



UNIVERSITAT  
POLITÈCNICA  
DE VALÈNCIA



Escuela Técnica Superior  
de Ingeniería Agronómica y del Medio Natural

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

ESCOLA TÈCNICA SUPERIOR D'ENGINYERIA AGRONÒMICA I DEL MEDI NATURAL (ETSIAMN)

**Optimización en la preparación de librerías para  
secuenciación masiva mediante amplicones:  
Estudio de bases genéticas de diabetes tipo 2**

GRADO EN BIOTECNOLOGÍA

ALUMNO/A: Claudia Jannone Pedro

TUTOR/A: José Gadea Vacas

*Curso académico: 2018-2019*

*Valencia, 8 de julio del 2019*

## **RESUMEN**

### **OPTIMIZACIÓN EN LA PREPARACIÓN DE LIBRERÍAS PARA SECUENCIACIÓN MASIVA MEDIANTE AMPLICONES: ESTUDIO DE BASES GENÉTICAS DE DIABETES TIPO 2**

La diabetes mellitus tipo 2 es una patología multifactorial, debida tanto a factores ambientales como genéticos y a su interacción. Para establecer parte de la base genética de esta enfermedad, se analizan 9 genes (regiones exómicas, incluyendo regiones intrónicas contiguas, y promotoras) y 81 regiones que contienen polimorfismos de una sola base (PSBs). En ambos casos, tanto los genes como los PSBs analizados han sido asociados con DM2 por el grupo de investigación UGD. El estudio abarca a una población de 14.000 muestras y se realiza en base a dos paneles desarrollados con características diferentes debido a su diseño. Por ello, es necesario optimizar las condiciones de generación de librerías para que las coberturas sean homogéneas, e idealmente poder analizarlas conjuntamente. Tras el proceso de optimización con distintos aditivos y concentraciones de estos y variaciones en la amplificación, se observan condiciones bastante óptimas para el panel diseñado en este trabajo (Panel 1), ya que permite el estudio del 85% de los fragmentos, aunque no del 90% que era el objetivo inicial. En cuanto al panel 2, los resultados muestran que no es válido para la enzima en condiciones estándar ni con aditivos, de manera que habría que estudiar si es mejor rediseñar el segundo panel o seguir con el proceso de optimización con otra enzima o aditivos, por todo ello, el panel 1+2 tampoco muestra resultados satisfactorios, con lo que no se podría realizar el estudio conjuntamente.

## **PALABRAS CLAVE**

Secuenciación masiva, amplicones, librerías, optimización, coberturas

Claudia Jannone Pedro

*Curso académico: 2018-2019*

*Valencia, 8 de julio del 2019*

Tutor académico : Prof. D. José Gadea Vacas

Cotutor colaborador: Felipe Javier Chaves

## **ABSTRACT**

OPTIMIZATION OF LIBRARY PREPARATION FOR HIGH-THROUGHPUT SEQUENCING USING AMPLICONS: STUDY OF GENETIC BASES OF TYPE 2 DIABETES.

Type 2 Diabetes mellitus is a multifactor pathology, due to both environmental and genetic factors and their interaction. To establish part of the genetic basis of this disease, 9 genes (exomic regions, including contiguous intronic regions, and promoters) and 81 regions containing SNPs will be analyzed. In both cases, the analyzed genes and the SNPs have been associated with DM2 by GDU research group. The study includes a population of 14,000 samples and will be based on two panels developed with different characteristics due to their design. Therefore, it is necessary to optimize the conditions for library generation so that coverage is homogeneous, and ideally to be able to analyze them together. After the optimization process with different additives and concentrations as well as PCR variations, admissible conditions are observed for the panel designed in this project (Panel 1), since it allows the study of 85% of the fragments, even though is not 90% as the initial aim. Regarding panel 2, results show that is not valid for the enzyme in standard conditions neither with additives, so it should be studied if it is better to redesign the second panel or keep going with the optimization process with other enzymes or additives. Due to this, panel 1+2 does neither show good results, so the study could not be done together.

## **KEYWORDS**

High-throughput sequencing, amplicons, libraries, optimization, coverage

Claudia Jannone Pedro

*Curso académico: 2018-2019*

*Valencia, 8 de julio del 2019*

Tutor académico : Prof. D. José Gadea Vacas

Cotutor colaborador: Felipe Javier Chaves

## AGRADECIMIENTOS

Por esos buñuelos después de largas tardes de biblioteca...

A Javier por darme esta oportunidad y sus correos a hora intempestivas,

Al laboratorio 23-1, Chavalines, Azahara por su incommensurable paciencia

Y Alicia y Kike con sus *"runs"*.

## ÍNDICE

1. INTRODUCCIÓN .....	1
1.1. Diabetes .....	1
1.1.1. Diabetes mellitus tipo 2 .....	1
1.2. Tecnología NGS .....	2
2. OBJETIVOS .....	3
3. MATERIALES Y MÉTODOS .....	5
3.1. Diseño de cebadores.....	5
3.2. Obtención de muestras .....	7
3.3. Optimización PCR 1-Multiplex .....	7
3.4. Optimización PCR 2-Anidada .....	8
3.5. Electroforesis capilar.....	9
3.6. Purificación con Magsi-prep .....	9
3.7. Cuantificación de las muestras mediante Fluorescencia .....	10
3.8. Generación de la mezcla de librerías para realizar las pruebas de NGS .....	10
3.9. Cuantificación de las MLs mediante una curva patrón.....	10
3.10. Secuenciación .....	10
3.11. Análisis bioinformático .....	13
4. RESULTADOS Y DISCUSIÓN .....	15
4.1. Diseño de cebadores.....	15
4.2. Optimización PCR 1-multiplex .....	15
4.3. Optimización PCR 2-anidada.....	24
4.4. Secuenciación de las pruebas con resultado positivo.....	26
5. CONCLUSIONES.....	35
6. BIBLIOGRAFÍA .....	36
7. ANEXOS.....	38

## ÍNDICE DE FIGURAS

Figura 1. Número estimado de personas con diabetes en el mundo y por región en 2015 y 2040 (20-79 años). Imagen extraída de Diabetes Atlas de la FID - 8ª edición. ....	1
Figura 2. Preparación de librería para NGS Illumina. Imagen adaptada de ILLUMINA®. ....	3
Figura 3. Electroforesis capilar de la amplificación de muestras control con las condiciones iniciales, del panel 1 y 2 juntos, el panel 1, y el panel 2.....	8
Figura 4. Ejemplo resultado de electroforesis con el sistema QIAxcel. ....	9
Figura 5. Esquema de purificación con Magsi-Prep. Imagen extraída de Seda, 2018.....	10
Figura 6. Cálculo del peso molecular del amplicón.....	10
Figura 7. Cálculo de la concentración en moles/L (para el peso molecular se utiliza el tamaño medio de los fragmentos presentes en la librería, como se indica en la ecuación anterior).....	11
Figura 8. Concentración en nanomol/L.....	11
Figura 9. Cálculo del volumen necesario para conseguir una concentración equimolar, siendo la concentración final de 10 nM y el volumen final 100 µL. ....	11
Figura 10. Cálculo del número de lecturas totales.....	11
Figura 11. Cálculo del volumen a cargar en el secuenciador. ....	11
Figura 12. Esquema para la obtención el producto desnaturalizado y diluido a la concentración recomendable. 4A. Esquema de la librería de interés. 4B. Esquema del fago de control Phix...12	12
Figura 13. Diagrama de flujo de secuenciación con la plataforma Miseq Illumina. A. Preparación de librería B. Generación de colonias locales. C. Secuenciación D. Obtención de imágenes. ....	13
Figura 14. Ejemplo de la cobertura de un amplicón. ....	14
Figura 15. Ejemplo de cebadores diseñados para el gen <i>MANBA</i> . ....	15
Figura 16. Representación de los dos multiplexes A y B. A representación de lo que ocurriría si no se separaran los cebadores. B Diseño de multiplexes. ....	15
Figura 17. Resultados de la electroforesis para las condiciones de Aditivo 1 6mM y 9 mM para los paneles 1+2,1 y 2.A. Resultados para aditivo 1 6mM B. Resultados para aditivo 1 9mM.....	17
Figura 18. Resultados de la electroforesis para las condiciones de Aditivo 1 6mM y 9 mM para los paneles 1+2,1 y 2.A. Resultados para aditivo 1 6mM B. Resultados para aditivo 1 9mM.....	18
Figura 19. Resultados de la electroforesis para las condiciones de Aditivo 1 6mM y 9 mM para los paneles 1+2,1 y 2.A. Resultados para aditivo 1 6Mm y aditivo 2 5% B. Resultados para aditivo 1 9mM y aditivo 2 5%.....	18
Figura 20. Resultados de la electroforesis para las condiciones de Aditivo 2 2,5% y 5% para los paneles 1+2.1 y 2. A. Resultados para aditivo 2 2,5% B.Resultados para aditivo 2 5%.....	19
Figura 21. Comparación de la electroforesis de una condición (6mM) entre las condiciones 'estándar' de PCR (A) y la modificación (B). ....	19
Figura 22. Comparación de una condición (Aditivo 1 6mM ) para el multiplex del panel 1+2 con los ciclos de PCR 'estándar', la modificación de un paso a 55°C y la modificación a 50°C, en ese orden, siendo estas últimas dos pruebas con una proporción similar entre cebadores. ....	21
Figura 23. Electroforesis de resultados positivos para los paneles 1, 2 en la prueba 3.A. Resultados para el panel 1 B. Resultados para el panel 2.....	22
Figura 24. Electroforesis de resultados positivos para la librería de los paneles 1 +2 para proporciones 1:1 y 1:4 en la prueba 3. A. Resultados para los paneles 1+2 con proporción 1:1 B. Resultados para los paneles 1+2 con proporción 1:4. ....	23
Figura 25. Electroforesis de las mejores condiciones observadas en la prueba 4 para los paneles 1, 2, 1+2 con proporción 1:1 y 1+2 con proporción 1:4. A. Resultados par a la librería del panel 1. B. para la librería del panel 2. C. Para librería del panel 1+2 con proporción 1:1 de cebadores.	

D. Para librería del panel 1+2 con proporción 1:4 de cebadores. Nota: P: panel/es, A: aditivo, [] ceb: concentración de cebador. ....	24
Figura 26. Ejemplo de muestra en la electroforesis capilar con las condiciones iniciales, la modificación de 64°C y la modificación de 90s. Nota: C. : condiciones. ....	25
Figura 27. Ejemplo de recta de regresión para extrapolar los valores de la cuantificación. ....	26
Figura 28. Ejemplo de comparación electroforesis entre antes de la purificación y después en el multiplex A. ....	26
Figura 29. Número de fragmentos que pueden ser analizados en cada condición para cada cobertura en el panel 1. En la figura se indica el número de fragmentos que tienen una profundidad mayor de 20x. ....	28
Figura 30. Número de fragmentos que pueden ser analizados en cada condición para cada cobertura en el panel 2. En la figura se indica el número de fragmentos que tienen una profundidad mayor de 20x. ....	28
Figura 31. Número de fragmentos que pueden ser analizados en cada condición para cada cobertura en el grupo de paneles 1+2 la figura se indica el número de fragmentos que tienen una profundidad mayor de 20x. ....	29
Figura 32. Distribución de lecturas en los distintos amplicones de las mejores condiciones de generación de las librerías para el panel 1. ....	29
Figura 33. Distribución de lecturas en los distintos amplicones de las condiciones óptimas de secuenciación para las librerías obtenidas del panel 2. ....	30
Figura 34. Distribución de lecturas en los distintos amplicones de las condiciones óptimas de secuenciación para el panel 1+2. ....	30
Figura 35. Número de fragmentos que pueden ser analizados en cada condición para cada cobertura en la librería del panel 1. En la figura se indica el número de fragmentos que tienen una profundidad mayor de 20x. ....	31
Figura 36. Número de fragmentos que pueden ser analizados en cada condición para cada cobertura en la librería del panel 2. En la figura se indica el número de fragmentos que tienen una profundidad mayor de 20x. ....	31
Figura 37. Número de fragmentos que pueden ser analizados en cada condición para cada cobertura en el grupo del panel 1+2. En la figura se indica el número de fragmentos que tienen una profundidad mayor de 20x. ....	32
Figura 38. Distribución de lecturas en los distintos amplicones de las condiciones óptimas de secuenciación para el multiplex del panel 1. ....	33
Figura 39. Distribución de lecturas en los distintos amplicones de las condiciones óptimas de secuenciación para el multiplex del panel 2. ....	33
Figura 40. Distribución de lecturas en los distintos amplicones de las condiciones óptimas de secuenciación para el multiplex del panel 1+2. ....	34
Figura 41. Representación de la cobertura obtenida para cada porcentaje de bases por amplicón (se muestra cada uno por separado en base a media de cobertura normalizada por condición en gris y en magenta la media de todos) en el caso de 50.000 lecturas. A. En el caso del panel 1. B. En el caso del panel 2 C. En el caso del panel 1+2. En el eje X, se muestra la cobertura de secuenciación media para las muestras de interés (profundidad) y en el eje Y el porcentaje de bases del amplicón por encima o igual que la cobertura indicada en el eje X (amplitud). ....	34

## ÍNDICE DE TABLAS

Tabla 1. Funciones de los genes a estudiar en el presente trabajo de los paneles 1 y 2.....	6
Tabla 2. Condiciones de aditivos probadas en la prueba 1.....	16
Tabla 3. Condiciones A de amplificación para la prueba 1. ....	17
Tabla 4. Condiciones B de amplificación para la prueba 1. ....	17
Tabla 5. Condiciones de aditivos analizadas en la prueba 2.....	20
Tabla 6. Condiciones A de amplificación para la prueba 2. ....	20
Tabla 7. Condiciones B de amplificación para la prueba 2. ....	20
Tabla 8. Condiciones de aditivos probadas en la prueba 3.....	21
Tabla 9. Condiciones de amplificación para la prueba 3 y resto del proyecto. ....	22
Tabla 10. Condiciones de aditivos probadas en la prueba 4.....	23
Tabla 11. Condiciones de amplificación para la prueba 1. ....	24
Tabla 12. Condiciones de amplificación para la prueba 1. ....	25
Tabla 13. Ejemplo de uno de los documentos obtenidos tras la secuenciación de la librería correspondiente a los paneles 1+2.....	27



## **ABREVIATURAS Y ACRÓNIMOS**

**ADN:** Ácido desoxirribonucleico.

**ARN:** Ácido ribonucleico.

**ARNm:** ARN mensajero.

**DM:** Diabetes mellitus.

**DM2:** Diabetes mellitus tipo 2.

**dNTP:** Desoxirribonucleótido trifosfato.

**ddNTP:** Didesoxirribonucleótido trifosfato.

**GWAS:** Del inglés, estudios de Asociación de Genoma Completo.

**MAF:** Del inglés, frecuencia del alelo menos común.

**NGS:** Del inglés, secuenciación de nueva generación.

**Pb:** Pares de bases.

**PCR:** Del inglés, reacción en cadena de la polimerasa.

**PSB:** Polimorfismo de una sola base.

**Tm:** Del inglés, temperatura de fusión.

# 1. INTRODUCCIÓN

## 1.1. Diabetes

La diabetes mellitus (DM) representa un grupo heterogéneo de desórdenes metabólicos que comparten ciertas características, con la hiperglucemia como elemento común y resultante de la falta y/o alteración de la secreción de insulina, de su acción o de ambas.

La DM está incrementándose a nivel mundial, alcanzando proporciones epidémicas. La prevalencia en adultos en el mundo era un 4% en 1995 y se estima un crecimiento hasta un 5,4% para el 2025. En término de pacientes, había unos 135 millones en 1995 y se estima que será de unos 300 millones por el año 2025 y 642 millones en 2040. Sin embargo, cálculos adicionales muestran que no habría un incremento homogéneo en todo el mundo; siendo del 42% (de 51 a 72 millones) en los países desarrollados y del 170% (de 84 a 228 millones) en los países en desarrollo. (Figura 1) (7ª edición del Diabetes Atlas de la FID, 2015 ; De Faria Maraschin, 2012).

