



BIOARRAY  
Diagnóstico Genético

UNIVERSIDAD MIGUEL HERNÁNDEZ

GRADO EN BIOTECNOLOGÍA

CURSO ACADÉMICO 2016/2017

TRABAJO DE FIN DE GRADO

TÍTULO:

**DESARROLLO Y VALIDACIÓN DE UN TEST  
GENÉTICO PRECONCEPCIONAL MEDIANTE  
SECUENCIACIÓN MASIVA**

AUTOR:

**EMILIO FUSTER COLLADOS**

TUTORES:

**LUIS A. ALCARAZ MAS**

**PEDRO ROBLES RAMOS**

## DECLARACIÓN DE LOS TUTORES

D. Luis A. Alcaraz Mas, socio cofundador de la empresa Bioarray S.L. y D. Pedro Robles Ramos, Profesor Titular del Área de Genética del Instituto de Bioingeniería de la Universidad Miguel Hernández de Elche,

CERTIFICAN:

Que D. Emilio Fuster Collados, estudiante de cuarto curso del Grado en Biotecnología de la Universidad Miguel Hernández de Elche, ha realizado bajo nuestra dirección el trabajo de investigación titulado “Desarrollo y validación de un test genético preconcepcional mediante secuenciación masiva” incluido en la memoria de este Trabajo Fin de Grado. El presente trabajo se ha llevado a cabo en la empresa Bioarray S.L. y el Área de Genética del Instituto de Bioingeniería de la Universidad Miguel Hernández de Elche.

Y para que conste a todos los efectos oportunos, expiden y firman la presente certificación,

En Elche, septiembre de 2017

Fdo.:

Luis A. Alcaraz Mas

Fdo.:

Pedro Robles Ramos

## AGRADECIMIENTOS

*A Luis, por permitirme realizar este trabajo, por su paciencia y comprensión.*

*A Pedro, por su especial motivación y dedicación en todo momento. Sin su ayuda, este trabajo no sería lo que es ahora.*

*Al equipo de Bioarray, por su acogida e interés. En el laboratorio me he sentido como en casa.*

*A mis padres, Emilio y Belén, y a mi hermana, Marta, por no dejar de luchar junto a mí.*

*Y a Irene, mi compañera de vida, por volar conmigo.*



# Índice

## ÍNDICE DE MATERIA

<b>I. ABREVIATURAS.....</b>	<b>3</b>
<b>II. RESUMEN .....</b>	<b>4</b>
<b>III. INTRODUCCIÓN.....</b>	<b>5</b>
<b>III. 1. Nuevas técnicas de secuenciación (NGS) .....</b>	<b>7</b>
<b>III. 2. Cariotipo.....</b>	<b>9</b>
<b>IV. ANTECEDENTES .....</b>	<b>10</b>
<b>V. OBJETIVOS.....</b>	<b>12</b>
<b>VI. MATERIALES Y MÉTODOS.....</b>	<b>13</b>
<b>VI. 1. Secuenciación masiva .....</b>	<b>13</b>
VI. 1. 1. Pacientes control .....	14
VI. 1. 2. Extracción de ADN .....	14
VI. 1. 3. Cuantificación y determinación de la calidad y la pureza del ADN.....	14
VI. 1. 4. Diseño del panel de cebadores.....	14
VI. 1. 5. Amplificación mediante PCR múltiple de las regiones genómicas de interés.....	15
VI. 1. 6. Digestión parcial de los amplicones .....	15
VI. 1. 7. Ligación de adaptadores a los amplicones y purificación .....	16
VI. 1. 8. Amplificación y cuantificación de las muestras.....	17
VI. 1. 9. Generación del molde .....	17
VI. 1. 10. Enriquecimiento de las ISP .....	18
VI. 1. 11. Secuenciación de las muestras .....	19
<b>VI. 2. Análisis bioinformático primario.....</b>	<b>19</b>
<b>VI. 3. Análisis bioinformático de las variantes .....</b>	<b>20</b>
VI. 3. 1. Clasificación e interpretación de las variantes.....	20
VI. 3. 2. Generación de informes.....	22
<b>VI. 4. Cálculo del riesgo residual.....</b>	<b>22</b>
VI. 4. 1. Determinación de la frecuencia de portador .....	22
VI. 4. 2. Determinación de la tasa de detección de portador .....	23
VI. 4. 3. Determinación del riesgo residual .....	23
<b>VI. 5. Cariotipado .....</b>	<b>23</b>
VI. 5. 1. Extracción de cromosomas a partir de linfocitos de sangre periférica.....	24
VI. 5. 2. Preparación de la muestra.....	24
VI. 5. 3. Bandeado con tripsina y Giemsa .....	25
VI. 5. 4. Análisis del cariotipo.....	25
<b>VII. RESULTADOS .....</b>	<b>26</b>
<b>VII. 1. Análisis de las variantes.....</b>	<b>26</b>
<b>VII. 2. Cálculo del riesgo residual .....</b>	<b>29</b>
<b>VII. 3. Análisis de cariotipo .....</b>	<b>31</b>
<b>VIII. DISCUSIÓN.....</b>	<b>34</b>
<b>VIII. 1. Secuenciación .....</b>	<b>34</b>
<b>VIII. 2. Cálculo del riesgo residual.....</b>	<b>35</b>
<b>VIII. 3. Análisis de cariotipo.....</b>	<b>36</b>
<b>IX. CONCLUSIONES Y PROYECCIÓN FUTURA .....</b>	<b>37</b>
<b>X. BIBLIOGRAFÍA.....</b>	<b>38</b>

# Índice

## ÍNDICE DE TABLAS

<b>Tabla 1.</b> Clasificación de las variantes según ClinVar <sup>47</sup> .....	21
<b>Tabla 2.</b> Cálculos de los riesgos de ser pareja portadora y de tener un descendiente afecto para el gen <i>CFTR</i> .....	31
<b>Tabla 3.</b> Cálculos de los riesgos de ser pareja portadora y de tener un descendiente afecto para el gen <i>GJB2</i> .....	31
<b>Tabla 4.</b> Media de los porcentajes de las distintas variantes encontradas en los pacientes analizados .....	34

## ÍNDICE DE FIGURAS

<b>Figura 1.</b> Apartados del protocolo experimental usado en la secuenciación masiva .....	13
<b>Figura 2.</b> Explicación del uso de dos <i>pools</i> de cebadores .....	15
<b>Figura 3.</b> Proceso de purificación con <i>Agencourt AMPure XP</i> paso a paso .....	16
<b>Figura 4.</b> Tipos de micelas originales sometidas a la amplificación .....	18
<b>Figura 5.</b> Sección de la tabla de cálculos del riesgo residual recogidos en el anexo 2 .....	30
<b>Figura 6.</b> Cariotipo de la paciente C1 .....	32
<b>Figura 7.</b> Cariotipo del paciente C2 .....	32
<b>Figura 8.</b> Cariotipo del paciente C3 .....	33
<b>Figura 9.</b> Cariotipo del paciente C4 .....	33

## Abreviaturas

### I. ABREVIATURAS

A: Adenina.

ACMG: del inglés *American College of Medical Genetics and Genomics* (Colegio Americano de Genética y Genómica Médica).

ADN: ácido desoxirribonucleico.

ARN: ácido ribonucleico.

C: Citosina.

dNTP: del inglés *deoxyNucleotide TriPhosphate* (desoxirribonucleótido trifosfato).

Fig: Figura.

G: Guanina.

h: Horas.

HBSS: del inglés *Hank's Balanced Salt Solution* (solución salina equilibrada de Hank).

HiFi: del inglés *High Fidelity* (alta fidelidad).

ISP: del inglés *Ion Sphere Particles* (partículas esféricas Ion).

*Ion Torrent™ PGM™*: del inglés *Ion Torrent™ Personal Genome Machine™*.

µl: Microlitros.

min: Minutos.

Ng: Nanogramos.

NGS: del inglés *Next Generation Sequencing* (secuenciación de próxima generación).

nm: Nanómetros.

PCR: del inglés *Polymerase Chain Reaction* (reacción en cadena de la polimerasa).

PEG: del inglés *PolyEthylene Glycol* (polietilenglicol).

pH: del inglés *potential of Hydrogen* (potencial de hidrógeno).

pM: Picomolar.

s: Segundos.

SNP: del inglés *Single Nucleotide Polymorphism* (polimorfismo de un solo nucleótido).

T: Timina.

TFG: Trabajo de Fin de Grado.

VOUS: del inglés *Variant Of Unknown Significance* (variante de significado desconocido).

x g: del inglés *times gravity* (veces gravedad).

### II. RESUMEN

Este trabajo de fin de grado consiste en el desarrollo y la validación de un test genético preconcepcional para la empresa Bioarray S.L. Por un lado, se ha llevado a cabo la puesta a punto del panel de detección de portadores mediante secuenciación masiva en el equipo *Ion Torrent PGM™*. El panel incluye 298 genes cuyas mutaciones causan enfermedades recesivas de herencia autosómica y ligada al cromosoma X. Tras la obtención de las secuencias de cinco pacientes, se han empleado herramientas bioinformáticas para detectar las posibles variantes nucleotídicas patogénicas presentes en ellos. También se ha calculado el riesgo residual para la mayoría de los genes del panel, con el fin de comunicar a los pacientes todos los riesgos posibles, cumpliendo así con el consentimiento informado. Por otro lado, como complemento del estudio del genoma, se ha realizado la tradicional técnica del cariotipado en más de veinte pacientes, incluyendo los pacientes secuenciados, para la búsqueda de anomalías cromosómicas que puedan afectar al riesgo reproductivo en una futura gestación de los pacientes, o dicho de otra forma, la probabilidad de tener un descendiente afecto.

**Palabras clave:** test genético preconcepcional, panel de detección de portadores, secuenciación masiva, riesgo residual, cariotipado.

This final degree project consists in the development and validation of a preconceptional genetic test for the company Bioarray S.L. On the one hand, the development of the carrier detection panel by massive sequencing in the Ion Torrent PGM™ has been carried out. The panel includes 298 genes whose mutations cause recessive autosomic and linked to the X chromosome disorders. After obtaining the sequences of five patients, bioinformatic tools have been used for detecting the possible pathogenic nucleotide variants present in them. The residual risk for the majority of the panel genes has been calculated as well, with the purpose of communicating every possible risk to the patients, accomplishing with the informed consent. On the other hand, as a complement to the study of the genome, the traditional technique of karyotyping has been carried out in more than twenty patients, including the sequenced patients, in order to detect chromosomal anomalies which can affect their reproductive risk in a future pregnancy of the patients, or in other words, the probability of having an affect descendant.

**Keywords:** preconceptional genetic test, carrier detection panel, massive sequencing, residual risk, karyotyping.

### III. INTRODUCCIÓN

El asesoramiento genético reproductivo consiste, en términos generales, en la evaluación del riesgo genético asociado a la reproducción de la pareja que acude a consulta, la comunicación del riesgo a la pareja y, si es necesario, el consejo sobre sus opciones reproductivas<sup>1</sup>. Todo ello constituye un importante componente del cuidado preconcepcional y prenatal.

En los últimos años, se aprecia un énfasis creciente por parte de los gobiernos y de organizaciones profesionales en el diseño de una estrategia para la provisión de consejo y cuidado de futuros padres e individuos en edad reproductiva. Esto es debido a la incidencia persistente de embarazos adversos, anomalías congénitas y condiciones genéticas que han dado lugar a una demanda de consejo preconcepcional en el sistema sanitario<sup>2</sup>.

El consejo preconcepcional se centra en la identificación de potenciales riesgos genéticos de los padres y de sus descendientes<sup>3</sup>. También tiene en cuenta circunstancias previas al nacimiento de un descendiente enfermo, como el historial clínico de abortos, un diagnóstico genético confirmado de los futuros padres, y un historial familiar conocido de desórdenes genéticos. Además, es necesario considerar la raza étnica de los futuros padres, ya que las probabilidades de sufrir enfermedades recesivas varían en las distintas poblaciones<sup>4</sup>.

Las enfermedades autosómicas recesivas se caracterizan porque presentan un efecto que sólo es reconocible en homocigosis o heterocigosis compuesta<sup>5</sup>. Por lo tanto, si un individuo es heterocigótico (portador de un alelo mutado y uno silvestre), será un individuo portador sano. Si un sujeto hereda dos copias mutadas, estará afectado por la enfermedad recesiva. El riesgo de heredar ambos alelos mutados, de padres portadores sanos, es del 25%. Por otro lado, la probabilidad de que el descendiente sea un portador sano (heterocigótico), es del 50%<sup>6</sup>.

De entre las enfermedades recesivas, también cabe destacar las de herencia recesiva ligada al cromosoma X. Las mujeres que presentan un solo alelo mutado recesivo en uno de los genes que residen en el cromosoma X, suelen ser portadoras sanas de estos desórdenes ya que el alelo silvestre del otro cromosoma X compensa la deficiencia del primero. Sin embargo, los varones estarán afectados por la enfermedad si han heredado el alelo mutado del cromosoma X de su madre, ya que son hemocigóticos para dicho cromosoma<sup>7</sup>. De entre las características habituales del patrón de herencia autosómica recesiva ligada al cromosoma X, cabe destacar las siguientes: la incidencia de la enfermedad es mucho mayor entre los hombres que entre las mujeres; un hombre enfermo siempre transmite el alelo mutante a todas sus hijas, pero éstas no presentarán la enfermedad si la madre no está enferma o es portadora; una mujer portadora sana transmite el alelo mutante a la mitad de sus hijos varones, que estarán enfermos, y a la



## Introducción

mitad de sus hijas, que serán portadoras sanas siempre que el padre esté sano; el alelo mutante nunca se transmite de padre a descendiente varón<sup>8</sup>.

Cada uno de nosotros somos portadores de una media de una a dos mutaciones recesivas, que pueden causar desórdenes genéticos severos o muerte prenatal cuando se encuentran en homocigosis<sup>9</sup>.

En el asesoramiento genético, la detección de mutaciones en portadores sanos es particularmente relevante. En estas situaciones, la mayoría de alelos mutantes se encuentran presentes en individuos heterocigotos no afectados por la condición, que mayoritariamente no son conscientes de su condición de portador<sup>10,11</sup>. En un árbol genealógico, la herencia autosómica recesiva sigue las siguientes reglas: afecta habitualmente tanto a hombres como a mujeres, el rasgo puede aparecer en algunas generaciones y en otras no, y normalmente padres sanos engendran descendientes afectados<sup>12</sup>.

