



UNIVERSIDAD TÉCNICA DE AMBATO

**FACULTAD DE CIENCIA E INGENIERÍA EN
ALIMENTOS Y BIOTECNOLOGÍA**



CARRERA DE BIOTECNOLOGÍA

Tema: Análisis de variantes de splicing y factores de transcripción (HSF1, CBP y Sp1) en la evolución de la enfermedad de Huntington.

Informe Final del Trabajo de Titulación, Opción Proyecto de Investigación, previo la obtención del Título de Ingeniero Biotecnólogo, otorgado por la Universidad Técnica de Ambato, a través de la Facultad de Ciencia e Ingeniería en Alimentos y Biotecnología.

Autor: Christopher Joel Silva Gamboa

Tutor: Mg. Cristian Fernando Galarza Galarza

Ambato – Ecuador

Febrero – 2025

APROBACIÓN DEL TUTOR

Mg. Cristian Fernando Galarza Galarza

CERTIFICA:

Que el presente Informe Final del Trabajo de Titulación ha sido prolijamente revisado. Por lo tanto, autorizo la presentación de este Informe Final del Trabajo de Titulación, Opción Proyecto de Investigación el mismo que responde a las normas establecidas en el reglamento de Títulos y Grados de la Facultad de Ciencia e Ingeniería en Alimentos y Biotecnología.

Ambato, 16 de Enero de 2025



Firmado electrónicamente por:
**CRISTIAN FERNANDO
GALARZA GALARZA**

Mg. Cristian Fernando Galarza Galarza

1803160272

TUTOR

AUTORÍA DEL TRABAJO DE TITULACIÓN

Yo, Christopher Joel Silva Gamboa manifiesto que los resultados obtenidos en el presente Informe Final del Trabajo de Titulación, Opción Proyecto de Investigación, previo a la obtención del título de Ingeniero Biotecnólogo, son absolutamente originales, auténticos y personales; a excepción de las citas bibliográficas.



Christopher Joel Silva Gamboa

1850506104

AUTOR

DERECHOS DE AUTOR

Autorizo a la Universidad Técnica de Ambato, para que haga uso de este Informe Final del Trabajo de Titulación o parte de él, un documento disponible para su lectura consulta y procesos de investigación, según las normas de la Institución.

Cedo los Derechos en línea patrimoniales de mi Informe Final del Trabajo de Titulación, con fines de difusión pública, además apruebo la reproducción de este, dentro de las regulaciones de la Universidad, siempre y cuando esta reproducción no suponga una ganancia económica y se realice respetando los derechos de autor.



Christopher Joel Silva Gamboa

1850506104

AUTOR

APROBACIÓN DEL TRIBUNAL DE GRADO

Los suscritos Profesores Calificadores, aprueban el presente Informe Final del Trabajo de Titulación, Opción Proyecto de Investigación, el mismo que ha sido elaborado de conformidad con las disposiciones emitidas por la Facultad de Ciencia e Ingeniería en Alimentos y Biotecnología de la Universidad Técnica de Ambato.

Para constancia firman:



Firmado electrónicamente por:
**DIEGO MANOLO
SALAZAR GARCÉS**

Presidente del Tribunal



Firmado electrónicamente por:
**MARIA EUGENIA
GARCIA PAZMIÑO**

Mg. María Eugenia García Pazmiño

2100248505



Firmado electrónicamente por:
**GABRIEL ALEJANDRO
MORENO TOASA**

Mg. Gabriel Alejandro Moreno Toasa

1500647910

Ambato, 30 de Enero de 2025

DEDICATORIA

“La Diferencia Entre La Estupidez Y El Genio Es Que El Genio Tiene Sus Límites”

- Albert Einstein

AGRADECIMIENTO

“La Diferencia Entre La Estupidez Y El Genio Es Que El Genio Tiene Sus Límites.”

- Albert Einstein

“El hombre más irremediablemente estúpido es aquel que ignora su sabiduría.”

- Isaac Asimov

ÍNDICE GENERAL DE CONTENIDOS

APROBACIÓN DEL TUTOR.....	ii
AUTORÍA DEL TRABAJO DE TITULACIÓN.....	iii
DERECHOS DE AUTOR	iv
APROBACIÓN DEL TRIBUNAL DE GRADO	v
DEDICATORIA	vi
AGRADECIMIENTO	vii
ÍNDICE GENERAL DE CONTENIDOS	viii
ÍNDICE DE TABLAS.....	x
ÍNDICE DE FIGURAS	xi
RESUMEN EJECUTIVO.....	xii
ABSTRACT.....	xiii
CAPÍTULO I.- MARCO TEÓRICO.....	1
1.1. Antecedentes investigativos	1
1.1.1. Generalidades de enfermedad de Huntington (EH).....	1
1.1.1.1. Bases Genéticas de la EH	2
1.1.2. Splicing alternativo en la enfermedad de Huntington.....	2
1.1.2.1. Posibles consecuencias funcionales de las variantes de splicing en la EH	3
1.1.2.2. Importancia del splicing alternativo en la progresión de la enfermedad.....	3
1.1.3. Factores de transcripción	3
1.1.3.1. HSF1 (Heat Shock Factor 1).....	3
1.1.3.2. CBP (CREB-binding protein)	4
1.1.3.3. Sp1 (Specificity Protein 1).....	4
1.1.4. Interacciones entre el splicing alternativo y los factores de transcripción	4
1.1.4.1. Implicación en la EH	5
1.1.5. Bioinformática para el análisis de variables en la EH	5
1.1.5.1. Tecnología de secuenciación.....	5
1.1.5.2. Análisis de expresión diferencial (DEG)	5
1.1.5.3. Herramientas bioinformáticas para el análisis de splicing alternativo.....	6
1.2. Objetivos	7
1.2.1. Objetivo general	7
1.2.2. Objetivos específicos	7
CAPÍTULO II.- METODOLOGÍA.....	8
2.1. Materiales	8
2.2. Métodos.....	9
2.2.1. Búsqueda de datos de secuenciación RNA-seq de muestras de Homo sapiens con EH... 9	9

2.2.2.	Control de calidad y análisis de expresión diferencial (DEG) de las muestras RNA-seq.	9
2.2.3.	Detección de variantes de splicing y visualización	10
2.2.4.	Construcción y análisis de redes de co-expresión e interacción funcional de los factores de transcripción HSF1, CBP y Sp1	10
2.2.5.	Cuantificación de variantes de Splicing y Factores de transcripción.....	11
2.2.6.	Correlación de las disfunciones genéticas y su contribución en la patogénesis de la EH	11
CAPÍTULO III.- RESULTADOS Y DISCUSIÓN		12
3.1.	Análisis y discusión de resultados	12
3.1.1.	Búsqueda de datos de secuenciación RNA-seq con EH	12
3.1.2.	Control de calidad de las muestras formato FASTQ	12
3.1.3.	Análisis de expresión diferencial.....	16
3.1.4.	Detección de variantes de splicing	18
3.1.5.	Construcción y análisis de redes de co-expresión e interacción funcional de los FT	25
3.1.6.	Interpretación de la influencia de las VS diferenciales	28
3.1.7.	Interpretación de la influencia de los FT	30
CAPÍTULO IV.- CONCLUSIONES Y RECOMENDACIONES		32
4.1.	Conclusiones	32
4.2.	Recomendaciones	33
MATERIALES DE REFERENCIA.....		34
ANEXOS		41

ÍNDICE DE TABLAS

Tabla 1. Herramientas bioinformáticas.....	8
Tabla 2. Recursos materiales.....	8
Tabla 3. Muestras del proyecto GSE124664.....	12
Tabla 4. Calidad de base por secuencia de muestras representativas	13
Tabla 5. Contenido de GC por secuencia de muestras representativas.....	15
Tabla 6. Variantes encontradas en 71 CAG y 109 CAG	19
Tabla 7. Eventos de splicing significativos comparación Control vs 71CAG	20
Tabla 8. Eventos de splicing significativos comparación Control vs 109CAG	21
Tabla 9. Variantes en zonas de interés en las condiciones 71CAG y 109CAG visualizadas en IGV .	23
Tabla 10. Redes globales de co-expresión e interacción funcional	26
Tabla 11. Puntuaciones de los FT y total de conexiones directas de interacción	27
Tabla 12. Nodos principales en la red de CBP	27
Tabla 13. Nodos principales en la red de Sp1	28
Tabla 14. Cuantificación de variantes VCF	29
Tabla 15. Cuantificación de eventos rMATS	29
Tabla 16. Correlación variantes halladas con existentes.....	29
Tabla 17. Correlación de redes existentes con las obtenidas	31

ÍNDICE DE FIGURAS

Figura 1. Flujo de trabajo de la metodología aplicada.....	11
Figura 2. Gráfico de análisis de componentes principales PCA	17
Figura 3. Gráfico de mapa de calor Heatmap	18
Figura 4. Representación gráfica de las variantes en contraste al genoma de referencia hg38	22

RESUMEN EJECUTIVO

La enfermedad de Huntington (EH) es un trastorno neurodegenerativo caracterizado por la expansión de repeticiones CAG en el gen HTT (Huntingtina), lo que conduce a disfunciones moleculares significativas. En el presente estudio se investigó el impacto de las variantes de splicing (VS) y factores de transcripción (FT) sobre la progresión de la enfermedad, para lo cual aplicamos un enfoque *in silico* mediante datos de secuenciación basados en tecnologías RNA-seq. Se identificaron variantes de splicing diferenciales y su posible influencia en la generación de isoformas aberrantes del gen HTT. De estas variantes, particularmente la condición 109CAG mostró un impacto más severo en el procesamiento del ARN mensajero, lo que afecta principalmente la estabilidad de la proteína huntingtina. Además, se construyeron y analizaron las redes funcionales de los factores HSF1, CBP y Sp1. CBP y Sp1 emergieron como nodos centrales, señalando que desempeñan roles clave en la regulación epigenética y genética global, mientras que, el HSF1 no muestra conexiones importantes con los demás. Estos resultados sugieren que las interacciones entre las VS y los FT exacerban la disfunción celular, lo que contribuye a la progresión de la EH. Finalmente, la investigación concluye que tanto las variantes de splicing como los factores de transcripción representan componentes críticos en la patogénesis de la EH. Estos hallazgos ofrecen una base sólida para explorar intervenciones terapéuticas dirigidas a modular estos procesos moleculares.

Palabras clave: Enfermedad de Huntington, variantes de splicing, factores de transcripción, gen HTT, bioinformática, CBP, Sp1, HSF1, redes funcionales.

ABSTRACT

Huntington's disease (HD) is a neurodegenerative disorder characterized by the expansion of CAG repeats in the HTT (Huntingtin) gene, leading to significant molecular dysfunctions. In the present study, the impact of splicing variants (SV) and transcription factors (TF) on disease progression was investigated using an in-silico approach with sequencing data based on RNA-seq technologies. Differential splicing variants were identified and their possible influence on the generation of aberrant HTT gene isoforms was analysed. Among these variants, the 109CAG condition showed a particularly severe impact on mRNA processing, primarily affecting huntingtin protein stability. In addition, the functional networks of the HSF1, CBP, and Sp1 factors were constructed and analysed. CBP and Sp1 emerged as central nodes, indicating that they play key roles in global epigenetic and genetic regulation, while HSF1 does not show significant connections with the others. These results suggest that interactions between SVs and TFs exacerbate cellular dysfunction, contributing to HD progression. Finally, the research concludes that both splicing variants and transcription factors represent critical components in HD pathogenesis. These findings provide a solid basis for exploring therapeutic interventions aimed at modulating these molecular processes.

Keywords: Huntington's disease, splicing variants, transcription factors, HTT gene, bioinformatics, CBP, Sp1, HSF1, functional networks.

CAPÍTULO I.- MARCO TEÓRICO

1.1. Antecedentes investigativos

La enfermedad de Huntington (EH), se presenta en los últimos años como un trastorno del neurodesarrollo, con mayor afectación en personas de mediana edad debido a su progresión incapacitante y neurodegenerativa (**Gatto et al., 2020**). La EH afecta la calidad de vida de las personas con síntomas psiquiátricos, déficit cognitivo y deterioro motor, actualmente, no existen enfoques novedosos para detener la progresión de la enfermedad y los tratamientos existentes están limitados al manejo de síntomas menores (**Lin et al., 2016**). La complejidad molecular de la EH y la necesidad de identificar nuevos enfoques terapéuticos que permitan manejar la progresión de la enfermedad hacen imperativa esta investigación.

Se ha evidenciado que las variantes de splicing (VS) pueden desempeñar un papel clave en la evolución de la EH, estas contribuyen a la generación de proteínas tóxicas implicadas en su patogénesis (**Elorza et al., 2021**). De la misma forma, otros estudios paralelos han destacado la importancia de los factores de transcripción (FT) HSF1, CBP y SP1 en la regulación génica asociada con la enfermedad, debido a su relación con la disfunción mitocondrial y la desregularización transcripcional (**Intihar et al., 2019**). Esta investigación se ha enfocado en entender la desregularización transcripcional y mitocondrial, buscando profundizar en el propósito de detectar las posibles variables genéticas asociadas (**Malla et al., 2021; Tabrizi et al., 2020**). En este contexto, la interacción entre estas variables podría ser determinante en la modulación de la expresión génica vinculada a la patología de la EH, lo que promueve nueva evidencia para comprender sus mecanismos moleculares.

Por tanto, considerando estos antecedentes la presente investigación estudia estas variables en conjunto para analizar el papel que desempeñan en la evolución de la EH, además, se puede utilizar como punto de partida para futuras investigaciones sobre las estrategias terapéuticas, buscando modular los procesos genéticos y moleculares de la enfermedad.

1.1.1. Generalidades de enfermedad de Huntington (EH)

La EH es un trastorno neurodegenerativo de herencia autosómica dominante que afecta a personas de mediana edad, generalmente de entre 20 y 50 años. Esta enfermedad se caracteriza por la degeneración progresiva del sistema nervioso central, especialmente en las regiones del cuerpo estriado, como el núcleo caudado y el putamen (**Croce & Yamamoto, 2019**). Los síntomas suelen incluir trastornos motores (como corea), deterioro cognitivo y alteraciones psiquiátricas que empeoran con el tiempo,

llevando eventualmente a la inmovilidad y la muerte, usualmente unos 15 a 20 años después de la aparición de los síntomas iniciales **(Bachoud-Lévi et al., 2019)**.

Aunque no existe una cura definitiva para la EH, las pautas actuales se centran en el manejo sintomático para mejorar la calidad de vida de los pacientes. Las estrategias incluyen tratamientos farmacológicos para manejar el corea y los síntomas psiquiátricos, así como intervenciones no farmacológicas para apoyar funciones cognitivas y motoras **(Bachoud-Lévi et al., 2019; Tabrizi et al., 2020)**. Las recomendaciones sugieren un enfoque multidisciplinario que integre el apoyo familiar y social, además de cuidados paliativos en las etapas avanzadas de la enfermedad.

1.1.1.1. Bases Genéticas de la EH

Una mutación genética en el gen HTT del cromosoma 4, que codifica la proteína huntingtina, causa la EH. La mutación se caracteriza por la expansión de repeticiones del trinucleótido CAG dentro del gen. En individuos sanos, el número de repeticiones CAG varía de 10 a 35, mientras que en aquellos con EH, estas repeticiones aumentan a 36 o más, lo que causa la síntesis de una proteína huntingtina anómala, que es tóxica para las células neuronales, particularmente en las áreas del cuerpo estriado del cerebro **(Ajitkumar & De Jesus, 2024)**.

La transmisión de la EH es autosómica dominante, lo que significa que basta con heredar una copia mutada del gen de uno de los progenitores para desarrollar la enfermedad **(Tabrizi et al., 2020)**. Además, existe un fenómeno conocido como anticipación genética, donde el número de repeticiones CAG puede aumentar cuando el gen defectuoso se hereda de un progenitor, especialmente del padre. Esto provoca que la enfermedad se presente a edades más tempranas en cada generación sucesiva **(Gatto et al., 2020)**.

1.1.2. Splicing alternativo en la enfermedad de Huntington

El splicing alternativo es un proceso clave que permite la generación de múltiples variantes de ARN mensajero (ARNm) a partir de un solo gen, lo que incrementa la diversidad proteica y funcional dentro de las células. En la EH, las alteraciones en el splicing alternativo del gen HTT han sido objeto de estudio, ya que se ha observado que estas variantes pueden influir significativamente en la expresión y función de la proteína huntingtina **(Elorza et al., 2021)**. Las investigaciones sugieren que ciertas variantes de splicing pueden llevar a la producción de isoformas de huntingtina con diferentes propiedades, algunas de las cuales podrían ser más propensas a formar agregados tóxicos en las neuronas **(Lin et al., 2016)**.

1.1.2.1. Posibles consecuencias funcionales de las variantes de splicing en la EH

Las VS del gen HTT pueden dar lugar a isoformas de la proteína huntingtina que varían en longitud y estructura, lo que puede tener diferentes consecuencias funcionales. Estudios recientes han identificado que algunas variantes resultan en fragmentos truncados de la proteína, más propensos a agregarse e inclusiones tóxicas en las células neuronales. Estas inclusiones están asociadas con la disfunción neuronal y la muerte celular, contribuyendo al deterioro progresivo observado en pacientes con EH (**Krach et al., 2022; Tano et al., 2023**). Además, se ha sugerido que el splicing alternativo defectuoso puede afectar la localización subcelular de la proteína huntingtina, alterando sus interacciones con otras proteínas esenciales para la función neuronal normal (**Elorza et al., 2021**).

1.1.2.2. Importancia del splicing alternativo en la progresión de la enfermedad

El splicing alternativo desempeña un papel crucial en la regulación de la expresión del gen HTT, y las variantes producidas por este proceso pueden influir en la severidad y progresión de la EH. Las investigaciones han demostrado que algunas variantes de splicing pueden actuar como moduladores de la toxicidad de la proteína huntingtina, afectando la tasa de agregación y la capacidad de la proteína para interactuar con otras moléculas dentro de la célula, ocasionando así una progresión más rápida de la enfermedad (**Gallardo & Delisle, 2023; Mason et al., 2020**).

1.1.3. Factores de transcripción

Los FT interactúan con patrones específicos de ADN uniéndose a regiones reguladoras, que aumentan o disminuyen la disfunción mitocondrial y la transcripción de genes, afectando varios procesos genéticos, entre ellos la síntesis del ARNm y proteínas (**Meem et al., 2023**).

1.1.3.1. HSF1 (Heat Shock Factor 1)

HSF1 es un factor de transcripción crucial en la respuesta celular al estrés, regulando la expresión de chaperonas que ayudan al correcto plegamiento de proteínas y la eliminación de agregados tóxicos. Su activación ha demostrado efectos neuro protectores en el contexto de patologías neurodegenerativas, donde se observa disfunción proteotóxica (**H. Kim & Gomez-Pastor, 2022**). Estudios sugieren que la activación sostenida de HSF1 podría mitigar la progresión de la EH, ya que su activación podría incrementar la capacidad de las células para lidiar con el estrés proteotóxico inducido por la proteína huntingtina mutada (**Burchfiel et al., 2021**).

1.1.3.2. CBP (CREB-binding protein)

La proteína de unión a CREB (CBP) es un coactivador transcripcional que desempeña un papel clave en la regulación de la expresión génica y la plasticidad neuronal. En el contexto de la EH, la CBP se une a proteínas como CREB para facilitar la transcripción de genes esenciales para la función y supervivencia neuronal (**Arancibia-Opazo et al., 2023; Spies et al., 2023**). Sin embargo, la mutación en la proteína huntingtina puede interferir con la función de la CBP, afectando la transcripción de genes neuro protectores y contribuyendo al deterioro neuronal característico de la EH (**Ivanenko et al., 2022**). Estudios recientes han sugerido que la modulación de CBP a través de la restauración de su interacción con CREB o el aumento de su actividad, podría mitigar los efectos neurotóxicos de la proteína huntingtina mutada (**C. Kim et al., 2021**).

1.1.3.3. Sp1 (Specificity Protein 1)

La Proteína de Especificidad 1 (Sp1) es un factor de transcripción que desempeña un papel fundamental en la regulación de la expresión génica y se ha implicado en diversas enfermedades neurodegenerativas, incluyendo la EH (**Ivanenko et al., 2022**). En el contexto de estas patologías, Sp1 puede influir en la expresión de genes relacionados con la supervivencia neuronal y la respuesta al estrés, lo que sugiere que su disfunción podría contribuir a la progresión de estas enfermedades (**Qu et al., 2022**). En particular, en la EH, se ha observado que Sp1 puede regular la expresión del gen HTT, y su actividad puede verse alterada por la presencia de huntingtina mutante, lo que provoca una cascada patológica que afecta la función neuronal (**Paul et al., 2022**).

1.1.4. Interacciones entre el splicing alternativo y los factores de transcripción

El splicing alternativo y los FT están interconectados a través de complejas redes de regulación que modulan la expresión y funcionalidad de los genes. Los FT pueden influir en el proceso de splicing alternativo al interactuar con componentes del complejo de splicing, afectando la selección de exones y la generación de isoformas específicas de ARNm (**Seefeldt & Kochanek, 2021**).

Este proceso permite a las células adaptarse a diferentes condiciones y responder a señales externas al regular la expresión de diferentes variantes proteicas. De manera recíproca, las variantes generadas por el splicing alternativo pueden regular la actividad de los FT, creando circuitos de retroalimentación que impactan procesos celulares críticos, como la diferenciación, el crecimiento celular y la respuesta al estrés (**Pogoda et al., 2021**).

1.1.4.1. Implicación en la EH

En la EH, las interacciones anormales entre el splicing alternativo y los FT contribuyen a la disfunción celular y la neurodegeneración. Se ha observado que la huntingtina mutada puede interferir con los mecanismos normales de splicing, afectando la producción de variantes de proteínas cruciales para la función neuronal **(Gallardo & Delisle, 2023)**. Además, la huntingtina mutada altera la función de los FT como CREB y Sp1, interfiriendo en la transcripción de genes necesarios para la supervivencia celular y el manejo del estrés. Esta disrupción en la regulación de splicing y transcripción exacerba la acumulación de proteínas mal plegadas y la disfunción mitocondrial, contribuyendo a la progresión de la EH **(Ivanenko et al., 2022; Tano et al., 2023)**.

1.1.5. Bioinformática para el análisis de variables en la EH

1.1.5.1. Tecnología de secuenciación

El transcriptoma comprende las regiones del genoma que han sido transcritas o expresadas como ARN. Existen varios tipos de ARN, entre ellos el ARN mensajero (ARNm), contiene la información necesaria para la síntesis de proteínas durante el proceso de traducción **(Buccitelli & Selbach, 2020)**. El estudio de transcriptomas se ha visto facilitado en los últimos años por el surgimiento de plataformas de secuenciación de próxima generación (NGS) y con la introducción de la secuenciación de ARN (RNA-seq) que permite la obtención de datos con mayor porcentaje de confiabilidad en contraste con datos antiguos generados por la tecnología de microarrays **(Stark et al., 2019)**.

1.1.5.2. Análisis de expresión diferencial (DEG)

El análisis de expresión diferencial (DEG) es una técnica que utiliza la secuenciación de ARN (RNA-Seq) para comparar niveles de expresión génica entre diferentes condiciones experimentales. Esta técnica permite identificar genes cuya expresión varía significativamente, proporcionando información clave sobre los mecanismos moleculares que subyacen a distintos fenómenos biológicos o enfermedades **(Bensalel et al., 2021)**. La identificación de DEGs se realiza mediante análisis estadísticos que evalúan la cantidad de lecturas asignadas a cada gen en las muestras secuenciadas, lo que permite detectar variaciones en la expresión génica que podrían ser relevantes para la condición estudiada **(Colpo et al., 2020)**.

Los resultados del análisis de DEG pueden visualizarse utilizando diversas herramientas gráficas que ayudan a interpretar las diferencias en la expresión génica. Entre estas, los mapas de calor destacan por su capacidad para mostrar patrones de expresión diferencial a lo largo de múltiples muestras, con

cambios significativos en la expresión y p-valores ajustados (**McDermaid et al., 2019**). Además, el análisis de componentes principales o gráficos de dispersión permiten visualizar y comparar directamente los niveles de expresión entre las condiciones, facilitando la identificación de tendencias y patrones significativos (**Li et al., 2022**).

1.1.5.3. Herramientas bioinformáticas para el análisis de splicing alternativo

Las herramientas bioinformáticas para el análisis de splicing alternativo permiten la identificación y caracterización de variantes de splicing a partir de datos de secuenciación de ARN (RNA-seq). Una de las plataformas más utilizadas es Galaxy, la cual facilita la gestión y análisis de grandes volúmenes de datos a través de flujos de trabajo personalizados y herramientas integradas. Para detectar variantes de splicing, los datos de RNA-seq se importan a entornos específicos dentro de Galaxy, donde se procesan utilizando metodologías diseñadas para identificar patrones de splicing alternativo, como la exclusión o inclusión de exones, y otras modificaciones en la estructura del ARN mensajero (**Gallardo & Delisle, 2023**). Esta plataforma es accesible y permite realizar análisis complejos sin necesidad de una infraestructura avanzada, lo que la convierte en una herramienta versátil para la bioinformática.

La visualización de las variantes detectadas es un paso crucial para confirmar y comprender las alteraciones en el splicing. Para ello, se utiliza el Integrative Genomics Viewer (IGV), que permite observar cómo las variantes se alinean con el genoma de referencia, en este caso, el ensamblaje HG38 de Homo sapiens. Esta herramienta facilita la interpretación y caracterización de las variantes al proporcionar una representación gráfica del alineamiento de RNA-seq, destacando regiones donde ocurren cambios en el splicing. El uso de IGV aumenta la confianza en las variantes detectadas y ayuda a focalizar el análisis en las que podrían tener mayor relevancia biológica (**Liang et al., 2022; Robinson et al., 2017**).

1.2. Objetivos

1.2.1. Objetivo general

- Analizar variantes de splicing y factores de transcripción (HSF1, CBP y Sp1) en la evolución de la enfermedad de Huntington.

1.2.2. Objetivos específicos

- Describir la expresión génica en la EH a partir de experimentos basados en RNA-seq obtenidos de bases de datos públicas.
- Identificar la influencia de las variantes de splicing diferenciales y factores de transcripción en genes expresados en la EH.
- Interpretar la influencia de las variantes de splicing diferenciales y factores de transcripción en la EH.

