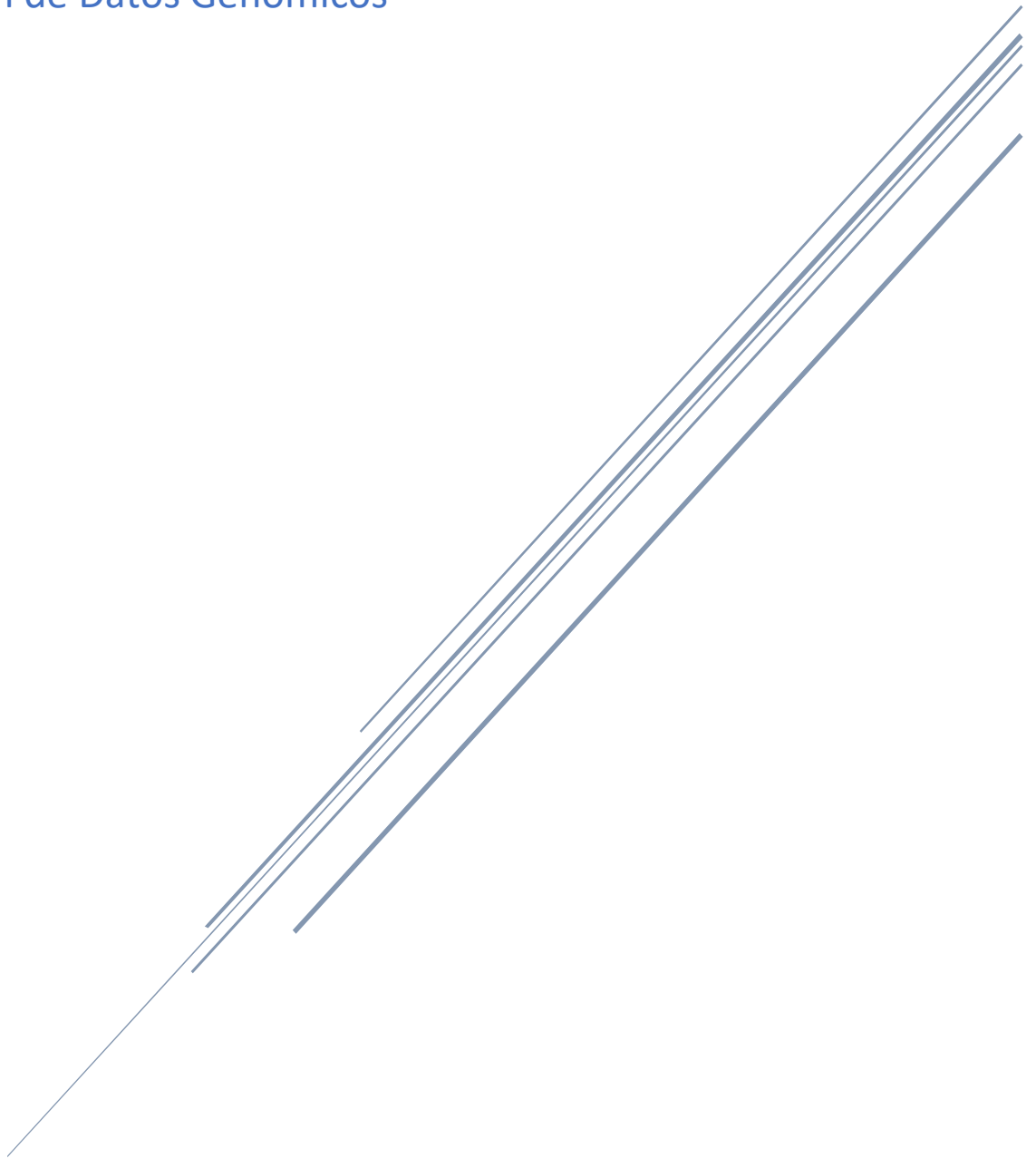


PRÁCTICA 1: INFORMACIÓN SOBRE GENES

Sistemas de Información Aplicados a la Bioinformática:
Gestión de Datos Genómicos



MITSS
Sergi Sanz

Índices:

EJERCICIO 4 – LOCALIZACIÓN CITOGÉNÉTICA.....	2
4.1.- ¿CUAL ES LA LOCALIZACIÓN CITOGÉNÉTICA DEL GEN ADAM10?	2
4.2.- ¿ES LA MISMA EN AMBAS FUENTES DE DATOS?	2
4.3.- ¿EN QUÉ CROMOSOMA SE ENCUENTRA EN GEN ADAM 10?	2
EJERCICIO 5 – COMPARACIÓN DE INFORMACIÓN ENTRE FUENTES DE DATOS	2
5.1.- ¿QUÉ ATRIBUTOS DE LA CLASE “GENE” DEL MODELO CONCEPTUAL SE PUEDEN ENCONTRAR EN NCBI GENE?	2
5.2.- ¿QUÉ ATRIBUTOS DE LA CLASE “GENE” DEL MODELO CONCEPTUAL SE PUEDEN ENCONTRAR EN HGNC?	2
EJERCICIO 6 – COMPARACIÓN DE TIPOS DE DATOS.....	3
6.1.- ENTRE LOS ATRIBUTOS COMUNES EN AMBAS FUENTES Y EL ESQUEMA CONCEPTUAL, DE DATOS HAY DOS QUE NO CONTIENEN LOS MISMOS VALORES ¿CUÁLES SON?.....	3
6.2.- ¿CUANTOS VALORES CONSIDERA CADA FUENTE DE DATOS PARA EL CAMPO “BIOTYPE”? PARA RESOLVER ESTA PREGUNTA SERÁ NECESARIO LOCALIZAR Y FAMILIARIZARSE CON LOS DISTINTOS TIPOS DE AYUDA QUE PROPORCIONA CADA FUENTE DE DATOS. .	3
EJERCICIO 7 – INFORMACIÓN SOBRE PROTEÍNAS.....	3
7.1.- ¿CUAL ES EL NOMBRE DE LA PROTEÍNA QUE CODIFICA EL GEN ADAM10?.....	3
EJERCICIO 8 – DISCREPANCIAS EN LOS VALORES DEL BIOTIPO	3
8.1.- ¿POR QUÉ SE PRODUCE ESTA DISCREPANCIA ENTRE LOS VALORES QUE CADA FUENTE DE DATOS ACEPTA PARA REPRESENTAR EL BIOTIPO?	3
EJERCICIO 9 – BIBLIOGRAFÍA ASOCIADA	4
9.1.- ¿CUAL DE LAS DOS FUENTES DE DATOS PROPORCIONA BIBLIOGRAFÍA MÁS EXHAUSTIVA SOBRE EL GEN?.....	4