La frecuencia de portador es la probabilidad de ser heterocigótico para una variante patogénica que en homocigosis causa una enfermedad autosómica recesiva concreta en una población dada<sup>13</sup>. La fibrosis quística, la talasemia, la anemia falciforme y la enfermedad Tay-Sachs representan ejemplos de condiciones autosómicas recesivas cuyas frecuencias de portador varían según las poblaciones. Por ejemplo, la frecuencia de portador para mutaciones patogénicas en el gen *HBB* (que codifica para la cadena beta de globina de la hemoglobina), que causan la beta talasemia, es alta en países mediterráneos, y en Chipre se estima en 1 de cada 7 individuos. Sin embargo, para la población mundial, la estimación de la frecuencia de portador de la beta talasemia es de 1 de cada 10000 individuos<sup>14</sup>. Otro desorden de la hemoglobina que también es causado por una mutación en el gen *HBB* es la anemia falciforme, que presenta una frecuencia de portador alta en poblaciones de origen africano<sup>15</sup>. Por otro lado, las mutaciones en los genes *HBA1* y *HBA2* causan alfa talasemia, y son frecuentes en las poblaciones asiáticas del sudeste<sup>16</sup>. La frecuencia de portador en el gen *CFTR*, cuyas mutaciones causan fibrosis quística, es alta en individuos de raza caucásica de poblaciones europeas, con una estimación de 1 de cada 25 individuos<sup>17</sup> y para la enfermedad de Tay-Sachs, una condición metabólica causada por deficiencia de la enzima hexosaminidasa A que conduce a una neurodegeneración progresiva, es de 1 de cada 30 individuos con pasado judío askenazí<sup>18</sup>. Un ejemplo de condición recesiva ligada al cromosoma X es la hemofilia A, causada por la mutación del gen *F8*, que codifica para el factor de coagulación VIII. La frecuencia de portador para mutaciones patogénicas en este gen se estima en 1 de cada 36 mujeres, ya que los varones que hereden de su madre un alelo mutante del gen *F8*, estarán afectados por la hemofilia A<sup>19</sup>.

La severidad de estas condiciones y la relativamente alta frecuencia en poblaciones específicas ha provocado la oferta de test genéticos para identificar futuros padres que puedan ser heterocigotos para mutaciones patogénicas. La confirmación del estatus de portador en la

## Introducción

etapa preconcepcional, permite a los futuros padres tomar decisiones informadas sobre sus opciones reproductivas, ya que podrían optar a un diagnóstico genético preimplantacional, mediante el cual los embriones producidos por técnicas reproductivas asistidas puedan ser examinados antes de ser implantados en el útero<sup>20</sup>. También hay otras opciones para los progenitores portadores como por ejemplo el uso de gametos donantes, evitar tener descendientes o la adopción. Claramente, estas opciones no serían contempladas si no se llega a conocer la condición de portadores de los padres previamente al embarazo. Otra alternativa para los futuros padres, sería proceder con el embarazo y optar a un diagnóstico prenatal temprano, que en caso de confirmar la afección del feto, les diese opción a una interrupción voluntaria del embarazo o a prepararse para el nacimiento de su descendiente afecto<sup>21</sup>.

En cualquier caso, cuando se ofrece un test genético preconcepcional o prenatal, debe acompañarse siempre de un apropiado consejo genético a través de un profesional y estar basado en el principio del consentimiento informado<sup>22</sup>. El consejo abarca la descripción de las enfermedades que el test examina y sus posibles tratamientos disponibles, la descripción del test en sí y de cualquiera de sus limitaciones, incluyendo cualquier riesgo post-test para un resultado negativo, además de un informe detallado de todas las opciones reproductivas disponibles<sup>21</sup>. Esto es debido a que en todos los casos de fecundación *in vitro*, la ley obliga a realizar todo lo técnicamente posible para implantar en la madre un embrión sano.

En los últimos años, gracias al avance de las tecnologías genómicas basadas en las nuevas técnicas de secuenciación masiva, se ha provisto a los profesionales genéticos de un método eficaz y rápido para llevar a cabo test genéticos a gran escala<sup>23</sup>. El desarrollo de paneles, que permiten analizar múltiples genes, y la secuenciación del genoma han hecho posible el diseño de nuevas estrategias para la investigación genética en las etapas preconcepcionales y prenatales con el objetivo de permitir a los futuros padres tomar mejores decisiones informadas<sup>24</sup>.

### III. 1. Nuevas técnicas de secuenciación (NGS)

Desde el descubrimiento de la estructura del ADN<sup>25</sup>, se han realizado grandes avances técnicos que han permitido analizar la complejidad y diversidad del genoma. Los distintos métodos de secuenciación del ADN han permitido analizar genomas completos o parciales a nivel de su secuencia nucleotídica. El método de Sanger de terminación de cadena se desarrolló en 1975 y se estableció rápidamente como método de referencia de secuenciación del ADN, permitiendo secuenciar fragmentos de hasta 900 pares de bases<sup>26</sup>. La técnica se fue optimizando hasta desarrollarse los secuenciadores automáticos que permitían analizar simultáneamente hasta 96 muestras de ADN con una longitud de lectura de 500 a 1000 pares de bases, lo que supuso la obtención de mayores cantidades de secuencia en menos tiempo y permitió secuenciar los primeros genomas. Sin embargo, el coste por megabase secuenciada encarece la obtención

## Introducción

de la secuencia completa de genomas, lo que impulsó el desarrollo de nuevas técnicas de secuenciación masiva, que también permitiesen la secuenciación del genoma de individuos concretos y así establecer la genómica personalizada. A día de hoy, el coste de la secuenciación del genoma humano ronda los 1000 € (teniendo en cuenta sólo el coste de los reactivos) y frecuentemente ya es utilizada como herramienta clínica<sup>27</sup>. Estas técnicas son capaces de realizar múltiples secuenciaciones de forma paralela, produciendo millones de secuencias de numerosos genes al mismo tiempo. Las ventajas de la secuenciación masiva respecto a la convencional son la amplitud del análisis, el uso de menor cantidad de material genético de partida, un menor coste y una mayor rapidez en la obtención de secuencia<sup>28</sup>.

Las nuevas técnicas de secuenciación masiva se clasifican en de secuenciación de lecturas cortas o largas<sup>29</sup>. Dentro de la secuenciación de lecturas cortas se diferencian dos tipos; la secuenciación por ligación y la secuenciación por síntesis. En ambas técnicas se produce la secuenciación paralela, generándose millones de copias de moléculas de ADN diferentes simultáneamente. En la secuenciación por ligación, una sonda que está unida a un fluoróforo hibrida con un fragmento de ADN y se une a un oligonucleótido adyacente para dar una imagen. La emisión del espectro del fluoróforo indica la identidad de la base o bases complementarias para posiciones específicas dentro de la sonda<sup>30</sup>. Esta técnica está prácticamente obsoleta, ya que ha sido sustituida por los otros métodos de secuenciación masiva. Por otro lado, la secuenciación de lecturas cortas por síntesis usa una polimerasa de ADN y mide las señales producidas por un fluoróforo (*Illumina*) o por un cambio de pH (*Ion Torrent PGM*), identificando la incorporación de un nucleótido concreto en la nueva cadena sintetizada<sup>31</sup>.

La plataforma *Pacific Biosciences*<sup>28</sup> realiza secuenciación de lecturas largas mediante la secuenciación a tiempo real de una única molécula de ADN (SMRT) y el uso de sondas guía modo-cero (ZMW) y nucleótidos fosfoligados. En la ZMW se fija una polimerasa de ADN con una cadena de ADN molde. Cada una de las cuatro bases nitrogenadas del ADN (A, C, G y T) se marcan con un fluoróforo diferente. Así, cuando se incorpora un nucleótido a la nueva cadena, se libera un fluoróforo y se observa una señal que será medida por un detector.

De todas las nuevas tecnologías de secuenciación masiva, la del *Ion Torrent PGM* traduce directamente la información químicamente codificada en las cuatro bases del ADN en información digital (0, 1) de un chip semiconductor, que sirve como soporte en el que se lleva a cabo la secuenciación. El resultado de esta secuenciación es más simple, rápido y eficaz que cualquier otra tecnología de secuenciación<sup>32</sup>. El secuenciador detecta los cambios de pH que se producen cuando un nucleótido concreto se incorpora a la molécula de ADN que se está sintetizando y los traduce en información digital al sistema informático. Cuando en la hebra de ADN a secuenciar hay dos bases idénticas adyacentes, el voltaje producido por el cambio de pH será doble, y el chip será capaz de detectar ambas bases. Debido a que se trata de un sistema de

## Introducción

detección directa, cada incorporación de un nucleótido se traduce en segundos<sup>33</sup>. Una vez finalizado el proceso de secuenciación, se procede a la detección y análisis de las variantes nucleotídicas. Para este análisis, es necesario el uso de la Bioinformática, que permite el empleo de herramientas computacionales para el estudio y la gestión de datos biológicos. Gracias a los análisis bioinformáticos, la secuenciación masiva se puede aplicar con éxito al diagnóstico genético.

### III. 2. Cariotipo

Si bien las técnicas de secuenciación masiva permiten detectar cambios nucleotídicos en el ADN analizado, las técnicas clásicas de citogenética como el cariotipado, se siguen utilizando en el diagnóstico genético ya que permiten identificar cambios importantes en la estructura o número de los cromosomas. Así pues, el análisis de cromosomas mediante la realización de un cariotipo, es una técnica convencional ampliamente usada para la detección de la inestabilidad de los mismos. Para dicho análisis se usa el bandeo G, que recibe su nombre por el colorante empleado, Giemsa, que genera un patrón de bandas claras y oscuras en los cromosomas y fue desarrollado por Gustav Giemsa en 1948<sup>34</sup>. Cuando las mutaciones involucran a grandes partes de los cromosomas, y se pueden observar con el microscopio, se denominan aberraciones cromosómicas. El cariotipado permite detectar células con aberraciones cromosómicas numéricas y/o estructurales<sup>35</sup>. En humanos, las aberraciones cromosómicas son muy comunes, ya que afectan a un 7,5% de todos los embarazos. Sin embargo, la mayoría de ellas ocasionan abortos espontáneos, si bien un 0,6% de los neonatos presentan una aberración cromosómica. De esta manera, el 60% de los abortos espontáneos tempranos son causados por una aberración cromosómica, así como el 5% de los abortos espontáneos tardíos y el 5% de los partos de feto muerto<sup>36</sup>. Con toda esta información, se puede afirmar que la técnica del cariotipado es un buen complemento de la secuenciación masiva a la hora de desarrollar un test genético preconcepcional.

### IV. ANTECEDENTES

Como se ha indicado en la introducción, cuando ambos padres son portadores de un alelo mutante causante en homocigosis o heterocigosis compuesta de una enfermedad autosómica recesiva, presentan un riesgo del 25% de concebir un descendiente afecto en cada embarazo. Se estima que entre una y dos parejas de cada cien en la población general tienen dicho riesgo<sup>37</sup>. Un test genético preconcepcional, con el que se detecten mutaciones para desórdenes recesivos, permite conocer a los miembros de las parejas si son portadores, y por lo tanto, les facilita la toma de decisiones reproductivas, como la aceptación del riesgo o la abstención de tener descendientes, el diagnóstico prenatal, el diagnóstico genético preimplantacional, usar gametos donantes o la adopción.

La detección de portadores se lleva a cabo en pocos países actualmente y en ocasiones, su realización está basada en el pasado ancestral de determinadas poblaciones. Por ejemplo, en la comunidad judía askenazí desde 1970 es frecuente llevar a cabo el test de detección de portadores para la enfermedad Tay-Sachs. En Chipre, la detección de portadores para la  $\beta$ -talasemia ha conseguido disminuir la frecuencia de dicha enfermedad. En la población caucásica es frecuente realizar la detección de portadores para la fibrosis quística<sup>38</sup>.

Gracias al desarrollo de las nuevas técnicas genómicas, ahora es posible realizar un test de detección para múltiples enfermedades recesivas simultáneamente. Estos tests se denominan paneles de detección de portadores, permiten la detección de variantes genéticas con alto riesgo para la población de interés y en la actualidad ya están siendo desarrollados y ofrecidos comercialmente.

En España, poco a poco se están implantando los paneles de detección de portadores. La técnica previa a la secuenciación masiva, y que se sigue utilizando en muchas empresas, son los *arrays* (matrices) basados en SNP. Estos *arrays* contienen sondas de oligonucleótidos que se corresponden a variantes alélicas de SNP seleccionados para su estudio. Un SNP es un polimorfismo de un solo nucleótido, una variante genética común en una población y sin consecuencias relevantes para la salud<sup>39</sup>. Para que una variante se considere un SNP, debe darse en al menos un 1% de la población, de manera que si su frecuencia es menor del 1%, pasa a considerarse una mutación puntual. La técnica de los *arrays* se utiliza habitualmente para poder encontrar mutaciones puntuales responsables de las variantes patogénicas y probablemente patogénicas más comunes de la población. Dichas variantes afectan a la correcta función del gen, causando una enfermedad. La mayor desventaja de los *arrays* frente a los paneles de secuenciación es que únicamente se analizan las variantes de las enfermedades con mayor

## Antecedentes

incidencia en la población. Sin embargo, la secuenciación masiva del panel de portadores permite analizar una mayor cantidad de genes y de manera completa.

La utilización de los paneles de portadores tiene una fuerte implicación legal. En julio de 2016, una condena al Instituto Valenciano de Infertilidad (IVI) por no informar de todos los riesgos posibles en un proceso de reproducción asistida, supuso a partir de ese momento la utilización masiva de paneles de detección de portadores por parte de las empresas dedicadas a la reproducción asistida. La condena no se basaba en el hecho de que no se detectara la condición de portador del ovocito donado, en este caso de hemofilia, sino en que no se facilitó información suficientemente detallada a los padres sobre los riesgos posibles en el proceso y los métodos utilizados para minimizar estos riesgos. Cabe destacar, que para este servicio, hay que firmar el consentimiento informado, en el que se asegura que: “se han realizado todas las pruebas necesarias para descubrir y excluir las enfermedades genéticas, infecciosas o hereditarias que se pueden detectar según el estado de la ciencia”. Para entonces, la hemofilia ya era perfectamente detectable mediante el panel de detección de portadores, al igual que muchas otras enfermedades.

