

## Genómica computacional

Enrique Blanco García y Jaume Sastre Tomàs

PEC1. Manejo de los navegadores genómicos	
Presentación y objetivos	Empleo de los navegadores genómicos para analizar información sobre los elementos funcionales de los genomas.
Fecha y formato de entrega	<b>21 de Octubre – 3 de Noviembre</b> (PDF o MS Word)
Criterios de corrección	Ejercicio 1/2 – 25% cada uno Ejercicio 3 – 40% Ejercicio 4 – 10%

### Presentación

Los navegadores genómicos permiten el acceso libre a una cantidad ingente de informaciones actualizadas, dotando al analista bioinformático de un entorno gráfico muy potente para la visualización de los genomas. Estas herramientas también permiten la descarga sin límite alguno de todas sus pistas de datos para el trabajo en nuestro propio ordenador. Utilizaremos varias funciones de los navegadores que permiten interrogar a gran escala sobre los datos existentes sin necesidad de la descarga directa.

### Objetivos

Se persigue que el estudiante adquiera experiencia en el uso de navegadores genómicos para visualizar las anotaciones existentes y extraer secuencias de trabajo. Del mismo modo, se pretende que los alumnos posean al acabar este trabajo nociones básicas sobre el mejor modo de interrogar a estos sistemas a la hora de recuperar información a escala genómica.

### Formato

**El informe final, como máximo, debe tener un tamaño de 20 páginas. Puede emplearse un apéndice para añadir el resto de informaciones.** El nombre y apellidos del estudiante deben indicarse claramente en la primera página del informe. Es necesario incluir capturas de pantalla que expliquen el modo de alcanzar la solución de los ejercicios propuestos. La solución de esta PEC1 debe entregarse en formato PDF o Word (pueden adjuntarse aparte ficheros adicionales si fuera necesario).

## Ejercicio 1. Anotar el gen de estudio [25%]

1. En primavera de 2017 conocimos la siguiente noticia a través de la mayoría de medios de comunicación. Leed el texto con atención, dado que utilizaremos esta información como base para nuestro ejercicio.

EL PAÍS

CIENCIA

SUSCRÍBETE

Identificado un gen relacionado con el envejecimiento del cerebro

Las personas que poseen algunas variantes del TMEM106B sufren un envejecimiento cerebral acelerado y son más propensas a padecer enfermedades neurodegenerativas

f

Twitter

Reddit

3

✉

📖

JOANA OLIVEIRA

29 MAR 2017 - 16:58 CEST



Investigadores de la Universidad de Columbia han descubierto un gen que tiene relación directa con el envejecimiento del cerebro. GETTY IMAGES

LA VERDAD,  
TODA LA VERDAD Y  
NADA MÁS QUE

Los científicos Asa Abeliovich y Herve Rhinn analizaron el

[https://elpais.com/elpais/2017/03/17/ciencia/1489749436\\_279057.html](https://elpais.com/elpais/2017/03/17/ciencia/1489749436_279057.html)

NEWSLETTERS

Recibe el boletín de Ciencia

2. Conectaos a PUBMED y localizad el resumen (en inglés, *Abstract*) de esta publicación mencionada en el artículo anterior:

<http://www.ncbi.nlm.nih.gov/pubmed>

3. Con el servidor genómico UCSC, buscad las anotaciones del gen *TMEM106B* (*Homo sapiens*, hg38). Utilizaremos la anotación suministrada por RefSeq (track RefSeq Genes llamada también UCSC RefSeq). Anotad la localización genómica y los genes más cercanos:

<http://genome.ucsc.edu/>

4. Dentro de la ficha del gen según RefSeq, accede al registro de Entrez Gene y de OMIM para describir que funciones desempeña según Gene Ontology y en que enfermedades puede estar involucrado.

5. Extraed de las anotaciones de RefSeq la región codificante (CDS, únicamente los exones) del gen *TMEM106B* humano. Repetid el mismo procedimiento para obtener la secuencia CDS ortóloga en el ratón (*Mus musculus*, mm10).

6. El programa *CLUSTAL Omega* realiza alineamientos globales de dos o más secuencias. Conectaos al servidor de *CLUSTAL Omega* para alinear las dos secuencias CDS (humana y de ratón) obtenidas en el paso anterior.

