

## DISEÑO CONCEPTUAL TAREA 2.1

Descripción del diseño conceptual de la Tarea 2.1 de la asignatura de Sistemas Bioinformáticos.

### RELACIÓN MUESTRA-RUN

La relación que utilizo para este caso es **Se ha secuenciado**. La cardinalidad entre **MUESTRA-RUN** es 0, 1 ya que se secuencia solamente una vez una muestra o todavía no se ha secuenciado. Así, un RUN puede tener 1 o N muestras. Es raro utilizar una muestra sola para secuenciar, pero puede ser que se utilice para completar un estudio en el que faltaba esa muestra o para predecir si esa muestra en concreto es sana o enferma tras haber creado un modelo predictivo previo.

Los atributos que utilizo en **MUESTRA** son “Código de identidad” de la muestra como atributo de identidad, y “Propiedades biológicas”.

Los atributos de **RUN** son “Fecha de envío”, “Fecha de secuenciación”, “Librerías utilizadas”, “Modelo del secuenciador” y “Número de copias”.

### RELACIÓN RUN-GEN

Utilizo la relación **Tiene lecturas de**, ya que un **RUN** tiene lecturas de 0 o N **GENES**, por lo que la cardinalidad es 0, N. Puede que la secuenciación haya ido mal en algún punto y no se hayan encontrado copias por degradación de la muestra, fallo en la secuenciación, etc. De la misma manera, un **GEN** puede estar presente en 0 o N **RUN**, es decir un gen de una especie que no se haya secuenciado no estará en ningún RUN, o un transcrito que se exprese esporádica (ej. solamente durante el desarrollo) no estará presente en individuos adultos del RUN.

Los atributos de **GEN** son “Entrez ID” y “Ensembl ID” como atributos de identidad ya que para cada gen en cada especie tiene un indicador propio y exclusivo. En cambio “Symbol” no lo pongo como atributo de identidad porque el mismo gen puede tener el mismo “Symbol” entre varias especies, como por ejemplo APOE (humanos, pandas, perros) o Apoe (ratones, hámsteres). También considero incluir “Secuencia” como atributo para buscar genes homólogos entre especies.

En la relación **Tiene lecturas de** pongo que tiene el atributo de “Numero de lecturas”. Esto es que en el **RUN** se han encontrado copias del **GEN** a estudiar. De esta manera, podemos ver diferencia cuantitativa entre distintos genes en el RUN según su número de copias encontradas.

### RELACIÓN GEN-ESPECIE

La relación es que un **GEN Pertene a una ESPECIE**. La cardinalidad **GEN-ESPECIE** lo considero 1, 1 porque un gen con su secuencia propia solamente puede ser de una especie. Otras especies pueden tener genes homólogos con más o menos similitud pero no exactamente iguales. La cardinalidad **ESPECIE-GEN** la considero 0, N porque puede haber una especie que no se haya secuenciado el genoma y no tengamos datos previos de los genes que contiene.

Pongo “Taxonomía” como atributo de Especie porque considero que es esencial para la comparación de genes homólogos y ver la distancia entre dos o más especies distintas.

### RELACIÓN MUESTRA-ESPECIE

Una muestra **Es de** una especie. La cardinalidad entre **MUESTRA-ESPECIE** la considero 0, 1, ya que podemos tener una muestra de una especie que no conozcamos todavía y no la tengamos anotada. Por otro lado, si la especie es conocida, la muestra solamente puede ser una especie. La cardinalidad entre **ESPECIE-MUESTRA**, es de 0, N. Podemos tener especies anotadas que no se hayan utilizado y de la misma manera muchas muestras pueden ser de la misma especie.

#### RELACIÓN GEN-RUTA METABÓLICA

Un gen **Pertenece** a ruta/s metabólicas. Considero la cardinalidad entre **GEN-RUTA METABÓLICA** de 1, N, ya que un gen puede estar en una o varias rutas metabólicas distintas. En la cambio la cardinalidad **RUTA METABÓLICA-GEN** la considero 2, N porque es necesario mínimo 2 moléculas para que sea una ruta.

Los atributos de **RUTA METABÓLICA** son “Código de ruta” como atributo de identidad, “Nombre” y “Descripción” de la ruta.

#### RELACIÓN ESPECIE-RUTA METABÓLICA

Una especie **Tiene** ruta/s metabólicas. La cardinalidad entre **ESPECIE-RUTA METABÓLICA** es de 0,N. Al igual que antes, considero 0, porque puede ser una especie que no tengamos información suficiente. Y N porque si conocemos la especie tiene muchas rutas metabólicas. Entre **RUTA METABÓLICA-ESPECIE** la cardinalidad es de 1,N porque una ruta puede estar mínimo en una especie.

#### RELACIÓN FÁRMACO-RUTA METABÓLICA

Un fármaco **Modifica** sobre ruta/s metabólicas. Entre **FÁRMACO-RUTA METABÓLICA** la cardinalidad es 1,N porque un fármaco puede actuar sobre una o varias rutas metabólicas. Entre **RUTA METABÓLICA-FÁRMACO** lo considero 0,N porque puede haber rutas que no sea afectadas por ningún fármaco, o que una ruta puede ser modulada por varios fármacos.

#### RELACIÓN FÁRMACO-GEN

Un fármaco **Actúa** sobre gen/es. La cardinalidad **FÁRMACO-GEN** considero 0,N porque un fármaco puede actuar sobre varios genes o ninguno directamente. Puede ser que un fármaco no actúe sobre un gen sino bloqueando una proteína que actúa en la inflamación, por ejemplo. Secundariamente o de manera colateral puede haber genes afectados por el fármaco, pero no directamente. De manera similar, la cardinalidad **GEN-FÁRMACO** también es 0,N por la misma teoría. Un gen puede ser modificado por varios fármacos o por ninguno.

En la relación **Actúa**, considero incluir un subtipo “Potencia” o “Inhibe” de manera disjunto, ya que o potencia o inhibe, pero no las dos a la vez. Lo incluyo en la relación porque creo que estos subtipos son exclusivos de la relación entre Gen y Fármaco, y no lo considero atributo propio de ninguno de ellos.