CAPÍTULO II.- METODOLOGÍA

2.1. Materiales

El trabajo experimental se realizó en las instalaciones de la Facultad de Ciencia e Ingeniería en Alimentos y Biotecnología de la Universidad Técnica de Ambato. En las **Tablas 1 y 2**, se lista las bases de datos biológicas, software y equipo que se usó durante la presente investigación.

Tabla 1.

Herramientas bioinformáticas

Recursos	Dirección web
Linux	
NCBI	https://www.ncbi.nlm.nih.gov/
Galaxy Australia	https://usegalaxy.org.au/
Galaxy server	https://usegalaxy.org/
FastQC	https://www.bioinformatics.babraham.ac.uk/projects/fastqc/
Trimmomatic 0.39	https://github.com/usadellab/Trimmomatic/releases
R	https://cran.r-project.org/bin/windows/base/
RStudio	https://posit.co/download/rstudio-desktop/
rMATs	https://github.com/Xinglab/rmats-turbo
Variant Effect Predictor (VEP)	https://www.ensembl.org/Tools/VEP
Integrative genoma viewer (IGV)	https://igv.org/
Cytoscape	https://cytoscape.org/
STRING	https://apps.cytoscape.org/apps/stringapp
GeneMANIA	https://apps.cytoscape.org/apps/genemania
cytoHUBBA	https://apps.cytoscape.org/apps/cytohubba
MCODE	https://apps.cytoscape.org/apps/mcode

Tabla 2.

Recursos materiales

Material	Cantidad
Computador portátil	1

2.2. Métodos

2.2.1. Búsqueda de datos de secuenciación RNA-seq de muestras de Homo sapiens con EH

Se han realizado distintos estudios sobre la EH, los cuales han almacenado sus resultados en distintas bases de datos biológicas públicas, en el presente proyecto se usaron datos obtenidos en el consorcio National Center of Biotechnology Information (NCBI), tales conjuntos se enfocan en dos condiciones de investigación distintas: células madre pluripotentes inducidas (iPSC) con EH y células de tejido estriado de pacientes con EH post-mortem.

La investigación realizada por **Świtońska et al., (2019)**, con código de acceso GSE124664 tuvo como objetivo definir y revelar las alteraciones transcripcionales en la EH temprana, mediante la investigación de los perfiles transcripcionales de líneas de iPSC con EH. Se aisló 9 muestras para el análisis de ARN, en las cuales 3 líneas iPSC son de control, 3 líneas iPSC clonales con 71 repeticiones CAG y 3 líneas iPSC clonales con 109 repeticiones CAG. La secuenciación se realizó utilizando la plataforma Illumina HiSeq 2000.

2.2.2. Control de calidad y análisis de expresión diferencial (DEG) de las muestras RNA-seq

Primero se descargó las muestras crudas en formato FASTQ correspondientes a cada estudio, se realizó el análisis de calidad de estas utilizando la plataforma Galaxy Australia (<https://usegalaxy.org.au>). Se utilizó FastQC para analizar la calidad de las muestras mediante la puntuación PHRED que representa la calidad de las bases, el contenido de CG, duplicados, adaptadores y contaminantes (**Sneha et al., 2022**). El recorte y filtrado se lo hizo mediante Trimmomatic, se recortaron las primeras 15 bases de las secuencias y se filtró los adaptadores con la secuencia de adaptadores TruSeq2 (paired-ended, for Illumina GA II) manteniendo la calidad y longitud de las muestras (**Ayyildiz et al., 2023**).

Se desarrolló dos réplicas del análisis DEG, el primero para discernir la semejanza entre las muestras y separar las muestras poco relacionadas y el segundo para obtener los resultados óptimos con las muestras filtradas del primer DEG, las muestras filtradas y seleccionadas se detallan en el **Anexo 1**. El análisis se realizó en un entorno de R versión 4.4.1 (2024-06-14 ucrt) -- "Race for Your Life", se alineó las lecturas utilizando "Rbowtie2" con el genoma de referencia GRCh38 donde se obtuvo archivos de salida SAM, que luego se convirtieron a archivos BAM mediante "Rsamtools" (**Sepulveda, 2020**). Para comparar y cuantificar los genes expresados entre las distintas muestras, se empleó "Rsubread" que cuenta el número de lecturas por cada gen comparándolo con el archivo de anotación de genes GFF correspondiente al genoma usado (**Malecka et al., 2024**).

Para la identificación de los genes diferencialmente expresados, se usó el paquete “DESeq2” seleccionando los conjuntos de datos referentes a cada una de las condiciones experimentales generando un resumen con información de la expresión génica, estadísticas significativas y puntuaciones de ajuste (S. Liu et al., 2021). Con los paquetes “ggplot2”, “heatmap3” y “pheatmap” se generó dos tipos de gráficos de resultados, un análisis de componentes principales (PCA) y un Heatmap filtrando los genes con un p_value ajustado < 0.05 y cambios absolutos mediante un $\log_2 > 1$ (Bensalel et al., 2021; Colpo et al., 2020). El script usado para el análisis DEG y la representación gráfica se encuentra en el **Anexo 2**.

2.2.3. Detección de variantes de splicing y visualización

Se creó una cuenta en Galaxy Australia (<https://usegalaxy.org.au>), dentro de esta se generaron dos entornos “History” para cada una de las condiciones de las muestras del proyecto seleccionado una con el nombre “71CAG” y la otra “109CAG”. En estos entornos, se cargó las muestras FASTQ seleccionadas previamente mediante el análisis DEG de cada condición y las muestras de control, luego se procedió al análisis siguiendo la metodología descrita por Gallardo & Delisle (2023), donde se obtuvo las variantes VCF para el gen HTT, con estos archivos se analizó mediante predicciones usando Variant Effect Predictor (VEP) el posible efecto de estas variantes (Livesey & Marsh, 2022).

Posteriormente, para complementar el análisis se realizó un análisis de splicing alternativo mediante el programa rMATS, donde se obtuvo resultados de los cinco patrones básicos de splicing alternativo: exón omitido (SE), sitios de empalme alternativos 5' (A5SS), sitios de empalme alternativos 3' (A3SS), exones mutuamente excluyentes (MXE) o intrón retenido (RI) para ambas condiciones de las muestras 71CAG y 109CAG (Jiang et al., 2023). Por último, se utilizó Integrative Genomics Viewer (IGV) para la visualización las variantes en el contexto del genoma HG38 y el alineamiento de las muestras RNA-seq para interpretar y caracterizar las semejanzas o diferencias entre las variantes de mayor interés entre ambas condiciones de las muestras analizadas (Liang et al., 2022; Robinson et al., 2017). El script usado para la detección de las variantes se presenta en el **Anexo 3 y 4**.

2.2.4. Construcción y análisis de redes de co-expresión e interacción funcional de los factores de transcripción HSF1, CBP y Sp1

Se utilizó STRING mediante su plugin en Cytoscape para obtener la información de los FT, usando los genes significativos previamente identificados mediante el DEG se generó las redes correspondientes para cada FT (Christodoulou & Papanicolaou, 2023). Una vez construidas las redes se usó los complementos cytoHubba y MCODE para la identificación de los módulos estrechamente conectados

en las redes y clasificar los mismos por sus características según los métodos de puntuación de Cytoscape: robustez, flujo, conectividad local y centralidad global (Meem et al., 2023).

2.2.5. Cuantificación de variantes de Splicing y Factores de transcripción

Se tabularon los datos correspondientes a los niveles de expresión de las diferentes VS y FT, centrándonos en las VS específicamente en los datos canónicos A5SS, A3SS, MXE, RI y SE (Tano et al., 2023) y para los FT en sus niveles de expresión, afinidad y tipo permitiendo identificar los patrones diferenciales de expresión entre las condiciones de estudio (Yan et al., 2021).

2.2.6. Correlación de las disfunciones genéticas y su contribución en la patogénesis de la EH

Se obtuvieron datos de hallazgos de investigaciones previas sobre las VS y los FT en la patogénesis de la EH en las bases de datos Invitae, ClinVar y Ensembl (Truty et al., 2021). Estos datos se correlacionaron con los datos obtenidos en el presente estudio, para determinar la contribución de la información obtenida y lograr identificar el papel, importancia e implicación de los resultados y como estos pueden aportar en futuras investigaciones (Hyeon et al., 2021).

En la **Figura 1** se presenta el diagrama de la metodología aplicada en el estudio.

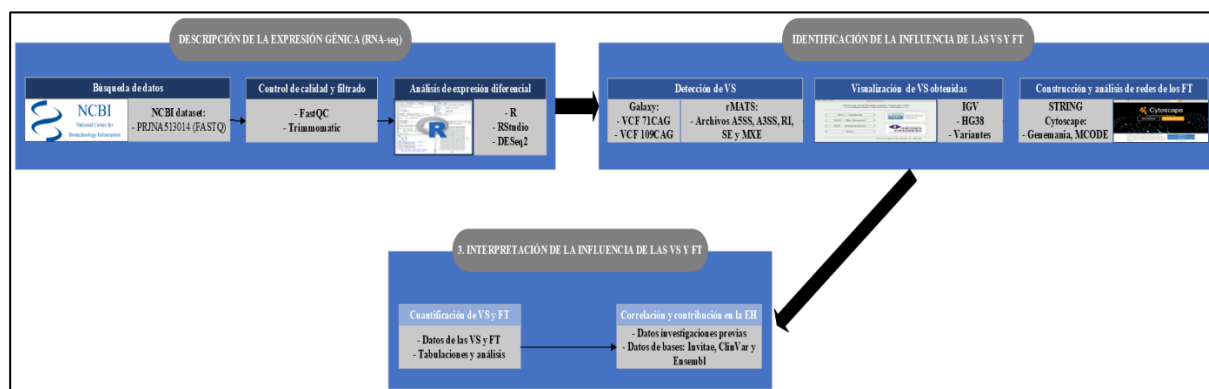


Figura 1. Flujo de trabajo de la metodología aplicada

Nota: Creado con Visio.

CAPÍTULO III.- RESULTADOS Y DISCUSIÓN

3.1. Análisis y discusión de resultados

3.1.1. Búsqueda de datos de secuenciación RNA-seq con EH

Los archivos de RNA-seq descargados en formato FASTQ para las condiciones se muestran en la **Tabla 3**. En esta tabla se indica los códigos de acceso, así como su respectiva condición, para cada una de las muestras se encuentran los archivos RNA-seq “forward” y “reverse”, indicando que la secuenciación corresponde a formato paired-end.

Tabla 3.

Muestras del proyecto GSE124664

Código	Condición
SRR8393695	Control
SRR8393696	Control
SRR8393697	Control
SRR8393698	Repetición de 71 CAG
SRR8393699	Repetición de 71 CAG
SRR8393700	Repetición de 71 CAG
SRR8393701	Repetición de 109 CAG
SRR8393702	Repetición de 109 CAG
SRR8393703	Repetición de 109 CAG

Nota: (Świtońska et al., 2019).

3.1.2. Control de calidad de las muestras formato FASTQ

Del control de calidad de las secuencias de cada muestra se observó una longitud promedio de 101 pares de bases (pb). Se analizó dos aspectos específicos que proporciona FastQC: calidad de base por secuencia y contenido de GC por secuencia. Se incluyen gráficos de estos dos aspectos tanto de datos crudos como filtrados y cortados, se seleccionaron muestras representativas para cada condición: SRR8393695 para el control, SRR8393698 para iPSC con repetición de 71 CAG y SRR8393701 para iPSC con repetición de 109 CAG. Los gráficos de control de calidad de todas las muestras de las tres condiciones de estudio se detallan en el **Anexo 5 y 6**.

La “Calidad de base por secuencia” mediante la puntuación PHRED asigna un valor a la probabilidad del proceso de identificación y asignación de una base (A, G, C o T) ya sea correctamente o no, en una

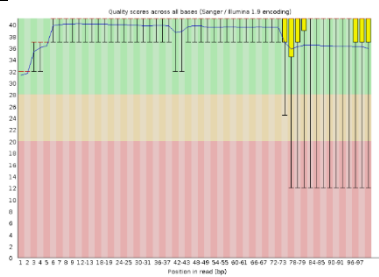
posición específica en la secuencia analizada (Wagner et al., 2021). Según De Sena Brandine & Smith (2021), la calidad de las muestras se clasifica en tres puntos: PHRED < 30 baja calidad, 30 < PHRED < 40 calidad media y PHRED > 40 calidad alta.

En la **Tabla 4** se presenta la “calidad por base de secuencia” de las muestras representativas crudas y cortadas. Se observó que las secuencias tanto forward como reverse presentan mayormente un promedio de calidad dentro de la zona de calidad alta según la puntuación PHRED (De Sena Brandine & Smith, 2021), sin embargo, los parámetros definidos en el código de control de calidad de FastQC y recorte de secuencias con Trimmomatic donde se limitó la longitud a 101 pb con lo que se eliminó las secuencias pequeñas de menor calidad, esto garantiza que las secuencias tengan una media en la puntuación PHRED mayor a 30 asegurando que al menos el 50% de las bases tienen una calidad óptima para analizar (Saif et al., 2021).

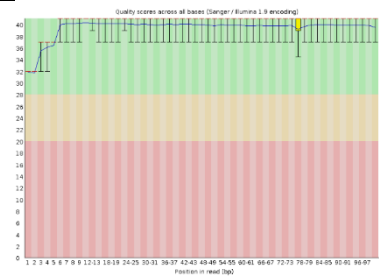
Tabla 4.

Calidad de base por secuencia de muestras representativas

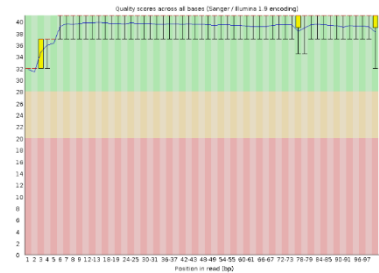
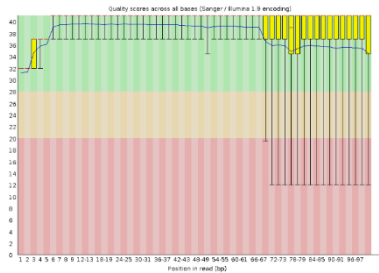
Control	
Muestras Crudas	Muestras Cortadas
SRR8393695_control Forward	SRR8393695_control Forward
SRR8393695_control Reverse	SRR8393695_control Reverse
Repetición 71CAG	
Muestras Crudas	Muestras Cortadas
SRR8393698_71CAG Forward	SRR8393698_71CAG Forward



SRR8393698_71CAG Reverse



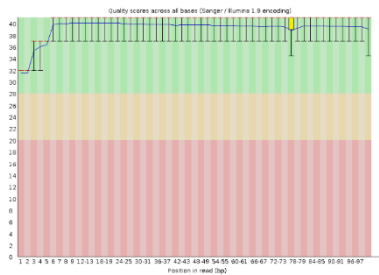
SRR8393698_71CAG Reverse



Repetición 109CAG

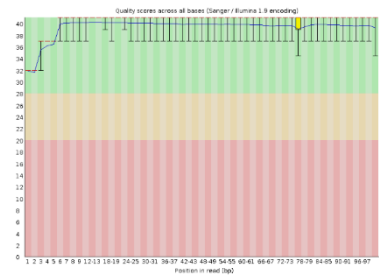
Muestras Crudas

SRR8393701_109CAG Forward

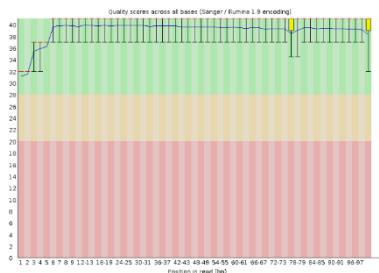


Muestras Cortadas

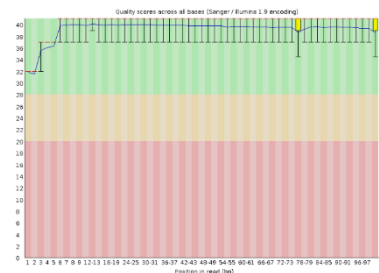
SRR8393701_109CAG Forward



SRR8393701_109CAG Reverse



SRR8393701_109CAG Reverse



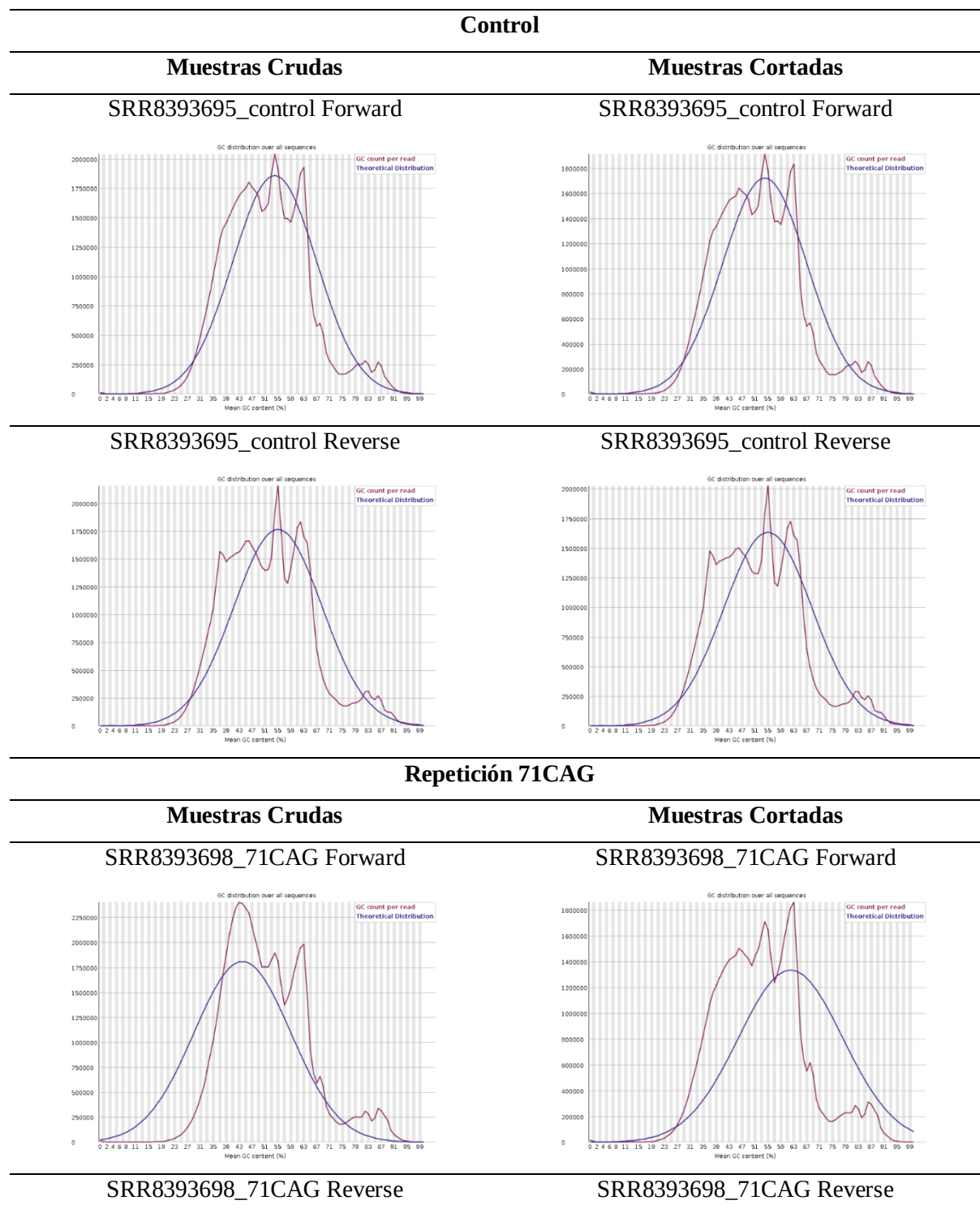
Nota: Gráficos creados por FastQC.

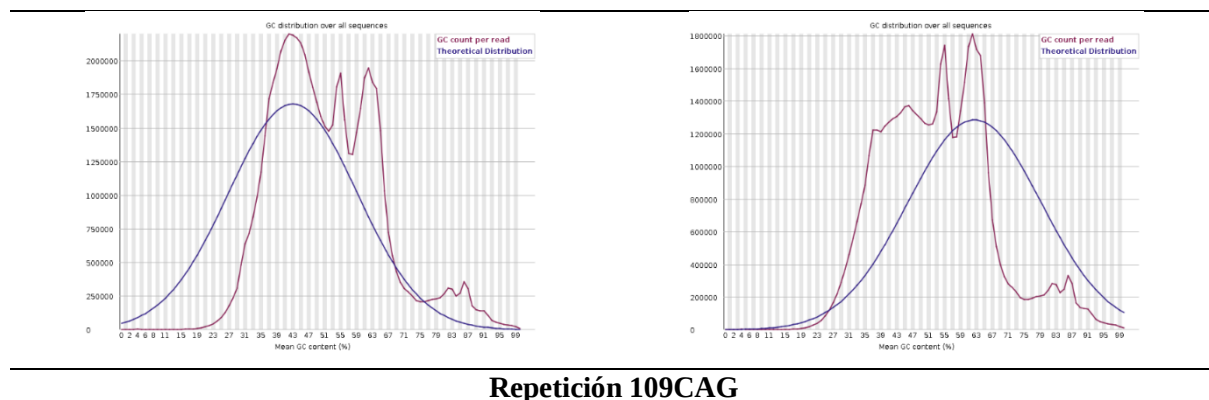
La **Tabla 5** presenta los gráficos del contenido de GC en las secuencias, se comparó el contenido de GC con la distribución normal esperada, mientras más se acerquen a la distribución esperada se consideran más aceptables, aunque suelen existir desviaciones y picos mayores o menores ya sea por contenido de contaminantes como cebadores, adaptadores o secuencias específicas usadas en la secuenciación de las muestras (**Browne et al., 2020**). Las muestras crudas, presentan una distribución desviada del contenido de GC, al realizar el filtrado y recorte de secuencias y adaptadores se muestra una mejora, aunque aún

existen picos y desviaciones, esto se debe a que las muestras procedentes de células iPSC con modificación de repeticiones CAG aumenta y varía el contenido de GC (Świtońska et al., 2019).

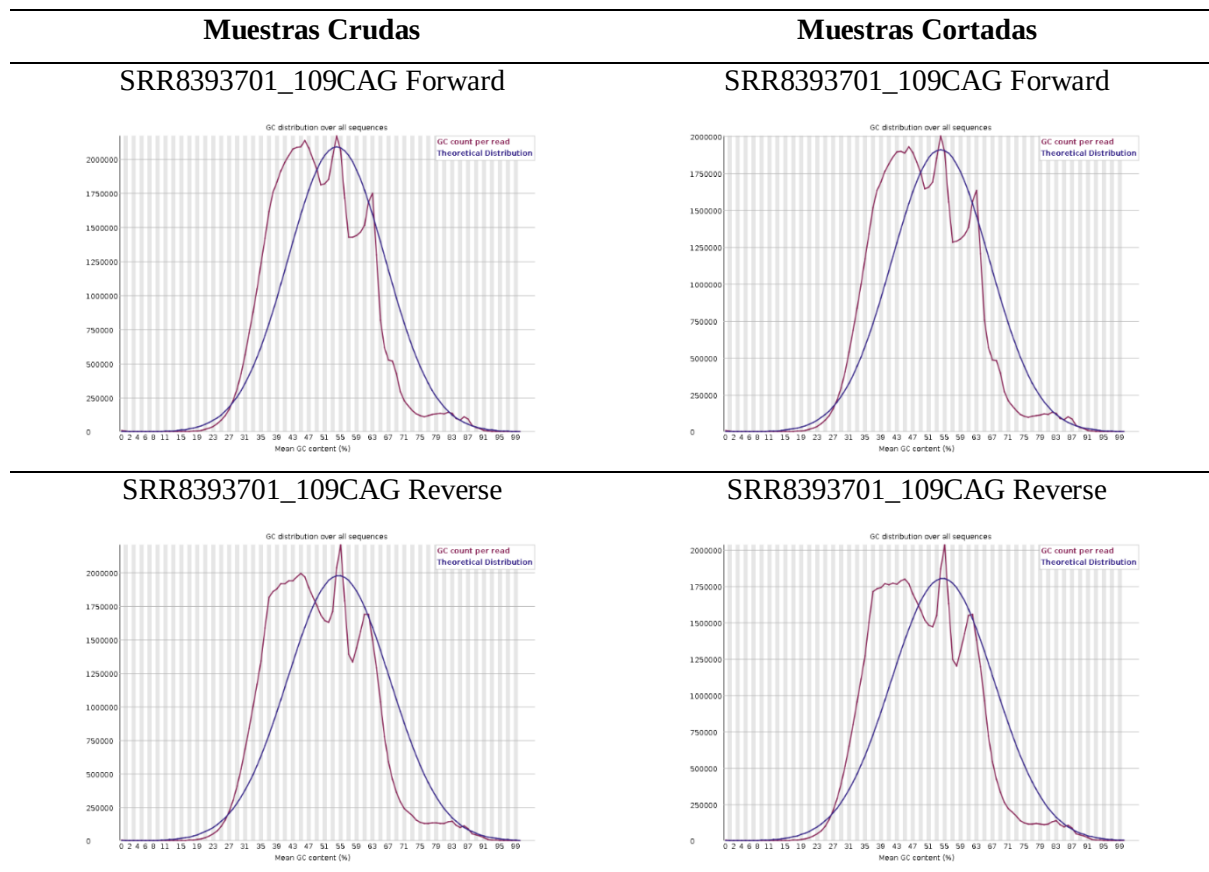
Tabla 5.

Contenido de GC por secuencia de muestras representativas





Repetición 109CAG



Nota: Gráficos creados por FastQC.

3.1.3. Análisis de expresión diferencial

Se desarrolló los dos análisis DEG descritos en la metodología para las tres condiciones de análisis. Del análisis mediante DESeq2 los resultados de las muestras filtradas de mayor semejanza para el estudio se representaron en gráficas PCA presentado en la **Figura 2** y mapas de calor Heatmap presentado en la **Figura 3**. Los gráficos de las muestras completas se presentan en el **Anexo 7**.

