



INSTITUTO POLITÉCNICO NACIONAL
ESCUELA SUPERIOR DE CÓMPUTO
BIOINFORMATICS



PRÁCTICA 7.
GENÓMICA

NOMBRE DEL ALUMNO: GARCÍA QUIROZ GUSTAVO IVAN

GRUPO: 7CV3

NOMBRE DEL PROFESOR: ROSAS TRIGUEROS JORGE LUIS

FECHA DE REALIZACIÓN DE LA PRÁCTICA: 12/02/2025

FECHA DE ENTREGA DEL REPORTE: 19/03/2025

Índice

1	Marco teórico.....	1
1.1	Introducción a la Genómica	1
1.2	Genome Data Viewer (GDV).....	1
1.3	Single Nucleotide Polymorphisms (SNPs)	1
1.4	SNPedia.....	1
1.5	CHOPCHOP	2
2	Material y equipo.	3
2.1.1	Hardware	3
2.1.2	Software.....	3
3	Desarrollo de la práctica.	4
3.1	Tarea 1: Exploración de genomas mediante Genome Data Viewer	4
3.1.1	Acceso y navegación en el Genome Data Viewer	4
3.1.2	Análisis del genoma de Drosophila melanogaster (Mosca de la.....	5
3.1.2	Análisis del genoma de Danio rerio (Pez cebra)	7
3.2	Análisis de SNPs mediante SNPedia.....	9
3.2.1	Navegación y utilización de la plataforma SNPedia.....	9
3.2.2	Análisis del SNP rs1815739 (ACTN3) - "El gen del sprint"	10
3.2.3	Análisis del SNP rs53576 (OXTR) - "El gen de la empatía"	11
3.2.4	Análisis del SNP rs4680 (COMT) - "Polimorfismo de la cognición" .	13
3.2.5	Análisis del SNP rs1799971 (OPRM1) - "Polimorfismo del receptor opioide" 15	
3.2.6	Análisis del SNP rs9939609 (FTO) - "Gen asociado a obesidad"	16
3.3	Tarea 3: Identificación de secuencias diana para CRISPR mediante CHOPCHOP	18
3.3.1	Procedimiento realizado	18
3.3.2	Resultados obtenidos	19
3.4	Desafíos y soluciones	20
4	Conclusiones y recomendaciones.	21
5	Referencias.....	22

1 Marco teórico.

1.1 Introducción a la Genómica

La genómica es una disciplina dentro de la biología molecular que estudia el genoma completo de los organismos, incluyendo la secuenciación, mapeo y análisis de todos los genes y elementos no codificantes del ADN. A diferencia de la genética tradicional, que se centra en genes individuales, la genómica examina todos los genes y sus interrelaciones para comprender su influencia conjunta sobre el crecimiento y desarrollo de un organismo.

1.2 Genome Data Viewer (GDV)

El Genome Data Viewer es una herramienta web desarrollada por el National Center for Biotechnology Information (NCBI) que permite la visualización, análisis y exploración de genomas completos de diversos organismos. Esta plataforma proporciona información detallada sobre:

- **Submitter:** Entidad o consorcio responsable de ensamblar y presentar el genoma.
- **Release date:** Fecha en que la versión actual del genoma fue publicada.
- **Number of Chromosomes:** Cantidad total de cromosomas presentes en el organismo.

Estas anotaciones son cruciales para establecer la procedencia y calidad de las secuencias genómicas con las que se trabaja, permitiendo a los investigadores contextualizar adecuadamente sus análisis.

1.3 Single Nucleotide Polymorphisms (SNPs)

Los SNPs (polimorfismos de nucleótido único) son variaciones en una sola posición nucleotídica del genoma que se encuentran en al menos el 1% de la población. Representan la forma más común de variación genética humana y tienen importantes implicaciones:

- Sirven como marcadores biológicos para localizar genes asociados con enfermedades.
- Pueden influir directamente en la función génica y el fenotipo resultante.
- Permiten estudiar la diversidad genética en poblaciones humanas.

1.4 SNPedia

SNPedia es una base de datos de acceso público tipo wiki que recopila y cataloga información sobre SNPs humanos, proporcionando:

- **Ubicación cromosómica:** El cromosoma donde se encuentra el SNP.

- **Efectos fenotípicos:** Consecuencias biológicas o clínicas asociadas con variantes específicas.
- **Incidencia estimada:** Frecuencia de las variantes en diferentes poblaciones.

Esta información es fundamental para la medicina personalizada, estudios de asociación genética y análisis de riesgo de enfermedades.

- **Edición Genómica CRISPR-Cas9**

CRISPR-Cas9 (Clustered Regularly Interspaced Short Palindromic Repeats and CRISPR-associated protein 9) es una tecnología revolucionaria de edición genómica que permite modificar con precisión secuencias específicas de ADN. Los componentes clave incluyen:

- **Cas9:** Enzima que funciona como "tijeras moleculares" para cortar el ADN en ubicaciones específicas.
- **gRNA (ARN guía):** Molécula de ARN diseñada para dirigir Cas9 a una secuencia específica del genoma.
- **Secuencia PAM** (Protospacer Adjacent Motif): Secuencia corta requerida adyacente al sitio objetivo para que Cas9 pueda realizar el corte.

1.5 CHOPCHOP

CHOPCHOP es una herramienta bioinformática diseñada para identificar secuencias diana óptimas para experimentos de edición genética mediante CRISPR. La plataforma evalúa:

- Especificidad de la secuencia diana (minimizando efectos fuera del objetivo)
- Eficiencia prevista del corte
- Accesibilidad de la cromatina en la región diana
- Presencia y posición de la secuencia PAM requerida

La selección de secuencias diana apropiadas es crucial para el éxito de experimentos de reparación génica con CRISPR, ya que determina la precisión, eficiencia y seguridad de la edición genética.

2 Material y equipo.

Para esta práctica se usaron las siguientes herramientas de software y hardware necesarias para realizar la práctica.

2.1.1 Hardware

- Computadora.

2.1.2 Software

- Navegador web
- <https://www.ncbi.nlm.nih.gov/genome/gdv/>
- <https://www.snpedia.com/>
- <https://chopchop.cbu.uib.no>

3 Desarrollo de la práctica.

3.1 Tarea 1: Exploración de genomas mediante Genome Data Viewer

El Genome Data Viewer (GDV) es una herramienta web interactiva desarrollada por el National Center for Biotechnology Information (NCBI) que permite visualizar, analizar y explorar genomas completos. Esta plataforma resulta invaluable para investigadores y estudiantes debido a su interfaz intuitiva y la riqueza de información genómica que proporciona.

Para esta tarea, se realizó un análisis comparativo de dos organismos modelo ampliamente utilizados en investigación biomédica: la mosca de la fruta (*Drosophila melanogaster*) y el pez cebra (*Danio rerio*). El objetivo fue identificar y documentar información clave sobre sus genomas, incluyendo el submitter (entidad responsable del ensamblaje), la fecha de publicación (release date) y el número de cromosomas.

