



<https://doi.org/10.18233/APM44No1pp29-552548>

## Anemia de Fanconi, Parte 2. Estrategia metodológica para el diagnóstico molecular en pacientes con anemia de Fanconi

### Fanconi anemia, Part 2. Methodological strategy for molecular diagnosis in patients with Fanconi anemia

Leda Torres,<sup>1</sup> Ulises Juárez,<sup>1,2</sup> Pedro Reyes,<sup>1,2</sup> Sara Frías<sup>1,3</sup>

#### Resumen

La anemia de Fanconi (AF) es una enfermedad rara, se presenta en 1-5/millón de nacidos vivos. A nivel celular presentan inestabilidad cromosómica, que es la base para su diagnóstico y aunque clínicamente son heterogéneos, hay tres características generales: alteraciones del desarrollo físico, pancitopenia y alto riesgo a desarrollar cáncer. Presenta heterogeneidad genética, hasta ahora se han reportado 22 genes responsables de la AF, 20 de estos genes se heredan de manera autosómica recesiva, uno autosómica dominante y uno ligada al X, sin embargo, existen genes por detectar, ya que a pesar de una minuciosa búsqueda, no en todos los pacientes se logra encontrar la variante patogénica responsable. Debido a esta heterogeneidad, el diagnóstico molecular es complicado, por lo que se necesita una estrategia con varias metodologías como el ensayo de amplificación de sondas dependiente de ligandos múltiples (MLPA), secuenciación de nueva generación, ya sea por panel dirigido (16 genes *FANC*), o por secuenciación del exoma completo, y microarreglos de alta resolución. Con estas metodologías se pueden detectar grandes deleciones o duplicaciones en los genes *FANC*, alteraciones puntuales y en el número de copias, así como regiones largas con homocigosidad, con el propósito de encontrar alelos homocigotos. En este artículo presentamos una estrategia detallada para realizar la genotipificación de los pacientes AF mexicanos, con un porcentaje de éxito del 80%.

**PALABRAS CLAVE:** Anemia de Fanconi; Variantes patogénicas; genes *FANC*; MLPA; Secuenciación de nueva generación; Estrategia de genotipificación.

#### Abstract

Fanconi anemia (FA) is a rare disease occurring in 1-5/million live births. Patients present chromosomal instability at the cellular level, which is the basis for their diagnosis, and although clinically they are heterogeneous, there are three general characteristics: alterations in physical development, pancytopenia and high risk of cancer development. To date, 22 genes responsible for FA have been reported, 20 of which are inherited in an autosomal recessive pattern, one autosomal dominant and one X-linked; however, there are genes to be detected, since despite a thorough search, the responsible pathogenic variant cannot be found in all patients. Due to this heterogeneity, the molecular diagnosis is complicated, so a strategy with several methodologies, such as multiple ligand-dependent probe amplification assay (MLPA) and next-generation sequencing, either by directed panel (16 *FANC* genes) or by whole exome sequencing and high-resolution microarrays, is necessary. With these methodologies, it is possible to detect long deletions or duplications in *FANC* genes, single nucleotide and copy number alterations, and long regions with homozygosity to find homozygous alleles. In this article, we present a detailed strategy for genotyping Mexican FA patients, with a success rate of 80%.

**KEYWORDS:** Fanconi anemia; Pathogenic variants; *FANC* genes; MLPA; Next generation sequencing; Genotyping strategy.

<sup>1</sup> Laboratorio de Citogenética, Instituto Nacional de Pediatría.

<sup>2</sup> Programa de Doctorado en Ciencias Biomédicas, Universidad Nacional Autónoma de México.

<sup>3</sup> Departamento de Medicina Genómica y Toxicología Ambiental, Instituto de Investigaciones Biomédicas, Universidad Nacional Autónoma de México.

#### ORCID

0000-0001-8355-5818  
0000-0001-6333-5738  
0000-0002-3097-6368

**Recibido:** 23 de junio de 2022

**Aceptado:** 02 de enero de 2023

#### Correspondencia

Sara Frías  
sarafrias@iibiomedicas.unam.mx

**Este artículo debe citarse como:** Torres L, Juárez U, Reyes P, Frías S. Anemia de Fanconi, Parte 2. Estrategia metodológica para el diagnóstico molecular en pacientes con anemia de Fanconi. Acta Pediatr Mex 2023; 44 (1): 29-55.

## INTRODUCCIÓN

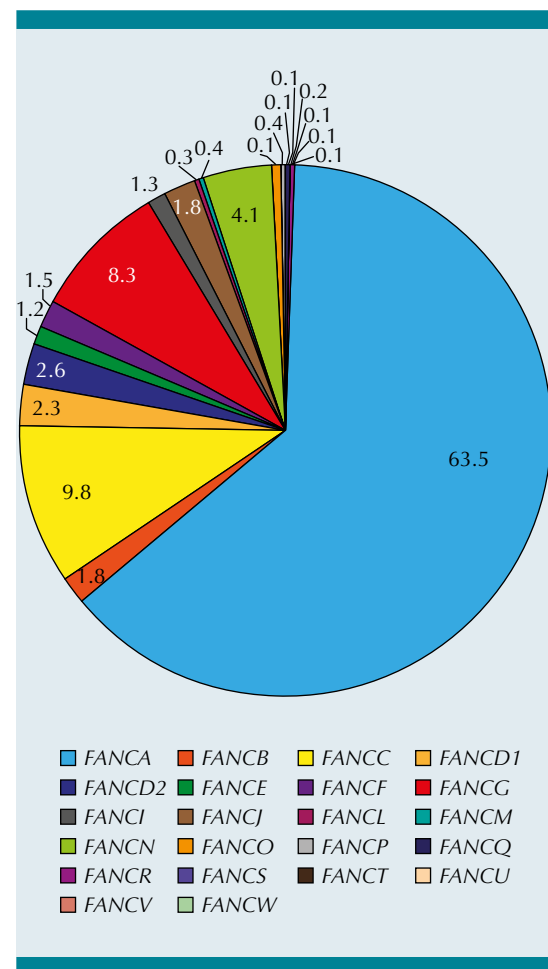
La anemia de Fanconi (AF) es una enfermedad hereditaria rara, caracterizada por inestabilidad genómica; el fenotipo clínico consiste en anomalías del desarrollo físico, insuficiencia de la médula ósea y predisposición a cáncer. La AF tiene heterogeneidad genética, se han identificado 22 genes *FANC* (**Cuadro 1**). Los productos de estos genes cooperan en la vía FA/BRCA, que regula la respuesta a los agentes genotóxicos y mantiene la integridad del genoma.<sup>1</sup> Todas las variantes patogénicas (VP) reportadas en

estos genes tienen herencia autosómica recesiva, excepto en *FANCB*, que están ligadas al cromosoma X, y en *FANCR/RAD51*, que son autosómicas dominantes.<sup>2</sup>

Se ha identificado una amplia variedad de mutaciones diferentes entre los genes de la vía FA/BRCA; las VP bialélicas en *FANCA* son la causa más frecuente de AF (60% a 70%), seguidas de las encontradas en *FANCC* (10-12%) y *FANCG* (6-8%)<sup>3</sup>. El porcentaje de individuos con VP en cada gen *FANC* se muestra en la **Figura 1** (Fanconi Anemia Mutation Database [\*\*Cuadro 1.\*\* Genes implicados en la anemia de Fanconi y localización cromosómica](https://www2.</a></p>
</div>
<div data-bbox=)

| Gene/Alias            | Localización Citogenética |
|-----------------------|---------------------------|
| <i>FANCA</i>          | 16q24.3                   |
| <i>FANCB</i>          | Xp22.2                    |
| <i>FANCC</i>          | 9q22.32                   |
| <i>FANCD1/BRCA2</i>   | 13q13.1                   |
| <i>FANCD2</i>         | 3p25.3                    |
| <i>FANCE</i>          | 6p21.31                   |
| <i>FANCF</i>          | 11p14.3                   |
| <i>FANCG/XRCC9</i>    | 9p13.3                    |
| <i>FANCI</i>          | 15q26.1                   |
| <i>FANCL/BRIP1</i>    | 17q23.2                   |
| <i>FANCL</i>          | 2p16.1                    |
| * <i>FANCM</i>        | 14q21.2                   |
| <i>FANCN/PALB2</i>    | 16q12.2                   |
| * <i>FANCO/RAD51C</i> | 17q22                     |
| <i>FANCP/SLX4</i>     | 16p13.3                   |
| <i>FANCQ/XPF</i>      | 16p13.12                  |
| * <i>FANCR/RAD51</i>  | 15q15.1                   |
| * <i>FANCS/BRCA1</i>  | 17q21.31                  |
| <i>FANCT/UBE2T</i>    | 1q32.1                    |
| <i>FANCU/XRCC2</i>    | 7q36.1                    |
| <i>FANCV/REV7</i>     | 1p36.22                   |
| <i>FANCW/RFWD3</i>    | 16q23.1                   |

\* Estos genes se denominan "Fanconi anemia-like" porque los pacientes que portan una VP en cualquiera de ellos no presentan insuficiencia medular<sup>1</sup>



**Figura 1.** Distribución de pacientes con anemia de Fanconi según el gen *FANC* afectado.

rockefeller.edu/fanconi/; consultada el 20 de enero de 2022).

El hecho de que varios genes y cientos de VP únicas estén asociados con el fenotipo AF, hace que la heterogeneidad de *locus* y de alelo sea la regla. Debido a esta complejidad, el diagnóstico molecular de la enfermedad puede ser desafiante. Distintos grupos de trabajo en el mundo han seguido estrategias diversas de acuerdo a su población y recursos disponibles.<sup>4-6</sup> Encontrar la estrategia molecular adecuada para estudiar un grupo de pacientes con AF, puede ser un trabajo largo y costoso.

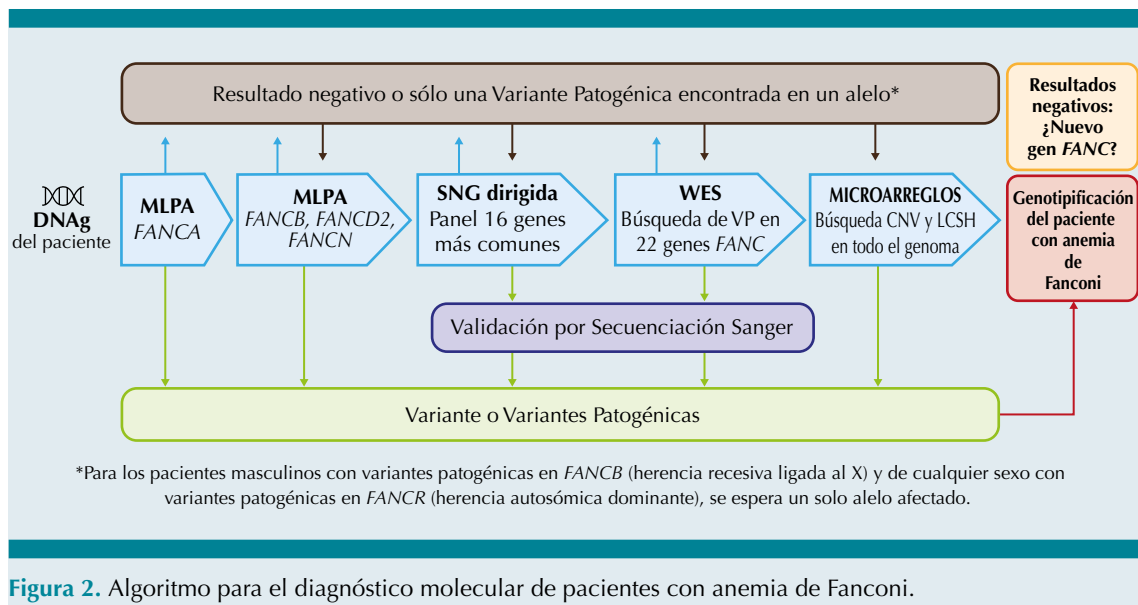
El objetivo de este artículo es presentar la estrategia seguida por nuestro grupo para determinar el genotipo de 50 pacientes con AF, en 40 de los cuales (80%), pudimos obtener el genotipo causal de la enfermedad. Esperamos que este trabajo contribuya a encontrar la estrategia de genotipificación de pacientes con AF más económica y efectiva, en laboratorios con características similares a las nuestras.

