

**Universidad ORT Uruguay**

**Facultad de Ingeniería**

**DESARROLLO DE UN *PIPELINE* DE  
TRABAJO PARA SECUENCIACIÓN  
DIRIGIDA POR CRISPR-CAS9**

Entregado como requisito para la obtención del título de Ingeniero en Biotecnología

**Ana María Ibáñez Oliver - 191615**

**Tutor: Dr. Bernardo Bertoni  
Co-tutor: Dra. Mónica Cappetta**

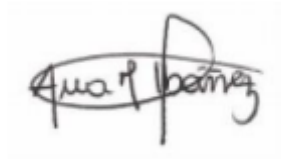
**2021**

## Declaración de autoría

---

Yo, Ana María Ibáñez, declaro que el trabajo que se presenta en esa obra es de mi propia mano. Puedo asegurar que:

- La obra fue producida en su totalidad mientras realizaba el trabajo final de carrera;
- Cuando he consultado el trabajo publicado por otros, lo he atribuido con claridad;
- Cuando he citado obras de otros, he indicado las fuentes. Con excepción de estas citas, la obra es enteramente mía;
- En la obra, he acusado recibo de las ayudas recibidas;
- Cuando la obra se basa en trabajo realizado conjuntamente con otros, he explicado claramente qué fue contribuido por otros, y qué fue contribuido por mi;
- Ninguna parte de este trabajo ha sido publicada previamente a su entrega, excepto donde se han realizado las aclaraciones correspondientes.



Ana María Ibáñez  
10 de Agosto 2021

## Agradecimientos

---

En primer lugar quiero agradecer a mi tutor Bernardo Bertoni y a mi co-tutora Mónica Cappetta por guiarme en esta etapa fundamental en mi carrera, por su dedicación y acompañamiento durante toda la tesis.

A todos mis compañeros de la Isla, Silvana, Lucía Bri, Lucía Be, Florencia y Angela por formar un equipo de trabajo excelente, donde encontré amabilidad, compañerismo y alegría.

A la Facultad de Medicina de la UDELAR, especialmente al departamento de Genética por dejarme realizar la tesis en su laboratorio.

A la Universidad ORT por estos años de aprendizaje.

Un agradecimiento especial a mis compañeros de carrera, por la amistad que forjamos durante estos años. Por esa palabra de aliento y compañerismo que tuvimos.

A mis amigas, por ser un pilar fundamental en mi vida, mis hermanas.

A mi familia por estar toda la vida acompañándome en todo. Gracias a mis padres por su fuerza, motivación y amor incondicional. Gracias a mis hermanos por siempre estar ahí para mí. Gracias a mis cuñadas. Y un gracias especial a Rafa, Marti, Pili y Bian, mis sobrinos, que son la alegría de esta familia.

## Resumen

---

Hace cuarenta años, el advenimiento de la secuenciación de Sanger fue revolucionario, ya que permitió descifrar por primera vez secuencias completas del genoma. Se produjo una segunda revolución cuando aparecieron las tecnologías de secuenciación de segunda generación, que hicieron que la secuenciación del genoma fuera mucho más barata y rápida. Sin embargo, estos métodos tienen varios inconvenientes, sobre todo sus lecturas cortas. Recientemente, aparecieron métodos de tercera generación que pueden producir lecturas largas del ADN. Además, estas tecnologías pueden detectar directamente modificaciones epigenéticas en el ADN nativo y permitir la secuenciación completa sin necesidad de amplificación por PCR. Esto marca la tercera revolución en la tecnología de secuenciación.

En este trabajo vamos a enfocarnos en una de las más recientes tecnologías de secuenciación de tercera generación, MinION de Oxford Nanopore, el dispositivo de secuenciación más pequeño disponible.

Nuestro objetivo fue desarrollar un *pipeline* de trabajo para la secuenciación dirigida de fragmentos de más de 10 kb en la plataforma MinION.

Para analizar los fragmentos, se realizó la secuenciación por minION de Oxford Nanopore mediante una técnica basada en CRISPR-Cas9. Dicha técnica utiliza la capacidad de la Cas9 para introducir cortes en ubicaciones específicas y poder luego ligar adaptadores de secuenciación directamente a esos sitios (nCATS). Esta técnica no necesita un proceso de amplificación por lo que se obtiene información directamente de la molécula nativa de ADN.

El desarrollo del pipeline incluyó ajustes a los protocolos de extracción de ADN para evitar fragmentación, desarrollo y puesta a punto de un método de evaluación de la eficiencia de corte del sistema CRISPR-Cas9 en la región de interés mediante PCR en tiempo real y secuenciación con 2 tipos de celdas de la región del gen *MC1R*. Además, se realizó un análisis económico de la secuenciación dirigida por nCATS.

Si bien se obtuvieron *reads* en la secuencia *target* en las pruebas de secuenciación, la cobertura no fue suficiente. Se detectaron algunos problemas posibles que podrían explicar esta baja cobertura. Es necesario más pruebas de secuenciación con mejores condiciones de las celdas para evaluar de manera adecuada la eficiencia del *pipeline* desarrollado.

El potencial de esta técnica, junto con la baja inversión del minION podría ser un gran avance en la investigación y la medicina translacional, ya que permitiría estudiar variantes génicas, estructurales, haplotipos y metilación de grandes regiones genómicas de los individuos a un bajo costo.

## Palabras claves

---

MinION, CRISPR-Cas9, nCATS, *MCIR*, secuenciación

## Abreviaturas

---

μl - microlitro

°C - grados Celsius

ADN - Ácido Desoxirribonucleico

ARN - Ácido ribonucleico

Buffer TE - buffer compuesto principalmente por Tris y EDTA

Cas9 - endonucleasa Cas9

cm - centímetro

crARN - ARN guía de la Cas9

CRISPR - *clustered regularly interspaced short palindromic repeats*

Ct - corte de la línea de amplificación con *threshold*

DSB - corte doble hebra

g - gramos

kb - kilobase

M - Molaridad

MC1R - Receptor melanocortina-1

mg - miligramo

mL - mililitro

mM - Milimolar

nCATS - Metodología de secuenciación dirigida por CRISPR-CAS9

nM - Nanomolar

ng/uL - Nanogramo por microlitro

nm - nanómetro

nt - nucleótidos

PAM - motivo adyacente de protoespaciador

pb - pares de bases

PCR - Reacción en Cadena de la Polimerasa

RNP - ribonucleoproteína

rpm - revoluciones por minuto

SGS - Secuenciación de segunda generación

SNP - Polimorfismo de nucleótido único

SV - Variantes estructurales del ADN

TCF25 - Factor de transcripción 25

TGS - Secuenciación de tercera generación

Tm - Temperatura de melting

tracrARN - ARN trans activador

TUBB3 - β-tubulina III

# Índice

---

Introducción .....	9
1. Metodologías de Secuenciación .....	9
1.1. MinION de Oxford Nanopore .....	11
1.1.1. Secuenciación de regiones específicas.....	13
2. Long PCR.....	13
3. CRISPR-Cas9 .....	14
3.1. nCATS .....	16
3.1.1. Análisis Bioinformático.....	17
4. Estudios de Melanoma cutáneo maligno en el Laboratorio de Epidemiología Genética 18	
4.1. Genes de la región del <i>MC1R</i> .....	19
4.1.1. <i>MC1R</i> .....	19
4.1.2. <i>TUBB3</i> y <i>TCF25</i> .....	20
Objetivos.....	21
1. Objetivo general .....	21
2. Objetivos Específicos .....	21
Diseño Experimental .....	22
Materiales y métodos.....	23
1. Extracción de ADN .....	23
2. Puesta a punto de PCR tiempo final para amplificar los genes <i>TCF25</i> , <i>MC1R</i> y <i>TUBB3</i> .....	23
2.1. Diseño de cebadores para Long PCR .....	23
2.2. Long PCR.....	25
2.3. Visualización de productos de PCR en gel de Agarosa.....	26
3. Secuenciación dirigida por Cas9 .....	26
3.1. Diseño de ARN guías.....	26
3.2. Evaluación del corte con Cas9 .....	29
3.2.1. Puesta a punto de PCR en tiempo real.....	30
3.2.1.1. Diseño de cebadores .....	30
3.2.1.2. PCR en tiempo real.....	30
3.2.1.3. PCR tiempo final .....	31
3.3. Secuenciación dirigida por Cas9 .....	31

3.3.1. Obtención de fragmentos de la región de interés .....	31
3.3.2. Preparación de librerías para secuenciación .....	33
3.4. Bioinformática .....	33
Resultados y discusión .....	34
1. Puesta a punto de PCR tiempo final para amplificar los genes <i>TCF25</i> , <i>MC1R</i> y <i>TUBB3</i> .....	34
1.1. Long PCR con el mix de Taq Polimerasa .....	34
1.2. Long PCR con la enzima Ranger ADN Polimerasa .....	34
2. Secuenciación dirigida por Cas9 .....	36
2.1. Puesta a punto de la evaluación del corte con CAS9 .....	36
2.1.1. Evaluación de los cortes de Cas9 por PCR en tiempo real .....	36
2.2. Secuenciación por MinION .....	38
Análisis Económico .....	49
Conclusiones y perspectivas a futuro .....	52
Bibliografía .....	53
Anexo .....	57



# Introducción

---

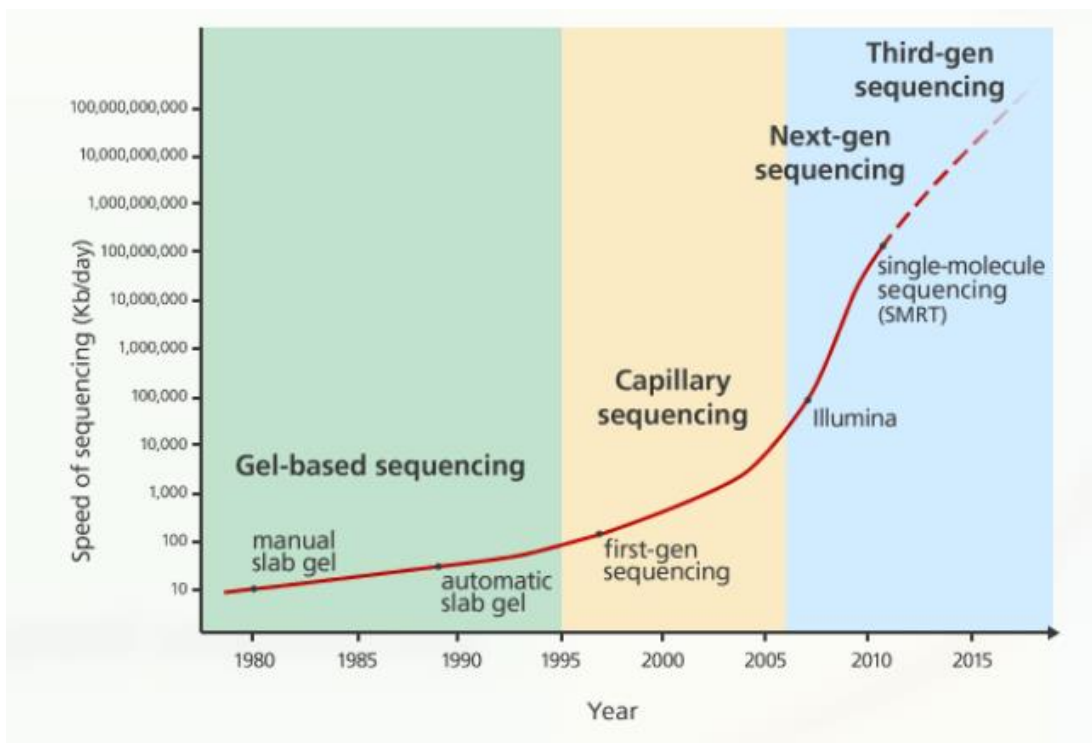
## 1. Metodologías de Secuenciación

Desde su introducción, la secuenciación del ADN ha revolucionado la comprensión de la genética humana, dando paso a grandes logros históricos como el Proyecto Genoma Humano y jugando un papel importante en los avances de la medicina. El Proyecto Genoma Humano duró aproximadamente 13 años y se completó por primera vez con un costo de casi 3 mil millones de dólares <sup>1,2</sup>.

Frederick Sanger inició, en 1977, el campo de la genómica con su revolucionario desarrollo de la secuenciación del ADN por el método de terminación de cadenas, conocida como secuenciación de primera generación. Esta técnica permitió la fundación de la Genómica gracias a la primera secuenciación completa de un genoma un organismo, el bacteriófago  $\phi$ X174 (5.386 pb). Se trata de un método de secuenciación por síntesis, es decir, basado en la replicación de una molécula de ADN por la ADN polimerasa. A partir de un oligonucleótido cebador la ADN polimerasa origina un conjunto de oligonucleótidos de diferentes tamaños y complementarios a la hebra molde. El método de Sanger se hizo con el monopolio de la secuenciación de ADN durante cerca de dos décadas que coinciden con los primeros avances importantes de la Genómica: secuenciación del primer procariota (*Haemophilus influenzae*; 1,83 Mb), el primer genoma eucariota (*Saccharomyces cerevisiae*; 12,1 Mb), el primer genoma de una planta (*Arabidopsis thaliana*; 157 Mb) o el genoma humano (*Homo sapiens*; 3,2 Gb). Además, la automatización de esta técnica fue definitiva para la consecución del Proyecto Genoma Humano. A pesar de los grandes avances en los métodos de secuenciación experimentados en los últimos quince años (Figura 1), que han cambiado los estudios genómicos, el método inventado por Sanger en 1977 sigue siendo el mejor para la secuenciación de un fragmento individual de ADN en el contexto de la manipulación de los ácidos nucleicos en el laboratorio. Sin este pionero de la Biología Molecular nada de lo hecho posteriormente existiría <sup>3</sup>.

Pero el bajo rendimiento, el alto costo y la necesidad de satisfacer la demanda de secuenciar genomas completos llevaron a un cambio fundamental en la metodología, implementando así la secuenciación de segunda generación (SGS). En 2004, el Instituto Nacional de Investigación del Genoma Humano (NHGRI) inició un programa para reducir el costo de la secuenciación del genoma completo a USD 1000 en 10 años. Esto aceleró el desarrollo de métodos más baratos y más rápidos <sup>4</sup>. En contraste con el método de Sanger, esta nueva tecnología (SGS) mejoró el rendimiento generando millones de secuencias de manera más rápida y redujo los costos a través de una variedad de métodos novedosos que incluyen pirosecuenciación (Roche 454), secuenciación por síntesis (Genome Analyzer, Illumina) y secuenciación por ligación (SOLiD) <sup>1,5</sup>. Sin embargo, la desventaja que presentan es la incapacidad de generar lecturas superiores a 700 pb. La SGS tiene un gran impacto en las mutaciones *de novo* en enfermedades raras en los últimos años. Sin embargo, muchas enfermedades raras todavía no son diagnosticadas completamente por la SGS debido a los métodos de lectura corta. Las variantes

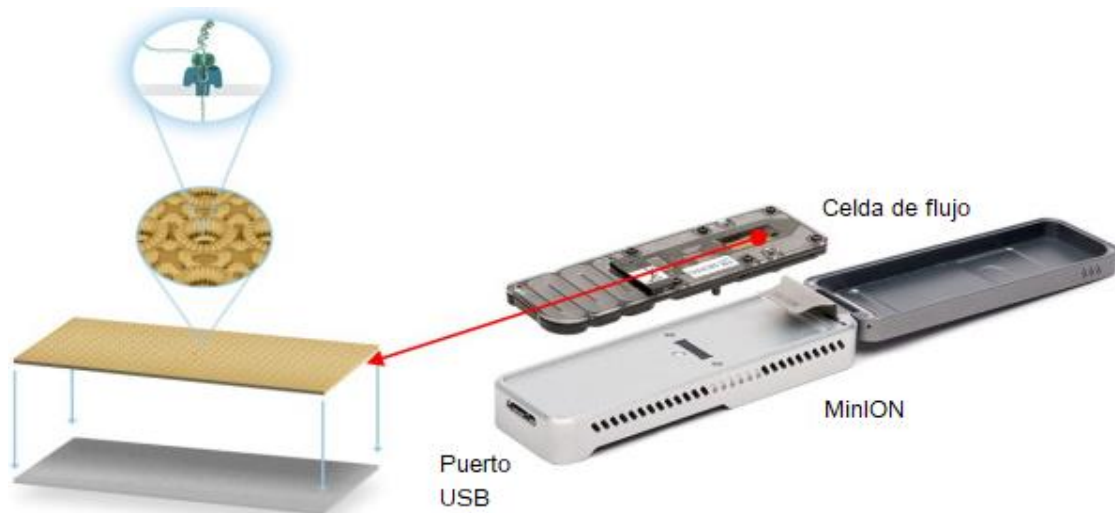
estructurales (SV), los elementos repetitivos, el contenido extremo de guanina-citosina (GC) o las secuencias con múltiples elementos homólogos en el genoma son difíciles de caracterizar mediante SGS, incluso con el uso de algoritmos bioinformáticos de última generación. Estos inconvenientes de las investigaciones de enfermedades humanas basadas en SGS han impulsado fuertemente la búsqueda de otros métodos para mejorar la precisión y reducir el tiempo de diagnóstico en enfermedades genéticas, dando entonces lugar a la creación de la secuenciación de tercera generación (TGS) <sup>2</sup>. Las principales diferencias de las plataformas de tercera generación con las de segunda generación radican en la no necesidad de realizar una PCR antes de la secuenciación, reduciendo el tiempo de análisis y los errores derivados de ésta, y en que la detección de nucleótidos se realiza en tiempo real. Además, se agregan las ventajas relacionadas con el menor costo, la velocidad de secuenciación y la posibilidad de obtener la lectura de grandes secuencias de ADN <sup>6</sup>. Las tres tecnologías de secuenciación de ADN de tercera generación disponibles comercialmente son la *Single Molecule Real Time* (SMRT) de Pacific Biosciences (PacBio), Illumina Tru-seq Synthetic Long-Read y la plataforma de secuenciación de Oxford Nanopore Technologies. Utilizando la secuenciación de una sola molécula o la amplificación clonal y la secuenciación de moléculas largas, las tres tecnologías pueden producir lecturas largas con un promedio de entre 5.000 pb y 15.000 pb, con algunas lecturas superiores a 100.000 pb <sup>7</sup>.



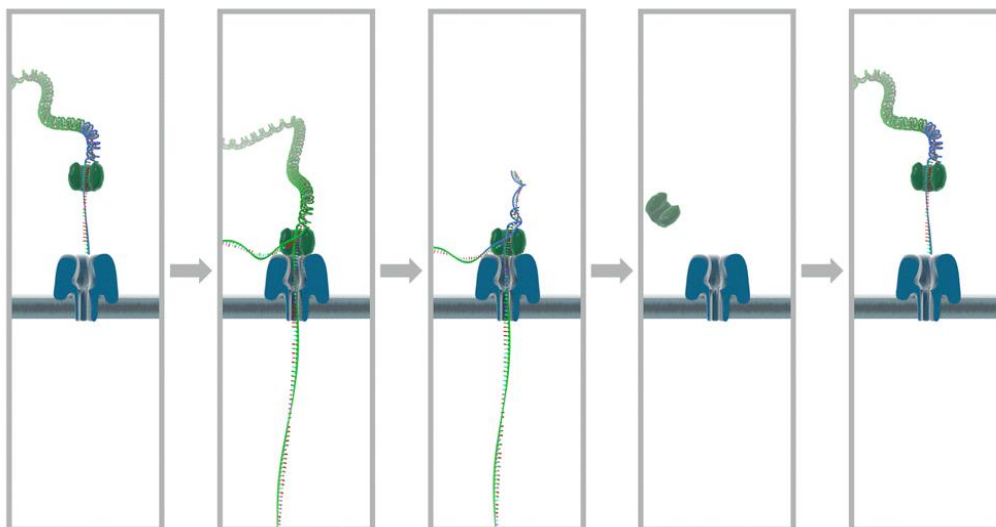
**Figura 1.** Gráfica de la velocidad de secuenciación en función de la plataforma tecnológica y el tiempo, que muestra cómo ha aumentado la velocidad de secuenciación de ADN desde las primeras técnicas en la década de 1980 <sup>8</sup>.