El análisis PCA se usó para el análisis de las muestras y sus distintas condiciones. En la gráfica de las muestras significativas, se observó que los dos componentes principales explican un 82 % de la

variabilidad total de los datos (PC1:56 % y PC2: 26 %), indicando que capturan gran parte de las diferencias entre las muestras. El porcentaje de variabilidad del 82%, presenta un umbral alto lo que da robustez al análisis, superando el umbral del 80% propuestos en la investigación de análisis PCA de **Camargo (2022)**. El análisis PCA evidencia que el incremento en el número de repeticiones CAG genera un impacto progresivo y diferenciado en las características analizadas. Este comportamiento es consistente con lo esperado en trastornos relacionados con repeticiones CAG, como lo es la EH, donde el aumento en las repeticiones se asocia con cambios severos en la función celular y fenotipos (**Ayyildiz et al., 2023**).

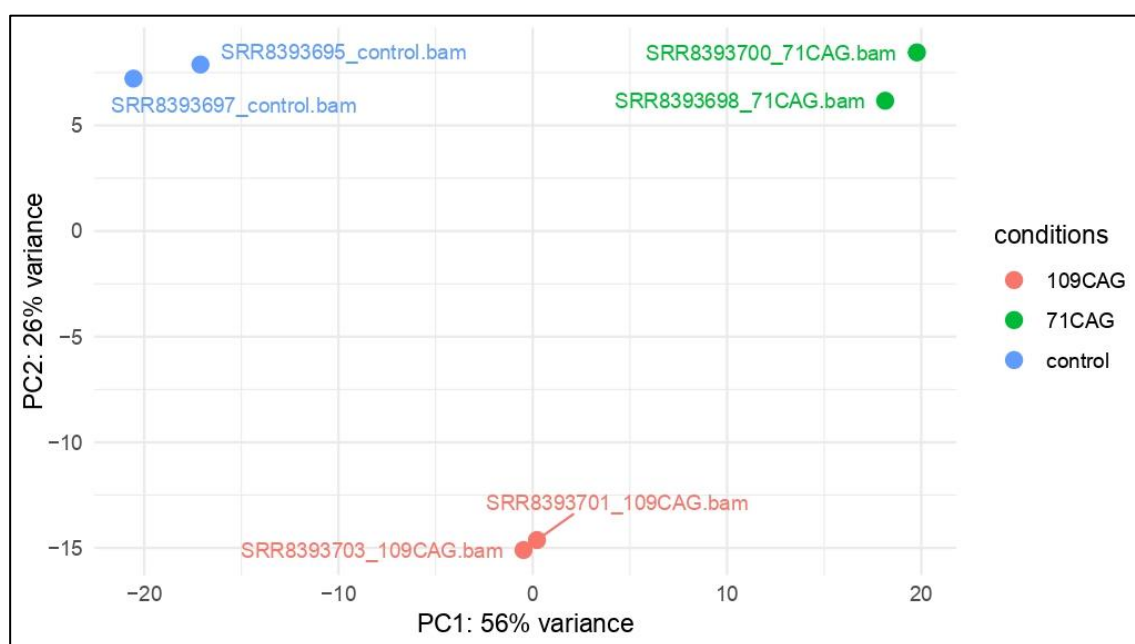


Figura 2. Gráfico de análisis de componentes principales PCA

Nota: Las muestras control se agrupan cercanamente lo que sugiere una alta consistencia y homogeneidad en los perfiles biológicos de estas. La separación clara entre 71CAG y control en el eje PC1 muestra que las diferencias en esta componente están altamente influenciadas por el incremento en las repeticiones CAG, mientras que la posición de las muestras 109CAG indica que las repeticiones extremadamente expandidas tienen un impacto mucho mayor en los perfiles biológicos, lo que podría estar relacionado con alteraciones masivas en la expresión génica, metabolismo o estructura celular.

En el Heatmap, según los criterios usados con la herramienta en R se presentan treinta genes que cumplen con las condiciones para ser significativos. El color azul muestra los genes que se inhiben mientras más intenso sea el color azul mayor es la inhibición, por otro lado, el color rojo representa la sobreexpresión que de igual manera mientras más intenso mayor es la sobreexpresión.

La transición de control a 71CAG y finalmente a 109CAG muestra un aumento progresivo en la regulación diferencial de los genes, reflejando la influencia directa de las repeticiones CAG en la

disfunción molecular (**Choi et al., 2024**). Este hallazgo es consistente con estudios previos que asocian las repeticiones CAG con alteraciones en la expresión génica, potencialmente mediadas por mecanismos epigenéticos o de regulación transcripcional (**Handsaker et al., 2024**).

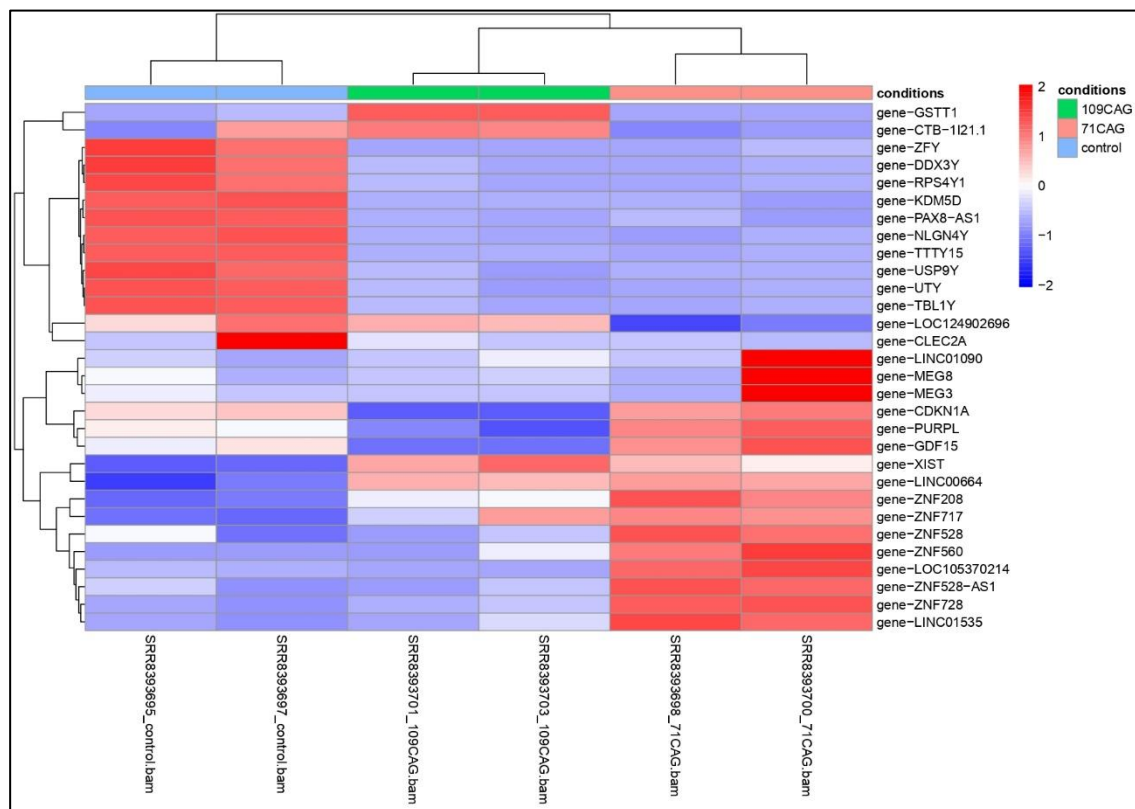


Figura 3. Gráfico de mapa de calor Heatmap

3.1.4. Detección de variantes de splicing

El análisis de variantes realizado en Galaxy comparando control contra las dos condiciones de estudio 71 CAG y 109 CAG revela diferencias significativas en las mutaciones observadas y posibles efectos en el splicing, en la **Tabla 6** a continuación se presentan algunas de las variantes encontradas para ambos casos. Las variantes completas se detallan en el **Anexo 8 y 9**.

Las variantes de 71 CAG predominan en regiones intrónicas, lo que sugiere un efecto más sutil pero aún disruptivo en el proceso de splicing. En contraste, las variantes del caso 109 CAG afectan tanto regiones intrónicas como exónicas, lo que puede tener consecuencias más severas (**Ayyildiz et al., 2023**). Las variantes al ser procesadas mediante VEP para predecir sus implicaciones funcionales, se observó que pueden afectar el reconocimiento del sitio de splicing, la codificación de la proteína, la codificación y no codificación de transcritos por parte de la maquinaria de empalme, esto posiblemente puede generar isoformas anómalas del gen HTT (**Fienko et al., 2022; Oka et al., 2021**). Además, podría contribuir a

una mayor acumulación de proteínas defectuosas asociadas con la patología, los datos de predicción del efecto de las variantes obtenidos de VEP se detallan en el **Anexo 10**.

Tabla 6.

Variantes encontradas en 71 CAG y 109 CAG

Cromosoma	Posición	ID	Referencia	Alteración	Qual
71CAG					
chr4	3135968	chr4:3135968_G/C	G	C	504,274
chr4	3177389	chr4:3177389_T/G	T	G	355,698
chr4	3180513	chr4:3180513_AGC/ATGC	AGC	ATGC	174,471
109CAG					
chr4	3135968	chr4:3135968_G/C	G	C	352,372
chr4	3223405	chr4:3223405_G/A	G	A	290,542
chr4	3180513	chr4:3180513_AGC/ATGC	AGC	ATGC	137,069

La expansión CAG más pronunciada en la condición 109 CAG (posición 3180513) podría amplificar la disfunción del splicing, reflejando un modelo en el que ambas alteraciones (genética y funcional) se potencian mutuamente (**Świtońska et al., 2019**). Según estos resultados se evidencia diferencias significativas entre los dos casos, que no solo reflejan la heterogeneidad clínica de la EH, sino también la necesidad de un análisis profundo y específico de las variantes genéticas para comprender mejor los mecanismos subyacentes y diseñar intervenciones terapéuticas más efectivas. Por ejemplo, oligonucleótidos anti sentidos podrían ser efectivos en el caso 109CAG para bloquear sitios de splicing aberrantes (**Aslesh & Yokota, 2020**).

3.1.4.1. Splicing alternativo rMATS

Del análisis realizado para las dos comparaciones control vs 71CAG y control vs 109CAG, se obtuvo como resultado un total de eventos de todos los genes del genoma humano (GRHg38), el total de eventos detallado se encuentra en el **Anexo 11**. De estos eventos aquellos específicos para el gen HTT en la primera comparación se obtuvo 64 y para la segunda comparación 48, estos eventos se detallan en el **Anexo 12 y 13** respectivamente. Nos enfocamos principalmente en aquellos con un p_value menor a 0,05 debido a que esta puntuación permite saber que dichos eventos son significativos (**Shen et al., 2014**).

Tabla 7.

Eventos de splicing significativos comparación Control vs 71CAG

Gen	Inicio	Final	Inicio	Final	Inicio	Final	Conteo lecturas	Conteo lecturas	Longitud	Longitud	P _ valor
	Exón_larg	Exón	Exón	Exón	Exón	Exón	incluido el	incluido el	transcrito	transcrito	
	o_0base	largo	corto	corto	débil	débil	evento_control	evento_71CAG	normal	alterado	
A3SS – JC											
HTT	3207280	3207357	3207285	3207357	3206806	3206983	38,5	53,29	105	100	0,0048195
A3SS – JCEC											
HTT	3207280	3207357	3207285	3207357	3206806	3206983	38,5	53,29	105	100	0,0048195
A5SS – JC											
HTT	3172907	3173331	3172907	3173131	3174720	3174799	7,11	2,3	200	100	0,0064002
HTT	3172907	3173218	3172907	3173131	3174720	3174799	6,9	3,2	187	100	0,0243922
A5SS – JCEC											
HTT	3172907	3173218	3172907	3173131	3174720	3174799	6,9	3,2	187	100	0,0243922
Gen	Inicio	Final	Inicio	Final	Inicio_E	Inicio_E	Conteo lecturas	Conteo lecturas	Longitud	Longitud	P _ valor
	1stExón_0	1stExón	2ndExón	2ndExón	xón_ant	xón_ant	incluido el	incluido el	transcrito	transcrito	
	base		_ 0base		erior (3')	erior (5')	evento_control	evento_71CAG	normal	alterado	
MXE – JC											
HTT	3099273	3099394	3103823	3103883	3086938	3105356	42,41	75,48	200	160	0,02363121
MXE – JCEC											
HTT	3099273	3099394	3103823	3103883	3086938	3105356	55,59	103,7	221	160	0,01149007
HTT	3086938	3087022	3099273	3099394	3074680	3103823	26,57	42,23	184	221	0,03733444

Tabla 8.

Eventos de splicing significativos comparación Control vs 109CAG

Gen	Inicio	Final	Inicio	Final	Inicio	Final	Conteo lecturas	Conteo lecturas	Longitud	Longitud	P _ valor
	Exón_larg	Exón	Exón	Exón	Exón	Exón	incluido el	incluido el	transcrito	transcrito	
	o_0base	largo	corto	corto	débil	débil	evento_control	evento_109CAG	normal	alterado	
A3SS – JC											
HTT	3207280	3207357	3207285	3207357	3206806	3206983	38,5	39,34	105	100	0,03586915
A3SS – JCEC											
HTT	3207280	3207357	3207285	3207357	3206806	3206983	38,5	39,34	105	100	0,03586915
Gen	Inicio	Final	Inicio	Final	Inicio_E	Inicio_E	Conteo lecturas	Conteo lecturas	Longitud	Longitud	P _ valor
	1stExón_0	1stExón	2ndExón	2ndExón	xón_ant	xón_ant	incluido el	incluido el	transcrito	transcrito	
	base		_ 0base		erior (3')	erior (5')	evento_control	evento_109CAG	normal	alterado	
MXE – JC											
HTT	3103823	3103883	3105356	3105436	3099273	3107284	32,3	49,48	160	180	0,00812202
HTT	3086938	3087022	3099273	3099394	3074680	3103823	26,57	22,33	184	200	0,03223897
MXE – JCEC											
HTT	3103823	3103883	3105356	3105436	3099273	3107284	32,3	49,48	160	180	0,00812202
HTT	3086938	3087022	3099273	3099394	3074680	3103823	26,57	22,33	184	221	0,03165829
Gen	Inicio	Final	Inicio_E	Final_E	Inicio_E	Final_E	Conteo lecturas	Conteo lecturas	Longitud	Longitud	P _ valor
	ri_Exón_0	ri_Exón	xón_ant	xón_ant	xón_ant	xón_ant	incluido el	incluido el	normal del	transcrito	
	base		erior (3')	erior (3')	erior (5')	erior (5')	evento_control	evento_109CAG	transcrito	alterado	
RI - JCEC											
HTT	3187650	3189093	3187650	3187886	3188950	3189093	74,65	65,62	1164	100	0,02970306

En la comparativa “Control vs 71CAG” los eventos A5SS y MXE destacan como categorías predominantes con un número mayor de eventos significativos como se observa en la **Tabla 7**, estos tipos de splicing ocasionan una alteración en la función de proteínas clave, producción de isoformas aberrantes y alteraciones en las rutas metabólicas neuronales (**L. Liu et al., 2023; Nguyen et al., 2023**), lo que sugiere un aumento en la neurodegeneración puesto que podrían alterar procesos críticos como el transporte axonal, la señalización sináptica o síntesis de proteínas (**Elorza et al., 2021**).

Por su parte, en los resultados de la comparativa “Control vs 109CAG” los eventos que destacan son los MXE y RI (**Tabla 8**), que ocasionan producción de isoformas aberrantes, estrés celular y muerte neuronal, producción de transcritos aberrantes y alteraciones en las rutas metabólicas neuronales (**Nguyen et al., 2023**). Así también, están involucrados en la plasticidad neuronal y en la regulación del citoesqueleto, sugiriendo una relación directa con la progresión acelerada de la EH observada, ya que el aumento de la repetición CAG es evidente en el caso 109 CAG (**Ayyildiz et al., 2023**).

En ambos casos se observan eventos MXE como categoría dominante, en el caso 109 CAG se exhibe la presencia de casos no detectados como los eventos RI y un ligero aumento en casos MXE. Esto sugiere que la mayor expansión de repeticiones CAG está asociada con un efecto más severo y complejo en el empalme alternativo, así como también una mayor disrupción de las funciones neuronales y metabólicas (**Choi et al., 2024**).

3.1.4.2. Visualización de variantes

La visualización en IGV se enfocó en el transcrito del gen HTT (ENST00000355072.11), focalizando las diferencias en variantes detectadas entre las condiciones 71CAG y 109CAG, resaltando su impacto en los exones que conforman el transcrito tal como se ve en la **Figura 4**. El transcrito posee un total de 67 exones, de los cuales los más importantes y críticos para la función y síntesis de la proteína huntingtina, corresponden a los que están cercanos al inicio y final del gen (**L. Liu et al., 2023**).

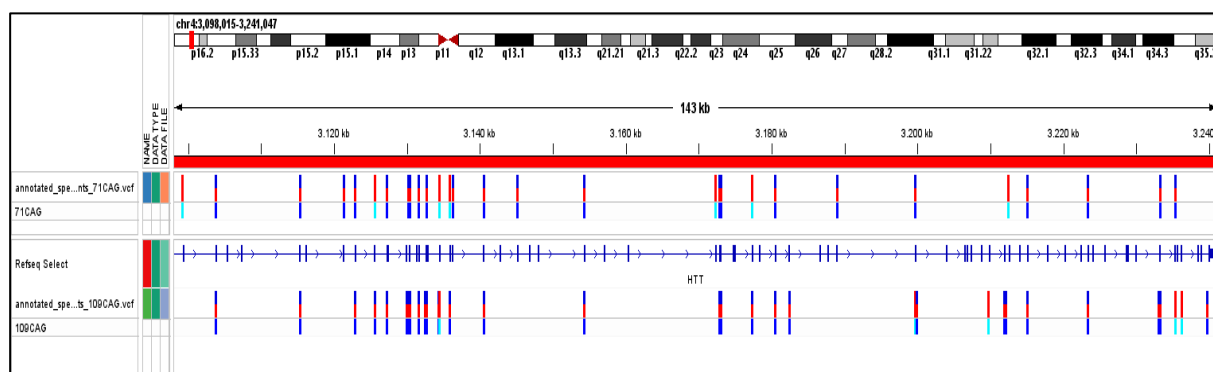
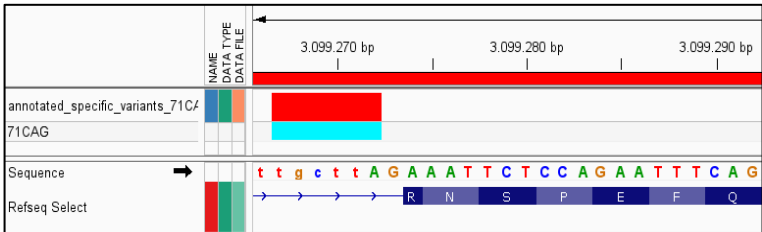
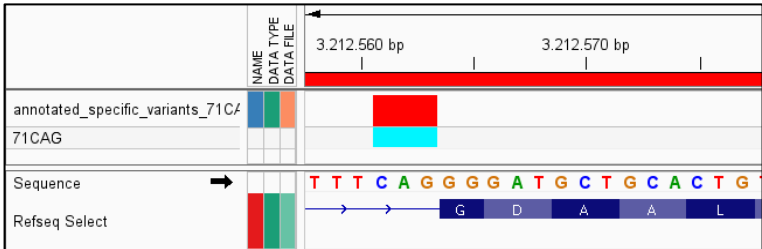
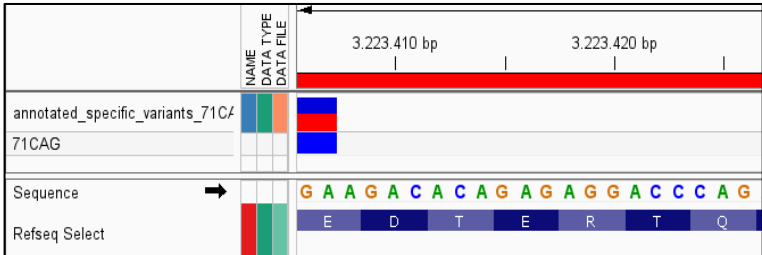


Figura 4. Representación gráfica de las variantes en contraste al genoma de referencia hg38

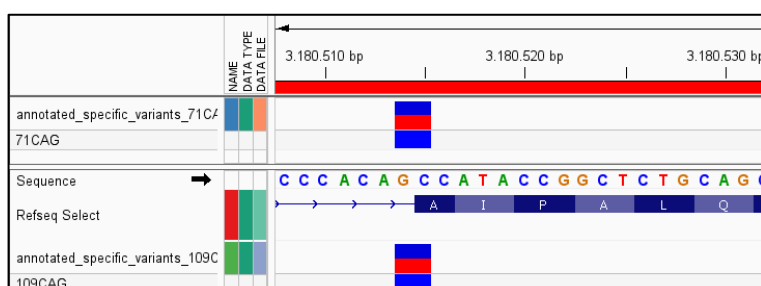
En la **Tabla 9** se observan las variantes en las zonas de interés basadas en el estudio y resultados de **Hillier et al. (2005)**. En 71CAG, la mayoría de las variantes se ubican en exones y regiones aledañas a los mismos, lo que genera un impacto directo en la codificación de la proteína, como la variante en la posición 3.099.260-3.099.272, que afecta al exón 1 del transcrito que está relacionado directamente con los niveles de la huntingtina tóxica (**Sogorb-Gonzalez et al., 2024**). Por su parte, en 109CAG las variantes están más distribuidas en las zonas intrónicas cercanas a los exones iniciales y finales afectando a las regiones donantes y promotoras de splicing, indicando un efecto más global en la transcripción y traducción del gen (**AlQazzaz et al., 2024; Pupak et al., 2023**).

Tabla 9.

Variantes en zonas de interés en las condiciones 71CAG y 109CAG visualizadas en IGV

Variante	Condición	Efecto
Posición: 3.099.260-3.099.272 	71 CAG Referencia: CTTTTTTTGCTTA* Alteración: CTTTTTTTGCTTA	Alteración en la codificación del transcrito.
Posición: 3.212.560-3.212.563 	71 CAG Referencia: TCAG* Alteración: TG	Codificación de la proteína.
Posición: 3.223.405-3.223.407 	71 CAG Referencia: GGA* Alteración: GAAGA	Alteración en el exón codificante.

Posición: 3.180.513-3.180.515



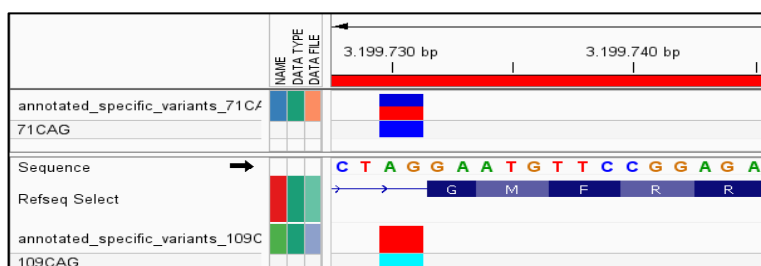
71 CAG y 109 CAG

Referencia: AGC*

Alteración: ATGC

Codificación
de la
proteína.

Posición: 3.199.729-3.199.731



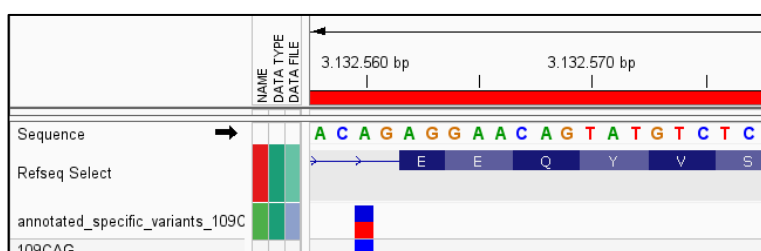
71 CAG y 109 CAG

Referencia: TAG*

Alteración: TG

Alteración
en la
codificación
del
transcrito.

Posición: 3.132.560



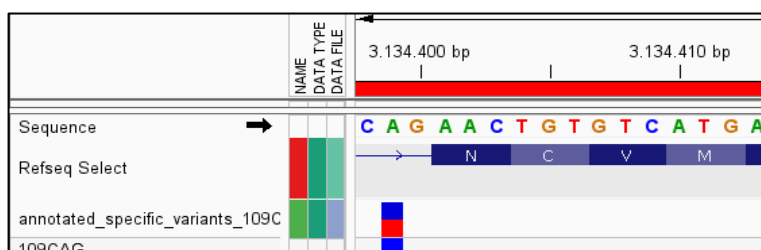
109 CAG

Referencia: A*

Alteración: T

Alteración
en el aceptor
de empalme.

Posición: 3.134.399



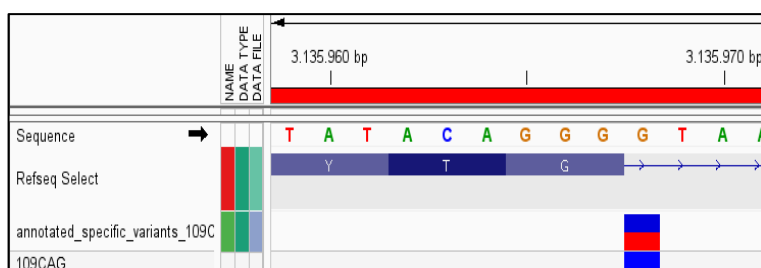
109 CAG

Referencia: A*

Alteración: G

Alteración
en el aceptor
de empalme.