En breve, partimos de muestras biológicas de pacientes con diagnóstico clínico y citogenético de AF. Usualmente, las muestras se obtienen de sangre periférica o de aspirado de médula ósea, de las que se extrae DNA genómico (DNAG). A partir de entonces, procedemos de acuerdo con el flujo de trabajo que se sugiere a continuación (**Figura 2**):

1. El primer abordaje es la detección de deleciones largas en el gen *FANCA*, a través de la amplificación de sondas dependiente de ligandos múltiples (MLPA, por sus siglas en inglés), esto debido a que este tipo de VP se observa con mayor frecuencia en este gen, involucrando uno o varios exones, o la pérdida del gen.
2. Para pacientes negativos a deleciones en *FANCA*, realizamos análisis por MLPA para los genes *FANCD2*, *FANCN*

(*PALB2*) y *FANCB*, que también presentan grandes deleciones y duplicaciones detectables por MLPA.

3. Si no se determinan VP mediante MLPA, o si sólo se logró detectar VP en uno de los dos alelos, continuamos la genotipificación a través de secuenciación de nueva generación (NGS, por sus siglas en inglés), ya sea por panel dirigido (en nuestro caso, compuesto por los 16 genes *FANC* más frecuentemente afectados), o por secuenciación del exoma completo (WES, por sus siglas en inglés), para detectar VP pequeñas o puntuales. Con esta estrategia hemos logrado obtener la mayoría de las VP presentes en nuestros pacientes.
4. Si todavía tenemos pacientes sin genotipo, el método de tamizaje que hemos seguido son los microarreglos de alta resolución, que constan de marcadores de polimorfismos de un sólo nucleótido (SNP, por sus siglas en inglés) a lo largo de todo el genoma y que nos permiten detectar variaciones en el número de copias (CNV, por sus siglas en inglés), grandes deleciones o duplicaciones en los genes *FANC*, y la detección de regiones con tramos largos contiguos de homocigosidad (LCSH, por sus siglas en inglés), con el propósito de encontrar alelos homocigotos ininterrumpidos con estado de número de copias neutro de 2 a nivel genómico. Los LCSH se observan en todo el genoma humano, se originan por endogamia (consanguinidad o aislamiento reproductivo, tanto geográfico como social) o por fuerzas evolutivas. Son típicas de las poblaciones consanguíneas, sin embargo, las LCSH también son comunes y en ocasiones, inesperadamente largas en el genoma de personas no consanguíneas;<sup>7-10</sup> la importancia de su detección es que podrían indicar



**Figura 2.** Algoritmo para el diagnóstico molecular de pacientes con anemia de Fanconi.

la presencia de genes recesivos y si se detectan en alguna zona que contenga alguno de los 20 genes *FANC* con herencia autosómica recesiva, se puede hacer una búsqueda intencionada para detectar VP en ese locus. **Figura 2**

## METODOLOGÍA

### Pacientes

Se incluyen pacientes con características clínicas de AF y prueba citogenética de aberraciones cromosómicas positiva. Se explica el procedimiento y se solicita el consentimiento informado por escrito antes de obtener las muestras biológicas.

### Extracción de DNA genómico

El aspirado de sangre o de médula ósea debe colectarse en tubos que contengan EDTA como anticoagulante. Una de las características clínicas más frecuentes en pacientes con AF es el bajo recuento de glóbulos blancos en la sangre; en consecuencia, se espera obtener poco material

de partida para la extracción de DNAG. Para la extracción se utiliza el kit Gentra Puregene Blood Kit (QIAGEN, cat 158389, Venlo, Limburg, NL) (<https://www.qiagen.com/> Gentra Puregene Handbook 12/2014 33), siguiendo las recomendaciones del fabricante, con mínimas modificaciones.

Este protocolo purifica DNAG de muestras frescas o congeladas, a partir de 3 ml de sangre completa. El protocolo también se puede utilizar para la purificación de DNAG proveniente de células centrifugadas, capa leucocitaria o médula ósea. Si se procesan muestras de capa leucocitaria, se debe escalar directamente el volumen de reactivos utilizados en proporción al volumen de la muestra de sangre original.

### Procedimiento

- a. Colocar 3 ml de sangre completa o médula ósea en un tubo cónico de 15 ml, agregar 9 ml de solución de lisis de glóbulos rojos y mezclar invirtiendo 10 veces.

- b.** Incubar 10 min a temperatura ambiente (15–25°C). Invertir al menos una vez durante la incubación.
- c.** Centrifugar durante 5 min a 2000 Fuerza Centrífuga Relativa (FCR) para obtener los glóbulos blancos; si quedan glóbulos rojos, repetir el paso de lisis de glóbulos rojos.
- d.** Desechar cuidadosamente el sobrenadante por pipeteo o decantación, conservando aproximadamente 200 µl del líquido residual y el concentrado de los glóbulos blancos obtenidos. Agitar vigorosamente el tubo para resuspender el concentrado celular en el líquido residual; el sedimento debe dispersarse por completo por agitación en vórtex.
- e.** Añadir 3 ml de solución de lisis celular y pipetear hacia arriba y hacia abajo para lisar las células, o agitar vigorosamente durante 10 s e incubar a 37 °C hasta que la solución sea homogénea.
- f.** Añadir 15 µl de solución de RNAasa A (4 mg/ml) y mezclar por inversión 25 veces. Incubar durante 30 min a 37°C. Luego, agregar 15 µl de proteinasa K (20 mg/ml) e incubar durante 30 min a 55 °C. Inmediatamente poner en hielo para enfriar la muestra.
- g.** Agregar 1 ml de solución de precipitación de proteínas, agitar vigorosamente con vórtex durante 20 s, e incubar en hielo durante 15 min.
- h.** Centrifugar durante 10 min/4°C a 2000 x FCR. Las proteínas precipitadas deben formar un sedimento compacto de color marrón oscuro. Si el sedimento de proteínas no está compacto, repetir la centrifugación.
- i.** Transferir el sobrenadante del paso anterior a un tubo de cónico nuevo de 15 ml. Asegurarse que el sedimento de proteína no se desprenda durante el vertido. Añadir 3 ml de isopropanol y 1 µl de glucógeno (20 mg/ml).
- j.** Mezclar invirtiendo suavemente 50 veces hasta que el DNAg sea visible como hilos o como un ovillo. Centrifugar durante 10 min a 2000 x FCR. El DNAg puede ser visible como un pequeño precipitado blanco.
- k.** Con la ayuda de una pipeta, retirar el sobrenadante. Agregar 3 ml de etanol al 70 %, e invertir varias veces para lavar el precipitado de DNAg.
- l.** Centrifugar durante 5 min a 2000 x FCR. Remover el sobrenadante, transferir el sedimento con el líquido restante a un tubo cónico de 1.5 ml y lavar nuevamente con 300 µl de etanol al 70 %. Centrifugar, retirar el sobrenadante y dejar secar el DNAg en un tubo abierto a temperatura ambiente o 65°C.
- m.** Resuspender el sedimento de DNAg en 50 µl de amortiguador Tris-EDTA (TE) de baja concentración de EDTA (Tris-HCl 10 mM (pH 8.0) + EDTA 0.1 mM), o en agua libre de nucleasas.
- n.** El DNAg obtenido se puede cuantificar en un espectrofotómetro Nanodrop™ (Nanodrop Technologies, Wilmington, DE). Para los ensayos de NGS, podría ser necesario un método de cuantificación más sensible, por ejemplo, a través de cuantificación fluorométrica (Qubit, Thermofisher).
- o.** Evaluar la integridad del DNAg mediante electroforesis en gel de agarosa estándar, al 1 % en amortiguador Tris Acetato-EDTA (TAE) 1X.

### Amplificación de sondas dependiente de ligandos múltiples (MLPA)

El MLPA es un ensayo basado en PCR multiplex para detectar CNV. En este método, sondas específicas de una región genómica hibridan con el DNAg y luego se amplifican mediante PCR, utilizando un sólo par de oligonucleótidos.<sup>11</sup> La gran ventaja de este método es que permite la amplificación simultánea y semi-cuantitativa de más de 60 sondas en una sola reacción, para la identificación rápida y precisa de deleciones y/o duplicaciones.<sup>12</sup>

Las VP encontradas en *FANCA* representan alrededor del 70% de los casos notificados en AF, de los cuales 15%-20% corresponden a grandes deleciones.<sup>13,14</sup> La mayoría de ellas se han encontrado principalmente a través de MLPA, a menudo abarcando múltiples exones.<sup>15</sup> En nuestro grupo, estas alteraciones se buscan con el kit de MLPA desarrollado por MRC Holland (<https://www.mrcholland.com/product/P057>). Además de *FANCA*, hay otros genes *FANC* en los cuales se ha reportado frecuentemente CNV, pudiendo ser evaluadas también a través de esta metodología, como *FANCD2* (~ 3% de los pacientes),<sup>16,17</sup> *FANCN* (<1% de pacientes)<sup>18</sup> y *FANCB* (~ 2% de los pacientes).<sup>19</sup>

Según nuestra experiencia, la cantidad óptima de DNAg que se utiliza por reacción debe ser 80 y 90 ng aproximadamente, diluida en amortiguador TE. Por otra parte, cada mezcla de sondas contiene fragmentos que son utilizados como control de calidad (generalmente 9), para evaluar la desnaturalización (fragmentos "D"), la cantidad de DNA y la eficiencia de la ligación (fragmentos "Q"). Es importante incluir muestras de DNA de referencia, a partir de: a) individuos sanos (tanto hombres como mujeres), b) muestra positiva para AF (por ejemplo, la línea celular FA-A-VU817) y c) muestra negativa (buffer TE sin DNAg); estos controles deben incluirse durante cada corrida. Se sugiere utilizar 3 muestras de referencia por cada 21 muestras.

Se sigue el mismo flujo de trabajo sugerido en el protocolo general de MRC Holland (<https://support.mrcholland.com/downloads/files/mlpa-general-protocol-one-tube>), que se resume en los siguientes pasos:

#### Procedimiento

- a. Desnaturalización del DNA e hibridación de las sondas: Calentar la muestra de DNA de interés (5 µl) durante 5 minutos a 98°C. Luego, agregar 3 µl de la mezcla de sondas a hibridar (1,5 µl de mezcla de sondas SALSA + 1,5 µl de amortiguador MLPA) a la muestra de DNA. Colocar el tubo durante 1 minuto a 95°C e hibridar durante 16 h a 60°C.
- b. Ligación: Una vez transcurrido el tiempo de hibridación, disminuir la temperatura del termociclador a 54°C. Sin retirar los tubos del equipo, agregar 32 µl de la mezcla de ligación e incubar a la misma temperatura durante 15 min. Para inactivar la Ligasa 65, incubar durante 5 minutos a 98 °C.
- c. Amplificación por PCR: Los tubos de reacción se enfrían a temperatura ambiente y se añaden 10 µl de la mezcla de polimerasa. Iniciar la reacción de PCR, utilizando el programa sugerido (95°C por 30 s, 60°C por 30 s y 72°C por 1 minuto, durante 35 ciclos), con una amplificación final a 72°C por 20 minutos. Guardar los tubos protegidos de la luz hasta la electroforesis capilar.
- d. Electroforesis capilar: los productos de PCR se cargan en un instrumento de electroforesis capilar para separarlos según su longitud. La proporción de los componentes de la mezcla de reacción (producto final de la reacción MLPA, marcador de peso y formamida), así como los parámetros de ejecución electroforética, varían según el instrumento disponible.