3.1.1 Acceso y navegación en el Genome Data Viewer

El proceso comenzó con el acceso al Genome Data Viewer a través de la URL proporcionada: <https://www.ncbi.nlm.nih.gov/genome/gdv/>

Al ingresar a la plataforma, se presenta una interfaz principal con diversas opciones de navegación y búsqueda. Las características más relevantes de esta interfaz incluyen:

- Un campo de búsqueda en la parte superior que permite la introducción de nombres científicos, comunes o identificadores de organismos
- Una sección de "Featured Genomes" que muestra genomas frecuentemente consultados
- Opciones avanzadas de filtrado y visualización

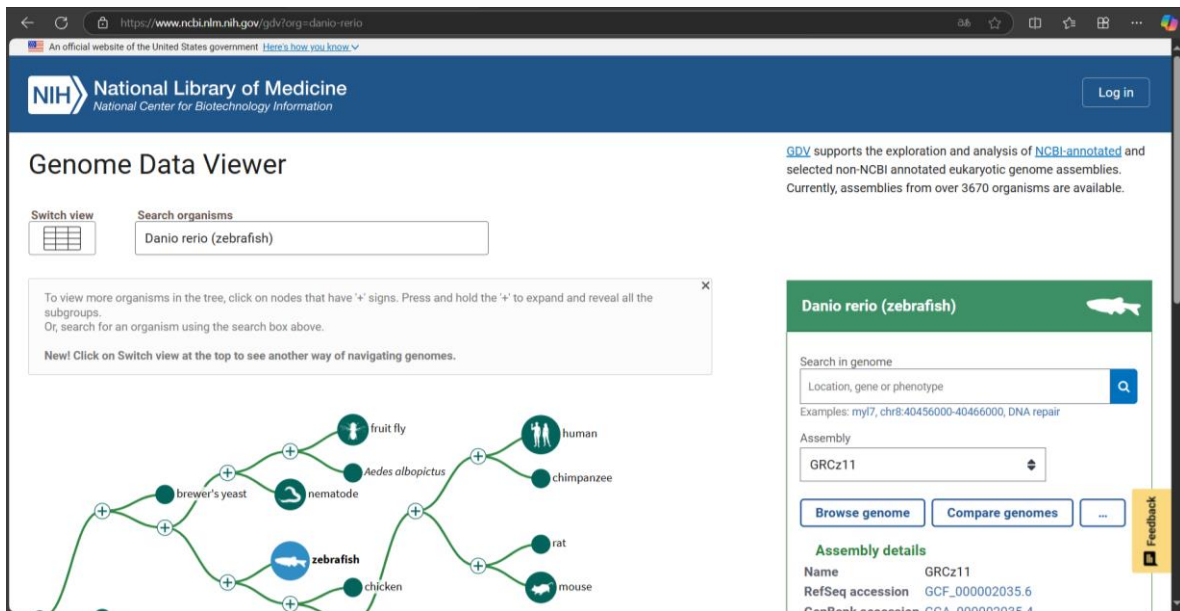


Figura 1 página principal del Genome Data Viewer

3.1.2 Análisis del genoma de *Drosophila melanogaster* (Mosca de la fruta)

La mosca de la fruta es uno de los organismos modelo más importantes en genética y biología del desarrollo debido a:

- Su ciclo de vida corto, que facilita estudios transgeneracionales
- Su genoma relativamente compacto y bien caracterizado
- La conservación de muchos genes y vías metabólicas con mamíferos, incluidos humanos

Proceso de búsqueda y selección

Para acceder al genoma de referencia de *Drosophila melanogaster*, se siguieron los siguientes pasos:

1. En el campo de búsqueda del GDV se introdujo "Drosophila melanogaster"
2. De los resultados obtenidos, se seleccionó la versión más actualizada del genoma de referencia
3. Se accedió a la vista detallada del genoma para extraer la información requerida

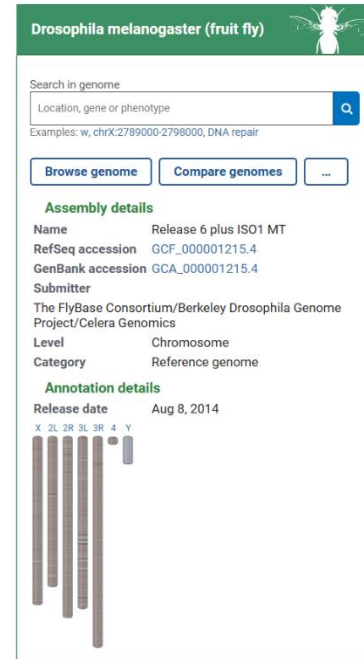
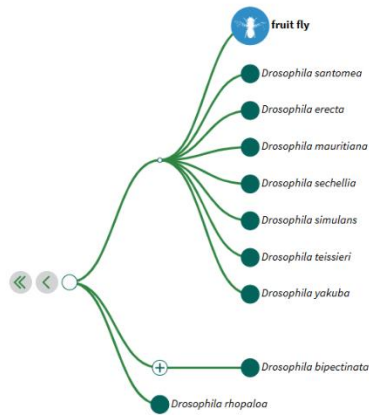


Figura 2 Imagen de los resultados de búsqueda para "Drosophila melanogaster" en el GDV

Información obtenida del genoma

Al explorar la información disponible para el genoma de referencia de *Drosophila melanogaster*, se identificaron los siguientes datos clave:

- **Entidad responsable del ensamblaje:** El Consorcio de FlyBase (The FlyBase Consortium/Berkeley Drosophila Genome Project/Celera Genomics)
- **Fecha de publicación:** 8 de agosto del 2014
- **Numero de Cromosomas:** 4 pares de cromosomas (X/Y, 2, 3, 4)

La información del submitter resulta particularmente relevante ya que FlyBase es un recurso especializado para la comunidad de investigación en *Drosophila*, lo que garantiza un alto nivel de curación y anotación del genoma.

Esta visualización permite comprender la organización espacial del genoma y la distribución de elementos funcionales a través de los cromosomas.

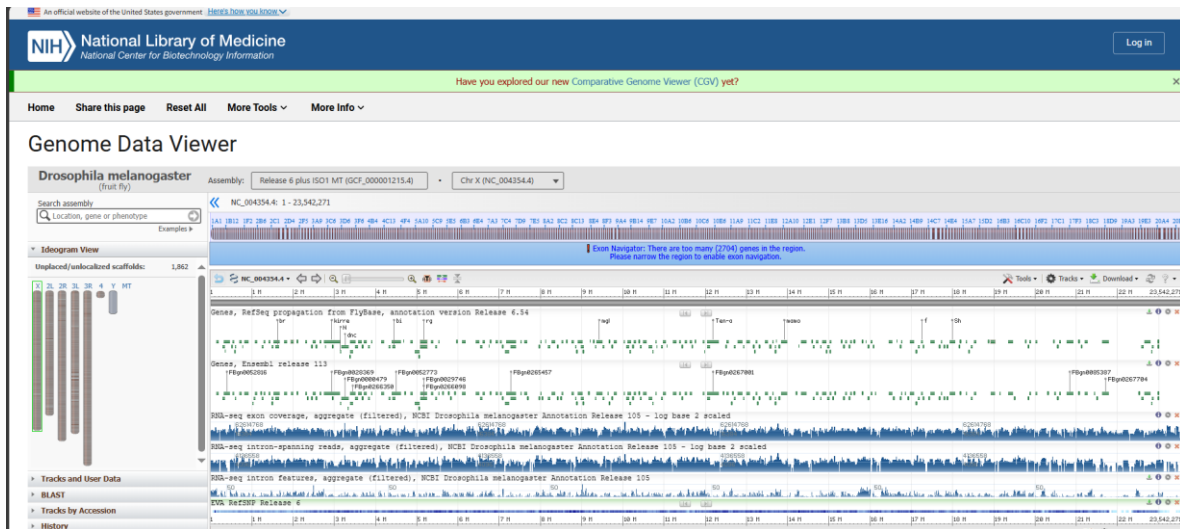


Figura 3 Visualización cromosómica de *Drosophila melanogaster* en el GDV mostrando la distribución de genes