## 1.1. MinION de Oxford Nanopore

En 2014, Oxford Nanopore Technologies lanzó una nueva plataforma TGS, el dispositivo MinION. Con solo  $10 \times 3 \times 2$  cm y un peso de 90 g, el MinION es el dispositivo de secuenciación más pequeño disponible actualmente (Figura 2). Se puede conectar directamente a un puerto USB estándar en una computadora con pocos requisitos de hardware y una configuración simple. El sistema consiste en hacer pasar una corriente eléctrica a través de membranas resistentes a la electricidad, donde están fijados una serie de nanoporos proteicos. A medida que el ADN pasa a través de los nanoporos, se producen alteraciones en la corriente eléctrica, emitiendo señales que definen la secuencia de bases (Figura 3) <sup>9</sup>.



**Figura 2.** Dispositivo de secuenciación MinION <sup>10</sup>.



**Figura 3.** De izquierda a derecha, el ADN doble hebra unido a la proteína motora se adhiere a una proteína del poro en la membrana. El potencial aplicado tira de una hebra a través del poro, mientras que la proteína motora descomprime el ADN. Una vez descomprimido el ADN, la proteína motora se desprende y el poro está listo para recibir otra hebra. Imagen tomada de Oxford Nanopore.

Al igual que la tecnología de SGS, es necesario la preparación de una biblioteca para diversas aplicaciones que usan MinION. Se diferencia de otras tecnologías en que puede utilizar ADN más largo y bicatenario (dsDNA) para poder realizar la secuenciación de ambas cadenas. El proceso de construcción de la biblioteca consta de una serie de pasos, que se llevan a cabo en el siguiente orden: Obtención de los fragmentos de ADN a secuenciar; reparación para crear extremos romos en fragmentos de ADN y PCR cortados; agregado de colas dA para agregar una base "A" al extremo 3' del fragmento; ligación de adaptadores y finalmente eliminación de nucleótidos libres y enzimas. La biblioteca generalmente contiene dos adaptadores, el adaptador líder y el adaptador de horquilla, cada uno ligado a un extremo del dsDNA. El adaptador líder se denomina "adaptador Y", ya que tiene una estructura en forma de "Y", mientras que el adaptador con estructura de horquilla se denomina "adaptador HP". Estos adaptadores facilitan la captura de la hebra y la carga de una enzima procesadora en el extremo 5' de una hebra. La enzima es necesaria para asegurar el desplazamiento unidireccional de un solo nucleótido a lo largo de la hebra a una escala de tiempo de milisegundos. Los adaptadores también concentran sustratos de ADN en la superficie de la membrana proximal al nanoporo, aumentando la tasa de captura de ADN en varios miles de veces. Además, el adaptador de horquilla permite la secuencia contigua de ambas hebras de una molécula dúplex uniendo covalentemente una cadena a la otra. Al capturar una molécula de ADN en el nanoporo, la enzima se procesa a lo largo de una hebra. Después de que la enzima pasa a través de la horquilla, este proceso se repite para la hebra complementaria <sup>11</sup>. A medida que el ADN pasa a través del poro, el sensor detecta cambios en la corriente iónica causados por diferencias en las secuencias de nucleótidos cambiantes que ocupan el poro. Estos cambios de corriente iónica se segmentan como eventos discretos que tienen una duración, amplitud media y varianza asociadas. Finalmente, el procesamiento de datos lo realiza el software MinKNOW, que se ocupa de la adquisición y el análisis de datos <sup>10</sup>.

El software especializado llamado MinKNOW se ejecuta en la computadora conectada al dispositivo MinION. MinKNOW realiza varias tareas centrales: adquisición de datos; análisis y retroalimentación en tiempo real; transmisión de datos mientras se proporciona control del dispositivo (incluida la selección de parámetros de ejecución); así como la identificación y el seguimiento de muestras, asegurando así que la química de la plataforma funcione correctamente en el procesamiento de las muestras <sup>11</sup>.

Existen dos tipos de celdas de flujo que se utilizan con el MinION, una celda de flujo estándar que contiene hasta 512 canales de nanoporos para secuenciar ADN en tiempo real y por otro lado ONT ha producido una celda de flujo llamada Flongle que es más pequeña y más barata que la celda de flujo normal. Flongle se monta en un adaptador Flongle que contiene la matriz de sensores de ONT y es compatible con MinION. Una Flongle alberga 126 canales de secuenciación, generalmente está diseñado para una sola muestra en lugar de múltiples y, por lo tanto, se usa para evaluar rápidamente la calidad de la muestra antes de escalar. Flongle está diseñado para ser el sistema de secuenciación más rápido, accesible y rentable para pruebas y experimentos más pequeños o realizados con mayor frecuencia <sup>12</sup>.

Las ventajas de MinION sobre otros dispositivos de secuenciación son numerosas. Comenzando por su tamaño, aproximadamente el de un teléfono celular, su costo de inversión

inicial (USD 1000), muy por debajo de sus competidores. Es razonablemente rentable, una secuenciación de 48 horas cuesta actualmente alrededor de 800 dólares y produce hasta 50 Gbases de datos de secuenciación sin procesar. Además, la técnica no se basa en ninguna técnica de marcado para reconocer diferentes bases, mientras que los métodos de secuenciación de Sanger, segunda generación y SMRT requieren alguna forma de marcado de nucleótidos. La amplificación por PCR es opcional para MinION, mientras que este paso es obligatorio para los métodos Sanger y SGS. La omisión de estos pasos no solo simplifica la preparación de muestras, sino que también ayuda a evitar errores y sesgos (por ejemplo, el sesgo CG producido por PCR) y permite la detección de modificaciones en bases. Finalmente, la longitud máxima de lectura (> 4Mb) producida por MinION es muchas veces mayor que la de la secuenciación de primera y segunda generación y solo es paralela a la secuenciación SMRT, lo cual es muy ventajoso <sup>13</sup>.

### **1.1.1 Secuenciación de regiones específicas**

En el desarrollo de estas técnicas se observa un gran desarrollo de métodos dirigidos a secuenciar completamente el genoma de un organismo. Dependiendo del campo de investigación, esto representa una ventaja como es la descripción de organismos con genomas pequeños (genomas bacterianos virales) o en la descripción de organismos más complejos. Las técnicas disponibles para obtener secuencias de fragmentos cromosómicos o genes previo a TGS implicaba un costo casi comparable a secuenciar el genoma completo de la muestra. La disponibilidad de estas plataformas de menor costo y con una química totalmente diferente parecen prometedoras. Hasta no hace mucho tiempo una posibilidad era analizar array hechos a medida sobre variantes conocidas en genes candidatos, otra posibilidad era analizar los exones de estos genes. Esto limita el estudio a variantes conocidas y en el caso del análisis de los exones solo a variantes limitadas a cambios en la proteína.

Sin embargo, las nuevas tecnologías permiten analizar un gen en toda su amplitud pudiendo explorar las regiones reguladoras, intrones y recientemente el estado de metilación de la secuencia analizada.

Para analizar genes candidatos en enfermedades o en el diagnóstico o seguimiento de la enfermedad, contar con sistemas eficiente resulta extremadamente necesario. En el momento actual se cuenta con dos aproximaciones para capturar la información de genes o regiones cromosómicas: por un lado amplificar por PCR fragmentos de más de 5kb y por otro lado aplicar la capacidad de Cas9 de reconocer y cortar el ADN en un sitio determinado previamente.

## **2. Long PCR**

Existen varios métodos de captura adaptados a secuenciación de fragmentos largos de ADN (long PCR, CATCH-seq, etc.), pero están limitados por la pérdida de modificaciones nativas, la longitud de lectura limitada, bajo rendimiento o protocolos largos (> 18 horas). La amplificación del ADN por la reacción en cadena de la polimerasa (PCR) se ha convertido

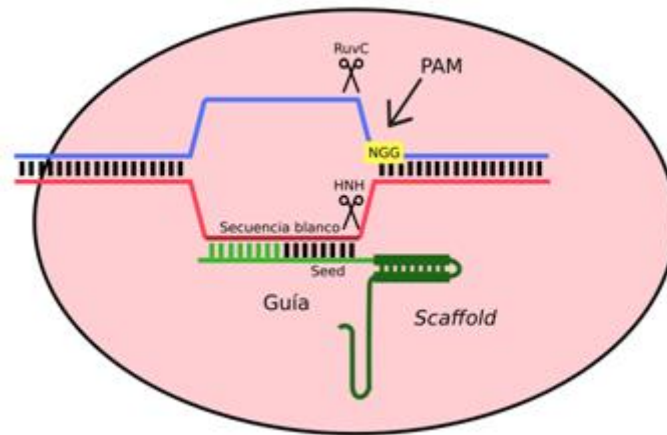
en una herramienta importante y generalizada de análisis genético desde la introducción de la ADN polimerasa termoestable de *Taq* (*Thermus aquaticus*). Es una forma simple y muy rápida de multiplicar el ADN presente en diferentes muestras biológicas, obteniéndose millones de copias de una determinada secuencia de ADN. En poco tiempo esta técnica ha conseguido ser ampliamente utilizada no sólo en el campo de la genética molecular, sino en otras ciencias. Pero este método convencional tiene dos limitaciones, la fidelidad del producto final y el tamaño de la extensión del producto que puede amplificarse (2 a 4 kb máximo) <sup>14</sup>. Si se quieren obtener secuencias más largas es necesario aplicar la técnica de *Long range PCR*. Se trata de una técnica que aumenta la sensibilidad de la PCR. En este caso se trabaja con 4 cebadores, en una primera ronda se amplifica de manera convencional con los dos cebadores más externos a la región que se desea amplificar. El producto de esta primera PCR se utiliza como molde para una segunda ronda que utiliza cebadores internos a la región previamente amplificada <sup>15</sup>.

Esta se basa en la utilización de la Long Amp Hot Start *Taq* DNA Polymerase que ofrece una fidelidad dos veces mayor que Hot Start *Taq* DNA Polymerase sola. Además, se puede generar una amplia gama de productos de PCR de hasta 30 kb de ADN genómico humano <sup>16</sup>.

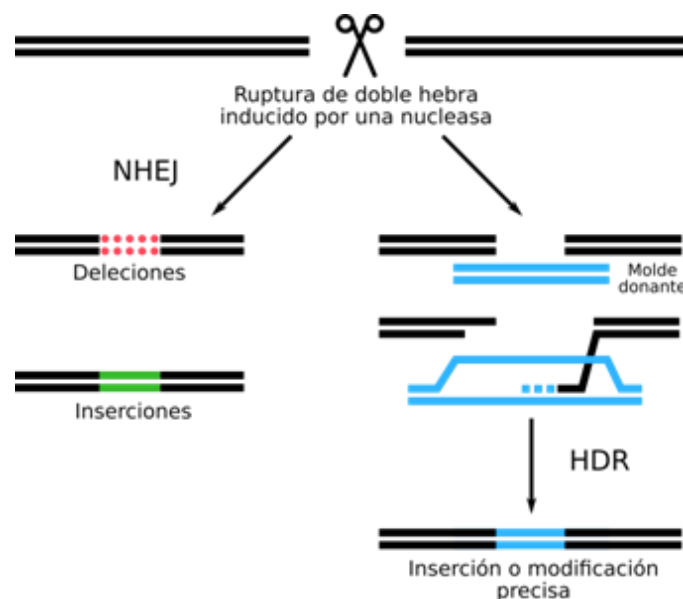
### 3. CRISPR-Cas9

Durante la última década, los avances tecnológicos han facilitado significativamente la edición y regulación del genoma. El sistema CRISPR-Cas9 (*Clustered Regularly Interspaced Short Palindromic Repeats - Cas9*) un mecanismo de defensa bacteriano contra la infección de fagos y la transferencia de plásmidos en la naturaleza, ha sido reutilizado como una potente plataforma guiada por ARN para la edición y regulación del genoma, altamente eficiente y específica en diversos organismos, creando así herramientas revolucionarias para la investigación médica y nuevas posibilidades para el tratamiento de trastornos genéticos. El sistema CRISPR se compone de ARNs cortos que reconocen secuencias de ADN invasor y guían un complejo enzimático para degradarlo. En general, la edición o regulación precisa de la información genómica a nivel de ADN requiere la acción de una maquinaria molecular compuesta de dos partes principales: un dominio de unión al ADN que media en el reconocimiento y unión del ADN específico de la secuencia, y un dominio efector que cliva el ADN o regula la transcripción cerca del sitio de unión. La creación de una ruptura bicatenaria mediante el uso de una endonucleasa específica de secuencia puede estimular la vía de reparación del ADN y aumentar considerablemente la tasa de modificación génica en la secuencia deseada. Por lo tanto, los enfoques mediados por nucleasas se han explorado ampliamente para la edición de genes específicos. En contraste con la mayoría de las proteínas de unión al ADN conocidas, Cas9 es una nucleasa guiada por ARN cuya especificidad de secuencia surge en gran medida del emparejamiento de bases entre su ARN guía y el sitio de *target* de ADN, además de una interacción directa entre la Cas9 y un protospacio adyacente corto (PAM) de ADN. Por lo tanto, Cas9 puede programarse para apuntar a nuevos sitios simplemente cambiando su secuencia de ARN guía, convirtiéndolo en una plataforma ideal para un alto rendimiento <sup>17,18</sup>.

El sistema CRISPR está compuesto por la nucleasa Cas9, que utiliza como guía a un crARN que se aparea con el llamado ARN trans activador (tracrARN), el cual cumple una función estructural. Estos ARNs son considerados el ARN guía (gARN) y forman un complejo con la proteína Cas9. Cada unidad de crARN contiene una secuencia guía de 20 nucleótidos y una repetición parcial directa, donde la primera dirige a Cas9 a un blanco de 20 nucleótidos a través del apareamiento de bases. Cas9 posee los dominios catalíticos RuvC y HNH, además del dominio de interacción con el motivo protoespaciador adyacente (PAM). Cas9 realiza cortes de doble hebra (DSB) utilizando los dominios RuvC y HNH (Figura 4 y 5) <sup>19</sup>.



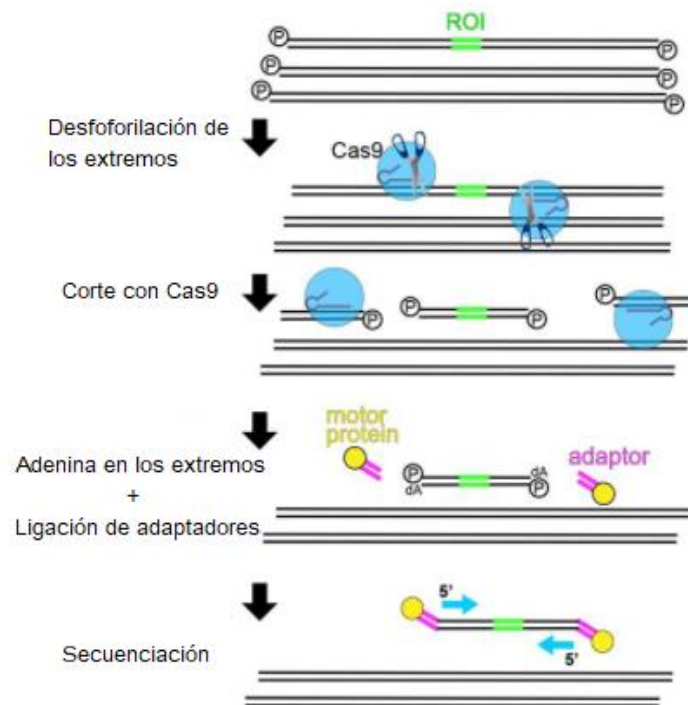
**Figura 4.** Esquema de la nucleasa Cas9 guiada por ARN. Cas9 (en rosado) es apuntada al ADN genómico a través de un crARN que consiste en una secuencia guía de 20 nucleótidos y una secuencia *scaffold*. La secuencia guía se aparea con el ARN blanco, directamente aguas arriba de un motivo adyacente 5'-NGG (PAM). Cas9 media una ruptura de doble hebra alrededor de 3 pb aguas arriba del PAM <sup>19</sup>.



**Figura 5.** Edición génica inducida por una nucleasa. Las rupturas de doble hebra (DSB) inducidas por una nucleasa pueden ser reparadas por vías de unión de extremos no homólogos (NHEJ) o de reparación dirigida por homología (HDR). Una reparación imprecisa mediada por NHEJ puede producir mutaciones de inserción y/o deleción de longitud variable en el sitio de la ruptura de doble hebra <sup>19</sup>.

### 3.1. nCATS

Recientemente se introdujo una técnica que utiliza la capacidad de la Cas9 para introducir cortes en ubicaciones específicas y poder luego ligar adaptadores de secuenciación de nanoporos directamente a esos sitios (nCATS) (Figura 6). El potencial de esta técnica, junto con la baja inversión del minION podría ser un gran avance en la investigación. Cuando Cas9 se une con el crARN idéntico a la secuencia *target* deseada, junto con el ARN CRISPR transactivador (tracrARN - un ARN estructural requerido para la actividad catalítica) forma un complejo de ribonucleoproteína (RNP). Este complejo de ARN y enzima busca en el ADN genómico la región *target* utilizando la secuencia de crARN. Si el crARN coincide y se empareja con la secuencia *target*, Cas9 corta ambas hebras de la secuencia *target*. Cas9 requiere un crARN de 20pb (el protospacer), más el motivo adyacente al protospacer (PAM - secuencia NGG). El PAM está inmediatamente al lado del extremo 3' del *target*. Cas9 corta ambas hebras de la secuencia *target* 3 pb aguas arriba del PAM. La producción de un extremo recientemente expuesto, lo que llamamos "desprotección", en un sitio definido es la base para la selectividad del método de enriquecimiento mediado por Cas. Para el enriquecimiento (sin PCR) mediado por Cas, se puede proteger y desproteger selectivamente los extremos utilizando CRISPR-Cas9 y enriquecer las regiones de interés en un segundo plano antes de la secuenciación de la molécula de ADN nativa <sup>20</sup>.



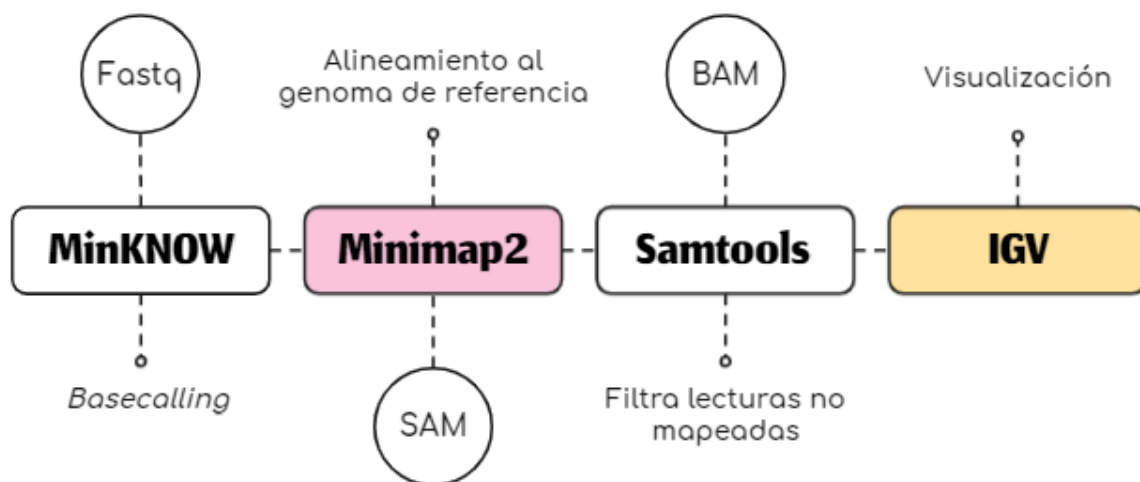
**Figura 6.** Esquema del enriquecimiento con Cas9 para secuenciación con MinION. ROI = región de interés. Los extremos del ADN se desfosforilan primero, se introducen nuevos cortes con el complejo Cas9/ARNguía, los adaptadores de secuenciación de nanoporos se ligan a cortes alrededor del ROI y la muestra se carga en el MinION <sup>20</sup>.



### 3.1.1. Análisis Bioinformático

El enriquecimiento libre de PCR mediado por Cas9 tiene como objetivo maximizar la recuperación de la secuencia en el objetivo al reducir tanto el fondo de ADN como las cantidades de ADN fuera del objetivo. Cuando funciona de manera óptima, el nCATS proporciona hasta un par de gigabases de datos de una celda de flujo MinION en 48 horas de secuenciación y brinda una gran cobertura en todas las regiones *target* <sup>21</sup>

Esta metodología utiliza el *software* Snakemake <sup>22</sup>, un sistema de gestión de flujo de trabajo implementado en Python. El objetivo de la herramienta Snakemake es permitir análisis de datos reproducibles y escalables. El flujo de trabajo de Snakemake define además los conjuntos de *software* requeridos (y las versiones de software cuando corresponda) y automatizará la instalación e implementación de este software. En este caso, el flujo de trabajo de Snakemake llamará a métodos que incluyen Minimap2 <sup>23</sup> y Samtools <sup>24</sup> (Figura 7).