Posición: 3.135.968

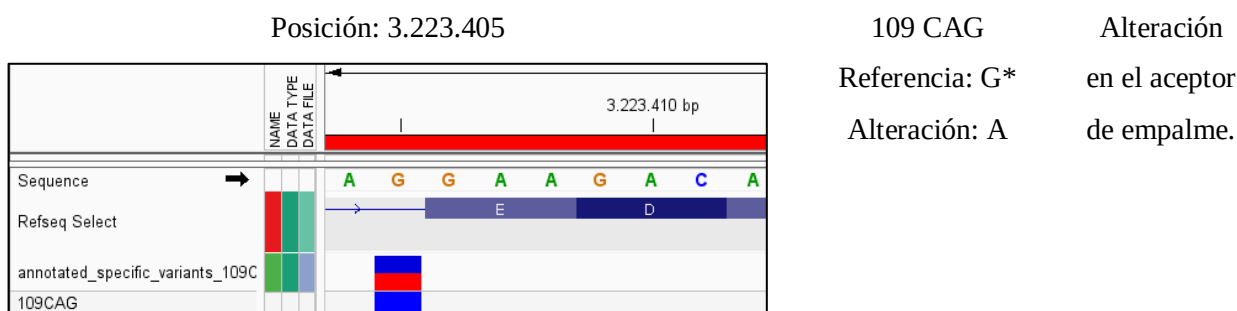


109 CAG

Referencia: G*

Alteración: C

Alteración
en el
donante de
empalme.



Nota: Información obtenida de **Hillier et al. (2005)**.

La visualización confirma que un mayor número de repeticiones CAG está asociado con una mayor densidad y diversidad de variantes genómicas. Esto afecta no solo la región codificante, sino que también interfiere en el splicing y procesamiento del ARN mensajero (**AlQazzaz et al., 2024**). Los eventos de splicing alterados observados en 109CAG podrían explicar las isoformas aberrantes que contribuyen a una mayor toxicidad proteica (**Ayyildiz et al., 2023**).

3.1.5. Construcción y análisis de redes de co-expresión e interacción funcional de los FT

Las redes obtenidas presentadas en la **Tabla 10** muestran las interacciones funcionales y de co-expresión de los FT en relación con los genes diferencialmente expresados obtenidos como resultado del análisis DEG. La red de HSF1, incluye conexiones relacionadas con la regulación del plegamiento proteico y el mantenimiento de la homeostasis celular, sin embargo, el FT aparece como un nodo periférico sin interacción directa con los genes (**C. Liu et al., 2022**).

Por otro lado, las redes de Sp1 y CBP presenta a estos FT como nodos centrales en sus respectivas redes, lo que indica un rol clave de estos en la regulación génica global. La densidad de conexiones y los genes implicados sugieren un rol en procesos homeostáticos neuronales y del sistema nervioso central.

Redes globales de co-expresión e interacción funcional

[illegible]

26

interacción, está involucrado en una regulación más amplia, siendo crucial para la homeostasis celular en condiciones normales y patológicas de la EH (Pogoda et al., 2021).

Tabla 11.

Puntuaciones de los FT y total de conexiones directas de interacción

Código	Nombre de FT	Puntuación	Número de conexiones
H_sapiens_9_-788660	HSF1	-0.478637	0
H_sapiens_8_-772833	CBP	0.593448	3
H_sapiens_10_-788768	Sp1	0.586794	10

Las **Tablas 12 y 13** muestran los nodos principales identificados en las redes correspondientes a los factores de transcripción CBP y Sp1, calculados mediante el plugin CytoHubba de Cytoscape. Este análisis evaluó la importancia de los nodos usando métricas como Bottleneck (Robustez y Flujo), MCC (Conectividad Local) y Betweenness (Centralidad Global) (Meem et al., 2023).

Tabla 12.

Nodos principales en la red de CBP

Rango	Gen	Puntaje
BottleNeck (Robustez y Flujo)		
1	UTY	10.0
2	GSTT1	5.0
2	ZBED1	5.0
MCC (Conectividad Local)		
1	RPS4Y1	951480.0
1	DDX3Y	951480.0
Betweenness (Centralidad Global)		
1	GSTT1	144.9089
2	KDM5D	112.7906

RPS4Y1 y DDX3Y son nodos centrales en términos de conectividad local (MCC) en ambas redes, lo que subraya su importancia funcional compartida, mientras que GSTT1 aparece como un mediador central, lo que indica que podría desempeñar un papel fundamental en varios procesos de la regulación genética en general (Wu et al., 2020). Los nodos identificados con altos valores en Bottleneck, MCC,

y Betweenness representan genes críticos para la funcionalidad de las redes de CREBBP, Sp1 y juegan un papel importante en la EH (El-Shamarka et al., 2022).

Estas diferencias en los patrones de los nodos entre CBP y Sp1 reflejan sus roles funcionales únicos: CBP se asocia más con la modulación epigenética, mientras que Sp1 parece tener un rol más diverso en la regulación génica (Ivanenko et al., 2022; M. Wang et al., 2022).

Tabla 13.

Nodos principales en la red de Sp1

Rango	Gen	Puntaje
BottleNeck (Robustez y Flujo)		
1	KDM5D	5.0
1	ZNF560	5.0
MCC (Conectividad Local)		
1	RPS4Y1	953880.0
2	DDX3Y	953760.0
Betweenness (Centralidad Global)		
1	GSTT1	140.8082
2	ZNF560	102.0103

3.1.6. Interpretación de la influencia de las VS diferenciales

Se identificaron 29 variantes en el gen HTT para la condición 71CAG y 34 variantes en la condición 109CAG, en la comparación Control vs. 71CAG se detectaron 187,472 eventos, para Control vs. 109CAG el total fue de 193,862 eventos, por lo que es evidente el incremento de casos en 109CAG (Tabla 14 y 15). El incremento de variantes genéticas en 109CAG sugiere una relación directa entre la expansión CAG y la acumulación de alteraciones genéticas en el gen HTT (Ayyildiz et al., 2023).

Aunque 109CAG muestra un mayor número de eventos y variantes genéticas, el número de eventos significativos no aumenta proporcionalmente, lo que podría indicar una saturación funcional del sistema de splicing en esta condición relacionada con la regulación de las isoformas transcripcionales debido a la disrupción global del splicing (Tano et al., 2023).

Tabla 14.

Cuantificación de variantes VCF

Condiciones	Variantes totales gen HTT	Variantes significativas
71 CAG	29	5
109 CAG	34	6

Tabla 15.

Cuantificación de eventos rMATS

Comparativas	Total, eventos	Eventos HTT	Eventos significativos
Control vs 71CAG	187472	64	8
Control vs 109CAG	193862	48	7

La **Tabla 16** muestra la comparación entre las variantes detectadas en las condiciones 71CAG y 109CAG del gen HTT con las variantes existentes en bases de datos genéticas. Entre todas estas, la variante específica (chr4:3135968_G/C) tiene correspondencia en la base de datos Ensembl con el código rs1278226838. Las variantes no reportadas previamente sugieren que podrían ser específicas del contexto patológico de la EH, posiblemente relacionadas con los mecanismos moleculares desencadenados por la expansión de repeticiones CAG (Ayyildiz et al., 2023; Choi et al., 2024).

La identificación de variantes únicas en 71CAG y 109CAG subraya la importancia de realizar un análisis detallado para identificar objetivos terapéuticos específicos, que podrían ser clave para comprender los mecanismos moleculares subyacentes a la progresión de la enfermedad y diseñar estrategias de intervención personalizadas.

Tabla 16.

Correlación variantes halladas con existentes

Variante detectada	Variante existente	
	Código acceso	Variante
71CAG		
chr4:3180513_AGC/ATGC	N/A	N/A
chr4:3099260_CTTTTTTTGCTTA/CTTTTTTGCTTA	N/A	N/A
chr4:3212560_TCAG/TG	N/A	N/A
chr4:3199729_TAG/TG	N/A	N/A
chr4:3223405_GGA/GAAGA	N/A	N/A

109CAG		
chr4:3135968_G/C	rs1278226838	chr4:3135968_G/C
chr4:3223405_G/A	N/A	N/A
chr4:3180513_AGC/ATGC	N/A	N/A
chr4:3132560_A/T	N/A	N/A
chr4:3134399_A/G	N/A	N/A
chr4:3199729_TAG/TG	N/A	N/A

3.1.7. Interpretación de la influencia de los FT

La **Tabla 12** presenta un análisis detallado de las redes específicas construidas en Cytoscape para los factores de transcripción Sp1 y CBP. Los datos se organizan por peso, tipo de interacción, y si existe o no un contraste con redes existentes.

Las redes de CBP están dominadas por interacciones de co-expresión y co-localización, lo que resalta su papel como un integrador de procesos epigenéticos y moleculares en múltiples contextos biológicos (**M. Wang et al., 2022**), como ejemplo significativo: La interacción entre los nodos H_sapiens__8__799470 y H_sapiens__8__7723831 tiene un peso de 0.00256711753037, siendo validada en redes existentes. Esto indica una co-expresión relevante y previamente estudiada. El peso de las conexiones es moderado pero significativo, lo que sugiere interacciones relevantes en procesos relacionados con la regulación transcripcional (**Arancibia-Opazo et al., 2023; Z.-M. Wang et al., 2020**).

Las redes de Sp1 presentan interacciones más variadas, con énfasis en interacciones genéticas y físicas. A diferencia de CREBBP, Sp1 tiene conexiones de mayor peso (indicando mayor relevancia relativa), pero muchas de estas no han sido validadas en estudios previos, lo que podría implicar nuevas rutas de investigación (**Ivanenko et al., 2022**). La inclusión de interacciones físicas y genéticas sugiere que Sp1 desempeña un papel crítico en procesos que van más allá de la regulación génica, posiblemente afectan la señalización celular y la codificación proteica (**Heinz et al., 2021**).

Tabla 17.

Correlación de redes existentes con las obtenidas

Código	Peso	Tipo	Contraste con red existente
CBP			
H__sapiens__8_- 799470 H__sapiens__8_- 772833	0.00256711750837 21233	Co – expresión	SI
H__sapiens__8_- 774353 H__sapiens__8_- 772833	0.00154912663842 9194	Co – expresión	SI
H__sapiens__8_- 778323 H__sapiens__8_- 772833	0.00088371001991 1505	Co – localización	SI
Sp1			
H__sapiens__10_- 778323 H__sapiens__10_- 788768	3.29387349853550 5E-8	Interacciones genéticas	SI
H__sapiens__10_- 788768 H__sapiens__10_- 779023	0.00675337965707 2836	Interacciones genéticas	NO
H__sapiens__10_- 789182 H__sapiens__10_- 788768	0.00675337965707 2836	Interacciones genéticas	NO
H__sapiens__10_- 789182 H__sapiens__10_- 788768	0.00281924606463 3025	Interacciones físicas	NO
H__sapiens__10_- 778323 H__sapiens__10_- 779023	0.00112413875512 329	Co – expresión	SI

Nota: (Elaborado por el autor).

CAPÍTULO IV.- CONCLUSIONES Y RECOMENDACIONES

4.1. Conclusiones

Esta investigación demuestra que la evolución de la EH está profundamente influenciada por las alteraciones en el splicing del gen HTT y las redes reguladoras de factores de transcripción clave. Los hallazgos subrayan la importancia de analizar estos elementos de manera integrada, abriendo oportunidades para el diseño de estrategias terapéuticas específicas que se enfoquen tanto en la corrección del splicing como en la modulación de los factores de transcripción implicados. Estos resultados aportan una perspectiva molecular más clara sobre los mecanismos subyacentes a la progresión de la EH y su heterogeneidad clínica.

El análisis de la expresión génica basado en datos RNA-seq permitió identificar alteraciones significativas en la regulación diferencial de genes asociadas a la EH. El incremento en las repeticiones CAG se correlaciona con una progresiva disfunción molecular, evidenciada en cambios en la expresión génica y splicing alternativo, así como en la producción de isoformas aberrantes.

Se logró identificar la influencia de las variantes de splicing diferenciales y factores de transcripción en los genes expresados en la EH, evidenciando que juegan un papel clave en los mecanismos moleculares subyacentes a la patogénesis de la enfermedad, el estudio confirma la influencia central de las variantes de splicing diferenciales y los factores de transcripción en la disfunción molecular observada en la enfermedad.

Este estudio destaca la relevancia de las nuevas variantes de splicing y los FT identificados como elementos centrales en la progresión de la EH, lo que ofrece nuevas perspectivas para la comprensión de su patogénesis y el desarrollo de estrategias terapéuticas dirigidas. La identificación de variantes específicas y nodos funcionales clave proporcionan un marco para futuras investigaciones orientadas a la modulación de estos procesos con fines terapéuticos.

4.2. Recomendaciones

- Ampliar la investigación mediante experimentos funcionales adicionales que validen el impacto de las variantes de splicing detectadas, particularmente en la condición 109CAG, donde se identificaron eventos significativos en regiones intrónicas y exónicas. Estas pruebas podrían incluir el uso de técnicas como la RT-PCR y la secuenciación de isoformas específicas para confirmar la presencia de transcritos aberrantes. Esto permitirá una comprensión más profunda de cómo estas variantes contribuyen a la disfunción molecular observada en la EH.
- Dado el rol central de los factores de transcripción como Sp1 y CREBBP en las redes de interacción, se recomienda la creación de modelos in vitro, como líneas celulares modificadas genéticamente, para evaluar como estos FT regulan la expresión génica bajo diferentes condiciones de expansión de repeticiones CAG. Esto facilitará la identificación de rutas específicas reguladas por los FT y sus posibles modulaciones terapéuticas.
- Incorporar algoritmos de aprendizaje automático que combinen los datos de variantes detectadas y los eventos de splicing con bases de datos existentes, permitiendo priorizar aquellas variantes con mayor probabilidad de impacto funcional. Esto optimizará el diseño de futuros experimentos y reducirá los costos asociados con la validación de todas las variantes detectadas, especialmente aquellas no reportadas previamente en bases de datos públicas.

MATERIALES DE REFERENCIA

Referencias bibliográficas

- Ajitkumar, A., & De Jesus, O. (2024). Huntington Disease. En *StatPearls*. StatPearls Publishing. <http://www.ncbi.nlm.nih.gov/books/NBK559166/>
- AlQazzaz, M. A., Ciamponi, F. E., Ho, J. C., Maron, M. I., Yadav, M., Sababi, A. M., MacLeod, G., Ahmadi, M., Bullivant, G., Tano, V., Langley, S. R., Sánchez-Osuna, M., Sachamitr, P., Kushida, M., Richards, L., Bardile, C. F., Pouladi, M. A., Pugh, T., Tyers, M., ... Prinos, P. (2024). *PRMT5 is required for full-length HTT expression by repressing multiple proximal intronic polyadenylation sites*. <https://doi.org/10.1101/2024.03.14.584861>
- Arancibia-Opazo, S., Contreras-Riquelme, J. S., Sánchez, M., Cisternas-Olmedo, M., Vidal, R. L., Martin, A. J. M., & Sáez, M. A. (2023). Transcriptional and Histone Acetylation Changes Associated with CRE Elements Expose Key Factors Governing the Regulatory Circuit in the Early Stage of Huntington's Disease Models. *International Journal of Molecular Sciences*, 24(13), 10848. <https://doi.org/10.3390/ijms241310848>
- Aslesh, T., & Yokota, T. (2020). Development of Antisense Oligonucleotide Gapmers for the Treatment of Huntington's Disease. En T. Yokota & R. Maruyama (Eds.), *Gapmers* (Vol. 2176, pp. 57–67). Springer US. https://doi.org/10.1007/978-1-0716-0771-8_4
- Ayyildiz, D., Bergonzoni, G., Monziani, A., Tripathi, T., Döring, J., Kerschbamer, E., Di Leva, F., Pennati, E., Donini, L., Kovalenko, M., Zasso, J., Conti, L., Wheeler, V. C., Dieterich, C., Piazza, S., Dassi, E., & Biagioli, M. (2023). CAG repeat expansion in the Huntington's disease gene shapes linear and circular RNAs biogenesis. *PLOS Genetics*, 19(10), e1010988. <https://doi.org/10.1371/journal.pgen.1010988>
- Bachoud-Lévi, A.-C., Ferreira, J., Massart, R., Youssov, K., Rosser, A., Busse, M., Craufurd, D., Reilmann, R., De Michele, G., Rae, D., Squitieri, F., Seppi, K., Perrine, C., Scherer-Gagou, C., Audrey, O., Verny, C., & Burgunder, J.-M. (2019). International Guidelines for the Treatment of Huntington's Disease. *Frontiers in Neurology*, 10, 710. <https://doi.org/10.3389/fneur.2019.00710>
- Bensalel, J., Xu, H., Lu, M. L., Capobianco, E., & Wei, J. (2021). RNA-seq analysis reveals significant transcriptome changes in huntingtin-null human neuroblastoma cells. *BMC Medical Genomics*, 14(1), 176. <https://doi.org/10.1186/s12920-021-01022-w>
- Browne, P. D., Nielsen, T. K., Kot, W., Aggerholm, A., Gilbert, M. T. P., Puetz, L., Rasmussen, M., Zervas, A., & Hansen, L. H. (2020). GC bias affects genomic and metagenomic reconstructions, underrepresenting GC-poor organisms. *GigaScience*, 9(2), giaa008. <https://doi.org/10.1093/gigascience/giaa008>

- Buccitelli, C., & Selbach, M. (2020). mRNAs, proteins and the emerging principles of gene expression control. *Nature Reviews Genetics*, 21(10), 630–644. <https://doi.org/10.1038/s41576-020-0258-4>
- Burchfiel, E. T., Vihervaara, A., Guertin, M. J., Gomez-Pastor, R., & Thiele, D. J. (2021). Comparative interactomes of HSF1 in stress and disease reveal a role for CTCF in HSF1-mediated gene regulation. *Journal of Biological Chemistry*, 296, 100097. <https://doi.org/10.1074/jbc.RA120.015452>
- Camargo, A. (2022). PCAtest: Testing the statistical significance of Principal Component Analysis in R. *PeerJ*, 10, e12967. <https://doi.org/10.7717/peerj.12967>
- Choi, D. E., Shin, J. W., Zeng, S., Hong, E. P., Jang, J.-H., Loupe, J. M., Wheeler, V. C., Stutzman, H. E., Kleinstiver, B., & Lee, J.-M. (2024). Base editing strategies to convert CAG to CAA diminish the disease-causing mutation in Huntington's disease. *eLife*, 12, RP89782. <https://doi.org/10.7554/eLife.89782.2>
- Christodoulou, C. C., & Papanicolaou, E. Z. (2023). Integrated Bioinformatics Analysis of Shared Genes, miRNA, Biological Pathways and Their Potential Role as Therapeutic Targets in Huntington's Disease Stages. *International Journal of Molecular Sciences*, 24(5), 4873. <https://doi.org/10.3390/ijms24054873>
- Colpo, G. D., Rocha, N. P., Furr Stimming, E., & Teixeira, A. L. (2020). Gene Expression Profiling in Huntington's Disease: Does Comorbidity with Depressive Symptoms Matter? *International Journal of Molecular Sciences*, 21(22), 8474. <https://doi.org/10.3390/ijms21228474>
- Croce, K. R., & Yamamoto, A. (2019). A role for autophagy in Huntington's disease. *Neurobiology of Disease*, 122, 16–22. <https://doi.org/10.1016/j.nbd.2018.08.010>
- De Sena Brandine, G., & Smith, A. D. (2021). Falco: High-speed FastQC emulation for quality control of sequencing data. *F1000Research*, 8, 1874. <https://doi.org/10.12688/f1000research.21142.2>
- Dong, X., & Cong, S. (2018). Identification of differentially expressed genes and regulatory relationships in Huntington's disease by bioinformatics analysis. *Molecular Medicine Reports*. <https://doi.org/10.3892/mmr.2018.8410>
- Elorza, A., Márquez, Y., Cabrera, J. R., Sánchez-Trincado, J. L., Santos-Galindo, M., Hernández, I. H., Picó, S., Díaz-Hernández, J. I., García-Escudero, R., Irimia, M., & Lucas, J. J. (2021). Huntington's disease-specific mis-splicing unveils key effector genes and altered splicing factors. *Brain*, 144(7), 2009–2023. <https://doi.org/10.1093/brain/awab087>
- El-Shamarka, M. E.-S., El-Sahar, A. E., Saad, M. A., Assaf, N., & Sayed, R. H. (2022). Inosine attenuates 3-nitropropionic acid-induced Huntington's disease-like symptoms in rats via the activation of the A2AR/BDNF/TrKB/ERK/CREB signaling pathway. *Life Sciences*, 300, 120569. <https://doi.org/10.1016/j.lfs.2022.120569>
- Fienko, S., Landles, C., Sathasivam, K., McAteer, S. J., Milton, R. E., Osborne, G. F., Smith, E. J., Jones, S. T., Bondulich, M. K., Danby, E. C. E., Phillips, J., Taxy, B. A., Kordasiewicz, H. B.,

- & Bates, G. P. (2022). Alternative processing of human *HTT* mRNA with implications for Huntington's disease therapeutics. *Brain*, 145(12), 4409–4424. <https://doi.org/10.1093/brain/awac241>
- Gallardo, C., & Delisle, L. (2023, junio 12). *Genome-wide alternative splicing analysis* [Text]. Galaxy Training Network; Galaxy Training Network. <https://training.galaxyproject.org/training-material/topics/transcriptomics/tutorials/differential-isoform-expression/tutorial.html>
- Gatto, E. M., Rojas, N. G., Persi, G., Etcheverry, J. L., Cesarini, M. E., & Perandones, C. (2020). Huntington disease: Advances in the understanding of its mechanisms. *Clinical Parkinsonism & Related Disorders*, 3, 100056. <https://doi.org/10.1016/j.prdoa.2020.100056>
- Handsaker, R. E., Kashin, S., Reed, N. M., Tan, S., Lee, W.-S., McDonald, T. M., Morris, K., Kamitaki, N., Mullally, C. D., Morakabati, N., Goldman, M., Lind, G., Kohli, R., Lawton, E., Hogan, M., Ichihara, K., Berretta, S., & McCarroll, S. A. (2024). *Long somatic DNA-repeat expansion drives neurodegeneration in Huntington disease*. <https://doi.org/10.1101/2024.05.17.592722>
- Heinz, A., Schilling, J., Van Roon-Mom, W., & Krauß, S. (2021). The MID1 Protein: A Promising Therapeutic Target in Huntington's Disease. *Frontiers in Genetics*, 12, 761714. <https://doi.org/10.3389/fgene.2021.761714>
- Hillier, L. W., Graves, T. A., Fulton, R. S., Fulton, L. A., Pepin, K. H., Minx, P., Wagner-McPherson, C., Layman, D., Wylie, K., Sekhon, M., Becker, M. C., Fewell, G. A., Delehaunty, K. D., Miner, T. L., Nash, W. E., Kremitzki, C., Oddy, L., Du, H., Sun, H., ... Wilson, R. K. (2005). Generation and annotation of the DNA sequences of human chromosomes 2 and 4. *Nature*, 434(7034), 724–731. <https://doi.org/10.1038/nature03466>
- Hyeon, J. W., Kim, A. H., & Yano, H. (2021). Epigenetic regulation in Huntington's disease. *Neurochemistry International*, 148, 105074. <https://doi.org/10.1016/j.neuint.2021.105074>
- Intihar, T. A., Martinez, E. A., & Gomez-Pastor, R. (2019). Mitochondrial Dysfunction in Huntington's Disease; Interplay Between HSF1, p53 and PGC-1 α Transcription Factors. *Frontiers in Cellular Neuroscience*, 13, 103. <https://doi.org/10.3389/fncel.2019.00103>
- Ivanenko, K. A., Prassolov, V. S., & Khabusheva, E. R. (2022). Transcription Factor Sp1 in the Expression of Genes Encoding Components of Mapk, JAK/STAT, and PI3K/Akt Signaling Pathways. *Molecular Biology*, 56(5), 756–769. <https://doi.org/10.1134/S0026893322050089>
- Jiang, M., Zhang, S., Yin, H., Zhuo, Z., & Meng, G. (2023). A comprehensive benchmarking of differential splicing tools for RNA-seq analysis at the event level. *Briefings in Bioinformatics*, 24(3), bbad121. <https://doi.org/10.1093/bib/bbad121>
- Kim, C., Yousefian-Jazi, A., Choi, S.-H., Chang, I., Lee, J., & Ryu, H. (2021). Non-Cell Autonomous and Epigenetic Mechanisms of Huntington's Disease. *International Journal of Molecular Sciences*, 22(22), 12499. <https://doi.org/10.3390/ijms222212499>

- Kim, H., & Gomez-Pastor, R. (2022). HSF1 and Its Role in Huntington's Disease Pathology. En K. Turksen (Ed.), *Cell Biology and Translational Medicine, Volume 19* (Vol. 1410, pp. 35–95). Springer Nature Switzerland. https://doi.org/10.1007/5584_2022_742
- Krach, F., Stemick, J., Boerstler, T., Weiss, A., Lingos, I., Reischl, S., Meixner, H., Ploetz, S., Farrell, M., Hehr, U., Kohl, Z., Winner, B., & Winkler, J. (2022). An alternative splicing modulator decreases mutant HTT and improves the molecular fingerprint in Huntington's disease patient neurons. *Nature Communications*, 13(1), 6797. <https://doi.org/10.1038/s41467-022-34419-x>
- Li, Y., Ge, X., Peng, F., Li, W., & Li, J. J. (2022). Exaggerated false positives by popular differential expression methods when analyzing human population samples. *Genome Biology*, 23(1), 79. <https://doi.org/10.1186/s13059-022-02648-4>
- Liang, D., Yu, J., Song, T., Zhang, R., Du, Y., Yu, M., Cao, H., Pan, X., Qiao, J., Liu, Y., Qi, Z., & Liu, Y. (2022). A Genome-Wide Alternative Splicing Landscape Specifically Associated with Durable Rice Blast Resistance. *Agronomy*, 12(10), 2414. <https://doi.org/10.3390/agronomy12102414>
- Lin, L., Park, J. W., Ramachandran, S., Zhang, Y., Tseng, Y.-T., Shen, S., Waldvogel, H. J., Curtis, M. A., Faull, R. L. M., Troncoso, J. C., Pletnikova, O., Ross, C. A., Davidson, B. L., & Xing, Y. (2016). Transcriptome sequencing reveals aberrant alternative splicing in Huntington's disease. *Human Molecular Genetics*, 25(16), 3454–3466. <https://doi.org/10.1093/hmg/ddw187>
- Liu, C., Fu, Z., Wu, S., Wang, X., Zhang, S., Chu, C., Hong, Y., Wu, W., Chen, S., Jiang, Y., Wu, Y., Song, Y., Liu, Y., & Guo, X. (2022). Mitochondrial HSF1 triggers mitochondrial dysfunction and neurodegeneration in Huntington's disease. *EMBO Molecular Medicine*, 14(7), e15851. <https://doi.org/10.15252/emmm.202215851>
- Liu, L., Malagu, K., Haughan, A. F., Khetarpal, V., Stott, A. J., Esmieu, W., Vater, H. D., Webster, S. J., Van De Poël, A. J., Clissold, C., Cosgrove, B., Sutton, B., Spencer, J. A., Breccia, P., Gancia, E., Bonomo, S., Ladduwahetty, T., Lazari, O., Patel, H., ... Dominguez, C. (2023). Identification and Optimization of RNA-Splicing Modulators as Huntingtin Protein-Lowering Agents for the Treatment of Huntington's Disease. *Journal of Medicinal Chemistry*, 66(18), 13205–13246. <https://doi.org/10.1021/acs.jmedchem.3c01173>
- Liu, S., Wang, Z., Zhu, R., Wang, F., Cheng, Y., & Liu, Y. (2021). Three Differential Expression Analysis Methods for RNA Sequencing: Limma, EdgeR, DESeq2. *Journal of Visualized Experiments*, 175, 62528. <https://doi.org/10.3791/62528-v>
- Livesey, B. J., & Marsh, J. A. (2022). Interpreting protein variant effects with computational predictors and deep mutational scanning. *Disease Models & Mechanisms*, 15(6), dmm049510. <https://doi.org/10.1242/dmm.049510>
- Małacka, A., Østlie, I., Trøen, G., Małacki, J., Delabie, J., Tierens, A., Munthe, L. A., Berentsen, S., & Tjønnfjord, G. E. (2024). Gene expression analysis revealed downregulation of complement