- e. **Análisis de datos:** La revisión de datos se realiza con el programa Coffalyzer, desarrollado por MRC Holland (<https://www.mrcholland.com/technology/software/coffalyser-net>). Durante este paso, el número relativo de copias se determina contrastando las alturas relativas de los picos de las sondas de referencia y de las regiones blanco en las muestras de prueba contra las de las muestras de referencia. Es fundamental configurar el estudio de las señales crudas obtenidas con la versión de cada mezcla de sondas y el número de lote con el que se realizó la reacción.

Una vez obtenido el electroferograma, se evalúa la altura de los picos de fluorescencia después de la fragmentación. El nivel de fluorescencia de cada pico en una muestra se compara con la señal observada en las muestras de DNA utilizadas como referencia. Por lo tanto, una delección se visualizará como una disminución en la altura del pico en comparación con el fragmento normal, mientras que una duplicación aumenta el tamaño del pico de ese fragmento. Posteriormente, las señales sin procesar se normalizan para obtener un valor que refleje la relación ajustada entre las señales contrastadas. Este valor, conocido como coeficiente de dosis (DQ), permite discriminar el número de copias. El DQ varía de 0 (delecciones homocigotas) a 2,15 (duplicaciones homocigotas). Los valores entre 0,8 y 1 son considerados normales (conjunto diploide) <sup>20</sup>.

### Secuenciación de nueva generación (NGS)

La NGS es un término genérico utilizado para describir el uso de diversas tecnologías, cuya característica común es la secuenciación masiva y paralela de múltiples fragmentos de DNA.<sup>21</sup> La combinación de alto rendimiento y secuenciación de una sola molécula de DNA son características distintivas de la NGS, indepen-

dientemente de la plataforma de secuenciación utilizada.<sup>22</sup> La evolución y diversificación de los ensayos de NGS ha permitido superar las limitaciones de la secuenciación de Sanger y generar una enorme cantidad de datos de secuencias de nucleótidos, haciéndolos dependientes de un análisis computacional intensivo para el ensamblaje de las secuencias producidas.<sup>23</sup>

Los pasos principales de la NGS de DNA implican la fragmentación del DNA, preparación de bibliotecas, secuenciación, análisis bioinformático, anotación de las variantes encontradas y su interpretación.<sup>24</sup> Sin embargo, debe considerarse que la elección de la longitud deseada del fragmento ("lecturas" cortas o largas) depende de la pregunta de investigación y su aplicación clínica.

La aplicación de la NGS ha sido especialmente útil en trastornos con alta variabilidad fenotípica, acelerando la identificación y descubrimiento de nuevas variantes causales, lo que ha impulsado su aplicación en diagnóstico molecular.<sup>25</sup> La reducción de costos y el potencial de la información obtenida han permitido la integración de datos de NGS a la práctica clínica en medicina.<sup>26</sup> Actualmente, existen diferentes enfoques para NGS, dependiendo de si el alcance involucra una gran cantidad de genes (paneles de genes dirigidos), regiones codificantes de proteínas (WES) o secuenciación del genoma completo (WGS, por sus siglas en inglés).

Según la última versión de la Guía de Atención Clínica de la AF, la NGS debe realizarse como parte del proceso de genotipificación, seleccionando un panel de genes *FANC* en una primera etapa, apelando a WES en caso de resultados negativos por prueba de panel.<sup>27</sup> Nuestro grupo ha seguido esta estrategia para la mayoría de los casos analizados, considerando los beneficios y limitaciones de cada enfoque.

El uso de paneles dirigidos se ha generalizado en el contexto de la AF, debido a la mejor

caracterización de la vía FA/BRCA, que ha permitido estudiar más genes *FANC* en un solo ensayo, además de ofrecer mayor profundidad de cobertura a un costo relativamente bajo. Sin embargo, la secuenciación de exoma completo, se ha vuelto cada vez más atractiva debido a que con ella, se pueden descubrir VP en regiones genómicas diferentes de los genes *FANC* conocidos y con esto, descubrir nuevos genes *FANC* o candidatos.<sup>27</sup> Un ejemplo que ilustra esta idea es el descubrimiento de VP bialélicas en *ERCC4/XP-F*, en un grupo de pacientes con AF sin clasificación genotípica; la pérdida de función de este gen se había asociado previamente a Xeroderma pigmentoso. Este hallazgo condujo a la reclasificación de este gen como *FANCQ*.<sup>28</sup> De manera similar, se descubrieron VP en *FANCT* (también conocido como *UBE2T*, en dos pacientes japoneses)<sup>29</sup> y en *FANCU*.<sup>30</sup>

Nuestra búsqueda de VP en la AF a través de NGS se encuentra resumida en el siguiente flujo de trabajo:

#### Procedimiento

- a. Diseño del panel de genes. El diseño *in silico* de los oligonucléidos flanqueadores, específicos para 16 genes *FANC* (*FANCA*, *FANCB*, *FANCC*, *FANCD1/BRCA2*, *FANCD2*, *FANCE*, *FANCF*, *FANCG*, *FANCI*, *FANCI/BRIP1*, *FANCL*, *FANCM*, *FANCN/PALB2*, *FANCO/RAD51C*, *FANCP/SLX4* y *FANCQ/ERCC4*) se realiza con la aplicación web SureDesign (Agilent Technologies). Se espera que el diseño final abarque el 99,5% de todos los exones codificantes de estos genes más 50 pb de sus regiones intrónicas 5' y 3' flanqueantes.
- b. La digestión, síntesis de las bibliotecas y la amplificación por PCR de los fragmentos seleccionados se logra mediante el sistema de enriquecimiento

Haloplex (Agilent Technologies Inc., 2013, Waghäusel-Wiesental, Alemania).

- c. Las lecturas de los extremos cortos pareados son obtenidas mediante secuenciación basada en tecnología de síntesis, en una plataforma Illumina Mi-Seq (Illumina, San Diego, CA, EE. UU.).
- d. En caso de que el análisis bioinformático de los datos de panel sea negativo, se procede a realizar WES, con las adaptaciones metodológicas sugeridas para este enfoque.

A diferencia del diseño de panel, en el caso de WES, existen diversas opciones comerciales enfocadas en la captura de regiones exónicas (por ejemplo, SureSelectXT Human All Exon V4 (Agilent Technologies), KAPA HyperExome (Roche), entre otras). Además, el uso de reactivos y formulaciones de kits compatibles con los sistemas antes mencionados se extiende a los pasos de enriquecimiento y construcción de bibliotecas. Finalmente, las plataformas de secuenciación también se adaptan para un rendimiento más escalable (por ejemplo, el sistema NovaSeq 6000 (Illumina)); debido a la amplificación de un mayor número de dianas de DNA.

Es importante señalar que, debido a la complejidad técnica e interpretativa de la NGS, usualmente su implementación está limitada a laboratorios públicos o privados con capacidad para cumplir con las recomendaciones y requerimientos sugeridos en los estándares internacionales.<sup>31</sup>

Debido a estas demandas, estos laboratorios ofrecen con frecuencia servicios de NGS a investigadores y médicos. Nuestro grupo ha utilizado esta modalidad para obtener datos de secuenciación a partir del envío de muestras de DNA de alta calidad. La selección del enfoque, así como de las plataformas, depende del propósito del proyecto y de los recursos



disponibles para el cribado molecular en los pacientes con AF.

### Análisis de datos de NGS

Debido a la gran cantidad de datos generados a partir de la NGS, incluso a partir de paneles pequeños de genes, se requiere de análisis computacionales para interpretar efectivamente la información de las secuencias.<sup>25</sup> El conjunto de algoritmos bioinformáticos ejecutados en una sucesión de pasos para procesar los datos iniciales generados a partir de la plataforma NGS se conoce como *pipeline* bioinformático.<sup>21</sup> A pesar de la heterogeneidad de las plataformas y el alcance del estudio, el flujo de trabajo bioinformático consta de cinco etapas principales: a) control de calidad de los datos sin procesar (a partir de un archivo en formato de texto que contiene las lecturas de las secuencias y los puntajes individuales de calidad de bases, denominado FASTQ), b) pre-procesamiento, c) mapeo y post-alineamiento, d) llamado de variantes, e) anotación y priorización.<sup>23,32,33</sup>

Desarrollamos un *pipeline* validado para agilizar el análisis computacional y maximizar la precisión en el llamado de variantes, según las buenas prácticas recomendadas en GATK (Genome Analysis Toolkit).<sup>34-36</sup> Los programas utilizados para la manipulación de archivos a menudo se empaquetan en una colección de herramientas basadas en línea de comandos para analizar datos de secuenciación de alto rendimiento (GATK "*toolkit*"), enfocadas principalmente en el descubrimiento de variantes. Es muy importante resaltar que estas herramientas sólo se ejecutan en sistemas operativos Linux o MacOSX.

### Procedimiento (las opciones de programas se nombran entre paréntesis)

- a. Realizar el control de calidad de las lecturas de secuencia (FastQC).
- b. Filtrar y recortar las lecturas que contienen bases de mala calidad, así como adaptadores de secuenciación (Trimmomatic, Cutadapt y Trim Galore).
- c. Las lecturas recortadas son alineadas para mapear con una secuencia de referencia (alineación de Burrows-Wheeler (BWA-MEM), usando las referencias que se encuentran en las bases de datos de UCSC o Ensembl).
- d. Transformación de los archivos SAM en archivos BAM; y marcaje de duplicados (Samtools y Picard).
- e. Comprobar la calidad de los archivos BAM (Qualimap).
- f. Seguir las recomendaciones de GATK, para obtener el archivo de llamado de variantes (VCF), en modo gVCF (genotipado conjunto).
- g. Finalmente, se anotan las variantes (VEP, SnpEff o ANNOVAR). Las variantes se describen siguiendo las pautas propuestas por la nomenclatura de la *Human Genome Variation Society* (HGVS).<sup>37</sup>

### Clasificación de variantes

Una vez filtradas las variantes candidatas por paciente, el siguiente paso consiste en clasificarlas según su grado de patogenicidad. El sistema más utilizado para esta tarea proviene de las recomendaciones sugeridas por el Colegio Americano de Genética Médica y la Asociación de Patología Molecular (ACMG-AMP, por sus siglas en inglés).<sup>38,39</sup> De acuerdo con esta guía, la clasificación de variantes está basada en cinco clases: benigna, probablemente benigna, variante de significado desconocido o incierto (VUS, por sus siglas en inglés), probablemente patógena y patógena.

Los criterios utilizados para esta clasificación se fundamentan en un marco de evidencia

en el que se consideran diferentes parámetros (funcionales, poblacionales, datos de segregación, algoritmos computacionales y predictivos, entre otros) para establecer un resultado final. Varias herramientas web facilitan el ejercicio de clasificación, ya que incorporan información valiosa de bases de datos (como ExAc, ClinVar, gnomAD, entre otras) en una sola plataforma (por ejemplo, Varsome).

#### Validación de variantes encontradas por NGS

Debido a que las técnicas de NGS no están libres de errores, la ACMG y la Sociedad Europea de Genética Humana (ESHG, por sus siglas en inglés) recomiendan enfáticamente validar las variantes seleccionadas. Existen factores tanto técnicos como bioinformáticos relacionados con la dificultad para confirmar variantes, principalmente en genes con pseudogenes, regiones repetitivas y con alto contenido de GC o variantes *hotspot*. La secuenciación de Sanger es el método más recomendable para la validación de los resultados de la NGS.