La empresa Bioarray S.L. fundada en 2008, decidió ofrecer el servicio de diagnóstico preconcepcional en 2016, y su puesta a punto se ha llevado a cabo mediante la realización de este trabajo de fin de grado (TFG), utilizando la secuenciación masiva del panel de portadores y complementariamente, la técnica del cariotipado.

Previamente, Bioarray S.L. ofrecía a las clínicas de reproducción asistida un test genético para donantes de óvulos y espermatozoides. El test incluía el análisis del cariotipo, la detección del síndrome X frágil<sup>40</sup> y la secuenciación de las 50 mutaciones más frecuentes del gen *CFTR* (*Cystic Fibrosis Transmembrane conductance Regulator*), causantes de la fibrosis quística<sup>17</sup>. Además, si la clínica solicitante lo demandaba, también se realizaba el test de detección de la delección del exón 7 del gen *SMN1* (*Survival Motor Neuron 1*), causante de la atrofia muscular espinal<sup>41</sup>.

Bioarray S.L. pretende ofrecer el nuevo test genético preconcepcional puesto a punto a través de este TFG a:

- La población en general: parejas sin antecedentes genéticos que desean conocer su riesgo de concebir un descendiente afecto por una enfermedad autosómica recesiva.
- Parejas en la que exista consanguinidad.
- Parejas que vayan a someterse a procedimientos de reproducción asistida, como complemento de los estudios diagnósticos.
- Donantes de espermatozoides u ovocitos, para la mejor selección de éstos y la disminución del riesgo de transmisión de enfermedades recesivas.

# Objetivos

## V. OBJETIVOS

El objetivo principal de este trabajo es el desarrollo y la validación de un test genético preconcepcional en la empresa Bioarray S.L. El test consiste, por un lado, en un estudio del genoma de los pacientes para encontrar variantes patogénicas y probablemente patogénicas mediante un panel de detección de portadores, y por otro, en el análisis del cariotipo para detectar anomalías cromosómicas estructurales y numéricas.

Su finalidad es poder ofrecer al público un diagnóstico genético preconcepcional, para que futuras parejas conozcan la probabilidad de tener un descendiente afecto por una enfermedad recesiva o una aberración cromosómica. En el caso de resultado positivo, la pareja puede optar por un diagnóstico genético preimplantacional y por último, asistir a una clínica de reproducción asistida.

Así mismo, los objetivos específicos a llevar a cabo durante la realización de este TFG han sido:

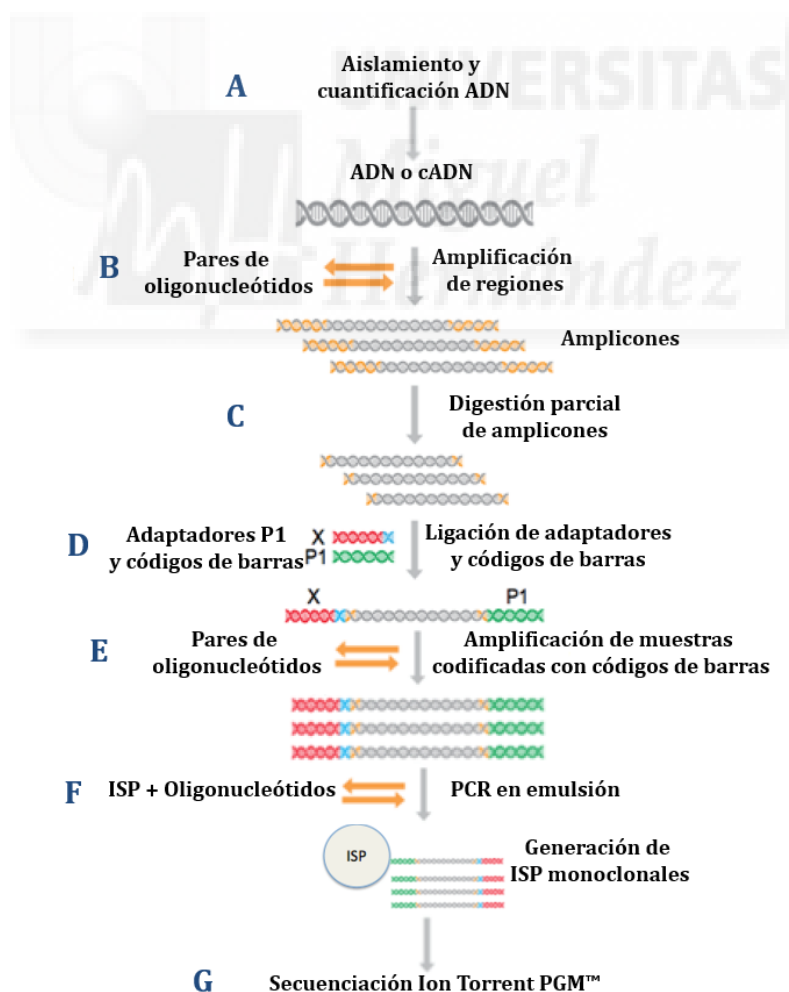
- La puesta a punto del panel de detección de portadores mediante secuenciación masiva con la plataforma *Ion Torrent PGM™*.
- El análisis bioinformático de las secuencias obtenidas mediante la secuenciación.
- El cálculo de los distintos riesgos para todas las situaciones posibles.
- La realización de la técnica del cariotipo y el análisis del mismo, como complemento al panel de detección de portadores.

### VI. MATERIALES Y MÉTODOS

El test genético preconcepcional desarrollado en este trabajo, incluye un análisis del genoma del paciente desde dos aproximaciones complementarias. Por una parte la secuenciación masiva del panel de portadores, que implica la extracción de ADN a partir de sangre, su secuenciación y el análisis de las secuencias obtenidas, incluyendo el cálculo del riesgo residual. Por otra, la técnica del cariotipado, mediante la cual a partir de una muestra de sangre se analiza con un microscopio el número y estructura de los cromosomas.

#### VI. 1. Secuenciación masiva

La plataforma de secuenciación masiva utilizada para el desarrollo y puesta a punto del test genético ha sido la del *Ion Torrent PGM™*. El protocolo de secuenciación utilizado en este TFG está basado en el *Ion AmpliSeq™ DNA Library Preparation*, ya que hemos modificado algunos de sus pasos para lograr una secuenciación óptima. El protocolo se explica a grandes rasgos en la figura 1.



**Figura 1. Apartados del protocolo experimental usado en la secuenciación masiva.** Se representan los diferentes pasos de la letra A a la G. Imagen modificada del protocolo *Ion AmpliSeq™ DNA and RNA library preparation*.



## Materiales y métodos

### VI. 1. 1. Pacientes

Para la puesta a punto del panel de portadores, la empresa seleccionó cinco muestras de sangre de pacientes de raza caucásica de población europea, siendo una paciente mujer, donante de óvulos, y los cuatro restantes pertenecientes a dos parejas, que habían acudido a la empresa para demandar otro tipo de test.

### VI. 1. 2. Extracción de ADN

La extracción de ADN (Fig.1-A) se realiza con el *Maxwell™ Tissue DNA Purification Kit* (Promega) a partir de sangre del paciente. La sangre se somete a un tratamiento con proteasas y ARNasas en el correspondiente tampón del kit con el objetivo de lisar las células y eliminar el ARN. Una vez lisadas las células, el ADN se somete a un proceso de lavado automatizado utilizando el equipo *Maxwell* (Promega). Éste último utiliza un sistema de esferas electromagnéticas para la captura de ADN a partir del extracto de células, el cual posteriormente se somete a una serie de lavados para eliminar todas las impurezas procedentes de la extracción: tampones, proteínas y otros restos orgánicos.

### VI. 1. 3. Cuantificación y determinación de la pureza del ADN

La cuantificación del ADN (Fig.1-A) se lleva a cabo mediante un *NanoDrop* (Thermo Fisher). Éste permite la cuantificación directa mediante la medición de absorbancia a 260 nm.

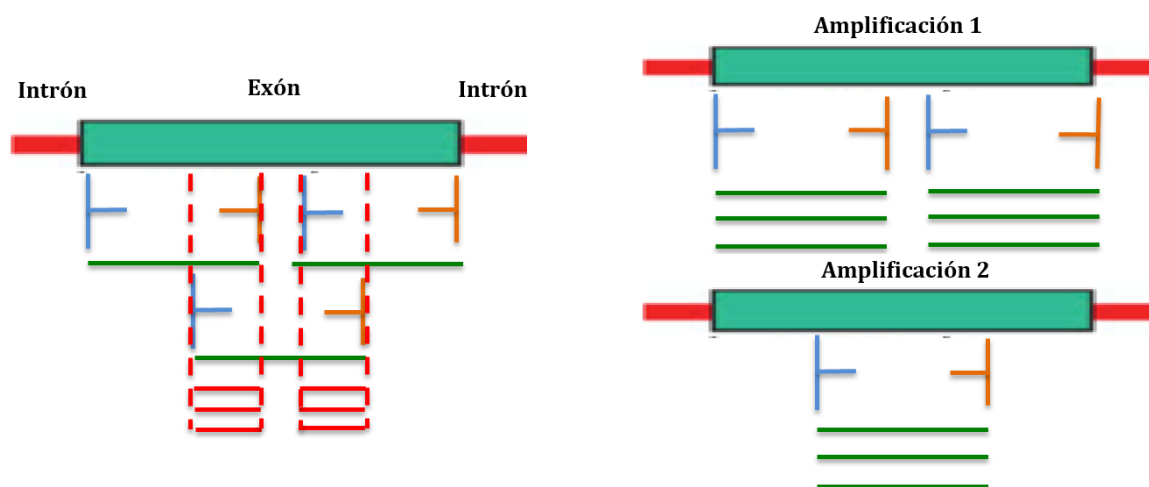
La pureza del ADN también se determina con el *NanoDrop*, estableciendo las relaciones de absorbancia 260/280 y 260/230 nm de la muestra, de modo que valores cercanos a 1,8 y 2,0 respectivamente, son indicativos de una correcta pureza del ADN.

### VI. 1. 4. Diseño del panel de cebadores

El panel de oligonucleótidos usado en esta memoria consta de 298 genes cuyas mutaciones causan diferentes desórdenes recesivos autosómicos o recesivos ligados al cromosoma X, e incluye un total de 6952 parejas de oligonucleótidos. El listado de los genes del panel se recoge en el Anexo 1, debido a que por su volumen si fuera incluido en esta memoria, el número de páginas de este documento excedería el máximo permitido por normativa. Thermo Fisher Scientific permite que el investigador diseñe su propio panel (<https://www.ampliseq.com/browse.action>). El diseño específico de cada pareja de oligonucleótidos fue realizado por los bioinformáticos de LifeTechnologies™ ya que el tamaño del panel era demasiado grande. Los paneles de Thermo Fisher se caracterizan por dividir los oligonucleótidos en varios conjuntos (*pools*). En este caso, el panel de portadores está compuesto por 2 *pools* de cebadores, cada uno con 3476 pares de oligonucleótidos. La finalidad de esta división es que los cebadores no solapen en la amplificación y no se produzcan

## Materiales y métodos

productos de amplificación (amplicones) no deseados. Este problema es habitual en la secuenciación masiva, y se ejemplifica en la figura 2.



**Figura 2. Explicación del uso de dos *pools* de cebadores.**  $\leftarrow$  representa el cebador *forward* de un par de oligonucleótidos y  $\rightarrow$  el *reverse*. A la izquierda, la amplificación se lleva a cabo con un solo *pool* de cebadores. Estos solapan en la región señalada con líneas discontinuas verticales rojas, generándose en la amplificación dos tipos de productos: los deseados en verde, y los no deseados en rojo, productos pequeños amplificados en gran cantidad. A la derecha, se llevan a cabo dos amplificaciones, cada una con un *pool* de cebadores distinto. En cada amplificación los cebadores no solapan, y los productos esperados son los únicos obtenidos. Por esta razón, se usan dos *pools* de cebadores en lugar de solo uno.

### VI. 1. 5. Amplificación mediante PCR múltiple de las regiones genómicas de interés

Primero, se preparan las reacciones de amplificación, añadiendo 5  $\mu$ l de 5x *Ion AmpliSeq™ HiFi Mix*, que es una mezcla que contiene dNTP y una *Taq* polimerasa de ADN de alta fidelidad, ADN del paciente a concentración 100 ng/ $\mu$ l y agua libre de nucleasa hasta alcanzar 12,5  $\mu$ l de volumen, creando una mezcla maestra o *master mix*. Se añaden 5  $\mu$ l de esta *master mix* a 5  $\mu$ l de cada *pool* de cebadores, quedando un volumen total de 10  $\mu$ l.

Seguidamente, con las muestras se ejecuta el siguiente programa en un termociclador (Fig.1-B): una primera etapa de 2 min a 99°C, en la que se produce la activación de la enzima y a continuación 10 ciclos de dos pasos; 15 s a 99°C (desnaturalización) y 8 min a 60°C (hibridación y extensión del cebador).

Una vez acabada la PCR, se combinan las reacciones amplificadas de los dos *pools* de oligonucleótidos, alcanzándose un volumen total de 20  $\mu$ l.

### VI. 1. 6. Digestión parcial de los amplicones

A los 20  $\mu$ l de la amplificación se añaden 2  $\mu$ l de reactivo FuPa, que contiene según su patente; una o más enzimas que degradan uracilo, como por ejemplo la uracilo ADN glicosidasa, formamidopiridina-ADN glicosilasa y una ADN polimerasa como pol1. En el termociclador se sigue el siguiente programa (Fig.1-C): 20 min a 50°C, 20 min a 55°C y 20 min a 60°C. La digestión parcial es un paso importante, ya que degrada los cebadores unidos al ADN amplificado, de

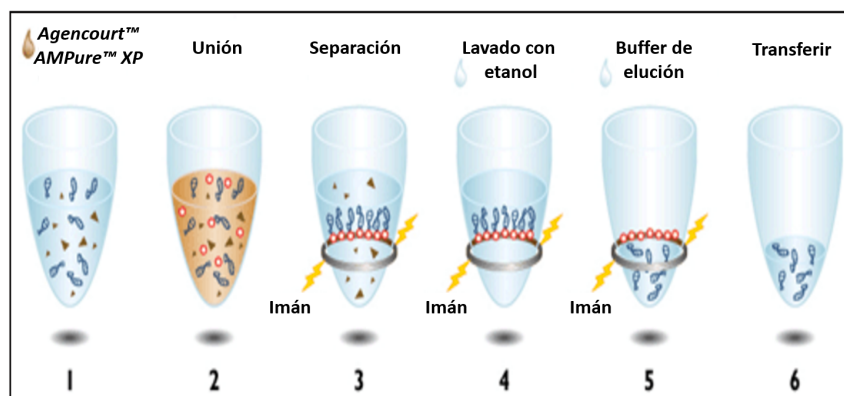
## Materiales y métodos

manera que posteriormente en la secuenciación, los cebadores no enmascararán mutaciones nucleotídicas que puedan existir en las posiciones de hibridación de los cebadores.