**Figura 7.** Pipeline bioinformático. MiniKnow realiza el *basecalling* generando un fastq, Minimap2 alinea las lecturas al genoma de referencia y exporta un archivo SAM, Samtools filtra las lecturas mapeadas y comprime a un BAM e IGV visualiza este archivo BAM.

El primer paso para el análisis de la estrategia de enriquecimiento de Cas9 es evaluar la distribución y la cobertura regional de las lecturas de secuencia en todo el genoma. Las secuencias fastq producidas durante la secuenciación del ADN se mapean en el genoma de referencia utilizando Minimap2 <sup>23</sup>. Los resultados del análisis de mapeo son tomados por Samtools <sup>24</sup> para filtrar las secuencias no mapeadas, convertir la salida en formato SAM de Minimap2 sin comprimir al formato BAM comprimido y ordenar las secuencias en el archivo BAM por sus coordenadas de mapeo. La indexación adicional del archivo BAM (nuevamente, usando Samtools) permite un acceso eficiente a las entradas BAM que corresponden a ubicaciones genómicas específicas <sup>21</sup>.

Las lecturas que brinda el enriquecimiento con Cas9 con la secuenciación con MinION se pueden clasificar en cuatro grupos diferentes que se excluyen mutuamente:

- *On target*: Lecturas que se alinean con las regiones de interés proporcionadas en el archivo de coordenadas de formato BED.
- *Off target*: Cortes de Cas9 en el ADN fuera de la región de interés. Estas regiones se clasifican como fuera del objetivo si la profundidad de cobertura es > 20 veces superior al nivel de fondo medio.
- *Background*: Lecturas que se alinean con el genoma de referencia, pero que no se incluyen en ninguna de las categorías anteriores.

Dada la cantidad de información obtenida se utiliza un entorno gráfico para facilitar la interpretación, la herramienta Integrative Genomics Viewer (IGV) que proporciona funcionalidad para la visualización de datos de mapeo de secuencias de archivos BAM que posteriormente se pueden superponer con "pistas" de información que pueden incluir profundidad de cobertura para datos de mapeo y anotaciones de genes <sup>25</sup>.

#### **4. Estudios de Melanoma cutáneo maligno en el Laboratorio de Epidemiología Genética**

El cáncer es la segunda causa de muerte más frecuente en todo el mundo, matando a más de 8 millones de personas cada año <sup>26</sup>. El melanoma cutáneo maligno (MM) es un cáncer de piel potencialmente mortal. Aunque el melanoma es menos común que otros tipos de cáncer de piel, es el más letal y representa casi el 73% de muertes <sup>27,28</sup>.

El melanoma se desarrolla a partir de melanocitos, pero también puede comenzar en otros tejidos pigmentados, como los ganglios linfáticos, los ojos o los intestinos <sup>27</sup>. La identificación de las personas con alto riesgo sería de gran interés para mejorar el diagnóstico temprano y, en última instancia, la supervivencia. Los factores de riesgo conocidos de MM incluyen sensibilidad al sol, exposición al sol, color de ojos y cabello claro, alto número de nevos melanocíticos, nevos atípicos y antecedentes familiares de melanoma. El conocimiento de los factores de riesgo de MM es la base para el desarrollo de herramientas de predicción de riesgos que pueden mejorar la comprensión y la toma de decisiones, lo que lleva a un cambio de comportamiento favorable y a la prevención de enfermedades. Dado que el melanoma parece estar determinado por interacciones complejas entre las características del huésped, la exposición ambiental y los factores genéticos, la inclusión y evaluación de marcadores genéticos en modelos de riesgo puede estar justificado y se ha considerado un paso importante para un mayor desarrollo y prueba de herramientas de predicción antes de que se puedan utilizar de forma rutinaria con confianza <sup>29</sup>.

Los melanomas esporádicos representan aproximadamente el 90% de todos los casos de melanoma en los que se involucran mutaciones somáticas, expresadas por alelos con una prevalencia alta y penetrancia baja, lo que nos orienta a reconocer los factores ambientales como actor importante en la transformación maligna de los melanocitos <sup>28</sup>. Se cree que los melanomas esporádicos surgen de mutaciones adquiridas al azar en los melanocitos. Estas mutaciones pueden deberse al medio ambiente, el envejecimiento, el daño solar crónico (exposición a los rayos UV), eventos fortuitos u otros eventos/factores no heredables.

Diferenciar la agrupación familiar de casos esporádicos de melanoma hereditario a menudo se confunde con factores de riesgo ambientales, y los rasgos con asociaciones débiles con el melanoma, como la exposición al sol intensa y prolongada en el tipo de piel susceptible, también influyen en la contribución relativa de los factores genéticos para el riesgo de melanoma. Además, la herencia conjunta de genes modificadores puede inhibir o mejorar la penetrancia aparente de una mutación de alto riesgo <sup>30</sup>.

En el laboratorio de Epidemiología Genética del Departamento de Genética de la Facultad de Medicina se analizaron los factores de riesgo, la contribución de las poblaciones parentales y los haplotipos del gen *MC1R* y su asociación con MM esporádicos en la población uruguaya. Este estudio es el primero en analizar los factores de riesgo y también las variantes de RHC y la estructura de haplotipos del gen *MC1R* en una población mestizada de América Latina y sus asociaciones con el melanoma esporádico. La presencia del fototipo I y II, la exposición solar en la infancia durante las vacaciones y los nevos atípicos son los principales factores de riesgo fenotípicos de MM en la población uruguaya <sup>27</sup>.

Actualmente el laboratorio decide avanzar en el estudio de melanoma aplicando técnicas de secuenciación con MinION con el fin de analizar grandes regiones que involucran el gen *MC1R* y regiones flanqueantes que incluyen los genes *TCF25* y *TUBB3*, que permitirán estudiar haplotipos enteros, variantes génicas, variantes estructurales y metilación del ADN, teniendo un panorama más amplio del estado genómico de la región cromosómica 16q24.

## **4.1. Genes de la región del *MC1R***

### **4.1.1. *MC1R***

A medida que avanza rápidamente el conocimiento de las asociaciones entre variantes genéticas comunes y un riesgo elevado de enfermedad, las expectativas son altas para la implementación de pruebas genéticas en la práctica clínica habitual. Para cánceres prevenibles como el cáncer de piel, la identificación de variantes genéticas comunes puede mejorar la precisión de los modelos de predicción de riesgos e informar estrategias de prevención y detección temprana estratificadas por riesgo para reducir potencialmente la mortalidad, la morbilidad y los costos de atención médica <sup>31</sup>.

El receptor melanocortina-1 (*MC1R*) se considera un gen de susceptibilidad con baja penetrancia. Este gen es un regulador primario de la pigmentación de la piel y codifica un receptor acoplado a una proteína G que se expresa en melanocitos. *MC1R* está ubicado en la región cromosómica 16q24, con un marco de lectura abierto de 951 pb (MIM # 155555). Numerosos estudios han demostrado que la variación genética en el *MC1R*, se asocia con un mayor riesgo de melanoma. Entre las personas con fenotipos “resistentes al sol” (buena respuesta al bronceado, cabello oscuro, pocas quemaduras), las asociaciones entre las variantes de *MC1R* y el cáncer de piel son a menudo más fuertes que entre las personas con fenotipos “sensibles al sol” (Ej. cabello rubio, piel quemada) <sup>27,31</sup>.

El *MC1R* es el gen más importante que desempeña un papel en la predisposición a la MM esporádica, y su asociación con la MM se ha replicado y confirmado mediante metanálisis y estudios de asociación de todo el genoma. Es altamente polimórfico en poblaciones de origen europeo, con más de 200 variantes de regiones codificantes descritas hasta la fecha y una prevalencia de cualquier variante de *MC1R* de ~ 60% en controles sanos. de estas variantes se ha demostrado que reducen la función del receptor, 24-26 dan como resultado un cambio cuantitativo de la síntesis de melanina de eumelanina a feomelanina, y determinan el fenotipo de pelo rojo, caracterizado por la co-ocurrencia de piel clara, pecas y sensibilidad a la radiación ultravioleta (mala respuesta al bronceado) <sup>29</sup>.

#### **4.1.2. *TUBB3* y *TCF25***

Numerosos estudios han sugerido que la expresión diferencial de varios genes, como la  $\beta$ -tubulina III (*TUBB3*), en los tejidos tumorales están estrechamente asociados con la quimiorresistencia y el pronóstico en pacientes con cáncer. *TUBB3* es un marcador de resistencia a taxanos, y los niveles altos de expresión de *TUBB3* se correlacionan con tasas de respuesta bajas en pacientes tratados con regímenes que contienen taxanos <sup>32,33</sup>. *TUBB3* está ubicado en la región cromosómica 16q24, con un marco de lectura abierto de 14 kb (MIM # 602661).

*TCF25* está ubicado en la región cromosómica 16q24, con un marco de lectura abierto de 37.775 pb (MIM # 612326). Este gen se encuentra en la región aguas arriba del gen *MC1R* y se incluye en este trabajo por caer parte de su secuencia en la región de interés.

# Objetivos

---

## 1. Objetivo general

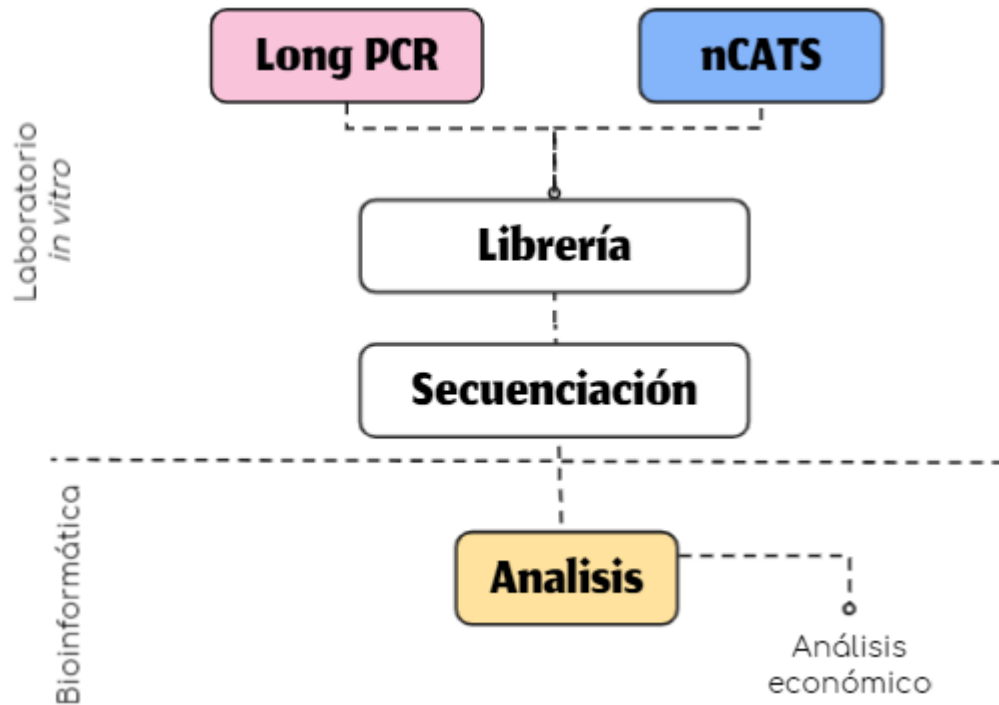
Desarrollo de un *pipeline* de trabajo para secuenciación dirigida de fragmentos de más de 10 kb en la plataforma MinION.

## 2. Objetivos Específicos

- Amplificación y secuenciación de fragmentos de 6 kb para la región que comprende los genes *TCF25*, *MC1R* y *TUBB3*.
- Desarrollar un método de evaluación de eficiencia de cortes de Cas9 en la región de interés.
- Lograr una buena eficiencia de la secuenciación dirigida por CRISPR-Cas9 de la región de los genes *TCF25*, *MC1R* y *TUBB3*.
- Comparar los resultados obtenidos en la secuenciación con ambos abordajes.
- Desarrollo en el laboratorio de un pipeline de trabajo para aplicar la secuenciación dirigida por CRISPR-Cas9 a otras regiones genómicas.
- Evaluación económica de la metodología nCATS.

## Diseño Experimental

Este trabajo comprende la puesta a punto en el laboratorio de la secuenciación dirigida de grandes fragmentos genómicos, utilizando como ejemplo la secuenciación de la región que contiene los genes *MC1R*, *TCF25* y *TUBB3*. Se abordaron dos técnicas para llegar a la secuenciación: por un lado, se realizaron ampliificaciones de secuencia larga por Long PCR, y por otro lado se realizó secuenciación dirigida por CRISPR-Cas9 (nCATS). La figura 8 muestra los dos abordajes trabajados en esta tesis.



**Figura 8.** Diseño experimental de la tesis.

## Materiales y métodos

---

### 1. Extracción de ADN

La puesta a punto de la secuenciación dirigida, utilizando ambos abordajes, fue realizada utilizando muestra de ADN de 2 individuos controles proporcionados por del Laboratorio de Epidemiología Genética, Departamento de Genética, de la Facultad de Medicina de la Universidad de la República. La extracción de ADN se realizó a partir de leucocitos de sangre periférica. Para ello, se mezcló 1 volumen de sangre con 2 volúmenes de buffer de lisis celular para glóbulos rojos, se dejó en hielo 10 minutos, se centrifugó a 2500 rpm por 10 minutos y al pellet se lo lavó con PBS. Luego se le agregó 500 µl de solución de lisis celular para glóbulos blancos y se dejó a 55°C *overnight*. A este lisado se le agregó 300 µl de solución de precipitación de proteínas y se centrifugó a 6000 rpm por 20 minutos. Se transfirió el sobrenadante a un tubo con 700 µl de Isopropanol y se volvió a centrifugar a 6000 rpm por 20 minutos. Posteriormente se lavó el pellet con Etanol 70% y se centrifugó a 6000 rpm por 10 minutos. Se dejó secar el pellet por inversión *overnight* y se resuspendió el ADN en 100 µl de agua estéril miliQ.

Con el objetivo de evitar la fragmentación del ADN, los pasos de centrifugación durante la extracción se realizaron a bajas rpm durante tiempos largos.

Se midió la concentración de ADN utilizando el Nanodrop 900 (Thermo Fisher Scientific), siguiendo el protocolo del proveedor. Solo se utilizaron los ADN cuyo valor de relación de absorbancia 260/280 fuera mayor a 1.80, y de 260/230 mayor a 1.2 (indicadores de pureza).

### 2. Puesta a punto de PCR tiempo final para amplificar los genes *TCF25*, *MC1R* y *TUBB3*

#### 2.1. Diseño de cebadores para Long PCR

Se realizó el diseño de cebadores para los tres genes *TCF25*, *MC1R* y *TUBB3* (Tabla 1).

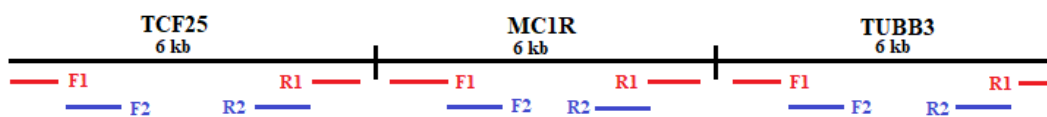
Para esto se buscó la secuencia de cada uno de los genes en el UCSC Genome Browser (<http://www.genome.ucsc.edu/>) y se diseñaron los cebadores utilizando el software Primer3 (<https://primer3.ut.ee/>). Se planteó una estrategia de PCR anidada (2 rondas) para amplificar fragmentos de 6kb que cubrieran cada gen, solapados entre sí por 1 kb. Esta estrategia permite obtener mayor amplificación y especificidad, ya que los cebadores para la ronda 2 de la PCR son más internos que los de la ronda 1 (Figura 9).

**Tabla 1.** Cebadores (IDT) correspondiente a la región de los tres genes (*TCF25*, *MC1R* y *TUBB3*). F corresponde a los cebadores *Forward* y R correspondiente a *Reverse*.

Nombre	Secuencia (5' - 3')
TCF25_F1	TTACAGCCCCTGTCTCCTCC
TCF25_R1	TCCTTTGAACTTGGGGGTGG
TCF25_F2	TTTCTGTTGGTGCTGATATTGCCACCTCCCCTGCCACCTC
TCF25_R2	ACTTGCCTGTCGCTCTATCTTCGGGTCTGGCTCAGTGGAAA
MC1R_F1	ATGACAAGAGTTGGCCCGAG
MC1R_R1	CACAGTCTGTCCTGGTCACC
MC1R_F2	TTTCTGTTGGTGCTGATATTGCCAGGAAGAGATCATGGGGGC
MC1R_R2	ACTTGCCTGTCGCTCTATCTTCGAGCGTGAAAGCGTCCAATG
TUBB3_F1	TGGCTTGTGAGTGGTGTGAG
TUBB3_R1	TTTAGAGACCAACCCGCACC
TUBB3_F2	TTTCTGTTGGTGCTGATATTGCACTGAAAACCCTGATGCGGT
TUBB3_R2	ACTTGCCTGTCGCTCTATCTTCGTTTTGGGTGGGGGAGGTAG

Para el diseño de los cebadores se siguieron los siguientes criterios:

- Largo de los cebadores entre 20 - 25 pares de bases
- La secuencia de los cebadores debe tener una composición de bases tal que su contenido GC sea entre 40% y 60% y ambos cebadores tengan un %GC similar.
- Evitar regiones de la secuencia molde donde contengan SNPs o regiones repetidas.
- La temperatura de melting ( $T_m$ ) no debe variar más de 5°C entre los cebadores y tratar de que se encuentre entre los 55°C - 60°C.
- Intentar que el extremo 3' contengan una G o C.
- Que no sean palindrómicos para evitar que formen una horquilla (plegarse sobre sí mismo por complementariedad).



**Figura 9.** Amplicones de los genes *TCF25*, *MC1R* y *TUBB3* de 6 kb cada uno de tamaño. En color rojo los cebadores para la ronda 1 y en azul los cebadores para la ronda 2 de la *long PCR*.



Finalmente, para evitar la amplificación de secuencias inespecíficas, se chequearon los cebadores en bases de datos de secuencias de ADN (GeneBank, <https://www.ncbi.nlm.nih.gov/genbank/>) la posibilidad de amplificación de otros blancos.

Los cebadores de la reamplificación (Ronda 2) requieren unas colas para poder ser utilizados en la secuenciación con MinION que tienen las siguientes secuencias:

5' TTTCTGTTGGTGCTGATATTGC-[*forward primer sequence*] 3'

5' ACTTGCCTGTCGCTCTATCTTC-[*reverse primer sequence*] 3'

## 2.2. Long PCR

Se realizaron dos tipos de *Long PCR*, con dos enzimas diferentes.

### 2.2.1. Prueba Long PCR con Mix Taq Polimerasa:

La reacción de PCR de la primera ronda se llevó a cabo como se describe a continuación: 0,9 µl de 11x Buffer de reacción, 0.12 µl de 1M Tris base, 0.3 µl de un mix 10 µM de ambos cebadores, 0.06 µl de Taq polimerasa mix (relación 20:1 de High Taq polimerasa 5U/µl de Bioline y Velocity Taq polimerasa 2U/µl de Bioline) y agua miliQ para llegar a un volumen total de 10 µl. Se utilizó 1 µl de 100 ng/µl de ADN como molde para la primera reacción de amplificación.

Las condiciones de ciclado de la primera ronda de PCR con los cebadores externos son las siguientes: 96°C por 20 segundos, desnaturalización; 55°C a 63°C por 30 segundos, *annealing*; 72°C por 10 minutos y 20 segundos, extensión, durante 35 ciclos; y 5 min de desnaturalización inicial y 1 min de extensión final.

La reacción de PCR de la segunda ronda se llevó a cabo como se describe a continuación: 1.8 µl de 11x Buffer de reacción, 0.25 µl de 1M Tris base, 0.8 µl de un mix 10 µM de ambos cebadores, 0.12 µl de Taq polimerasa mix y la correspondiente agua miliQ para llegar a un volumen total de 20 µl. Se utilizó 0.5 µl del producto de PCR de la primera ronda como molde para la segunda reacción de amplificación

En la segunda ronda (*nested*) se mantienen las mismas condiciones de ciclado a excepción de que se hacen 32 ciclos. Ambas amplificaciones fueron realizadas en un termociclador Applied Biosystems de Thermo Fisher Scientific.

