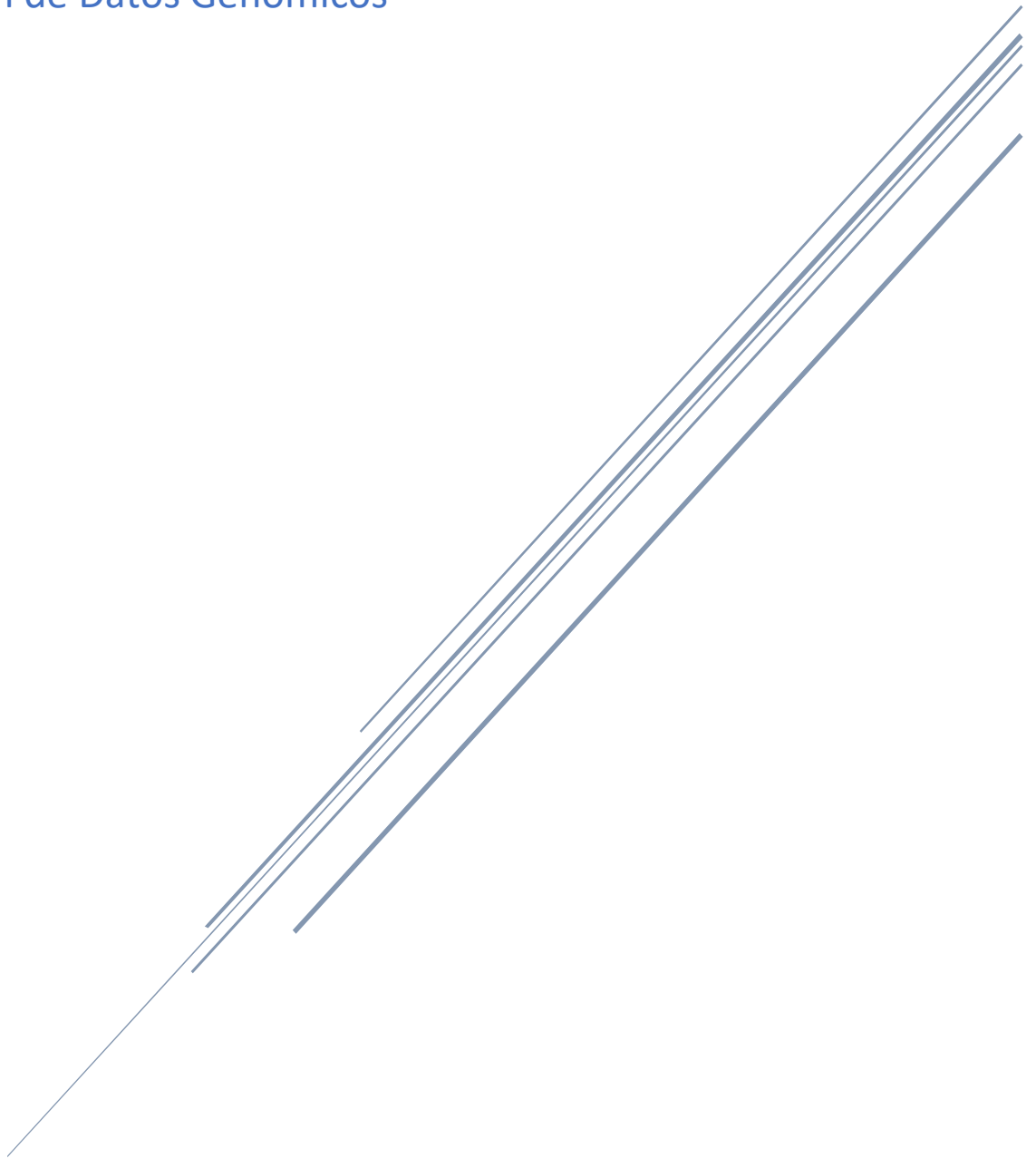


PRÁCTICA 2: INFORMACIÓN SOBRE GENES

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



MITSS
Sergi Sanz

Índices:

EJERCICIO 2 – TIPOS DE SECUENCIAS DE REFERENCIA.....	2
2.1.-¿A QUÉ ELEMENTO CROMOSÓMICO CORRESPONDE LA SECUENCIA DE REFERENCIA NC_000023.10?	2
2.2.- ¿CUÁLES SON LOS TIPOS DE SECUENCIAS DE REFERENCIA QUE PUEDEN SER UTILIZADOS COMO PREFIJO?	2
2.3.- LAS SECUENCIAS DE REFERENCIA SE REPRESENTAN POR UN IDENTIFICADOR LLAMADO CONOCIDO COMO “ACCESSION IDENTIFIER”. ¿CUÁLES SON LOS FORMATOS APROBADOS PARA ESTE IDENTIFICADOR?	2
2.4.- SEGÚN LAS RECOMENDACIONES GENERALES ¿QUÉ PREFIJO SE UTILIZA PARA INDICAR QUE LA SECUENCIA DE REFERENCIA SE CORRESPONDE CON UNA SECUENCIA DE ADN CODIFICANTE?	3
EJERCICIO 3 – TIPOS DE VARIACIONES.....	3
3.1.- ¿CUÁNTOS TIPOS DE VARIACIONES DE ADN SE PUEDEN REPRESENTAR CON LA NOMENCLATURA HGVS?	3
EJERCICIO 4 – REPRESENTACIÓN DE VARIACIONES	4
4.1.- ¿CÓMO SE REPRESENTA LA POSICIÓN EN LA QUE SUCEDE LA INSERCIÓN?	4
EJERCICIO 5 – REPRESENTACIÓN COMPLETA DE UNA VARIACIÓN.....	4
5.1.- ¿CÓMO SE REPRESENTARÍA LA ELIMINACIÓN DE LA SECUENCIA DE NUCLEÓTIDOS “GTC” EN LA POSICIÓN 89838 DE LA SECUENCIA GÉNICA NG_033876.1?	4
EJERCICIO 6 – RESOLVER LAS SIGUIENTES PREGUNTAS DE INVESTIGACIÓN.....	4
6.1.- EXPLICAR EL SIGNIFICADO DE LAS SIGUIENTES EXPRESIONES HGVS: c.419_420ins[T;401_419] , c.77_80inv, c.20_23dup	4
6.2.- ES CORRECTA LA SIGUIENTE EXPRESIÓN HGVS? g.9002_9009delinsTTT EN CASO AFIRMATIVO ¿QUÉ REPRESENTA?	4
6.3.- ¿ES LA SIGUIENTE AFIRMACIÓN VERDADERA O FALSA? LOS NÚMEROS DE LOS NUCLEÓTIDOS BASADOS EN SECUENCIAS DE REFERENCIA GENÓMICAS NO INCLUYEN LOS PREFIJOS “+”, “-” Y “*”	4
6.4.- ¿QUÉ SIGNIFICA EL “*” EN UNA EXPRESIÓN HGVS?	5
6.5.- ¿DEBERÍAN LAS EXPRESIONES SOBRE PROTEÍNAS DESCRIBIR LOS CAMBIOS SOLO A ESTE NIVEL O DEBERÍAN INCLUIR TAMBIÉN INFORMACIÓN SOBRE LOS CAMBIOS A NIVEL DE ADN?	5

Ejercicio 2 – Tipos de secuencias de referencia

2.1.-¿A qué elemento cromosómico corresponde la secuencia de referencia NC_000023.10?

Corresponde a *Homo sapiens* chromosome X.

2.2.- ¿Cuáles son los tipos de secuencias de referencia que pueden ser utilizados como prefijo?

Usando un prefijo que precede a la descripción de la variante. Los tipos de secuencia de referencia aprobados:

ADN

- g. = secuencia de referencia genómica lineal.
- o. = secuencia de referencia genómica circular.
- m. = referencia mitocondrial (caso especial de una secuencia de referencia genómica circular).
- c. = secuencia de referencia de ADN codificante (basada en una transcripción de codificación de proteínas).
- n. = secuencia de referencia de ADN no codificante (basada en una transcripción que no codifica una proteína).

ARN

- r. = Secuencia de referencia de ARN.

Proteína

- p. = secuencia de referencia de proteínas.

2.3.- Las secuencias de referencia se representan por un identificador llamado conocido como “accession identifier”. ¿Cuáles son los formatos aprobados para este identificador?

El formato aprobado para este identificador es el de una secuencia de referencia que debe contener tanto el número de acceso como el de versión, un ejemplo sería NG_012232.1 el cual sería correcto, ya que el “.” se utiliza para indicar la versión, por lo que NG_012232 no sería correcto.

Otro identificador es “:” se utiliza como separador entre el identificador del archivo de secuencia de referencia y la descripción real de una variante. Un ejemplo sería: NC_000011.9:g.12345611G>A

2.4.- Según las recomendaciones generales ¿qué prefijo se utiliza para indicar que la secuencia de referencia se corresponde con una secuencia de ADN codificante?

Se utiliza el prefijo “c” para una secuencia de referencia de ADN codificante.