### VI. 1. 7. Ligación de adaptadores a los amplicones y purificación

El protocolo usado está diseñado para que en una misma reacción de secuenciación masiva se pueda secuenciar el ADN de dos o más pacientes, y para ello se añade al ADN de cada paciente una secuencia distintiva, o código de barras. El código de barras es una secuencia de 6 a 14 nucleótidos que se une a los amplicones de cada paciente en el extremo 3', de modo que cuando éstos se secuencian, el secuenciador reconoce el código de barras, y por lo tanto, el paciente del que se trate. Por otro lado, por el extremo 5' se une a los amplicones el adaptador P1 de 17 nucleótidos de longitud. Tanto el código de barras como el P1 servirán como secuencia complementaria para los oligonucleótidos posteriormente utilizados en amplificación. Para poder unir éstos a los amplicones (Fig.1-D), se prepara una *master mix* con 2 µl de adaptador *Ion P1*, 2 µl de *Ion Xpress™ Barcode*, y 4 µl de agua libre de nucleasa, añadiendo 2 µl de esta *master mix* a los amplicones. Para ligar los adaptadores a los amplicones de la muestra se añaden 2 µl de ADN ligasa y se ejecuta el siguiente programa en el termociclador: 30 min a 22°C, 5 min a 68°C y 5 min a 72°C.

Para la purificación de las muestras (Figura 3), se usa el reactivo *Agencourt™ AMPure™ XP*, que contiene bolas paramagnéticas denominadas *Beckman beads*. Cada bola está hecha de poliestireno recubierto por una capa de magnetita, a su vez cubierta con moléculas carboxílicas. Esto permite al ADN unirse reversiblemente a las bolas en presencia de PEG y sales. El PEG provoca que el ADN con carga negativa se pueda unir a los grupos carboxilos de la superficie. Finalmente, se realizan dos lavados con etanol al 70%. Tras cada lavado, las bolas se dirigen hacia los laterales del tubo con la ayuda de un rack magnético, para que las impurezas se separen de la muestra. Por último, se añade tampón de elución para separar el ADN de las bolas y poder transferir la muestra a un nuevo tubo.



**Figura 3. Proceso de purificación con *Agencourt AMPure XP* paso a paso.** Los elementos azules corresponden a los amplicones de los pacientes. Las bolas rojas son las *Beckman beads*. Las formas marrones son impurezas. Imagen modificada de <http://www.beckman.com/nucleic-acid-sample-prep/purification-clean-up/pcr-purification>.

## Materiales y métodos

### VI. 1. 8. Amplificación y cuantificación de las muestras

Primeramente, es necesario amplificar las muestras (Fig.1-E) para obtener suficiente material para una cuantificación precisa. Para la reacción de amplificación se usa la mezcla *Platinum™ PCR HiFi*, que contiene una ADN *Taq* polimerasa, sales, magnesio y dNTP, además de una mezcla de cebadores de secuencia complementaria a los códigos de barras y los adaptadores P1. La mezcla *Platinum™* permite realizar una amplificación con baja tasa de errores y en la mitad de tiempo. El programa utilizado en el termociclador para la amplificación es el siguiente: una primera etapa de activación de la enzima durante 2 min a 98°C, y a continuación 5 ciclos de dos pasos; 15 s a 98°C para la desnaturalización del ADN y 1 min a 64°C, para la hibridación de los cebadores y su extensión.

A continuación, se procede a una nueva purificación de las muestras amplificadas con *AgentCourt™ AMPure™ XP*. Ésta se realiza en dos rondas; la primera con una razón de volumen de reactivo/muestra de 0,5, donde el ADN de alto peso molecular se unirá a las bolas del reactivo, siendo éste en su mayor parte productos no deseados de la amplificación, mientras que los amplicones y oligonucleótidos permanecen en la solución, que es conservada. La segunda se realiza con una razón 1,2, donde los amplicones se unirán a las bolas y los oligonucleótidos permanecerán en la solución, por lo tanto, se guarda el precipitado. Al igual que en la purificación previa, se realizan dos lavados con etanol al 70%. Ya que el etanol puede inhibir la posterior secuenciación, es importante asegurarse de que se evapore sin llegar al secado en exceso. Una vez realizados los lavados, se añade tampón *Low TE*, que se utiliza para liberar el ADN de las bolas y resuspender las muestras finales purificadas. Con el rack magnético se separan las bolas para poder transferir el sobrenadante a un tubo Eppendorf, que contiene las muestras amplificadas de los pacientes.

Finalmente, se realiza la cuantificación mediante la *TapeStation*, determinando la molaridad de cada muestra para poder igualar concentraciones y combinarlas en un *pool*. La molaridad ideal de ADN de cada muestra es alrededor de 1000 pM. Una vez combinadas las muestras, se analiza de nuevo el *pool* mediante la *TapeStation* para conocer su molaridad, ya que posteriormente habrá que diluirlo.

### VI. 1. 9. Generación del molde

Para generar el molde, que será posteriormente secuenciado, se lleva a cabo una PCR en emulsión (Fig.1-F). Para ello, el *pool* de las muestras de pacientes se diluye a una concentración de 13 pM. Los reactivos y sus volúmenes usados son: 800 µl de la mezcla *Ion PGM™ Hi-Q Reagent*, que contiene dNTP, cebadores y tampón, 50 µl de la mezcla *Ion PGM™ Hi-Q Enzyme*, que incluye una ADN *Taq* polimerasa, *Ion PGM™ Hi-Q View ISP (Ion Sphere Particles)*, que son partículas esféricas con carga positiva, a las que se unen las muestras de ADN, con carga negativa, por su

## Materiales y métodos

extremo 5' ocupado por el adaptador P1, 25  $\mu$ l del *pool* de las muestras de ADN y 25  $\mu$ l de agua libre de nucleasa. La PCR en emulsión se realiza en el instrumento *Ion OneTouch™ 2* (Anexo 3, figura 1) y para que la amplificación clonal se produzca, se añade aceite de reacción, que permite la formación de tres tipos de micelas en medio acuoso (Figura 4).



**Figura 4. Tipos de micelas originales sometidas a la amplificación.** Las bolas negras son las ISP, mientras que las cadenas unidas a éstas, son diferentes secuencias de ADN de los pacientes.

Tras la PCR en emulsión tendremos de nuevo tres tipos de micelas:

- **Monoclonales:** micelas que incluyen una única ISP que se encuentra asociada con millones de copias debido a la amplificación de un solo tipo de molécula de ADN.
- **Policlonales:** micelas que incluyen una única ISP con varias moléculas de ADN unidas. Cada tipo de molécula de ADN unida a la ISP presentará millones de copias.
- **No clonales:** micelas que incluyen varias ISP a las que no se han unido moléculas de ADN.

Las micelas de interés para la posterior secuenciación son las de tipo monoclonal, ya que incluirán millones de copias de un solo tipo de moléculas de ADN. Para favorecer la formación de estas micelas, previamente a la PCR, hemos disminuido la cantidad de ADN, utilizando en este trabajo 13 pM en lugar de 25 pM como recomienda el protocolo original de PCR. Al añadir menor cantidad de ADN y muchas ISP, se favorece la formación de micelas monoclonales y se evitan las policlonales. Sin embargo, siempre permanece un pequeño porcentaje de ISP policlonales, que el secuenciador no será capaz de leer. Las no clonales tampoco son de interés y consecuentemente, se lleva a cabo un lavado de las mismas en el siguiente paso.

### VI. 1. 10. Enriquecimiento de las ISP

Tras la PCR en emulsión, hay que recuperar las ISP. Para ello se realiza un lavado con solución de lavado y una posterior centrifugación durante 2,5 min a 15000 x g. Cabe destacar que el código de barras del extremo 3' está biotinilado. El enriquecimiento de las ISP se realiza en el instrumento *Ion OneTouch™ ES* y se utilizan *Dynabeads™ MyOne™ Streptavidin C1*, que son bolas superparamagnéticas de 1  $\mu$ m de diámetro, recubiertas de una monocapa de

## Materiales y métodos

estreptavidina unida covalentemente a la superficie. Las *Dynabeads™* forman el complejo estreptavidina-biotina con el código de barras de las ISP monoclonales y policlonales. Dicho complejo resulta ser una de las interacciones no covalentes más fuertes que se conocen y es capaz de unirse a un imán del *Ion OneTouch™ ES* (Anexo 3, figura 2), que permite el lavado de las ISP no clonales mediante detergente *Tween* y *Wash solution*.

### VI. 1. 11. Secuenciación de las muestras

El instrumento utilizado para la secuenciación (Fig.1-C) es el *Ion Torrent PGM™* (Anexo 3, figura 3), basado en la detección de protones liberados en el proceso de replicación de ADN (véase el apartado III. 1. de la introducción). A partir de las ISP, que contienen el ADN amplificado de los pacientes, se añaden los oligonucleótidos de secuenciación, de secuencia complementaria al código de barras y al adaptador P1, para que posteriormente, la ADN polimerasa pueda actuar. Para la hibridación de los oligonucleótidos a las muestras, se ejecuta el siguiente programa en el termociclador: 95°C durante 2 min y 37°C durante 2 min. Cuando acaba el proceso de hibridación, se añade la ADN polimerasa de secuenciación y se procede a la carga del chip, que es el soporte físico donde se lleva a cabo la secuenciación. El chip usado para este trabajo es el *Ion 316™ Chip v2 BC* (Anexo 3, figura 3), que arroja aproximadamente de 2 a 3 millones de lecturas, lo que es suficiente para secuenciar dos pacientes con cobertura adecuada. Dicho chip está formado por micropocillos, en los que se cargan las ISP. Tras la carga y el llenado del chip, éste se introduce en el *Ion Torrent™ PGM™* y se realiza la reacción de secuenciación, que tiene una duración de 5 h. Una vez acabada la secuenciación, las señales generadas son transmitidas directamente al ordenador, con el que se procede al análisis bioinformático.

### VI. 2. Análisis bioinformático primario

Antes de proceder al análisis de datos y obtener conclusiones, es importante realizar un control de calidad para asegurarse de que el proceso de secuenciación ha transcurrido de manera correcta y sin errores que puedan afectar al estudio de los datos. Para este control, se emplea el software *Torrent Suite™* y lo realiza automáticamente el secuenciador.

Por otro lado, también se hace un alineamiento de las secuencias obtenidas con el genoma de referencia. En la empresa Bioarray S.L., se emplea como referencia del genoma humano la versión GRCh37/hg19. Este paso también está automatizado por el secuenciador y el software utilizado es el *Ion Torrent TMAP™*.

Por último, antes de proceder al análisis bioinformático, se emplea el software *Torrent Variant Caller* para la generación del archivo de variantes de los datos procedentes de la secuenciación.

## Materiales y métodos

### VI. 3. Análisis bioinformático de las variantes

Para el análisis de las variantes secuenciadas, se emplean diferentes herramientas bioinformáticas que han sido desarrolladas por y para la investigación científica y la secuenciación clínica de la empresa, siendo dichas herramientas propiedad de Bioarray S.L.

Estas herramientas bioinformáticas permiten anotar funcionalmente las variantes genómicas detectadas, en este caso, en el panel de detección de portadores, mediante el uso de distintas bases de datos públicas y privadas. El empleo de estas herramientas permite diferentes tipos de anotación que se explican a continuación:

- Anotación por gen, mediante el estándar de la HGVS<sup>42</sup> (*Human Genome Variation Society*) que proporciona información sobre el gen en el que se localiza la variante, el cromosoma dónde se encuentra y su posición en éste, los distintos transcritos en los que aparece, el tipo de variante (sinónima, no codificante, etc.), su localización en el gen (exón, intrón, secuencias del splicing, regiones UTR no traducidas), el cambio nucleotídico y aminoacídico que produce, y la frecuencia de la variante en el paciente, indicando si se encuentra en homocigosis o en heterocigosis. Además, esta base de datos proporciona información clínica sobre el gen y el fenotipo que causa su mutación según otras bases de datos como OMIM<sup>43</sup>, Orphanet<sup>44</sup> y Biomart<sup>45</sup>.
- Anotación por filtro, mediante dbSNP<sup>46</sup>, ClinVar<sup>47</sup>, ExAC<sup>48</sup>, Proyecto 1000 genomas<sup>49</sup> y Comparative Genomics<sup>50</sup>, que proporcionan información sobre la frecuencia alélica de la variante según los pacientes de cada base de datos, además de indicar el número de homocigotos con la variante encontrada según ExAC<sup>48</sup>. También se emplean predictores *in silico* de la patogenicidad de los cambios aminoacídicos (SIFT<sup>51</sup>, PolyPhen<sup>52</sup>, MutationTaster<sup>53</sup>) y de la conservación de residuos entre especies (PhyloP<sup>54</sup>, SiPhy<sup>55</sup>).
- Por último, estas herramientas bioinformáticas proporcionan información sobre la cobertura, definida como el número de lecturas que produce el secuenciador de cada posición nucleotídica, y la calidad de las lecturas obtenidas para cada variante, que resultan de gran utilidad para filtrar las variantes de cara a su posterior clasificación.

#### VI. 3. 1. Clasificación e interpretación de las variantes

Con la ayuda de las herramientas bioinformáticas de la empresa, se procede a clasificar las diferentes variantes genómicas detectadas en secuenciación. Para ello, se hace uso de distintos filtros que permiten reducir significativamente el número de variantes:

## Materiales y métodos

- Filtro de calidad: se aplica para descartar todas las variantes cuya cobertura obtenida por el secuenciador sea menor a 10X. Se considera que un número de lecturas tan bajo no es suficiente para realizar un buen análisis.
- Filtro de frecuencia alélica: se aplica para clasificar como benignas aquellas variantes que superen el 5% de frecuencia alélica en la base de datos Proyecto 1000 genomas<sup>49</sup>, siempre y cuando no hayan sido descritas antes como patogénicas.
- Filtro de secuencias no codificantes: se aplica para clasificar como VOUS (*Variant of Unknown Significance*) aquellas variantes que sean intrónicas o no codificantes y que estén alejadas  $\pm 6$  nucleótidos de la zona codificante, y no aparezcan descritas en las bases de datos o bibliografía científica. Aquellas variantes que se encuentran en una posición  $\pm 5$  nucleótidos de la zona codificante sí que serán analizadas, ya que según el *American College of Medical Genetics and Genomics* (ACMG), estas variantes pueden afectar al procesamiento y empalme de los exones.
- Filtro de homocigotos: se aplica para clasificar como benignas aquellas variantes para las que se hayan descrito al menos 20 personas homocigóticas según la base de datos ExAC<sup>48</sup>.