En nuestro caso, se diseñan pares de oligonucleótidos flanqueantes específicos para validar la VP encontrada por NGS, utilizando el algoritmo Primer3 (<http://bioinfo.ut.ee/primer3-0.4.0/>), una de las herramientas de diseño de cebadores más utilizadas.<sup>40</sup> La amplificación por PCR se lleva a cabo a partir de 100 ng de DNA molde. Los productos de PCR se purifican con el kit QIAquick PCR Purification (QIAGEN, Venlo, Limburg, NL) según las instrucciones del fabricante. Los amplicones purificados se secuencian bidireccionalmente utilizando el kit de secuenciación Big Dye Terminator y se resuelven en un secuenciador Applied Biosystems 3130 Genetic Analyzer (Applied Biosystems, Foster City, CA, EE. UU).

Los electroferogramas se analizan con el software Chromas V2.6.6 ([www.technelysium.com.au](http://www.technelysium.com.au)) y las secuencias blanco se comparan con las secuencias de referencia correspondientes

desde GenBank (<https://www.ncbi.nlm.nih.gov>). Existen otros programas alternativos de análisis (<https://www.genewiz.com/Public/Resources/Tools-for-Viewing-Sequencing-Data/>) o aplicaciones web como GEARgenomics (<https://www.gear-genomics.com>, apoyado por EMBL-EBI, EMBL-European Bioinformatics Institute).

#### La secuenciación de Sanger como estrategia de genotipificación

La secuenciación de Sanger podría realizarse para buscar VP en genes específicos, es decir, cuando se estudian miembros de una familia con una VP previamente identificada o en individuos pertenecientes a una comunidad donde se ha detectado una mutación fundadora, particularmente en poblaciones genéticamente aisladas, como *FANCA* c.295C>T en gitanos españoles,<sup>41</sup> *FANCC* c.456+4A>T en judíos Askenazí<sup>42</sup>, *FANCC* c.67delG en la comunidad menonita<sup>43</sup> y *FANCG* c.511-3\_511-2delCA en una comunidad Mixe mexicana.<sup>44</sup>

#### Microarreglos de alta resolución

Realizamos el análisis del DNAG con microarreglos CytoScan 750K (Affymetrix, Santa Clara, CA, EE. UU.). Estos microarreglos proporcionan cobertura de todo el genoma, incluidos 550,000 marcadores para detectar CNV y 200,436 sondas SNP.

Los pasos de digestión, ligación, amplificación por PCR, fragmentación, etiquetado, desnaturalización e hibridación del DNA en el microarreglo, se realizan de acuerdo con los protocolos y pautas de control de calidad proporcionados por el proveedor.

Luego, los microarreglos se tiñen y lavan en la estación de fluidos Affymetrix GeneChip Fluidics Station 450 y se escanean en un Affymetrix GeneChip Scanner 3000 7G (Affymetrix, Santa Clara, CA, EE. UU.); los archivos obtenidos, se analizan con el programa Chromosome Analysis

Suite (ChAS) versión 4.1, proporcionado por Affymetrix (Affymetrix, Santa Clara, CA, EE. UU.) (Anotación NetAffx 33 hg19) (<http://www.affymetrix.com>). Para el análisis del número de copias (CN), se utilizan los parámetros de “alta resolución”: ganancias y pérdidas en 100Kb con  $\geq 50$  marcadores (sondas).

### Análisis de segmentos largos contiguos de homocigosidad (LCSH)

Los umbrales mínimos para los LCSH en análisis clínicos generalmente se establecen entre 3 y 10 Mb.<sup>45</sup>

Los LCSH se visualizan en el programa ChAS, versión 4.1, proporcionado por Affymetrix (Affymetrix, Santa Clara, CA, EE. UU.). Para el análisis, se utilizan los archivos de anotación NetAffx 33 hg19 (<http://www.affymetrix.com>). Para LCSH>3Mb, la configuración de análisis se establece en LOH con recuento de marcadores (sondas) = 50 y tamaño = 3000 Kbp, y para LCSH>5Mb se establece en recuento de marcadores = 50 y tamaño = 5000 Kbp.

### Reporte de resultados

Una vez clasificadas y validadas las variantes, el siguiente paso es redactar un informe o reporte de resultados. Además de la ACMG, otras organizaciones, como ClinGen, (siglas en inglés de *Clinical Genome Resources*), la ESHG y la Asociación para la Ciencia Genética Clínica (ACGS siglas en inglés de *Association for Clinical Genetic Science*) han desarrollado guías para asesorar a los laboratorios en la notificación y descripción de variantes.<sup>38,39,46-48</sup> Estos documentos son emitidos por un laboratorio hacia el médico tratante y describen los resultados del análisis genético. Deben ser claros, concisos y seguir las recomendaciones sugeridas en las guías existentes.<sup>48</sup>

### Recomendaciones

- a. En general, el formato del informe incluye información sobre el médico solicitante, el laboratorio donde se realizó el análisis y el paciente.
- b. Para la descripción e interpretación de los resultados, la nomenclatura de las variantes se redacta según la Organización del Genoma Humano (HUGO, siglas en inglés de *Human Genome Organization*, <https://www.genenames.org/>) y la Sociedad de Variación del Genoma Humano (HGVS, siglas en inglés de *Human Genome Variation Society* <https://varnomen.hgvs.org/>). El informe debe incluir una interpretación clara de los resultados de la prueba, basada en una revisión del contexto clínico del paciente y la información familiar disponible.<sup>48</sup> El objetivo debe ser generar una declaración simple e inequívoca que explique las implicaciones clínicas del resultado.
- c. Los detalles metodológicos (principio de la tecnología utilizada, sensibilidad, resolución, limitaciones, herramientas bioinformáticas, entre otros) relevantes para la interpretación de los resultados deben incluirse en el informe. Por ejemplo, en el caso del uso de datos de secuenciación masiva para variantes constitucionales, es recomendable consultar las guías más recientes desarrolladas por la ACMG.
- d. Cada laboratorio debe documentar la evidencia utilizada para clasificar las variantes de acuerdo al impacto que pueden tener en la condición estudiada y establecer políticas precisas para definir cuales variantes se excluirán del informe.<sup>39</sup>

En 2008, se creó una base de datos de mutaciones de la AF como un esfuerzo cooperativo para acelerar el acceso a la información sobre las variantes que se encuentran en los genes *FANC* (<https://www2.rockefeller.edu/fanconi/mutate/>). La adición de una variante (nueva o previamente notificada) debe hacerse como institución y no de forma individual. Compartir datos con la comunidad científica aumenta significativamente la cantidad de evidencia existente para una variante particular en las bases de datos públicas, y contribuye a mejorar los enfoques en la búsqueda de correlaciones genotipo-fenotipo.<sup>49</sup> Además, se debe fomentar esta práctica porque facilita la reclasificación de variantes, siendo especialmente importante en el caso de VUS, donde el uso de esta información se suma a la evidencia generada a través de algoritmos bioquímicos predictivos y validación de ensayos funcionales.<sup>50</sup>

- e. Resultados negativos en un paciente con estudio de aberraciones cromosómicas positivo.

Los resultados negativos de las pruebas moleculares deben interpretarse cuidadosamente para un individuo con una prueba de aberraciones cromosómicas dentro del rango de FA. Una explicación para un resultado negativo puede ser la presencia de variantes en un gen *FANC* no descubierto o un tipo de variante en un gen que no se puede identificar con las tecnologías empleadas. Otra posibilidad es el mosaicismo somático.<sup>51,52</sup> En estos casos, se puede considerar el análisis de un tipo de muestra alternativo a la sangre periférica o médula ósea, como fibroblastos, en individuos que presenten un fenotipo de AF y estudios genéticos negativos en sangre periférica.

## DISCUSIÓN

En un porcentaje variable de pacientes (en nuestro caso fue el 20%), no es posible llegar a la genotipificación. Las posibles razones son

a) podrían ser variantes que no se detectan por secuenciación de exoma, como aquellas que se encuentran en regiones intrónicas o regulatorias de los genes y que requerirían otra estrategia más costosa, como la secuenciación del genoma completo; b) alternativamente, podría deberse a que las variantes se encuentren en genes aún no descritos como genes *FANC*, por lo que es recomendable realizar un análisis de los genes asociados funcionalmente a la vía FA/BRCA.

La gran heterogeneidad genética de la AF dificulta la implementación de técnicas de detección clásicamente utilizadas en otro tipo de enfermedades de herencia mendeliana. Por ello, consideramos que la divulgación detallada de los aspectos metodológicos inherentes a la genotipificación en la AF, es fundamental para establecer los requerimientos biológicos y tecnológicos mínimos para identificar las variantes responsables de la enfermedad, en especial en zonas geográficas en donde este abordaje no se ha extendido o se encuentra limitado. Esta explicación, que abarca desde la obtención del DNAG hasta la generación del reporte, pone en contexto la importancia de la información obtenida (como la frecuencia, localización y el tipo de variante) para desarrollar una estrategia de genotipificación costo-efectiva y apropiada a la población AF.

## CONCLUSIONES

Describir el genotipo de AF en una población es relevante por varias razones: en primer lugar, mejora la caracterización de los pacientes y por lo tanto, contribuye a la identificación de correlaciones genotipo-fenotipo de AF. En segundo lugar, refina la adaptación de las estrategias de diagnóstico genético en grupos específicos de pacientes, incluidas las selecciones adecuadas de donantes de médula ósea de hermanos sin AF y el asesoramiento genético adecuado. Finalmente, la descripción de casos con genotipos raramente descritos aporta nuevos datos a

considerar cuando se enfrentan a variantes con antecedentes de clasificación de patogenicidad ambigua.

### Agradecimientos

Esta investigación fue parcialmente financiada por el Consejo Nacional de Ciencia y Tecnología (CONACyT), número de proyecto SALUD-2014-1-233721; Instituto Nacional de Pediatría, Proyecto Recursos Fiscales E022 INP 2014/41; Programa de Apoyo a Proyectos de Investigación e Innovación Tecnológica (PAPIIT), Universidad Nacional Autónoma de México (UNAM) Grant Number IN205120. Los autores agradecen al M. en C. Francisco Zúñiga-Frias por la revisión del inglés de este manuscrito.