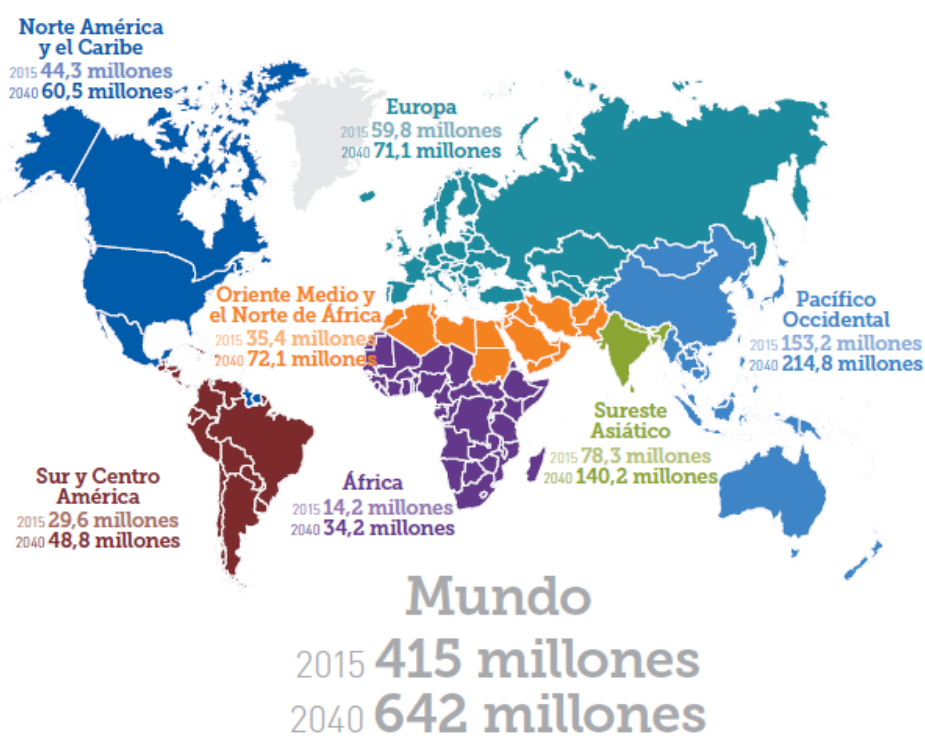


Figura 1. Número estimado de personas con diabetes en el mundo y por región en 2015 y 2040 (20-79 años). Imagen extraída de Diabetes Atlas de la FID, 2015.

### 1.1.1. Diabetes mellitus tipo 2

La diabetes mellitus tipo 2 (DM2) es el tipo más común de diabetes, correspondiendo al 85-95% de todos los casos en el mundo y la que mayor aumento tendrá los próximos años debido a los cambios de hábitos que se están produciendo a nivel mundial (nutrición, nivel de ejercicio, protección frente al medio, contaminación, etc.). Esta, se caracteriza por un aumento de la hiperinsulinemia, la

resistencia a la insulina y el fallo de las células  $\beta$  pancreáticas, con una pérdida celular de hasta el 50% en el momento del diagnóstico.

La DM2 es más común después de los 40 años de edad, pero debido al crecimiento global de obesidad y vida sedentaria se están incrementando también los casos entre jóvenes adultos y adolescentes.

Los individuos afectados, tienen un alto riesgo de sufrir complicaciones microvasculares (incluyendo retinopatía, nefropatía y neuropatía) y macrovasculares (como comorbilidades cardiovasculares). Esto se debe a la hiperglucemia y componentes individuales del síndrome de resistencia a la insulina, lo que también aumenta el riesgo de muerte (DeFronzo *et al.*, 2015 ; De Faria Maraschin, 2012 ; Chatterjee *et al.*, 2017).

La causa específica de este tipo de diabetes no se conoce, pero se sabe que es el resultado de la interacción entre numerosos factores genéticos y ambientales, dejando un fenotipo muy heterogéneo (De Faria Maraschin, 2012).

Los estudios de asociación de genoma completo, en inglés "*Genome wide association studies*" (GWAS), han permitido la identificación de variantes comunes asociadas a aspectos relacionados con el metabolismo glucídico y DM2. Éstas solo representan el 10% de la varianza total de los rasgos, lo que podría implicar que las variantes raras tuvieran un papel relevante en el componente genético de los rasgos relacionados con la DM2 (Chatterjee *et al.*, 2017). Estos estudios se basan en la teoría denominada de "enfermedades comunes, polimorfismos comunes" que indicaría que una enfermedad frecuente estaría producida por variantes genéticas comunes (alelo de menor frecuencia o MAF>0.05). Los resultados obtenidos hasta ahora en los estudios GWAS indican que hay una gran parte del componente genético calculado que no se ha encontrado. Por lo que estos resultados han hecho que se tome más en consideración la teoría de "enfermedades comunes, polimorfismos raros" lo que sugiere que las variantes raras pueden explicar una parte importante del componente genético de estas enfermedades comunes.

Dentro de este ámbito de estudio se enmarca el presente proyecto, dado que se está intentando localizar variantes poco frecuentes en diferentes genes previamente asociados con DM2 mediante estudios de exoma en el grupo de investigación que puedan explicar una parte de esta enfermedad.

## 1.2. Tecnología NGS

En 1977 se introdujo el método de secuenciación Sanger, que supuso un gran avance respecto a métodos anteriores que implicaban el uso de agentes tóxicos o radioactivos. Esta metodología de secuenciación permite secuenciar en una reacción fragmentos de ADN relativamente largos (500-1000 pb) y tiene una baja tasa de error, por lo que se le ha considerado el "*gold standard*" de la secuenciación.

Con la aparición de la tecnología de pirosecuenciación de 454 Life Science en 2005, empezó la revolución de las tecnologías de secuenciación de nueva generación (NGS). Este hecho ha cambiado radicalmente el conocimiento y la comprensión sobre el genoma.

La mayoría de los métodos e instrumentos de secuenciación de nueva generación comparten muchos aspectos con la tecnología Sanger, la diferencia principal es que no separan la incorporación de nucleótidos a la cadena (realizada enzimáticamente) de la obtención de los datos.

Estos instrumentos realizan ambos pasos de manera secuencial, lo que permite que los datos se generen de muchas secuencias a la vez, de ello que se conozca como secuenciación ‘masiva’ (McCombie *et al.*, 2018a ; Besser *et al.*, 2018).

A pesar de que existen distintas plataformas, el procedimiento hasta la secuenciación comparte varios pasos: (i) Extracción de ADN, (ii) preparación de la librería que supone la fragmentación del ADN o la amplificación de los fragmentos de interés, como en este caso, agregando adaptadores (necesarios para que se produzcan las reacciones de secuenciación) y etiquetas moleculares (necesarios para identificar la muestra a la que pertenecen las secuencias generadas a partir de cada fragmento) (Fig. 2) (Besser *et al.*, 2018).

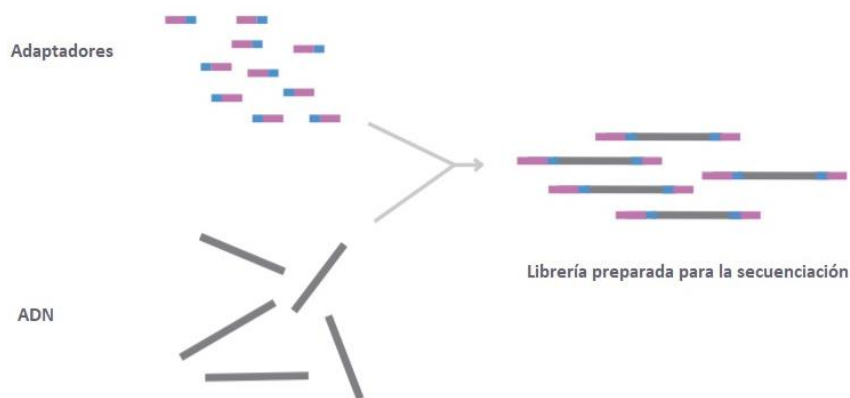


Figura 2. Preparación de librería para NGS Illumina. Imagen adaptada de ILLUMINA®.

## 2. OBJETIVOS

El grupo de investigación Unidad de Genómica y Diabetes (UGD) ha realizado varios proyectos para describir la genética de la enfermedad. Para ello, ha realizado en primer lugar un estudio de exoma de casos y controles, que incluye las regiones incluidas en los exomas (aquellas incluidas en los ARNs del genoma) y regiones intrónicas contiguas. El estudio se centra en el exoma dado que en él se encuentran el 90-95% de las alteraciones genéticas identificadas que son causantes en enfermedades humanas hereditarias (Botstein y Risch, 2003 ; Stenson *et al.*, 2003). A diferencia con el GWAS, el estudio de exoma sí que permite detectar polimorfismos poco frecuentes, los cuales son muy numerosos en el genoma y representan aproximadamente el 50% de los polimorfismos totales (Bodmer y Bonilla, 2008 ; So *et al.*, 2011 ; Fu *et al.*, 2013 ; Tennessen *et al.*, 2012).

Este estudio permitió identificar más de 5.000 variantes potencialmente asociadas con DM2 (Galan-Chilet, 2016), de ellas se seleccionaron 150 (MAF < 0.01) en 140 genes que por su impacto funcional pueden tener un efecto importante en la modulación del riesgo de desarrollar DM2 en obesidad moderada, muchos de los cuales no se habían descrito como implicados en la DM2. A nivel biológico estos genes se relacionan con procesos que pueden favorecer o impedir el desarrollo de la DM2 como la transcripción, funcionamiento el sistema de Golgi, modificación del metabolismo hidrocarbonado, excitosis, etc. (Galan-Chilet, 2016).

Dado que el estudio inicial es en una muestra limitada, el segundo paso del estudio se centra en la verificación de estas asociaciones mediante el estudio de los genes completos en 14.000 muestras procedentes de estudios caso/control para DM2 y estudios poblacionales que se han realizado entorno a factores de riesgo cardiovascular y que incluyen el diagnóstico de la DM2. Para ello, la UGD seleccionó como método para el análisis el estudio mediante librerías de amplicones debido al alto presupuesto que supondrían los kits comerciales para realizar el estudio mediante librerías de enriquecimiento.

Por todo lo anterior, **el objetivo principal** del presente trabajo es mejorar el rendimiento de la generación de librerías de NGS mediante amplicones previo al estudio de las muestras mencionadas.

Con esta finalidad se plantean los siguientes objetivos parciales:

- Diseñar cebadores propios para un panel (Panel 1) que incluye los genes *MANBA* y *TCF7L2* y polimorfismos de una sola base, PSBs que será el central para optimizar un nuevo procedimiento para reacciones de secuenciación en 150 pb en ambos sentidos.
- Utilizar un panel diseñado previamente (Panel 2) para condiciones de otras mezclas de reacción y reacciones de secuenciación de lecturas de 250 bp en ambos sentidos) con unos parámetros de diseño específico (diferentes a las usadas para el panel principal del presente estudio).
- A partir de la metodología publicada para generar librerías por amplicones y la experiencia previa obtenida, plantear un protocolo que permita obtener librerías de amplicones.
- Optimizar el protocolo mediante la comprobación de los resultados obtenidos con dos aditivos.
- Verificar si el cambio de las condiciones de reacción permite mejorar los resultados en los dos grupos de amplicones por separado y conjuntamente.
- Seleccionar una condición, por lo menos, para hacer el estudio posterior previsto.

### 3. MATERIALES Y MÉTODOS

#### 3.1. Diseño de cebadores

Para diseñar los cebadores de los genes *MANBA* y *TCF7L2* y los PSBs incluidos en el Panel 1 (Tabla 1) se hizo:

- En el caso de los genes, se identificaron los transcritos de referencia en base a los datos de las páginas COSMIC (<https://cancer.sanger.ac.uk/cosmic>) y ensembl (<http://www.ensembl.org/index.html>), se obtuvo la secuencia genómica y en la versión del genoma GRCh38, visualizando aquellas variantes con una MAF < 0.01 para evitar diseñar oligonucleótidos que incluyeran estas variantes.
- Para el estudio de PSBs: se identificaron los polimorfismos y las secuencias flanqueantes (150 bp); igualmente se visualizaron los polimorfismos presentes en estas regiones.

El primer programa de elección fue el “ExonPrimer” (<https://ihg.helmholtz-muenchen.de/ihg/ExonPrimer.html>) ya que permite el diseño de los cebadores de todos los exones de un gen de forma simultánea. Para ello se estableció como criterios 5 pb de mínima distancia entre el cebador y el límite del intrón o exón, para un máximo de 150 pb de región diana para que después fuera visible toda la región en la secuenciación con una superposición de 50 pb, una temperatura de anillamiento de 60°C y un mínimo de 18 pb y máximo de 32 pb entre otros. Para aquellas regiones en las que no se obtuvo resultados y para los PSB, se utilizó el programa “Primer3” (<http://bioinfo.ut.ee/primer3-0.4.0/>), estableciendo los mismos criterios.

Para comprobar que los cebadores obtenidos eran específicos, se hicieron una serie de pruebas informáticas posteriormente, con los programas de PCR *in silico* (<https://genome.ucsc.edu/cgi-bin/hgPcr>), el cual evalúa las posiciones del genoma en las que ensambla una pareja de cebadores y BlastN ([https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastnyPAGE\\_TYPE=BlastSearchyLINK\\_LOC=blasthome](https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastnyPAGE_TYPE=BlastSearchyLINK_LOC=blasthome)) que es similar al anterior, pero funciona con cada cebador individualmente.

Por último, se analizaron todos los cebadores con los adaptadores universales con el programa FastPCR para comprobar las interacciones fuertes que ocurrían en la región terminal 3' con ellos mismos y con todos los demás para en ese caso rediseñarlos o separarlos en los multiplexes. Como interacciones fuertes se consideraron aquellas que tenían una Tm superior a 20 °C. Se incluyeron los adaptadores universales dado que se incluyen en la región 5' de todos los oligonucleótidos debido a que son necesarios para la formación de colonias locales de ADN o “clusters” que posteriormente se secuenciarán.

A partir de la información obtenida del último programa y teniendo en cuenta que no se pueden juntar cebadores cuyos amplicones solapan, se dividieron en dos reacciones (multiplex A y B, panel 1) para evitar solapamientos de cebadores en una reacción que pudiera impedir la amplificación del fragmento de interés. En el caso del panel 2 (Tabla 1), los diseños de oligonucleótidos se realizaron previamente mediante otras condiciones de diseño y las multiplexes se realizaron de forma similar.

**Tabla 1.** Funciones de los genes a estudiar en el presente trabajo de los paneles 1 y 2.

<b>Genes estudiados completos</b>	<b>Función del gen</b>
<i>MANBA</i> (panel 1, 37 fragmentos)	La enzima $\beta$ -manosidasa participa en el último paso de descomposición de los oligosacáridos unidos a proteínas en los lisosomas.
<i>TCF7L2</i> (panel 1, 25 fragmentos)	Está implicado en la homeostasis de la glucosa en sangre.
<i>ACAD10</i> (panel 2, 29 fragmentos)	Codifica un miembro de la familia de enzimas acil-CoA deshidrogenasas (ACAD) que participan en la $\beta$ -oxidación de ácidos grasos en las mitocondrias.
<i>ACSM3</i> (Panel 2, 21 fragmentos)	Proteínas relacionadas con el metabolismo y $\beta$ -oxidación de ácidos grasos (peroxisoma).
<i>SREBF2</i> (panel 2, 34 fragmentos)	Codifica el del factor de transcripción SREBP2 expresado de forma ubicua que controla la homeostasis del colesterol y el metabolismo glucídico.
<i>CYP1A1</i> (panel 2, 12 fragmentos)	Miembro de la superfamilia de enzimas del citocromo P450, las cuales catalizan muchas reacciones involucradas en el metabolismo de fármacos y la síntesis de colesterol, esteroides y otros lípidos.
<i>EBLN2</i> (panel 2, 14 fragmentos)	Puede actuar como una proteína de unión a ARN.
<i>GCKR</i> (panel 2, 19 fragmentos)	Proteína que pertenece a la subfamilia de GCKR de la familia de proteínas azúcar isomerasas, inhibe la glucoquinasa en las células de los islotes pancreáticos y del hígado al unirse para formar un complejo inactivo con la enzima.
<i>HABP2</i> (panel 2, 17 fragmentos )	Miembro de la familia peptidasa S1 de serinproteasas, puede desempeñar un papel en los sistemas de coagulación y fibrinólisis.
<b>Polimorfismos de una sola base (PSB)</b> (panel 1, 81 fragmentos)	<b>Función del gen</b>
<i>BDH2</i>	Proteína relacionada con el metabolismo y el metabolismo corporal de la cetona.
<i>CENPE</i>	Proteína E asociada a centrosoma, necesaria para el alineamiento de los cromosomas durante la prometafase.

<i>NFKB1</i>	Proteína que es una subunidad de unión al ADN del complejo de proteínas NF-kappa-B (NFKB).
<i>SLC9B1</i>	Intercambiador de sodio / hidrógeno y una proteína transmembrana.
<i>SLC9B2</i>	Proteína relacionada con el transporte de glucosa y otros azúcares, sales biliares y ácidos orgánicos, iones metálicos y compuestos de aminas y transporte de canales iónicos.
<i>UBE2D3</i>	Miembro de la familia de enzimas conjugadoras de ubiquitina E2.

### 3.2. Obtención de muestras

Se utilizaron muestras de sangre de las cuales se extrajo el ADN genómico mediante el sistema Maxwell® (Promega®, USA) y se dejaron a una concentración de 10 ng/μL.