Después de aplicar estos filtros, el número inicial de variantes encontradas se reduce significativamente. A continuación, se analizan una por una las variantes restantes sin clasificar, valorando toda la información obtenida por las herramientas bioinformáticas, consultando en bases de datos de referencia como HGMD<sup>56</sup> (*Human Gene Mutation Database*) y ClinVar<sup>47</sup>, y revisando la bibliografía científica mediante búsquedas en PubMed<sup>57</sup>, para poder interpretar dichas variantes. La base de datos ClinVar<sup>47</sup> clasifica en seis grupos las variantes (Tabla 1) detectadas a partir de toda la información relevante que se encuentra depositada en ella, tomando siempre como referencia las publicaciones del ACMG. Según las directrices del ACMG, una variante es patogénica o benigna, cuando causa o no causa una enfermedad, respectivamente, con una certeza del 100%. Y lo es probablemente patogénica o benigna cuando la certeza es superior al 90% e inferior al 100%<sup>58</sup>.

**Tabla 1. Clasificación de las variantes según ClinVar<sup>47</sup>**

Código	Significancia ClinVar <sup>47</sup>
0	VOUS ( <i>Variant Of Unknown Significance</i> )
1	No clasificadas
2	Benigna
3	Probablemente benigna
4	Probablemente patogénica
5	Patogénica



## Materiales y métodos

### VI. 3. 2. Generación de informes

Los informes se realizan de acuerdo a las recomendaciones del ACMG<sup>58</sup> y en ellos se dan a conocer a los pacientes únicamente las variantes patogénicas y probablemente patogénicas detectadas en su genoma, y que por tanto afectarían a la correcta función del gen o genes dañados. Las variantes de significado desconocido (VOUS) no serán objeto de informe ya que, según las buenas prácticas de laboratorio, no se han de tener en cuenta para un posterior diagnóstico genético preimplantacional hasta que su significado sea clasificado como benigno o patogénico.

### VI. 4. Cálculo del riesgo residual

Se denomina riesgo residual, o riesgo de portador post-test, a la probabilidad que tiene un paciente de ser portador aun habiéndose obtenido un resultado negativo para las variantes que pueden ser detectadas por el test. La evaluación del riesgo residual es un componente esencial para el consejo genético. Se calcula mediante el método Bayesiano de Ogino y Wilson<sup>59</sup>. La estimación del riesgo residual se realiza para una población específica, en nuestro caso la caucásica, y depende de la frecuencia de portador y la tasa de detección de portador<sup>13</sup>, que se definen a continuación:

- Frecuencia de portador (riesgo de portador pre-test): probabilidad de que un individuo de una población dada presente en heterocigosis una variante patogénica causante de una enfermedad autosómica recesiva concreta.
- Tasa de detección de portador: porcentaje de portadores que serían detectados por un determinado test. No es del 100% debido a, entre otros fenómenos, que no todas las variantes patogénicas pueden ser detectadas por un determinado test. Para nuestro test habría un porcentaje de portadores que no podría ser detectado por secuenciación debido a la cobertura del secuenciador y/o, a inserciones o deleciones grandes, de uno o dos exones, que podrían ocurrir en un gen. Éstas últimas se pueden detectar mediante otras técnicas como la PCR cuantitativa, los *arrays* o MLPA (*Multiplex Ligation-dependent Probe Amplification*)<sup>60</sup>.

#### VI. 4. 1. Determinación de la frecuencia de portador

Las frecuencias de portadores de la población caucásica pueden obtenerse directamente de la literatura científica, en caso de estar disponibles, dando preferencia a aquellas publicadas en las líneas guía del ACMG. En el caso de que la frecuencia de portador no estuviera disponible, se puede estimar a partir de los datos bibliográficos sobre la incidencia de una enfermedad en la población caucásica, si se asume el equilibrio poblacional y se aplica la ley de Hardy-Weinberg del siguiente modo:

## Materiales y métodos

$$x = 2\sqrt{J}(1 - \sqrt{J})$$

siendo  $x$  la frecuencia de portador y  $J$  la incidencia de la enfermedad. No se debe confundir incidencia con prevalencia. La incidencia de una enfermedad contabiliza el número de casos nuevos que aparecen en un período de tiempo determinado. Sin embargo, la prevalencia describe la proporción de la población que padece la enfermedad en un momento concreto.

### VI. 4. 2. Determinación de la tasa de detección de portador

Al igual que para la frecuencia de portador, la tasa de detección de portador también es específica para la población caucásica. El valor de esta tasa depende de la cobertura del secuenciador para los genes del panel de portadores y de la tasa de detección teórica. Esta última se obtiene de la literatura científica, con preferencia para aquellos datos que han sido publicados como parte de las líneas guía del ACMG y que han tenido en cuenta las inserciones y deleciones grandes que puedan darse en el gen y que por tanto no podrían ser detectadas por secuenciación. De esta manera, la tasa de detección de portador es el producto de la cobertura del secuenciador por la tasa de detección teórica. Resumiendo, la tasa de detección de portador viene determinada por el tipo de mutaciones que puede detectar el secuenciador, en nuestro caso las puntuales, ya que no es capaz de detectar las inserciones y deleciones de uno o dos exones en un gen.

### VI. 4. 3. Determinación del riesgo residual

Para aquellos genes que se ha podido calcular la frecuencia de portador y la tasa de detección de portador, hemos estimado el riesgo residual mediante el método Bayesiano publicado por Ogino y Wilson<sup>59</sup> del siguiente modo:

$$p(\text{test negativo de portador}) = \frac{(1 - c)x}{1 - xc}$$

Siendo  $x$  la frecuencia de portador, y  $c$  la tasa de detección de portador. Cabe destacar que el riesgo residual calculado sería para un miembro de la pareja que hubiese dado negativo en el test de portadores. En caso de que el otro miembro de la pareja también diese un resultado negativo para el test, se multiplicarían ambos riesgos, para obtener el riesgo residual de ser una pareja portadora. Finalmente, este riesgo se multiplicaría por 1/4, que es la probabilidad de transmitir la enfermedad recesiva a la siguiente generación, obteniéndose así el riesgo residual reproductivo de tener un descendiente afecto.

## VI. 5. Cariotipado

El análisis de cromosomas es una técnica convencional ampliamente usada para la detección de la inestabilidad de los cromosomas. Para dicho análisis se usa el bandedo G, que se puede complementar con técnicas como la hibridación fluorescente *in situ* (FISH), la hibridación

## Materiales y métodos

genómica comparativa (CGH) o el cariotipo espectral (SKY) para los casos en los que sean necesarias estas técnicas. El análisis de cromosomas permite detectar células con aberraciones que involucran la pérdida o ganancia de partes del genoma y/o reordenamientos que suceden en uno o más cromosomas. El protocolo usado en este trabajo se basa en el videoartículo de Howe *et al*<sup>61</sup>, y se detalla en los siguientes apartados.

### VI. 5. 1. Extracción de cromosomas a partir de linfocitos de sangre periférica

En primer lugar, se extrae sangre periférica del paciente con el objetivo de estimular los linfocitos de dicha muestra. El cultivo se hace en un medio convencional con fitohemaglutinina, una lectina que tiene capacidad para aglutinar leucocitos y estimular inespecíficamente la proliferación de células T. Tras 72 h de la adición de la fitohemaglutinina, aproximadamente el 50% de las células se encuentran en fase S, momento de máxima actividad mitótica y óptimo para recolectar los cromosomas para su posterior estudio. Así pues, se recoge 1 ml de sangre periférica del cultivo en un tubo con heparina sódica que inhibe la coagulación de la sangre. De dicho tubo, se alicuotan 0,25 ml de sangre para añadirlos a 10 ml de medio RPMI con L-glutamina. Se incuba la mezcla del medio con la sangre a 37°C con un 5% de CO<sub>2</sub> durante 45 min. Para provocar un choque hipotónico, se añaden al cultivo 10 µl de 0,075 M KCl. De esta manera se aumenta el volumen celular y los cromosomas se separan. Se vuelve a centrifugar a 200 x g durante 5 min y se elimina el sobrenadante, excepto unos 0,5 µl para resuspender. Por último, se añaden 10 µl de fijador ácido Carnoy (3 metanol : 1 ácido acético), para preservar los ácidos nucleicos. La propia fijación deshidrata la muestra, causando la muerte de las células y fortaleciendo los cromosomas. Después se centrifuga a 200 x g durante 5 min, se descarta el sobrenadante y el precipitado se resuspende en 0,5 µl de sobrenadante, para volver a añadir Carnoy. Este proceso se repite tres veces.

### VI. 5. 2. Preparación de la muestra

Se centrifuga de nuevo a 200 x g durante 5 min y se quita el sobrenadante. El precipitado se resuspende en 0,5 µl de sobrenadante, y con una pipeta Pasteur se deja caer una gota en medio del portaobjetos. Seguidamente, se añade al portaobjetos una gota de Carnoy para la fijación de los cromosomas al mismo y se deja secar durante 10 min. Mientras tanto, se prepara la solución Giemsa (3 tampón Gurr : 1 tinte Giemsa). Se cubren los portaobjetos con la solución Giemsa y tras 5 min se lavan con agua y se dejan secar. Finalmente, se analizan las muestras al microscopio óptico a 100 aumentos. Si los cromosomas metafásicos están bien extendidos y son abundantes, se puede continuar. En caso de que los cromosomas se encuentren poco extendidos, se repite el proceso de lavado con Carnoy.

## Materiales y métodos

### VI. 5. 3. Bandeado con tripsina y Giemsa

La tripsina es una enzima proteolítica que desnaturaliza histonas eucromatínicas en regiones de ADN con alta actividad transcripcional. Tras la tinción con Giemsa, dichas regiones aparecen como bandas claras. La cromatina altamente condensada con poca o ninguna actividad transcripcional (heterocromatina) tiene una gran cantidad de sus histonas protegidas contra la acción de la tripsina, y por lo tanto, son oscuras después de la tinción Giemsa. Es esencial hacer el bandedo en una muestra y monitorizar el proceso, para poder ajustar el tiempo de actuación de la tripsina en las siguientes muestras. Siguiendo con el proceso, a 4 frascos Coplin se añaden las siguientes soluciones: 30 ml de HBSS 1x y 4 ml de tripsina 10x al frasco 1, 50 ml de HBSS 1x al frasco 2, 45 ml de HBSS 1x y 5 ml de suero fetal bovino, que aporta diferentes factores de crecimiento para el mantenimiento de los cromosomas, al frasco 3, y por último, 50 ml de HBSS 1x al frasco 4. Cada portaobjetos se sumerge en el frasco 1 durante 30 s, se introduce y se saca del frasco 2, se sumerge en el frasco 3 durante 30 s y finalmente se introduce y se saca rápidamente del frasco 4. Seguidamente, los portaobjetos se cubren con solución Giemsa durante 5 min, tras lo cual se lavan con agua y se dejan secar durante 10 min.

### VI. 5. 4. Análisis del cariotipo

Una vez los portaobjetos están secos, se procede al análisis de la muestra con un microscopio óptico *Olympus bx51* (Anexo 3, figura 4), que alcanza 1000 aumentos y al que se encuentra acoplada una cámara de vídeo conectada que permite exportar la imagen de la muestra al ordenador. El cariotipo se analiza mediante el software *Leica Cytovision* (Anexo 3, figura 4) y se ordenan los cromosomas, apareando cada uno de ellos con su homólogo y estudiando, en el caso de que las haya, las anomalías cromosómicas como translocaciones, deleciones y aneuploidías.

## Resultados

### VII. RESULTADOS

A continuación, se presentan los resultados obtenidos en este TFG que han sido divididos en tres grandes secciones: el análisis de las variantes obtenidas en secuenciación (VII. 1), el cálculo del riesgo residual con diversos ejemplos de las distintas situaciones posibles (VII. 2), y el cariotipado de pacientes (VII. 3).

#### VII. 1. Análisis de las variantes

Hemos convenido mostrar en este TFG únicamente las variantes patogénicas y probablemente patogénicas encontradas tras analizar las secuencias obtenidas de cinco pacientes. Éstos se han denotado con una S, de secuenciación, seguida de un número ordinal que identifica a cada paciente, del 1 al 5. La nomenclatura usada para describir las variantes identificadas en los pacientes, sigue las recomendaciones de la HGVS<sup>42</sup>. Primero aparece el nombre abreviado del gen afectado, seguido de una *c* que indica que la variante afecta a la región codificante. Después aparece la posición e identidad del nucleótido afectado en la región codificante y el tipo de mutación producida, en el caso de nuestros pacientes sustitución (>) o delección (*del*). En caso de tratarse de una sustitución, aparece a continuación del símbolo > el nucleótido sustituto. Por último, con una *p.* se indica el cambio ocurrido en la proteína que. A continuación de la *p.* se especifica el aminoácido de referencia, utilizando el código de tres letras, y su posición en la proteína silvestre. Las modificaciones en la proteína ocasionadas por la mutación nucleotídica se representan del siguiente modo: terminación prematura (\*), cambio en la pauta de lectura (*fs*), o sustitución, indicada con la abreviatura del nuevo aminoácido. A continuación se especifican las variantes encontradas en los pacientes:

- **Paciente S1:** en esta paciente, que es mujer y pareja del paciente S2, se han encontrado dos variantes patogénicas. Las variantes son:
  - ❖ ***AMPD1:c.133C>T p.(Gln45\*)***

Esta primera alteración se encuentra en heterocigosis, en el exón 2 del gen *AMPD1* y se trata de una transición de C a T en la posición 133 de la región codificante del gen. Dicha alteración supone la aparición de un codón de terminación prematuro, que puede ocasionar el truncamiento de la proteína, generando una versión de solo 45 aminoácidos, en lugar de los 780 de la versión silvestre. Aparece descrita en las bases de datos de variantes con una frecuencia de 3,81% (Proyecto 1000 Genomas<sup>49</sup>) y 8,7% (ExAC<sup>48</sup>). Además, ha sido clasificada como patogénica según la base de datos clínica de variantes ClinVar<sup>47</sup> (<https://www.ncbi.nlm.nih.gov/clinvar/variation/18271/>). El gen *AMPD1* codifica la enzima

## Resultados

adenosina monofosfato desaminasa1 (AMPD1), implicada en la producción de energía en el músculo esquelético. Las mutaciones en este gen son la causa de la deficiencia de AMP desaminasa con patrón de herencia autosómico recesivo. Dado su patrón de herencia y la presencia de la alteración p.(Gln45\*) en heterocigosis, la paciente es portadora asintomática de la deficiencia de AMP desaminasa.