3.1.2 Análisis del genoma de Danio rerio (Pez cebra)

El pez cebra representa otro organismo modelo fundamental en la investigación biomédica contemporánea, con aplicaciones particulares en:

- Biología del desarrollo y embriogénesis
- Modelado de enfermedades humanas
- Estudios de toxicología y farmacología
- Regeneración tisular y medicina regenerativa

Proceso de búsqueda y selección

Para acceder al genoma de referencia de Danio rerio, se siguió un procedimiento similar al anterior:

1. Se regresó a la página principal del GDV utilizando el botón de navegación
2. Se introdujo "Danio rerio" en el campo de búsqueda
3. De los resultados obtenidos, se seleccionó el ensamblaje más reciente del genoma
4. Se accedió a la vista detallada de la información genómica

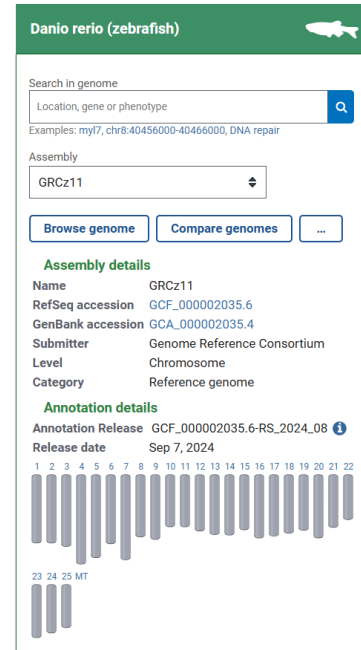
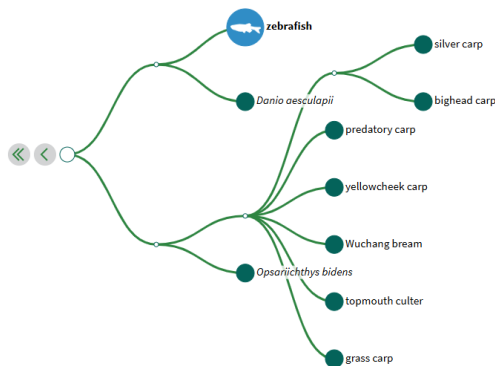


Figura 4 Imagen de los resultados de búsqueda para "Danio rerio" en el GDV

Información obtenida del genoma

Al explorar la información disponible para el genoma de referencia de Danio rerio, se identificaron los siguientes datos clave:

- **Entidad responsable del ensamblaje:** Genome Reference Consortium (GRC)
- **Fecha de publicación:** 7 de septiembre del 2024
- **Numero de Cromosomas:** 25 pares de cromosomas

La participación del Genome Reference Consortium como submitter indica el carácter colaborativo internacional del proyecto de secuenciación y anotación del genoma del pez cebra.

Esta visualización permite apreciar las similitudes y diferencias en la organización genómica entre un invertebrado (*Drosophila*) y un vertebrado (*Danio*).

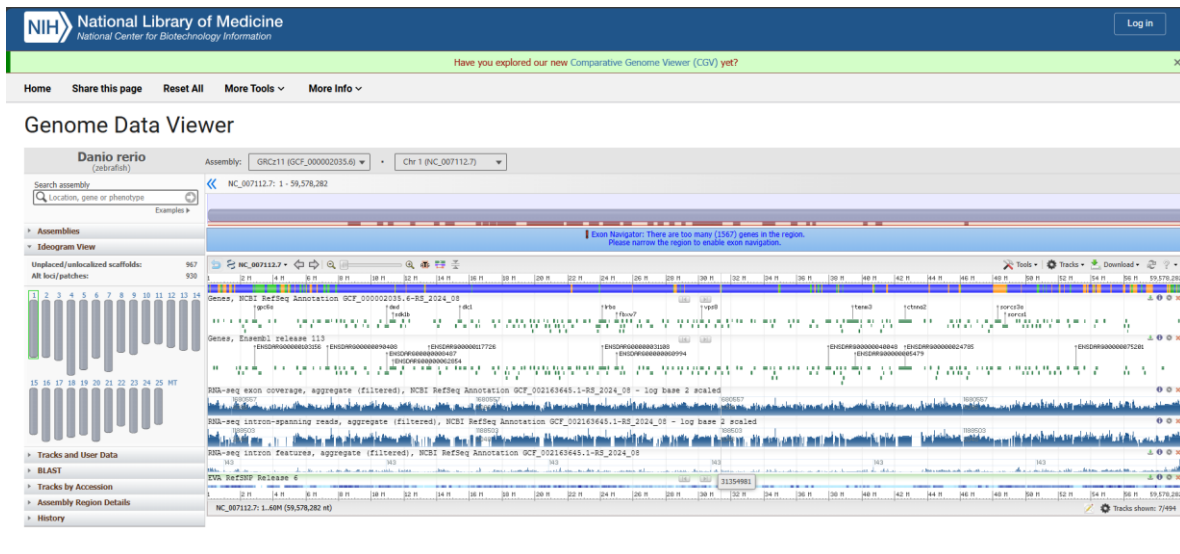


Figura 5 Visualización cromosómica de *Danio rerio* en el GDV mostrando la distribución de genes

3.2 Análisis de SNPs mediante SNPedia

Los polimorfismos de nucleótido único (SNPs) representan la forma más común de variación genética en el genoma humano. Estas variaciones de un solo nucleótido ocurren en promedio cada 300 bases a lo largo del genoma humano, resultando en millones de posiciones variables que contribuyen a la diversidad genética entre individuos. Para explorar y analizar el impacto de estos SNPs en fenotipos humanos, la plataforma SNPedia ofrece un recurso valioso tipo wiki que recopila información científica actualizada sobre variantes genéticas específicas.

En esta tarea, se realizó un análisis detallado de cinco SNPs relevantes utilizando la base de datos SNPedia, con el objetivo de documentar su ubicación cromosómica, efectos fenotípicos asociados e incidencia estimada en poblaciones humanas.

3.2.1 Navegación y utilización de la plataforma SNPedia

SNPedia (<https://www.snpedia.com/>) es una enciclopedia wiki colaborativa que recopila información sobre variantes genéticas humanas y sus asociaciones con fenotipos y condiciones específicas. A diferencia de bases de datos institucionales como dbSNP del NCBI, SNPedia incorpora un enfoque más accesible y orientado a la interpretación clínica y funcional de las variantes.

Para iniciar el análisis, se accedió a la página principal de SNPedia y se familiarizó con las principales características de la plataforma:

- Barra de búsqueda para introducir identificadores rs (reference SNP ID)
- Categorización por temas, enfermedades y rasgos
- Índices alfabéticos de variantes