Ejercicio 4 – Localización Citogenética

4.1.- ¿Cual es la localización citogenética del gen ADAM10?

La localización citogenética del gen ADAM10 es 15q21.3

4.2.- ¿Es la misma en ambas fuentes de datos?

Si es la misma en ambas bases de datos

4.3.- ¿En qué cromosoma se encuentra en gen ADAM 10?

La ubicación del gen ADAM 10 se encuentra en el cromosoma 15.

Ejercicio 5 – Comparación de información entre fuentes de datos

5.1.- ¿Qué atributos de la clase “gene” del modelo conceptual se pueden encontrar en NCBI Gene?

Los atributos de la clase “gene” del modelo conceptual que podemos encontrar en NCBI Gene son:

- Id_symbol representado como **Official Symbol**
- Id_hugo representado como **Official Full Name**
- Oficial_name representado como **Primary source**
- Biotype representado como **Gene type**
- Description representado como **summary**
- Status representado como **RefSeq status**
- Gene_synonym representado como **also known as**

El único atributo que no esta representado de la clase “gene” es gc_percentage

5.2.- ¿Qué atributos de la clase “gene” del modelo conceptual se pueden encontrar en HGNC?

Los atributos de la clase “gene” del modelo conceptual que podemos encontrar en HGNC son:

- Id_symbol representado como **Approved symbol**
- Id_hugo representado como **Approved name**
- Oficial_name representado como **HGNC ID**
- Biotype representado como **Locus type**
- Status representado como **Symbol status**
- Gene_synonym representado como **also Alias symbols**

El único atributo que no está representado de la clase “gene” es `gc_percentage`, y `description`

Ejercicio 6 – Comparación de tipos de datos

6.1.- Entre los atributos comunes en ambas fuentes y el esquema conceptual, de datos hay dos que no contienen los mismos valores ¿cuáles son?

En el atributo `gene_synonym` en el caso de HGNC obtenemos los atributos: *kuz*, *MADM*, *HsT18717*, *CD156C*. Mientras que en el caso de NCBI Gene obtenemos: *RAK*; *kuz*; *AD10*; *AD18*; *MADM*; *CD156c*; *CDw156*; *HsT18717*. Por tanto, en el NCBI Gene se incluyen atributos adicionales como son *RAK*, *AD10*, *CDw156* y *AD18*.

Otro caso de discrepancia entre los valores es el que se produce con el campo `biotype`, ya que en el caso de HGNC los valores son: *O14672*, *InterPro*, *PDBe*, *Reactome*. Mientras que en el caso de NCBI los valores son: *tRNA*, *rRNA*, *snRNA*, *scRNA*, *snoRNA*, *miscRNA*, *ncRNA*, *protein coding*, *pseudo*, *other*, and *unknown*.

6.2.- ¿Cuántos valores considera cada fuente de datos para el campo “*biotype*”? Para resolver esta pregunta será necesario localizar y familiarizarse con los distintos tipos de ayuda que proporciona cada fuente de datos.

En la fuente de datos HGNC se aprecian los valores de datos del campo `biotype`: *O14672*, *InterPro*, *PDBe*, *Reactome*.

Mientras que en la fuente de datos NCBI los valores de datos del campo `biotype` son: *tRNA*, *rRNA*, *snRNA*, *scRNA*, *snoRNA*, *miscRNA*, *ncRNA*, *protein coding*, *pseudo*, *other*, and *unknown*.

Ejercicio 7 – Información sobre proteínas

7.1.- ¿Cuál es el nombre de la proteína que codifica el gen *ADAM10*?

- ENSP00000260408 (*ADAM10-201*)
- ENSP00000391930 (*ADAM10-204*)
- ENSP00000452704 (*ADAM10-212*)

Ejercicio 8 – Discrepancias en los valores del biotipo

8.1.- ¿Por qué se produce esta discrepancia entre los valores que cada fuente de datos acepta para representar el biotipo?

La discrepancia entre los distintos valores de la fuente de datos se produce por la procedencia de estos, ya que NCBI Gene se basa en la nomenclatura proporcionada por la International Nucleotide Sequence Databases (INSDC). Mientras que el HGNC utiliza la nomenclatura

proporcionada por Sequence Ontology (junto a cada tipo se encuentra un link del tipo SO:0001217 que lleva a la respectiva página web de SO).

Ejercicio 9 – Bibliografía asociada

9.1.- ¿Cual de las dos fuentes de datos proporciona bibliografía más exhaustiva sobre el gen?

La fuente de datos en la que podemos apreciar mayor cantidad de información sobre el gen ADAM10 es NCBI Gene, donde podemos apreciar un gran número de referencia bibliográficas.