### ❖ **GLDC:c.1175delC p.(Ala392fs)**

Esta alteración en heterocigosis, se encuentra en el exón 9 del gen *GLDC* y supone la deleción de la C que ocupa la posición 1175 de la región codificante del gen. Dicha alteración genera un cambio en la pauta de lectura y por tanto un cambio en la secuencia de la proteína a partir del aminoácido 392 incluido. Su frecuencia no aparece en las bases de datos de variantes. La alteración ha sido clasificada como patogénica según la base de datos clínica de variantes ClinVar<sup>47</sup> (<https://www.ncbi.nlm.nih.gov/clinvar/variation/56039/>). El gen *GLDC* codifica la enzima descarboxilasa piridoxal fosfato dependiente de glicina (proteína P). Las mutaciones en este gen son responsables de la encefalopatía por glicina o hiperglicinemia no cetósica. Dado el patrón de herencia autosómico recesivo y la presencia de la alteración p.(Ala392fs) en heterocigosis, la paciente es portadora asintomática de hiperglicinemia no cetósica.

- **Paciente S2:** en este paciente, que es varón y pareja de la paciente S1, se han encontrado dos variantes patogénicas:

### ❖ **ACADS:c.511C>T p.(Arg171Trp)**

Se detecta una mutación de cambio de sentido, en heterocigosis, en el exón 5 del gen *ACADS*, que se trata de una transición de C a T en la posición 511 de la región codificante del gen. Dicha alteración origina la substitución de una arginina por un triptófano. Aparece descrita en las bases de datos de variantes con una frecuencia de 1,55% (Proyecto 1000 Genomas<sup>49</sup>) y 3,1% (ExAC<sup>48</sup>). Esta alteración ha sido clasificada controvertidamente como patogénica, probablemente patogénica, e incluso, probablemente benigna, según la base de datos clínica de variantes ClinVar<sup>47</sup> (<https://www.ncbi.nlm.nih.gov/clinvar/variation/3830/>). El gen *ACADS* codifica una flavoproteína tetramérica mitocondrial, que pertenece a la familia de las acil CoA-deshidrogenasas. Esta enzima cataliza el paso inicial en la ruta de beta-oxidación mitocondrial de los ácidos grasos mitocondriales. Las mutaciones en este gen se asocian con la deficiencia de acil CoA-deshidrogenasa de cadena corta. Dado el patrón de herencia autosómico recesivo y la presencia de la alteración p.(Arg171Trp) en heterocigosis, el paciente es portador asintomático de la deficiencia de acil CoA-deshidrogenasa de cadena corta.

### ❖ **SERPINA1:c.863A>T p.(Glu288Val)**

Se detecta una alteración en heterocigosis en el exón 3 del gen *SERPINA1*, encontrando una transversión de A a T en la posición 863 de la región codificante del gen, lo que ocasiona un cambio de ácido glutámico por valina en la posición 288 de la proteína. Esta variante aparece en

## Resultados

las bases de datos de variantes con una frecuencia de 1,95% (Proyecto 1000 Genomas<sup>49</sup>) y 2,007% (ExAC<sup>48</sup>). El gen *SERPINA1* codifica la glicoproteína alfa 1-antitripsina, un tipo de inhibidor de las serín proteasas (serpina) de amplia especificidad. Las serpinas controlan las cascadas proteolíticas durante procesos biológicos como la coagulación y la inflamación. Las mutaciones en el gen *SERPINA1* producen deficiencia de alfa 1-antitripsina, y consecuentemente, degradación tisular durante la inflamación. De acuerdo a las recomendaciones más recientes del ACMG, la variante se clasifica como patogénica. Sin embargo, su presencia en heterocigosis en el paciente le confiere el estado de portador.

- **Paciente S3:** Se ha encontrado una variante probablemente patogénica en esta paciente, que es mujer y pareja del paciente S4:

- ❖ ***GJB2:c.101T>C p.(Met34Thr)***

Esta alteración se encuentra en heterocigosis en el exón 2 del gen *GJB2* y supone una transición de T a C en la posición 101 de la región codificante del gen, lo que produce la sustitución de una metionina por una treonina en la posición 34 de la proteína. Aparece en las bases de datos de variantes con una frecuencia de 0,6% (Proyecto 1000 Genomas<sup>49</sup>) y menor del 1% (ExAC<sup>48</sup>). El gen *GJB2* codifica el miembro beta 2 de la familia de proteínas conocidas como conexinas, cuya función es establecer puentes de comunicación entre las células, los cuales facilitan la transferencia de iones y pequeñas moléculas entre las células. Las mutaciones en este gen se asocian con la sordera tipo 1A. De acuerdo a las recomendaciones más recientes del ACMG, la variante detectada en el gen *GJB2* se clasifica como probablemente patogénica, por lo que la paciente es portadora.

- **Paciente S4:** en este paciente, que es varón y pareja de la paciente S3, se han encontrado dos variantes conflictivas: una probablemente patogénica y una patogénica:

- ❖ ***AGL:c.907C>T p.(Gln303\*)***

Se ha detectado una mutación sin sentido, en heterocigosis en el exón 7 del gen *AGL*, que supone una transición de C a T, generando un codón de parada prematuro que da lugar a una proteína truncada de sólo 302 aminoácidos. Aparece en las bases de datos de variantes con una frecuencia de 0,0008% (ExAC<sup>48</sup>). El gen *AGL* codifica una enzima implicada en la degradación de glucógeno, con actividad 4-alfa-glucotransferasa y amilo-1,6-glucosidasa. Las mutaciones en este gen se asocian con la enfermedad de almacenamiento de glucógeno. La variante encontrada no ha sido previamente descrita en la literatura médica. Sin embargo, de acuerdo a las recomendaciones más recientes del ACMG, la variante detectada se clasifica como probablemente patogénica, por lo que el paciente es portador.

- ❖ ***SLC26A4:c.2162C>T p.(Thr721Met)***

Esta alteración consiste en una transición de C a T en la posición 2162 de la región codificante del gen, en heterocigosis, en el exón 19 del gen *SLC26A4* y supone la sustitución de

## Resultados

una treonina por una meteonina en la posición 721 de la proteína. Aparece en las bases de datos de variantes con una frecuencia de 0,0058% (ExAC<sup>48</sup>). El gen *SLC26A4* codifica el miembro 4 de la familia de transportadores de solutos número 26. Las mutaciones en este gen se asocian con la sordera autosómica recesiva 4. De acuerdo a las recomendaciones del ACMG, la variante detectada en el gen *SLC26A4* se clasifica como patogénica, por lo que el paciente es portador.

- **Paciente S5:** se ha encontrado una variante probablemente patogénica en esta paciente, que es mujer donante de óvulos y a diferencia de los anteriores pacientes, su pareja no ha sido analizada.

### ❖ *DUOX2:c.3901C>T p.(Gln1301\*)*

Se ha detectado una alteración en heterocigosis, en el exón 30 del gen *DUOX2*, debida a una transición de C a T, lo que genera un codón de parada prematuro dando lugar a una proteína truncada de 1300 aminoácidos. Aparece en las bases de datos de variantes con una frecuencia de 0,0008% (ExAC<sup>48</sup>). El gen *DUOX2* codifica una glicoproteína, miembro de la familia de NADPH oxidadas. Las mutaciones en este gen se asocian con la dishormogénesis tiroidea tipo 6, enfermedad autosómica recesiva, debido a defectos en los pasos de síntesis y secreción de hormona tiroidea. De acuerdo con las recomendaciones del ACMG, la variante detectada en el gen *DUOX2* se clasifica como probablemente patogénica por lo que la paciente es portadora.

## VII. 2. Cálculo del riesgo residual

Los cálculos del riesgo residual para los 298 genes del panel de detección de portadores, se recogen en el Anexo 2, debido a que si fueran incluidos en esta memoria su número de páginas excedería de 40. Los datos usados para los distintos genes y los resultados obtenidos están enfocados a pacientes de raza caucásica de población europea. Además, cabe destacar, como se ha explicado en el apartado de Materiales y métodos, que el riesgo residual para un paciente dado es la probabilidad de ser portador de una variante patogénica, o probablemente patogénica, para un gen concreto aun habiéndose obtenido un resultado negativo en el test.

Aunque cada paciente secuenciado presenta un valor de cobertura distinto, y por tanto un valor diferente de riesgo residual, que aparecería reflejado específicamente en su informe individual, en este TFG para simplificar su presentación se ha optado por adjuntar una sola tabla de cálculos. Para ello se ha utilizado para cada gen su media de cobertura considerando los cinco pacientes analizados. Los cálculos del riesgo mostrados en el anexo 2 se presentan de dos formas (Figura 5): la primera, como el valor numérico resultante del cálculo, y la segunda, como 1 en N, siendo N el número de casos entre los que aparecería 1 para la situación contemplada (ser portador, establecer una pareja portadora o concebir un descendiente afecto siendo pareja portadora, asumiendo siempre haber obtenido un resultado negativo en el test). Además, se redondea dicho número al múltiplo superior exacto.



## Resultados

En la tabla del anexo 2 aparecen los valores de la incidencia (Columna B) y/o de la frecuencia de portador (Columnas C y D). Cuando no se han encontrado disponibles en la literatura científica los hemos referenciado como N/A (no aplicable). En estos casos tampoco es posible calcular los riesgos posteriores.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1			Frecuencia de portador					Riesgo residual de ser portador		Riesgo de establecer una pareja portadora		Riesgo de tener un descendiente afecto siendo	
2	Gen	Incidencia	Valor	1 en N	Cobertura	Tasa de detección (teórica)	Tasa de detección de portador	Valor	1 en N	Valor	1 en N	Valor	1 en N
3	ABCB11	N/A	N/A	N/A	0,9814	0,95	0,93233	N/A	N/A	N/A	N/A	N/A	N/A
4	ABCC2		0,00200	510	0,9701	0,95	0,921595	0,000156785	6380	2,45817E-08	40680730	6,14542E-09	162722900
5	ABCC8		0,01500	70	0,9438	0,9	0,84942	0,00228785	440	5,23426E-06	191050	1,30856E-06	764200
6	ABCD1	0,0005	0,04372	30	0,8516	0,93	0,791988	0,009420779	110	8,87511E-05	11270	2,21878E-05	45070
7	ACADM		0,00200	500	0,8507	0,78	0,663546	0,000673802	1490	4,54009E-07	2202600	1,13502E-07	8810400
8	ACADS		0,01000	100	0,7797	0,99	0,771903	0,002298714	440	5,28409E-06	189250	1,32102E-06	757000
9	ACADVL		0,01150	90	0,9408	0,9	0,84672	0,001780053	570	3,16859E-06	315600	7,92147E-07	1262400
10	ACAT1		0,00200	500	0,8241	0,9	0,74169	0,000517387	1940	2,6769E-07	3735670	6,69225E-08	14942670

**Figura 5. Sección de la tabla de cálculos del riesgo residual recogidos en el anexo 2.** Se incluyen los cálculos de riesgo residual para los ocho primeros genes, ordenados alfabéticamente, del panel de portadores.

En la tabla, también se recoge la tasa de detección de portador (Columna G) que es el producto de la cobertura (Columna E) por la tasa de detección teórica (Columna F).

A continuación, se encuentra el cálculo del riesgo residual de ser portador para cada paciente (Columnas H e I). El riesgo residual de establecer una pareja portadora (Columnas J y K) es el producto de los riesgos residuales de cada miembro de la pareja. Finalmente, se ha calculado el riesgo residual de concebir un descendiente afecto (Columnas L y M), que es el producto del riesgo residual de establecer una pareja portadora por la probabilidad de transmitir una enfermedad autosómica recesiva a un descendiente, es decir, 1/4.

Si bien en los anteriores cálculos del riesgo residual solo se contemplaban situaciones en las que los pacientes hayan dado negativo en el test, esto no siempre es así, como se ha comprobado en la secuenciación de los cinco pacientes. En las tablas 2 y 3 se ejemplifican diferentes situaciones posibles, tanto pre-test como post-test, para varias parejas teóricas, en las que ninguno, uno o los dos miembros hayan sido analizados por el test para un gen concreto, y en caso de que hayan sido analizados, uno o los dos hayan dado positivo en el test. Como ejemplo, podemos señalar el caso en el que ambos miembros hubiesen sido analizados y dado positivo en el test para una variante patogénica o probablemente patogénica del mismo gen, lo que implicaría un riesgo de ser pareja portadora del 100% o de 1 en 1. Para estos ejemplos, se han escogido dos genes del panel: *CFTR*, cuyas mutaciones causan la fibrosis quística, una de las enfermedades autosómicas recesivas con mayor incidencia en poblaciones caucásicas, y el gen *GJB2* por haber detectado una de sus variantes probablemente patogénica en el paciente S3. Estos resultados se expresan de acuerdo a las líneas guía para laboratorios de genética clínica del ACMG<sup>62</sup>:

## Resultados

**Tabla 2. Cálculos de los riesgos de ser pareja portadora y de tener un descendiente afecto para el gen *CFTR*.**

	Resultados del test para la pareja		Riesgo de ser pareja portadora (1 en N)	Riesgo de tener un descendiente afecto (1 en N)	
	No analizado	Analizado Negativo    Positivo			
<b>Gen <i>CFTR</i></b>	XX		770	3080	
		XX	788300	3153170	
			XX	1	4
	X	X	24700	98800	
	X		X	30	120
		X	X	890	3560

Cada X representa a cada uno de los miembros de la pareja. Para los riesgos referentes a resultados no analizados, se tiene en cuenta la frecuencia de portador de una variante patogénica o probablemente patogénica del gen *CFTR* en la población caucásica. Los resultados negativos expresan no haber encontrado variantes patogénicas o probablemente patogénicas para el gen *CFTR* una vez realizado el test. Los resultados positivos indican que el test ha determinado que el paciente es portador de una variante patogénica o probablemente patogénica en el gen *CFTR*.