<http://www.ebi.ac.uk/Tools/msa/clustalo/>

7. Repetid este mismo alineamiento global, utilizando ahora las correspondientes proteínas de cada gen (que previamente debéis recuperar de la entrada de RefSeq). Valorad el grado de homología entre estas dos secuencias tanto a nivel genómico como a nivel de proteína.

8. Emplead la herramienta LiftOver para averiguar las coordenadas del mismo gen en la versión hg19 del genoma humano y en la versión mm9 del genoma del ratón, respectivamente.

<https://genome.ucsc.edu/cgi-bin/hgLiftOver>

## **Ejercicio 2. La herramienta BLAT del navegador de UCSC [25%]**

1. La aplicación BLAT es una herramienta muy popular disponible dentro del servidor genómico de UCSC. Localizad su página web y definid en pocas palabras cuál es la función principal de este programa:

<http://genome.ucsc.edu>

2. Emplead BLAT para identificar la ubicación de la secuencia CDS humana del gen *TMEM106B* dentro del genoma humano (hg38, cromosoma, coordenadas).

3. Emplead BLAT para localizar con la secuencia CDS humana del gen *TMEM106B* dónde se encuentra la ubicación de este mismo gen en el genoma del ratón (mm10, cromosoma, coordenadas).

4. Emplead BLAT para localizar con la secuencia CDS humana del gen *TMEM106B* dónde se encuentra la ubicación de este mismo gen en el genoma del pollo (galGal4, cromosoma, coordenadas).

5. Ahora emplead BLAT con la proteína humana del gen *TMEM106B* sobre el mismo genoma del pollo (galGal4, cromosoma, coordenadas). Razonad sobre las diferencias entre este resultado y el obtenido en el punto anterior.

### Ejercicio 3. La herramienta *Table browser* del navegador de UCSC [40%]

1. La aplicación *Table browser* es una herramienta muy útil para acceder a los datos de las pistas que constituyen el entorno gráfico disponible dentro del servidor genómico de UCSC. Localizad su página web y definid en pocas palabras cuál es la función principal de este programa:

<http://genome.ucsc.edu>

2. Imaginemos un escenario real: estamos trabajando con el genoma humano (ensamblado hg38) y se nos plantean una serie de cuestiones prácticas. Encontrad la manera de responder a estas preguntas empleando el navegador de tablas sobre la pista *RefSeq genes* (*track NCBI RefSeq, table UCSC RefSeq*). Debéis añadir una breve descripción en el informe de cómo habéis logrado llegar a la solución.

- Número de pares de bases del genoma completo.
- Número de transcritos en total a lo largo del genoma.
- Listado de todos los transcritos en el genoma (únicamente captura de los primeros cinco). Debéis mostrar solo los siguientes atributos: código de RefSeq del transcrito, cromosoma, hebra, inicio/final del transcrito, número de exones y nombre del gen.
- Número total y listado de los transcritos en la región chr7:12000000-13000000.
- Extraed la *custom track* con los datos de la pregunta anterior (en formato BED), para añadirle una cabecera (*track name=...*) que os permita darle el nombre y el color que os resulte más atractivo. Cargar posteriormente la pista en el navegador (hg38) y comprobad que los exones encajan con los visualizados en la pista RefSeq original.
- Extraed ahora la secuencia CDS de todos estos transcritos con el *Table browser*. Escoged una de las secuencias al azar y realizad un BLAT con ella para comprobar que encaja correctamente con el transcrito de RefSeq anotado en esa localización.

### Ejercicio 4. La herramienta Biomart del navegador ENSEMBL [10%]

1. Estudiad el modo de funcionamiento de la herramienta *Biomart*. Mostrad un ejemplo de cómo interrogar el genoma humano (hg38) para la búsqueda de datos sobre la región chr12:7,680,240-7,905,217. Por ejemplo, cómo obtener el listado de términos de *Gene ontology* para los genes que se encuentran en ese lugar del genoma o el listado de SNPs anotados en los mismos genes.

<http://www.ensembl.org/biomart/martview>