### 3.3. Optimización PCR 1-Multiplex

Se realizó de manera separada con los dos multiplexes A y B para amplificar los fragmentos diana, con los cebadores diseñados según se ha indicado anteriormente con el genoma GRCh38 (llamados a partir de ahora panel 1) formados por 73 pares de cebadores para el multiplex A y 70 pares de cebadores para el multiplex B, y los obtenidos en una fase anterior del estudio con el genoma GRCh37 (llamados a partir de ahora panel 2) formados por 76 pares de cebadores para el multiplex A y 63 para el multiplex B.

Debido a que no amplificaban estos paneles con las condiciones “estándar” de la Unidad de Genómica y Diabetes basadas en experiencia previa (Fig. 3), se probaron distintas condiciones de concentración de cebadores, aditivos, condiciones de PCR (tanto para la PCR1 como la PCR2), para determinar las mejores condiciones para obtener resultados en el mayor número de fragmentos y con las coberturas más homogéneas en la secuenciación. La falta de amplificación de estos paneles, tanto el 1 como el 2, seguramente se debe a las diferentes condiciones de uso de programas bioinformáticos de diseño, especialmente respecto el tamaño, que supusieron que los resultados obtenidos (los oligonucleótidos diseñados) sean muy diferentes. En general con el protocolo que se indica a continuación se obtiene (bajo las condiciones de diseño usadas como estándar en el UGD) la cobertura de un 85-90% de las regiones incluidas en paneles similares e incluyendo fragmentos de tamaño variado (100 bp a 250 bp).



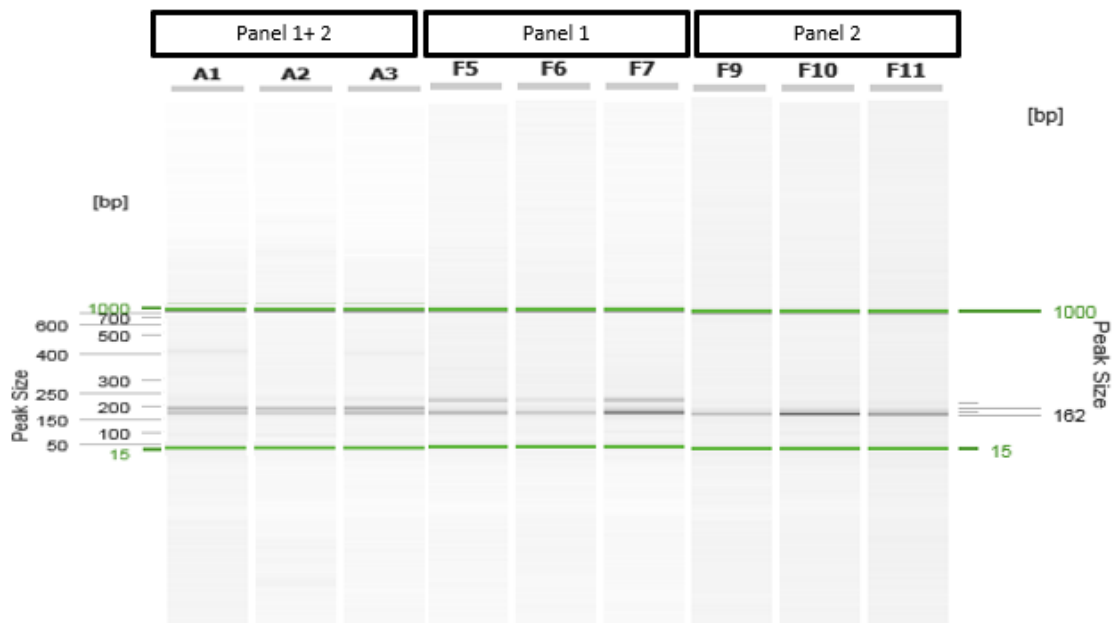


Figura 3. Electroforesis capilar de la amplificación de muestras control con las condiciones iniciales, del panel 1 y 2 juntos, el panel 1, y el panel 2.

Para resolver el problema, además de utilizar la mezcla de ambos grupos de cebadores, se utilizaron también por separado dado que, si no se obtenían unas buenas condiciones para un estudio conjunto, el posterior estudio se realizaría con ambos grupos por separado.

Para el proceso de amplificación se utilizaron termocicladores de placas 384 Veriti® (Applied Biosystems®, California).

El procedimiento general consistió en la preparación de *mixes* de reacción diferentes, variando las concentraciones de cebadores de los aditivos 1 y 2 para la PCR 1 y utilizando estas condiciones con cambios en las condiciones de reacción (temperaturas y tiempos).

### 3.4. Optimización PCR 2-Anidada

Se realizó a partir de los productos amplificados por la primera PCR. Durante este proceso se añaden adaptadores, que son oligonucleótidos que contienen la secuencia de la etiqueta molecular o “barcode” (uno para el 3’ y otro para el 5’) además de otras secuencias necesarias para que sea compatible con la tecnología Illumina y permite la amplificación específica del producto de la PCR1.

En este caso se optimizaron los ciclos de PCR2, probando distintos tiempos y temperaturas en las fases de unión y extensión.

### 3.5. Electroforesis capilar

Se realizó para comprobar si se había producido la amplificación y seleccionar aquellas que presentaban el perfil electroforético adecuado (contenían los fragmentos esperados y ausencia de dímero de cebador o fragmentos de tamaños diferentes a los deseados). Para ello se utilizó el sistema QIAxcel (QIAxcel DNA Screening Kit, Qiagen®) en condiciones estándar utilizando el cartucho de "Screening" (Fig 4). Este paso nos permitió eliminar de las reacciones de secuenciación todas aquellas condiciones que no fueran apropiadas.

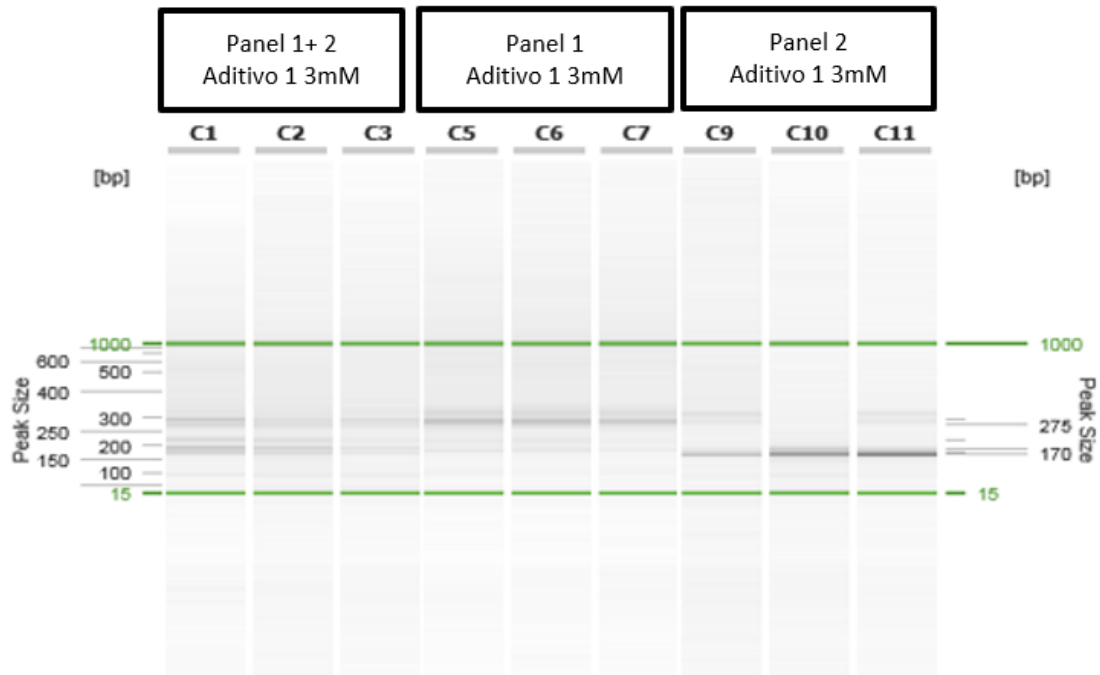


Figura 4. Ejemplo resultado de electroforesis con el sistema QIAxcel.

Siguiendo los resultados que salieron en la electroforesis capilar, se escogieron las condiciones que mejores resultados daban para secuenciar y, tras realizar la secuenciación de cada producto de reacción, decidir en un paso posterior en base a las coberturas de cada fragmento.

### 3.6. Purificación con Magsi-prep

Para purificar las librerías obtenidas en las reacciones de PCR (eliminar sales, dNTPS, enzimas, oligonucleótidos y dímeros) y obtener los fragmentos amplificados, se hizo uso del kit de purificación Magsi-NGS PREP®(MagnaMedics Diagnostics B.V., Países Bajos) basado en un sistema de esferas magnéticas. Estas esferas cuentan con unos terminadores con carga positiva en su superficie, lo que hace que tengan afinidad por las cargas negativas de los ácidos nucleicos. Cuanto más larga sea la cadena, más cargas negativas tiene y más afinidad tiene por las esferas, de manera que las moléculas largas son las que se quedan adheridas a las esferas, y las cortas, las que no interesan, se quedarán en el eluyente. Con distintos lavados con etanol se van eliminando las moléculas que se unen con menor fuerza. Finalmente, el ADN se eluye con agua, ya que compete con las esferas por el DNA; lo que produce que se liberen de las esferas y permanezcan en el agua (Fig. 5).



Figura 5. Esquema de purificación con Magsi-Prep. Imagen extraída de Seda, 2018.

### 3.7. Cuantificación de las muestras mediante Fluorescencia

Para ello se hizo uso del kit QuantiFluor® dsDNA(Promega, USA) y el fluorímetro GloMax®-Multi+ (Promega®, USA), siguiendo el procedimiento estándar que incluye una curva patrón con un ADN de fago lambda para poder extrapolar la concentración del ADN de interés a partir de la fluorescencia.

### 3.8. Generación de la mezcla de librerías para realizar las pruebas de NGS

Se escogieron aquellas condiciones cuyos resultados eran favorables en un tubo eppendorf para su cuantificación y posterior secuenciación. Este paso es posible debido a que cada muestra cuenta con un código de barras específico que permite posteriormente identificar cada muestra por separado.

Se procedió de forma similar para generar las mezclas de librerías (MLs) en las 3 reacciones de NGS con el sistema Miseq® (Illumina®, California) que se han realizado. En todos los casos, se calculó la concentración de cada librería y luego se mezclaron de forma que todas tuvieran una concentración equimolar en base a la concentración, tamaño medio de los fragmentos y número total.

### 3.9. Cuantificación de las MLs mediante una curva patrón

Finalmente, se cuantificó la librería para verificar su concentración y añadir la cantidad adecuada.

### 3.10. Secuenciación

Para comprobar las condiciones que proporcionaban una cobertura más homogénea, se utilizó el secuenciador Miseq de la Unidad de Genotipado y Diabetes en INCLIVA.

Para ello, se calculó una cobertura de 1000 lecturas de cada base por fragmento.

Las librerías que iban a ser secuenciadas se prepararon a una concentración de 10 nM para evitar sesgos en el número de lecturas, teniendo en cuenta el número y longitud de amplicones, la concentración y el peso molecular del ADN, calculando para obtener unos 100 µL finales (Figuras 6,7,8,9).

$$\text{Peso molecular (g/mol)} = \text{Peso par de nucleótido} \left( \frac{\text{g}}{\text{mol} * \text{pb}} \right) * \text{Longitud amplicón(pb)}$$

Figura 6. Cálculo del peso molecular del amplicón.

$$\text{Concentración} \left( \frac{\text{moles}}{\text{L}} \right) = \frac{\text{Concentración} \left( \frac{\text{g}}{\text{L}} \right)}{\text{peso molecular} \left( \frac{\text{g}}{\text{mol}} \right)}$$

**Figura 7.** Cálculo de la concentración en moles/L (para el peso molecular se utiliza el tamaño medio de los fragmentos presentes en la librería, como se indica en la ecuación anterior).

$$\text{Concentración} \left( \frac{\text{nanomoles}}{\text{L}} \right) = \text{Concentración} \left( \frac{\text{moles}}{\text{L}} \right) * 10^9 \left( \frac{\text{nanomoles}}{\text{mol}} \right)$$

**Figura 8.** Concentración en nanomol/L.

$$\text{Volumen necesario} (\mu\text{L}) = \frac{\text{Concentración final}(\text{nanomol/L}) * \text{Volumen final}(\mu\text{L})}{\text{Concentración inicial} (\text{nanomol/L})}$$

**Figura 9.** Cálculo del volumen necesario para conseguir una concentración equimolar, siendo la concentración final de 10 nM y el volumen final 100  $\mu\text{L}$ .

Posteriormente, se calcula el volumen de librería que se ha de secuenciar, mirando el número de lecturas totales necesarias que varía según el número de muestras y cuántas lecturas por fragmento se desea (Figuras 10,11).

*Número de lecturas totales(millones de lecturas) =*

*Número de fragmentos \* número de lecturas/fragmento \* número de individuos*

**Figura 10.** Cálculo del número de lecturas totales.

$$\text{Volumen a cargar en el secuenciador}(\mu\text{L}) = \frac{\text{Número de lecturas totales} * 100}{22 (\text{millones de lecturas})}$$

**Figura 11.** Cálculo del volumen a cargar en el secuenciador.

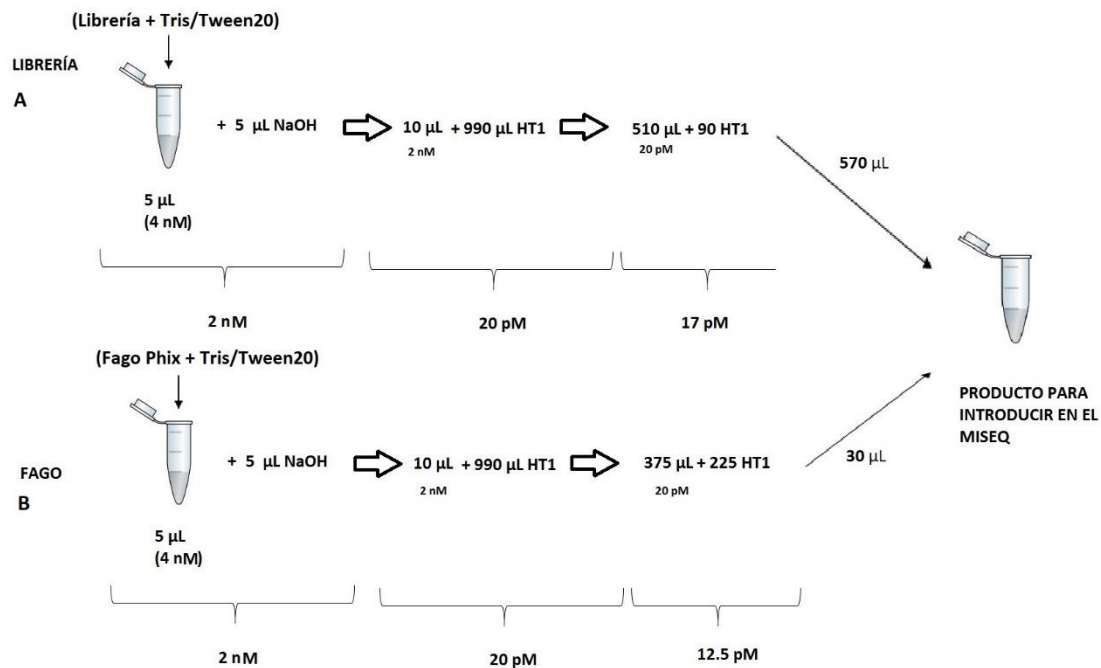
Para realizar la reacción de secuenciación:

Se utilizó en todos los casos kits de Miseq V2 (Illumina®, California) de 300bp que recomienda cargar 600  $\mu\text{L}$  librería con una concentración de 6-20 pM, requiriendo al menos una concentración de 4nM de la librería antes de diluir y desnaturalizar.

En primer lugar, se diluyó la librería de 10nM a 4nM con Tris/Tween20, se desnaturalizó con NaOH y se añadió el tampón de hibridación (HT1) hasta obtener 1mL a una concentración de 20 pM. Después se volvió a diluir en HT1, obteniendo una concentración final de 17 pM (Fig. 12A).

Además, se utilizó una librería del fago Phix de Illumina como control de secuencia. Esta biblioteca, proporciona un control de calidad para la generación de “clusters”, la secuenciación o el alineamiento. Puede ser alineado rápidamente para estimar la tasa de error entre otros. Este fago, también fue diluido y desnaturalizado como se ha indicado anteriormente en la librería de interés, obteniendo una concentración final de 12.5 pM (Fig. 12B).

Finalmente, se mezclan 570  $\mu\text{L}$  de librería con 30  $\mu\text{L}$  de fago para obtener los 600  $\mu\text{L}$  recomendados para cargar en el secuenciador y utilizar el 5% de fago en la reacción.