Ejercicio 3 – Tipos de variaciones

3.1.- ¿Cuántos tipos de variaciones de ADN se pueden representar con la nomenclatura HGVS?

Los tipos de variaciones son:

- **Sustitución:** Una letra (nucleótido) del código de ADN se reemplaza (sustituye) por otra letra. Una sustitución se indica usando ">".
 - Ejemplo: **c.4375C>T** (el nucleótido C en la posición c.4375 cambió a una T).
- **Borrado:** Faltan una o más letras del código de ADN (eliminado). Una eliminación se indica usando "del".
 - Ejemplo: **c.4375_4379del** (faltan los nucleótidos de la posición c.4375 a c.4379 (CGATT) (eliminado). A menudo se informa como c.4375_4379delCGATT) .
- **Duplicados:** Una o más letras del código de ADN están presentes dos veces (duplicadas, duplicadas). Una duplicación se indica usando "dup".
 - Ejemplo: **c.4375_4385dup** (los nucleótidos de la posición c.4375 a c.4385 (CGATTATCCA) están presentes dos veces (duplicados)).
- **Insertar:** Una o más letras en el código de ADN son nuevas (insertadas). Una inserción se indica con "ins".
 - Ejemplo: **c.4375_4376insACCT** (la nueva secuencia "ACCT" se encontró insertada entre las posiciones c.4375 y c.4376).
- **Eliminación / inserción (indel) :** faltan una o más letras en el código de ADN y son reemplazadas por varias letras nuevas. Una eliminación / inserción se indica usando "delins".
 - Ejemplo: **c.4375_4376delinsAGTT** (faltan los nucleótidos de la posición c.4375 a c.4376 (CG) (se eliminan) y se reemplazan por la nueva secuencia "AGTT").

Ejercicio 4 – Representación de variaciones

4.1.- ¿Cómo se representa la posición en la que sucede la inserción?

La posición que sucede en la inserción se representa entre los dos rangos que se establecen para introducir esa inserción, por ejemplo c.4375_4376insACCT (la nueva secuencia "ACCT" se encontró insertada entre las posiciones c.4375 y c.4376).

Ejercicio 5 – Representación completa de una variación

5.1.- ¿Cómo se representaría la eliminación de la secuencia de nucleótidos "GTC" en la posición 89838 de la secuencia génica NG_033876.1?

La representación de la eliminación de la secuencia "GTC" en la posición 89838 se representaría como NG_033876.1:g.89838_89840delGTC.

Ejercicio 6 – Resolver las siguientes preguntas de investigación

6.1.- Explicar el significado de las siguientes expresiones HGVS: c.419_420ins[T;401_419] , c.77_80inv, c.20_23dup

En el caso de c.419_420ins[T;401_419] se inserta en las posiciones c.419 y c.420 la secuencia T y las secuencias comprendidas entre las posiciones 401_419.

En el caso de c.77_80inv indica inversión, aunque generalmente no se utiliza a nivel de proteína, generalmente se describe como "delins".

En el caso de c.20_23dup los nucleótidos de la posición 20 a 23 están duplicados.

6.2.- Es correcta la siguiente expresión HGVS? g.9002_9009delinsTTT En caso afirmativo ¿qué representa?

Si, y representa que en g.9002_9009delinsTTT faltan los nucleótidos de la posición g.9002_9009 (se eliminan) y se reemplazan por la nueva secuencia "TTT".

6.3.- ¿Es la siguiente afirmación verdadera o falsa? Los números de los nucleótidos basados en secuencias de referencia genómicas no incluyen los prefijos "+" , "-" y "*" .

Falsa, ya que no se deben utilizar los símbolos "+" o "-" en las secuencias de referencias genómicas. Sin embargo, la numeración de los nucleótidos de las secuencias de referencias de ADNc está basada en la isoforma de la proteína anotada más grande que se traduce.

Los nucleótidos posicionados aguas arribas del codón de iniciación ATG están nombrados con el signo menos "-" y son numerados como c.-1, c.- 2, c.-3, etc. y los posicionados aguas abajo del codón de terminación están nombrados con un asterisco "*"

6.4.- ¿Qué significa el "*" en una expression HGVS?

"*" (Asterisco) se utiliza en la numeración de nucleótidos y para indicar un codón de terminación de la traducción (parada).

6.5.- ¿Deberían las expresiones sobre proteínas describir los cambios solo a este nivel o deberían incluir también información sobre los cambios a nivel de ADN?

Las expresiones sobre proteínas deberían describir la mayor cantidad de información significativa, por tanto, deberían incluir también información sobre los cambios a nivel de ADN.