**Tabla 3. Cálculos de los riesgos de ser pareja portadora y de tener un descendiente afecto para el gen *GJB2*.**

	Resultados del test para la pareja		Riesgo de ser pareja portadora (1 en N)	Riesgo de tener un descendiente afecto (1 en N)	
	No analizado	Analizado Negativo    Positivo			
<b>Gen <i>GJB2</i></b>	XX		2500	10000	
		XX	24000000	96000000	
			XX	1	4
	X	X	245000	980000	
	X		X	50	200
		X	X	4900	19600

Se siguen los mismos criterios que en la tabla 2.

### VII. 3. Análisis de cariotipo

Este análisis permite obtener una imagen del cariotipo del paciente, que contiene los cromosomas de una célula en metafase ordenados de acuerdo a su tamaño y morfología, ya sean metacéntricos, submetacéntricos, telocéntricos, subtlocéntricos o acrocéntricos. El cariotipo humano normal presenta 46 cromosomas, de los cuales 23 proceden de la madre y 23 del padre, y están organizados en 22 pares autosómicos y 1 par sexual (XY en el hombre y XX en la mujer). Una vez ordenado el cariotipo, se procede al análisis del mismo.

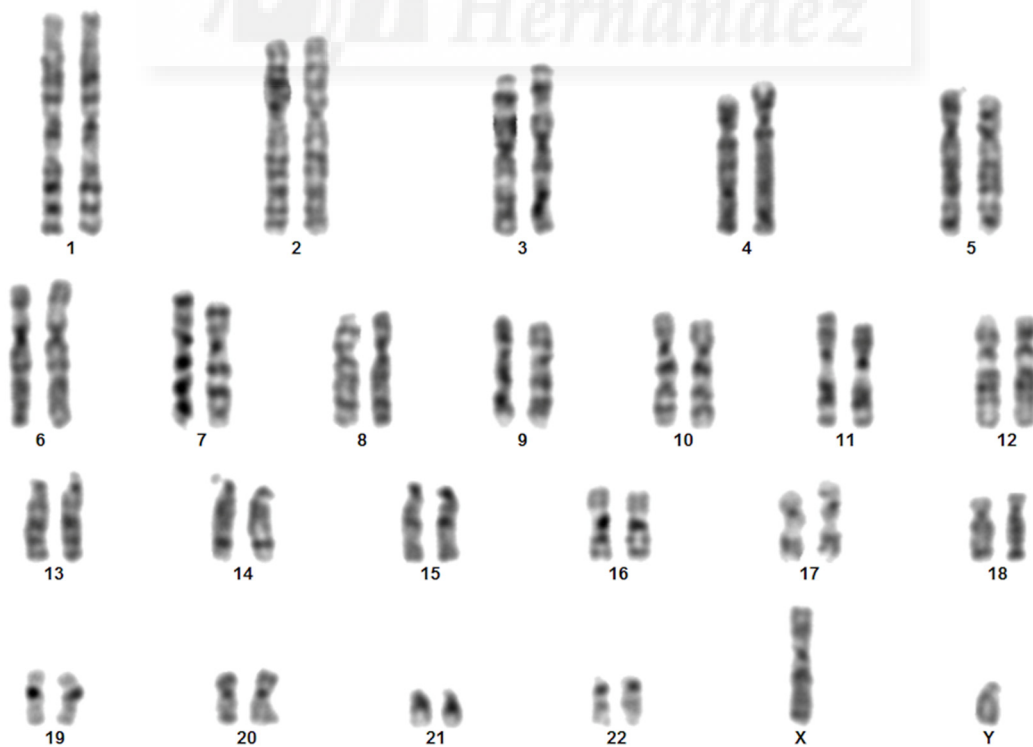
Durante la realización de este TFG, he participado en el cariotipado de más de 20 pacientes, con el objetivo de aprender esta técnica, entre los que se incluyen los de los cinco pacientes que fueron analizados mediante el panel de portadores. La nomenclatura usada en este trabajo para denominar los diferentes cariotipos se basa en la obra de referencia “*An international System for Human Cytogenetic Nomenclature*”<sup>63</sup>. Los pacientes cariotipados han sido denotados con una C, de cariotipo, seguida de un número ordinal que identifica al paciente. Los cariotipos de los pacientes han sido en su mayoría normales, como los de C1 y C2, que corresponden a los pacientes S1 y S2 analizados mediante el panel de portadores y se muestran en las figuras 5 y 6, respectivamente, a modo de ejemplo de cariotipos silvestres. También se han encontrado dos pacientes, denotados como C3 y C4, que no se corresponden con los S3 y S4, que

## Resultados

presentan anomalías cromosómicas, cuyos cariotipos se muestran en las figuras 7 y 8, respectivamente. Por ello, se ha convenido mostrar únicamente estos cariotipos, junto a los dos ejemplos de cariotipos normales.

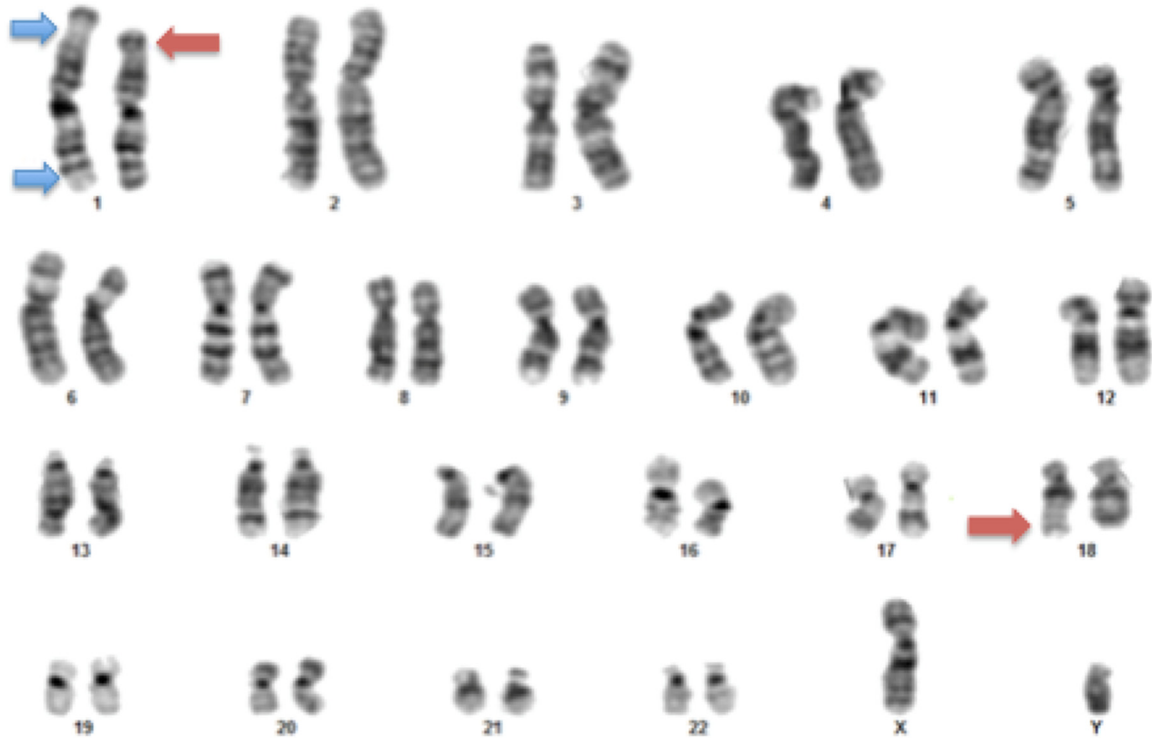


**Figura 6. Cariotipo de la paciente C1.** El cariotipo 46, XX corresponde al de una mujer que no presenta ninguna aberración cromosómica.



**Figura 7. Cariotipo del paciente C2.** El cariotipo 46, XY corresponde al de un varón que no presenta ninguna aberración cromosómica.

## Resultados



**Figura 8. Cariotipo del paciente C3.** El cariotipo corresponde al de un varón 46, XY, inv (1), t (1;18) caracterizado por una inversión en uno de sus cromosomas 1 y una translocación 1;18 en el otro miembro del par 1. Las flechas azules indican las regiones del cromosoma 1 en las que se ha producido la inversión. Las flechas rojas indican la translocación 1;18.



**Figura 9. Cariotipo del paciente C4.** El cariotipo corresponde al de un varón 47, XXY que debido al cromosoma X extra presenta síndrome de Klinefelter. La flecha azul indica el cromosoma X en exceso.

### VIII. DISCUSIÓN

En este TFG, se ha desarrollado y validado un test genético preconcepcional basado en el estudio del genoma mediante la secuenciación de un panel de detección de portadores y la técnica del cariotipado.

El empleo de las herramientas bioinformáticas permite el análisis de los datos procedentes de secuenciación y ha resultado fundamental en el proceso de desarrollo de este test genético. El cálculo del riesgo residual permite a los pacientes conocer las probabilidades de su riesgo reproductivo tanto para las variantes patogénicas, y probablemente patogénicas identificadas con el test, como para aquellas que no hayan podido ser detectadas por la secuenciación, que supone una parte esencial del consentimiento informado. Por otro lado, la técnica convencional del cariotipado permite complementar el estudio del genoma de los pacientes, con la búsqueda de anomalías cromosómicas que también pueden afectar en una futura gestación de los pacientes.

#### VIII. 1. Secuenciación

Tras los resultados obtenidos al secuenciar el panel de detección de portadores en los 5 pacientes analizados, la media de las variantes nucleotídicas detectadas en los pacientes es de 1300, y tras el cálculo de los porcentajes de las distintas variantes encontradas (Tabla 4), se puede afirmar que la gran mayoría de las variantes son benignas. La rápida aplicación de los distintos filtros usados en el análisis bioinformático permite descartar dichas variantes, para poder analizar más detenidamente aquellas más conflictivas como las patogénicas, probablemente patogénicas y VOUS. Basándose en las bases de datos ClinVar<sup>47</sup> y HGMD<sup>56</sup> y en publicaciones científicas mediante búsquedas en PubMed<sup>57</sup>, se interpretan las variantes conflictivas siguiendo siempre los criterios y recomendaciones del ACMG. Cabe destacar que todas las alteraciones clasificadas como patogénicas o probablemente patogénicas han sido verificadas y comprobadas por Bioarray S.L. utilizando secuenciación capilar mediante el método de Sanger<sup>26</sup>.

**Tabla 4. Media de los porcentajes de las distintas variantes encontradas en los pacientes analizados.**

Clasificación de las variantes	Porcentaje
Variantes benignas	96%
Variantes probablemente benignas	1,55%
Variantes patogénicas y probablemente patogénicas	0,15%
VOUS	0,75%
Variantes filtradas	1,55%

Las variantes filtradas son aquellas con una cobertura menor de 10, y que son descartadas al aplicar el filtro de calidad.

## Discusión

Los cinco pacientes analizados tienen la condición de portador para una o dos variantes nucleotídicas patogénicas y/o probablemente patogénicas. En las dos parejas analizadas, las respectivas variantes no coinciden en ambos miembros de la pareja, por lo que el riesgo reproductivo de tener un descendiente afecto es muy reducido, lo que queda plasmado y es informado a la pareja a través del cálculo del riesgo residual. Aun así, es recomendable el consejo genético para las dos parejas incluidas en el estudio, en una consulta especializada para transmitirles la información sobre las alteraciones detectadas. Sin embargo, la pareja de la paciente S5 no ha sido analizada, y teniendo en cuenta que una variante nucleotídica probablemente patogénica ha sido detectada en el gen *DUOX2*, se recomienda el estudio del genoma de su pareja. Así mismo, también es recomendable el consejo genético especializado en las alteraciones detectadas.

### VIII. 2. Cálculo del riesgo residual

Mediante el método Bayesiano publicado por Ogino y Wilson<sup>59</sup>, se ha calculado el riesgo residual para la inmensa mayoría de los 298 genes del panel de detección de portadores. En concreto, se ha calculado el riesgo para 245 genes, mientras que para los 53 restantes no ha podido ser calculado debido a la falta de los datos de incidencia o de frecuencia de portador de dichos genes en la literatura científica. Los resultados obtenidos para los 298 genes se muestran en el anexo 2, y cabe destacar de nuevo que éstos se expresan suponiendo un teórico resultado negativo en la detección de variantes mediante el panel de portadores para estos genes. En el apartado de Resultados, se ha ilustrado con diferentes ejemplos cómo se informaría a los pacientes en cualquier situación posible, tanto si se han encontrado en ellos variantes patogénicas o probablemente patogénicas como si no se ha hecho, cumpliendo así con el consentimiento informado al poder comunicar a los pacientes todos los riesgos posibles después de haberse realizado el test genético preconcepcional.

En cuanto a los datos mostrados en el anexo 2, cabe destacar el valor de cobertura de la secuenciación, expresado en tanto por 1, como la fracción del gen para la que se han obtenido por secuenciación al menos 20 lecturas de cada uno de sus nucleótidos. Este dato juega un papel importante en el cálculo del riesgo residual, ya que cuanto mayor sea la cobertura, más se reduce el posterior riesgo reproductivo. La cobertura es más alta cuanto mejor es el resultado de la secuenciación, aunque siempre hay regiones de los genes que no podrán cubrirse por el secuenciador, debido a las inserciones y deleciones de uno o dos exones que puedan ocurrir en el gen.

En nuestro caso, la cobertura de los 298 genes ha estado comprendida entre el 57% (gen *HBA1*) y el 100%, si bien para el 88,6% de los genes comprendidos en el panel, la cobertura ha sido igual o superior al 85%.

## Discusión

### VIII. 3. Análisis de cariotipo

Como ya se ha comentado previamente en el apartado de Resultados, durante la realización de este proyecto, realicé el cariotipado de más de 20 pacientes, siendo la mayoría de ellos normales. Los pacientes analizados mediante secuenciación también fueron cariotipados y todos ellos presentaron cariotipos normales. Los pacientes S1 y S2 analizados, que corresponden a los pacientes C1 y C2 cariotipados, presentan cariotipos normales, que no afectan al riesgo reproductivo de la pareja. Sin embargo, los pacientes C3 y C4 fueron los únicos en los que se encontraron anomalías cromosómicas al realizar el análisis del cariotipo. En concreto, el paciente C3 sufre una inversión en uno de los cromosomas del par 1, intercambiándose los extremos de dicho cromosoma. Además, este paciente también sufre una translocación 1;18, en el otro cromosoma del par 1. Como se indica en la figura 7, la translocación ocurre en un extremo de dicho cromosoma del par 1 con un extremo de un cromosoma del par 18. A pesar de presentar dos anomalías cromosómicas en su cariotipo, dicho paciente es únicamente portador de éstas, ya que es un individuo sano. Atendiendo a su riesgo reproductivo, la formación de gametos sufrirá problemas en la segregación de los cromosomas del par 1, ya que ambos cromosomas sufren anomalías. Por lo tanto, es altamente recomendable que el paciente sea aconsejado en una clínica especializada en diagnóstico genético preimplantacional junto a su pareja.