**Figura 12.** Esquema para la obtención el producto desnaturalizado y diluido a la concentración recomendable. **A.** Esquema de la librería de interés. **B.** Esquema del fago de control Phix.

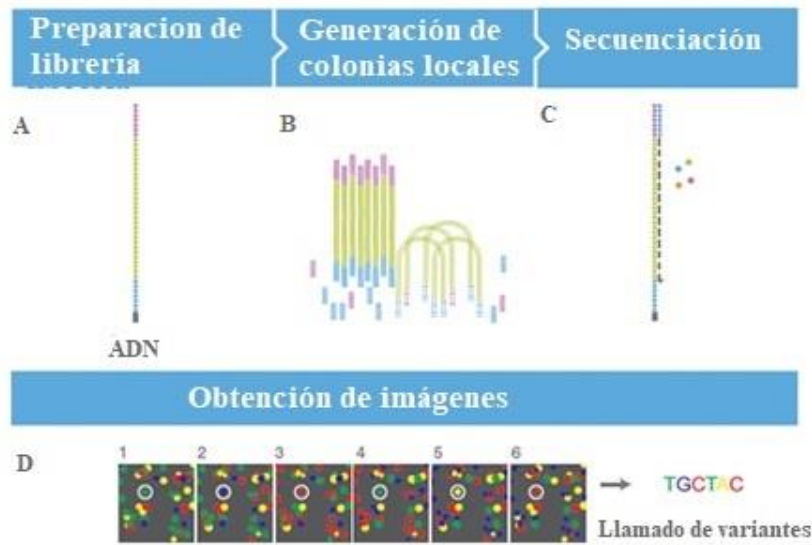
En este caso, al tratarse de la plataforma Miseq de Illumina, todo el proceso de secuenciación concurre en una celda de flujo de un carril. La plataforma de secuenciación de esta casa comercial utiliza la amplificación por puente para la generación de colonias locales de ADN o “clusters”, y sigue una secuenciación por síntesis.

El primer paso para llevar a cabo el proceso consiste en la desnaturalización de los fragmentos de DNA de doble cadena, para que, en la celda de flujo, se pueda hibridar los adaptadores añadidos en la PCR2 (Fig. 13A) a los oligonucleótidos fijos en la superficie, los cuales son utilizados como cebadores para formar una copia inicial de la molécula molde.

Las hebras originales se descartan, y la copiada, unida a la celda de flujo, se utiliza para generar colonias de moléculas de DNA idénticas, utilizando amplificación isotérmica durante la cual, los extremos 3' de las hebras copiadas pueden hibridar a los cebadores complementarios en la celda de flujo, inclinándose, formando una estructura de puente (Fig. 13B). El paso final es eliminar una de las dos hebras y bloquear todos los extremos 3' con ddNTPs para que no actúen como cebadores de secuenciación.

Durante la secuenciación, se incorpora un dNTP marcado con un fluoróforo cada vez, durante ciclos repetitivos. Cada uno de los distintos dNTPs cuenta con un fluoróforo distinto, uno para cada nucleótido (A, C, G, T) que sirve para su posterior identificación, y actúa como un terminator reversible para evitar la incorporación de más dNTPs (Fig. 13C).

Después de captar la imagen, el grupo fluorescente se elimina, el terminador reversible se desactiva y la hebra puede, de nuevo, incorporar un dNTPs. La secuencia se lee siguiendo la señal fluorescente de la extensión en cada paso en cada colonia (Fig. 13D) (Buermans y den Dunnen, 2014).



**Figura 13.** Diagrama de flujo de secuenciación con la plataforma Miseq Illumina. **A.** Preparación de librería **B.** Generación de colonias locales. **C.** Secuenciación **D.** Obtención de imágenes.

### 3.11. Análisis bioinformático

Se realizó un paso de preprocesado del archivo FASTQ obtenido tras la secuenciación, en el que se eliminaron los nucleótidos de baja calidad en los extremos y posteriormente se mapearon con el genoma de referencia humano para obtener un archivo en formato BAM a partir del cual se pudo realizar el cálculo de cobertura por base. Estas secuencias se analizaron de tal forma que permitía saber la profundidad o veces que cada fragmento se había secuenciado. Además, se obtuvo un archivo de control de calidad por cada secuenciación con el número de lecturas total por muestra y la calidad de la secuenciación. Se descartaron así muestras que tenían una muy baja calidad de secuenciación o muy poco número de lecturas (16 en el primero, 18 en el segundo, y 4 en el tercero).

Por otra parte, para el multiplex A se obtuvieron únicamente dos archivos con los resultados, uno por cada *mix* de librerías (paneles 1 y 2 juntos y separados), uno donde indica la cobertura mínima y máxima, y otro donde muestra el porcentaje del amplicón que está por encima del límite establecido (Fig. 14 ), que en un primer momento se estableció en 100x y la mínima y máxima cobertura conseguida.

A partir de estos datos, por una parte, se estudió qué condición era la mejor para utilizarse posteriormente con las muestras de varios estudios:

Para ello, para cada condición (muestra en el caso de haber fallado una de las muestras o media de muestras) se normalizaron las lecturas de cada amplicón, dividiéndolo las lecturas de cada amplicón individual entre la suma total de todas las lecturas y se multiplicó el número total de lecturas para cada prueba de generación de librerías cuyos resultados planteaban: 10000, 20000, 30000, 50000, y 100000 en el caso de la librería de los paneles 1 + 2 por el número de amplicones, contabilizándose

únicamente aquellos amplicones con una cobertura superior a 20 lecturas. Esta normalización permite una mejor comparación de los resultados dado que cada librería obtiene un número diferente de lecturas.

Por otra parte, se decidió cuáles eran las mejores condiciones, siguiendo:

Para una menor cobertura cuáles eran las condiciones que cubrían mayor número de amplicones, y mayores lecturas en cada amplicón.

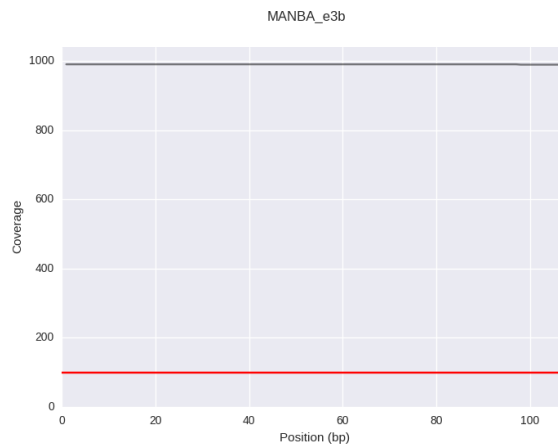


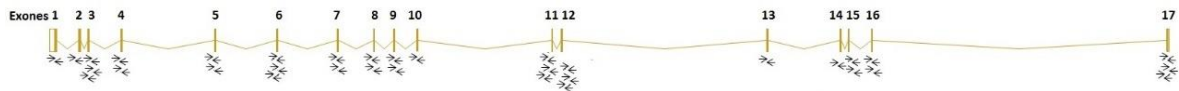
Figura 14. Ejemplo de la cobertura de un amplicón.

## 4. RESULTADOS Y DISCUSIÓN

### 4.1. Diseño de cebadores

El diseño de cebadores se realizó con el programa “ExonPrimer” como primera opción, que genera cebadores para todos los exones en un mismo proceso, y si no ofrecía resultados favorables, con el programa “Primer 3”, si bien en aquellas regiones que no podían ser analizadas solo mediante un amplicón se cubrieron mediante fragmentos solapantes (Fig. 15).

Figura 15. Ejemplo de cebadores diseñados para el gen *MANBA*.



En total, para el panel 1, se han obtenido 143 pares de cebadores, que cubren 2 genes completos: *MANBA* y *TCF7L2*, y PSBs en 6 genes más (Anexo I).

Dos amplicones que solapan no pueden ir en el mismo multiplex dado que se amplificaría la región solapante dando fragmentos de pequeño tamaño (Fig. 16A). Por lo tanto, estos fragmentos se deben amplificar en reacciones por separado pudiéndose utilizar 2 reacciones para amplificar todos los amplicones sin que haya solapamientos (Fig. 16B).

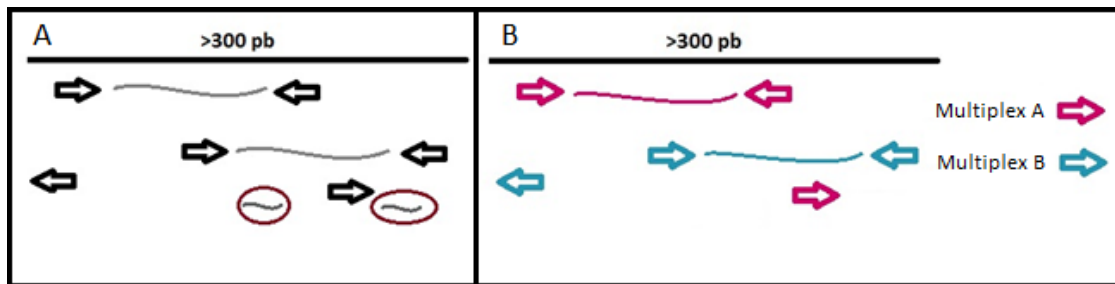


Figura 16. Representación de los dos multiplexes A y B. A representación de lo que ocurriría si no se separaran los cebadores. B Diseño de multiplexes.

Por ejemplo, el cebador del gen *TCF7L2* exón 14c en dirección reversa, interaccionaba con el cebador del gen *MANBA* exón 11a en dirección reversa, lo que hacía que la pareja del exón 14c del gen *TCF7L2* estuviera en el multiplex A y los cebadores del gen *MANBA* exón 11a se incluyeran en el multiplex B. Si bien, para sumar los paneles 1 y 2 no se modificó la distribución de oligonucleótidos dado que los grupos ya estaban definidos y se quería comprobar que pese a haber cebadores muy diferentes y posibles problemas de interacciones o diseño, modificando las condiciones se mejoraba la cobertura de los diferentes fragmentos.

### 4.2. Optimización PCR 1-multiplex

A partir de pruebas previas en el grupo de investigación, se tenían las condiciones de partida de la enzima utilizada en el laboratorio con concentración de *mix* de cebadores a 0,4  $\mu$ M.

Para realizar la optimización, evitando un aumento de costes, se realizó solamente con la multiplex A de los diseños 1 y 2, y más tarde se probarían las mejores condiciones para el multiplex B.



Por otro lado, se probó la posibilidad de amplificar, como se ha indicado anteriormente, los paneles 1+2 conjuntamente pese a que los diseños eran muy diferentes (parámetros de diseño, Tms obtenidas y tamaños de los amplicones). Si la modificación de las condiciones permitiera mejorar la amplificación de estos dos multiplexes conjuntamente y obtener una mejora significativa supondría un importante adelanto al poder combinar cebadores y amplicones muy diferentes

Para ello se realizaron diferentes pruebas en base a los resultados que se fueron obteniendo. En estas pruebas se incluyeron 2 aditivos (aditivo 1 y aditivo 2) que pueden mejorar los resultados de las amplificaciones según los datos de la UGD:

### Prueba 1

En la prueba 1 se evaluó, para una misma concentración del *mix* de cebadores (0,04  $\mu\text{M}$  , siendo una dilución de 1/10 del *mix* original), diferentes concentraciones de aditivos (Tabla 2) y para el multiplex 1,2 y 1+2. Esto se probó con las condiciones habituales usadas en la UGD (Tabla 3) y una modificación (Tabla 4).

Tabla 2. Condiciones de aditivos probadas en la prueba 1.

Cebadores a 0,1X	ADITIVO 1 (mM)					
		0	1,5	3	6	9
ADITIVO 2 (%)	0	Cebadores: 0,1X	Cebadores: 0,1X Aditivo 1: 1,5 mM	Cebadores 0,1X Aditivo 1: 3mM	Cebadores 0,1X Aditivo 1: 6mM	Cebadores 0,1X Aditivo 1: 9mM
	2,5	Cebadores: 0,1X Aditivo 2: 2,5 %	Cebadores: 0,1X Aditivo 1: 1,5 Mm Aditivo 2: 2,5%	Cebadores 0,1X Aditivo 1: 3Mm Aditivo 2: 2,5%	Cebadores 0,1X Aditivo 1: 6Mm Aditivo 2: 2,5%	Cebadores 0,1X Aditivo 1: 9Mm Aditivo 2: 2,5%
	5	Cebadores: 0,1X Aditivo 2: 5%	Cebadores: 0,1X Aditivo 1: 1,5 Mm Aditivo 2: 5%	Cebadores 0,1X Aditivo 1: 3 Mm Aditivo 2: 5%	Cebadores 0,1X Aditivo 1: 6 Mm Aditivo 2: 5%	Cebadores 0,1X Aditivo 1: 9 Mm Aditivo 2: 5%

**Tabla 3.** Condiciones A de amplificación para la prueba 1.

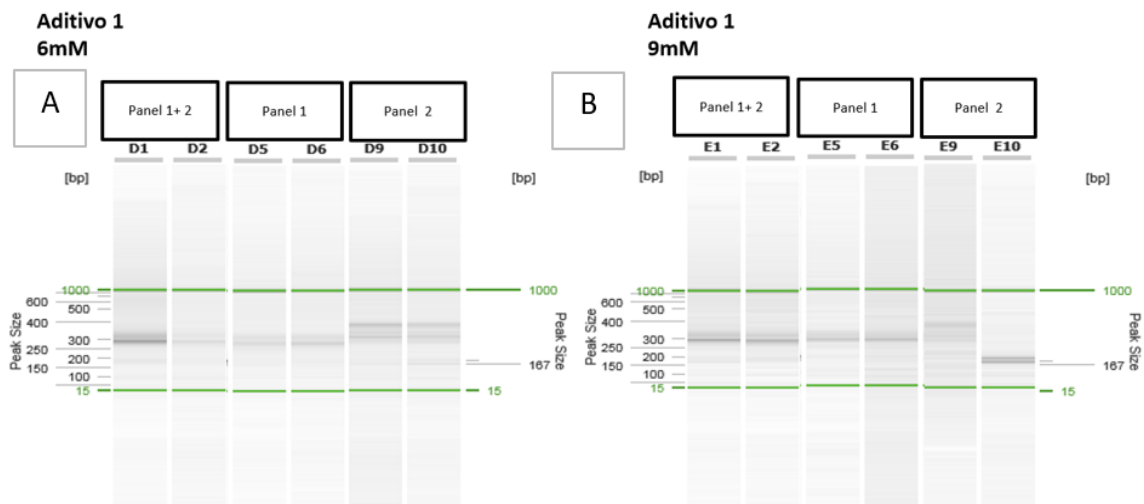
Programa	Temperatura	Tiempo	Ciclos
Activación inicial	95 °C	15 min	
Desnaturalización	96 °C	15 s	25x
Unión	55 °C	30 s	
	60 °C	30 s	
Extensión	72 °C	60 s	
Extensión final	72 °C	5 min	
Hold	15 °C	∞	

**Tabla 4.** Condiciones B de amplificación para la prueba 1.

Programa	Temperatura	Tiempo	Ciclos
Activación inicial	95 °C	15 min	
Desnaturalización	96 °C	15 s	25x
Unión	60 °C	30 s	
Extensión	72 °C	60 s	
Extensión final	72 °C	5 min	
Hold	15 °C	∞	

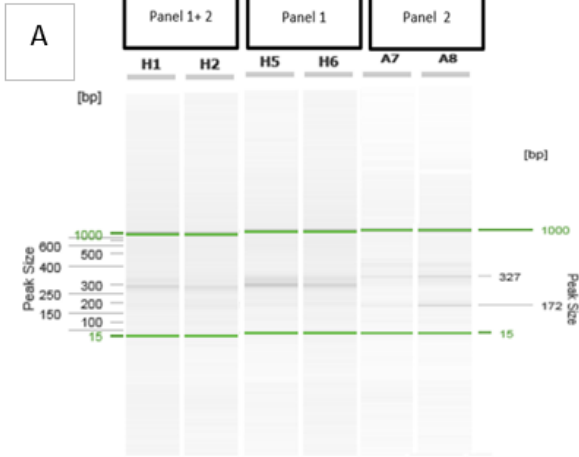
Los resultados indicaron que las mejores condiciones eran:

- Para los paneles 1,2, y 1+2: Aditivo 1: 6 mM y 9 mM y aditivo 1: 6 mM y 9 mM con aditivo 2: 2,5% y 5%; con las condiciones de A PCR1 (Figuras 17,18 y 19).