### REFERENCIAS

- Fiesco-Roa MÓ, García-de Teresa B, Leal-Anaya P, van 't Hek R, Wegman-Ostrosky T, Frías S, et al. Fanconi anemia and dyskeratosis congenita/telomere biology disorders: Two inherited bone marrow failure syndromes with genomic instability. *Front Oncol* [Internet]. 2022;12. Available from: <https://www.frontiersin.org/articles/10.3389/fonc.2022.949435>
- García-de-Teresa B, Rodríguez A, Frias S. Chromosome Instability in Fanconi Anemia: From Breaks to Phenotypic Consequences. *Genes (Basel)*. 2020 Dec;11(12).
- Niraj J, Färkkilä A, D'Andrea AD. The Fanconi Anemia Pathway in Cancer. *Annu Rev cancer Biol*. 2019 Mar;3:457-78.
- Pilonetto D V, Pereira NF, Bonfim CMS, Ribeiro LL, Bitencourt MA, Kerkhoven L, et al. A strategy for molecular diagnostics of Fanconi anemia in Brazilian patients. *Mol Genet genomic Med*. 2017 Jul;5(4):360-72.
- Mori M, Hira A, Yoshida K, Muramatsu H, Okuno Y, Shiraishi Y, et al. Pathogenic mutations identified by a multimodality approach in 117 Japanese Fanconi anemia patients. *Haematologica*. 2019 Oct;104(10):1962-73.
- Ben Haj Ali A, Messaoud O, Elouej S, Talmoudi F, Ayed W, Mellouli F, et al. FANCA Gene Mutations in North African Fanconi Anemia Patients. *Front Genet* [Internet]. 2021;12. Available from: <https://www.frontiersin.org/articles/10.3389/fgene.2021.610050>
- Li LH, Ho SF, Chen CH, Wei CY, Wong WC, Li LY, et al. Long contiguous stretches of homozygosity in the human genome. *Hum Mutat*. 2006 Nov;27(11):1115-21.
- Pajusalu S, Žilina O, Yakoreva M, Tammur P, Kuuse K, Mölter-Väär T, et al. The Diagnostic Utility of Single Long Contiguous Stretches of Homozygosity in Patients without Parental Consanguinity. *Mol Syndromol*. 2015 Sep;6(3):135-40.
- Wang JC, Ross L, Mahon LW, Owen R, Hemmat M, Wang BT, et al. Regions of homozygosity identified by oligonucleotide SNP arrays: evaluating the incidence and clinical utility. *Eur J Hum Genet*. 2015 May;23(5):663-71.
- Chaves TF, Oliveira LF, Ocampos M, Barbato IT, de Luca GR, Barbato Filho JH, et al. Long contiguous stretches of homozygosity detected by chromosomal microarrays (CMA) in patients with neurodevelopmental disorders in the South of Brazil. *BMC Med Genomics*. 2019 Mar;12(1):50.
- Stuppia L, Antonucci I, Palka G, Gatta V. Use of the MLPA assay in the molecular diagnosis of gene copy number alterations in human genetic diseases. *Int J Mol Sci*. 2012;13(3):3245-76.
- Hömig-Hölzel C, Savola S. Multiplex ligation-dependent probe amplification (MLPA) in tumor diagnostics and prognostics. *Diagnostic Mol Pathol Am J Surg Pathol part B*. 2012 Dec;21(4):189-206.
- Ameziane N, Errami A, Léveillé F, Fontaine C, de Vries Y, van Spaendonck RML, et al. Genetic subtyping of Fanconi anemia by comprehensive mutation screening. *Hum Mutat*. 2008 Jan;29(1):159-66.
- Castella M, Pujol R, Callén E, Trujillo JP, Casado JA, Gille H, et al. Origin, functional role, and clinical impact of Fanconi anemia FANCA mutations. *Blood*. 2011 Apr;117(14):3759-69.
- Kimble DC, Lach FP, Gregg SQ, Donovan FX, Flynn EK, Kamat A, et al. A comprehensive approach to identification of pathogenic FANCA variants in Fanconi anemia patients and their families. *Hum Mutat*. 2018 Feb;39(2):237-54.
- Kalb R, Neveling K, Hoehn H, Schneider H, Linka Y, Batish SD, et al. Hypomorphic mutations in the gene encoding a key Fanconi anemia protein, FANCD2, sustain a significant group of FA-D2 patients with severe phenotype. *Am J Hum Genet*. 2007 May;80(5):895-910.
- Flynn EK, Kamat A, Lach FP, Donovan FX, Kimble DC, Narisu N, et al. Comprehensive analysis of pathogenic deletion variants in Fanconi anemia genes. *Hum Mutat*. 2014 Nov;35(11):1342-53.
- Xia B, Dorsman JC, Ameziane N, de Vries Y, Roomans MA, Sheng Q, et al. Fanconi anemia is associated with a defect in the BRCA2 partner PALB2. *Nat Genet*. 2007 Feb;39(2):159-61.
- Jung M, Ramanagoudr-Bhojappa R, van Twest S, Rosti RO, Murphy V, Tan W, et al. Association of clinical severity with FANCB variant type in Fanconi anemia. *Blood*. 2020 Apr;135(18):1588-602.
- Coffa J, van den Berg J. Analysis of MLPA Data Using Novel Software Coffalyser.NET by MRC-Holland. In: Eldin AB, editor. *Modern Approaches To Quality Control* [Internet]. Rijeka: IntechOpen; 2011. Available from: <https://doi.org/10.5772/21898>
- Roy S, Coldren C, Karunamurthy A, Kip NS, Klee EW, Lincoln SE, et al. Standards and Guidelines for Validating

- Next-Generation Sequencing Bioinformatics Pipelines: A Joint Recommendation of the Association for Molecular Pathology and the College of American Pathologists. *J Mol Diagn.* 2018 Jan;20(1):4–27.
22. Hu T, Chitnis N, Monos D, Dinh A. Next-generation sequencing technologies: An overview. *Hum Immunol.* 2021 Nov;82(11):801–11.
  23. Voelkerding K V, Coonrod EM, Durtschi JD, Margraf RL. Next-Generation Sequencing: Principles for Clinical Application BT - Molecular Pathology in Clinical Practice. In: Leonard DGB, editor. Cham: Springer International Publishing; 2016. p. 889–909. Available from: [https://doi.org/10.1007/978-3-319-19674-9\\_59](https://doi.org/10.1007/978-3-319-19674-9_59)
  24. Qin D. Next-generation sequencing and its clinical application. *Cancer Biol Med.* 2019 Feb;16(1):4–10.
  25. den Dunnen JT. Efficient variant data preparation for Human Mutation manuscripts: Variants and phenotypes. Vol. 40, Human mutation. United States; 2019. p. 1009.
  26. Rizzo JM, Buck MJ. Key principles and clinical applications of “next-generation” DNA sequencing. *Cancer Prev Res (Phila).* 2012 Jul;5(7):887–900.
  27. Sroka I, Frohnmayer L, Wirkkula L. Fanconi Anemia Clinical Care Guidelines [Internet]. Fifth Edit. Vol. 1, Fanconi Anemia Research Fund, Inc. 2020. 1–276 p. Available from: <http://marefateadyan.nashriyat.ir/node/150>
  28. Bogliolo M, Schuster B, Stoecker C, Derkunt B, Su Y, Raams A, et al. Mutations in ERCC4, encoding the DNA-repair endonuclease XPF, cause Fanconi anemia. *Am J Hum Genet.* 2013 May;92(5):800–6.
  29. Hira A, Yoshida K, Sato K, Okuno Y, Shiraishi Y, Chiba K, et al. Mutations in the gene encoding the E2 conjugating enzyme UBE2T cause Fanconi anemia. *Am J Hum Genet.* 2015 Jun;96(6):1001–7.
  30. Shamseldin HE, Elfaki M, Alkuraya FS. Exome sequencing reveals a novel Fanconi group defined by XRCC2 mutation. *J Med Genet.* 2012 Mar;49(3):184–6.
  31. Rehm HL, Bale SJ, Bayrak-Toydemir P, Berg JS, Brown KK, Deignan JL, et al. ACMG clinical laboratory standards for next-generation sequencing. *Genet Med Off J Am Coll Med Genet.* 2013 Sep;15(9):733–47.
  32. McCombie WR, McPherson JD, Mardis ER. Next-Generation Sequencing Technologies. *Cold Spring Harb Perspect Med.* 2019 Nov;9(11).
  33. Bao R, Huang L, Andrade J, Tan W, Kibbe WA, Jiang H, et al. Review of current methods, applications, and data management for the bioinformatics analysis of whole exome sequencing. *Cancer Inform.* 2014;13(Suppl 2):67–82.
  34. DePristo MA, Banks E, Poplin R, Garimella K V, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet.* 2011 May;43(5):491–8.
  35. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, et al. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinforma.* 2013;43(1110):11.10.1-11.10.33.
  36. Franke KR, Crowgey EL. Accelerating next generation sequencing data analysis: an evaluation of optimized best practices for Genome Analysis Toolkit algorithms. *Genomics Inform.* 2020 Mar;18(1):e10.
  37. den Dunnen JT, Dalgleish R, Maglott DR, Hart RK, Greenblatt MS, McGowan-Jordan J, et al. HGVS Recommendations for the Description of Sequence Variants: 2016 Update. *Hum Mutat.* 2016 Jun;37(6):564–9.
  38. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, 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 Off J Am Coll Med Genet.* 2015 May;17(5):405–24.
  39. Rehder C, Bean LJH, Bick D, Chao E, Chung W, Das S, et al. Next-generation sequencing for constitutional variants in the clinical laboratory, 2021 revision: a technical standard of the American College of Medical Genetics and Genomics (ACMG). *Genet Med Off J Am Coll Med Genet.* 2021 Aug;23(8):1399–415.
  40. Kumar A, Chordia N. In Silico PCR Primer Designing and Validation BT - PCR Primer Design. In: Basu C, editor. New York, NY: Springer New York; 2015. p. 143–51. Available from: [https://doi.org/10.1007/978-1-4939-2365-6\\_10](https://doi.org/10.1007/978-1-4939-2365-6_10)
  41. Callén E, Casado JA, Tischkowitz MD, Bueren JA, Creus A, Marcos R, et al. A common founder mutation in FANCA underlies the world’s highest prevalence of Fanconi anemia in Gypsy families from Spain. *Blood.* 2005 Mar;105(5):1946–9.
  42. Whitney MA, Jakobs P, Kaback M, Moses RE, Grompe M. The Ashkenazi Jewish Fanconi anemia mutation: incidence among patients and carrier frequency in the at-risk population. *Hum Mutat.* 1994;3(4):339–41.
  43. García-de Teresa B, Frias S, Molina B, Villarreal MT, Rodríguez A, Carnevale A, et al. FANCC Dutch founder mutation in a Mennonite family from Tamaulipas, México. *Mol Genet genomic Med.* 2019 Jun;7(6):e710.
  44. Reyes P, García-de Teresa B, Juárez U, Pérez-Villatoro F, Fiesco-Roa MO, Rodríguez A, et al. Fanconi Anemia Patients from an Indigenous Community in Mexico Carry a New Founder Pathogenic Variant in FANCG. *Int J Mol Sci.* 2022 Feb;23(4).
  45. Kearney HM, Thorland EC, Brown KK, Quintero-Rivera F, South ST. American College of Medical Genetics standards and guidelines for interpretation and reporting of postnatal constitutional copy number variants. *Genet Med Off J Am Coll Med Genet.* 2011 Jul;13(7):680–5.
  46. Rehm HL, Berg JS, Brooks LD, Bustamante CD, Evans JP, Landrum MJ, et al. ClinGen—the Clinical Genome Resource. *N Engl J Med.* 2015 Jun;372(23):2235–42.
  47. Matthijs G, Souche E, Alders M, Corveleyn A, Eck S, Feenstra I, et al. Guidelines for diagnostic next-generation sequencing. *Eur J Hum Genet.* 2016 Oct;24(10):1515.

48. Smith K, Martindale J, Wallis Y, Bown N, Leo N, Creswell L, et al. General genetic laboratory reporting recommendations. *Birmingham Assoc Clin Genet Sci.* 2015;11.
49. Directors ABO. Laboratory and clinical genomic data sharing is crucial to improving genetic health care: a position statement of the American College of Medical Genetics and Genomics. *Genet Med Off J Am Coll Med Genet.* 2017 Jul;19(7):721–2.
50. Federici G, Soddu S. Variants of uncertain significance in the era of high-throughput genome sequencing: a lesson from breast and ovary cancers. *J Exp Clin Cancer Res.* 2020 Mar;39(1):46.
51. Gregory JJ, Wagner JE, Verlander PC, Levran O, Batish SD, Eide CR, et al. Somatic mosaicism in Fanconi anemia: evidence of genotypic reversion in lymphohematopoietic stem cells. *Proc Natl Acad Sci U S A.* 2001 Feb;98(5):2532–7.
52. Ikeda H, Matsushita M, Waisfisz Q, Kinoshita A, Oostra AB, Nieuwint AWM, et al. Genetic reversion in an acute myelogenous leukemia cell line from a Fanconi anemia patient with biallelic mutations in BRCA2. *Cancer Res.* 2003 May;63(10):2688–94.