### 2.2.2. Prueba long PCR con Ranger DNA polimerasa:

Se realizó la amplificación de la región de los tres genes mediante una PCR con la enzima Ranger DNA Polymerase (Bioline, LDN, UK). La reacción de PCR se llevó a cabo como se describe a continuación: 25 µl de 2x Ranger Mix (Bioline, LDN, UK), 2 µl de un mix 10 µM

de ambos cebadores y la correspondiente agua miliQ para llegar a un volumen total de 50  $\mu$ l. Se utilizó 2  $\mu$ l de 100 ng/ $\mu$ l de ADN como molde para la reacción de amplificación.

Las condiciones de amplificación fueron: 3 minutos 95°C, 35 ciclos de 10 segundos a 98°C y 9 minutos a 62°C. La amplificación fue realizada en un termociclador Applied Biosystems de Thermo Fisher Scientific.

### 2.3. Visualización de productos de PCR en gel de Agarosa

Se realizaron electroforesis en gel de Agarosa para la separación y visualización de los productos de las distintas PCR.

Se prepararon geles de agarosa 1% en buffer TBE 0.5x. Los geles fueron suplementados con el agente intercalante GoodView™ (SBS Genetech Co. Ltd.). Las muestras de ADN fueron cargadas en los geles con buffer de carga Gel Loading Dye (NEB®). Se utilizó el marcador de peso molecular 1kb DNA Ladder Mix (NEB®, #S3232S) o el marcador de peso molecular Fast Ruler DNA LADDER, High range 1kb (Fermentas Life Sciences, #SM1128). Utilizando buffer de electroforesis TBE 0.5x, las muestras se corrieron en la cuba electroforética *Fast Mini Horizontal gel unit H1-SET* (SCIE-PLAS) a 100V por aproximadamente 1 hora. La visualización del gel se realizó utilizando el transiluminador (Bio DOC-IT).

## 3. Secuenciación dirigida por Cas9

### 3.1. Diseño de ARN guías

Se realizó el diseño de los crARN para los genes *TCF25*, *MC1R* y *TUBB3*, utilizando la herramienta Alt-R HDR *Design Tool* de Integrate DNA Technologies (IDT). Este es un programa que asiste en el diseño de estrategias de CRISPR-Cas9 evaluando posibles crARN que pueden dirigir un corte en el ADN en los tres genes mencionados anteriormente.

Para esto se buscó la secuencia de cada uno de los genes en el UCSC Genome Browser y se la ingresó en la herramienta de IDT. Según la secuencia proporcionada a la herramienta, junto con la información del PAM, ésta encuentra PAMs sobre la secuencia de interés y predice las posibles secuencias de ARN guías, desde la que tiene mayor especificidad a la que menos tiene. Además, predice su eficiencia y la cantidad de *off target* en base a los *mismatches* presentes en ellos.

Se verificó que los crARN no estuvieran en zonas de la secuencia molde con SNPs y zonas de repetidos. Los crARN se diseñaron según los siguientes criterios:

- Tiene que contener la secuencia PAM 5'-NGG-3'
- Un tamaño de 20 nucleótidos
- Valores de *Off target* cercanos al 100%

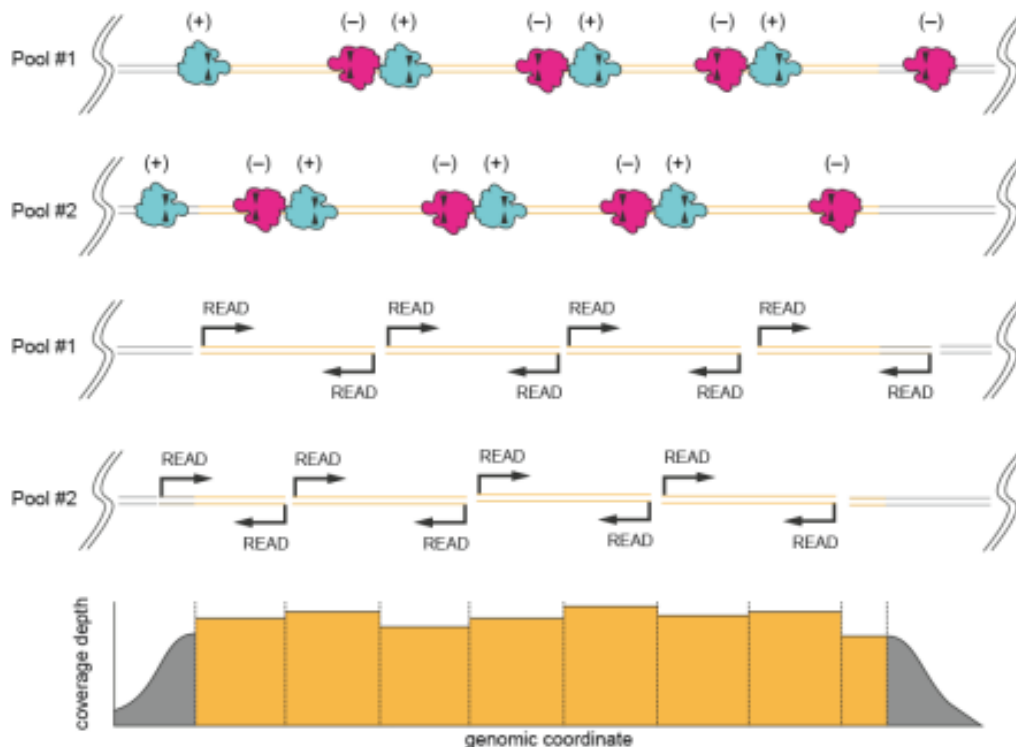
Los valores *off target* con una puntuación del 100% predice una baja actividad fuera del *target*. En una primera instancia se diseñaron tres crARN que se encuentran en la Tabla 2.

**Tabla 2.** Secuencias para los tres crARN para cortar los genes *TCF25*, *MC1R* y *TUBB3*.

Posición	Hebra	Secuencia (5' - 3')	PAM	<i>ON target score</i>	<i>OFF target score</i>
89885895	+	AGCACTGGGTTGAACCGTCC	CGG	58	81
89919670	-	GATGTAGCGGTCCACGGCGA	TGG	52	92
89934753	+	GGCCAAGGGTCACTACACGG	AGG	83	81

Método *tiling* (Figura 10):

Este método implica diseñar dos grupos de varios crARN que deben prepararse como dos experimentos separados y agruparse en el paso final (antes de la purificación final) y cargarse juntos en la misma celda de flujo. Para este caso se realizó un enfoque de "albañil", donde cada grupo contiene crARN (+) y (-) que están diseñadas para superponerse entre sí.



**Figura 10.** Diseño de los crARN para un enfoque *tiling*. Se diseñaron dos grupos de sondas que cubren una gran región de interés (> 20 kb). Esto divide el ROI (región de interés) en secciones superpuestas, asegurando una buena cobertura de toda la longitud del ROI. Imagen tomada de Oxford Nanopore.

En una segunda instancia se diseñaron 18 crARN de aproximadamente 10 kb entre corte y corte divididos en dos pools distintos (Tabla 3 y 4). Las sondas de crARN fueron diseñadas a lo

largo de la región de interés en fragmentos superpuestos de 5-10 kb que permiten una cobertura uniforme en toda la región de interés.

**Tabla 3.** Secuencias de crARN del *pool 1* del *tiling*. En color azul las secuencias que se utilizaron para las primeras pruebas.

<b>POOL 1</b>							
Hebra	Secuencia (5' - 3')	PAM	ON target score	OFF target score	Largo (kb)	Distancia de - a +	GEN
+	AGCACTGGGTTGAACCGTCC	CGG	58	81	9.1	37 pb	TCF25
-	AGGAGTGAGTCAACGTGGTA	AGG	83	73			TCF25
+	GAGATGGCTCGAGACCTCGT	AGG	51	87	11.1	328 pb	TCF25
-	GGATATTCCTGGGTGCACGC	TGG	51	80			TCF25
+	CACAAGCTGAGTCGTCACCG	TGG	78	82	9.5	1.9 Kb *	TCF25
-	GGCGTGAACAGACCACGTCT	GGG	73	84			MC1R
+	CTGCGGAGCACCACGTGCGA	CGG	62	93			MC1R
-	TAACGCGGACAGCATGGACA	TGG	52	76	11.4		MC1R

\*Contiene al *MC1R*.

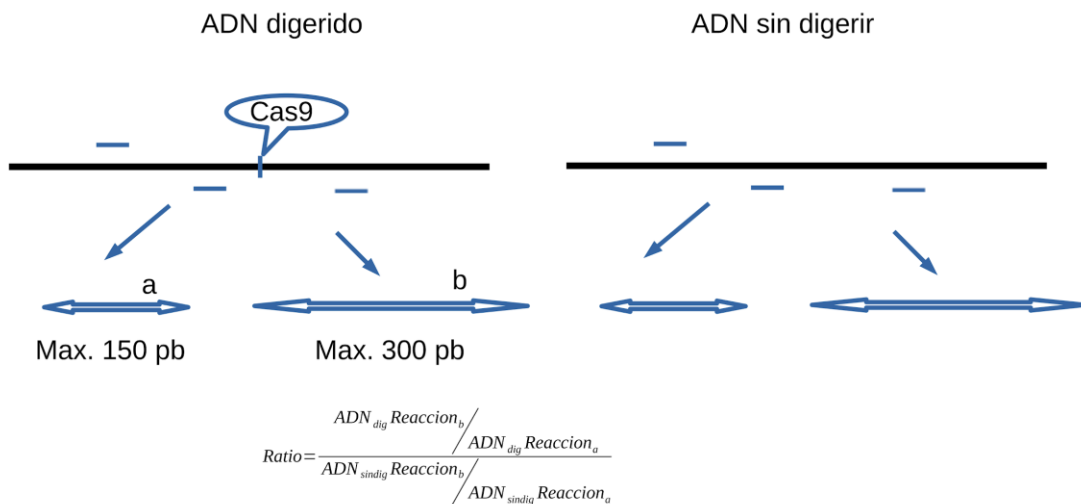
**Tabla 4.** Secuencias de crARN del *pool 2* del *tiling*.

<b>POOL 2</b>							
Hebra	Secuencia (5' - 3')	PAM	ON target score	OFF target score	Largo (kb)	Distancia de - a +	GEN
+	TGAACCTTAGTGCATCCGAT	TGG	56	85	9	32 pb	TCF25
-	CGTAACTCTACTGAAGGCAC	TGG	78	71			TCF25
+	ACCTTCTTACTCGTAGTCGA	CGG	62	94	10	790 pb	TCF25
-	GATAACTTAAAACGTCCACC	TGG	59	82			TCF25
+	AGCTTCGATCCACGCACATG	TGG	56	87	10.3	710 pb	TCF25
-	TTCAAGGAGGACGTGCACCC	AGG	87	74			TCF25
+	TCGGACCAAGACCCAGTCT	TGG	80	77	10.5	650 pb	MC1R
-	CAAAGCGTCGGATCAGAGCT	CGG	59	83			MC1R
+	ACCAGATCGGGGCAAGGTG	AGG	58	56			TUBB3
-	AGGCACGTACTIONTGTGAGCTG	AGG	71	70	10		TUBB3

### 3.2. Evaluación del corte con Cas9

Para analizar la presencia y eficiencia de corte en una reacción de Cas9 usamos un método basado en la cuantificación de la expresión génica por PCR en tiempo real. De esta manera se evita ingresar en la etapa de la construcción de la biblioteca sin una evaluación previa.

El diseño se describe en la figura 11 y se basa en la elaboración de tres cebadores, un cebador *Forward* y dos *Reverse*. El cebador *Reverse* se encuentra a aproximadamente 150pb de del *Forward*, antes del corte del crARN. El segundo *Reverse* se encuentra a 300 pb aproximadamente del cebador *Forward* y dicho amplicón contiene dentro el corte del crARN.



**Figura 11.** Cebadores que amplifican 300 pb con el corte del CRISPR-Cas9 interior y cebadores que amplifican 150 pb antes del corte.

Utilizando la lógica de la cuantificación relativa en base a una referencia podemos estimar cuánto ha sido la eficiencia del corte (*Pffafl method*).

$$RC = \frac{2^{\Delta Ct \text{ dig}}}{2^{\Delta Ct \text{ s/dig}}}$$

$$\begin{aligned} \Delta Ct \text{ dig} &= Ct \text{ frag largo dig} - Ct \text{ frag corto dig} \\ \Delta Ct \text{ s/dig} &= Ct \text{ frag largo s/dig} - Ct \text{ frag corto s/dig} \end{aligned}$$

$$RC = 2^{(\Delta Ct \text{ dig} - \Delta Ct \text{ s/dig})}$$

El RC que se tomará como valor mínimo aceptado de una buena eficiencia de corte es  $RC = 5$ . Lo que significa que hay 5 veces más cantidad de ADN cortado por CAS9 que ADN sin cortar en la región evaluada.

### 3.2.1. Puesta a punto de PCR en tiempo real

#### 3.2.1.1. Diseño de cebadores

Para evaluar la eficiencia de los cortes con CRISPR-Cas9 se diseñaron tres cebadores (IDT). Para esto se tomó como corte a evaluar el del ARN guía (5'CTGCGGAGCACCACGTGCGA3'). A partir de esta secuencia se eligieron cebadores que amplificaran aproximadamente 150 pb antes del corte (F1R2) y 300 pb (F1R1) con el corte incluido en el medio como muestra la Figura 11. Los cebadores se diseñaron con el Primer3 (<https://primer3.ut.ee/>) y se chequearon en GeneBank (<https://www.ncbi.nlm.nih.gov/genbank/>) para evitar amplificación en secuencias inespecíficas. Las secuencias de los tres cebadores se observan en la Tabla 5.

**Tabla 5.** Cebadores para evaluar los cortes de ADN con el sistema CRISPR-Cas9.

Nombre	Dirección	Secuencia (5' - 3')	Tamaño	Tm	%GC	Largo (pb)
F1	<i>Forward</i>	CTTTCCTGCAGCTGTAAATGG	21	59.9	47.6	
R1	<i>Reverse</i>	GCATCTTCCTGGGTTCACTC	20	59.5	55	121
R2	<i>Reverse</i>	TGCACCAACAGCCACATC	18	60.3	55.6	260

#### 3.2.1.2. PCR en tiempo real

Luego del diseño de los cebadores se hizo una puesta a punto de la amplificación por PCR en tiempo real.

Esta PCR utiliza el mix de *2x Sensi Fast Mix* (Bioline, LDN, UK). La reacción de PCR en tiempo real se llevó a cabo como se describe a continuación: 10 µl de *2x Sensi FAST Mix* (Bioline, LDN, UK), 0.8 µl del cebador *Forward* 10 µM, 0.8 µl d del cebador *Reverse R1 o R2* 10 µM y agua miliQ para llegar a un volumen total de 20 µl.

Se realizaron pruebas con concentraciones de 50ng/µl y de 100 ng/µl de ADN molde, y se probaron distintas condiciones de ciclado y temperaturas de *annealing*.

Las condiciones del ciclado para la amplificación del F1R1 fueron: 2 minutos 95°C, 40 ciclos de 15 segundos a 95°C y 30 segundos a 57°C - 60°C. Las condiciones del otro ciclado de amplificación para el F1R2 fueron: 3 minutos 95°C, 40 ciclos de 5 segundos a 95°C, 10 segundos a 57°C - 61°C, 20 segundos a 72°C.

Las amplificaciones fueron realizadas en un termociclador Eco Real-Time PCR System de Illumina.

### 3.2.1.3. PCR tiempo final

Con el fin de chequear los cebadores diseñados para la PCR en tiempo real, se probaron con la PCR en tiempo final. La reacción de PCR se llevó a cabo como se describe a continuación: 2.5 µl de 10x Buffer de reacción, 0.6 µl de 10mM de dNTPs, 0.75 µl de 100 µM de cada uno de los cebadores, 0.2 µl de Taq Polimerasa 5U/µl (TIANGEN) y agua miliQ para llegar a un volumen total de 26 µl. Se utilizó 2 µl de 100 ng/µl de ADN como molde para la reacción de amplificación.

Las condiciones de amplificación fueron: 5 minutos 95°C, 30 ciclos de 30 segundos a 95°C, 30 segundos a 57°C y 60°C y 40 segundos a 72°C, 10 minutos a 72°C y se mantiene a 15°C. La amplificación fue realizada en un termociclador Applied Biosystems de Thermo Fisher Scientific.

Los productos se visualizaron por electroforesis en gel de Agarosa al 2% en buffer TBE 0,5x. Los geles fueron suplementados con el agente intercalante GoodView™ (SBS Genetech Co. Ltd.). Las muestras de ADN fueron cargadas en los geles con buffer de carga Gel Loading Dye (NEB®). Se utilizó el marcador de peso molecular 100 pb DNA Ladder no stain (Bioron, #304005). Utilizando buffer de electroforesis TBE 0.5x, las muestras se corrieron en la cuba electroforética *Fast Mini Horizontal gel unit HI-SET* (SCIE-PLAS) a 100V por aproximadamente 30 minutos. La visualización del gel se realizó utilizando el transiluminador (Bio DOC-IT).

Con el fin de verificar mediante un método alternativo que lo que se está amplificando es realmente lo que se quería, se secuenciaron los productos de PCR en Macrogen (Corea del Sur).

## 3.3. Secuenciación dirigida por Cas9

### 3.3.1. Obtención de fragmentos de la región de interés

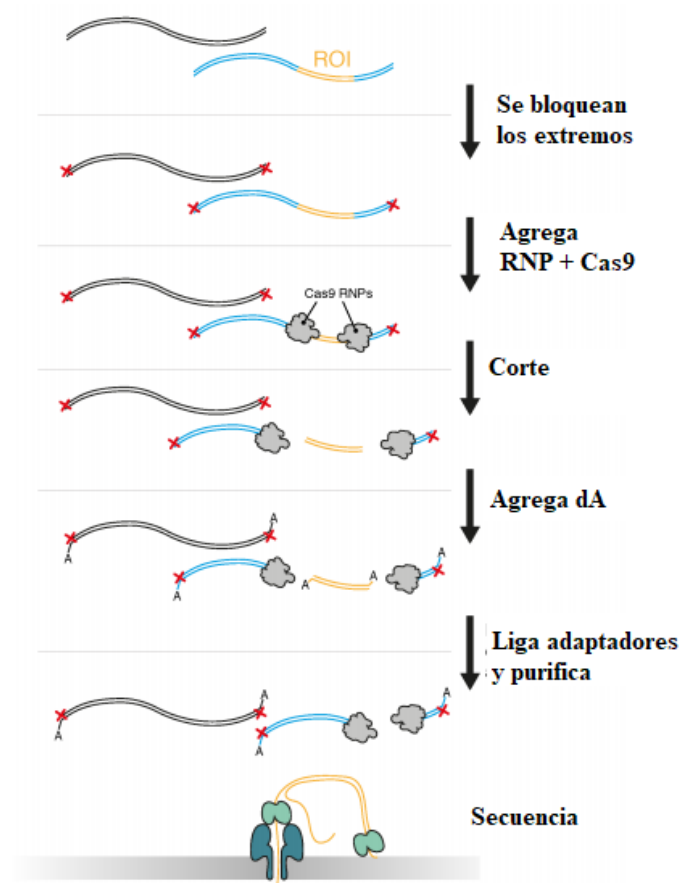
Se utilizó y modificó el protocolo de Gilpatrick et al., 2020 (Figura 12) <sup>20</sup>.

En un primer paso, se resuspendieron los crARN (IDT) y el tracrARN (IDT) a una concentración final de 100 µM en buffer TE pH 7.5 (2 nmol crARN en 20 µl; 5 nmol de tracrARN en 50 µl). Luego se realizó un mix equimolar de los tres crARN agregando 0.75 µl de cada uno. Y se ensambló el dúplex de gARN (8 µl de agua libre de nucleasas, 1µl de 100 µM de tracrARN y 1 µl de crARN mix). Este ensamblado se calentó a 95°C por 5 minutos.

Por otra parte, se diluyó el 10x Cut Smart Buffer (NEB) 1:8 en agua libre de nucleasas (NFW) para hacer 1.25x Cut Smart buffer y la Hifi Cas9 (IDT) 1:5 utilizando este 1.25x de Cut Smart buffer.

Siguiendo con el protocolo, se ensambló el complejo RNP, agregando 23.7  $\mu\text{l}$  de NFW, 2.9  $\mu\text{l}$  de 1.25x Cut Smart buffer, 2.4  $\mu\text{l}$  del gARN dúplex y 1  $\mu\text{l}$  de la dilución 1:5 de Hifi Cas9. Este complejo RNP se incubó a temperatura ambiente por 20 minutos y luego de la incubación se mantuvo en hielo.