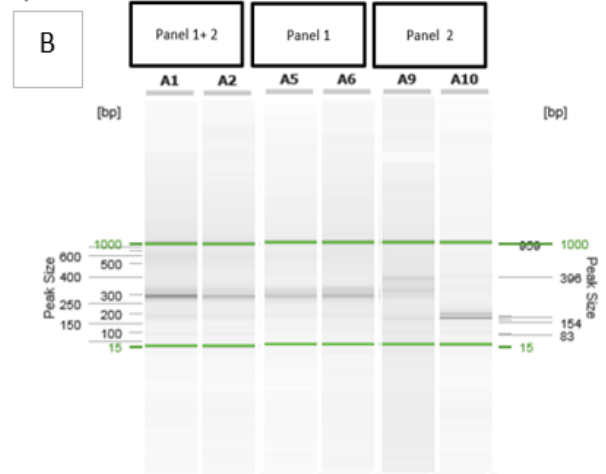


**Figura 17.** Resultados de la electroforesis para las condiciones de Aditivo 1 6mM y 9 mM para los paneles 1+2,1 y 2.A. Resultados para aditivo 1 6mM B. Resultados para aditivo 1 9mM.

**Aditivo 1  
6mM  
Aditivo 2  
2,5%**

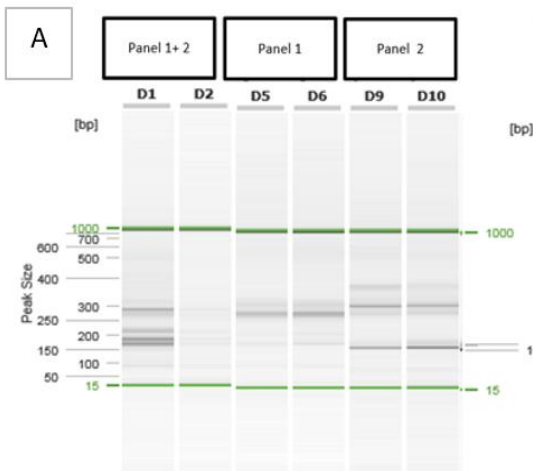


**Aditivo 1  
9mM  
Aditivo 2  
2,5%**

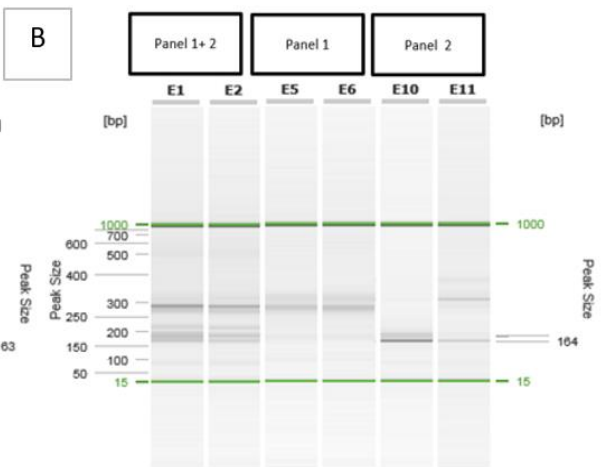


**Figura 18.** Resultados de la electroforesis para las condiciones de Aditivo 1 6mM y 9 mM y aditivo 2 2,5% para los paneles 1+2,1 y 2. **A.** Resultados para aditivo 1 6mM y aditivo 2 2,5% **B.** Resultados para aditivo 1 9mM y aditivo 2 2,5%

**Aditivo 1  
6mM  
Aditivo 2  
5%**

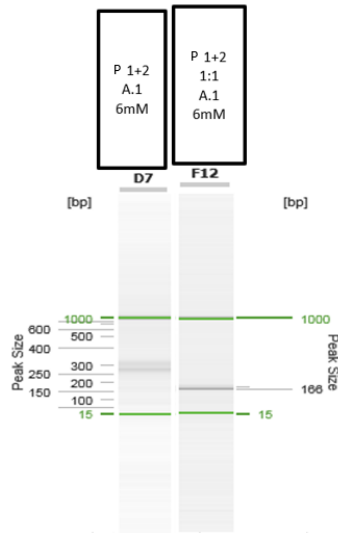


**Aditivo 1  
9mM  
Aditivo 2  
5%**

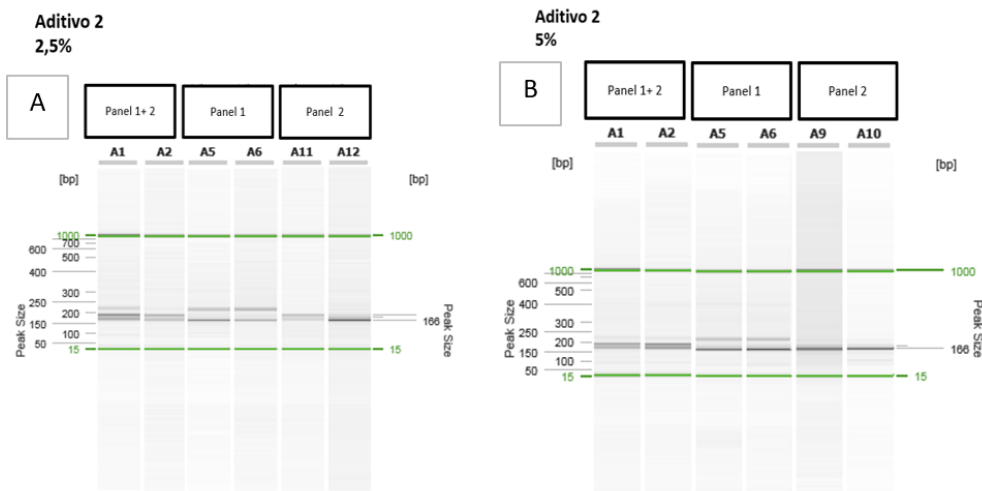


**Figura 19.** Resultados de la electroforesis para las condiciones de Aditivo 1 6mM y 9 mM para los paneles 1+2,1 y 2. **A.** Resultados para aditivo 1 6mM y aditivo 2 5% **B.** Resultados para aditivo 1 9mM y aditivo 2 5%.

Se observaba que tanto las condiciones B de PCR, eliminando una temperatura de unión como solo el aditivo 2, en todos los casos suponía una pérdida de intensidad en la amplificación y se aumentaba la cantidad de dímero (Figura 20,21).



**Figura 20.** Comparación de la electroforesis de una condición (6mM) entre las condiciones 'estándar' de PCR (condiciones A) y la modificación (condiciones B). **Nota:** P:panel A:Aditivo



**Figura 21.** Resultados de la electroforesis para las condiciones de Aditivo 2 2,5% y 5% para los paneles 1+2.1 y 2. **A.** Resultados para aditivo 2 2,5% **B.** Resultados para aditivo 2 5%

### Prueba 2

En este caso, se repitieron las condiciones favorables obtenidos de la fase anterior (Tabla 5), con una nueva condición de amplificación: aumentando la temperatura de unión a 55°C en la condición de PCR en la que se elimina una temperatura de unión, para comprobar si en este caso salían bandas específicas y utilizando la condición 'de partida' como control (Tabla 6,7).

Dadas las diferentes características de los cebadores y amplicones del panel 2, se realizó una prueba para ver si aumentando la cantidad de cebadores de este panel permitía mejorar su amplificación conjuntamente con el panel 1.

**Tabla 5.** Condiciones de aditivos analizadas en la prueba 2.

Cebadores a 0,1X	ADITIVO 1 (mM)		
ADITIVO 2 (%)		6	9
	0	Cebadores 0,1X Aditivo 1: 6mM	Cebadores 0,1X Aditivo 1: 9mM
	2,5	Cebadores 0,1X Aditivo 1: 6Mm Aditivo 2: 2,5%	Cebadores 0,1X Aditivo 1: 9Mm Aditivo 2: 2,5%
	5	Cebadores 0,1X Aditivo 1: 6 Mm Aditivo 2: 5%	Cebadores 0,1X Aditivo 1: 9 Mm Aditivo 2: 5%

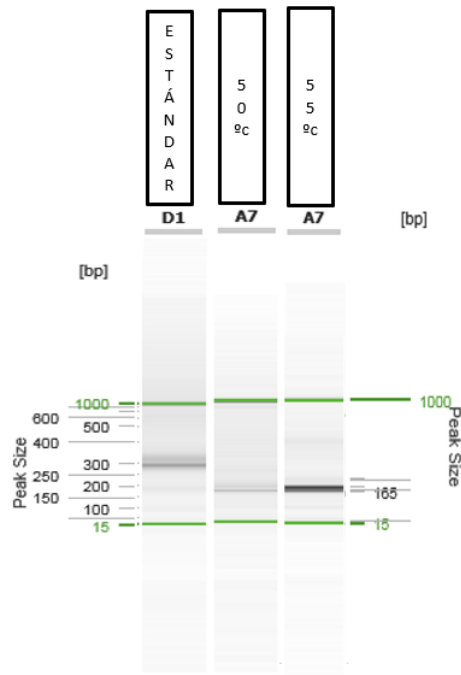
**Tabla 6.** Condiciones A de amplificación para la prueba 2.

Programa	Temperatura	Tiempo	Ciclos
Activación inicial	95°C	15 min	
Desnaturalización	96 °C	15 s	25x
Unión	55°C	30 s	
Extensión	72 °C	60 s	
Extensión final	72 °C	5 min	
Hold	15 °C	∞	

**Tabla 7.** Condiciones B de amplificación para la prueba 2.

Programa	Temperatura	Tiempo	Ciclos
Activación inicial	95°C	15 min	
Desnaturalización	96 °C	15 s	25x
Unión	50°C	30 s	
Extensión	72 °C	60 s	
Extensión final	72 °C	5 min	
Hold	15 °C	∞	

Se descartaron todos los resultados debido a que no se observaba amplificación en la electroforesis capilar (Figura 22).



**Figura 22.** Comparación de una condición (Aditivo 1 6mM ) para el multiplex del panel 1+2 con los ciclos de PCR 'estándar', la modificación de un paso a 55°C (A) y la modificación a 50°C (B) , en ese orden, siendo estas últimas dos pruebas con una proporción similar entre cebadores.

### Prueba 3

En vista de los resultados previos se procedió a estudiar si cambios entre 5 y 9 mM tenían un efecto relevante y si una gradación aumentar la cantidad de aditivo 1 puede aportar mejores resultados (Tabla 8). A partir de esta prueba, se utilizó únicamente una condición de PCR, la seleccionada como óptima (Tabla 9).

**Tabla 8.** Condiciones de aditivos probadas en la prueba 3.

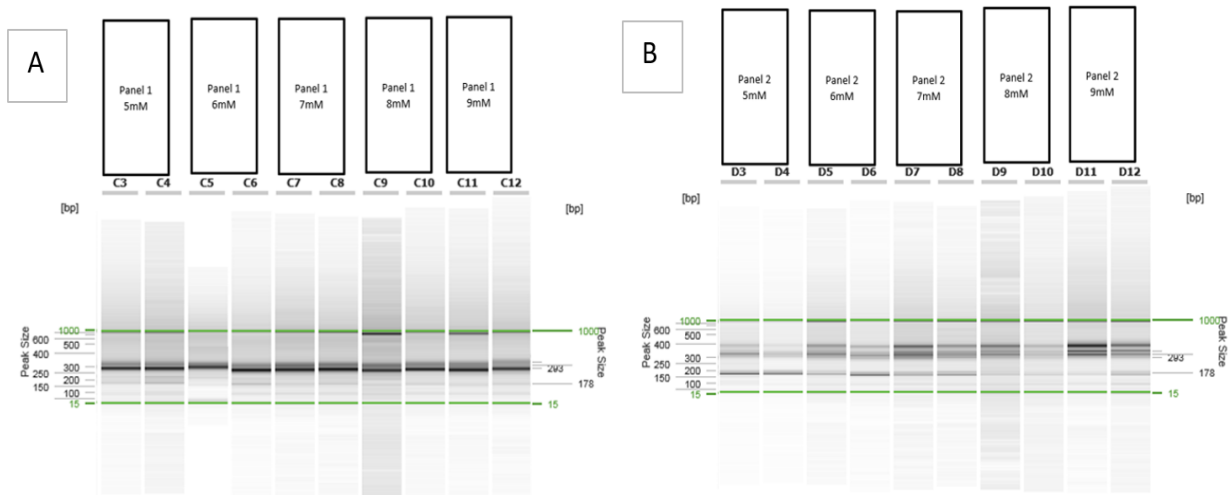
Cebadores a 0,1X	ADITIVO 1 (mM)						
	3	4	5	6	7	8	9
Cebadores 0,1X Aditivo 1: 3mM	Cebadores 0,1X Aditivo 1: 4mM	Cebadores 0,1X Aditivo 1: 5mM	Cebadores 0,1X Aditivo 1: 6mM	Cebadores 0,1X Aditivo 1: 7mM	Cebadores 0,1X Aditivo 1: 8mM	Cebadores 0,1X Aditivo 1: 9mM	Cebadores 0,1X Aditivo 1: 9mM

**Tabla 9.** Condiciones de amplificación para la prueba 3 y resto del proyecto.

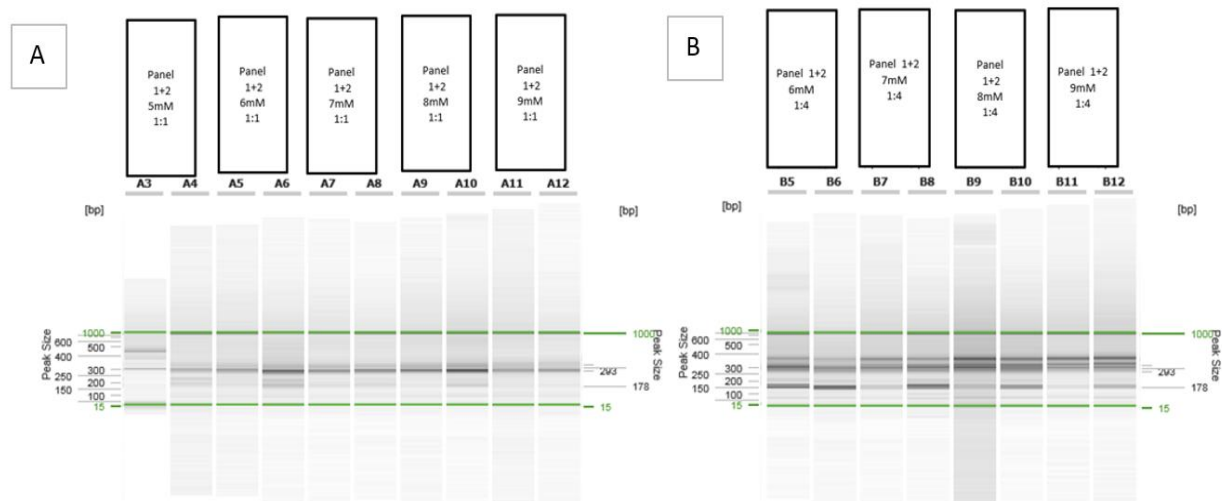
Programa	Temperatura	Tiempo	Ciclos
Activación inicial	95°C	15 min	
Desnaturalización	96 °C	15 s	25x
Unión	55°C	30 s	
	60°C	30 s	
Extensión	72 °C	60 s	
Extensión final	72 °C	5 min	
Hold	15 °C	∞	

Las condiciones que mejor resultados dieron fueron:

- Del multiplex de paneles 1 y 2: Aditivo 1 5mM, 6mM, 7mM, 8mM, 9mM (Fig. 23A, 23B).
- Del multiplex de paneles 1+2 con una proporción 1:1: Aditivo 1 5mM, 6mM, 7mM, 8mM, 9mM (Fig. 24A).
- Del multiplex de paneles 1+2 con una proporción 1:4: Aditivo 1 6mM, 7mM, 8mM, 9mM (Fig. 24B).



**Figura 23.** Electroforesis de resultados positivos para los paneles 1, 2 en la prueba 3. **A.** Resultados para el panel 1 **B.** Resultados para el panel 2.



**Figura 24.** Electroforesis de resultados positivos para la librería de los paneles 1 +2 para proporciones 1:1 y 1:4 en la prueba 3. **A.** Resultados para los paneles 1+2 con proporción 1:1 **B.** Resultados para los paneles 1+2 con proporción 1:4.

