

## **Actividades de priorización de genes**

### **Endeavour**

#### **Ejercicio 1**

En un estudio sobre una enfermedad hereditaria autosómica recesiva de la que se conocen mutaciones en genes que causan la enfermedad, se encuentra un grupo de pacientes en los que no aparece ninguna de estas mutaciones por lo que se decide realizar secuenciación de exoma de los pacientes y sus familiares. Tras la filtración de variantes según su frecuencia alélica y la co-segregación con la enfermedad se llega a una lista de genes de los cuáles alguno puede ser causante de dicha enfermedad. Para escoger con qué gen comenzar los análisis de validación se realizó una priorización de genes.

Realiza la priorización de genes con Endeavour utilizando los siguientes datos:

- Especie: Homo sapiens
  - Genes de entrenamiento: seed\_genes\_01.txt
  - Fuentes de datos: Gene Ontology, Text-minig, GAD, OMIM, String, PaGenBase, Blast
  - Genes candidatos: list\_genes\_01.txt
1. ¿Qué gen de la lista de genes candidatos tiene mayor probabilidad de estar implicado en la enfermedad?
  2. ¿Cuántos genes aparecen entre los p-valores 0.0 y 0.15? ¿Aparece algún gen candidato?
  3. Busca los genes PIEZO1 y PIEZO2, ¿qué p-valores tienen en Text-mining? ¿Qué crees que significa? ¿Cómo crees que afecta al p-valor global?

#### **Ejercicio 2**

En un grupo de pacientes con hernia diafragmática congénita se ha realizado un estudio de exomas y tras la filtración de variantes se ha llegado a una lista de genes candidatos que es necesario validar.

Realiza la priorización de genes con Endeavour utilizando los siguientes datos:

- Especie: Homo sapiens
  - Genes de entrenamiento: seed\_genes\_02.txt
  - Fuentes de datos: selecciona todas las fuentes de datos excepto las de los grupos "Chemical information" y "Bio-molecular pathways".
  - Genes Candidatos: genes\_list\_02.txt
1. ¿Qué gen de la lista de genes candidatos tiene mayor probabilidad de estar implicado este proceso?
  2. Guarda la lista completa de genes. ¿Qué información aparece en el fichero? Pista: "Export".
  3. Escoge los primeros 5 primeros genes candidatos en la clasificación y descarga el fichero con sus p-valores. Para ello selecciona los genes arrastrando el cursor por la imagen donde aparece la columna p-valor.

### Ejercicio 3

Un grupo de investigadores está estudiando los genes relacionados con la reparación de rotura de una hebra de ADN (*single strand break repair* GO:0000012). Los análisis que han estado haciendo han dado lugar a una lista de genes candidatos que pueden estar relacionados con este proceso (fichero genes\_list\_03.txt). Utiliza la herramienta web Endeavour para descubrir qué genes tienen una mayor probabilidad de estar implicados en dicho proceso.

- Especie: Homo sapiens.
  - Genes de entrenamiento: Gene Ontology GO:0000012
  - Fuentes de datos: Gene Ontology, Uniprot, Text-mining, String, iRefIndex, GeneRIF, PaGenBase, Blast.
  - Genes candidatos: list\_genes\_03.txt
1. ¿Qué gen tiene una mayor probabilidad de estar relacionado con este proceso biológico?
  2. ¿Por qué aparece una línea verde?
  3. Guarda la clasificación en un fichero en el que solo aparezcan los nombres de los genes y el p-valor global. Pista: combina “keep” y “Export”.
  4. Busca el gen ACR. ¿Cuál es su p-valor? ¿Qué base de datos le está afectando más negativamente? ¿Qué ocurrirá si eliminamos esa base de datos?

### Ejercicio 4

En un estudio sobre microsomía hemifacial los investigadores han localizado una duplicación de 1.3 Mb en 14q22.3, exactamente la región chr14: 56,799,905 - 58,298,139, ahora se preguntan cuál de los genes incluidos en esta región cromosómica está afectando al desarrollo de los pacientes con microsomía hemifacial.

Utiliza Endeavour para priorizar los genes candidatos con los siguientes datos:

- Especie: Homo sapiens
  - Genes de entrenamiento: CHD7, ZIC3, SALL1.
  - Fuentes de datos: Gene Ontology, Uniprot, Textmining, Annot GAD, Inter String, PaGenBase, BLAST
  - Genes candidatos: región cromosómica chr14:56,799,905-58,298,139
1. ¿Cuál es el gen que tiene una mayor probabilidad de ser el causante de esta malformación? ¿Qué p-valor global tiene?
  2. Ahora realiza la misma priorización pero utilizando todas las fuentes de datos. ¿Qué ha ocurrido con el gen que antes aparecía el primero en el ranking de los candidatos? ¿Por qué crees que ha ocurrido esto? ¿Qué fuentes de datos le están afectando de manera negativa?