El siguiente paso fue desfosforilar los extremos libres del ADN y para esto fue necesario tener una concentración suficiente de ADN de 3  $\mu\text{g}$  en menos de 33  $\mu\text{l}$ . En este caso se realizó para los dos ADN extraídos previamente. La reacción de desfosforilación contiene 4  $\mu\text{l}$  de buffer AP 10x (Thermo Scientific), 3  $\mu\text{g}$  de ADN, 3  $\mu\text{l}$  de Fosfatasa AP Ultrafast (Thermo Scientific) y agua ultra pura hasta llegar a un volumen de 40  $\mu\text{l}$ . Esto se incubó a 37°C por 30 minutos para el paso de desfosforilación y se inactivó la enzima a 72°C por 5 minutos, luego se mantuvo a 4°C.



**Figura 12.** Pasos para la secuenciación por la técnica nCATs. Imagen tomada y modificada de Gilpatrick et al., 2020.

Luego de tener el ADN desfosforilado y el complejo RNP que contiene el ARN guía con la Cas9, se procedió a realizar los cortes y el agregado de Adenina a los extremos. Para esto se diluyó dATP (Bioron) 100 mM en 1:10. Se colocó el ADN desfosforilado con 10  $\mu\text{l}$  del complejo RNP, 1  $\mu\text{l}$  1:10 dATP y 1  $\mu\text{l}$  de la Taq Polimerasa (TIANGEN). Se incubó a 37°C por 30 minutos para que la Cas9 corte el ADN y se inactivó la enzima a 72°C por 5 minutos. Luego se mantuvo a 4°C. Finalmente, la eficiencia de los cortes se evaluó por PCR en tiempo real de acuerdo a la metodología mencionada anteriormente.



### 3.3.2. Preparación de librerías para secuenciación

La elaboración de las librerías se realizó con el kit Ligation sequencing kit SQK-LSK109 (Oxford Nanopore) siguiendo el protocolo del proveedor.

Luego de la ligación de los adaptadores de secuenciación, se juntaron los dos ADN que se vinieron tratando previamente por separado y se realiza el paso final de purificación con Ampure beads (Beckman Coulter, Brea, CA) según el protocolo de Oxford Nanopore.

La secuenciación de la librería fue evaluada en 2 celdas distintas: MinION flow cell FLO-MIN106 y Flongle flow cell FLO\_FLG001 (Oxford Nanopore). Cuando se utilizó la celda de secuenciación FLO-MIN106 se utilizó el Flow Cell Priming Kit EXP-FLP0002 (Oxford Nanopore). Mientras que cuando se utilizó la celda FLO-FLG001, se utilizó el Flongle Sequencing Expansion Kit EXP-FSE0001 (Oxford Nanopore). Con ambas celdas se siguieron protocolos de Oxford Nanopore.

## 3.4. Bioinformática

Para realizar el análisis bioinformático de los datos obtenidos en la secuenciación, fue requerido una estación de trabajo con un sistema operativo Linux. El control del dispositivo de secuenciación, la adquisición de datos y la *basecalling* en tiempo real se llevaron a cabo mediante el software MinKNOW (Oxford Nanopore).

Lo primero que se realizó fue instalar todos los softwares a utilizar. Para evitar que errores de procedimiento afectaran el servidor, primero se instaló Conda que es el software que gestiona los paquetes que se utilizaran a posteriori en el análisis bioinformáticas. Se trabajó sobre un *script* ofrecido por la comunidad de Oxford Nanopore (ont\_tutorial\_cas9). A través de Snakemake<sup>22</sup> se gestiona el trabajo evitando de esta manera la realización de cada programa en forma manual. Este flujo de trabajo incluye la alineación de las secuencias al genoma de referencia con Minimap2<sup>23</sup>, manejo de los archivos SAM/BAM con Samtools<sup>24</sup>, identificación de las secuencias *On target*, *Off target* y presentación de los resultados, mediante los paquetes RSamtools, GenomicAlignments y ggplot2 del software R<sup>34</sup>.

Para visualizar los resultados se utilizó Integrative Genomics Viewer (IGV)<sup>25</sup>. IGV es una herramienta que permite la visualización de datos genómicos en una forma sencilla y de fácil interpretación. El software proporciona funcionalidad para la visualización de datos de mapeo de secuencias de archivos BAM que posteriormente se pueden superponer con "pistas" de información que pueden incluir profundidad de cobertura para datos de mapeo y anotaciones de genes.

Las principales métricas utilizadas para mostrar la eficiencia de un experimento dirigido por Cas9 son el rendimiento general, el % de lecturas acertadas, la cobertura de la región de interés y el agotamiento del ADN no objetivo.

En el Anexo se puede acceder al pipeline.

## Resultados y discusión

---

### 1. Puesta a punto de PCR tiempo final para amplificar los genes *TCF25*, *MC1R* y *TUBB3*

Con el objetivo de amplificar la región que comprende los genes *TCF25*, *MC1R* y *TUBB3* se realizó 2 rondas de PCR (*Nested PCR*) utilizando oligonucleótidos específicos diseñados previamente. El diseño de cebadores para dicha región comprendía aproximadamente 6 kb de cada gen, en fragmentos solapados de aproximadamente 1kb para la primer ronda y cebadores internos de aproximadamente el mismo tamaño para la ronda 2 del *Nested PCR*. Se diseñaron solapados para evitar perder parte de la secuencia en la región de los cebadores y las uniones de los tres amplicones, y así reconstruir la secuencia completa de la región cromosómica.

#### 1.1. Long PCR con el mix de Taq Polimerasa

Se probaron diferentes condiciones tal como se describió previamente. Las diferentes temperaturas de *annealing* evaluadas estuvieron en un rango de 55°C a 63°C para ambas rondas. En todas las pruebas con las distintas temperaturas de *annealing* no se observaron bandas en el gel de agarosa del tamaño esperado.

Los productos de PCR fueron visualizados por electroforesis en gel de agarosa, evaluándose porcentajes de 0.8 % y 1 %. No se observaron diferencias significativas entre los dos tipos de geles, por lo que se decidió a seguir trabajando con 1%.

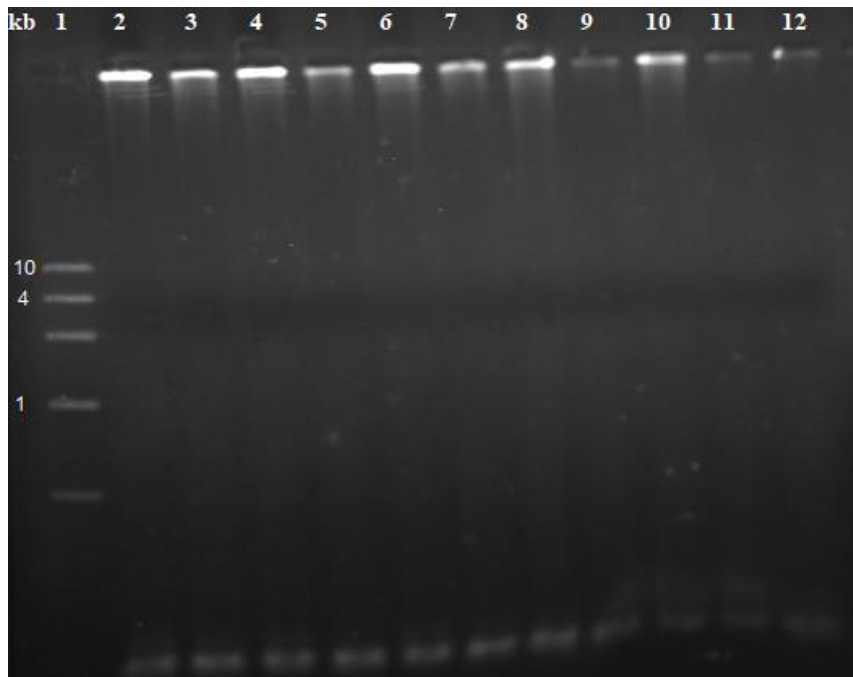
Los resultados obtenidos que se observaron en la electroforesis en gel de Agarosa fueron inconclusos. No se observaron bandas de 6 kb en ninguna de las pruebas realizadas con 2 ADNs distintos y a distintas temperaturas. Para lograr amplificar los fragmentos de 6 kb esperados se realizaron pruebas adicionales. Primero, se ajustó la reacción de PCR modificando la concentración de MgCl<sub>2</sub>, y luego eliminando el Tris. En ambos casos, el resultado continuó siendo el mismo, no se detectó amplificación.

Ante la posibilidad de que fuera producto de estructuras secundarias, se probó a desnaturalizar por calor (5 minutos a 95°C) los productos de PCR y luego realizar la electroforesis en gel de Agarosa, pero el resultado siguió incambiado, no se pudo detectar bandas del tamaño esperado. Este resultado se puede deber a que la zona en donde se encuentra los genes *TCF25*, *MC1R* y *TUBB3*, es una región genómica muy difícil de amplificar porque contiene grandes zonas de secuencias repetidas, como se discute más adelante.

#### 1.2. Long PCR con la enzima Ranger ADN Polimerasa

Al no lograr amplificar la región de interés con el mix de ADN polimerasas, se realizaron pruebas con la enzima *Ranger* ADN Polimerasa (Bioline, LDN, UK).

Se probaron diferentes condiciones de la PCR como se describió en la metodología. Las temperaturas de *annealing* utilizadas variaron entre 55°C y 63°C.



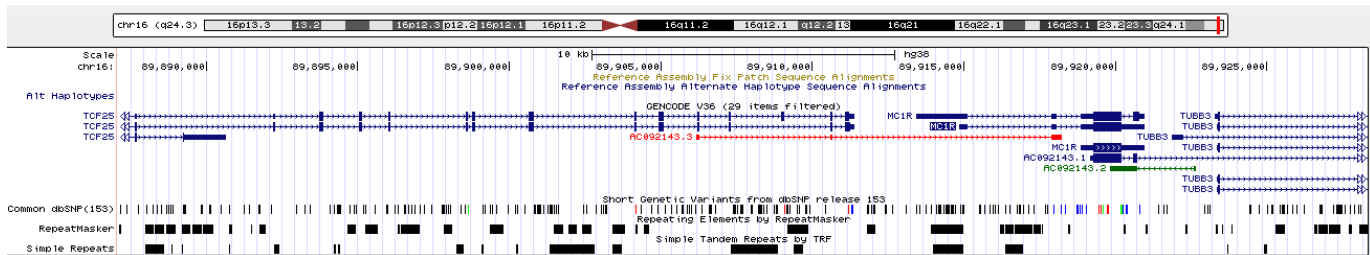
**Figura 13.** Gel de electroforesis de Agarosa 1% de los productos de PCR. Carril 1: Marcador de peso molecular Fast Ruler DNA LADDER, High range 1kb (Fermentas Life Sciences, #SM1128). Carril 2 al 12: PCR con la enzima Ranger DNA Polymerase con temperaturas de *annealing* de 53°C al 63°C en orden.

Se realizaron electroforesis en gel de Agarosa 1% para visualizar los productos de la PCR (Figura 13). No se observaron productos de PCR a un tamaño de 6kb para ninguna de las diferentes temperaturas de *annealing* evaluadas. En este caso, se observó bandas en los pocillos del gel. Ante la posibilidad de que se deba a una gran concentración de ADN que no le permitía migrar, se probaron concentraciones más bajas de ADN. Sin embargo, se continuó sin observar la banda de interés.

Con ambos abordajes de Long PCR no se logró amplificar la región del gen *MC1R* y zonas flanqueantes. Esto puede ser en parte porque como se observa en la Figura 14 la región genómica que comprende el *MC1R*, *TUBB3* y *TCF25* tiene una gran cantidad de SNPs y secuencias repetidas que dificultarían su amplificación en grandes fragmentos.

Anteriormente algunos investigadores han reportado dificultades para amplificar regiones genómicas en grandes fragmentos. Por ejemplo, Barnes y colaboradores intentaron amplificar una secuencia de ADN de 6,4 a 18kb y se encontraron con que utilizando el ciclo de desnaturalización de 1 o 2 minutos a 95°C o 98°C no lograban amplificar la secuencia esperada, por lo que probó reduciendo el tiempo a 2 o 20 segundos a 93°C o 94°C y si lograba obtener un buen producto de amplificación. Además, probaron el paso de desnaturalización a 94°C a 20, 60 y 180 segundos y el producto de amplificación disminuía a medida que aumentaba el

tiempo, por lo que concluían que un componente de la reacción no era termoestable <sup>35</sup>. De manera que también esto podría estar influyendo en los resultados obtenidos.



**Figura 14.** Región cromosómica chr16 (q24.3) que comprende los genes *TCF25*, *MC1R* y *TUBB3*. Visualización en la región de la presencia de SNPs y bloques de secuencias repetidas mediante Genome Browser.

En el mismo momento en que nos encontramos con estos inconvenientes se publica una estrategia que prometía ser más simple de implementar: la secuenciación dirigida por Cas9. Por esta razón no se siguió profundizando en esta metodología y comenzamos a implementar en el laboratorio esta nueva estrategia.

## 2. Secuenciación dirigida por Cas9

En este estudio, presentamos un flujo de trabajo para la secuenciación dirigida por CRISPR-Cas9 para una región de aproximadamente 10 kb que contiene el gen *MC1R*. Es un método que combina la escisión dirigida CRISPR-Cas9 de una región de interés, seguida de enriquecimiento y secuenciación de lectura larga utilizando la tecnología Oxford Nanopore (ONT).

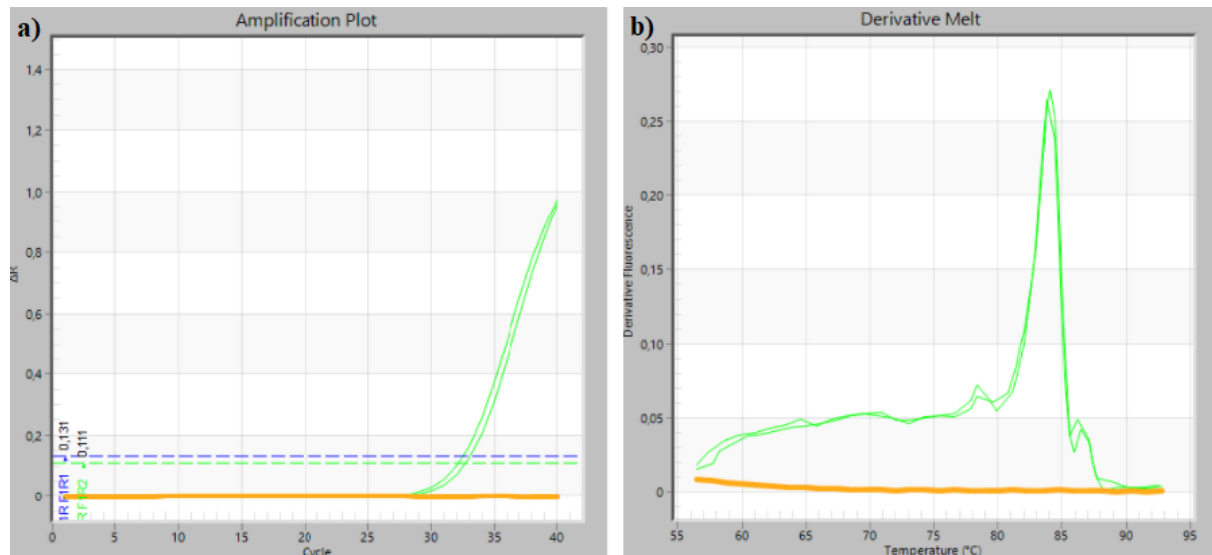
### 2.1. Puesta a punto de la evaluación del corte con CAS9

#### 2.1.1. Evaluación de los cortes de Cas9 por PCR en tiempo real

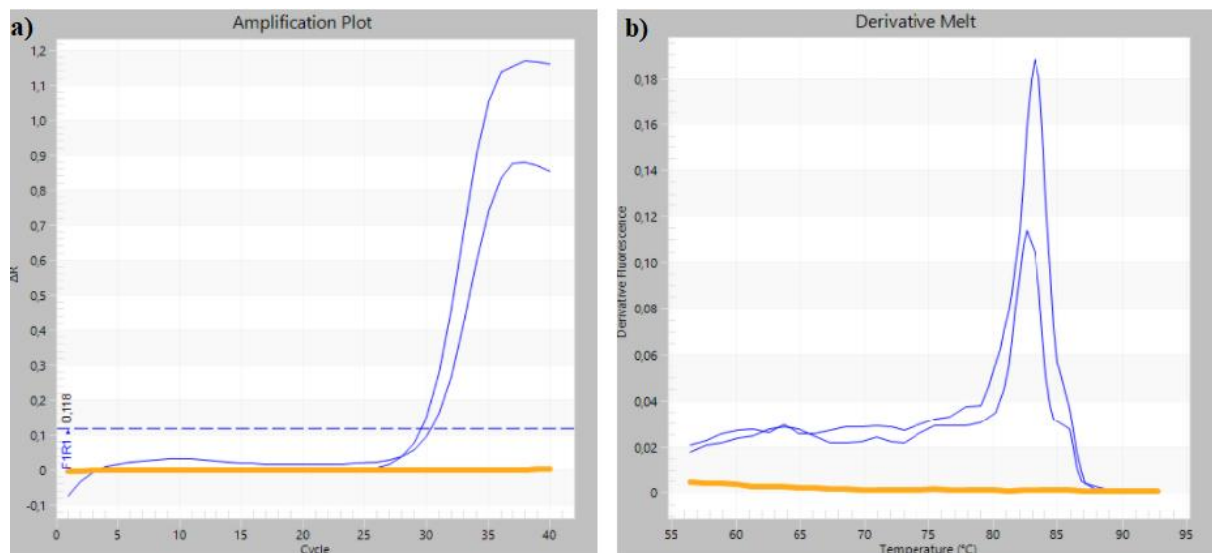
Debido a que mediante la metodología nCATS se llegaba a la secuenciación sin ningún tipo de información acerca de la eficiencia de corte de la Cas9 en el ADN, fue necesario pensar en un método que evaluara el corte que realizaba la enzima. Para esto se diseñó la estrategia de medir la eficiencia de los cortes con un sistema basado en evaluar la amplificación de la región de corte por PCR en tiempo real. Si bien la enzima CAS9 digería 3 cortes en la región del gen *MC1R*, consideramos que evaluando sólo uno de los cortes nos daría una idea de la eficiencia de este proceso previo a la secuenciación.

Se diseñaron los cebadores para la evaluación de corte de la Cas9: dos pares de cebadores que amplificaban una secuencia de 121pb externa al sitio de corte y otra secuencia de 260 pb que incluye dentro del amplicón el corte con el ARN guía (5'CTGCGGAGCACACGTGCGA 3').

Luego de ajustar las condiciones de amplificación, se implementó en el laboratorio una metodología sencilla y rápida para evaluar la eficiencia de cortes dirigidos por CAS9 en la región de interés. En las Figuras 15 y 16 se observan las amplificaciones por PCR en tiempo real del fragmento corto (121 pb) externo al corte y del fragmento largo (260 pb) que incluye uno de los cortes dirigido por CAS).