#### Prueba 4

Dado que el aditivo 1 a una concentración 9 mM permitía obtener resultados aparentemente correctos según la Figuras 17,23 y 24, donde se observa la amplificación correcta de los amplicones mediante electroforesis con el sistema QIAxcel, se decidió probar una concentración mayor para ver el efecto. Por otro lado, se quiso verificar el efecto de la disminución de cebadores en la librería 1 (0.05x) (Tabla 10):

**Tabla 10.** Condiciones de aditivos probadas en la prueba 4.

ADITIVO 1 (Mm)				
9	10	11	12	15
Cebadores 0,05X Aditivo 1: 9mM	Cebadores 0,1X Aditivo 1: 10mM	Cebadores 0,1X Aditivo 1: 11mM	Cebadores 0,1X Aditivo 1: 12mM	Cebadores 0,1X Aditivo 1: 15mM

Las pruebas que mejor resultados dieron, fueron:

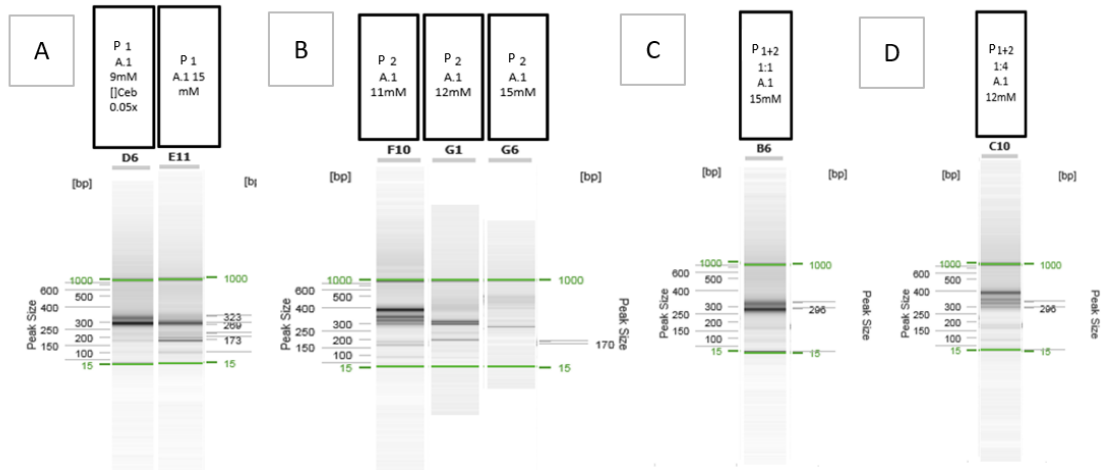
De la librería del panel 1: Aditivo 1 9mM (Cebadores a 0,05x), Aditivo 1 15mM (Fig. 25A).

De la librería del panel 2: Aditivo 1 11mM, Aditivo 1 12mM, Aditivo 1 15mM (Fig. 25B).

De la librería panel 1+2 preparados con una proporción 1:1: Aditivo 1 15mM (Fig 25C).

De la librería panel 1+2 preparados con una proporción 1:4: Aditivo 1 12mM, 15mM (Fig 25D).





**Figura 25.** Electroforesis de las mejores condiciones observadas en la prueba 4 para los paneles 1, 2, 1+2 con proporción 1:1 y 1+2 con proporción 1:4. **A.** Resultados par a la librería del panel 1. **B.** para la librería del panel 2. **C.** Para librería del panel 1+2 con proporción 1:1 de cebadores. **D.** Para librería del panel 1+2 con proporción 1:4 de cebadores. Nota: P: panel/es, A: aditivo, [ ] ceb: concentración de cebador.

### 4.3. Optimización PCR 2-anidada

#### Prueba 1

Dado que los amplicones de los cebadores del panel 2 son de mayor tamaño de los que se suelen utilizar en este laboratorio, se planteó la posibilidad de favorecer su amplificación mediante el incremento en el tiempo de extensión en la PCR 2, para ello se realizó una prueba que permitiera ver si aumentar este tiempo nos permite aumentar la amplificación de los amplicones de mayor tamaño. Para ello, se comprobaron las siguientes condiciones (Tabla 11, 12)

**Tabla 11.** Condiciones A de amplificación para la prueba 1.

Programa	Temperatura	Tiempo	Ciclos
Activación inicial	95 °C	15 min	
Desnaturalización	96 °C	15 s	35x
Unión	64°C	30 s	
Extensión	72 °C	1 min	
Extensión final	72 °C	5 min	
Hold	15 °C	∞	

Tabla 12. Condiciones B de amplificación para la prueba 1.

Programa	Temperatura	Tiempo	Ciclos
Activación inicial	95 °C	15 min	
Desnaturalización	96 °C	15 s	35x
Unión	64°C	30 s	
Extensión	72 °C	90 s	
Extensión final	72 °C	5 min	
Hold	15 °C	∞	

Los resultados indican que las mejores condiciones son la amplificación aumentando a 64º la temperatura de unión y 90s el tiempo en el paso de extensión a 72ºC (condiciones B) , ya que se observa que en la tercera carrera de la imagen hay más amplificación que en la primera y menos dímero que en la segunda (Fig. 26).

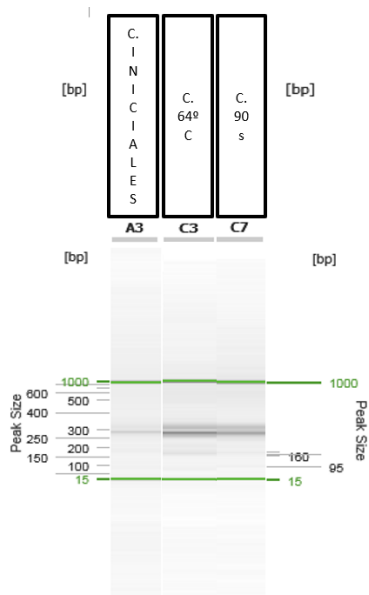


Figura 26. Ejemplo de muestra en la electroforesis capilar con las condiciones iniciales, la modificación de 64ºC y la modificación de 90s. Nota: C. : condiciones.

En cualquier caso, excepto por aquellas condiciones en las que no se obtenían los amplicones de interés o se obtenía mucho dímero, es necesario realizar un estudio mediante NGS para ver qué fragmentos son amplificados correctamente y la diferencias en la amplificación y secuenciación de cada uno de ellos.

#### 4.4. Secuenciación de las pruebas con resultado positivo.

##### Purificación y cuantificación.

Se obtuvieron distintas concentraciones de ADN muy variables ya que, al ser un trabajo de optimización, los resultados dependían de las condiciones.

##### Cuantificación

Se realizaron varias cuantificaciones a partir de una curva estándar (Fig. 27).

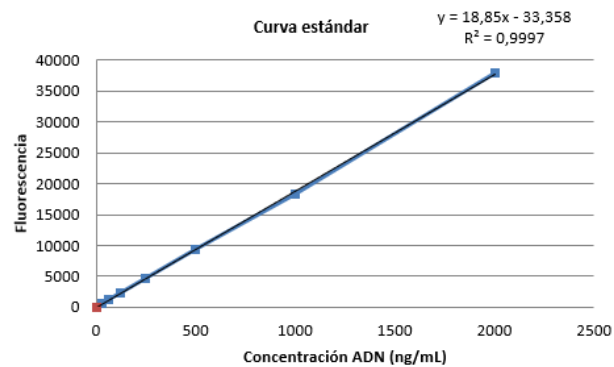


Figura 27. Ejemplo de recta de regresión para extrapolar los valores de la cuantificación.

##### Verificación de la purificación mediante electroforesis en QIAxcel.

Se realizó una electroforesis tras la purificación para comprobar que no se había perdido la muestra durante el procedimiento (Fig. 28).

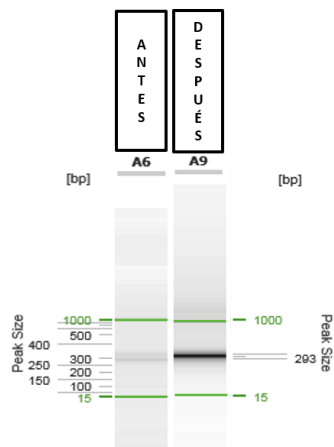


Figura 28. Ejemplo de comparación electroforesis entre antes de la purificación y después en el multiplex A.

### Reacción de secuenciación.

Se realizó siguiendo procedimiento estándar de Illumina, incluyendo las librerías con una amplificación correcta o casi correcta y teniendo en cuenta el número de amplicones: En el caso del multiplex A se hizo una única librería para el panel 1 y 2 ya que tenían un número de amplicones muy similar (70 y 73) y en el caso de la multiplex B se hicieron las 3 librerías por separado.

En los archivos obtenidos tras la secuenciación, se observaron algunas regiones cubiertas al 100% según el límite inferior que se había establecido al principio (100x), y otras que no alcanzaban este límite en las reacciones realizadas (indicadas como 0 en la *tabla 15*).

**Tabla 13.** Ejemplo de uno de los documentos obtenidos tras la secuenciación de la librería correspondiente a los paneles 1+2.

muestra	amplicón	longitud	cob_min	cob_max	%cob_encima_100x
1	MANBA-e6c	106	673	673	100
1	MANBA-e7a	104	159	167	100
1	MANBA-e8a	105	342	343	100
1	MANBA-e9a	105	550	550	100
1	MANBa-e6a	108	0	0	0
1	SREBF2-E10a	195	0	0	0
1	SREBF2-E11a	186	2	4	0
1	SREBF2-E12a	161	1	2	0
1	SREBF2-E13a	172	1167	2368	100

La diferencia en la amplificación entre los diferentes fragmentos puede deberse a diferentes factores (secuencia de los cebadores, de los amplicones,  $T_m$ , interacciones entre cebadores e incluso con los amplicones, etc.). Si bien, en la tabla 15 queda claro que los cebadores pertenecientes al panel 1 (*MANBA*) se amplifican y secuencian mejor que los del 2 (*SREBF2*).

### MULTIPLEX A

Los resultados muestran que algunas de las condiciones probadas para el panel 1 tiene unos buenos resultados a partir de 20.000 lecturas dado que se obtienen porcentajes del 80% de los fragmentos analizables llegando a más del 85% en algunas condiciones con 30.000 o 50.000 lecturas (Figura 29). Los resultados obtenidos para los paneles 2 y 1+2 son de baja calidad dado que los porcentajes son mucho menores (Figuras 30, 31).

En el análisis de la librería 1+2 se aumentó el número de lecturas totales hasta 100.000 dado que cuenta con el doble de fragmentos que las librerías por separado. Pese a ello, los resultados siguen sin ser adecuados.

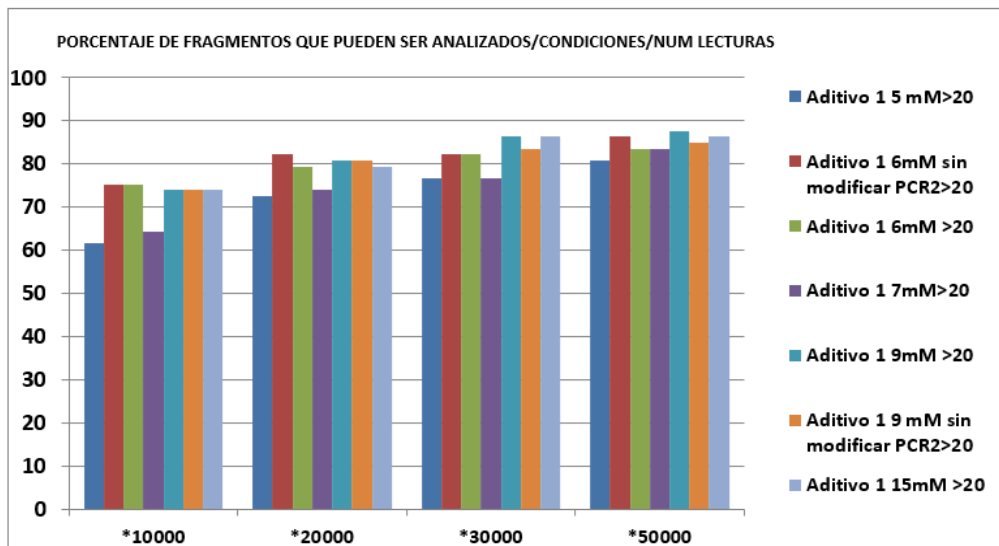


Figura 29. Número de fragmentos que pueden ser analizados en cada condición para cada cobertura en el panel 1. En la figura se indica el número de fragmentos que tienen una profundidad mayor de 20x.

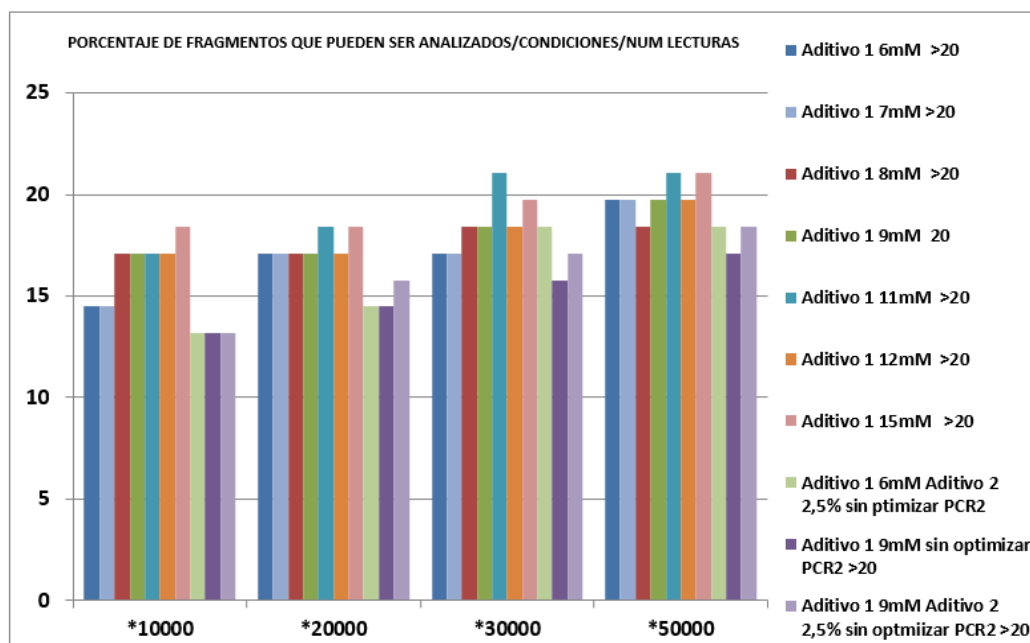


Figura 30. Número de fragmentos que pueden ser analizados en cada condición para cada cobertura en el panel 2. En la figura se indica el número de fragmentos que tienen una profundidad mayor de 20x.

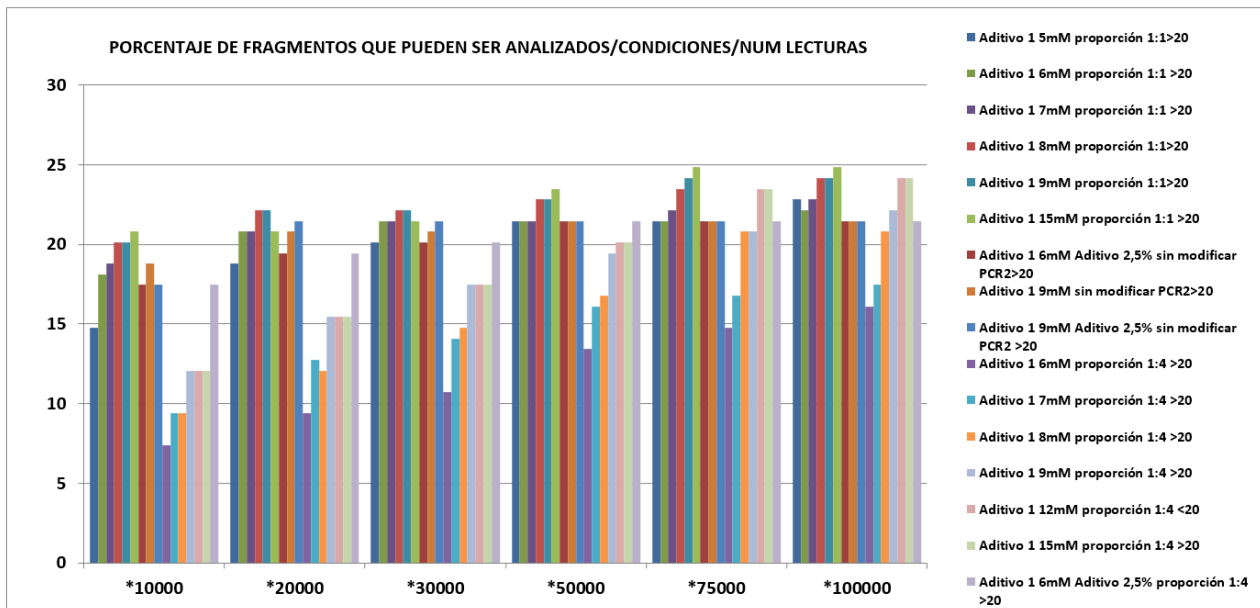


Figura 31. Número de fragmentos que pueden ser analizados en cada condición para cada cobertura en el grupo de paneles 1+2 la figura se indica el número de fragmentos que tienen una profundidad mayor de 20x.

Por otro lado, la visualización de las lecturas de cada amplicón permite ver cómo se han distribuido las lecturas y a qué amplicón corresponden. Para ello, se han creado figuras representando para las tres o cuatro condiciones que mejores resultados han dado, el número de lecturas que se obtendría para la cobertura que se considera idónea (calculándose como lecturas normalizadas por la cobertura) frente a cada amplicón, representados como números (1-73,1-76,1-149 respectivamente):

- En el caso de las librerías del panel 1, se observaron como condiciones óptimas: Aditivo 1 a 6 mM y 9mM y aditivo 1 a 6mM sin optimizar las condiciones de PCR2 (Fig. 32).