- Referencias a la literatura científica asociada a cada SNP

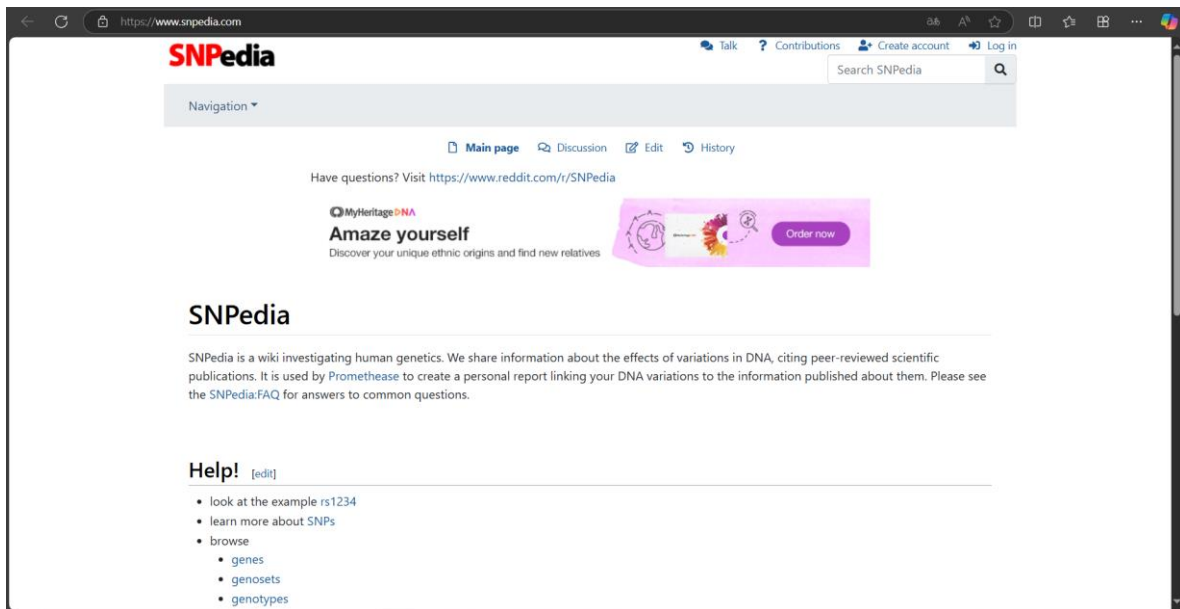


Figura 6 Página principal de SNpedia

3.2.2 Análisis del SNP rs1815739 (ACTN3) - "El gen del sprint"

El gen ACTN3 codifica para la proteína α -actinina-3, un componente estructural clave en las fibras musculares de contracción rápida (tipo II). Esta proteína está implicada en:

- Estabilización del aparato contráctil muscular
- Anclaje de filamentos delgados (actina) en la línea Z del sarcómero
- Regulación de la contracción muscular durante actividades de alta potencia y velocidad

rs1815739

This SNP, in the ACTN3 gene, encodes a premature stop codon in a muscle protein called alpha-actinin-3. The polymorphism alters position 577 of the alpha-actinin-3 protein. In publications the (C,C) genotype is often called RR, whereas the (T,T) genotype is often called XX.

According to [PMID 18043716] (T,T) is under-represented in elite strength athletes, consistent with previous reports indicating that alpha-actinin-3 deficiency appears to impair muscle performance. However in 2016 [PMID 26824906] failed to replicate casting some doubt.

The most common nucleotide at this position, (C), encodes an arginine (amino acid code R), the alternative T allele encodes a stop codon (X). Hence, the SNP is referred to as R577X, with homozygotes being either RR or XX and heterozygotes being RX. XX individuals completely lack the expression of alpha-actinin-3.

One of the earliest report studying a relatively small number of Australian elite (i.e. ~Olympic) athletes found that, at least in females, the R allele (ie **rs1815739**(C)) is associated with sprinters, while the X allele (**rs1815739**(T)) is associated with endurance athletes. No female or Olympic-level sprinters were XX homozygotes (**rs1815739**(T,T)). The association tended the same way but was statistically weaker in males. [PMID 12879365]. There have been several subsequent studies, but few with large sample sizes and thus few with much statistical power. An example of a typical study: no increase in endurance ability was associated with the X allele in elite male cyclists. [PMID 16612741]

muscle performance			
Orientation	plus		
Stabilized	plus		
Geno	Mag	Summary	
(C,C)	2.2	Better performing muscles. Likely sprinter.	
(C,T)	2.1	Mix of muscle types. Likely sprinter.	
(T,T)	2.2	Impaired muscle performance. Likely endurance athlete.	
Reference	GRCh38 38.1/141		
Chromosome	11		
Position	66560624		
Gene	ACTN3		
is a	snp		
is mentioned by			
dbSNP	rs1815739		

Figura 7 Imagen de la página de resultados para rs1815739 en SNpedia

Información obtenida

Del análisis de la página de rs1815739 en SNPedia, se extrajeron los siguientes datos clave:

- **Cromosoma:** 11
- **Efecto del SNP en el fenotipo:**
 - La variante R577X (C→T) resulta en una proteína truncada no funcional
 - Individuos homocigotos XX (TT) carecen completamente de α -actinina-3 en fibras musculares tipo II
 - Asociación con rendimiento deportivo: el genotipo RR (CC) favorece actividades de potencia/velocidad, mientras que XX (TT) se asocia con mejor rendimiento en resistencia
 - No se asocia con patologías musculares o enfermedades
- **Incidencia estimada:**

Esta información revela la frecuencia del alelo T entre poblaciones, siendo notablemente menos común en poblaciones africanas (17.65%) en comparación con poblaciones europeas (43.73%) y asiáticas (42.77%).

Population	Group	Sample Size	Ref Allele	Alt Allele	Ref HMOZ	Alt HMOZ	HTRZ	HWE P
Total	Global	306864	C=0.569083	T=0.430917	0.329097	0.190931	0.479972	32
European	Sub	252096	C=0.562746	T=0.437254	0.316546	0.191054	0.4924	0
African	Sub	15754	C=0.82347	T=0.17653	0.680716	0.033769	0.285515	2
African Others	Sub	542	C=0.895	T=0.105	0.811808	0.02214	0.166052	2
African American	Sub	15212	C=0.82093	T=0.17907	0.676045	0.034184	0.289771	1
Asian	Sub	3774	C=0.5641	T=0.4359	0.318495	0.190249	0.491256	0
East Asian	Sub	2406	C=0.5723	T=0.4277	0.328346	0.183707	0.487947	0
Other Asian	Sub	1368	C=0.5497	T=0.4503	0.30117	0.201754	0.497076	0
Latin American 1	Sub	1160	C=0.6431	T=0.3569	0.405172	0.118966	0.475862	0
Latin American 2	Sub	6902	C=0.3479	T=0.6521	0.125761	0.43002	0.444219	1
South Asian	Sub	500	C=0.4000	T=0.6000	0.100000	0.300000	0.400000	0

Figura 8 Frecuencia del alelo ALFA de rs1815739

3.2.3 Análisis del SNP rs53576 (OXTR) - "El gen de la empatía"

El gen OXTR codifica para el receptor de oxitocina, un receptor acoplado a proteína G que media las acciones de la hormona oxitocina. Este sistema está involucrado en:

- Regulación del comportamiento social y afiliativo

- Procesos de vinculación emocional y empatía
- Respuestas al estrés y ansiedad social
- Comportamiento parental y cuidado de la descendencia

rs53576

rs53576 is a silent G to A change in the oxytocin receptor (OXTR) gene. Studies have demonstrated that individuals with the G allele are more empathetic, feel less lonely, employ more sensitive parenting techniques, and have lower rates of autism (discussed in [PMID 20724662]).

The blog *Not Exactly Rocket Science* discusses that Americans with (G,G) tend to be more sensitive parents, more empathetic and less lonely than those with an 'A'. In a Korean population people with (G,G) were less likely to seek support from their peers.

[PMID 20724662] Culture, distress, and oxytocin receptor polymorphism (OXTR) interact to influence emotional support seeking.