Por último, el paciente C4 sufre el síndrome de Klinefelter<sup>64</sup>, debido a la presencia de un cromosoma X en exceso. Entre las características de este síndrome, se encuentra habitualmente la infertilidad, de manera que la única opción viable para poder gestar un descendiente junto a su pareja, sería acudiendo a una clínica de diagnóstico genético preimplantacional.

### IX. CONCLUSIONES Y PROYECCIÓN FUTURA

El objetivo general de este trabajo era el desarrollo y la validación de un test genético preconcepcional para la empresa Bioarray S.L., a través de la puesta a punto del panel de detección de portadores mediante la secuenciación con el *Ion Torrent PGM™*, el cálculo del riesgo residual para los 298 genes del panel, y como complemento al estudio del genoma de los pacientes, la realización de la técnica del cariotipado. Las conclusiones que se pueden obtener de este trabajo son las siguientes:

- Se ha logrado poner a punto el panel de detección de portadores, mediante el seguimiento y optimización del protocolo *Ion AmpliSeq DNA Library Preparation*, de Thermo Fisher . Se han realizado algunas modificaciones en dicho protocolo para que los resultados de la secuenciación del ADN de los pacientes sean óptimos. Además, gracias a las herramientas bioinformáticas de la empresa, el análisis de las variantes resulta rápido y preciso, y para la interpretación de éstas se han seguido siempre las líneas guía del ACMG.
- Se han realizado con éxito los cálculos del riesgo residual para la mayor parte de los genes del panel de detección de portadores, pudiéndose informar a los pacientes de todos los riesgos reproductivos posibles, cumpliendo así con el consentimiento informado.
- Se ha realizado la técnica tradicional del cariotipado, que se utiliza en la empresa desde la creación de la misma, a más de 20 pacientes, incluidos los cinco sometidos al test del panel de detección de portadores. Esta técnica es un buen complemento a la secuenciación del panel de portadores, en la búsqueda de anomalías cromosómicas que puedan afectar al riesgo reproductivo de los pacientes.

Respecto a la proyección futura, cabe destacar que este test genético preconcepcional se va a ofrecer como servicio de la empresa incluyendo también el análisis del gen *SMN1* y del síndrome X frágil, que no han sido objeto de este trabajo. En un futuro cercano, la empresa Bioarray S.L. quiere ampliar el número de genes incluido en el panel de detección de portadores. Además, este test genético está enfocado a la población de raza caucásica, por lo que la empresa también está interesada en ofrecer nuevos paneles de detección de portadores para otras razas. En lo que a mí respecta, este trabajo me ha permitido familiarizarme con un amplio espectro de técnicas y herramientas usadas en una empresa dedicada a la genética clínica, como la aplicación de diferentes protocolos, el cariotipado, la secuenciación con el *Ion Torrent PGM™* y la realización del análisis de variantes llevadas a cabo mediante la bioinformática.



### X. BIBLIOGRAFÍA

1. Pritchard, D. J. & Korf, B. R. Asesoramiento genético, diagnóstico y tratamiento de las enfermedades y asuntos éticos y sociales. En *Genética médica: Lo esencial en un vistazo*. Editorial médica Panamericana. Madrid, España (2013).
2. Temel, S. *et al.* The Dutch national summit on preconception care: a summary of definitions, evidence and recommendations. *J. Community Genet.* **6**, 107–115 (2015).
3. Wille, M. C., Weitz, B., Kerper, P. & Frazier, S. Advances in preconception genetic counseling. *J. Perinat. Neonatal Nurs.* **18**, 28–40
4. Fowler, J. & Hughes, J. *Preconception Counseling. StatPearls* (2017).
5. Delgado, A. *et al.* Herencia mendeliana. En *Asesoramiento Genético en la Práctica médica*. Editorial médica Panamericana. Madrid, España (2012).
6. Genetic Alliance & The New York-Mid-Atlantic Consortium for Genetic and Newborn Screening Services. *Understanding Genetics: A New York, Mid-Atlantic Guide for Patients and Health Professionals.* (2009).
7. Pinto, L. L., Vieira, T. A., Giugliani, R. & Schwartz, I. VD. Expression of the disease on female carriers of X-linked lysosomal disorders: a brief review. *Orphanet J. Rare Dis.* **5**, 14 (2010).
8. Pritchard, D. J. & Korf, B. R. Herencia ligada a los cromosomas X e Y. En *Genética médica: Lo esencial en un vistazo*. Editorial médica Panamericana. Madrid, España (2013).
9. Gao, Z., Waggoner, D., Stephens, M., Ober, C. & Przeworski, M. An estimate of the average number of recessive lethal mutations carried by humans. *Genetics* **199**, 1243–54 (2015).
10. Castellani, C. *et al.* Benchmarks for cystic fibrosis carrier screening: a European consensus document. *J. Cyst. Fibros.* **9**, 165–78 (2010).
11. Hussein, N., Weng, S. F., Kai, J., Kleijnen, J. & Qureshi, N. Preconception risk assessment for thalassaemia, sickle cell disease, cystic fibrosis and Tay-Sachs disease. *Cochrane database Syst. Rev.* CD010849 (2015). doi:10.1002/14651858.CD010849.pub2
12. Pritchard, D. J. & Korf, B. R. Herencia autosómica recesiva: principios. En *Genética médica: Lo esencial en un vistazo*. Editorial médica Panamericana. Madrid, España (2013).
13. Wu, S. Estimating carrier frequency, carrier detection rate, and post-test carrier risk for recessive disorders. *23andme White Pap.* (2015).
14. Galanello, R. & Origa, R. Beta-thalassemia. *Orphanet J. Rare Dis.* **5**, 11 (2010).
15. Bender, M. *Sickle Cell Disease. GeneReviews*(®) (1993).
16. Origa, R. & Moi, P. *Alpha-Thalassemia. GeneReviews*(®) (1993).
17. Ong, T. *et al.* *Cystic Fibrosis and Congenital Absence of the Vas Deferens. GeneReviews*(®) (1993).
18. Kaback, M. M. & Desnick, R. J. *Hexosaminidase A Deficiency. GeneReviews*(®) (1993).
19. Konkle, B. A., Huston, H. & Nakaya Fletcher, S. *Hemophilia A. GeneReviews*(®) (1993).
20. Petrou, M. Preimplantation genetic diagnosis. *Hemoglobin* **33 Suppl 1**, S7–S13 (2009).
21. Ioannides, A. S. Preconception and prenatal genetic counselling. *Best Pract. Res. Clin. Obstet. Gynaecol.* **42**, 2–10 (2017).
22. Marteau, T. M., Dormandy, E. & Michie, S. A measure of informed choice. *Health Expect.* **4**, 99–108 (2001).
23. Koboldt, D. C., Steinberg, K. M., Larson, D. E., Wilson, R. K. & Mardis, E. R. The next-generation sequencing revolution and its impact on genomics. *Cell* **155**, 27–38 (2013).
24. Edwards, J. G. *et al.* Expanded carrier screening in reproductive medicine—points to consider: a joint statement of the American College of Medical Genetics and Genomics, American College of Obstetricians and Gynecologists, National Society of Genetic Counselors, Perinatal Qual. *Obstet. Gynecol.* **125**, 653–62 (2015).
25. WATSON, J. D. & CRICK, F. H. The structure of DNA. *Cold Spring Harb. Symp. Quant. Biol.* **18**, 123–31 (1953).
26. Sanger, F. & Coulson, A. R. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *J. Mol. Biol.* **94**, 441–8 (1975).
27. Mardis, E. R. The impact of next-generation sequencing technology on genetics. *Trends*

## Bibliografía

- Genet.* **24**, 133–41 (2008).
28. Goodwin, S., McPherson, J. D. & McCombie, W. R. Coming of age: ten years of next-generation sequencing technologies. *Nat. Rev. Genet.* **17**, 333–51 (2016).
  29. Liu, L. *et al.* Comparison of next-generation sequencing systems. *J. Biomed. Biotechnol.* **2012**, 251364 (2012).
  30. Tomkinson, A. E., Vijayakumar, S., Pascal, J. M. & Ellenberger, T. DNA ligases: structure, reaction mechanism, and function. *Chem. Rev.* **106**, 687–99 (2006).
  31. Guo, J. *et al.* Four-color DNA sequencing with 3'-O-modified nucleotide reversible terminators and chemically cleavable fluorescent dideoxynucleotides. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 9145–50 (2008).
  32. Ivanova, M., Shivarov, V., Pavlov, I., Lilakos, K. & Naumova, E. Clinical Evaluation of a Novel Nine-Gene Panel for Ion Torrent PGM Sequencing of Myeloid Malignancies. *Mol. Diagn. Ther.* **20**, 27–32 (2016).
  33. Malapelle, U. *et al.* Ion Torrent next-generation sequencing for routine identification of clinically relevant mutations in colorectal cancer patients. *J. Clin. Pathol.* **68**, 64–68 (2015).
  34. Alturkistani, H. A., Tashkandi, F. M. & Mohammedsaleh, Z. M. Histological Stains: A Literature Review and Case Study. *Glob. J. Health Sci.* **8**, 72–9 (2015).
  35. Hwang, S. M. *et al.* The application of an in situ karyotyping technique for mesenchymal stromal cells: a validation and comparison study with classical G-banding. *Exp. Mol. Med.* **45**, e68 (2013).
  36. Connor, J. M. & Ferguson-Smith, M. A. Essential medical genetics (3rd edn). *Trends Genet.* **7**, 236 (1991).
  37. Ropers, H.-H. On the future of genetic risk assessment. *J. Community Genet.* **3**, 229–236 (2012).
  38. Holtkamp, K. C. A. *et al.* Factors for successful implementation of population-based expanded carrier screening: learning from existing initiatives: Table 1. *Eur. J. Public Health* ckw110 (2016). doi:10.1093/eurpub/ckw110
  39. Pritchard, D. J. & Korf, B. R. Lugar de la genética en la medicina. En *Genética médica: Lo esencial en un vistazo*. Editorial médica Panamericana. Madrid, España (2013).
  40. Ciaccio, C. *et al.* Fragile X syndrome: a review of clinical and molecular diagnoses. *Ital. J. Pediatr.* **43**, 39 (2017).
  41. Zarkov, M. *et al.* Association between the SMN2 gene copy number and clinical characteristics of patients with spinal muscular atrophy with homozygous deletion of exon 7 of the SMN1 gene. *Vojnosanit. Pregl.* **72**, 859–63 (2015).
  42. den Dunnen, J. T. & Antonarakis, S. E. Mutation nomenclature extensions and suggestions to describe complex mutations: a discussion. *Hum. Mutat.* **15**, 7–12 (2000).
  43. Hamosh, A. *et al.* Onlined Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Res.* **30**, 52–55 (2002).
  44. Pavan, S. *et al.* Clinical Practice Guidelines for Rare Diseases: The Orphanet Database. *PLoS One* **12**, e0170365 (2017).
  45. Smedley, D. *et al.* The BioMart community portal: an innovative alternative to large, centralized data repositories. *Nucleic Acids Res.* **43**, W589–W598 (2015).
  46. Sherry, S. T. *et al.* dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.* **29**, 308–11 (2001).
  47. Landrum, M. J. *et al.* ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Res.* **44**, D862–8 (2016).
  48. Lek, M. *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285–291 (2016).
  49. Auton, A. *et al.* A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
  50. Roten, C. A., Gamba, P., Barblan, J. L. & Karamata, D. Comparative Genometrics (CG): a database dedicated to biometric comparisons of whole genomes. *Nucleic Acids Res.* **30**, 142–144 (2002).
  51. Kumar, P., Henikoff, S. & Ng, P. C. Predicting the effects of coding non-synonymous

## Bibliografía

- variants on protein function using the SIFT algorithm. *Nat. Protoc.* **4**, 1073–81 (2009).
52. Adzhubei, I. A. *et al.* A method and server for predicting damaging missense mutations. *Nat. Methods* **7**, 248–9 (2010).
  53. Schwarz, J. M., Cooper, D. N., Schuelke, M. & Seelow, D. MutationTaster2: mutation prediction for the deep-sequencing age. *Nat. Methods* **11**, 361–362 (2014).
  54. Pollard, K. S., Hubisz, M. J., Rosenbloom, K. R. & Siepel, A. Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res.* **20**, 110–121 (2010).
  55. Garber, M. *et al.* Identifying novel constrained elements by exploiting biased substitution patterns. *Bioinformatics* **25**, i54–i62 (2009).
  56. Stenson, P. D. *et al.* Human Gene Mutation Database (HGMD): 2003 update. *Hum. Mutat.* **21**, 577–81 (2003).
  57. PubMed. Disponible en: <https://www.ncbi.nlm.nih.gov/pubmed/>.
  58. Richards, S. *et al.* Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet. Med.* **17**, 405–24 (2015).
  59. Ogino, S. & Wilson, R. B. Bayesian analysis and risk assessment in genetic counseling and testing. *J. Mol. Diagn.* **6**, 1–9 (2004).
  60. Fedida, J. *et al.* Contribution of exome sequencing for genetic diagnostic in arrhythmogenic right ventricular cardiomyopathy/dysplasia. *PLoS One* **12**, e0181840 (2017).
  61. Howe, B., Umrigar, A. & Tsien, F. Chromosome Preparation From Cultured Cells. *J. Vis. Exp.* (2014). doi:10.3791/50203
  62. Richards, C. S. *et al.* Standards and guidelines for CFTR mutation testing. *Genet. Med.* **4**, 379–91 (2008).
  63. Shaffer, L. & Tommerup, N. *An international system for human cytogenetic nomenclature.* Editorial Karger. Basel, Suiza (2005).
  64. Crawford, D. & Dearmun, A. Klinefelter syndrome. *Nurs. Child. Young People* **29**, 19 (2017).