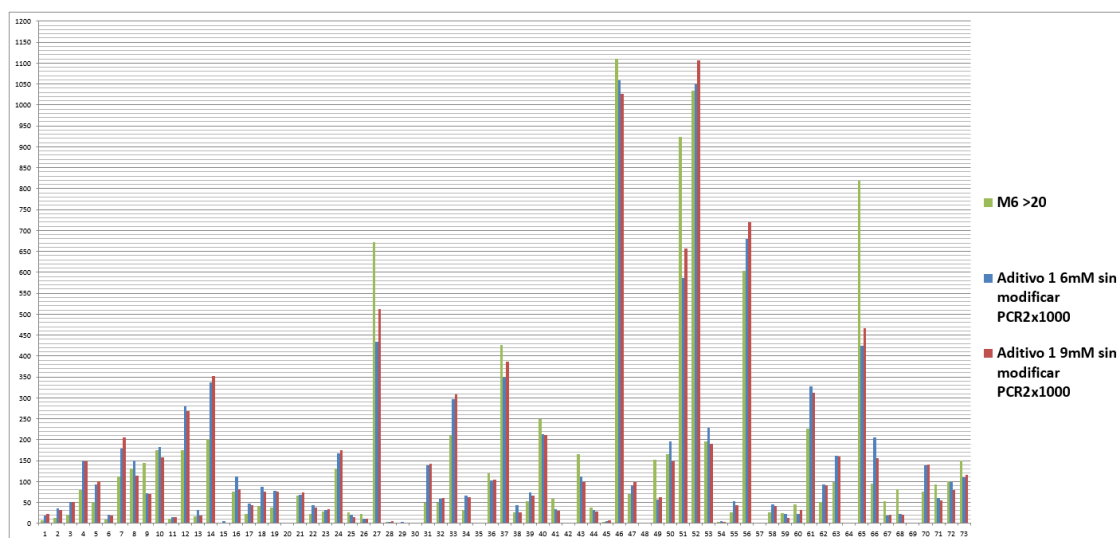


Figura 32. Distribución de lecturas en los distintos amplicones de las mejores condiciones de generación de las librerías para el panel 1.

- En el caso de las librerías del panel 2, las condiciones que mejores resultados dieron fueron: Aditivo 1 a 9 mM, 11 mM, 12 mM y 15 mM (Fig. 33).

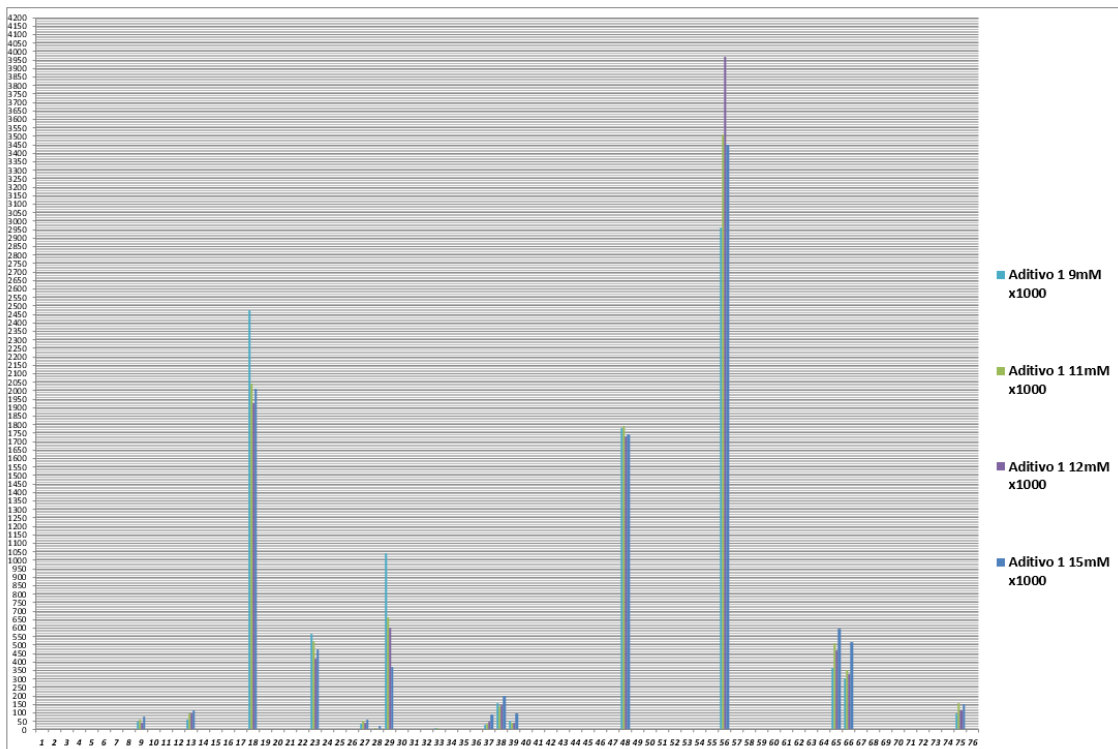


Figura 33. Distribución de lecturas en los distintos amplicones de las condiciones óptimas de secuenciación para las librerías obtenidas del panel 2.

- Los resultados para los paneles 1+2, las mejores condiciones fueron: 1 8mM, 9mM y 15mM (Figura 34).

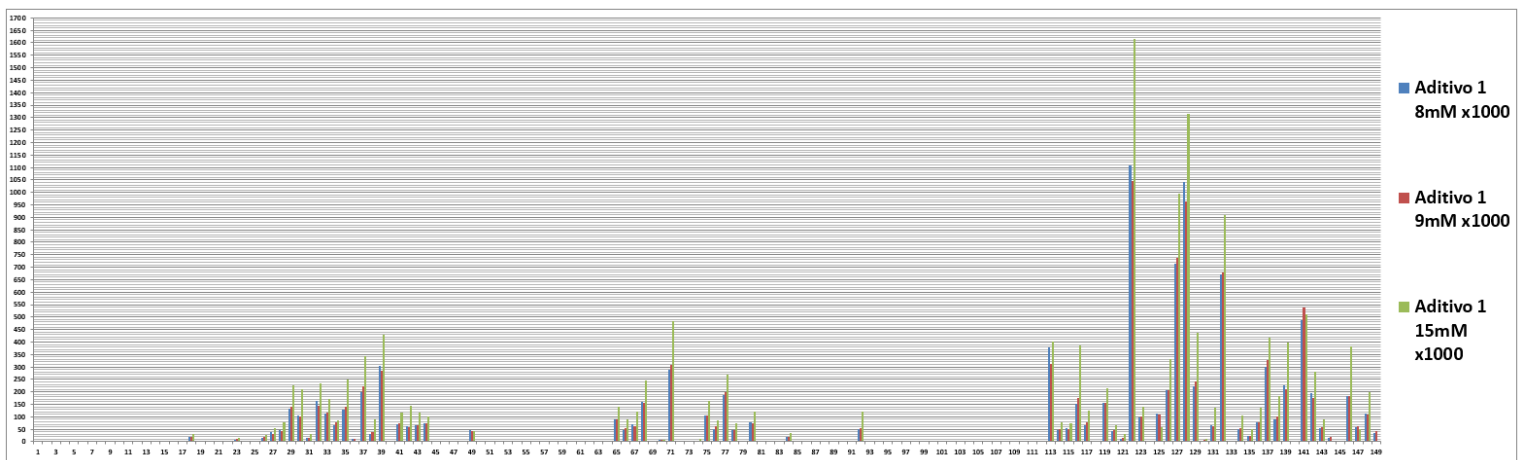


Figura 34. Distribución de lecturas en los distintos amplicones de las condiciones óptimas de secuenciación para el panel 1+2.

## MULTIPLEX B

En este caso, el análisis de los resultados de coberturas llevó a diferentes conclusiones, dependiendo de cada panel. En el caso de realizarse por separado:

- Para la multiplex B del panel 1 sería conveniente utilizar las condiciones que incluyen el aditivo 1 a 9 mM teniendo resultados aceptables con 20.000 lecturas que mejoran con 50.000, llegando a tener casi un 90% de cobertura ya que sí que aumenta considerablemente el número de amplicones que se pueden analizar (Fig. 35).
- En el caso de las librerías de los paneles 2 y 1+2 los resultados son similares a los obtenidos en la Multiplex A (Fig. 36,37); si bien las condiciones probadas no suponen grandes cambios.

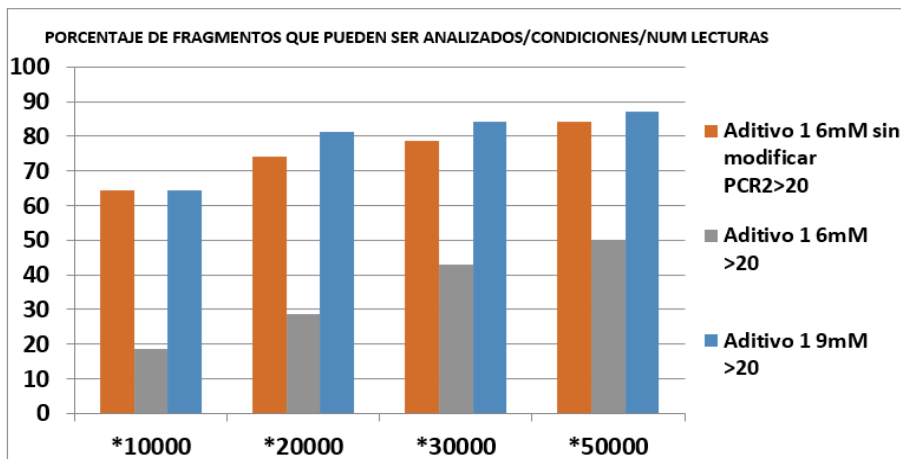


Figura 35. Número de fragmentos que pueden ser analizados en cada condición para cada cobertura en la librería del panel 1. En la figura se indica el número de fragmentos que tienen una profundidad mayor de 20x.

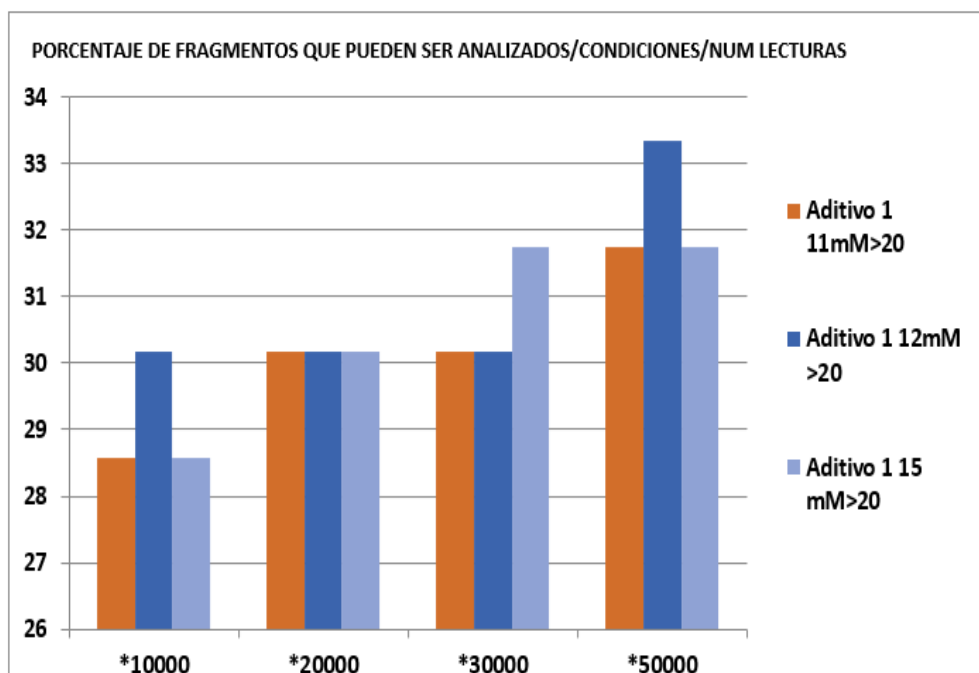


Figura 36. Número de fragmentos que pueden ser analizados en cada condición para cada cobertura en la librería del panel 2. En la figura se indica el número de fragmentos que tienen una profundidad mayor de 20x.



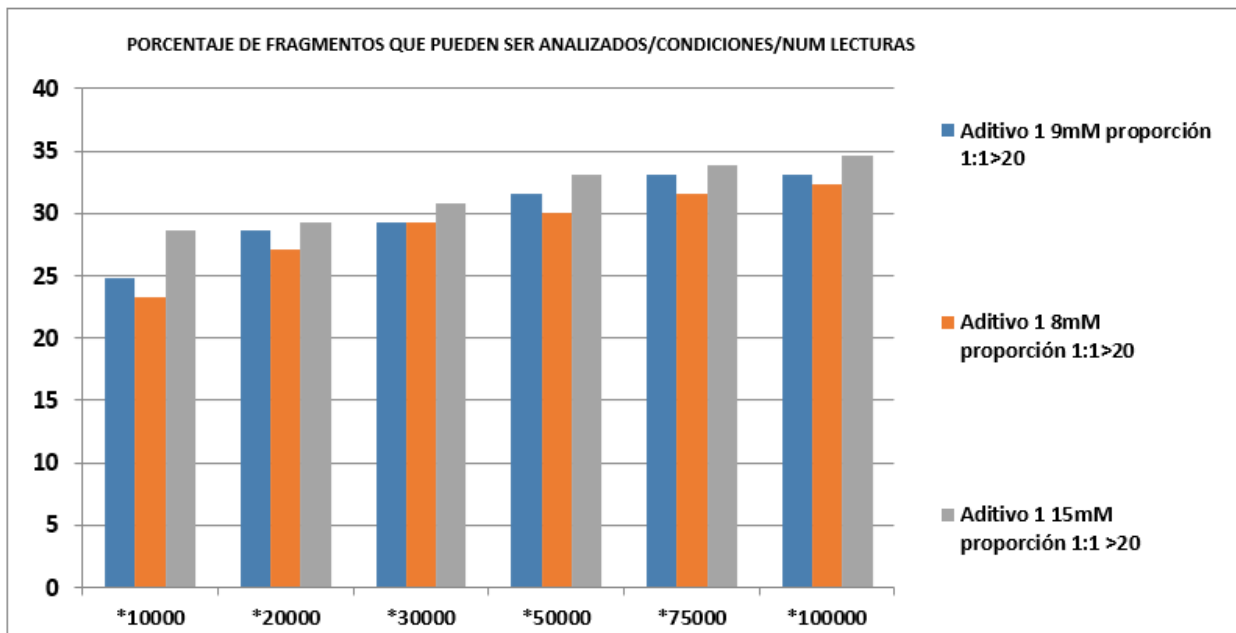


Figura 37. Número de fragmentos que pueden ser analizados en cada condición para cada cobertura en el grupo del panel 1+2. En la figura se indica el número de fragmentos que tienen una profundidad mayor de 20x.

- Dados los resultados obtenidos, el estudio de los paneles debería hacerse por separado y utilizar condiciones diferentes para el segundo panel (optimizado anteriormente en la Unidad para este estudio). Trabajos previos del laboratorio mostraron que se mejoraban los resultados de este panel, alcanzándose un 85-90% de cobertura con un aumento importante de polimerasa en la reacción, lo que aumentaría en gran medida el coste, con lo que no sería aplicable para un estudio a gran escala.

Por otro lado, en cuanto a las condiciones que mejores resultados han dado:

- En el panel 1 (multiplex B), se observó como condición aditivo 1 a 9mM, ya que permite analizar mayor número de amplicones, y a pesar de que la condición de aditivo 1 6mM sin modificar la PCR2 obtiene más lecturas en muchos amplicones, no hay una distribución tan homogénea como la otra condición (Fig. 38). Esto supondría una mayor necesidad de lecturas para analizar el mismo número de fragmentos.