- receptor 1 in clonal B cells in cold agglutinin disease. *Clinical and Experimental Immunology*, 216(1), 45–54. <https://doi.org/10.1093/cei/uxad135>
- Malla, B., Guo, X., Senger, G., Chasapopoulou, Z., & Yildirim, F. (2021). A Systematic Review of Transcriptional Dysregulation in Huntington's Disease Studied by RNA Sequencing. *Frontiers in Genetics*, 12, 751033. <https://doi.org/10.3389/fgene.2021.751033>
- Mason, M. A., Gomez-Paredes, C., Sathasivam, K., Neueder, A., Papadopoulou, A.-S., & Bates, G. P. (2020). Silencing Srsf6 does not modulate incomplete splicing of the huntingtin gene in Huntington's disease models. *Scientific Reports*, 10(1), 14057. <https://doi.org/10.1038/s41598-020-71111-w>
- McDermaid, A., Monier, B., Zhao, J., Liu, B., & Ma, Q. (2019). Interpretation of differential gene expression results of RNA-seq data: Review and integration. *Briefings in Bioinformatics*, 20(6), 2044–2054. <https://doi.org/10.1093/bib/bby067>
- Meem, T. M., Khan, U., Mredul, M. B. R., Awal, M. A., Rahman, M. H., & Khan, M. S. (2023). A Comprehensive Bioinformatics Approach to Identify Molecular Signatures and Key Pathways for the Huntington Disease. *Bioinformatics and Biology Insights*, 17, 11779322231210098. <https://doi.org/10.1177/11779322231210098>
- Nguyen, T. B., Miramontes, R., Chillon-Marin, C., Maimon, R., Vazquez-Sanchez, S., Lau, A. L., McClure, N. R., England, W. E., Singha, M., Stocksdales, J. T., Jang, K.-H., Jung, S., McKnight, J. I., Ho, L. N., Faull, R. L. M., Steffan, J. S., Reidling, J. C., Jang, C., Lee, G., ... Thompson, L. M. (2023). *Aberrant splicing in Huntington's disease via disrupted TDP-43 activity accompanied by altered m6A RNA modification*. <https://doi.org/10.1101/2023.10.31.565004>
- Oka, M., Xu, L., Suzuki, T., Yoshikawa, T., Sakamoto, H., Uemura, H., Yoshizawa, A. C., Suzuki, Y., Nakatsura, T., Ishihama, Y., Suzuki, A., & Seki, M. (2021). Aberrant splicing isoforms detected by full-length transcriptome sequencing as transcripts of potential neoantigens in non-small cell lung cancer. *Genome Biology*, 22(1), 9. <https://doi.org/10.1186/s13059-020-02240-8>
- Paul, B. D., Sbodio, J. I., & Snyder, S. H. (2022). Mutant Huntingtin Derails Cysteine Metabolism in Huntington's Disease at Both Transcriptional and Post-Translational Levels. *Antioxidants*, 11(8), 1470. <https://doi.org/10.3390/antiox11081470>
- Pogoda, A., Chmielewska, N., Maciejak, P., & Szyndler, J. (2021). Transcriptional Dysregulation in Huntington's Disease: The Role in Pathogenesis and Potency for Pharmacological Targeting. *Current Medicinal Chemistry*, 28(14), 2783–2806. <https://doi.org/10.2174/0929867327666200705225821>
- Pupak, A., Navarro, I. R., Sathasivam, K., Essmann, A., Singh, A., Del Toro, D., Ginés, S., Bates, G. P., Vang Ørom, U. A., Marti, E., & Brito, V. (2023). *m6A RNA modification of m Htt intron 1 regulates the generation of Htt1a in Huntington's Disease*. <https://doi.org/10.1101/2023.11.10.566530>

- Qu, J., Yang, D., Wu, Y., & Yu, H. (2022). *The reduced GTH-gene expression caused by mHtt on Sp1 will ultimately result in cell death*. 020032. <https://doi.org/10.1063/5.0112750>
- Robinson, J. T., Thorvaldsdóttir, H., Wenger, A. M., Zehir, A., & Mesirov, J. P. (2017). Variant Review with the Integrative Genomics Viewer. *Cancer Research*, 77(21), e31–e34. <https://doi.org/10.1158/0008-5472.CAN-17-0337>
- Saif, R., Ejaz, A., Mahmood, T., & Zia, S. (2021). *Differential Gene Expression Pipeline for Whole Transcriptome RNA-Seq Data using Personal Computer*. <https://doi.org/10.1101/2021.01.26.428352>
- Seefelder, M., & Kochanek, S. (2021). A meta-analysis of transcriptomic profiles of Huntington's disease patients. *PLOS ONE*, 16(6), e0253037. <https://doi.org/10.1371/journal.pone.0253037>
- Sepulveda, J. L. (2020). Using R and Bioconductor in Clinical Genomics and Transcriptomics. *The Journal of Molecular Diagnostics*, 22(1), 3–20. <https://doi.org/10.1016/j.jmoldx.2019.08.006>
- Shahib, A. K., Rastegar, M., Van Wijnen, A. J., & Davie, J. R. (2024). Neurodevelopmental functions and activities of the KAT3 class of lysine acetyltransferases. *Biochemistry and Cell Biology*, 102(6), 430–447. <https://doi.org/10.1139/bcb-2024-0156>
- Shen, S., Park, J. W., Lu, Z., Lin, L., Henry, M. D., Wu, Y. N., Zhou, Q., & Xing, Y. (2014). rMATS: Robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. *Proceedings of the National Academy of Sciences*, 111(51). <https://doi.org/10.1073/pnas.1419161111>
- Sneha, N. P., Dharshini, S. A. P., Taguchi, Y.-H., & Gromiha, M. M. (2022). Integrative Meta-Analysis of Huntington's Disease Transcriptome Landscape. *Genes*, 13(12), 2385. <https://doi.org/10.3390/genes13122385>
- Sogorb-Gonzalez, M., Landles, C., Caron, N. S., Stam, A., Osborne, G., Hayden, M. R., Howland, D., Van Deventer, S., Bates, G. P., Vallès, A., & Evers, M. (2024). Exon 1-targeting miRNA reduces the pathogenic exon 1 HTT protein in Huntington's disease models. *Brain*, 147(12), 4043–4055. <https://doi.org/10.1093/brain/awae266>
- Spies, J., Covarrubias-Pinto, A., Carcamo, C., Arancibia, Y., Salazar, F., Paredes-Martinez, C., Otth, C., Castro, M., & Zambrano, A. (2023). Modulation of Synaptic Plasticity Genes Associated to DNA Damage in a Model of Huntington's Disease. *Neurochemical Research*, 48(7), 2093–2103. <https://doi.org/10.1007/s11064-023-03889-w>
- Stark, R., Grzelak, M., & Hadfield, J. (2019). RNA sequencing: The teenage years. *Nature Reviews Genetics*, 20(11), 631–656. <https://doi.org/10.1038/s41576-019-0150-2>
- Świtońska, K., Szlachcic, W. J., Handschuh, L., Wojciechowski, P., Marczak, Ł., Stelmaszczuk, M., Figlerowicz, M., & Figiel, M. (2019). Identification of Altered Developmental Pathways in Human Juvenile HD iPSC With 71Q and 109Q Using Transcriptome Profiling. *Frontiers in Cellular Neuroscience*, 12, 528. <https://doi.org/10.3389/fncel.2018.00528>

- Tabrizi, S. J., Flower, M. D., Ross, C. A., & Wild, E. J. (2020). Huntington disease: New insights into molecular pathogenesis and therapeutic opportunities. *Nature Reviews Neurology*, 16(10), 529–546. <https://doi.org/10.1038/s41582-020-0389-4>
- Tano, V., Utami, K. H., Yusof, N. A. B. M., Bégin, J., Tan, W. W. L., Pouladi, M. A., & Langley, S. R. (2023). Widespread dysregulation of mRNA splicing implicates RNA processing in the development and progression of Huntington's disease. *eBioMedicine*, 94, 104720. <https://doi.org/10.1016/j.ebiom.2023.104720>
- Truty, R., Ouyang, K., Rojahn, S., Garcia, S., Colavin, A., Hamlington, B., Freivogel, M., Nussbaum, R. L., Nykamp, K., & Aradhya, S. (2021). Spectrum of splicing variants in disease genes and the ability of RNA analysis to reduce uncertainty in clinical interpretation. *The American Journal of Human Genetics*, 108(4), 696–708. <https://doi.org/10.1016/j.ajhg.2021.03.006>
- Wagner, D. D., Carleton, H. A., Trees, E., & Katz, L. S. (2021). Evaluating whole-genome sequencing quality metrics for enteric pathogen outbreaks. *PeerJ*, 9, e12446. <https://doi.org/10.7717/peerj.12446>
- Wang, M., Chen, Z., & Zhang, Y. (2022). CBP /p300 and HDAC activities regulate H3K27 acetylation dynamics and zygotic genome activation in mouse preimplantation embryos. *The EMBO Journal*, 41(22), e112012. <https://doi.org/10.15252/embj.2022112012>
- Wang, Z.-M., Dong, X.-Y., & Cong, S.-Y. (2020). Bioinformatic analysis of a microRNA regulatory network in Huntington's disease. *Journal of Integrative Neuroscience*, 19(4), 641. <https://doi.org/10.31083/j.jin.2020.04.203>
- Wu, Z., Parry, M., Hou, X.-Y., Liu, M.-H., Wang, H., Cain, R., Pei, Z.-F., Chen, Y.-C., Guo, Z.-Y., Abhijeet, S., & Chen, G. (2020). Gene therapy conversion of striatal astrocytes into GABAergic neurons in mouse models of Huntington's disease. *Nature Communications*, 11(1), 1105. <https://doi.org/10.1038/s41467-020-14855-3>
- Yan, J., Qiu, Y., Ribeiro Dos Santos, A. M., Yin, Y., Li, Y. E., Vinckier, N., Nariai, N., Benaglio, P., Raman, A., Li, X., Fan, S., Chiou, J., Chen, F., Frazer, K. A., Gaulton, K. J., Sander, M., Taipale, J., & Ren, B. (2021). Systematic analysis of binding of transcription factors to noncoding variants. *Nature*, 591(7848), 147–151. <https://doi.org/10.1038/s41586-021-03211-0>

ANEXOS

Anexo 1. Muestras de cada condición experimental filtradas representativas en el análisis DEG

Condición	Código de muestra
Control	SRR8393695
	SRR8393697
Repetición de 71 CAG	SRR8393698
	SRR8393700
Repetición de 109 CAG	SRR8393701
	SRR8393703

Nota: (Elaborado por el autor).

Anexo 2. Script empleado en el análisis de expresión diferencial (DEG) y representación gráfica

- **Activación de paquetes**

```
library(annotate)
library (BiocManager)
library (GOstats)
library(gplots)
library (R. utils)
library (Rbowtie2)
library (Rsamtools)
library (Rsubread)
library (DESeq2)
library(dplyr)
library(ggplot2)
library(ggrepel)
library(heatmap3)
library(pheatmap)
```

- **Creación de índices en el genoma de referencia mediante el paquete Bowtie2**

```
> bowtie2_build ("chromosomas/GCF_000001405.40_GRCh38.p14_genomic.fna", bt2Index =
"chromosomas/human")
```

- **Alineamiento de los archivos FASTQ sobre el genoma de referencia**

```
> dir. create("bam")
> bowtie2(bt2Index = " chromosomas/human ", samOutput = "bam/ SRR8393695_control.sam ", seq1
= " SRR8393695_control iPSC line / Trim_SRR8393695_1_(R1_paired).fastq ", seq2 = "
SRR8393695_control iPSC line / Trim_SRR8393695_2_(R2_paired).fastq ", "--threads=3")
```

```

> command = "C:/Users/User/Documents/R/win-library/4.0/Rbowtie2/bowtie2-align-s" --threads 3 -x
"chromosomas/human" -1 " SRR8393695_control iPSC line /Trim_SRR8393695_1_(R1_paired).fastq"
-2 " SRR8393695_control iPSC line /Trim_SRR8393695_2_(R2_paired).fastq" -S
"bam/SRR8393695_control.sam"

> result=system(command, intern = TRUE, show.output.on.console = TRUE)

> command = "C:/Users/User/Documents/R/win-library/4.0/Rbowtie2/bowtie2-align-s" --threads 3 -x
"chromosomas/human" -1 " SRR8393696_control iPSC line /Trim_SRR8393696_1_(R1_paired).fastq"
-2 " SRR8393696_control iPSC line /Trim_SRR8393696_2_(R2_paired).fastq" -S
"bam/SRR8393696_control.sam"

> result=system(command, intern = TRUE, show.output.on.console = TRUE)

> command = "C:/Users/User/Documents/R/win-library/4.0/Rbowtie2/bowtie2-align-s" --threads 3 -x
"chromosomas/human" -1 " SRR8393697_control iPSC line /Trim_SRR8393697_1_(R1_paired).fastq"
-2 " SRR8393697_control iPSC line /Trim_SRR8393697_2_(R2_paired).fastq" -S
"bam/SRR8393697_control.sam"

> result=system(command, intern = TRUE, show.output.on.console = TRUE)

> command = "C:/Users/User/Documents/R/win-library/4.0/Rbowtie2/bowtie2-align-s" --threads 3 -x
"chromosomas/human" -1 " SRR8393698_clonal iPSC line with 71 CAG
repeats/Trim_SRR8393698_1_(R1_paired).fastq" -2 " SRR8393698_clonal iPSC line with 71 CAG
repeats/Trim_SRR8393698_2_(R2_paired).fastq" -S "bam/SRR8393698_71CAG.sam"

> result=system(command, intern = TRUE, show.output.on.console = TRUE)

> command = "C:/Users/User/Documents/R/win-library/4.0/Rbowtie2/bowtie2-align-s" --threads 3 -x
"chromosomas/human" -1 " SRR8393699_clonal iPSC line with 71 CAG
repeats/Trim_SRR8393699_1_(R1_paired).fastq" -2 " SRR8393699_clonal iPSC line with 71 CAG
repeats/Trim_SRR8393699_2_(R2_paired).fastq" -S "bam/SRR8393699_71CAG.sam"

> result=system(command, intern = TRUE, show.output.on.console = TRUE)

> command = "C:/Users/User/Documents/R/win-library/4.0/Rbowtie2/bowtie2-align-s" --threads 3 -x
"chromosomas/human" -1 " SRR8393700_clonal iPSC line with 71 CAG
repeats/Trim_SRR8393700_1_(R1_paired).fastq" -2 " SRR8393700_clonal iPSC line with 71 CAG
repeats/Trim_SRR8393700_2_(R2_paired).fastq" -S "bam/SRR8393700_71CAG.sam"

> result=system(command, intern = TRUE, show.output.on.console = TRUE)

> command = "C:/Users/User/Documents/R/win-library/4.0/Rbowtie2/bowtie2-align-s" --threads 3 -x
"chromosomas/human" -1 " SRR8393701_clonal iPSC line with 109 CAG
repeats/Trim_SRR8393701_1_(R1_paired).fastq" -2 " SRR8393701_clonal iPSC line with 109 CAG
repeats/Trim_SRR8393701_2_(R2_paired).fastq" -S "bam/SRR8393701_109CAG.sam"

> result=system(command, intern = TRUE, show.output.on.console = TRUE)

> command = "C:/Users/User/Documents/R/win-library/4.0/Rbowtie2/bowtie2-align-s" --threads 3 -x
"chromosomas/human" -1 " SRR8393702_clonal iPSC line with 109 CAG

```

```
repeats/Trim_SRR8393702_1_(R1_paired).fastq" -2 " SRR8393702_clonal iPSC line with 109 CAG
repeats/Trim_SRR8393702_2_(R2_paired).fastq" -S "bam/SRR839702_109CAG.sam"
```

```
> result=system(command, intern = TRUE, show.output.on.console = TRUE)
```

```
> command = "'C:/Users/User/Documents/R/win-library/4.0/Rbowtie2/bowtie2-align-s' --threads 3 -x
'chromosomas/human' -1 ' SRR8393703_clonal iPSC line with 109 CAG
repeats/Trim_SRR8393703_1_(R1_paired).fastq' -2 ' SRR8393703_clonal iPSC line with 109 CAG
repeats/Trim_SRR8393703_2_(R2_paired).fastq' -S 'bam/SRR839703_109CAG.sam'"
```

```
> result=system(command, intern = TRUE, show.output.on.console = TRUE)
```

- **Convertir los archivos .sam en .bam (limpieza y optimización de los alineamientos**

```
> asBam ("bam/ SRR8393695_control.sam ")
```

```
> asBam ("bam/ SRR8393696_control.sam ")
```

```
> asBam ("bam/ SRR8393697_control.sam ")
```

```
> asBam ("bam/ SRR8393698_71CAG.sam ")
```

```
> asBam ("bam/ SRR8393699_71CAG.sam ")
```

```
> asBam ("bam/ SRR8393700_71CAG.sam ")
```

```
> asBam ("bam/ SRR8393701_109CAG.sam ")
```

```
> asBam ("bam/ SRR8393702_109CAG.sam ")
```

```
> asBam ("bam/ SRR8393703_109CAG.sam ")
```

- **Función que permite almacenar todos los paths de los archivos bam en "bamFiles" usando la función "list.files"**

```
> bamFiles = list.files(path="bam", pattern="*.bam$", full.names=TRUE)
```

- **Cuantificar "genes" con función "featureCounts" y almacenar en un objeto R "data"**

```
> data=featureCounts(files = bamFiles,annot.ext = "genomic.gff",isGTFAnnotation =
TRUE,GTF.featureType = "gene",GTF.attrType = "ID",isPairedEnd = TRUE,requireBothEndsMapped
= TRUE,minMQS = 20,strandSpecific = 2)
```

- **Creamos un data.frame para dar un nombre a los grupos replicados. En el mismo orden "bamFiles":**

```
> samples=data.frame(conditions=c("control","control","control","71CAG","71CAG","71CAG","109
CAG","109CAG","109CAG"))
```

- **Usar "DESeqDataSetFromMatrix" para reparar todos los raw counts y el diseño experimental:**

```
> dds=DESeqDataSetFromMatrix(data$counts,colData = samples,design = ~conditions)
```

- **Usar "rowSums" para que los counts > 10 sean legibles para la próxima función:**

```
> dds=dds[rowSums(counts(dds)) > 10,]
```

- **Usamos la función "DESeq" para normalizar y calcular la dispersión de los datos de expresión génica**

```
> dds=DESeq(dds)
```

- **Usamos la función "results" para calcular: log2Ratios, pvalues... para las comparaciones de interés:**

```
> controlvs71CAG=results(dds,contrast=c("conditions","control","71CAG"))
```

```
> controlvs109CAG=results(dds,contrast=c("conditions","control","109CAG"))
```

```
> CAG71vs109CAG=results(dds,contrast=c("conditions","71CAG","109CAG"))
```

- **Generar una tabla con formato con todos los Resultados obtenidos (data.frame):**

```
>FINAL_controlvs71CAG=data.frame(GENEID=row.names(controlvs71CAG),log2BaseMean=log2(
controlvs71CAG$baseMean),log2Ratio=controlvs71CAG$log2FoldChange,STDERR_log2Ratio=con
trolvs71CAG$lfcSE,pvalue=controlvs71CAG$pvalue,adjust=controlvs71CAG$padj)
```

```
>FINAL_controlvs109CAG=data.frame(GENEID=row.names(controlvs109CAG),log2BaseMean=log
2(controlvs109CAG$baseMean),log2Ratio=controlvs109CAG$log2FoldChange,STDERR_log2Ratio
=controlvs109CAG$lfcSE,pvalue=controlvs109CAG$pvalue,adjust=controlvs109CAG$padj)
```

```
>FINAL_CAG71vs109CAG=data.frame(GENEID=row.names(CAG71vs109CAG),log2BaseMean=log2(
CAG71vs109CAG$baseMean),log2Ratio=CAG71vs109CAG$log2FoldChange,STDERR_log2R
atio=CAG71vs109CAG$lfcSE,pvalue=CAG71vs109CAG$pvalue,adjust=CAG71vs109CAG$padj)
```

- **Usamos la función "write.table" para crear un archivo con el contenido del análisis:**

```
> write.table(FINAL_controlvs71CAG,file = "RESULTS_controlvs71CAG.tsv",sep = "/",row.names
= FALSE)
```

```
> write.table(FINAL_controlvs109CAG,file = "RESULTS_controlvs109CAG.tsv",sep =
"/",row.names = FALSE)
```

```
> write.table(FINAL_CAG71vs109CAG,file = "RESULTS_CAG71vs109CAG.tsv",sep =
"/",row.names = FALSE)
```

Gráfico de componentes principales PCA

- **Aseguramos la normalización del objeto dds**

```
> vsd = vst(dds, blind=FALSE)
```

- **Filtrar genes significativos con p-value < 0.05**

```
> controlvs71CAG_noNA = controlvs71CAG[!is.na(controlvs71CAG$pvalue), ]
```

```
> significant_genes = rownames(controlvs71CAG_noNA[controlvs71CAG_noNA$pvalue < 0.05, ])
```

```
> vsd_filtered = vsd[significant_genes, ]
```

- **Generamos el PCA con solo los valores significativos**

```
> pcaData = plotPCA(vsd_filtered, intgroup=c("conditions"), returnData=TRUE)
```

```
> percentVar = round(100 * attr(pcaData, "percentVar"))
```

- **Aseguramos que los nombres de las muestras sean visibles**

```
> pcaData$sample = rownames(pcaData)
```

- **Graficar PCA con ggplot2**

```
> ggplot(pcaData, aes(PC1, PC2, color=conditions, label=sample)) + geom_point(size=3) +
geom_text_repel(size=3, max.overlaps = 10, box.padding = 0.5, point.padding = 0.5) +
xlab(paste0("PC1: ", percentVar[1], "% variance")) + ylab(paste0("PC2: ", percentVar[2], "%
variance")) + theme_minimal()
```

HeatMap

- **Filtrar los genes eliminando filas con valores NA**

```
> controlvs71CAG_noNA = controlvs71CAG[!is.na(controlvs71CAG$pvalue), ]
```

- **Filtrar genes significativos con p-value < 0.05**

```
> significant_genes = rownames(controlvs71CAG_noNA[controlvs71CAG_noNA$pvalue < 0.05, ])
```

- **Asegúrate de que el objeto 'vsd' contiene los datos de expresión para los genes significativos**

```
> vsd_filtered = vsd [significant_genes,]
```

- **Seleccionar los genes más variables entre los genes significativos**

```
> topVarGenes = head(order(rowVars(assay(vsd_filtered))), decreasing=TRUE), 30)
```

- **Extraer la matriz de expresión de estos genes**

```
> mat = assay(vsd_filtered) [topVarGenes,]
```

- **Centrar los datos de expresión en cada fila**

```
> mat = mat - rowMeans(mat)
```

- **Configura las filas del objeto 'samples' para que coincidan con las columnas de 'vsd'**

```
> rownames(samples) = colnames(mat)
```

- **Definir una paleta de colores explícita**

```
> my_palette = colorRampPalette(c("blue", "white", "red"))(50)
```

- **Generar el heatmap con los datos filtrados**

```
> pheatmap(mat, annotation_col = samples, scale = "row", clustering_distance_rows = "euclidean",
clustering_distance_cols = "euclidean", clustering_method = "complete", color = my_palette,
show_rownames = TRUE, show_colnames = TRUE)
```

Anexo 3. Script para la detección de variantes de splicing

Mapeado de lecturas contra el genoma de referencia (Map with BWA-MEM)

control

```
>set -o | grep -q pipefail && set -o pipefail;      bwa mem -t "${GALAXY_SLOTS:-1}" -v 1
-R          '@RG\tID:Control.CAG\tSM:Control\tPL:ILLUMINA'          -I          '250,50'
'/cvmfs/data.galaxyproject.org/byhand/hg38/hg38full/bwa_index_v0.7.10-r789/hg38full.fa' '/mnt/user-
data-volD/data23/9/e/6/dataset_9e695b5f-c695-4d3e-816d-4463ccfd3fd4.dat'          '/mnt/user-data-
volD/data23/5/1/7/dataset_5171b041-6a85-412f-91b3-1203c22e410b.dat'          | samtools sort -
@${GALAXY_SLOTS:-2}          -T          "${TMPDIR:-.}"          -O          bam          -o
'/mnt/scratch/job_working_directory/010/143/10143393/outputs/dataset_01680a45-bf90-411f-a759-
d2d9f810016c.dat'
```