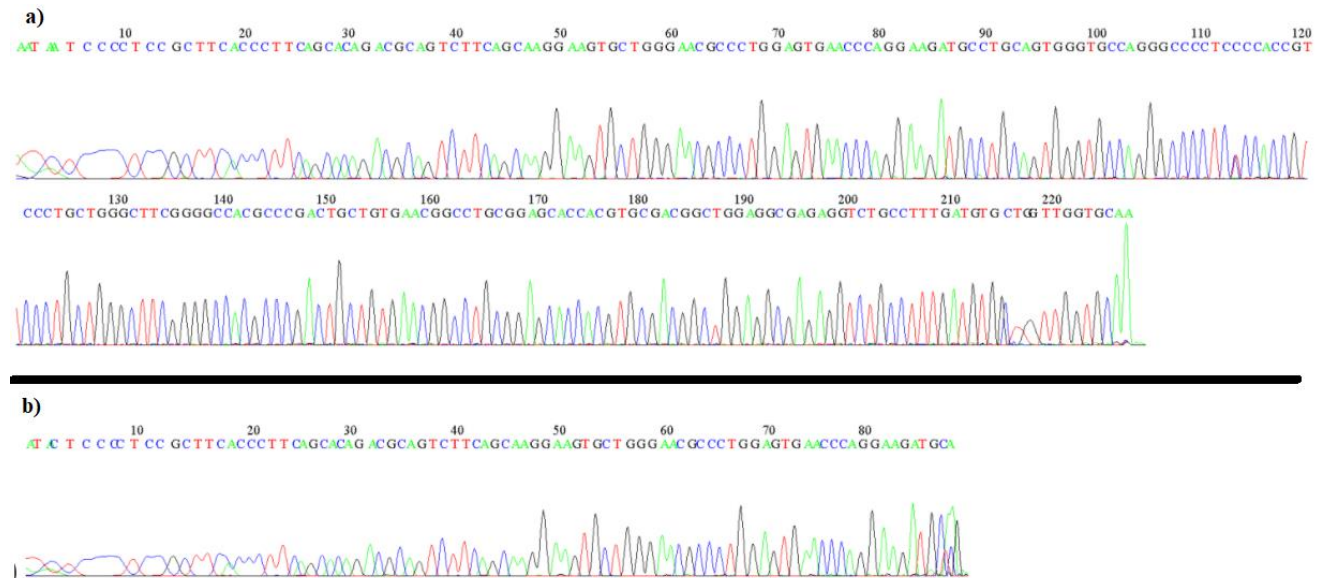
**Figura 15.** Amplificación por PCR en tiempo real de fragmento externo al corte dirigido por CAS9. a) Curva de amplificación con 2 ADNs y control negativo (Fluorescencia vs ciclos). b) Curvas de melting de los productos amplificados (Fluorescencia vs derivada segunda de la temperatura). En naranja: control negativo, en verde: muestras de ADN de prueba.



**Figura 16.** Amplificación por PCR en tiempo real de fragmento interno al corte dirigido por CAS9. a) Curva de amplificación con 2 ADNs y control negativo (Fluorescencia vs ciclos). b) Curvas de melting de los productos amplificados (Fluorescencia vs derivada segunda de la temperatura). En naranja: control negativo, en azul: muestras de ADN de prueba.

En ambos casos se logró una amplificación eficiente ( $C_t = 28$ ) y específica ya que se observa un único pico de *melting* en los amplicones. A su vez, se amplificó por PCR en tiempo final ambos fragmentos y se visualizaron en gel de agarosa para confirmar la presencia de fragmentos específicos del tamaño esperado.

Los productos de PCR de ambos fragmentos fueron secuenciados en Macrogen (Corea del Sur) con el fin de verificar que la secuencia corresponde a la región de interés. Los electroferogramas correspondientes a las secuenciaciones se observan en la Figura 17.



**Figura 17.** a) Electroferograma de la secuenciación del fragmento externo al corte dirigido por CAS9. b) Electroferograma de la secuenciación del fragmento interno al corte dirigido por CAS9.

Se realizó un BLAST de ambas secuenciaciones en NCBI (<https://www.ncbi.nlm.nih.gov/>) para chequear que sean las secuencias esperadas. El resultado del BLAST indicó que la secuencia es la esperada.

Este método de evaluación de los cortes dirigidos por Cas9 tiene una gran importancia en el laboratorio ya que nos brinda una información valiosa para poder seguir con los siguientes pasos de secuenciación. Permite saber si el sistema CRISPR-Cas9 está funcionando correctamente, evitando así pasar a la parte de secuenciación sin información de corte. Además, se evita pérdidas de tiempo y dinero, ya que en caso de no tener corte no se prosigue a la secuenciación. Vale destacar que este método de evaluación de corte es bastante novedoso, ya que no se encontró en bibliografía un enfoque similar.

## 2.2. Secuenciación por MinION

En una primera instancia se probaron un grupo de ARN guías (como se describen en la Tabla 2) con la metodología de nCATS, se secuenció utilizando la Flow Cell FLO-MIN106 y no se obtuvieron secuencias *target*. El diseño de estos tres crARN generaban dos fragmentos de un tamaño entre 30 a 50 kb.

Cabe destacar que en ese momento no se había implementado aún en el laboratorio la evaluación de la eficiencia de los cortes por lo cual no pudimos definir dónde estaba el problema de la ausencia de *reads* para la secuencia *target*. Por lo que se plantearon distintos problemas y soluciones. Uno de los posibles problemas es que los fragmentos eran muy grandes y eso disminuía la eficiencia de la secuenciación. Otro problema puede deberse a que no estaban funcionando los ARN guías o la Cas9, ya que no era posible evaluar el corte.

Para resolver esto, se implementó en el laboratorio la evaluación de corte de los ARN guías descrita anteriormente y se procedió a diseñar nuevos ARN guías (Tabla 3 y 4).

Estos nuevos crARN están basados en una nueva técnica de diseño que es la de *Tiling*, donde están diseñadas a lo largo de la región de interés en fragmentos superpuestos de 5-10 kb que permiten una cobertura uniforme en toda la región de interés.

En una primera prueba se eligieron tres ARN guías para probar si CRISPR-Cas9 podría escindir específicamente un segmento de 13409 pares de bases (bp) del gen *MC1R* y *TUBB3*. Los tres crARN tenían puntuaciones de especificidad del 76 al 96% (Tabla 3) para la producción de longitudes de fragmentos específicos basados en el ensamblaje del genoma de referencia. Los valores de *off target* te dan la especificidad, con una puntuación del 100% predice una baja actividad fuera del *target*.

Los nuevos crARN que se utilizaron como prueba para las siguientes secuenciaciones fueron:

5' GGCGTGAACAGACCACGTCT 3'

5' CTGCGGAGCACCACGTGCGA 3'

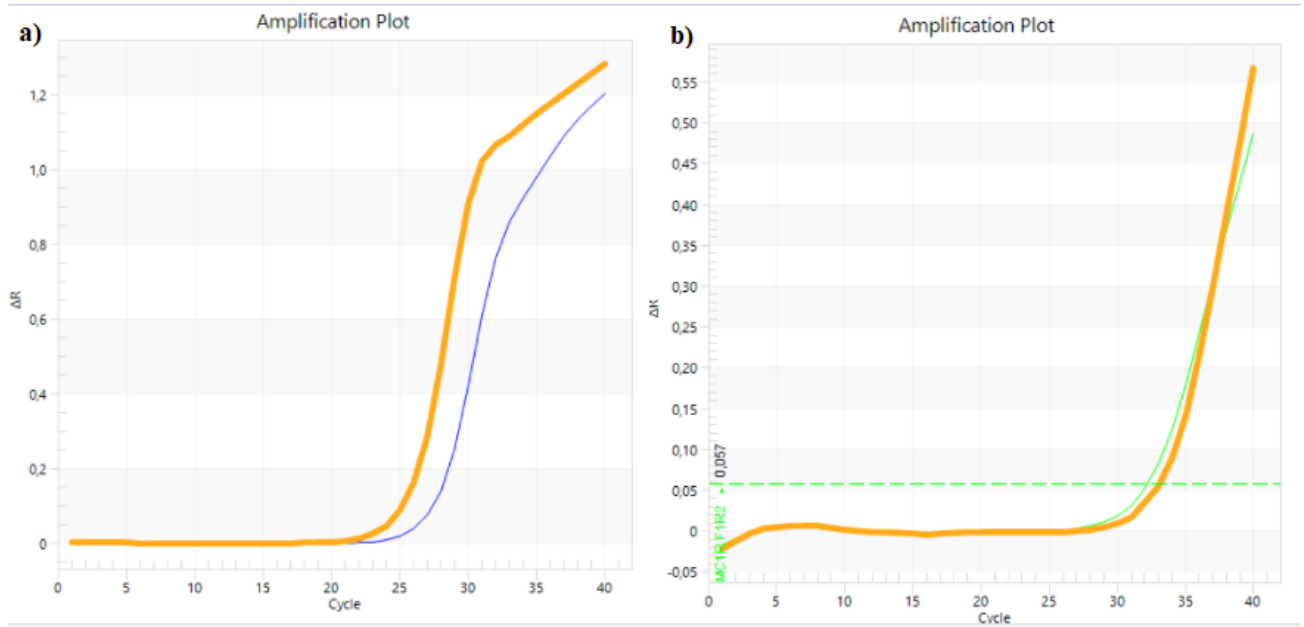
5' TAACGCGGACAGCATGGACA 3'

Se probó secuenciar la región de interés con los nuevos ARN guías en 2 celdas distintas. Por una parte, se realizó la metodología de nCATS para la celda MinION flow cell FLO-MIN106 (*Secuenciación 1*) y por otro lado se utilizó la celda Flonge flow cell FLO\_FLG001 (*Secuenciación 2*).

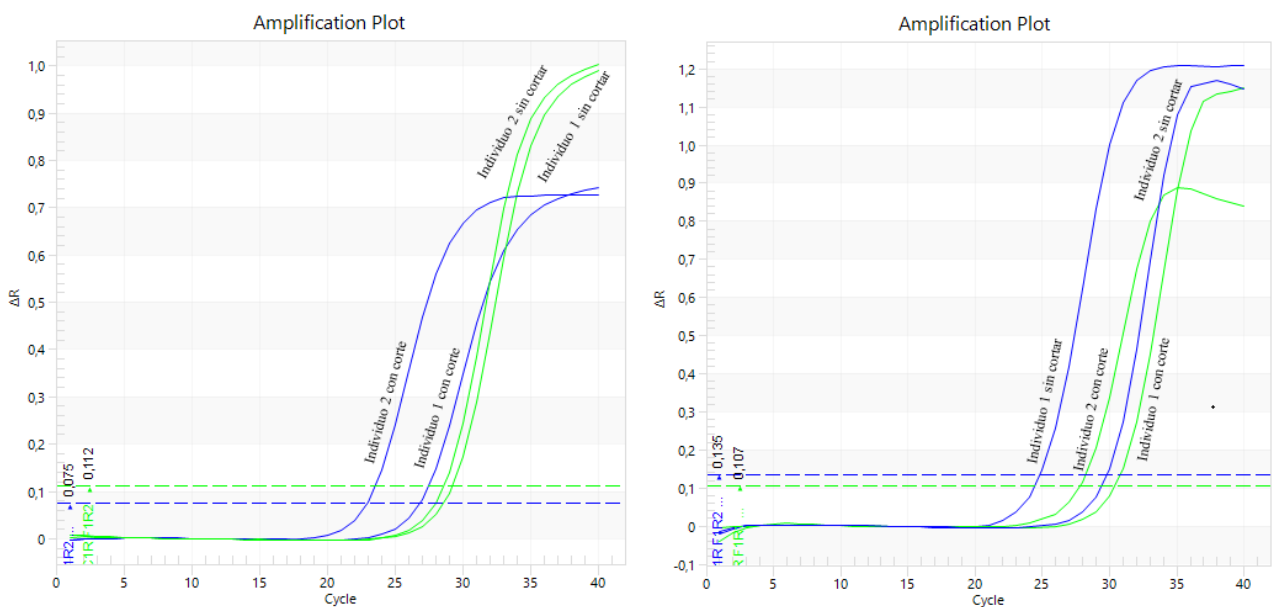
Para ambas secuenciaciones se obtuvieron los fragmentos de la región de interés mediante cortes con la metodología CRISPR-Cas9. Luego se evaluaron estos cortes por PCR en tiempo real, se preparó la librería de secuenciación y se secuenció por MinION.

Para la *Secuenciación 1* se realizó la secuenciación dirigida por CRISPR-Cas9 de una sola muestra de ADN, mientras que para la *Secuenciación 2* se realizó la secuenciación dirigida por CRISPR-Cas9 de dos muestras de ADN. Para esta última secuenciación, si bien uno de los ADN es el mismo que para la primera, las extracciones se realizaron un poco diferentes. Para la segunda secuenciación se realizó una extracción donde se cuidó más que el ADN no se fragmentara, evitando vortexear y centrifugaciones de mayor duración pero a menor rpm, incluso se evitó el vortexeo durante la preparación de la librería.

Se evaluó el corte del ARN guía (5'CTGCGGAGCACCACGTGCGA 3') por PCR en tiempo real con las condiciones mencionadas en la puesta a punto de dicho método. En las Figura 18 y 19 se observan la amplificación por tiempo real de los fragmentos externo e interno al corte dirigido por CAS9 para la *Secuenciación 1* y *2* respectivamente.



**Figura 18.** a) Amplificación por PCR en tiempo real de fragmento externo al corte dirigido por CAS9 para la *Secuenciación 1* (Fluorescencia vs ciclos). En naranja: ADN sin cortar, en azul: ADN cortado. b) Amplificación por PCR en tiempo real de fragmento interno al corte dirigido por CAS9 para la *Secuenciación 1* (Fluorescencia vs ciclos). En naranja: ADN sin cortar, en azul: ADN cortado.



**Figura 19.** a) Amplificación por PCR en tiempo real de fragmento externo al corte dirigido por CAS9 para la *Secuenciación 2* (Fluorescencia vs ciclos). En azul: ADN de los dos individuos sin cortar, en verde: ADN de ambos individuos cortado. b) Amplificación por PCR en tiempo real de fragmento interno al corte dirigido por CAS9 para la *Secuenciación 2* (Fluorescencia vs ciclos). En verde: ADN de los dos individuos sin cortar, en azul: ADN de ambos individuos cortado.



Se realizó el cálculo de RC (Tabla 6) utilizando la lógica de la cuantificación relativa en base a una referencia. Como se puede observar en la Tabla 6 todos los cortes tuvieron una buena eficiencia de corte, ya que todos tienen un valor por encima del valor aceptado como correcto (RC = 5).

**Tabla 6.** Valores de Ct y RC de los ADN digeridos y sin digerir para ambas secuenciaciones.

		Individuo 1			Individuo 2		
		Ct ADN digerido	Ct ADN sin digerir	RC	Ct ADN digerido	Ct ADN sin digerir	RC
Secuenciación 1	Fragmento Largo	24.91	27.15	8.63	-	-	-
	Fragmento Corto	33.03	32.19		-	-	-
Secuenciación 2	Fragmento Largo	30.38	24.77	171.25	27.82	29.73	8.93
	Fragmento Corto	26.8	29.19		22.9	28.55	

Además, se puede apreciar que para el individuo 1 el RC es muy distinto en las pruebas. Para este individuo en la *Secuenciación 1* tiene una menor eficiencia de corte (RC = 8.63) que la *Secuenciación 2* (RC = 171.25). Esto puede explicarse a que el ADN de la segunda prueba estaría menos fragmentado que el otro ADN. Lo único que cambió para el individuo 1 en ambas pruebas fue una mejor extracción del ADN, mucho más cuidada, buscando evitar la fragmentación.

Se puede observar que pequeños cambios en la metodología pueden dar resultados muy distintos. Una buena extracción de ADN cuidando que no se fragmente puede cambiar por completo la eficiencia del corte.

Como todos los cortes tuvieron una buena eficiencia, se prosiguió a la elaboración de las librerías y secuenciación por MinION con la celda correspondiente a cada prueba.

Como se mencionó anteriormente en la *Secuenciación 1* se realizó una secuenciación en la MinION flow cell FLO-MIN106 durante 24hs con una cantidad de 982 poros activos iniciales. Mientras que para la *Secuenciación 2* se utilizó la Flonge flow cell FLO\_FLG001 durante 24hs con una cantidad de 70 poros activos iniciales.

Los datos obtenidos en ambas secuenciaciones como el *Background*, *off target* y *on target* se encuentran en la Tabla 7.

**Tabla 7.** Características de todo el mapeo incluyendo el *on-target*, *target-flanking* y *off-target* de ambas secuenciaciones.

		<i>Background</i>	<i>Off Target</i>	<i>Target flanking</i>	<i>On Target</i>
<i>Secuenciación 1</i>	<i>Reads</i> totales	444	3,312	1	3
	<i>Reads</i> mapeados	383	3,312	1	3
	Bases secuenciadas	1,810,496	15,277,533	4,67	12,914
	Bases mapeadas	1,787,259	15,277,533	4,67	12,914
	Fracción del genoma (%)	99441%	558%	1%	0%
	Promedio de cobertura	0	0.99	0.66	0.79
<i>Secuenciación 2</i>	<i>Reads</i> totales	29	956	0	1
	<i>Reads</i> mapeados	14	956	0	1
	Bases secuenciadas	78,33	4,476,661	0	9,671
	Bases mapeadas	74,062	4,476,661	0	9,671
	Fracción del genoma (%)	99847%	0.15%	2%	0%
	Promedio Cobertura	0	0.98	0.14	0.21

La información presentada en la tabla resume las métricas clave para comparar el rendimiento de una secuenciación de ADN siguiendo el protocolo de enriquecimiento libre de PCR mediado por Cas9.

Los *reads* del *background* resultan de la desfosforilación incompleta del ADN genómico seguida de una ligación no específica de la secuencia adaptadora. Los *reads off target* son el resultado de que la proteína Cas9 corta el ADN en una ubicación genómica fuera de la región *target*.

Se obtuvo un valor de 3,312 *reads off target* para la *Secuenciación 1* y un valor de 956 *reads off target* para la *Secuenciación 2*, unos valores muy alto, por lo que es recomendable revisar el diseño de los crARN para evaluar posibles SNP y *mismatches* en secuencias candidatas.

Para obtener la mejor información sobre el rendimiento del protocolo de enriquecimiento sin PCR mediado por Cas9, es preferible considerar el rendimiento de cada *target* por separado. La Tabla 8 destaca las características de las diferentes regiones *target* definidas dentro del archivo BED inicial de ambas secuenciaciones.

**Tabla 8.** Mapeo de la región de interés para ambas secuenciaciones.

<i>Corrida</i>	Gen objetivo	Tamaño <i>target</i> (nt)	Cobertura media	<i>Read</i>	Bases	Tamaño del <i>read</i>	Promedio calidad del <i>read</i>	Promedio calidad mapeado
<i>1</i>	<i>MC1R</i>	6105	0.79	3	12914	4305	8.53	60
<i>2</i>	<i>MC1R</i>	6105	0.21	1	9671	9671	10.71	60

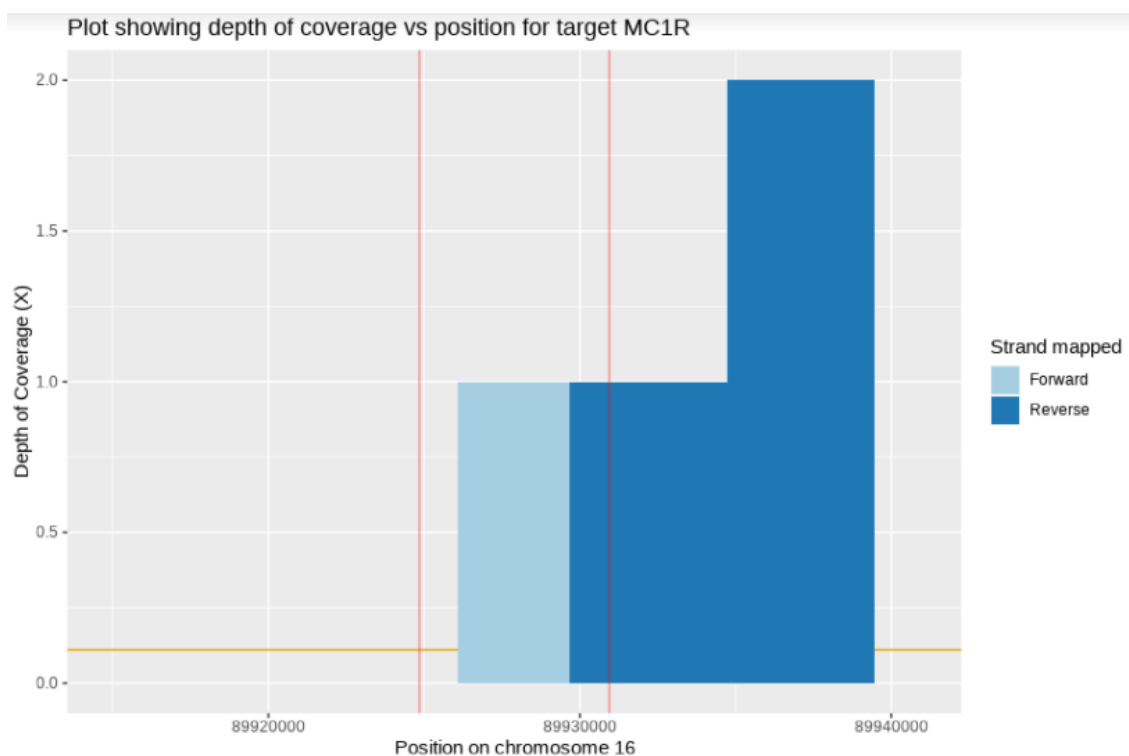
En la Tabla 8 se puede ver que la cobertura media del *target* de ambas secuenciaciones es muy por debajo de la esperada, con valores de 0.79 y 0.21.