## INTRODUCTION

Fanconi anemia (FA) is a rare hereditary disease characterized by genomic instability. The clinical phenotype consists of developmental abnormalities, bone marrow failure, and cancer predisposition. FA presents genetic heterogeneity, 22 *FANCA* genes have been identified to date (**Table 1**). The products of these genes cooperate in the FA/BRCA pathway, which regulates the response to genotoxic agents and maintains genome integrity.<sup>1</sup> All pathogenic variants (PV) in these genes have an autosomal recessive inheritance, except in *FANCB*, which is linked to the X chromosome, and *FANCR/RAD51*, which is autosomal dominant.<sup>2</sup>

Many different mutations have been identified among the FA/BRCA pathway genes. These include the biallelic PV in *FANCA* as the most frequent cause of FA (60% to 70%), followed by *FANCC* (10-12%) and *FANCG* (6-8%).<sup>3</sup> The percentage of individuals with PV in each FA gene is shown in **Figure 1** (Fanconi Anemia Mutation Database <https://www2.rockefeller.edu/fanconi/>; consulted 20th January 2022).

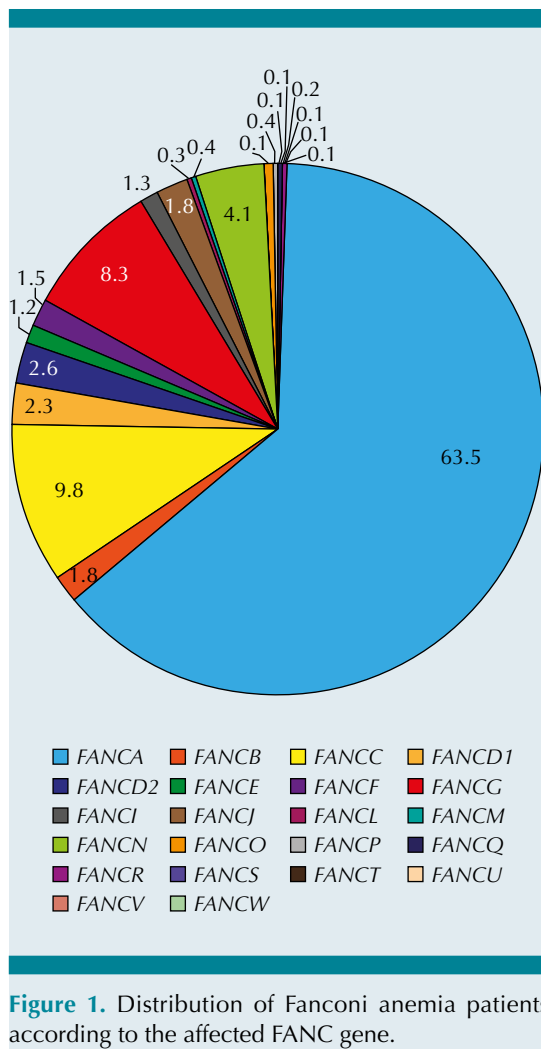
**Table 1.** Fanconi anemia genes and chromosomal location

| Gene/Alias            | Cytogenetic location |
|-----------------------|----------------------|
| <i>FANCA</i>          | 16q24.3              |
| <i>FANCB</i>          | Xp22.2               |
| <i>FANCC</i>          | 9q22.32              |
| <i>FANCD1/BRCA2</i>   | 13q13.1              |
| <i>FANCD2</i>         | 3p25.3               |
| <i>FANCE</i>          | 6p21.31              |
| <i>FANCF</i>          | 11p14.3              |
| <i>FANCG/XRCC9</i>    | 9p13.3               |
| <i>FANCI</i>          | 15q26.1              |
| <i>FANCI/BRIP1</i>    | 17q23.2              |
| <i>FANCL</i>          | 2p16.1               |
| * <i>FANCM</i>        | 14q21.2              |
| <i>FANCN/PALB2</i>    | 16q12.2              |
| * <i>FANCO/RAD51C</i> | 17q22                |
| <i>FANCP/SLX4</i>     | 16p13.3              |
| <i>FANCP/XPF</i>      | 16p13.12             |
| * <i>FANCR/RAD51</i>  | 15q15.1              |
| * <i>FANCS/BRCA1</i>  | 17q21.31             |
| <i>FANCT/UBE2T</i>    | 1q32.1               |
| <i>FANCU/XRCC2</i>    | 7q36.1               |
| <i>FANCV/REV7</i>     | 1p36.22              |
| <i>FANCW/RFWD3</i>    | 16q23.1              |

\* These genes are called “Fanconi anemia-like” because patients who carry a PV in any of them, do not present bone marrow failure 1.

The fact that several genes and hundreds of unique PV are associated with the FA phenotype makes locus and allelic heterogeneity the canon in this disease. Because of this complexity, molecular diagnosis of the disease can be challenging.

Different working groups in the world have followed different strategies according to their population and available resources.<sup>4-6</sup> Finding the appropriate molecular strategy to study a group of FA patients can be a long and costly task.



The objective of this communication is to present the strategy followed by our group to determine the genotype of 50 patients with FA, in 40 of whom (80%), we were able to obtain the causal genotype of the disease. We hope that this work contributes to finding the most economical and effective genotyping strategy for patients with FA, in laboratories with characteristics such as ours.

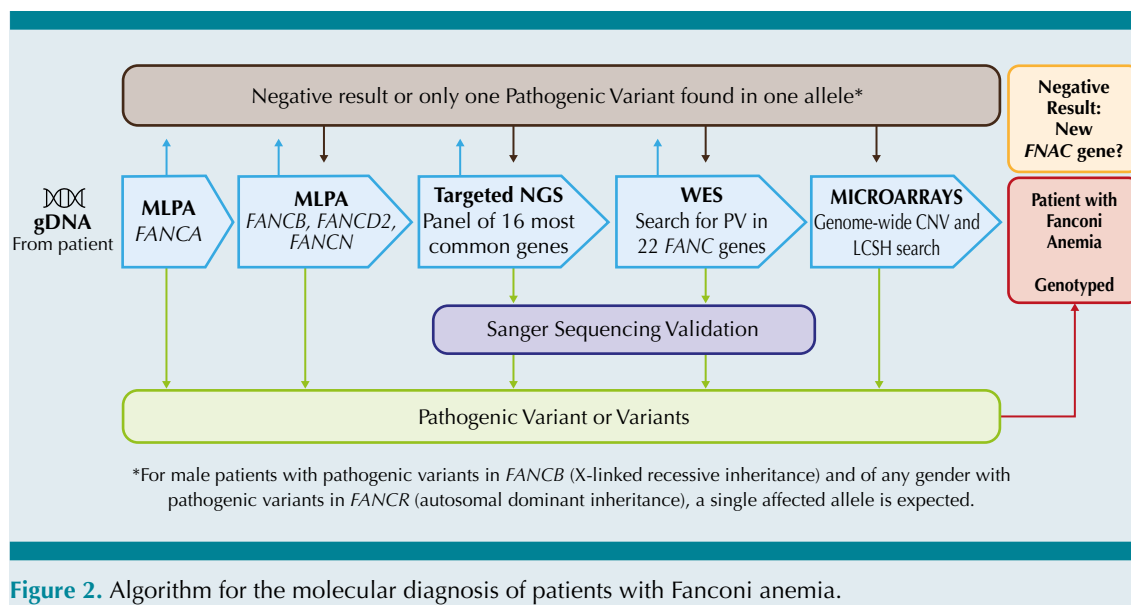
We start with biological samples from patients with a clinical and cytogenetic diagnosis of FA. The samples usually are obtained from peripheral

blood or bone marrow aspirate, from which we extract genomic DNA. Thereafter, we proceed according to the workflow suggested below (**Figure 2**):

1. We screen the *FANCA* gene through a Multiplex Ligation-Dependent Probe Amplification (MLPA) assay to detect long deletions, because *FANCA* deletions are one of the most frequent PV reported, and these deletions can range from one or multiple exons to the entire gene.
2. For negative patients, we perform MLPA analysis for *FANCD2*, *FANCN* (*PALB2*) and *FANCB* genes, which also present large deletions and duplications detectable by MLPA.
3. If no PV are found, or if they are detected in only one of the two alleles, we continue genotyping through next-generation sequencing (NGS), either by a targeted panel (in our case, composed of the 16 *FANCA* genes more commonly affected), or by whole-exome sequencing (WES), to detect small or point PV. In this way we found the variants responsible for the disease in most of the patients studied.
4. If we still have patients without genotype, the screening method we have followed is high-resolution microarrays with genome-wide markers for single nucleotide polymorphism (SNP) and copy number variants (CNV), for the search of large deletions or duplications at the *FANCA* genes. Microarrays also allow the detection of regions with long contiguous stretches of homozygosity (LCSH), in order to find uninterrupted homozygous alleles with neutral copy number state of 2 at the genomic level.

LCSH are observed throughout the human genome due to endogamy (consanguinity or reproductive isolation, both geographic and





**Figure 2.** Algorithm for the molecular diagnosis of patients with Fanconi anemia.

social) or evolutionary forces. Although LCSH are typical for inbred populations, they are also common and sometimes unexpectedly long in the genome of outbred people.<sup>7-10</sup> The importance of their detection is that they could indicate the presence of genes in a recessive state and if they are detected in an area that contains any of the 20 FANC genes with autosomal recessive inheritance, an intentional search can be made to detect VP in that locus (**Figure 2**).

## Methodology

### Patients

We include patients with clinical features of FA who tested positive for cytogenetic chromosome breakage. We explain the procedure and ask for written informed consent before obtaining the biological samples.

### Genomic DNA extraction

Blood or bone marrow aspirate should be collected in tubes containing EDTA as an anticoagulant. One of the most frequent clinical features in FA patients is the low count of white

cells in blood. Consequently, we expect to have a small starting material for the genomic DNA (gDNA) extraction.

We use the Gentra Puregene Blood Kit (QIAGEN, cat 158389, Venlo, Limburg, NL) (<https://www.qiagen.com/Gentra-Puregene-Handbook-12/2014-33>) with a few modifications. This protocol purifies gDNA from fresh or frozen samples of 3 ml whole blood. The protocol can also be used for gDNA purification from packed cells, buffy coat, or bone marrow. If processing packed cells or buffy coat samples, directly scale the volume of reagents used in proportion to the volume of the original blood sample.

### Procedure

- a. Dispense 3 ml whole blood or bone marrow in a 15 ml centrifuge tube, add 9 ml Red Blood Cell (RBC) Lysis Solution, and mix by inverting 10 times.
- b. Incubate 10 min at room temperature (15–25°C). Invert at least once during the incubation.

- c. Centrifuge for 5 min at 2000 Relative Centrifugal Force (RCF) to pellet the white blood cells; if there are remaining red cells, repeat the RBC lysis step.
- d. Discard the supernatant carefully by pipetting or pouring, leaving approximately 200  $\mu$ l of the residual liquid and the white blood cell pellet. Vortex the tube vigorously to resuspend the pellet in the residual liquid, the pellet should be completely dispersed after vortexing.
- e. Add 3 ml Cell Lysis Solution, and pipet up and down to lyse the cells or vortex vigorously for 10 s and incubate at 37°C until the solution is homogeneous.
- f. Add 15  $\mu$ l RNase A [4 mg/ml] Solution and mix by inverting 25 times. Incubate for 30 min at 37°C. Then add 15  $\mu$ l Proteinase K [20mg/ml] and incubate for 30 min at 55°C. Immediately put in ice to cool the sample.
- g. Add 1 ml Protein Precipitation Solution, vortex vigorously for 20 s at high speed and incubate on ice for 15 min.
- h. Centrifuge for 10 min/4°C at 2000 RCF. The precipitated proteins should form a tight, dark brown pellet. If the protein pellet is not tight, repeat the centrifugation.
- i. Dispense the supernatant from the previous step in a new 15 ml centrifuge tube. Be sure the protein pellet is not dislodged during pouring. Add 3 ml of isopropanol and 1  $\mu$ l of glycogen [20 mg/ml]
- j. Mix by inverting gently 50 times until the gDNA is visible as threads or a clump. Centrifuge for 10 min at 2000 RCF. The gDNA may be visible as a small white pellet.
- k. Carefully discard the supernatant or remove with a pipette tip. Add 3 ml of 70% ethanol and invert several times to wash the gDNA pellet.
- l. Centrifuge for 5 min at 2000 x RCF. Remove the supernatant, transfer the pellet with the remaining liquid to a 1.5 ml microcentrifuge tube, and wash again with 300  $\mu$ l of 70% ethanol. Centrifuge, remove the supernatant, and let dry the DNA in an open tube at room temperature or 65°C.
- m. Resuspend the gDNA pellet in 50  $\mu$ l of nuclease-free low Tris-EDTA (TE) buffer or nuclease-free water.
- n. The gDNA obtained can be quantified by a Nanodrop™ spectrophotometer (Nanodrop Technologies, Wilmington, DE). For NGS assays a more sensitive quantification could be required, such as fluorometric quantification (Qubit, Thermofisher).
- o. The integrity of the gDNA is evaluated by standard agarose gel electrophoresis 1% in Tris Acetate EDTA (TAE) 1X buffer.