[PMID 19934046] Oxytocin receptor genetic variation relates to empathy and stress reactivity in humans. In brief, people with the G/G genotype were better able to discern the emotional state of others than those who carried the A-allele (blog summary found here). The study encompassed 192 participants of both sexes and mixed ethnicities. The study subjects underwent a number of tests to determine their level of empathy and stress reactivity. They found that G/G individuals performed significantly better on the behavioral measure of empathy and were 22.7% less likely to make a mistake on the "Reading the Mind in the Eyes Test" (RMET) (a behavioural measure of empathic accuracy) than A/A/G individuals ($P = 0.005$). Similarly, G/G individuals reported higher levels of dispositional empathy than A/A/G individuals: mean (SE) = 3.69 (0.06) and 3.53 (0.04) for G/G and A/A/G, respectively ($P = 0.025$), and were less affected by stress (as measured by their heart rates): mean (SE) = 72.1 (0.54) and 78.4 (1.19) for G/G and A/A/G, respectively.

influences social behavior and personality			
Orientation	plus		
Stabilized	plus		
Geno	Mag	Summary	
(A,A)	2.8	Lack of empathy?	
(A,G)	2.8	Lack of empathy?	
(G,G)	2.5	Optimistic and empathetic; handle stress well	
Reference	GRCh38 38.1/141		
Chromosome	3		
Position	8762685		
Gene	OXTR		
is a	snp		
is	mentioned by		
dbSNP	rs53576		
dbSNP (classic)	rs53576		
ClinGen	rs53576		
ebi	rs53576		

Figura 9 Imagen de la página de resultados para rs53576 en SNPedia

Información obtenida

Del análisis detallado de rs53576 en SNPedia, se extrajeron los siguientes datos:

- **Cromosoma:** 3
- **Efecto del SNP en el fenotipo:**
 - Polimorfismo A/G en el tercer intrón del gen OXTR
 - Individuos con genotipo GG muestran:
 - Mayor sensibilidad emocional y empatía
 - Mejor reconocimiento de expresiones faciales emocionales
 - Menor reactividad al estrés social
 - Mayor capacidad de vinculación social
 - Individuos con uno o dos alelos A (AA o AG) muestran tendencias hacia:
 - Menor optimismo y autoestima
 - Mayor sensibilidad al estrés social
 - Procesamiento emocional menos eficiente
- **Incidencia estimada:**

Esta información revela la frecuencia del alelo A entre poblaciones, siendo notablemente menos común en poblaciones africanas (21.4 %) en comparación con poblaciones del este de Asia (58.1%).

Population	Group	Sample Size	Ref Allele	Alt Allele	Ref HMOZ	Alt HMOZ	HTRZ	HWE P
Total	Global	74968	A=0.32646	G=0.67354, T=0.00000	0.108126	0.455208	0.436666	1
European	Sub	62252	A=0.32577	G=0.67423, T=0.00000	0.106246	0.4547	0.439054	0
African	Sub	4144	A=0.2379	G=0.7621, T=0.0000	0.06612	0.590251	0.343629	4
African Others	Sub	154	A=0.214	G=0.786, T=0.000	0.051948	0.623377	0.324675	0
African American	Sub	3990	A=0.2388	G=0.7612, T=0.0000	0.066667	0.588972	0.344361	4
Asian	Sub	188	A=0.559	G=0.441, T=0.000	0.297872	0.180851	0.521277	0
East Asian	Sub	148	A=0.581	G=0.419, T=0.000	0.324324	0.162162	0.513514	0
Other Asian	Sub	40	A=0.47	G=0.53, T=0.00	0.2	0.25	0.55	0
Latin American 1	Sub	242	A=0.256	G=0.744, T=0.000	0.090909	0.578512	0.330579	1
Latin American 2	Sub	1236	A=0.3657	G=0.6343, T=0.0000	0.131068	0.399676	0.469256	0

Figura 10 Frecuencia del alelo ALFA de rs53576

3.2.4 Análisis del SNP rs4680 (COMT) - "Polimorfismo de la cognición"

El gen COMT codifica para la enzima catecol-O-metiltransferasa, responsable de la degradación de catecolaminas como la dopamina, epinefrina y norepinefrina. Su función es particularmente importante en:

- Regulación de los niveles de dopamina en la corteza prefrontal
- Modulación de procesos cognitivos como memoria de trabajo y función ejecutiva
- Procesamiento de estímulos dolorosos
- Respuesta al estrés y regulación emocional

rs4680

rs4680 (Val158Met) is a well studied SNP in the **COMT** gene. The **COMT** gene codes for the **COMT** enzyme, which breaks down dopamine in the brain's prefrontal cortex. The wild-type allele is a (G), coding for a valine amino acid; the (A) substitution polymorphism changes the amino acid to a methionine. This alters the structure of the resultant enzyme such that its activity is only 25% of the wild type. As a result, A allele carriers have more dopamine in their prefrontal cortex, which may be responsible for many of the neuropsychological associations listed below.

23andMe blog summarizes the alleles at this SNP as

- **rs4680(A)** = Worrier. Met, more exploratory, lower COMT enzymatic activity, therefore higher dopamine levels; lower pain threshold, enhanced vulnerability to stress, yet also more efficient at processing information under most conditions
- **rs4680(G)** = Warrior. Val, less exploratory, higher COMT enzymatic activity, therefore lower dopamine levels; higher pain threshold, better stress resiliency, albeit with a modest reduction in executive cognition performance under most conditions

Roughly speaking, the predominant wisdom (known colloquially as the warrior/worrier hypothesis; summary at [PMID 17008817]) posits that people with Val alleles have increased COMT activity and lower prefrontal extracellular dopamine compared with those with the Met substitution. Val158 alleles may be associated with an advantage in the processing of aversive stimuli (warrior strategy), while Met158 alleles may be associated with an advantage in memory and attention tasks (worrier strategy). Under conditions of increased dopamine release (eg, stress), individuals with Val158 alleles may have improved dopaminergic transmission and better performance, while individuals with Met158 alleles may have less efficient neurotransmission and worse performance. Some evidence suggests that Val158 alleles are associated with schizophrenia, while Met158 alleles are associated with anxiety.

warrior vs worrier; number of other associations

Orientation plus

Stabilized plus

Geno • Mag • Summary

(A/A)

2.5

(warrior) advantage in memory and attention tasks

(A/G)

Intermediate dopamine levels, other effects

(G/G)

2.5

(warrior) multiple associations, see details

Reference GRCh38 38.1/141

Chromosome22

Position 19963748

Gene COMT, MIR4761

is a snp

is mentioned by

dbSNP rs4680

dbSNP rs4680 (classic)

ClinGen rs4680

...4680

Figura 11 Imagen de la página de resultados para rs4680 en SNPedia

Información obtenida

Del análisis de rs4680 en SNPedia, se obtuvieron los siguientes datos:

- **Cromosoma:** 22
- **Efecto del SNP en el fenotipo:**
 - Polimorfismo Val158Met (G→A) que resulta en un cambio de valina a metionina en la posición 158
 - Impacto en la actividad enzimática:
 - Genotipo GG (Val/Val): Alta actividad enzimática, degradación rápida de dopamina
 - Genotipo AA (Met/Met): Actividad enzimática reducida (3-4 veces menor), niveles más altos de dopamina prefrontal
 - Efectos cognitivos y psicológicos:
 - GG (Val/Val): Mejor rendimiento bajo estrés, mayor resiliencia, pero menor eficiencia en tareas cognitivas en condiciones normales
 - AA (Met/Met): Mejor memoria de trabajo y función ejecutiva en condiciones normales, mayor sensibilidad al estrés
 - Impacto en la percepción del dolor:
 - AA (Met/Met): Mayor sensibilidad al dolor y respuesta emocional al dolor
- **Incidencia estimada:**

Esta información revela la frecuencia del alelo G entre poblaciones, siendo notablemente menos común en poblaciones europeas (49.1334%) en comparación con poblaciones de Asia (75.57%).