71CAG

```
>set -o | grep -q pipefail && set -o pipefail;      bwa mem -t "${GALAXY_SLOTS:-1}" -v 1
-R          '@RG\tID:71CAG.CAG\tSM:71CAG\tPL:ILLUMINA'          -I          '250,50'
'/cvmfs/data.galaxyproject.org/byhand/hg38/hg38full/bwa_index_v0.7.10-r789/hg38full.fa' '/mnt/user-
data-volD/data23/c/4/7/dataset_c471a7e1-7418-4351-85ca-063d86017a22.dat'          '/mnt/user-data-
volD/data23/f/9/c/dataset_f9c06288-55e5-4429-b5ca-aa4094be2448.dat'          | samtools sort -
@${GALAXY_SLOTS:-2}          -T          "${TMPDIR:-.}"          -O          bam          -o
'/mnt/scratch/job_working_directory/010/150/10150393/outputs/dataset_cedaaefe-706e-413f-b30e-
0f390519ddd0.dat'
```

109CAG

```
>set -o | grep -q pipefail && set -o pipefail;      bwa mem -t "${GALAXY_SLOTS:-1}" -v 1
-R          '@RG\tID:109CAG.CAG\tSM:109CAG\tPL:ILLUMINA'          -I          '250,50'
'/cvmfs/data.galaxyproject.org/byhand/hg38/hg38full/bwa_index_v0.7.10-r789/hg38full.fa' '/mnt/user-
data-volD/data23/0/f/b/dataset_0fb44fce-9df2-4668-95e8-d1a87a11ba34.dat'          '/mnt/user-data-
volD/data23/8/5/c/dataset_85c2164e-1fd7-461e-8f65-85c782487709.dat'          | samtools sort -
@${GALAXY_SLOTS:-2}          -T          "${TMPDIR:-.}"          -O          bam          -o
'/mnt/scratch/job_working_directory/010/168/10168663/outputs/dataset_730b89df-bc89-4451-8b62-
8e9f9c76b06d.dat'
```

Ordenar el conjunto de datos SAM/BAM (SortSam)

control

```
>          _JAVA_OPTIONS=${_JAVA_OPTIONS:-"-Xmx2048m          -Xms256m          -
Djava.io.tmpdir=${TMPDIR:-${_GALAXY_JOB_TMPDIR}}"} && export _JAVA_OPTIONS &&
ln          -sf          '/mnt/usuario-datos-volD/data23/0/1/6/conjunto-de-datos_01680a45-bf90-411f-a759-
d2d9f810016c.dat' 'SRR8393695_fastq' && picard SortSam --ENTRADA 'SRR8393695_fastq' --
SALIDA          '/mnt/scratch/directorio_trabajo/010/149/10149019/salidas/conjunto_de_datos_dfb4ea0e-
```

```
8168-4f1e-8112-5657bbeec842.dat' --ORDEN_DE_CLASIFICACIÓN 'coordenada' --QUIET
verdadero --ERROR DE VERBOSIDAD --STRINGENCIA_DE_VALIDACIÓN SILENCIOSO
> _JAVA_OPTIONS=${_JAVA_OPTIONS:-"-Xmx2048m -Xms256m -
Djava.io.tmpdir=${TMPDIR:-${_GALAXY_JOB_TMPDIR}}"} && export _JAVA_OPTIONS &&
ln -sf '/mnt/datos-usuario-volD/data23/5/5/a/dataset_55a4d3ae-6615-47c6-901c-5b9b841bed93.dat'
'SRR8393697_fastq' && picard SortSam --ENTRADA 'SRR8393697_fastq' --SALIDA
'/mnt/scratch/directorio_trabajo/010/149/10149020/salidas/conjunto_de_datos_df688fbb-4734-4178-
a6a6-ec997e33b057.dat' --ORDEN_DE_CLASIFICACIÓN 'coordenada' --QUIET verdadero --
ERROR DE VERBOSIDAD --STRINGENCIA_DE_VALIDACIÓN SILENCIOSO
```

71CAG

```
> _JAVA_OPTIONS=${_JAVA_OPTIONS:-"-Xmx2048m -Xms256m -
Djava.io.tmpdir=${TMPDIR:-${_GALAXY_JOB_TMPDIR}}"} && export _JAVA_OPTIONS &&
ln -sf '/mnt/user-data-volD/data23/c/e/d/dataset_cedaaefe-706e-413f-b30e-0f390519ddd0.dat'
'SRR8393698_fastq' && picard SortSam --INPUT 'SRR8393698_fastq' --OUTPUT
'/mnt/scratch/job_working_directory/010/151/10151111/outputs/dataset_274c4ad5-8a32-42d7-b1b4-
815cfc29fa4e.dat' --SORT_ORDER 'coordinate' --QUIET true --VERBOSITY ERROR --
VALIDATION_STRINGENCY SILENT
```

```
> _JAVA_OPTIONS=${_JAVA_OPTIONS:-"-Xmx2048m -Xms256m -
Djava.io.tmpdir=${TMPDIR:-${_GALAXY_JOB_TMPDIR}}"} && export _JAVA_OPTIONS &&
ln -sf '/mnt/user-data-volD/data23/6/b/2/dataset_6b26eb95-0d52-408c-8994-82e0de26888a.dat'
'SRR8393700_fastq' && picard SortSam --INPUT 'SRR8393700_fastq' --OUTPUT
'/mnt/scratch/job_working_directory/010/151/10151112/outputs/dataset_7652b365-cf9a-478f-bc71-
4627f4e22e4b.dat' --SORT_ORDER 'coordinate' --QUIET true --VERBOSITY ERROR --
VALIDATION_STRINGENCY SILENT
```

109CAG

```
> _JAVA_OPTIONS=${_JAVA_OPTIONS:-"-Xmx2048m -Xms256m -
Djava.io.tmpdir=${TMPDIR:-${_GALAXY_JOB_TMPDIR}}"} && export _JAVA_OPTIONS &&
ln -sf '/mnt/user-data-volD/data23/7/3/0/dataset_730b89df-bc89-4451-8b62-8e9f9c76b06d.dat'
'SRR8393701_fastq' && picard SortSam --INPUT 'SRR8393701_fastq' --OUTPUT
'/mnt/scratch/job_working_directory/010/169/10169580/outputs/dataset_971eb3db-6bba-4465-b133-
e76cb6aa0693.dat' --SORT_ORDER 'coordinate' --QUIET true --VERBOSITY ERROR --
VALIDATION_STRINGENCY SILENT
```

```
> _JAVA_OPTIONS=${_JAVA_OPTIONS:-"-Xmx2048m -Xms256m -
Djava.io.tmpdir=${TMPDIR:-${_GALAXY_JOB_TMPDIR}}"} && export _JAVA_OPTIONS &&
ln -sf '/mnt/user-data-volD/data23/7/e/c/dataset_7ec4f059-23b3-454d-bb28-e4c90a65b449.dat'
'SRR8393703_fastq' && picard SortSam --INPUT 'SRR8393703_fastq' --OUTPUT
'/mnt/scratch/job_working_directory/010/169/10169581/outputs/dataset_98c523fa-80c3-46da-aa11-
```

```
d3f046ee7be4.dat' --SORT_ORDER 'coordinate' --QUIET true --VERBOSITY ERROR --
VALIDATION_STRINGENCY SILENT
```

Generar pileup del conjunto de datos BAM (Generate pileup)

control

```
>ln -s '/mnt/user-data-volD/data23/d/f/b/dataset_dfb4ea0e-8168-4f1e-8112-5657bbeec842.dat'
input1.bam && ln -s '/mnt/user-data-volD/data23/_metadata_files/d/f/7/metadata_df7c6232-efad-4ccd-
adab-95d6ae70a20f.dat' 'input1.bam.bai' && samtools pileup -M 60 -c -T 0.85 -N 2 -r 0.001 -I 40 -f
'/cvmfs/data.galaxyproject.org/byhand/hg38/sam_index/hg38.fa' input1.bam >
'/mnt/scratch/job_working_directory/010/149/10149040/outputs/dataset_a32f6a22-0bba-4c0c-8f92-
760548c193a7.dat'
```

```
> ln -s '/mnt/user-data-volD/data23/d/f/6/dataset_df688fbb-4734-4178-a6a6-ec997e33b057.dat'
input1.bam && ln -s '/mnt/user-data-volD/data23/_metadata_files/1/c/8/metadata_1c86bc0e-ed45-
47b8-ae72-99ce57615005.dat' 'input1.bam.bai' && samtools pileup -M 60 -c -T 0.85 -N 2 -r 0.001 -I 40
-f '/cvmfs/data.galaxyproject.org/byhand/hg38/sam_index/hg38.fa' input1.bam >
'/mnt/scratch/job_working_directory/010/149/10149041/outputs/dataset_b70439d8-4319-425b-8ae5-
abc26a8975af.dat'
```

71CAG

```
> ln -s '/mnt/user-data-volD/data23/2/7/4/dataset_274c4ad5-8a32-42d7-b1b4-815cfc29fa4e.dat'
input1.bam && ln -s '/mnt/user-data-volD/data23/_metadata_files/7/0/2/metadata_7029d40d-379a-
406b-92f9-a3be249ef03d.dat' 'input1.bam.bai' && samtools pileup -M 60 -c -T 0.85 -N 2 -r 0.001 -I 40
-f '/cvmfs/data.galaxyproject.org/byhand/hg38/sam_index/hg38.fa' input1.bam >
'/mnt/scratch/job_working_directory/010/151/10151486/outputs/dataset_e92bee1d-847b-4463-aa78-
62e060ac8f43.dat'
```

```
> ln -s '/mnt/user-data-volD/data23/7/6/5/dataset_7652b365-cf9a-478f-bc71-4627f4e22e4b.dat'
input1.bam && ln -s '/mnt/user-data-volD/data23/_metadata_files/0/3/3/metadata_0339adf4-10f8-
46cb-99a1-41c52fd7bef1.dat' 'input1.bam.bai' && samtools pileup -M 60 -c -T 0.85 -N 2 -r 0.001 -I 40
-f '/cvmfs/data.galaxyproject.org/byhand/hg38/sam_index/hg38.fa' input1.bam >
'/mnt/scratch/job_working_directory/010/151/10151487/outputs/dataset_f740f392-3c85-47ee-8213-
bdfb75f78278.dat'
```

109CAG

```
> ln -s '/mnt/user-data-volD/data23/9/7/1/dataset_971eb3db-6bba-4465-b133-e76cb6aa0693.dat'
input1.bam && ln -s '/mnt/user-data-volD/data23/_metadata_files/1/1/c/metadata_11c76e74-0746-
43ec-a1a4-9f6e2afce5b2.dat' 'input1.bam.bai' && samtools pileup -M 60 -c -T 0.85 -N 2 -r 0.001 -I 40
-f '/cvmfs/data.galaxyproject.org/byhand/hg38/sam_index/hg38.fa' input1.bam >
'/mnt/scratch/job_working_directory/010/169/10169713/outputs/dataset_58c23dcf-045d-47c3-9dea-
a92c2452a709.dat'
```

```
> ln -s '/mnt/user-data-volD/data23/9/8/c/dataset_98c523fa-80c3-46da-aa11-d3f046ee7be4.dat'
input1.bam && ln -s '/mnt/user-data-volD/data23/_metadata_files/0/7/f/metadata_07f77858-1730-
4623-b4fa-d79f0e1a4e81.dat' 'input1.bam.bai' && samtools pileup -M 60 -c -T 0.85 -N 2 -r 0.001 -I 40
-f '/cvmfs/data.galaxyproject.org/byhand/hg38/sam_index/hg38.fa' input1.bam >
'/mnt/scratch/job_working_directory/010/169/10169714/outputs/dataset_f25367c2-14cf-44a7-a558-
ad20d3e4a779.dat'
```

Acumulación de filtros sobre cobertura y SNP (Filter pileup)

control

```
> perl
/mnt/tools/shed_tools/toolshed.g2.bx.psu.edu/repos/devteam/pileup_parser/85bedbea8a12/pileup_pars
er/pileup_parser.pl /mnt/user-data-volD/data23/a/3/2/dataset_a32f6a22-0bba-4c0c-8f92-
760548c193a7.dat "3" "9" "10" "8" 20 20 Yes No "2"
/mnt/scratch/job_working_directory/010/149/10149127/outputs/dataset_9819b793-ec40-46a3-ac18-
6ba91d512605.dat No Yes
```

71CAG

```
> perl
/mnt/tools/shed_tools/toolshed.g2.bx.psu.edu/repos/devteam/pileup_parser/85bedbea8a12/pileup_pars
er/pileup_parser.pl /mnt/user-data-volD/data23/e/9/2/dataset_e92bee1d-847b-4463-aa78-
62e060ac8f43.dat "3" "9" "10" "8" 20 20 Yes No "2"
/mnt/scratch/job_working_directory/010/153/10153574/outputs/dataset_6e64a714-e36e-4393-9bac-
b5855554ede4.dat No Yes
```

109CAG

```
> perl
/mnt/tools/shed_tools/toolshed.g2.bx.psu.edu/repos/devteam/pileup_parser/85bedbea8a12/pileup_pars
er/pileup_parser.pl /mnt/user-data-volD/data23/5/8/c/dataset_58c23dcf-045d-47c3-9dea-
a92c2452a709.dat "3" "9" "10" "8" 20 20 Yes No "2"
/mnt/scratch/job_working_directory/010/171/10171988/outputs/dataset_e1cd733b-aabe-475b-9658-
f85454007955.dat No Yes
```

Detectar las variantes genéticas (FreeBayes)

control

```
> ln -s -f '/mnt/pulsar/files/staging/10149129/inputs/dataset_dfb4ea0e-8168-4f1e-8112-
5657bbeec842.dat' 'b_0.bam' && ln -s -f
'/mnt/pulsar/files/staging/10149129/inputs/metadata_metadata_df7c6232-efad-4ccd-adab-
95d6ae70a20f.dat' 'b_0.bam.bai' && ln -s -f
'/mnt/pulsar/files/staging/10149129/inputs/dataset_df688fbb-4734-4178-a6a6-ec997e33b057.dat'
'b_1.bam' && ln -s -f '/mnt/pulsar/files/staging/10149129/inputs/metadata_metadata_1c86bc0e-ed45-
47b8-ae72-99ce57615005.dat' 'b_1.bam.bai' && samtools view -H b_0.bam | grep '^@SQ' | cut -f 2 - |
```

```
awk '{ gsub("^SN:", "", $1); gsub("^LN:", "", $2); print $1"\t0\t"$2; }' >> regions_all.bed && samtools
view -H b_1.bam | grep '^@SQ' | cut -f 2- | awk '{ gsub("^SN:", "", $1); gsub("^LN:", "", $2); print
$1"\t0\t"$2; }' >> regions_all.bed && sort -u regions_all.bed > regions_uniq.bed && mkdir vcf_output
failed_alleles trace && for i in `cat regions_uniq.bed | awk '{print $1":"$2".."$3}'`; do echo "
freebayes --region '$i' --bam 'b_0.bam' --bam 'b_1.bam' --fasta-reference
'/cvmfs/data.galaxyproject.org/byhand/hg38/sam_index/hg38.fa' --vcf './vcf_output/part_${i}.vcf' --
standard-filters "; done > freebayes_commands.sh && cat freebayes_commands.sh | parallel --will-
cite -j ${GALAXY_SLOTS:-1} && grep "^#" './vcf_output/part_${i}.vcf' > header.txt && for i in `cat
regions_uniq.bed | awk '{print $1":"$2".."$3}'`; do cat './vcf_output/part_${i}.vcf' | grep -v "^#" || true ;
done | sort -k1,1 -k2,2n -k5,5 -u | cat header.txt - >
'/mnt/pulsar/files/staging/10149129/outputs/dataset_900eade1-c507-4e0b-9a0c-b13fda78ec77.dat'
```

71CAG

```
> ln -s -f '/mnt/user-data-volD/data23/2/7/4/dataset_274c4ad5-8a32-42d7-b1b4-815cfc29fa4e.dat'
'b_0.bam' && ln -s -f '/mnt/user-data-volD/data23/_metadata_files/7/0/2/metadata_7029d40d-379a-
406b-92f9-a3be249ef03d.dat' 'b_0.bam.bai' && ln -s -f '/mnt/user-data-
volD/data23/7/6/5/dataset_7652b365-cf9a-478f-bc71-4627f4e22e4b.dat' 'b_1.bam' && ln -s -f
'/mnt/user-data-volD/data23/_metadata_files/0/3/3/metadata_0339adf4-10f8-46cb-99a1-
41c52fd7bef1.dat' 'b_1.bam.bai' && samtools view -H b_0.bam | grep '^@SQ' | cut -f 2- | awk '{
gsub("^SN:", "", $1); gsub("^LN:", "", $2); print $1"\t0\t"$2; }' >> regions_all.bed && samtools view -H
b_1.bam | grep '^@SQ' | cut -f 2- | awk '{ gsub("^SN:", "", $1); gsub("^LN:", "", $2); print $1"\t0\t"$2; }'
>> regions_all.bed && sort -u regions_all.bed > regions_uniq.bed && mkdir vcf_output failed_alleles
trace && for i in `cat regions_uniq.bed | awk '{print $1":"$2".."$3}'`; do echo " freebayes --region
'$i' --bam 'b_0.bam' --bam 'b_1.bam' --fasta-reference
'/cvmfs/data.galaxyproject.org/byhand/hg38/sam_index/hg38.fa' --vcf './vcf_output/part_${i}.vcf' --
standard-filters "; done > freebayes_commands.sh && cat freebayes_commands.sh | parallel --will-
cite -j ${GALAXY_SLOTS:-1} && grep "^#" './vcf_output/part_${i}.vcf' > header.txt && for i in `cat
regions_uniq.bed | awk '{print $1":"$2".."$3}'`; do cat './vcf_output/part_${i}.vcf' | grep -v "^#" || true ;
done | sort -k1,1 -k2,2n -k5,5 -u | cat header.txt - >
'/mnt/scratch/job_working_directory/010/153/10153578/outputs/dataset_87da2fe7-a339-4860-9611-
11ce9f4d66c1.dat'
```

109CAG

```
> ln -s -f '/mnt/user-data-volD/data23/9/7/1/dataset_971eb3db-6bba-4465-b133-e76cb6aa0693.dat'
'b_0.bam' && ln -s -f '/mnt/user-data-volD/data23/_metadata_files/1/1/c/metadata_11c76e74-0746-
43ec-a1a4-9f6e2afce5b2.dat' 'b_0.bam.bai' && ln -s -f '/mnt/user-data-
volD/data23/9/8/c/dataset_98c523fa-80c3-46da-aa11-d3f046ee7be4.dat' 'b_1.bam' && ln -s -f
'/mnt/user-data-volD/data23/_metadata_files/0/7/f/metadata_07f77858-1730-4623-b4fa-
d79f0e1a4e81.dat' 'b_1.bam.bai' && samtools view -H b_0.bam | grep '^@SQ' | cut -f 2- | awk '{
```

```

gsub("^SN:", "", $1); gsub("^LN:", "", $2); print $1"\t0\t"$2; }' >> regions_all.bed && samtools view -H
b_1.bam | grep '^@SQ' | cut -f 2- | awk '{ gsub("^SN:", "", $1); gsub("^LN:", "", $2); print $1"\t0\t"$2; }'
>> regions_all.bed && sort -u regions_all.bed > regions_uniq.bed && mkdir vcf_output failed_alleles
trace && for i in `cat regions_uniq.bed | awk '{print $1":"$2".."$3}'`; do echo " freebayes --region
'$i' --bam 'b_0.bam' --bam 'b_1.bam' --fasta-reference
'/cvmfs/data.galaxyproject.org/byhand/hg38/sam_index/hg38.fa' --vcf './vcf_output/part_${i}.vcf' --
standard-filters "; done > freebayes_commands.sh && cat freebayes_commands.sh | parallel --will-
cite -j ${GALAXY_SLOTS:-1} && grep "^#" './vcf_output/part_${i}.vcf' > header.txt && for i in `cat
regions_uniq.bed | awk '{print $1":"$2".."$3}'`; do cat './vcf_output/part_${i}.vcf' | grep -v "^#" || true ;
done | sort -k1,1 -k2,2n -k5,5 -u | cat header.txt - >
'/mnt/scratch/job_working_directory/010/171/10171993/outputs/dataset_5ae94801-8c56-4e07-ae2c-
42746ceeaafc.dat'

```

Anotar las variantes (SnpEff eff)

control

```

> snpEff -Xmx${GALAXY_MEMORY_MB:-8192}m eff -i vcf -o vcf -upDownStreamLen 5000 -
spliceSiteSize "2" -stats
'/mnt/scratch/job_working_directory/010/149/10149161/outputs/dataset_32f65510-5200-4ef9-b79e-
04d2672d8e11.dat' -noLog -dataDir "$PWD/temp" -download 'hg38' '/mnt/user-data-
volD/data23/9/0/0/dataset_900eade1-c507-4e0b-9a0c-b13fda78ec77.dat' >
'/mnt/scratch/job_working_directory/010/149/10149161/outputs/dataset_8a176d4b-e1b0-4de4-b954-
56333cbb79ea.dat' && mkdir
'/mnt/scratch/job_working_directory/010/149/10149161/outputs/dataset_32f65510-5200-4ef9-b79e-
04d2672d8e11_files' && mv
'/mnt/scratch/job_working_directory/010/149/10149161/outputs/dataset_32f65510-5200-4ef9-b79e-
04d2672d8e11.dat.genes.txt'
'/mnt/scratch/job_working_directory/010/149/10149161/outputs/dataset_32f65510-5200-4ef9-b79e-
04d2672d8e11_files/dataset_32f65510-5200-4ef9-b79e-04d2672d8e11.dat.genes.txt'

```

71CAG

```

> snpEff -Xmx${GALAXY_MEMORY_MB:-8192}m eff -i vcf -o vcf -upDownStreamLen 5000 -
spliceSiteSize "2" -stats
'/mnt/scratch/job_working_directory/010/153/10153957/outputs/dataset_0530a7d4-b892-44c4-a09f-
8f62a16d69c6.dat' -noLog -dataDir "$PWD/temp" -download 'hg38' '/mnt/user-data-
volD/data23/8/7/d/dataset_87da2fe7-a339-4860-9611-11ce9f4d66c1.dat' >
'/mnt/scratch/job_working_directory/010/153/10153957/outputs/dataset_93d99eb7-ade0-4020-b7e4-
5205160315c6.dat' && mkdir
'/mnt/scratch/job_working_directory/010/153/10153957/outputs/dataset_0530a7d4-b892-44c4-a09f-
8f62a16d69c6_files' && mv

```

```
'/mnt/scratch/job_working_directory/010/153/10153957/outputs/dataset_0530a7d4-b892-44c4-a09f-8f62a16d69c6.dat.genes.txt'
```

```
'/mnt/scratch/job_working_directory/010/153/10153957/outputs/dataset_0530a7d4-b892-44c4-a09f-8f62a16d69c6_files/dataset_0530a7d4-b892-44c4-a09f-8f62a16d69c6.dat.genes.txt'
```

109CAG

```
> snpEff -Xmx${GALAXY_MEMORY_MB:-8192}m eff -i vcf -o vcf -upDownStreamLen 5000 -  
spliceSiteSize "2" -stats  
'/mnt/scratch/job_working_directory/010/172/10172013/outputs/dataset_8a7fdaad-c966-4683-ba4e-  
205190046957.dat' -noLog -dataDir "$PWD/temp" -download 'hg38' '/mnt/user-data-  
volD/data23/d/8/5/dataset_d854e6ed-5b8b-4479-8df3-51775fa1291f.dat' >  
'/mnt/scratch/job_working_directory/010/172/10172013/outputs/dataset_66fead9e-011e-48c3-b2e1-  
a6039e5b64eb.dat' && mkdir  
'/mnt/scratch/job_working_directory/010/172/10172013/outputs/dataset_8a7fdaad-c966-4683-ba4e-  
205190046957_files' && mv  
'/mnt/scratch/job_working_directory/010/172/10172013/outputs/dataset_8a7fdaad-c966-4683-ba4e-  
205190046957.dat.genes.txt'  
'/mnt/scratch/job_working_directory/010/172/10172013/outputs/dataset_8a7fdaad-c966-4683-ba4e-  
205190046957_files/dataset_8a7fdaad-c966-4683-ba4e-205190046957.dat.genes.txt'
```

Llamado de variantes específicas de la huntingtina

Paquetes requeridos

```
library(BiocManager)
```

```
library(VariantAnnotation)
```

```
library(GenomicFeatures)
```

Lectura del archivo vcf

```
> vcf=readVcf("SnpEff_eff_control.vcf", "hg38")
```

```
> vcf=readVcf("SnpEff_eff_71CAG.vcf", "hg38")
```

```
> vcf=readVcf("SnpEff_eff_109CAG.vcf", "hg38")
```

Creación del objeto TxDb a partir de archivo GTF

```
> txdb=makeTxDbFromGFF("hg38.ensGene.v78.gtf")
```

Localiza las variantes en regiones específicas

```
> splicing_variants=locateVariants(vcf, txdb, SpliceSiteVariants())
```

Filtra solo las variantes que afectan a tu proteína específica

```
> protein_variants=splicing_variants [which (splicing_variants$GENEID == "ENSG00000197386")]
```

Extraer las variantes originales del VCF que coinciden con los rangos en protein_variants

```
>selected_variants=subsetByOverlaps(vcf, protein_variants)
```

Escribir el archivo VCF

```
> writeVcf(selected_variants, "annotated_specific_variants_control.vcf")
> writeVcf(selected_variants, "annotated_specific_variants_71CAG.vcf")
> writeVcf(selected_variants, "annotated_specific_variants_109CAG.vcf")
```

Anexo 4. Script análisis de splicing alternativo en rMATS

- **Se creo archivos txt que se utilizarán para la agrupación de entradas rMATS de las condiciones de investigación:**

Control

```
>/path/to/SRR8393695_1.fastq:/path/to/SRR8393695_2.fastq,/path/to/SRR8393697_1.fastq:/path/to/
SRR8393697_2.fastq -o /path/to/s1.txt
```

71 CAG

```
>/path/to/SRR8393698_1.fastq:/path/to/SRR8393698_2.fastq,/path/to/SRR8393700_1.fastq:/path/to/
SRR8393700_2.fastq -o /path/to/s2.txt
```