#### Multiplex Ligation-Dependent Probe Amplification (MLPA)

The MLPA is a multiplex PCR-based assay for detecting CNV. In this method, probes specific to a genomic region hybridize to gDNA and then are amplified by PCR using a single pair of oligonucleotides.<sup>11</sup> The great advantage of this method is that it allows the simultaneous and semi-quantitative amplification of more than 60 probes in a single reaction for the rapid and accurate identification of deletions and/or duplications.<sup>12</sup>

PV found in *FANCA* accounts for approximately 70% of reported cases, of which 15% to 20% correspond to large deletions.<sup>13,14</sup> Most of them have been found mainly through MLPA, often spanning multiple exons.<sup>15</sup> We search these alterations with the MLPA kit developed by MRC Holland (<https://www.mrcholland.com/product/P057>).

In addition to *FANCA*, there are other *FANC* genes in which CNV have been frequently reported and that can also be assessed with this methodology: *FANCD2* (~ 3% of patients),<sup>16,17</sup> *FANCN* (<1% of patients)<sup>18</sup> and *FANCB* (~ 2% of patients).<sup>19</sup>

In our experience, the optimal amount of gDNA to be used for each reaction should be between 80-90 ng, diluted in a low TE buffer. On the other hand, each probe mix contains fragments that are used for quality control (usually 9), to evaluate parameters such as denaturation (“D” fragments), the amount of DNA, and the efficiency of the ligation step (“Q” fragments). It is important to include reference DNA samples from a) healthy individuals (both male and female), b) a FA positive sample (e.g., we use the FA cell line FA-A-VU817) and c) a negative sample (low TE buffer without DNA); these controls should be included during each run. It is suggested to use 3 reference samples for every 21 samples.

We follow the same workflow suggested in the general protocol by MRC Holland (<https://support.mrcholland.com/downloads/files/mlpa-general-protocol-one-tube>), which can be summarized in the following steps:

#### Procedure

- a. Denaturation of DNA and probe hybridization: Heat the DNA sample of interest (5 µl) for 5 minutes at 98°C. Then, add 3 µl of the probe master mix (1.5 µl SALSA probe mix + 1.5 µl MLPA buffer) to the DNA sample. Place the tube for 1 minute at 95°C, and hybridize for 16 h at 60°C.
- b. Ligation: After the hybridization time has elapsed, decrease the temperature of the thermal cycler to 54°C. Without removing the tubes from the equipment, add 32 µl of the ligation master mix and incubate at the same temperature for 15 min. To inactivate Ligase 65, incubate for 5 minutes at 98°C.
- c. PCR amplification: The reaction tubes are cooled to room temperature and 10 µl of the polymerase master mix is added. Start the PCR reaction, using the suggested program (95°C for 30 s, 60°C for 30 s and 72°C for 1 minute, for 35 cycles), with a final amplification at 72°C for 20 minutes. Store the tubes protected from light until capillary electrophoresis.
- d. Capillary electrophoresis: PCR products are loaded into a capillary electrophoresis instrument to be separated according to their length. The ratio of the components of the reaction mixture (final product of the MLPA reaction, weight marker, and formamide), as well as the electrophoretic run parameters, vary depending on the instrument available.
- e. Data analysis: Data review must be performed using the Coffalyzer program, developed by MRC Holland (<https://www.mrcholland.com/technology/software/coffalyser-net>). During this step, the relative copy number is determined by contrasting the relative peak heights of the reference and target probes in the test samples against those of the reference samples. It is essential to set up the study of the raw signals obtained with the version of each probe mix and the batch number with which the reaction was performed.

Once the electropherogram is obtained, the height of the fluorescence peaks after fragmentation is evaluated. The fluorescence level of each peak in a sample is compared with the signal observed in the DNA samples used as reference. Thus, deletion will be visualized as a decrease in peak height compared with the wild-type fragment, while a duplication increases the size of that fragment. Subsequently, the raw signals are normalized to obtain a value that reflects the adjusted ratio between the contrasting signals. This value is known as the dose coefficient

(DQ) and is used to discriminate the number of copies. The DQ ranges from 0 (homozygous deletions) to 2.15 (homozygous duplications). Values between 0.8 and 1 are considered normal (diploid set).<sup>20</sup>

### Next-Generation Sequencing (NGS)

NGS is a generic term used to describe the use of various technologies whose common feature is the massively parallel sequencing of multiple DNA fragments.<sup>21</sup> The combination of high throughput and single DNA molecule sequencing are hallmarks of NGS, regardless of the sequencing platform used.<sup>22</sup> The evolution and diversification of NGS assays have made it possible to overcome the limitations of Sanger sequencing and generate a huge amount of data, making it dependent on intensive computational analysis for the assembly of the sequences produced.<sup>23</sup>

The main steps of DNA NGS involve DNA fragmentation, library preparation, sequencing, bioinformatics analysis, annotation of the variants found, and interpretation.<sup>24</sup> However, it should be noted that the choice of the desired fragment length (short vs. long “reads”) depends on the research question and its clinical application.

The application of NGS technologies has been especially useful in disorders with high phenotypic variability, accelerating the identification and discovery of new causal variants, which has boosted its application in molecular diagnostics.<sup>25</sup> The reduction of costs and the potential of the information obtained have allowed the integration of NGS data into the clinical practice of medicine.<sup>26</sup> Currently, there are different approaches to NGS, depending on whether the scope of sequencing involves a large number of genes (targeted gene panels), protein-coding regions (whole-exome sequencing (WES)), or the entire genome (whole-genome sequencing (WGS)).

According to the most recent version of the AF Clinical Care Guideline, NGS should be performed as part of the genotyping process, selecting a panel of *FANC* genes in the first stage, appealing to WES in case of negative results by panel test.<sup>27</sup> We have followed this strategy for most of the cases analyzed, considering the benefits and limitations of each approach.

The use of dedicated panels has become widespread in the FA context, because the better characterization of the FA/BRCA pathway has allowed studying more *FANC* genes in a single assay, in addition to offering greater depth of coverage at a relatively low cost. However, WES has become increasingly attractive and affordable due to its capacity to sequence genetic regions in candidate or novel *FANC* genes.<sup>27</sup> An example that illustrates this idea is given by the description of bi-allelic PV in a group of patients with FA without genotype classification in *ERCC4/XPF*, whose loss of function had previously been associated with Xeroderma pigmentosum. This finding led to this gene being reclassified as *FANCC*.<sup>28</sup> Similarly, PV were discovered in *FANCT* (also known as *UBE2T*, in two Japanese patients)<sup>29</sup> and *FANCU*.<sup>30</sup>

Our search for PV in FA by NGS is summarized in the following bioinformatics workflow:

### Procedure

- a. Gene panel design. The in-silico design of specific probes for 16 *FANC* genes (*FANCA*, *FANCB*, *FANCC*, *FANCD1/BRCA2*, *FANCD2*, *FANCE*, *FANCF*, *FANCG*, *FANCI*, *FANCI/BRIP1*, *FANCL*, *FANCM*, *FANCN/PALB2*, *FANCO/RAD51C*, *FANCP/SLX4*, and *FANCCQ/ERCC4*) is performed with the SureDesign web-based application (Agilent Technologies). The final design is expected to cover 99.5% of all coding exons of these genes plus 50 bp of their flanking 5' and 3' intronic regions.

- b.** Digestion, library synthesis, and PCR amplification of targeted fragments are achieved using the Enrichment Haloplex System (Agilent Technologies Inc., 2013, Waghäusel-Wiesental, Germany).
- c.** Short paired-end reads are obtained through sequencing by synthesis technology, on an Illumina MiSeq platform (Illumina, San Diego, CA, USA).
- d.** In case the bioinformatics analysis of the panel data is negative, we proceed to use the WES technique following the suggested methodological adaptations.

Unlike panel design, in the case of WES, there are different commercial options focused on the capture of exonic regions (e.g., SureSelectXT Human All Exon V4 (Agilent Technologies), KAPA HyperExome (Roche), among others). In addition, the use of reagents and kit formulations compatible with the aforementioned systems extends to the enrichment and library construction steps. Finally, the sequencing platforms are adapted for a more scalable throughput (e.g., NovaSeq 6000 system, Illumina), due to the amplification of a larger number of DNA targets.

Note that the technical and interpretative complexity of NGS usually restricts its use to highly qualified public or private laboratories which can accomplish the recommendations and requirements suggested in international standards.<sup>31</sup>

Because of these demands, these laboratories frequently offer NGS services to researchers and clinicians. We have used this modality to obtain sequencing data from the submission of high-quality DNA samples. The selection of the approach, as well as the platforms, is dependent on the project purpose and resources available for screening FA patients.

#### NGS data analysis

Due to the large amounts of data generated from NGS, even from small gene panels, computa-

tional analyses are required to effectively process and evaluate the sequence information.<sup>25</sup> The set of bioinformatics algorithms that are executed in a succession of defined steps to process the massive raw data generated from the NGS platform is known as the bioinformatics pipeline.<sup>21</sup> Despite the heterogeneity of platforms and the scope of the study, the bioinformatics workflow consists of five major steps: quality control of raw data (from a text-based file format that contains sequence reads and individual base quality scores, known as FASTQ), preprocessing, mapping and postalignment, variant calling, and annotation and prioritization.<sup>23,32,33</sup>

We developed an analytical pipeline (validated for data analysis) to streamline computational analysis and maximize variant calling accuracy, based on GATK best practice recommendations.<sup>34-36</sup> The programs used for file manipulation are often packaged in the GATK Toolkit, a collection of command-line tools for analyzing high-throughput sequencing data with a primary focus on variant discovery. It is very important to highlight that these tools only run in Linux or MacOSX operating systems.

#### *Procedure (the program's options are named in parentheses):*

- a.** Review quality control of sequence reads (FastQC).
- b.** Filter and trim the reads with base poor quality and sequencing adapters (Trimomatic, Cutadapt, and Trim Galore).
- c.** Trimmed reads are aligned to map with a reference sequence (Burrows-Wheeler Alignment (BWA-MEM), using the references found at UCSC or Ensembl databases).
- d.** SAM files are transformed to BAM files and mark duplicates (Samtools and Picard).
- e.** Check the quality of BAM files (Qualimap).

**f.** Follow the recommendations of the GATK to get the variant calling file (VCF), in gVCF mode.

**g.** Finally, we annotate the variants (VEP, SnpEff, or ANNOVAR). Variants are described following the guidelines proposed by the Human Genome Variation Society (HGVS) nomenclature.<sup>37</sup>

### Variant Classification

Once we have filtered the candidate variants for each patient, the next step is to classify them according to their degree of pathogenicity. The most widely used system to perform this task comes from the recommendations suggested by the American College of Medical Genetics and the Association for Molecular Pathology (ACMG-AMP).<sup>38,39</sup> According to this guide, the criteria for variant classification is a score based on five classes: benign, likely benign, variant of unknown significance (VUS), likely pathogenic, and pathogenic.