Population	Group	Sample Size	Ref Allele	Alt Allele	Ref HMOZ	Alt HMOZ	HTRZ	HWE P
Total	Global	298100	G=0.512929	A=0.487071	0.268178	0.242321	0.4895	33
European	Sub	242776	G=0.491334	A=0.508666	0.243459	0.260792	0.495749	5
African	Sub	16330	G=0.69571	A=0.30429	0.481568	0.090141	0.428291	1
African Others	Sub	554	G=0.722	A=0.278	0.509025	0.064982	0.425993	1
African American	Sub	15776	G=0.69479	A=0.30521	0.480603	0.091024	0.428372	0
Asian	Sub	3820	G=0.7351	A=0.2649	0.536649	0.066492	0.396859	0
East Asian	Sub	2420	G=0.7231	A=0.2769	0.517355	0.071074	0.41157	1
Other Asian	Sub	1400	G=0.7557	A=0.2443	0.57	0.058571	0.371429	0
Latin American 1	Sub	976	G=0.571	A=0.429	0.315574	0.17418	0.510246	1
Latin American 2	Sub	8888	G=0.5842	A=0.4158	0.343384	0.175068	0.481548	0

Figura 12 Frecuencia del alelo ALFA de rs4680

3.2.5 Análisis del SNP rs1799971 (OPRM1) - "Polimorfismo del receptor opioide"

El gen OPRM1 codifica para el receptor μ -opioide, principal sitio de acción de opioides endógenos y exógenos. Este receptor está involucrado en:

- Modulación de la percepción del dolor
- Mediación de efectos analgésicos de opioides
- Sistemas de recompensa cerebral
- Susceptibilidad a dependencia de sustancias

rs1799971

The **rs1799971**(G) allele in exon 1 of the mu opioid receptor **OPRM1** gene causes the normal amino acid at residue 40, asparagine (Asn), to be replaced by aspartic acid (Asp). In the literature this SNP is also known as A118G, N40D, or Asn40Asp.

Carriers of at least one **rs1799971**(G) allele appear to have stronger cravings for alcohol than carriers of two **rs1799971**(A) alleles, and are thus hypothesized to be more at higher risk for [alcoholism](#). [PMID 17207095]

However, subsequent research results are mixed, and there are other studies both agreeing or disagreeing with this finding [PMID 15525999, PMID 9399694, PMID 12960749]

Among 200+ alcoholics treated with [naltrexone](#), **rs1799971**(G) carriers receiving the drug (even without behavioral intervention) had an increased percentage of days abstinent ($p = .07$) and a decreased percentage of heavy drinking days ($p = .04$) if treated with naltrexone vs. placebo, whereas **rs1799971**(A;A) homozygotes showed no medication differences. Upon treatment with [naltrexone](#), 87% of **rs1799971**(G) carriers had a good clinical outcome, compared with only 55% of individuals with the (A;A) genotype (odds ratio, 5.75, CI: 1.88-17.54)[PMID 18250251]

This SNP may also influence the response to opioids such as [heroin](#), [codeine](#) or [morphine](#). A 2015 meta-analysis (totaling 5,902 patients) concluded that the carriers of a **rs1799971**(G) allele consumed more opioids for analgesia (SMD = -0.17, CI: -0.25, -0.10, $p < 0.00001$), but still reported higher pain scores ($p = 0.002$) and less nausea and vomiting (odds ratio 1.30, CI: 1.08-1.55, $p = 0.005$) than homozygous (A;A) patients during the first 24 hour, but not 48 hour, postoperative period.[PMID 25794200]

Orientation	plus		
Stabilized	plus		
Geno	Mag	Summary	
(A;A)	0	normal	
(A;G)	2.5	stronger cravings for alcohol. if alcoholic, naltrexone treatment 2x more successful	
(G;G)	2.6	more pain	
Reference	GRCh38 38.1/141		
Chromosome	6		
Position	154039662		
Gene	OPRM1		
is a	snp		
is	mentioned by		
dbSNP	rs1799971		
dbSNP (classic)	rs1799971		
ClinGen	rs1799971		
ebi	rs1799971		
HLI	rs1799971		
Ensembl	rs1799971		

Figura 13 Imagen de la página de resultados para rs1799971 en SNPedia

- **Información obtenida**

Del análisis de rs1799971 en SNPedia, se extrajeron los siguientes datos:

- **Cromosoma:** 6
- **Efecto del SNP en el fenotipo:**
 - Polimorfismo A118G que resulta en el cambio Asn40Asp en el receptor
 - El alelo G (Asp40) se asocia con:
 - Menor expresión del receptor μ -opioide
 - Reducción en la eficacia analgésica de opioides
 - Mayor dosis requerida para manejo del dolor
 - Posible protección contra dependencia a opioides (aunque los estudios muestran resultados contradictorios)
 - Diferencias en la respuesta al alcohol y otras sustancias
- **Incidencia estimada:**

Esta información revela la frecuencia del alelo A entre poblaciones, siendo notablemente menos común en poblaciones africanas (99.8%) en comparación con poblaciones del este de Asia (63.14%).

Population	Group	Sample Size	Ref Allele	Alt Allele	Ref HMOZ	Alt HMOZ	HTRZ	HWEP
Total	Global	387988	A=0.861204	G=0.138796	0.745111	0.022702	0.232188	32
European	Sub	329210	A=0.867641	G=0.132359	0.753841	0.01856	0.227599	8
African	Sub	13908	A=0.96657	G=0.03343	0.935145	0.002013	0.062842	3
African Others	Sub	510	A=0.998	G=0.002	0.996078	0.0	0.003922	0
African American	Sub	13398	A=0.96537	G=0.03463	0.932826	0.00209	0.065084	3
Asian	Sub	6974	A=0.6382	G=0.3618	0.406367	0.129911	0.463722	0
East Asian	Sub	5006	A=0.6314	G=0.3686	0.395525	0.132641	0.471834	0
Other Asian	Sub	1968	A=0.6555	G=0.3445	0.433943	0.122967	0.443089	0
Latin American 1	Sub	1486	A=0.8890	G=0.1110	0.795424	0.017497	0.187079	1
Latin American 2	Sub	7244	A=0.8049	G=0.1951	0.646052	0.036168	0.31778	0
South Asian	Sub	5336	A=0.6571	G=0.3429	0.443193	0.138205	0.438203	4

Figura 14 Frecuencia del alelo ALFA de rs1799971

3.2.6 Análisis del SNP rs9939609 (FTO) - "Gen asociado a obesidad"

El gen FTO (Fat mass and obesity-associated protein) codifica para una demetilasa de ácidos nucleicos que regula procesos epigenéticos. Este gen está involucrado en:

- Regulación del balance energético

- Control del apetito y saciedad
- Metabolismo lipídico
- Desarrollo del tejido adiposo

rs9939609

rs9939609 is a SNP in the fat mass and obesity associated *FTO* gene, aka the "Fat Gene" [1]. The original paper describing it is here [PMID 17434869].

2015 research suggests rs1421085 may be the causal snp.