109 CAG

```
>/path/to/SRR8393701_1.fastq:/path/to/SRR8393701_2.fastq,/path/to/SRR8393703_1.fastq:/path/to/
SRR8393703_2.fastq -o /path/to/s3.txt
```

- **Ejecución de rMATS de las entradas para control vs 71CAG y control vs 109CAGG:**

Control vs 71CAG

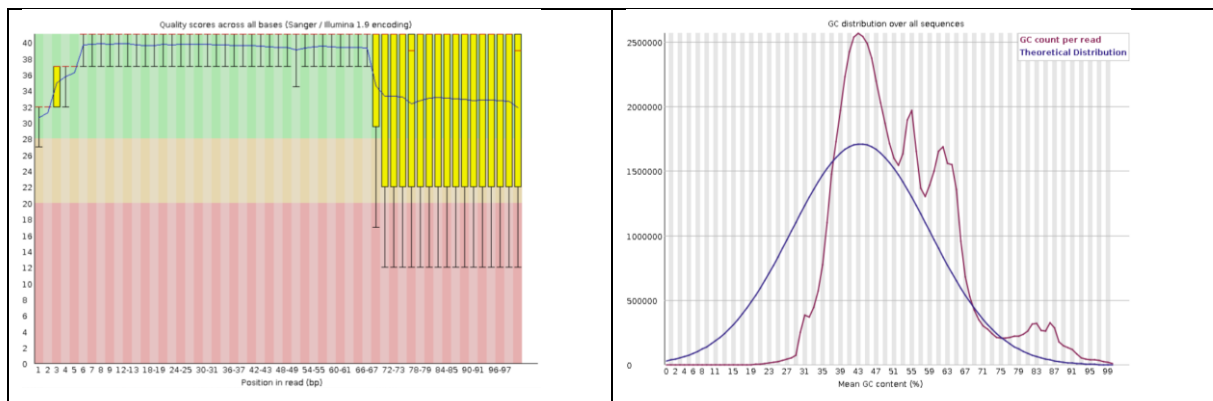
```
>python rmats.py --s1 /path/to/s1.txt --s2 /path/to/s2.txt --gtf /path/to/the.gtf --bi
/path/to/STAR_binary_index -t paired --readLength 101 --nthread 4 --od /path/to/output --tmp
/path/to/tmp_output
```

Control vs 109CAG

```
>python rmats.py --s1 /path/to/s1.txt --s2 /path/to/s3.txt --gtf /path/to/the.gtf --bi
/path/to/STAR_binary_index -t paired --readLength 50 --nthread 4 --od /path/to/output --tmp
/path/to/tmp_output
```

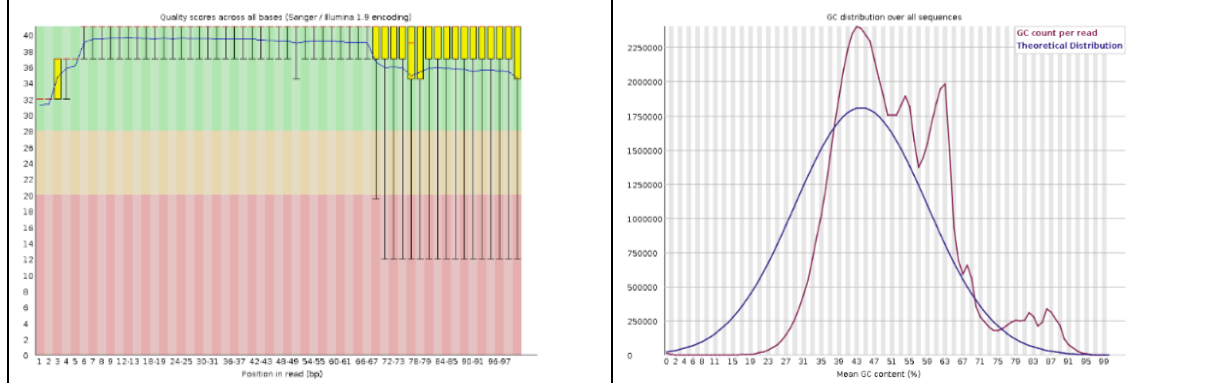
Anexo 5. Control de calidad de muestras crudas de las condiciones de estudio

Calidad de base por secuencia	Contenido de GC por secuencia
SRR8393695_CONTROL	
- Forward	
- Reverse	
SRR8393697_CONTROL	
- Forward	
- Reverse	

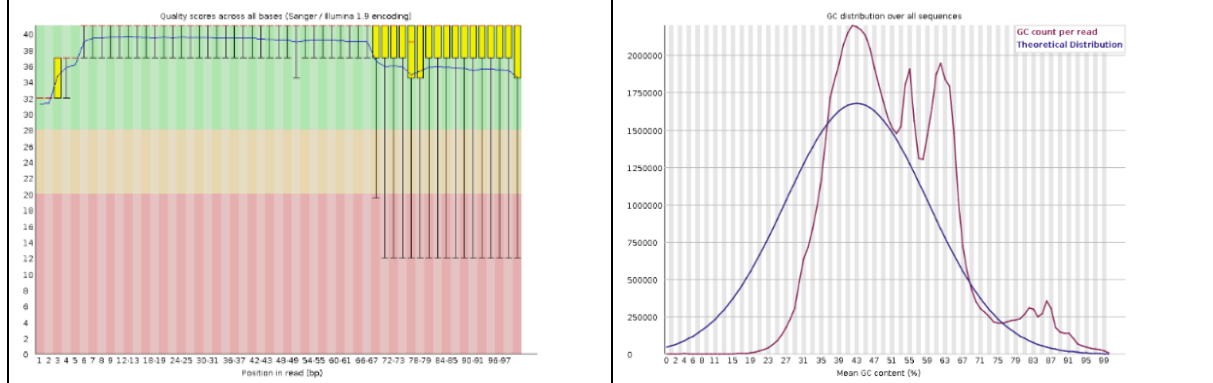


SRR8393698_71CAG

- Forward

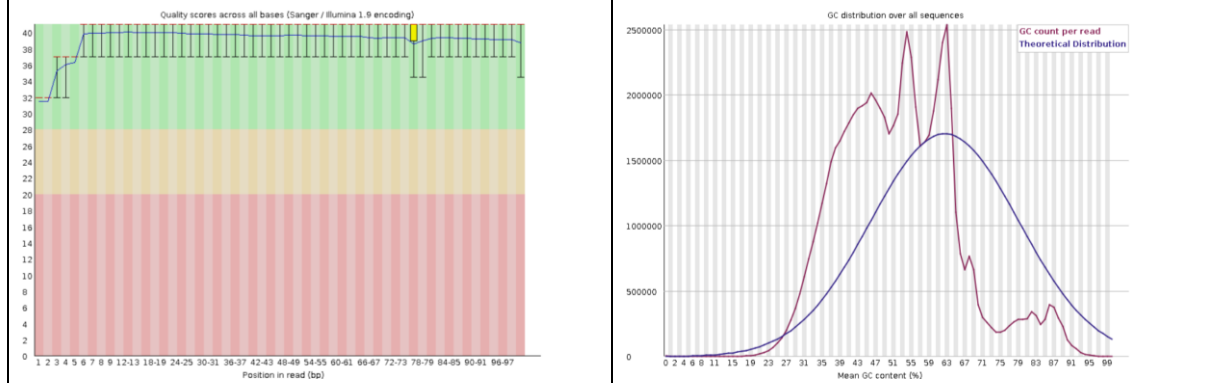


- Reverse

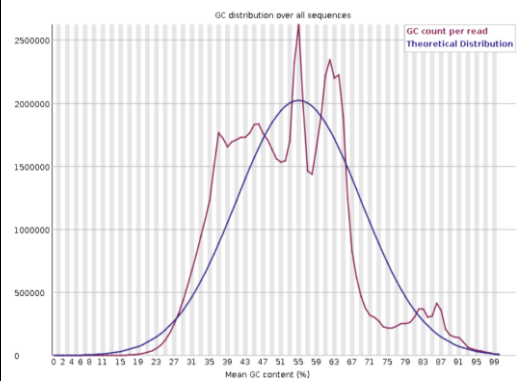
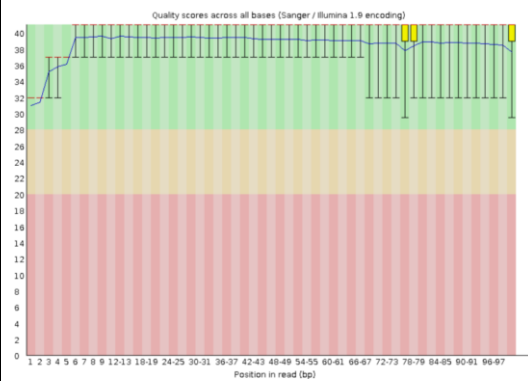


SRR8393700_71CAG

- Forward

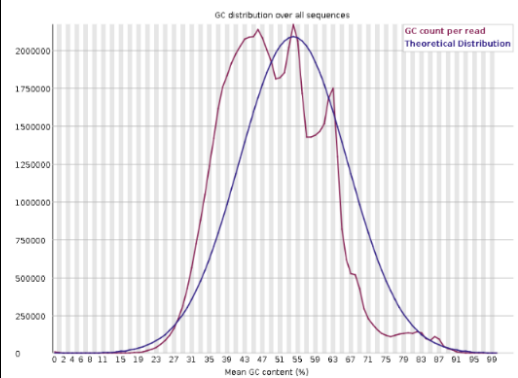
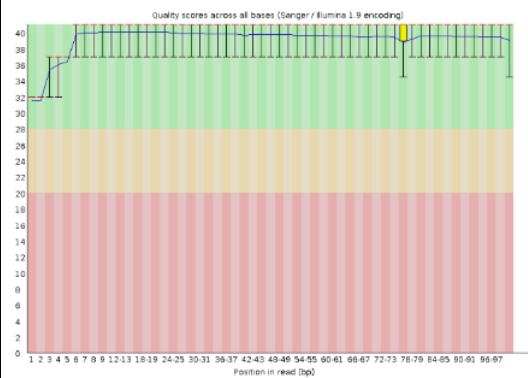


- **Reverse**

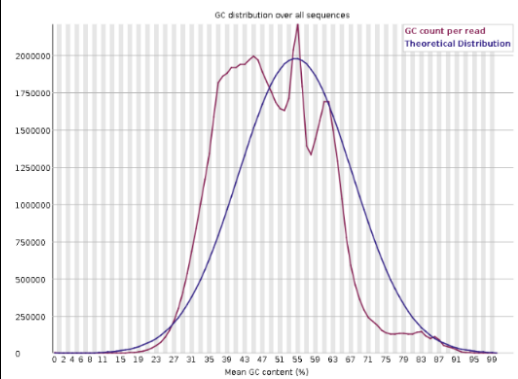
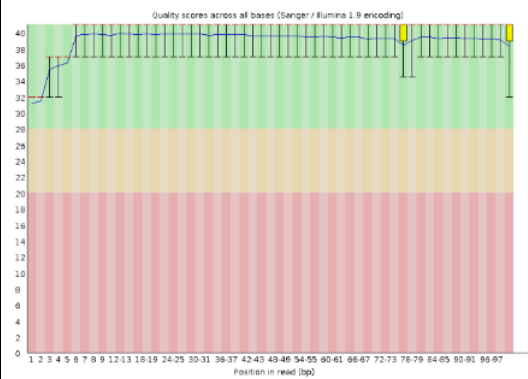


SRR8393701_109CAG

- **Forward**

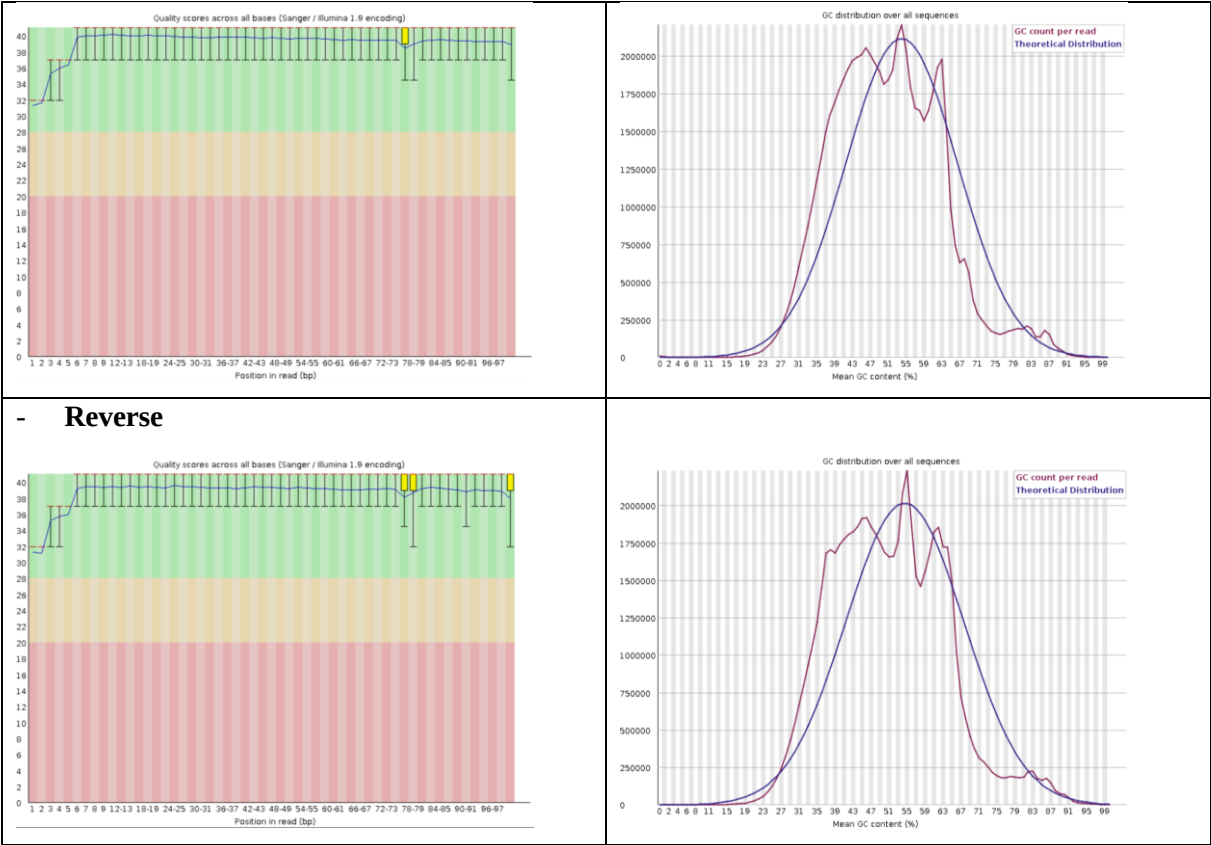


- **Reverse**

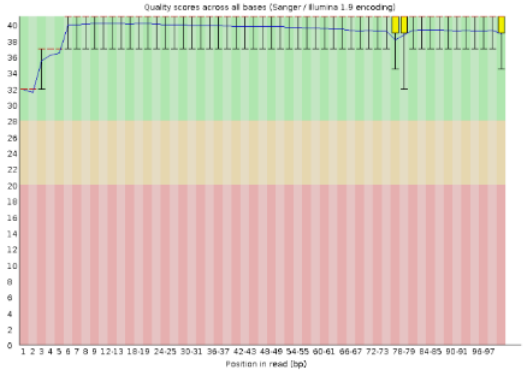
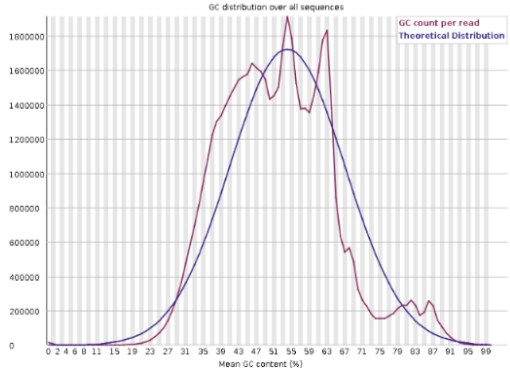
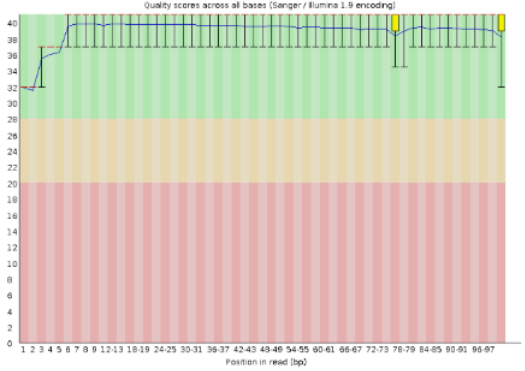
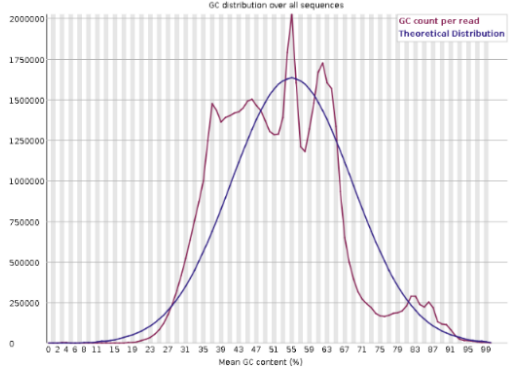
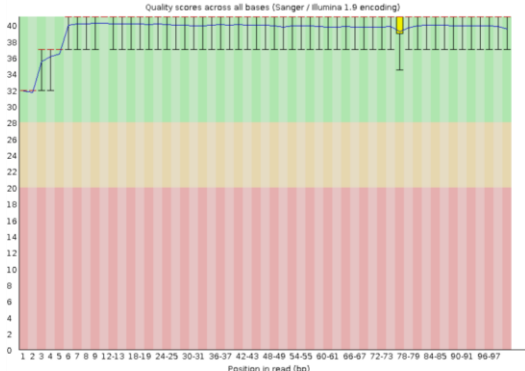
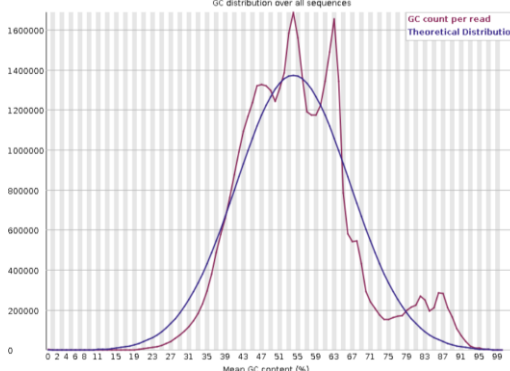


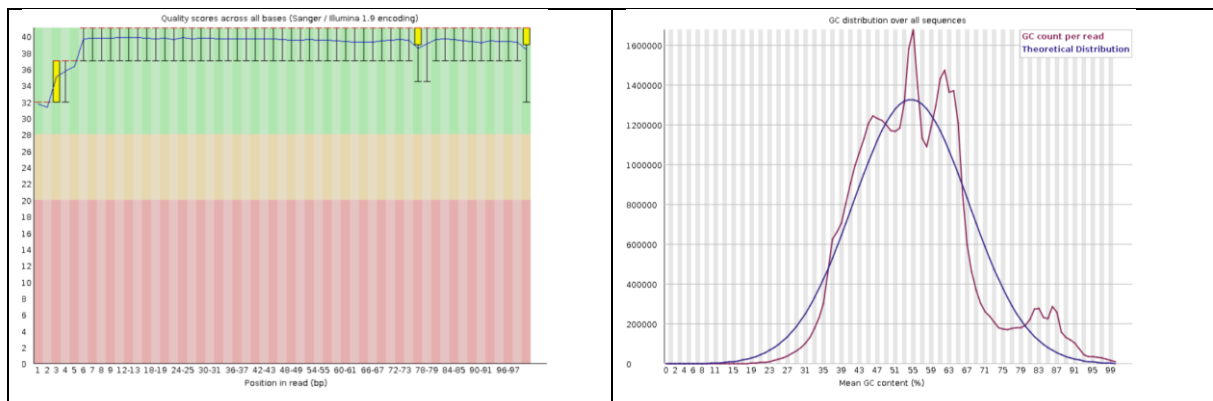
SRR8393703_109CAG

- **Forward**



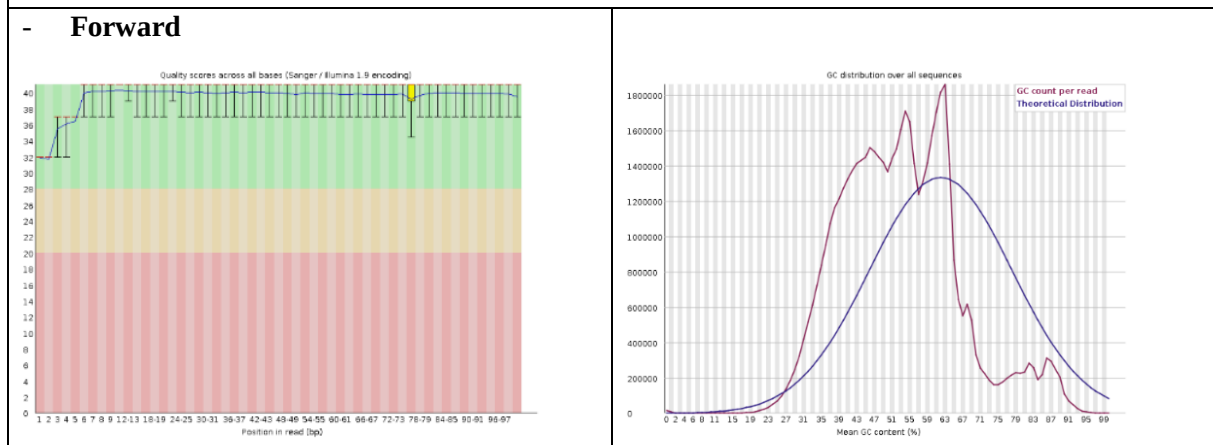
Anexo 6. Control de calidad de muestras cortadas de las condiciones de estudio

Calidad de base por secuencia	Contenido de GC por secuencia
SRR8393695_CONTROL	
- Forward 	
- Reverse 	
SRR8393697_CONTROL	
- Forward 	
- Reverse	

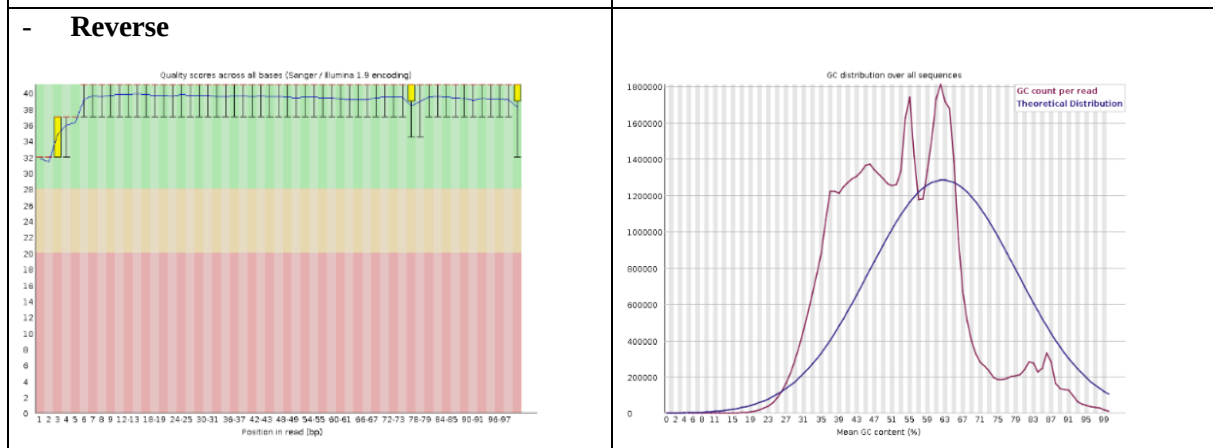


SRR8393698_71CAG

- Forward

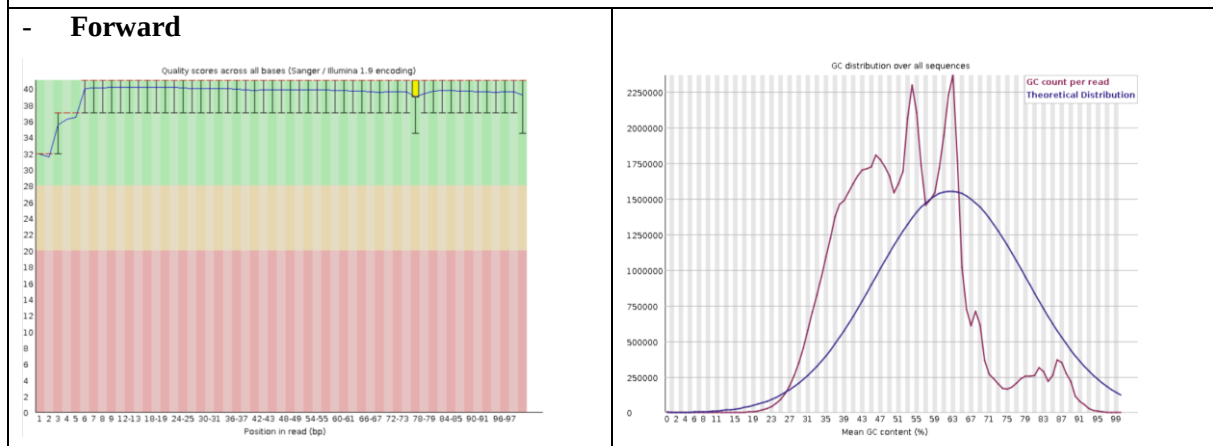


- Reverse

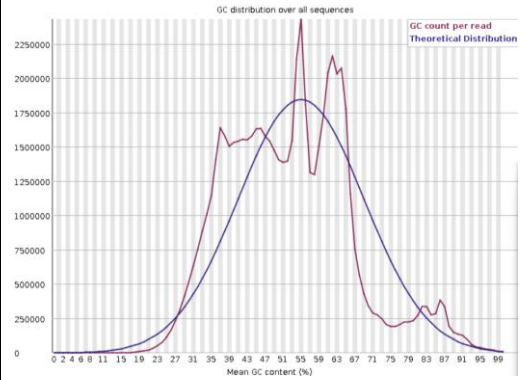
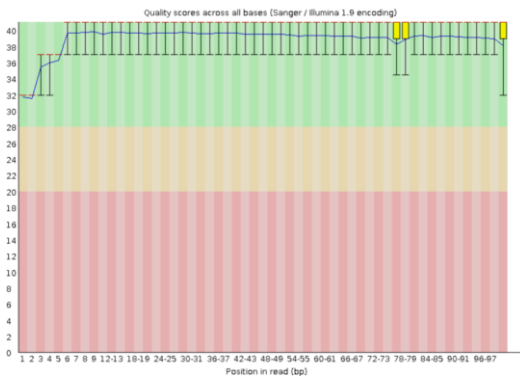