Una calidad de mapeo perfecta debería ser 60. Un valor de 60 indica que las lecturas se están mapeando en una única ubicación en el genoma (la ubicación de destino). Las calidades de mapeo más bajas pueden indicar mapeo fragmentado (bloques de secuencia intercalados por regiones sin mapeo en una sola ubicación genómica) o mapeo múltiple (las secuencias se pueden mapear en múltiples ubicaciones en el genoma) que conducen a efectos fuera del objetivo. En este caso se obtuvo una calidad de mapeo de 60 para ambas secuenciaciones por lo que está correcto.

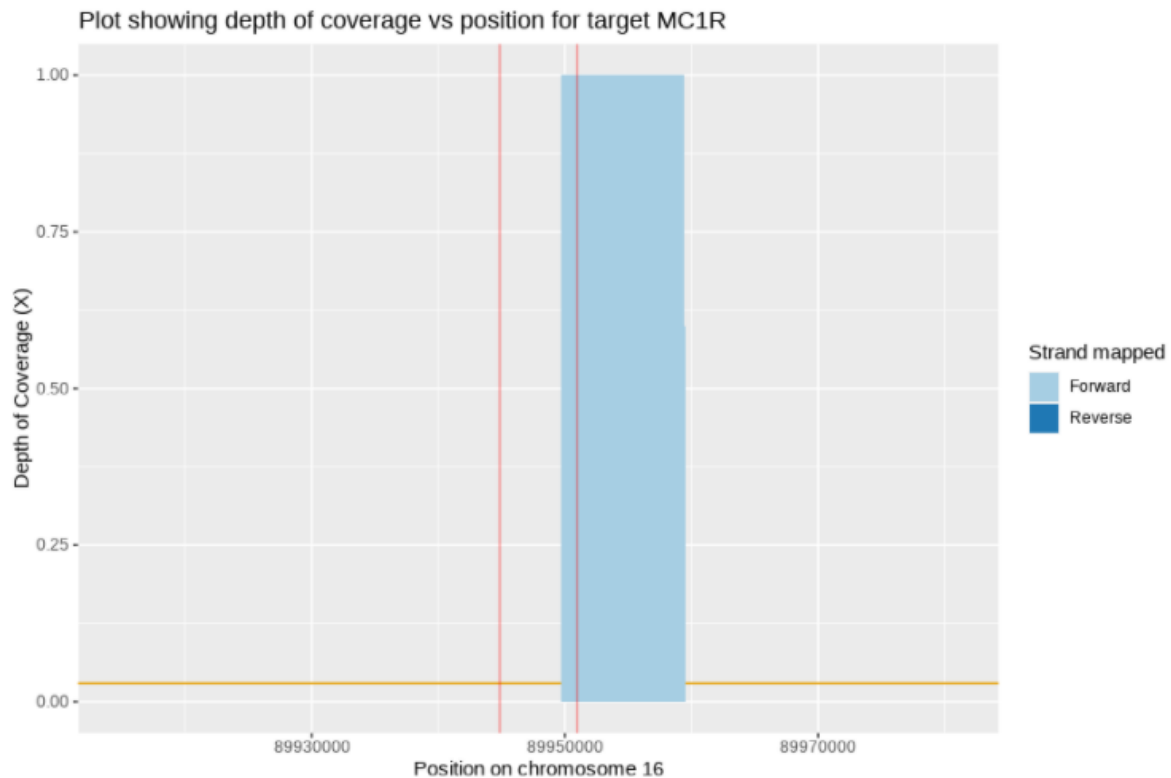
La comparación de las longitudes de los *reads* del *target* se puede utilizar para identificar los *target* (y sus crARN) que permiten la lectura completa. También se debe considerar la relación entre la longitud de lectura media y el tamaño del *target*.

Los *reads* de ONT se mapearon al genoma de referencia. Un total de 3699 (98 %) *reads* para la *Secuenciación 1* y 971 (98%) *reads* para la *Secuenciación 2* se alinearon con éxito con calidad de mapeo promedio de 60 utilizando Minimap2.

Las dos tablas presentadas anteriormente han proporcionado un resumen de las características generales del mapeo y las estadísticas sobre el *target*. Trazar la profundidad de cobertura en las regiones *target* también permite una evaluación del rendimiento de los crARN utilizados. Las gráficas correspondientes a las Figuras 20 y 21 revisan la profundidad de la cobertura de la secuencia más allá de los límites de la región *target* para ambas secuenciaciones respectivamente.



**Figura 20.** Grafica de la profundidad de cobertura en función de la posición del *target* para la *Secuenciación 1*.

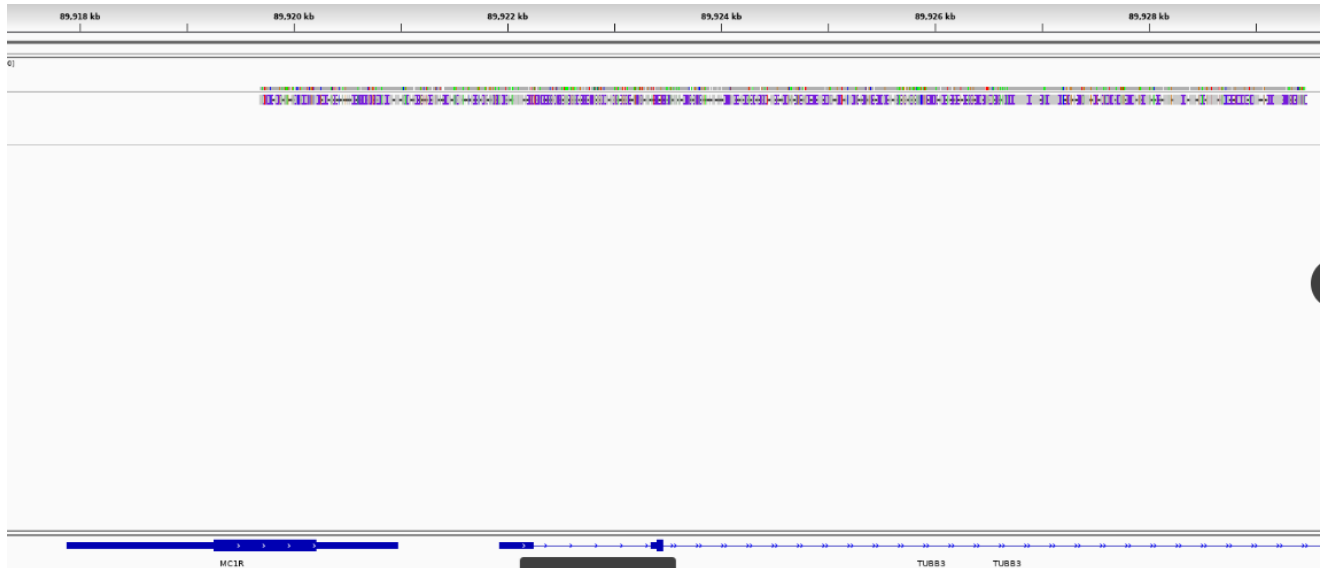


**Figura 21.** Grafica de la profundidad de cobertura en función de la posición del *target* para la *Secuenciación 2*.

Las Figuras 20 y 21 presentan la profundidad de cobertura, que se encuentra sombreada por la hebra (*Forward* o *Reverse*) a la que se asignan los *reads* mapeados. Estas figuras se pueden utilizar para observar las desviaciones de la distribución 50:50 esperada del mapeo entre las hebras + y -. Las secuencias que se extienden desde las regiones *target* y hacia las regiones proximales al *target* pueden indicar un rendimiento subóptimo de una secuencia guía de crRNA.

Como se puede observar en la Figura 20 no hay una distribución 50:50 entre las dos hebras, hay un porcentaje mayor de *reads* mapeados con la *Reverse*. Mientras que en la Figura 21 se observa también que no hay una distribución 50:50 entre las dos hebras, hay solamente un *read* en la hebra positiva. Si bien en la *Secuenciación 1* se observa un sesgo hacia la hebra negativa y en la *Secuenciación 2* hacia la hebra positiva, es difícil afirmar si el problema está en el diseño de los crARN debido a que hay pocas secuencias *target* obtenidas.

Finalmente se observó los *reads* de la secuencia *on target* de ambas secuenciaciones en el Integrative Genomics Viewer (IGV). En la Figura 22 se muestra el resultado para la *Secuenciación 2*, donde se observa el tamaño de este *read* que corresponde a 9671 pb.

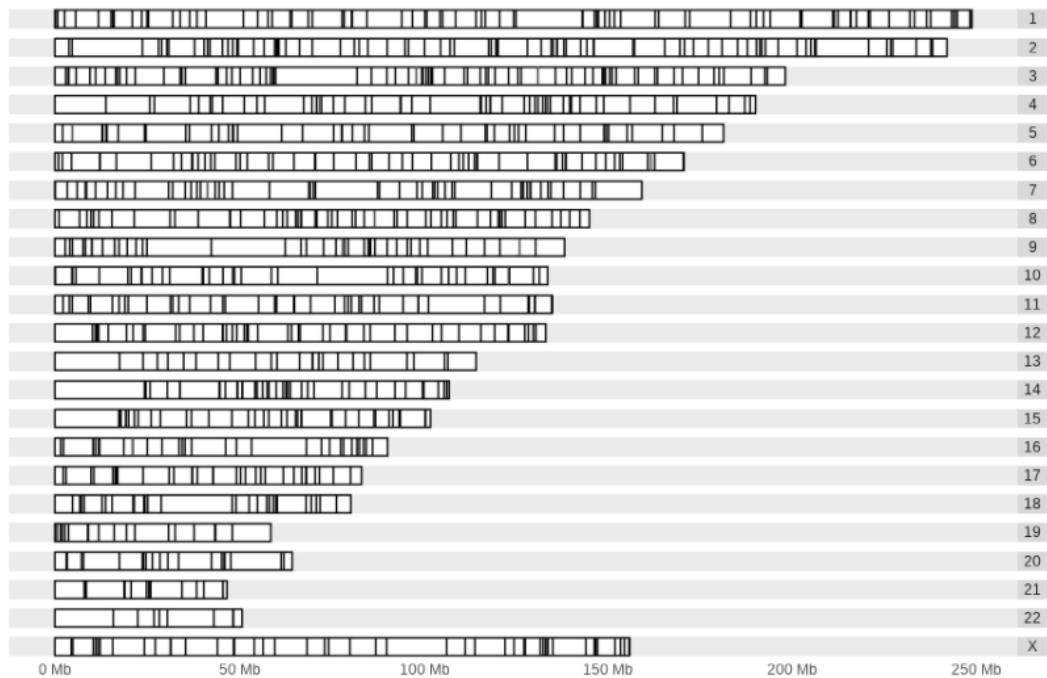


**Figura 22.** Visualización en Integrative Genomics Viewer (IGV) de *reads* mapeados en ONT. IGV mostrando el único *read* en la zona del *MC1R* y *TUBB3* de 9671 pb de la *Secuenciación 2*.

Se encontraron secuencias *off target* en distintas regiones en todos los cromosomas en ambas secuenciaciones (Figuras 23 y 24). Según trabajos previamente reportados, el análisis de cobertura de todo el genoma muestra que las lecturas *off target* se distribuyen aleatoriamente en el genoma, lo que indica que son el resultado principalmente de la unión de los adaptadores de Nanopore a puntos de rotura aleatorios <sup>20</sup>.



**Figura 23.** Ideograma de las ubicaciones *off target* ubicadas en los distintos cromosomas para la *Secuenciación 1*.



**Figura 24.** Ideograma de las ubicaciones *off target* ubicadas en los distintos cromosomas para la *Secuenciación 2*.

Los valores esperados a continuación definidos según Oxford Nanopore son para una ejecución de MinION de 24 horas con la Flow cell FLO-MIN106 para un panel bien diseñado que mira 10 regiones *target* que van desde 5-15 kb:

- Rendimiento general 1 Gb, la producción será menor siguiendo un protocolo de enriquecimiento mediado por Cas9 en comparación con un experimento de secuenciación Nanopore promedio
- 1-10% de los datos secuenciados deben estar en el *target*
- La cobertura media por *target* debe ser > 100X
- Debe observarse un agotamiento 3000X del ADN no *target*.

Para genomas y regiones de interés más grandes o más pequeños, la cobertura disminuirá o aumentará respectivamente.

Sin embargo, reportes de otros investigadores que trabajan con la misma técnica han tenido grandes variaciones en la cobertura que se obtiene que va desde 20X a 400X<sup>20,36-40</sup>.

Por ejemplo, los reportes informados por Gilpatrick y colaboradores indican que en un principio realizaron pruebas con un ARN guía de cada lado de la región de interés y obtuvieron una cobertura media que variaba de 18X a 846X para la flow cell FLO-MIN106, y una variación de 8X y 65X para la Flong flow cell. Cambiando el diseño de los ARN guías por varios para la misma región de interés observaron un gran aumento en la cobertura, > 400X para la flow cell FLO-MIN106 y > 40X para la Flong flow cell. Además, con el primer diseño obtenían un porcentaje de 1.8% *reads on target* y con el cambio al panel de varios ARN guías aumentaron a 4.6%<sup>20</sup>.

La región enriquecida *on target* tenía una profundidad de lectura mucho menor que las regiones fuera del *target*. Existen varios factores que podrían haber afectado la eficiencia de escisión. Uno de ellos puede ser que la presencia de un SNP en la secuencia PAM, puede remover el locus *target* para la escisión *off target*. Otro factor es la orientación *Forward* de dos de los crARN que proporcionan direccionalidad para las hebras que se alejan de la región de interés como Cas9, y permanece estrechamente unida después de la escisión en el extremo del PAM, bloqueando la ligación de los adaptadores ONT. Si hay presencia de SNP en la secuencia del protoespaciador, algunas hebras podrían haberse liberado del enlace de Cas9 con el resultado de que se ligaron con adaptadores <sup>40</sup>.

Según la bibliografía, la principal consideración para utilizar con éxito este enfoque de captura de secuencias es la calidad del ADN. Es necesario obtener de la mejor forma el ADN genómico a utilizar, evitando la fragmentación del mismo. Si el ADN se fragmenta hay mayor probabilidad de que los adaptadores de secuenciación se unan a estos fragmentos y se generen muchas secuencias *off target*. Los inhibidores que afectan la digestión del ADN y reducen la secuenciación de lecturas muy largas se eliminan mediante un buen protocolo de extracción de ADN. Aunque implique algunos pasos más que las preparaciones estándar de ADN genómico total, vale la pena optimizar el protocolo de extracción del ADN para evitar desperdiciar datos de secuencia valiosos <sup>40</sup>.

Gilpatrick y colaboradores descubrieron que con el uso del kit Circulomics NanoBind kit para la extracción de ADN aumentaban significativamente los *reads* en la región de interés <sup>20</sup>. Este kit sirve para mejorar la pureza y reducir el tiempo de extracción de 2-3 días a un solo día. Otro método que se podría utilizar que es compatible con MinION es el de Circulomics Short Read Eliminator (Circulomic) este método mejora las longitudes de lectura larga de secuenciación eliminando el ADN corto. Los kits de eliminación de lectura corta utilizan un método simple de precipitación selectiva por tamaño de 1 hora <sup>41</sup>.

Según López Girona y colaboradores otro aspecto crítico para una buena secuenciación dirigida por CRISPR-Cas9 es la digestión eficaz del ADN mediante los complejos de RNP. Se sabe que los ARN guía no siempre funcionan tan eficientemente como se predijo. Una estrategia alternativa podría ser utilizar varios RNP dirigidos a una sola región y luego combinarlos para el enriquecimiento y la secuenciación del ADN. Es probable que este enfoque sea particularmente valioso si los sitios *target* de CRISPR no se han secuenciado completamente (por ejemplo, secuenciando un locus de muchas muestras), o cuando hay varios haplotipos (por ejemplo, aislando un locus de especies poliploides) <sup>40</sup>.

Otro punto crítico que puede haber interferido en la eficiencia del resultado es el estado de las celdas de flujo. Las celdas de flujo tienen un vencimiento en donde los poros se van degradando con el tiempo, lo que hace que tengas que utilizarlas dentro de un período de tiempo especial dependiendo de la celda. Lo ideal es que las celdas de flujo Flongle se utilicen dentro de las 4 semanas posteriores a su recepción, mientras que las celdas de flujo normales tienen un período de 3 meses posteriores a su recepción. Debido a la situación sanitaria durante la realización del trabajo de tesis, que durante largos períodos de tiempo no pudimos trabajar en el laboratorio, las secuenciaciones se realizaron con celdas fuera del tiempo de uso óptimo. Si bien las celdas utilizadas pasaron el control de calidad inicial de poros activos, notamos que los poros se

inactivaron en un corto tiempo iniciada la secuenciación lo que influyó en la cantidad de Gb obtenidos de secuencias de buena calidad.

Se ha documentado que la secuenciación ONT es propensa a errores, lo que requiere una corrección de errores cuidadosa, generalmente mediante el uso de una alta profundidad de secuenciación o pulido con lecturas cortas más precisas. La profundidad de lectura obtenida por nuestro método de enriquecimiento no fue suficiente para corregir el error de las lecturas.

Otra metodología que se utiliza para la secuenciación dirigida por CRISPR-Cas9 es la de Wallace y colaboradores. Este es un método que es eficaz y útil para secuenciar *targets* grandes y estructuralmente complejos. El método CaBagE aprovecha la unión estable de la Cas9 a su ADN diana para proteger los fragmentos deseados de la eliminación del *background* por la digestión con exonucleasa. El ADN flanqueado por Cas9 se conserva en medio de una extensa digestión del ADN genómico por exonucleasas, lo que permite un enriquecimiento de la diana altamente específico sin PCR. Los fragmentos de ADN enriquecidos se secuencian luego con la tecnología de secuenciación de lectura larga MinION de Oxford Nanopore. El enriquecimiento con CaBagE dio como resultado una media de cobertura de 116X (rango 39–416) del *target* cuando se probó en cinco *targets* de 4 a 20 kb de longitud <sup>38</sup>. A finales del 2020 salió un nuevo software que implementarlo mejoraría los tiempos y la cobertura. Este software se llama UNCALLED que es un mapeador de código abierto que hace coincidir rápidamente en tiempo real la transmisión de señales de corriente de nanoporos con una secuencia de referencia. UNCALLED considera probabilísticamente k-meros que podrían estar representados por la señal y luego eligen los candidatos basándose en la referencia codificada. Así el software lee en tiempo real una secuencia y si no es parte de la secuencia *target* la rechaza del poro y permite que se otra secuencia nueva al poro <sup>37</sup>.

Si bien a lo largo del trabajo de tesis se fueron resolviendo y ajustando protocolos, es necesario realizar más pruebas de secuenciado, fundamentalmente con celdas de calidad óptima, para realmente hacer una evaluación de la técnica y el *pipeline* de trabajo desarrollado.



## Análisis Económico

En el marco de esta tesis se realizó un análisis económico aproximado, en donde se estudió el costo de los insumos necesarios para realizar una secuenciación dirigida por CRISPR-Cas9 por la plataforma MinION. Hay que destacar que en este análisis no se tuvieron en cuenta los costos fijos correspondiente al equipamiento del laboratorio.

En el desarrollo de esta tesis se utilizó un método de extracción de ADN tradicional, con soluciones preparadas en el propio laboratorio. No se incluyó en el análisis económico esta parte ya que no se utilizó ningún kit de extracción, solamente se utilizaron materiales comunes del laboratorio.

Se utilizaron dos tipos de celdas de flujo para la secuenciación con MinION. El MinION tiene un costo inicial de USD 1000, el starter pack contiene un MinION, una Flow cell, un Ligation sequencing kit y un cell Wash kit.

La secuenciación con la celda de flujo Flonge flow cell requiere de un adaptador para ser utilizado en el dispositivo MinION. Este starter pack tiene un costo de USD 1460 y contiene un adaptador Flonge y doce Flonge flow cells. Por lo que si se utilizan las celdas Flonge es necesario una inversión inicial de USD 2460.

Se realizó el análisis para secuenciar mediante dicho método una secuencia de 10kb de 1 individuo conteniendo 2 crARN con la celda de Flonge (Tabla 9). Este método tiene un costo aproximado de USD 171.