[2] The researchers identified 10 different *FTO* SNPs in the first intron of the gene that associated with both BMI and type-2 diabetes. Because they all correlated with each other, they chose to examine one of the SNPs (**rs9939609**) in detail.

rs9939609 has also been reported in a large study to be associated with type-2 diabetes. The risk allele (oriented to the dbSNP entry) is (A); the odds ratio associated with heterozygotes is 1.34 (CI 1.17-1.52), and for homozygotes, 1.55 (CI 1.3-1.84). [PMID 17554300]

In a different study, this SNP was found in a linkage block in the *FTO* gene with rs1121980; see rs1121980 for the association of this block with early onset obesity since it showed the strongest association. [PMID 18159244]

The increases in body mass index associated with **rs9939609**(A) carriers appears to begin at a young age and are maintained throughout adulthood, according to a study of 5,600+ Utah families. [PMID 18239580]

This SNP did not show any association with obesity, type-2 diabetes or any other related traits in a study of 3,210 Chinese subjects. Furthermore, the minor allele frequency was much lower in Chinese populations compared to populations of European descent. [PMID 17959933]

Obesity related. Raises T2D risk through obesity.		
Orientation	plus	
Stabilized	plus	
Geno	Mag	Summary
(A;A)	3	obesity risk and 1.6x risk for T2D
(A;T)	2.4	1.3x risk for T2D; obesity risk
(T;T)	1.5	lower risk of obesity and Type-2 diabetes
Reference	GRCh38 38.1/141	
Chromosome	16	
Position	53786615	
Gene	FTO	
is a	snp	
is	mentioned by	
dbSNP	rs9939609	
dbSNP (classic)	rs9939609	

Figura 15 Imagen de la página de resultados para rs9939609 en SNPedia

Información obtenida

Del análisis de rs9939609 en SNPedia, se obtuvieron los siguientes datos:

- **Cromosoma:** 16
- **Efecto del SNP en el fenotipo:**
 - Polimorfismo T→A en el primer intrón del gen *FTO*
 - Cada copia del alelo A se asocia con:
 - Aumento promedio de 1.5 kg en el peso corporal
 - Incremento del 20-30% en el riesgo de obesidad
 - Mayor índice de masa corporal (IMC)
 - Preferencia por alimentos calóricamente densos
 - El genotipo AA se asocia con:
 - Mayor dificultad para sensación de saciedad
 - Mayor ingesta calórica
 - Tendencia a comportamientos alimentarios impulsivos
- **Incidencia estimada:**

Esta información revela la frecuencia del alelo T entre poblaciones, siendo notablemente menos común en poblaciones africanas (45.2%) en comparación con poblaciones del este de Asia (87.54%).

Population	Group	Sample Size	Ref Allele	Alt Allele	Ref HMOZ	Alt HMOZ	HTRZ	HWE P
Total	Global	84190	T=0.60143	A=0.39857	0.368072	0.165222	0.466706	16
European	Sub	73860	T=0.58975	A=0.41025	0.351313	0.171812	0.476875	5
African	Sub	3744	T=0.5256	A=0.4744	0.264957	0.213675	0.521368	3
African Others	Sub	166	T=0.452	A=0.548	0.204819	0.301205	0.493976	0
African American	Sub	3578	T=0.5291	A=0.4709	0.267747	0.209614	0.522638	3
Asian	Sub	3274	T=0.8699	A=0.1301	0.753818	0.01405	0.232132	1
East Asian	Sub	2648	T=0.8754	A=0.1246	0.765106	0.01435	0.220544	0
Other Asian	Sub	626	T=0.847	A=0.153	0.70607	0.01278	0.28115	1
Latin American 1	Sub	436	T=0.573	A=0.427	0.321101	0.174312	0.504587	0
Latin American 2	Sub	928	T=0.747	A=0.253	0.575431	0.081897	0.342672	3
South Asian	Sub	270	T=0.713	A=0.287	0.538413	0.183841	0.267647	1

Figura 16 Frecuencia del alelo ALFA de rs9939609

3.3 Tarea 3: Identificación de secuencias diana para CRISPR mediante CHOPCHOP

En esta tarea se utilizó la herramienta bioinformática CHOPCHOP para identificar secuencias diana óptimas para la edición del gen FTO (Fat mass and obesity-associated protein) mediante la tecnología CRISPR-Cas9. Este gen está asociado con el SNP rs9939609, el cual ha sido vinculado con mayor predisposición a la obesidad.

3.3.1 Procedimiento realizado

Para la identificación de secuencias diana en el gen FTO, se siguieron estos pasos:

1. Se accedió a la plataforma CHOPCHOP (<https://chopchop.cbu.uib.no>)
2. Se seleccionó el organismo Homo sapiens (GRCh38/hg38)
3. Se configuró el modo de búsqueda para CRISPR/Cas9
4. Se introdujo el nombre del gen FTO

Figura 17 Imagen de la interfaz de CHOPCHOP con los parámetros configurados para el gen FTO

3.3.2 Resultados obtenidos

Tras el análisis, CHOPCHOP generó una lista de 10 secuencias diana potenciales. Los resultados se presentan ordenados según la eficiencia predicha por el algoritmo:

Rank	Target sequence	Genomic location	Strand	GC content (%)	Self-complementarity	P00	P01	P02	P03	Efficiency
1	CGGTGGGTGGAACAAACGAGG	chr16:53873833	-	60	0	0	0	0	2	80.81
2	ACTAAACCGAGGTGTGAACCG	chr16:53873822	-	45	2	0	0	0	2	55.37
3	GTTTAGTTCACCCACCGAGTG	chr16:53873838	+	55	0	0	0	0	3	67.51
4	CACTTCATCTTGTCCGTTGTAGG	chr16:53826291	-	45	3	0	0	1	2	33.81
5	ACTTACTCTGCCACTCGGTGG	chr16:53873849	-	55	1	0	0	0	4	62.24
6	TTTTGGCCGGTTCACAACCTCG	chr16:53873816	+	50	1	0	0	0	4	53.89
7	GTGTCAACCATGGCTCAACTGG	chr16:53888900	+	55	0	0	0	0	4	36.00
8	CTACAACGGACAAGATGAAGTG	chr16:53826292	+	45	3	0	0	0	5	59.16
9	GGATGAACCATCCCAACCTCG	chr16:53826270	-	60	0	0	0	1	4	55.99
10	TACCTCTGCCACTCGGTGGTG	chr16:53873846	-	65	2	0	0	1	4	54.59

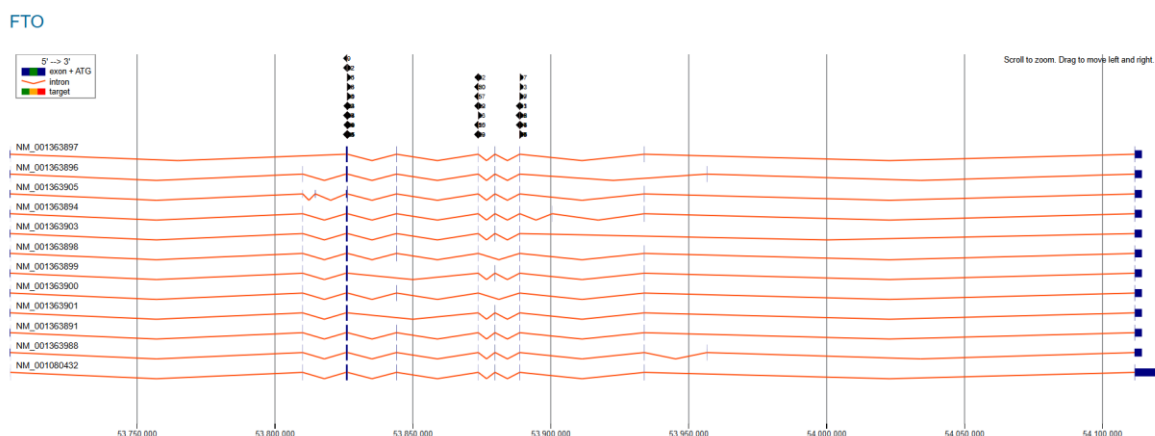


Figura 18 Resultados obtenidos

Es importante destacar que, aunque se ha identificado una secuencia diana óptima desde el punto de vista técnico, la aplicación terapéutica de la edición genómica

para FTO requeriría consideraciones éticas, regulatorias y de seguridad adicionales antes de cualquier implementación clínica.

