ligase” y “enzyme”. ¿Qué hemos descartado con esta nueva búsqueda más restringida?
- Utiliza ahora bases de datos de proteínas UniProt para realizar la misma búsqueda de ubiquitinas en humano. ¿Cuántos resultados obtienes? De manera similar a las restricciones en Entrez, ¿qué términos filtrarías para reducir los resultados a aquellos relevantes?

Ejercicio 3

Encuentra el número de acceso de una lipocalina (por ejemplo: la lactoglobulina en la vaca, la apolipoproteína D, una lipocalina bacteriana o la glicodelina).

Usa Entrez, UniGene y OMIM. :

- ¿Cuál de las tres estrategias es más efectiva? ¿Cuáles son sus aportaciones a la información sobre la lipocalina?
- ¿Cuál es la función de la proteína buscada?

Ejercicio 4

Estamos interesados en saber más sobre el virus del Nilo Occidental (West Nile virus). Qué pasa si buscamos por “West Nile virus” en

- NCBI (Entrez, PubMed, OMIM)
- EBI (Sequence Retrieval System –SRS-)
- Universidad de Kyoto (DBGET)

Describe qué encuentras con cada búsqueda y sus diferencias.

Ejercicio 5

Nos gustaría saber qué artículos se han publicado en la revista BMC Bioinformatics relacionados con el virus del ejercicio 4. Haz esta búsqueda mediante PubMed y comenta los resultados más relevantes.