SRR8393700_71CAG

- Forward

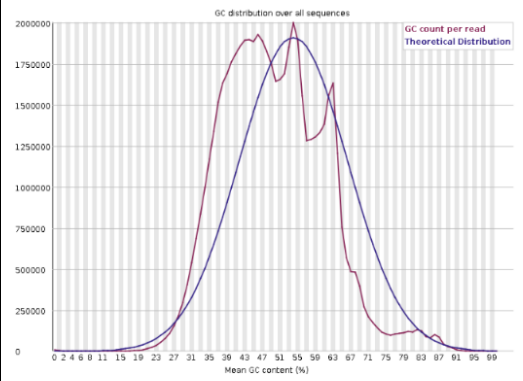
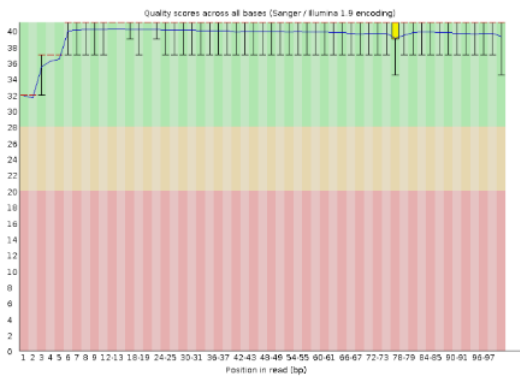


- Reverse

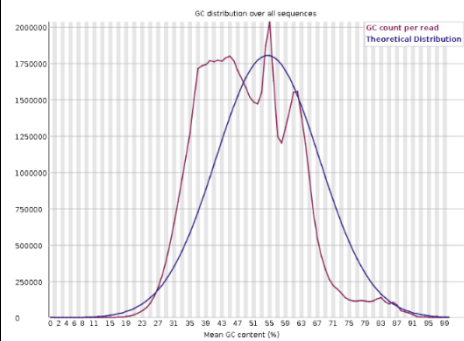
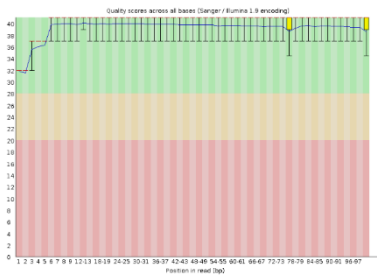


SRR8393701_109CAG

- Forward

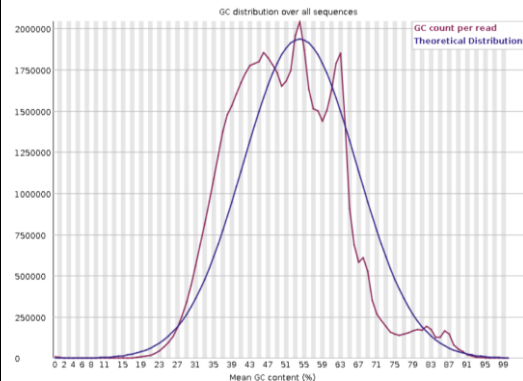
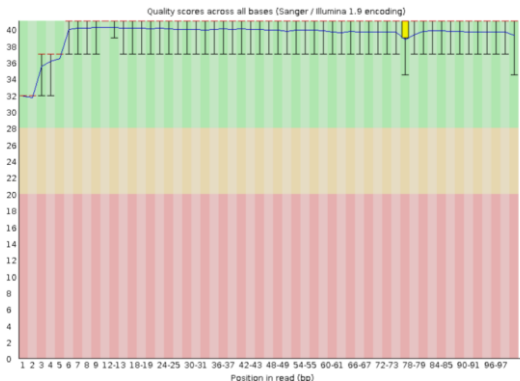


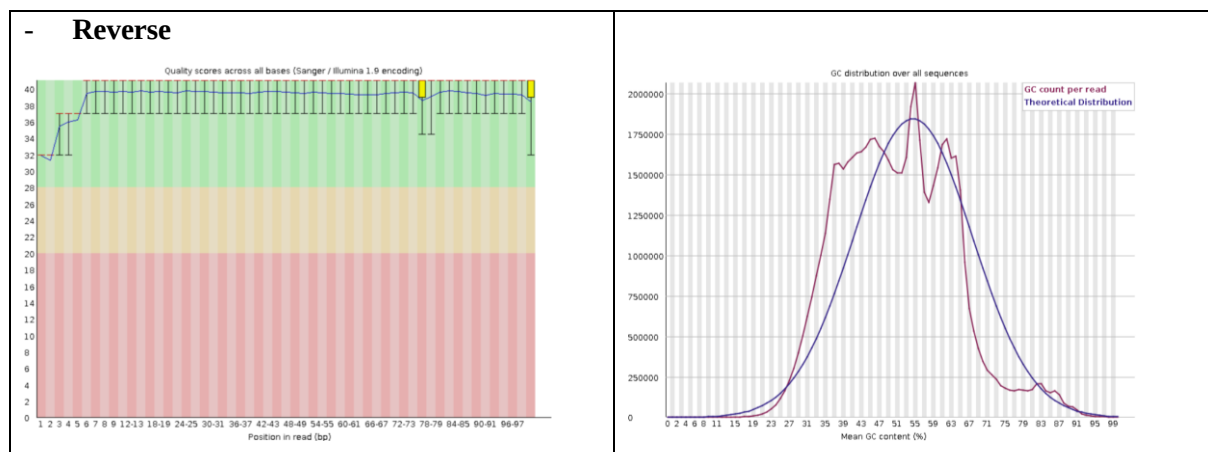
- Reverse



SRR8393703_109CAG

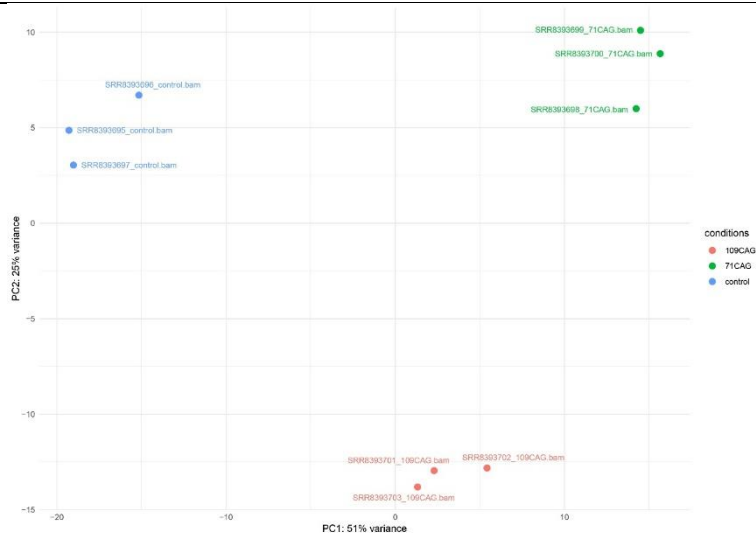
- Forward



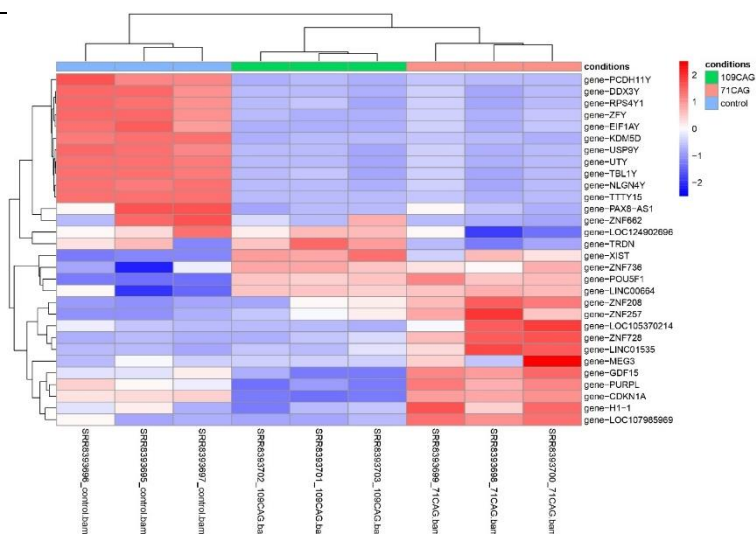


Anexo 7. Gráficos de análisis DEG de las muestras completas

Análisis de componentes principales PCA



Mapa de calor Heatmap



Anexo 8. Variantes detectadas en las muestras 71CAG

Cromosoma	Posición	ID	Referencia	Alteración	Qual	Formato	71CAG
chr4	3135968	chr4:3135968_	G	C	504,2	GT:GL:DP:AD:RO:Q	1/1:-62.8959,-
		G/C			74	R:AO:QA	1.44706,0:25:2,23:2:64:23:763
chr4	3177389	chr4:3177389_	T	G	355,6	GT:GL:DP:AD:RO:Q	1/1:-41.0177,-
		T/G			98	R:AO:QA	3.61236,0:12:0,12:0:0:12:452
chr4	3180513	chr4:3180513_	AGC	ATGC	174,4	GT:GL:DP:AD:RO:Q	0/1:-23.7855,0,-
		AGC/ATGC			71	R:AO:QA	13.8053:17:5,8:5:197:8:308
chr4	3140657	chr4:3140657_	G	A	167,5	GT:GL:DP:AD:RO:Q	0/1:-25.304,0,-
		G/A			09	R:AO:QA	18.3058:18:7,10:7:260:10:338
chr4	3099260	chr4:3099260_	CTTTTTTTGC	CTTTTTTTGCT	130,9	GT:GL:DP:AD:RO:Q	1/1:-15.1499,-
		CTTTTTTTTGC			39	R:AO:QA	1.20412,0:4:0,4:0:0:4:164
		TTA/CTTTTTTT					
chr4	3233354	chr4:3233354_	G	C	120,0	GT:GL:DP:AD:RO:Q	0/1:-16.9488,0,-
		G/C			52	R:AO:QA	1.98736:7:1,6:1:41:6:208
chr4	3125630	chr4:3125630_	G	C	104,3	GT:GL:DP:AD:RO:Q	1/1:-14.049,-
		G/C			66	R:AO:QA	1.20412,0:4:0,4:0:0:4:152
chr4	3173132	chr4:3173132_	G	A	89,35	GT:GL:DP:AD:RO:Q	0/1:-14.7,0,-
		G/A			85	R:AO:QA	19.2088:11:6,5:6:246:5:196
chr4	3134542	chr4:3134542_	T	C	82,25	GT:GL:DP:AD:RO:Q	1/1:-11.8375,-
		T/C			11	R:AO:QA	1.20412,0:4:0,4:0:0:4:128

chr4	3127257	chr4:3127257_ TCTTCA/CCT TAACA	TCTTCA	CCTTAACA	80,16 38	GT:GL:DP:AD:RO:Q R:AO:QA	0/1:-9.86279,0,- 0.995812:4:1,3:1:22:3:120
chr4	3212560	chr4:3212560_ TCAG/TG	TCAG	TG	61,98 93	GT:GL:DP:AD:RO:Q R:AO:QA	1/1:-7.40278,- 0.60206,0:2:0,2:0:0:2:78
chr4	3172398	chr4:3172398_ G/T	G	T	59,73 77	GT:GL:DP:AD:RO:Q R:AO:QA	1/1:-7.77968,- 0.60206,0:2:0,2:0:0:2:82
chr4	3215108	chr4:3215108_ A/C	A	C	56,17 99	GT:GL:DP:AD:RO:Q R:AO:QA	0/1:-11.9117,0,- 34.5015:16:11,5:11:433:5:182
chr4	3103823	chr4:3103823_ G/A	G	A	54,42 42	GT:GL:DP:AD:RO:Q R:AO:QA	0/1:-10.6435,0,- 22.5312:11:7,4:7:283:4:151
chr4	3136328	chr4:3136328_ T/C	T	C	49,41 56	GT:GL:DP:AD:RO:Q R:AO:QA	0/1:-8.58953,0,- 2.89045:4:1,3:1:41:3:105
chr4	3130424	chr4:3130424_ G/C	G	C	44,23 67	GT:GL:DP:AD:RO:Q R:AO:QA	0/1:-11.548,0,- 12.4406:9:4,5:4:164:5:155
chr4	3235399	chr4:3235399_ GTGA/ATGT	GTGA	ATGT	39,75 66	GT:GL:DP:AD:RO:Q R:AO:QA	0/1:-9.18701,0,- 5.86507:6:2,3:2:82:3:119
chr4	3131636	chr4:3131636_ A/T	A	T	21,02 17	GT:GL:DP:AD:RO:Q R:AO:QA	0/1:-6.9476,0,- 25.7621:11:8,3:8:319:3:110
chr4	3199729	chr4:3199729_ TAG/TG	TAG	TG	20,32 54	GT:GL:DP:AD:RO:Q R:AO:QA	0/1:-6.25909,0,- 6.07428:5:2,2:2:80:2:82
chr4	3189095	chr4:3189095_ T/A	T	A	19,67 9	GT:GL:DP:AD:RO:Q R:AO:QA	0/1:-6.27453,0,- 9.95964:5:3,2:3:123:2:82

chr4	3132722	chr4:3132722_ T/A	T	A	16,23 88	GT:GL:DP:AD:RO:Q R:AO:QA	0/1:-5.92475,0,- 9.25034:6:3,2:3:119:2:82
chr4	3223405	chr4:3223405_ GGA/GAAGA	GGA	GAAGA	12,37 06	GT:GL:DP:AD:RO:Q R:AO:QA	0/1:-5.44205,0,- 6.17602:6:2,2:2:82:2:74
chr4	3122938	chr4:3122938_ T/C	T	C	11,11 36	GT:GL:DP:AD:RO:Q R:AO:QA	0/1:-6.69861,0,- 30.2411:14:10,4:10:379:4:118
chr4	3115447	chr4:3115447_ T/C	T	C	7,006 63	GT:GL:DP:AD:RO:Q R:AO:QA	0/1:-5.37144,0,- 19.3845:8:6,2:6:238:2:82
chr4	3172907	chr4:3172907_ G/A	G	A	4,516 38	GT:GL:DP:AD:RO:Q R:AO:QA	0/1:-4.57354,0,- 9.95964:5:3,2:3:123:2:64
chr4	3154420	chr4:3154420_ G/C	G	C	4,516 35	GT:GL:DP:AD:RO:Q R:AO:QA	0/1:-4.57354,0,- 9.58935:5:3,2:3:119:2:64
chr4	3145229	chr4:3145229_ G/A	G	A	2,515 8	GT:GL:DP:AD:RO:Q R:AO:QA	0/1:-3.75164,0,- 3.19148:3:1,2:1:41:2:49
chr4	3121433	chr4:3121433_ G/C	G	C	0,396 799	GT:GL:DP:AD:RO:Q R:AO:QA	0/1:-3.54198,0,- 26.88:10:8,2:8:328:2:69
chr4	3130303	chr4:3130303_ A/G	A	G	0,157 83	GT:GL:DP:AD:RO:Q R:AO:QA	0/1:-4.16732,0,- 32.0188:12:10,2:10:392:2:82

Nota: (Elaborado por el autor).

Anexo 9. Variantes detectadas en las muestras 109CAG

Cromosoma	Posición	ID	Referencia	Alteración	Qual	Formato	109CAG
chr4	3135968	chr4:3135968 _G/C	G	C	352,372	GT:GL:DP:AD:RO:QR:A O:QA	0/1:-50.2134,0,- 6.22768:27:4,19:4:146:19:635
chr4	3223405	chr4:3223405 _G/A	G	A	290,542	GT:GL:DP:AD:RO:QR:A O:QA	0/1:-39.2428,0,- 2.71477:14:2,12:2:73:12:479
chr4	3180513	chr4:3180513 _AGC/ATG C	AGC	ATGC	137,069	GT:GL:DP:AD:RO:QR:A O:QA	0/1:-20.063,0,- 21.5949:21:7,7:7:287:7:270
chr4	3134542	chr4:3134542 _T/C	T	C	132,457	GT:GL:DP:AD:RO:QR:A O:QA	1/1:-16.3632,- 1.50515,0:5:0,5:0:0:5:178
chr4	3132560	chr4:3132560 _A/T	A	T	112,501	GT:GL:DP:AD:RO:QR:A O:QA	0/1:-22.4679,0,- 4.46835:11:2,9:2:82:9:283
chr4	3127257	chr4:3127257 _TCTTCA/C CTTAACA	TCTTCA	CCTTAACA	105,658	GT:GL:DP:AD:RO:QR:A O:QA	0/1:-16.4547,0,- 5.48436:7:2,5:2:80:5:202
chr4	3134399	chr4:3134399 _A/G	A	G	90,7457	GT:GL:DP:AD:RO:QR:A O:QA	0/1:-17.3096,0,- 3.74536:9:2,7:2:68:7:219
chr4	3235399	chr4:3235399 _GTGA/AT GT	GTGA	ATGT	87,248	GT:GL:DP:AD:RO:QR:A O:QA	1/1:-10.4798,- 0.90309,0:3:0,3:0:0:3:116
chr4	3236256	chr4:3236256 _T/A	T	A	83,853	GT:GL:DP:AD:RO:QR:A O:QA	1/1:-11.0945,- 0.90309,0:3:0,3:0:0:3:119

chr4	3233354	chr4:3233354 _G/C	G	C	83,2534	GT:GL:DP:AD:RO:QR:A O:QA	0/1:-13.4333,0,- 3.87256:7:2,5:2:63:5:169
chr4	3140657	chr4:3140657 _G/A	G	A	78,088	GT:GL:DP:AD:RO:QR:A O:QA	0/1:-13.1082,0,- 4.11268:9:2,5:2:69:5:169
chr4	3122938	chr4:3122938 _T/C	T	C	72,2314	GT:GL:DP:AD:RO:QR:A O:QA	0/1:-13.9014,0,- 25.7964:15:9,6:9:333:6:201
chr4	3212144	chr4:3212144 _T/G	T	G	71,2272	GT:GL:DP:AD:RO:QR:A O:QA	0/1:-12.1486,0,- 5.9735:6:2,4:2:82:4:151
chr4	3199729	chr4:3199729 _TAG/TG	TAG	TG	59,7377	GT:GL:DP:AD:RO:QR:A O:QA	1/1:-7.77968,- 0.60206,0:2:0,2:0:0:2:82
chr4	3177389	chr4:3177389 _T/G	T	G	56,5845	GT:GL:DP:AD:RO:QR:A O:QA	0/1:-10.2607,0,- 2.89045:4:1,3:1:41:3:123
chr4	3209826	chr4:3209826 _G/A	G	A	42,7281	GT:GL:DP:AD:RO:QR:A O:QA	1/1:-6.07869,- 0.60206,0:2:0,2:0:0:2:64
chr4	3125630	chr4:3125630 _G/C	G	C	40,0586	GT:GL:DP:AD:RO:QR:A O:QA	0/1:-9.21906,0,- 6.27453:5:2,3:2:82:3:115
chr4	3103884	chr4:3103884 _G/A	G	A	37,1587	GT:GL:DP:AD:RO:QR:A O:QA	0/1:-6.87659,0,- 2.29622:3:1,2:1:32:2:82
chr4	3215108	chr4:3215108 _A/C	A	C	35,4078	GT:GL:DP:AD:RO:QR:A O:QA	0/1:-9.50531,0,- 39.0936:16:12,4:12:484:4:155
chr4	3154420	chr4:3154420 _G/C	G	C	30,201	GT:GL:DP:AD:RO:QR:A O:QA	0/1:-6.87659,0,- 3.19148:3:1,2:1:41:2:82

chr4	3239845	chr4:3239845 _G/T	G	T	21,019	GT:GL:DP:AD:RO:QR:A O:QA	0/1:-5.72507,0,- 5.72507:4:2,2:2:73:2:73
chr4	3173132	chr4:3173132 _G/A	G	A	17,6544	GT:GL:DP:AD:RO:QR:A O:QA	0/1:-9.59933,0,- 39.5133:17:12,5:12:492:5:160
chr4	3211927	chr4:3211927 _A/C	A	C	17,364	GT:GL:DP:AD:RO:QR:A O:QA	0/1:-5.34816,0,- 5.25051:4:2,2:2:68:2:69
chr4	3115447	chr4:3115447 _T/C	T	C	15,1718	GT:GL:DP:AD:RO:QR:A O:QA	0/1:-6.71583,0,- 33.3472:13:10,3:10:410:3:114
chr4	3131636	chr4:3131636 _A/T	A	T	13,5737	GT:GL:DP:AD:RO:QR:A O:QA	0/1:-5.04713,0,- 9.12407:5:3,2:3:114:2:69
chr4	3130424	chr4:3130424 _G/C	G	C	13,3907	GT:GL:DP:AD:RO:QR:A O:QA	0/1:-5.1756,0,- 3.19148:3:1,2:1:41:2:64
chr4	3172907	chr4:3172907 _G/A	G	A	8,24944	GT:GL:DP:AD:RO:QR:A O:QA	0/1:-6.24881,0,- 18.6265:9:6,3:6:233:3:96
chr4	3233161	chr4:3233161 _A/T	A	T	7,31958	GT:GL:DP:AD:RO:QR:A O:QA	0/1:-4.87457,0,- 6.57556:4:2,2:2:82:2:64
chr4	3103823	chr4:3103823 _G/A	G	A	6,92744	GT:GL:DP:AD:RO:QR:A O:QA	0/1:-6.55891,0,- 39.4539:16:12,4:12:488:4:123
chr4	3130303	chr4:3130303 _A/G	A	G	6,44198	GT:GL:DP:AD:RO:QR:A O:QA	0/1:-7.69913,0,- 59.3081:22:18,4:18:729:4:155
chr4	3199940	chr4:3199940 _G/A	G	A	3,60213	GT:GL:DP:AD:RO:QR:A O:QA	0/1:-5.07041,0,- 23.1332:9:7,2:7:283:2:82

chr4	3132722	chr4:3132722 _T/A	T	A	3,41842	GT:GL:DP:AD:RO:QR:A O:QA	0/1:-3.78738,0,- 9.61733:6:3,2:3:123:2:59
chr4	3182471	chr4:3182471 _G/A	G	A	0,334313	GT:GL:DP:AD:RO:QR:A O:QA	0/1:-3.67045,0,- 19.2912:8:6,2:6:237:2:64
chr4	3130048	chr4:3130048 _G/C	G	C	6,43898E- 06	GT:GL:DP:AD:RO:QR:A O:QA	0/1:-0.961184,0,- 48.6778:17:15,2:15:594:2:64

Nota: (Elaborado por el autor).

Anexo 10. Predicciones del efecto de las variantes 71CAG y 109CAG utilizando VEP

[https://utaedu-](https://utaedu-my.sharepoint.com/:x:/g/personal/csilva6104_uta_edu_ec/EaQX_wQRsYdGkDKjpAHu6pgBdAJhAi2DSyr4AyU2Cf4FJw?e=ZMwHh5)

[my.sharepoint.com/:x:/g/personal/csilva6104_uta_edu_ec/EaQX_wQRsYdGkDKjpAHu6pgBdAJhAi2DSyr4AyU2Cf4FJw?e=ZMwHh5](https://utaedu-my.sharepoint.com/:x:/g/personal/csilva6104_uta_edu_ec/EaQX_wQRsYdGkDKjpAHu6pgBdAJhAi2DSyr4AyU2Cf4FJw?e=ZMwHh5)

Anexo 11. Eventos totales de splicing alternativo

Tipo de evento		Total, Eventos JC	TotalEventos JCEC	Sig.Even- tos JC	Sig.Even- tosJC Control inclusió n alta	Sig.Even- tos JC 71CAG inclusió n alta	Signific- antEve- nts JCEC	Sig.Even- tos JCEC Control inclusió n alta	Sig.Even- tos JCEC 109CA G inclusió n alta
CONTROL VS 71 CAG									
SE	skipped exon	59832	61688	1466	740	726	1484	737	747
A5SS	alternati- ve 5' splice sites	5480	5503	169	78	91	189	89	100
A3SS	alternati- ve 3' splice sites	8468	8496	194	95	99	226	109	117
MXE	mutuall y exclusiv e exons	7992	8297	139	48	91	153	70	83
RI	retained intron	6209	6267	240	109	131	360	165	195
CONTROL VS 109 CAG									
SE	skipped exon	62142	63918	1450	832	618	1522	863	659
A5SS	alternati- ve 5'	5567	5587	167	93	74	198	118	80

	splice sites								
A3SS	alternati ve 3' splice sites	8593	8615	144	66	78	163	71	92
MXE	mutuall y exclusiv e exons	8458	8751	194	117	77	194	119	75
RI	retained intron	6236	6301	326	155	171	489	260	229

Nota: (Elaborado por el autor).

Anexo 12. Eventos de splicing alternativo Control vs 71CAG del gen HTT

https://utaedu-my.sharepoint.com/:x/g/personal/csilva6104_uta_edu_ec/EflLhkPVuJNgOdRRrNdU6YBUbSAys9QNv9gWT-IiHM5tQ?e=QcQNbG

Anexo 13. Eventos de splicing alternativo Control vs 109CAG del gen HTT

https://utaedu-my.sharepoint.com/:x/g/personal/csilva6104_uta_edu_ec/EaCfTAnbpqhLuvTGA7J23ogBfFZv5mnWwsalm43lJHi1fQ?e=8dT5r