**Tabla 9.** Insumos necesarios para secuenciar 10 kb de 1 individuo con la celda Flonge cell de MinION.

INSUMO	PRECIO (USD)	PRESENTACION	CONC. FINAL	VOLUMEN (uL)	UNIDAD	CANTIDAD UTILIZADA c/u (uL)	COSTO (USD)
Alt-R® S.p. HiFi Cas9 Nuclease V3	245	100ug	20ug	50	1	0.33	1.61
Alt-R® CRISPR-Cas9 crRNA	95	2nmol	100uM	20	2	0.0625	0.59
Alt-R® CRISPR-Cas9 tracrRNA	95	5nmol	100uM	50	1	0.083	0.16
CutSmart® Buffer	9,4	10x	1.25x	40000	1	1.76	0.0004
Fosfatasa AP Ultrafast	140	1000U	1U/ul	1000	1	3	0.42
Taq Polimerasa	100	500U		200	1	1	0,33
dATP	50	100mM	10uM	10000	1	1	0.005
Kit SQK-LSK109 + EXP-FLP002	599				6	1	99.83
Flonge flow cell FLO-FLOG001	810				12	1	67.5
<b>TOTAL</b>							<b>170.5</b>

El costo para secuenciar 10 kb de 1 individuo con al flow cell (Tabla 10) es de USD 1003. Estos valores serían si se utilizara una corrida en una celda solo para un individuo, sería mucho más rentable realizar varios individuos y aprovechar al máximo cada celda, lo que llevaría a que se bajen los costos notoriamente. Si se utilizara el kit de barcoding (Native Barcoding

expansion 1-12 EXP-NBD104), se podría secuenciar 12 individuos en la celda común y tendría un costo aproximado de USD 181 los 10 kb por individuo (Tabla 11).

Para evaluar el corte de los dos crARN se realizó una PCR en tiempo real. El costo de dicha PCR es de aproximadamente USD 0.8 (Tabla 12).

**Tabla 10.** Insumos necesarios para secuenciar 10 kb de 1 individuo con la Flow cell de MinION.

INSUMO	PRECIO (USD)	PRESENTACION	CONC. FINAL	VOLUMEN (uL)	UNIDAD	CANTIDAD UTILIZADA c/u (uL)	COSTO (USD)
Alt-R® S.p. HiFi Cas9 Nuclease V3	245	100ug	20ug	50	1	0.33	1.62
Alt-R® CRISPR-Cas9 crRNA	95	2nmol	100uM	20	2	0.063	0.59
Alt-R® CRISPR-Cas9 tracrRNA	95	5nmol	100uM	50	1	0.083	0.16
CutSmart® Buffer	9,4	10x	1.25x	40000	1	1.76	0.00042
Fosfatasa AP Ultrafast	140	1000U	1U/ul	1000	1	3	0.42
Taq Polimerasa	100	500U		200	1	1	0.33
dATP	50	100mM	10uM	10000	1	1	0.005
Kit SQK-LSK109 + EXP-FLP002	599					0.17	99.43
Flow cell FLO-MIN106	900				1	1	900
<b>TOTAL</b>							<b>1002.5</b>

**Tabla 11.** Insumos necesarios para secuenciar 10 kb de 1 individuo con la Flow cell de MinION utilizando el *barcoding* kit.

INSUMO	PRECIO (USD)	PRESENTACION	CONC. FINAL	VOLUMEN (uL)	UNIDAD	CANTIDAD UTILIZADA c/u (uL)	COSTO (USD)
Alt-R® S.p. HiFi Cas9 Nuclease V3	245	100ug	20ug	50	1	0.33	1.62
Alt-R® CRISPR-Cas9 crRNA	95	2nmol	100uM	20	2	0.063	0.59
Alt-R® CRISPR-Cas9 tracrRNA	95	5nmol	100uM	50	1	0.083	0.16
CutSmart® Buffer	9,4	10x	1.25x	40000	1	1.76	0.00042
Fosfatasa AP Ultrafast	140	1000U	1U/ul	1000	1	3	0.42
Taq Polimerasa	100	500U		200	1	1	0.33
dATP	50	100mM	10uM	10000	1	1	0.005
Kit SQK-LSK109 + EXP-FLP002	599					0.17	99.43
Flow cell FLO-MIN106	900				1	1/12	75
Native Barcoding Expansion (1-12) EXP -NBD104	288				1	1/12	4
<b>TOTAL</b>							<b>181.5</b>

**Tabla 12.** Insumos necesarios para realizar una PCR en tiempo real para la evaluación de los cortes con CRISPR-Cas9.

INSUMO	PRECIO (USD)	PRESENTACION	CONC. FINAL	VOLUMEN (uL)	UNIDAD	CANTIDAD UTILIZADA c/u (uL)	COSTO (USD)
Sensi fast HRM mix	150	2X		5000	1	20	0.6
Primer Forward	8	100uM	10uM	300	1	3.2	0.085
Primer Reverse	8	100uM	10uM	300	2	1.6	0.085
<b>TOTAL</b>							<b>0.770</b>

Además se puede desglosar del costo final que utilizando la celda de flujo normal para MinION con *barcoding* o la Flonge cell quedaría un costo de USD 0.02 la base nucleotídica, lo que hace un costo interesante para el laboratorio.

La implementación del método de manera interna permite independizarse de servicios externos y poder contar con los resultados en un menor tiempo.

No es posible comparar los precios con otras secuenciaciones que se realizan en Uruguay o en el exterior ya que utilizan otra metodologías y equipos. Además, al ser una técnica nueva y que todavía está en evaluación no hay nadie que lo realice.

A pesar de que estos costos son estimados, ya que deberían de contarse otros parámetros, como la mano de obra, envío de las muestras, gasto de energía, materiales descartables entre otros, podría decirse que es una técnica que bajaría de manera notoria los costos de las secuenciaciones actuales. Comparando entre las dos celdas, si se quisiera realizar la secuenciación para un solo individuo utilizaría la Flonge ya que tiene menor capacidad que la otra pero suficiente para un individuo y un valor muy por debajo de la otra celda. En cambio, si se quiere secuenciar varios individuos a la vez, optaría por utilizar la Flow cell con *barcoding* ya que es la opción más barata.

## Conclusiones y perspectivas a futuro

---

Durante esta tesis se desarrolló un *pipeline* de trabajo para la secuenciación dirigida por CRISPR-Cas9 de fragmentos de más de 10 kb en la plataforma MinION.

Se determinó el protocolo óptimo para la secuenciación dirigida mediante la técnica nCATS, donde se realizaron los ajustes adecuados para la extracción de ADN evitando la fragmentación del mismo.

Se desarrollo un método para la evaluación de los cortes que realiza el sistema CRISPR-Cas9 en el ADN mediante PCR en tiempo real. Este método es novedoso e importante para tener un control de corte antes de seguir con la secuenciación, y así evitar posibles pérdidas de tiempo y dinero.

Se realizó la secuenciación dirigida por CRISPR-Cas9 en dos celdas de flujo distintas del MinION, la celda normal y la Flonge cell. En ambas se lograron obtener *reads* en la secuencia *target*.

Es necesario realizar más pruebas de secuenciación para poder realizar una evaluación real del *pipeline* desarrollado.

Luego de mejorar la metodología de nCATS, las mejoras futuras en nuestra metodología se trasladarán al enriquecimiento de multiplexación CRISPR-Cas9. El uso de *barcoding* para indexar diferentes muestras permite un uso más eficiente de las celdas de flujo de secuenciación ONT, aumentando la ocupación de los poros y, por lo tanto, expandiendo la vida de los poros durante la secuenciación.

Se realizó la evaluación económica de los insumos requeridos para utilizar la metodología nCATs comparando entre las dos celdas de flujo y en la utilización de *barcoding* para mejorar los costos.

Esta metodología es una buena herramienta porque nos permitiría estudiar variantes génicas, variantes estructurales, haplotipos y metilación de grandes regiones genómicas en una única secuenciación.

## Bibliografía

---

1. Patel A, Belykh E, Miller EJ, et al. MinION rapid sequencing: Review of potential applications in neurosurgery. *Surgical Neurology International*. 2018;9(1). doi:10.4103/sni.sni-55-18
2. Xiao T, Zhou W. The third generation sequencing: the advanced approach to genetic diseases. *Translational Pediatrics*. 2020;9(2):163. doi:10.21037/TP.2020.03.06
3. Valderrama J, Ortigosa F, Cañas R. Métodos de secuenciación de ácidos nucleicos: Primera generación.
4. van Dijk EL, Jaszczyszyn Y, Naquin D, Thermes C. The Third Revolution in Sequencing Technology. *Trends in Genetics*. 2018;34(9):666-681. doi:10.1016/J.TIG.2018.05.008
5. Mordoh A. Secuenciación masiva de ADN: la próxima generación. *Dermatol Argent*. 2019;25(1).
6. López de Heredia U. Las técnicas de secuenciación masiva en el estudio de la diversidad biológica. *Munibe Ciencias Naturales*. 2016;64. doi:10.21630/MCN.2016.64.07
7. Lee H, Gurtowski J, Yoo S, et al. Third-generation sequencing and the future of genomics. *bioRxiv*. Published online April 13, 2016:048603. doi:10.1101/048603
8. Third generation sequencing | Stories | yourgenome.org. Accessed July 28, 2021. <https://www.yourgenome.org/stories/third-generation-sequencing>
9. Tosco Herrera E. Análisis de datos en Biocomputación: tecnología MinION y herramientas bioinformáticas. Published online 2019. Accessed July 28, 2021. <https://riull.ull.es/xmlui/handle/915/15764>
10. Lu H, Giordano F, Ning Z. Oxford Nanopore MinION Sequencing and Genome Assembly. *Genomics, proteomics & bioinformatics*. 2016;14(5):265-279. doi:10.1016/J.GPB.2016.05.004
11. Jain M, Olsen H, Paten B, Akeson M. The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. *Genome biology*. 2016;17(1). doi:10.1186/S13059-016-1103-0

12. Androniki BC, Anastasios K, Maria A, et al. DNA Preparation in *Vitis vinifera* L. For Third Generation Sequencing. *bioRxiv*. Published online March 12, 2021:2021.03.11.435003. doi:10.1101/2021.03.11.435003
13. de Lannoy C, de Ridder D, Risse J. The long reads ahead: de novo genome assembly using the MinION. *F1000Research*. 2017;6. doi:10.12688/F1000RESEARCH.12012.2
14. Mas E, Poza J, Ciriza J, Zaragoza P, Osta R, Rodellar C. Fundamento de la Reacción en Cadena de la Polimerasa (PCR). *Revista AquaTIC*. 2016;0(15). Accessed July 28, 2021. <http://www.revistaaquatic.com/ojs/index.php/aquatic/article/view/139>
15. Gambogi A, Jackson E, de los Santos J, Ferrari M, Umpiérrez M. Introducción a las metodologías utilizadas en biotecnología.
16. England Biolabs N. Datasheet for LongAmp® Taq DNA Polymerase (M0323; Lot 0101206). Accessed July 28, 2021. <http://frodo.wi.mit.edu/primer3>
17. Jiang F, Doudna JA. CRISPR–Cas9 Structures and Mechanisms. <https://doi.org/10.1146/annurev-biophys-062215-010822>. 2017;46:505-529. doi:10.1146/ANNUREV-BIOPHYS-062215-010822
18. Wang H, Russa M la, Qi LS. CRISPR/Cas9 in Genome Editing and Beyond. <http://dx.doi.org/10.1146/annurev-biochem-060815-014607>. 2016;85:227-264. doi:10.1146/ANNUREV-BIOCHEM-060815-014607
19. Riveros-Maidana R, Méndez-Ferreira A, Benítez-Candia N, et al. Sistema CRISPR/Cas: Edición genómica de precisión. *Memorias del Instituto de Investigaciones en Ciencias de la Salud*. 2020;18(1):97-107. doi:10.18004/mem.iics/1812-9528/2020.018.01.97-107
20. Gilpatrick T, Lee I, Graham JE, et al. Targeted nanopore sequencing with Cas9-guided adapter ligation. *Nature Biotechnology*. 2020;38(4):433-438. doi:10.1038/S41587-020-0407-5
21. Oxford Nanopore. Evaluation of read-mapping characteristics from a Cas-mediated targeted sequencing experiment.
22. Köster J, Rahmann S. Snakemake—a scalable bioinformatics workflow engine. *Bioinformatics*. 2012;28(19):2520-2522. doi:10.1093/BIOINFORMATICS/BTS480
23. Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*. 2018;34(18):3094-3100. doi:10.1093/BIOINFORMATICS/BTY191

24. Li H, Handsaker B, Wysoker A, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009;25(16):2078-2079. doi:10.1093/BIOINFORMATICS/BTP352
25. Thorvaldsdóttir H, Robinson J, Mesirov J. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Briefings in bioinformatics*. 2013;14(2):178-192. doi:10.1093/BIB/BBS017
26. Consortium TIP-CA of WG. Pan-cancer analysis of whole genomes. *Nature*. 2020;578(7793):82. doi:10.1038/S41586-020-1969-6
27. Hochmann Valls J, Cappetta Sapriza M, Pérez J, et al. Melanoma, ancestry and MC1R variations in the uruguayan admixed population. *Issue 1 | Article 1-research Journal of Basic & Applied Genetics*. 2016;27. Accessed July 28, 2021. <https://www.colibri.udelar.edu.uy/jspui/handle/20.500.12008/24784>
28. Tamariz Campillo AF, Corona Deschamps A, Mellado Orellana R, Díaz Greene EJ, Rodríguez Weber FL. Inhibidores de puntos de control inmunitario en estadios avanzados de melanoma. *Acta Médica Grupo Ángeles*. 2021;19(1):108-115. doi:10.35366/98579
29. Tagliabue E, Gandini S, Bellocco R, et al. MC1R variants as melanoma risk factors independent of at-risk phenotypic characteristics: a pooled analysis from the M-SKIP project. *Cancer management and research*. 2018;10:1143-1154. doi:10.2147/CMAR.S155283
30. Abdo JF, Sharma A, Sharma R. Role of Heredity in Melanoma Susceptibility: A Primer for the Practicing Surgeon. *Surgical Clinics of North America*. 2020;100(1):13-28. doi:10.1016/J.SUC.2019.09.006
31. Smit AK, Collazo-Roman M, Vadaparampil ST, et al. MC1R variants and associations with pigmentation characteristics and genetic ancestry in a Hispanic, predominately Puerto Rican, population. *Scientific Reports 2020 10:1*. 2020;10(1):1-9. doi:10.1038/s41598-020-64019-y
32. Person F, Wilczak W, Hube-Magg C, et al. Prevalence of  $\beta$ III-tubulin (TUBB3) expression in human normal tissues and cancers: <https://doi.org/10.1177/1010428317712166>. 2017;39(10):1-11. doi:10.1177/1010428317712166
33. Li J, Sun P, Huang T, He S, Li L, Xue G. Individualized chemotherapy guided by the expression of *ERCC1*, *RRM1*, *TUBB3*, *TYMS* and *TOP2A* genes versus classic chemotherapy in the treatment of breast cancer: A comparative effectiveness study. *Oncology Letters*. 2021;21(1):1-1. doi:10.3892/OL.2020.12282

34. Schliep KP. phangorn: phylogenetic analysis in R. *Bioinformatics*. 2011;27(4):592-593. doi:10.1093/BIOINFORMATICS/BTQ706
35. Barnes WM. PCR amplification of up to 35-kb DNA with high fidelity and high yield from lambda bacteriophage templates. *Proceedings of the National Academy of Sciences*. 1994;91(6):2216-2220. doi:10.1073/PNAS.91.6.2216
36. Wongsurawat T, Jenjaroenpun P, de Loose A, et al. A novel Cas9-targeted long-read assay for simultaneous detection of IDH1/2 mutations and clinically relevant MGMT methylation in fresh biopsies of diffuse glioma. *Acta Neuropathologica Communications* 2020 8:1. 2020;8(1):1-13. doi:10.1186/S40478-020-00963-0
37. Kovaka S, Fan Y, Ni B, Timp W, Schatz MC. Targeted nanopore sequencing by real-time mapping of raw electrical signal with UNCALLED. *Nature Biotechnology* 2020 39:4. 2020;39(4):431-441. doi:10.1038/s41587-020-0731-9
38. Wallace AD, Sasani TA, Swanier J, et al. CaBagE: A Cas9-based Background Elimination strategy for targeted, long-read DNA sequencing. *PLoS ONE*. 2021;16(4). doi:10.1371/JOURNAL.PONE.0241253
39. Höijer I, Johansson J, Gudmundsson S, et al. Amplification-free long-read sequencing reveals unforeseen CRISPR-Cas9 off-target activity. *Genome Biology*. 2020;21(1). doi:10.1186/S13059-020-02206-W
40. López-Girona E, Davy MW, Albert NW, et al. CRISPR-Cas9 enrichment and long read sequencing for fine mapping in plants. *Plant Methods*. 2020;16(1). doi:10.1186/S13007-020-00661-X
41. Vaillancourt B, Buell CR. High molecular weight DNA isolation method from diverse plant species for use with Oxford Nanopore sequencing. *bioRxiv*. Published online October 9, 2019:783159. doi:10.1101/783159



## Anexo

---

### 1. Soluciones de la extracción de ADN:

#### **Buffer de lisis para glóbulos rojos (10X):**

NH<sub>4</sub>Cl – 40.10gr

NaHCO<sub>3</sub> – 4.20gr

EDTA disódico – 1.85gr

Se disuelve en Agua destilada y se lleva a un volumen final de 500 ml.

La solución de trabajo es de 1X (1:10).

#### **Buffer de lisis para glóbulos blancos:**

	<u>Concentración inicial</u>	<u>Volumen Inicial</u>	<u>Concentracion final</u>
Tris HCl pH = 8	1M	5.0 ml	50 nM
EDTA pH = 8	0.5M	20 ml	100 mM
SDS	10%	5.0 ml	0.5%
Proteinasas K			0.6 mg/ml

Se enrasa con Agua destilada hasta un volumen final de 100 ml.

#### **Precipitación de proteínas:**

Acetato de Amonio 7.5 M en Agua destilada

#### **TE (Tris- EDTA):**

	<u>Concentración inicial</u>	<u>Volumen Inicial</u>	<u>Concentración final</u>
Tris pH = 8	1M	500 µl	10 mM
EDTA pH = 8		100 µl	1.0 mM

## 2. Pipeline del análisis bioinformático

El primer paso es instalar todos los softwares (Conda, Minimap2, Snakemake, Samtools)

Iniciar el ambiente Conda

```
source activate ont_tutorial_cas9  
R CMD javareconf > /dev/null 2>&1
```

*SCRIPT* ont\_tutorial\_cas9 :

Identificar el flujo de trabajo al que pertenece este archivo de configuración

Pipeline: “Nanopore Cas9 enrichment MC1R”

Define un nombre al flujo de trabajo, es una etiqueta que se usa para identificar el análisis y la etiqueta se usa para nombrar los archivos producidos durante el análisis

study\_name:”cas9\_FAL12180”

Se refiere al genoma contra el cual se mapeará la secuencia de lecturas.

reference\_genome: “/ReferenceData/Homo\_sapiens.GRCh38.dna.primary\_assambly.fa”

Apunta a un archivo en formato BED que describe las coordenadas genómicas para cada uno de los objetivos que se evalúan. El formato de archivo BED es un archivo delimitado por tabulaciones con columnas sin nombre que describen, en orden, el cromosoma, la posición inicial, la posición final y el nombre del objetivo.

target\_regions: “RawData/MC1R.bed”

Se agrega el archivo fastq para ser mapeado

fastq:”RawData/FAL12180/fastq\_pass/FAL12180\_pass\_46a826db\_0.fastq”

gstride es el tamaño del compartimento para la profundidad de cobertura contra el genoma de referencia

gstride:100

target\_proximity es la distancia arriba y abajo de on target BED para definir el mapeo del target proximal

target\_proximity: 30000

Se define el treshold para mapear los off target

offtarget\_level: 20

Se define los threads que se utilizaran en el proceso

threads: 8

tutorialText: TRUE

Luego se ejecuta el archivo de flujo de trabajo de snakemake

```
snakemake -j 4 all
```

El script Rmarkdown se puede ejecutar usando el cuadro de diálogo de tejido en el software Rstudio.

genomicranges filtrar todo lo que está en la secuencia on target y off target.

```
R --slave -e 'rmarkdown::render("ont_tutorial_cas9.Rmd",  
"html_document")'
```

Se llama al IGV para visualizar los reads

```
igv -g ./ReferenceData/Homo_sapiens.GRCh38.dna.chromosome.4.fa \  
./Analysis/Minimap2/cas9_FAK76554.bam,./RawData/enrichment_targets.bed  
\4:49091201-49156600
```