3.4 Desafíos y soluciones

Durante el desarrollo de la práctica se enfrentaron diversos desafíos técnicos y conceptuales. En la exploración del Genome Data Viewer, la principal dificultad fue la navegación a través de la compleja interfaz y la identificación precisa de la información requerida para los genomas de *Drosophila melanogaster* y *Danio rerio*, lo cual se solucionó mediante una exploración sistemática de las secciones de metadatos y el uso de filtros específicos. Al trabajar con SNPedia, el desafío principal consistió en interpretar correctamente la terminología especializada y correlacionar las variantes alélicas con sus respectivos efectos fenotípicos, lo que se resolvió consultando la literatura científica complementaria citada en las páginas de cada SNP. Finalmente, en la identificación de secuencias diana para CRISPR mediante CHOPCHOP, las limitaciones incluyeron la evaluación comparativa de múltiples parámetros simultáneamente (eficiencia, especificidad, ubicación) para seleccionar la secuencia óptima, lo que se abordó estableciendo una jerarquía clara de criterios con énfasis en la relevancia terapéutica y la minimización de efectos off-target.

4 Conclusiones y recomendaciones.

La práctica de laboratorio realizada ha permitido explorar tres herramientas bioinformáticas fundamentales para el análisis genómico, mostrando interfaces más accesibles que facilitan la manipulación y análisis de datos genómicos.

A través del Genome Data Viewer, se ha observado la enorme diversidad estructural entre genomas de diferentes organismos modelo. La comparación entre la mosca de la fruta (*Drosophila melanogaster*) y el pez cebra (*Danio rerio*) revela no solo diferencias en la organización y complejidad genómica, reflejando la distancia evolutiva entre invertebrados y vertebrados. Resulta evidente que, a pesar de estas diferencias, ambos organismos mantienen su relevancia como modelos experimentales para la investigación biomédica debido a la conservación de genes y vías fundamentales a lo largo de la evolución.

El análisis de SNPs mediante SNPedia ha permitido apreciar la complejidad de la variabilidad genética humana y su impacto en diversos fenotipos. Los cinco polimorfismos estudiados muestran cómo variaciones puntuales en el genoma pueden asociarse con diferencias significativas en características tan diversas como rendimiento muscular, comportamiento social, cognición, percepción del dolor y metabolismo.

La identificación de secuencias diana para CRISPR en el gen CFTR demuestra el potencial revolucionario de las tecnologías de edición genómica para abordar enfermedades monogénicas como la fibrosis quística. La precisión con que CHOPCHOP permitió identificar secuencias diana específicas, evaluando simultáneamente su eficiencia y posibles efectos evidenciando la madurez que están alcanzando las herramientas bioinformáticas en este campo.

En base a la experiencia adquirida durante esta práctica, se pueden formular las recomendaciones para futuras investigaciones y aplicaciones en el campo de la genómica como complementar análisis de contextualización poblacional. Considerar siempre la variabilidad poblacional al interpretar datos de SNPs, evitando generalizaciones excesivas basadas en estudios realizados principalmente en poblaciones europeas.

Desde una perspectiva integradora, esta práctica ha demostrado cómo la bioinformática se ha convertido en el puente indispensable entre los datos genómicos crudos y su aplicación biomédica, facilitando la transición desde la secuenciación masiva hasta intervenciones terapéuticas personalizadas. La democratización de estas herramientas, ahora accesibles a través de interfaces web, está acelerando el ritmo de la investigación genómica y expandiendo su alcance más allá de laboratorios especializados.

5 Referencias

- [1].Buniello, A., MacArthur, J. A. L., Cerezo, M., Harris, L. W., Hayhurst, J., Malangone, C., McMahon, A., Morales, J., Mountjoy, E., Sollis, E., Suveges, D., Vrousitou, O., Whetzel, P. L., Amode, R., Guillen, J. A., Riat, H. S., Trevanion, S. J., Hall, P., Junkins, H., ... Parkinson, H. (2019). The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Research*, 47(D1), D1005–D1012. <https://doi.org/10.1093/nar/gky1120>
- [2].Cokus, S. J., Rose, C. J., Haynor, D. R., Gronbech-Jensen, N., & Pellegrini, M. (2023). CRISPR Guide RNA Design: Software, Data, and Analytics. *Annual Review of Biomedical Data Science*, 6, 79-103. <https://doi.org/10.1146/annurev-biodatasci-110921-083119>
- [3].Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P. D., Wu, X., Jiang, W., Marraffini, L. A., & Zhang, F. (2013). Multiplex genome engineering using CRISPR/Cas systems. *Science*, 339(6121), 819–823. <https://doi.org/10.1126/science.1231143>
- [4].Howe, K., Clark, M. D., Torroja, C. F., Torrance, J., Berthelot, C., Muffato, M., Collins, J. E., Humphray, S., McLaren, K., Matthews, L., McLaren, S., Sealy, I., Caccamo, M., Churcher, C., Scott, C., Barrett, J. C., Koch, R., Rauch, G. J., White, S., ... Stemple, D. L. (2013). The zebrafish reference genome sequence and its relationship to the human genome. *Nature*, 496(7446), 498–503. <https://doi.org/10.1038/nature12111>
- [5].Labun, K., Montague, T. G., Krause, M., Torres Cleuren, Y. N., Tjeldnes, H., & Valen, E. (2019). CHOPCHOP v3: expanding the CRISPR web toolbox beyond genome editing. *Nucleic Acids Research*, 47(W1), W171–W174. <https://doi.org/10.1093/nar/gkz365>
- [6].Montague, T. G., Cruz, J. M., Gagnon, J. A., Church, G. M., & Valen, E. (2014). CHOPCHOP: a CRISPR/Cas9 and TALEN web tool for genome editing. *Nucleic Acids Research*, 42(Web Server issue), W401–W407. <https://doi.org/10.1093/nar/gku410>
- [7].Ramos, E. M., Din-Lovinescu, C., Berg, J. S., Brooks, L. D., Duncanson, A., Dunn, M., Good, P., Hubbard, T. J., Jarvik, G. P., O'Donnell, C., Sherry, S. T., Aronson, N., Biesecker, L. G., Blumberg, B., Calonge, N., Colhoun, H. M., Epstein, R. S., Flicek, P., Gordon, E. S., ... Williams, M. S. (2012). Characterizing genetic variants for clinical action. *American Journal of Medical Genetics. Part C, Seminars in Medical Genetics*, 160C(1), 93–104. <https://doi.org/10.1002/ajmg.c.31386>
- [8].Ranz, J. M., & Casals, F. (2020). The *Drosophila melanogaster* genome. In *eLS* (pp. 1-9). John Wiley & Sons, Ltd. <https://doi.org/10.1002/9780470015902.a0001072.pub3>
- [9].Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W. W., Hegde, M., Lyon, E., Spector, E., Voelkerding, K., Rehm, H. L., &

- ACMG Laboratory Quality Assurance Committee (2015). Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genetics in Medicine*, 17(5), 405–424. <https://doi.org/10.1038/gim.2015.30>
- [10]. Roberts, R. G. (2014). CRISPR and the Cas9 revolution. *Science*, 345(6197), 612-613. <https://doi.org/10.1126/science.345.6197.612-b>