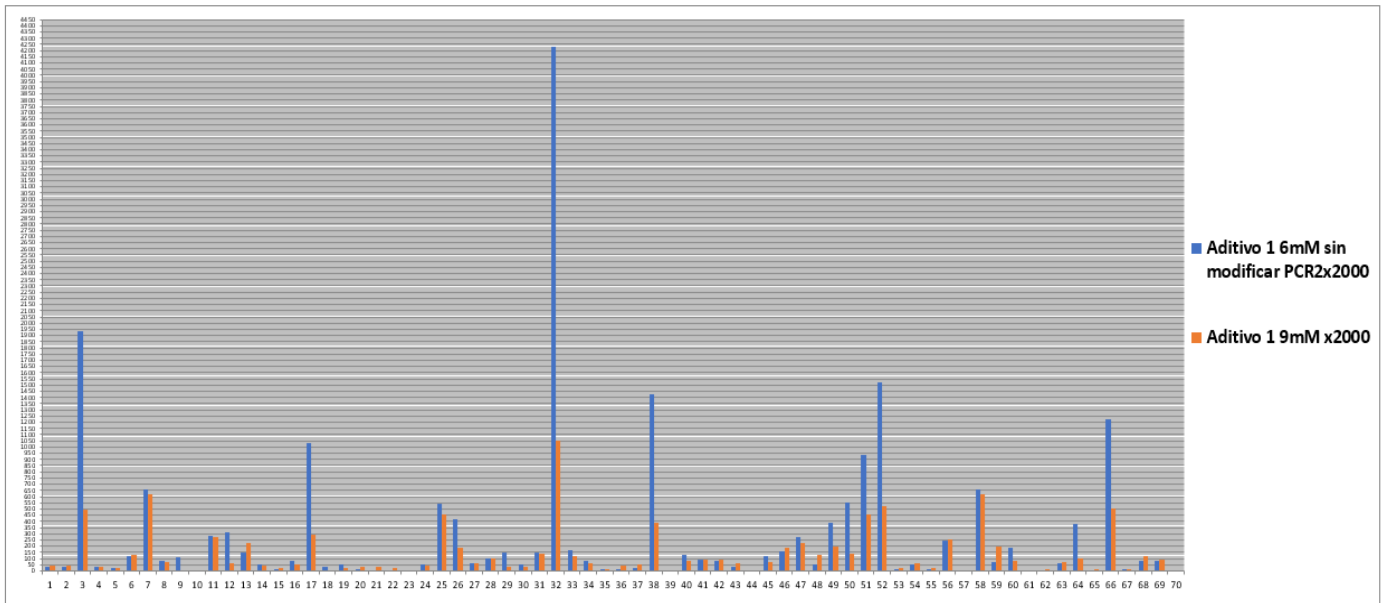


Figura 38. Distribución de lecturas en los distintos amplicones de las condiciones óptimas de secuenciación para el multiplex del panel 1.

- En el caso del panel 2, la condición óptima es aditivo 1 a 12mM, ya que permite analizar un mayor número de amplicones, y en comparación con otras de las condiciones (Fig. 39), las lecturas son mucho más homogéneas.

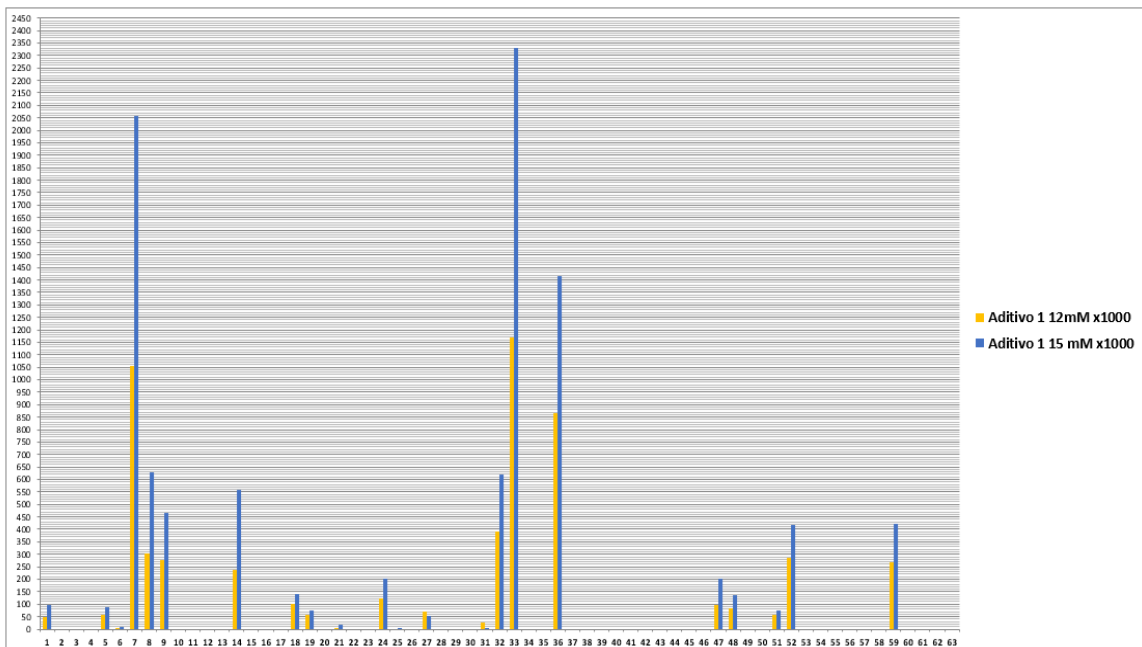


Figura 39. Distribución de lecturas en los distintos amplicones de las condiciones óptimas de secuenciación para el multiplex del panel 2.

- En el caso del panel 1+2, se observa que los mejores resultados se obtienen con aditivo 1 a 15mM ya que permite un mayor análisis de fragmentos (Fig. 40).

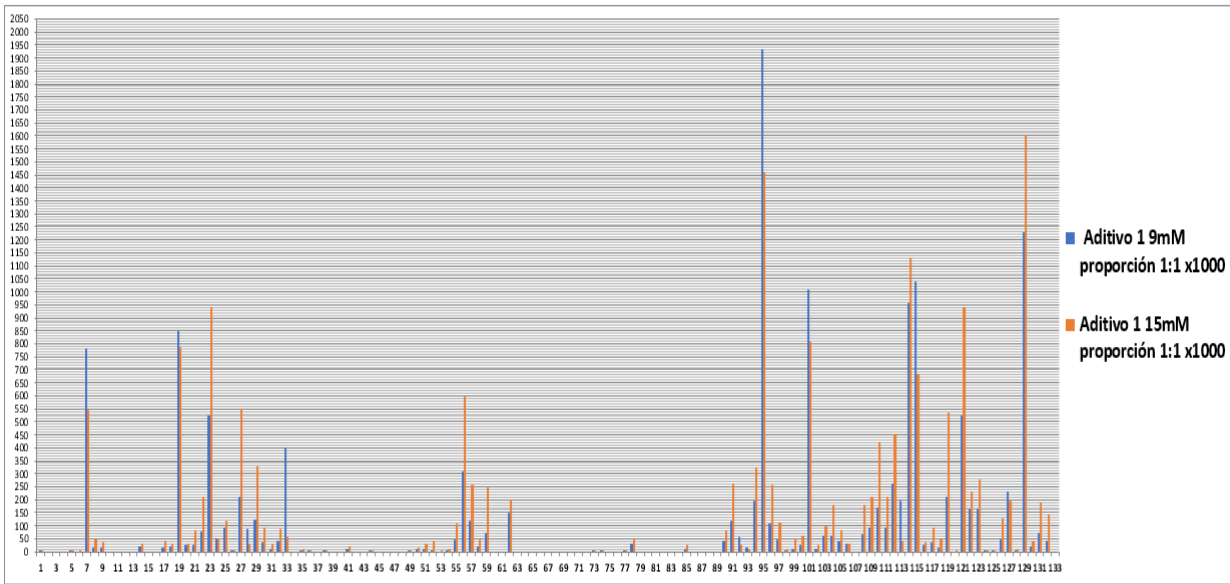


Figura 40. Distribución de lecturas en los distintos amplicones de las condiciones óptimas de secuenciación para el multiplex del panel 1+2.

Para visualizar de otra manera los resultados, se realizaron unas gráficas (Fig. 41) que incluían ambos multiplexes de cada panel. En ella se observa la buena optimización del panel 1 respecto al panel 2. En el primer caso, panel 1, a 20x se consigue un estudio de prácticamente el 90% de los amplicones, mientras que en el caso de paneles 2 está debajo del 80%. Siguiendo la gráfica 41.C que representa amplitud/profundidad de cobertura del panel 1+2, la cobertura es inferior a los resultados de ambos paneles por separado, lo que podría indicar que alguna de las interacciones que se producen entre ambos paneles resultan en la deficiencia de amplificación de fragmentos de ambos paneles, ya que para la cobertura que se planteaba (20x) solo se podría estudiar el 60% de los fragmentos.

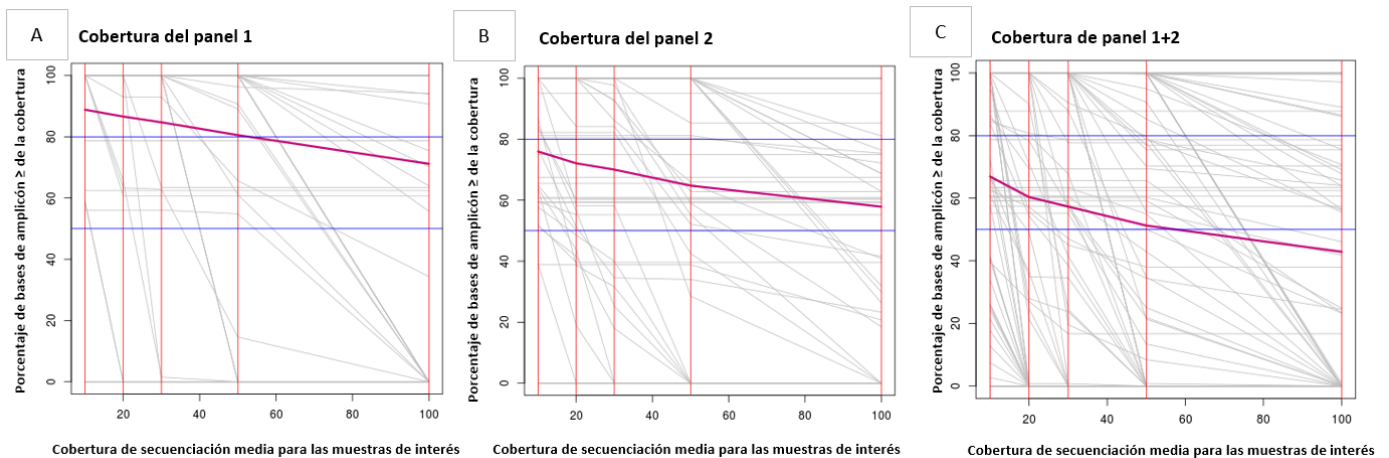


Figura 41. Representación de la cobertura obtenida para cada porcentaje de bases por amplicón (se muestra cada uno por separado en base a media de cobertura normalizada por condición en gris y en magenta la media de todos) en el caso de 50.000 lecturas. **A.** En el caso del panel 1. **B.** En el caso del panel 2 **C.** En el caso del panel 1+2. En el eje X, se muestra la cobertura de secuenciación media para las muestras de interés (profundidad) y en el eje Y el porcentaje de bases del amplicón por encima o igual que la cobertura indicada en el eje X (amplitud).

## 5. CONCLUSIONES

Se puede concluir que:

1. Existe una gran diferencia de los resultados obtenidos para los cebadores para los paneles 1 y 2, obteniendo los primeros resultados mucho más eficientes en sus dos multiplexes.
2. Las condiciones para el panel 1 son adecuadas, aunque podría ser interesante mejorarlas para poder alcanzar el 90% de amplificación. Con estas condiciones se alcanza un 85% de los fragmentos en el multiplex A y B respectivamente. Por tanto, debería continuarse con el proceso de optimización.
3. El diseño del panel 2 no es válido para esta enzima en condiciones estándar ni con los aditivos probados.
4. La combinación de las dos multiplexes no permite amplificar de forma apropiada ambos grupos de amplicones dadas las diferencias entre los mismos.
5. El trabajo de mejora de resultados ha supuesto que los resultados con el programa "ExonPrimer" mejoren enormemente dado que en las pruebas realizadas con condiciones estándar de la UGD no se obtenían productos de PCR que contuvieran librerías que pudieran ser secuenciadas. Este trabajo ha permitido llegar al 85% de cobertura al 20x de fragmentos para una enzima determinada, el uso de otros aditivos o de otras enzimas podría mejorar estos resultados.
6. Se deberá sopesar el beneficio que supone la reducción de tiempo en el diseño de cebadores. Esto puede implicar la búsqueda de un consenso entre el uso de programas bioinformáticos para la realización de multiplexes y el trabajo que supone el cambio de condiciones de reacción y su optimización.

## 6. BIBLIOGRAFÍA

- BESSER, J.; CARLETON, H. A.; GERNER-SMIDT, P.; LINDSEY, R. L. Y TREES, E. (2018). Next-generation sequencing technologies and their application to the study and control of bacterial infections. *Clinical Microbiology and Infection*, 24(4), 335-341. <https://doi.org/10.1016/j.cmi.2017.10.013>
- BODMER, W. Y BONILLA, C. (2008). Common and rare variants in multifactorial susceptibility to common diseases. *Nature genetics*, 40(6), 695-701. <https://doi.org/10.1038/ng.f.136>
- BOTSTEIN, D. Y RISCH, N. (2003). Discovering genotypes underlying human phenotypes: Past successes for mendelian disease, future approaches for complex disease. *Nature Genetics*, 33(S3), 228-237. <https://doi.org/10.1038/ng1090>
- BUERMANS, H. P. J. Y DEN DUNNEN, J. T. (2014). Next generation sequencing technology: Advances and applications. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease*, 1842(10), 1932-1941. <https://doi.org/10.1016/j.bbadis.2014.06.015>
- CHATTERJEE, S.; KHUNTI, K. Y DAVIES, M. J. (2017). Type 2 diabetes. *The Lancet*, 389(10085), 2239-2251. [https://doi.org/10.1016/S0140-6736\(17\)30058-2](https://doi.org/10.1016/S0140-6736(17)30058-2)
- DE FARIA MARASCHIN, J. (2012). Classification of diabetes, en: Diabetes, and old disease, a new insight. Edición Landes Bioscience:12-18
- DEFRONZO, R. A.; FERRANNINI, E.; GROOP, L.; HENRY, R. R.; HERMAN, W. H.; HOLST, J. J.; HU, F. B.; KAHN, C.R.; RAZ, I.; SHULMAN G. I.; SIMONSON D. C.; TESTA M. A. Y WEISS, R. (2015). Type 2 diabetes mellitus. *Nature Reviews Disease Primers*, 15019. <https://doi.org/10.1038/nrdp.2015.19>
- FID, D. (2015). Atlas de la Diabetes de la FID. Vol 7ª edición. Versión online del Atlas de la Diabetes de la FID2015.144 pp.
- FU, W., O'CONNOR, T. D., JUN, G., KANG, H. M., ABECASIS, G., LEAL, S. M., GABRIEL, S ; ALTSHULER, D ; SHENDURE, J.; NICKERSON, D. A.; BAMSHAD, M. J. Y AKEY, J. M. (2013). Analysis of 6,515 exomes reveals a recent origin of most human protein-coding variants. *Nature*, 493(7431), 216-220. <https://doi.org/10.1038/nature11690>
- GALAN-CHILET, I.(2016). Identificación de variantes genéticas poco frecuentes y raras en Diabetes mellitus tipo 2 mediante secuenciación de exoma (tesis doctoral), Universidad de Valencia, Valencia.

- MCCOMBIE, W. R.; MCPHERSON, J. D. Y MARDIS, E. R. (2018). Next-Generation Sequencing Technologies. *Cold Spring Harbor Perspectives in Medicine*. <https://doi.org/10.1101/cshperspect.a036798>
- SEDA, E. (2018). Desarrollo de un sistema de detección de mutaciones somáticas de interés clínico mediante secuenciación masiva en muestras oncológicas (Tesis doctoral). Universidad de Valencia, Valencia.
- SO, H.-C.; GUI, A. H. S.; CHERNY, S. S. Y SHAM, P. C. (2011). Evaluating the heritability explained by known susceptibility variants: A survey of ten complex diseases. *Genetic Epidemiology*, 35(5), 310-317. <https://doi.org/10.1002/gepi.20579>
- STENSON, P. D.; BALL, E. V.; MORT, M.; PHILLIPS, A. D.; SHIEL, J. A.; THOMAS, N. S. T.; ABEYSHINGHE, S.; KRAWCZAK, M. Y COOPER, D. N. (2003). Human Gene Mutation Database (HGMD®): 2003 update. *Human Mutation*, 21(6), 577-581. <https://doi.org/10.1002/humu.10212>
- TENNESSEN, J. A.; BIGHAM, A. W.; O'CONNOR, T. D.; FU, W.; KENNY, E. E.; GRAVEL, S.; MCGEE, S.; DO, R.; LIU, X; JUN, G.; MIN KANG, H.; JORDAN, D.; M.LEAL, S; GABRIEL, S.; RIEDER, M. J.; ABECASIS, G.; ALTSHULER, D.; NICKERSON, D. A.; BOERWINKLE, E.; SUNYAEV, S.; BUSTAMANTE, C. D.; BAMSHAD, M. J. Y AKEY, J. M. on behalf of the NHLBI Exome Sequencing Project. (2012). Evolution and Functional Impact of Rare Coding Variation from Deep Sequencing of Human Exomes. *Science*, 337(6090), 64-69. <https://doi.org/10.1126/science.1219240>