The criteria used for this classification are based on an evidence framework, in which different parameters (functional, population, segregation data, computational and predictive algorithms, among others) are considered to establish a final statement. Several web tools facilitate the classification exercise, as they incorporate valuable information from databases (such as ExAc, ClinVar, gnomAD, among others) into a single platform (e.g. Varsome).

### NGS validation

Because NGS is not an error-free technique, the ACMG and the European Society of Human Genetics (ESHG) strongly recommend validating the selected PV. Technical as well as bioinformatics factors are related to the difficulty of confirming variants, mainly in genes with pseudogenes, repetitive regions, and with high GC content or hotspot variants. Sanger sequencing can solve this problem.

In our case, we design specific flanking primer pairs to validate the PV founded by NGS, using the Primer3 algorithm (<http://bioinfo.ut.ee/primer3-0.4.0/>), one of the most widely used primer designing tools<sup>40</sup>. PCR amplification is carried out from 100 ng of DNA template. PCR products are purified with QIAquick PCR Purification kit (QIAGEN, Venlo, Limburg, NL) according to manufacturer instructions. Purified amplicons are bidirectional sequenced using the Big Dye Terminator sequencing kit and resolved on an Applied Biosystems 3130 Genetic Analyzer (Applied Biosystems, Foster City, CA, USA).

The electropherograms are analyzed with Chromas V2.6.6 software ([www.technelysium.com.au](http://www.technelysium.com.au)) and target sequences are compared to the corresponding reference sequences from GenBank (<https://www.ncbi.nlm.nih.gov>). There are other alternative programs for analysis (<https://www.genewiz.com/Public/Resources/Tools-for-Viewing-Sequencing-Data/>) or web applications such as GEARgenomics (<https://www.gear-genomics.com/>), supported by EMBL-EBI, EMBL-European Bioinformatics Institute).

### Sanger sequencing as genotyping strategy

Sanger sequencing could be performed to search PV in specific genes, i.e., when members of a family with a previously identified PV are studied or in individuals belonging to a community where a founder mutation has been reported, particularly, in genetically isolated populations such as the FANCA c.295C>T in Spanish Gypsies,<sup>41</sup> FANCC c.456+4A>T in Ashkenazi Jews,<sup>42</sup> FANCC c.67delG in the Mennonite community<sup>43</sup> and FANCG c.511-3\_511-2delCA in a Mexican Mixe community.<sup>44</sup>

### High-resolution microarray analysis

We perform the analysis of the gDNA with Affymetrix CytoScan750K arrays (Affymetrix, Santa Clara, CA, USA). This array provides a genome-wide coverage, including 550,000 markers for

detecting copy number variation and 200,436 SNP probes.

The steps for DNA digestion, ligation, PCR amplification, fragmentation, labeling, denaturing, and hybridization into the array are performed according to the protocols and QC guidelines provided by the supplier.

Arrays are then stained and washed in the Affymetrix GeneChip Fluidics Station 450 and scanned using an Affymetrix GeneChip Scanner 3000 7G (Affymetrix, Santa Clara, CA, USA).

The files obtained were analyzed with the software Chromosome Analysis Suite (ChAS) software version 4.1, provided by Affymetrix (Affymetrix, Santa Clara, CA, USA). (NetAffx 33 hg19 annotation files) (<http://www.affymetrix.com>). The reporting threshold of the copy number (CN) is performed at "high resolution" settings: gains or losses in 100Kb with marker count  $\geq 50$  probes.

#### Analysis of long-contiguous stretch of homozygosity (LCSH)

Minimal thresholds for LCSH calls in clinical analyses are generally set at 3 to 10 Mb.<sup>45</sup> We visualize the LCSH in ChAS software, version 4.1, provided by Affymetrix (Affymetrix, Santa Clara, CA, USA). For the analysis, the NetAffx 33 hg19 annotation files are used, (<http://www.affymetrix.com>). For LCSH $>3$  Mb, the analysis configuration is set at LOH with marker count = 50 and size = 3000 Kbp, and for LCSH $>5$ Mb is set at marker count = 50 and size = 5000 Kbp.

#### Results report

Once the variants have been classified and validated, the next step is to write a report. In addition to the ACMG, other organizations, such as Clinical Genome Resource (ClinGen), European Society of Human Genetics (ESHG), and the Association for Clinical Genetic Science

(ACGS) have developed guidelines to advise laboratories in reporting and describing variants<sup>38,39,46-48</sup>. These documents are issued from a laboratory to a treating physician and describe the results of genetic analysis. They should be clear, concise, and follow the recommendations suggested in existing guidelines<sup>48</sup>.

#### Recommendations

- a. In general, the report format includes information on the requesting clinician, the laboratory where the analysis was performed, and the patient.
- b. For the description and interpretation of the results, the nomenclature of the variants should be written according to the Human Genome Organization (HUGO, <https://www.genenames.org/>) and the Human Genome Variation Society (HGVS, <https://varnomen.hgvs.org/>).

The report should include a clear interpretation of the test results based on a review of the patient's clinical context and available family information.<sup>48</sup> The goal should be to generate a simple, unambiguous statement that explains the clinical implications of the result.

- c. Methodological details (principle of the technology used, sensitivity, resolution, limitations, bioinformatics tools, among others) relevant to the interpretation of the results should be included in the report. For example, in the case of massive sequencing data for constitutional variants, it is advisable to refer to the most recent guidelines developed by the ACMG.
- d. Each laboratory should document the evidence used to classify variants concerning the impact they may have on the condition studied and establish precise policies to define which variants will be excluded from reporting.<sup>39</sup>

In 2008, the Fanconi Anemia Mutation Database was created as a cooperative effort to accelerate access to information on variants found in the FANC genes. These data are available using the Leiden Open-Source Variation Database (<https://www2.rockefeller.edu/fanconi/mutate/>). The addition of a variant (new or previously reported) should be done as an institution and not on an individual basis. Sharing data with the scientific community significantly increases the amount of existing evidence for a particular variant in public databases and contributes to improving approaches in the search for genotype-phenotype correlations.<sup>49</sup> Additionally, this practice should be encouraged because it facilitates variant reclassification, which is especially important in the case of VUS, where the use of this information adds to the evidence generated through predictive biochemical algorithms and functional assay validation.<sup>50</sup>

- e. Negative results in a patient with positive chromosomal breakage study.

Negative molecular test results should be carefully interpreted for an individual with a chromosomal breakage test within the FA range. One explanation for a negative result is the presence of a variant(s) in an undiscovered FANC gene or a type of variant in a gene that cannot be identified with the technology employed; another possibility is somatic mosaicism.<sup>51,52</sup> Analysis of a tissue alternative to the peripheral blood or bone marrow, such as fibroblasts, may be considered in individuals presenting with an FA phenotype and negative genetic studies on peripheral blood.

## DISCUSSION

In a variable percentage of patients (in our case it was 20%), it is not possible to reach the genotype. Possible reasons are that they could be variants that are not detected by exome sequencing, such as those found in intronic or regulatory regions of genes, and that would

require another more expensive strategy, whole genome sequencing. Alternatively, it could be because the variants are found in genes not yet described as FANC genes, so it is advisable to analyze the genes functionally associated with the FA/BRCA pathway.

The great genetic heterogeneity of FA makes it difficult to implement detection techniques classically used in other types of diseases of Mendelian inheritance. For this reason, we consider that the detailed disclosure of the methodological aspects inherent to genotyping in FA is essential to establish the minimum biological and technological requirements to identify the variants responsible for the disease, especially in geographical areas where this approach is not used or is limited. This methodological strategy, which ranges from obtaining the gDNA to the generation of the report, puts in context the importance of the information obtained (such as the frequency, location and type of variant) to develop a cost-effective and appropriate genotyping strategy for the FA population.

## CONCLUSIONS

Describing the FA genotype in a population is relevant for several reasons: first, it improves the characterization of patients and thus contributes to the identification of FA genotype-phenotype correlations. Second, it refines the tailoring of genetic diagnostic strategies in specific groups of patients, including adequate selections of non-FA sibling bone marrow donors and appropriate genetic counseling. Finally, the description of further cases with seldomly described genotypes brings new data to consider when confronted with variants with a history of ambiguous pathogenicity classification

## Acknowledgements

This research was partially funded by Consejo Nacional de Ciencia y Tecnología (CONACyT), grant number SALUD-2014-1-233721; Insti-



tuto Nacional de Pediatría, Recursos Fiscales E022 project INP 2014/41; Programa de Apoyo a Proyectos de Investigación e Innovación Tecnológica (PAPIIT), Universidad Nacional Autónoma de México (UNAM) Grant Number IN205120. The authors thank M. en C. Francisco Zúñiga-Frias for the English revision of this manuscript.

## REFERENCIAS

- Fiesco-Roa MÓ, García-de Teresa B, Leal-Anaya P, van 't Hek R, Wegman-Ostrosky T, Frias S, et al. Fanconi anemia and dyskeratosis congenita/telomere biology disorders: Two inherited bone marrow failure syndromes with genomic instability. *Front Oncol* [Internet]. 2022;12. Available from: <https://www.frontiersin.org/articles/10.3389/fonc.2022.949435>
- García-de-Teresa B, Rodríguez A, Frias S. Chromosome Instability in Fanconi Anemia: From Breaks to Phenotypic Consequences. *Genes (Basel)*. 2020 Dec;11(12).
- Niraj J, Färkkilä A, D'Andrea AD. The Fanconi Anemia Pathway in Cancer. *Annu Rev cancer Biol*. 2019 Mar;3:457–78.
- Pilonetto D V, Pereira NF, Bonfim CMS, Ribeiro LL, Bitencourt MA, Kerkhoven L, et al. A strategy for molecular diagnostics of Fanconi anemia in Brazilian patients. *Mol Genet genomic Med*. 2017 Jul;5(4):360–72.
- Mori M, Hira A, Yoshida K, Muramatsu H, Okuno Y, Shiraishi Y, et al. Pathogenic mutations identified by a multimodality approach in 117 Japanese Fanconi anemia patients. *Haematologica*. 2019 Oct;104(10):1962–73.
- Ben Haj Ali A, Messaoud O, Elouej S, Talmoudi F, Ayed W, Mellouli F, et al. FANCA Gene Mutations in North African Fanconi Anemia Patients. *Front Genet* [Internet]. 2021;12. Available from: <https://www.frontiersin.org/articles/10.3389/fgene.2021.610050>
- Li LH, Ho SF, Chen CH, Wei CY, Wong WC, Li LY, et al. Long contiguous stretches of homozygosity in the human genome. *Hum Mutat*. 2006 Nov;27(11):1115–21.
- Pajusalu S, Žilina O, Yakoreva M, Tammur P, Kuuse K, Mõlter-Väär T, et al. The Diagnostic Utility of Single Long Contiguous Stretches of Homozygosity in Patients without Parental Consanguinity. *Mol Syndromol*. 2015 Sep;6(3):135–40.
- Wang JC, Ross L, Mahon LW, Owen R, Hemmat M, Wang BT, et al. Regions of homozygosity identified by oligonucleotide SNP arrays: evaluating the incidence and clinical utility. *Eur J Hum Genet*. 2015 May;23(5):663–71.
- Chaves TF, Oliveira LF, Ocampos M, Barbato IT, de Luca GR, Barbato Filho JH, et al. Long contiguous stretches of homozygosity detected by chromosomal microarrays (CMA) in patients with neurodevelopmental disorders in the South of Brazil. *BMC Med Genomics*. 2019 Mar;12(1):50.
- Stuppia L, Antonucci I, Palka G, Gatta V. Use of the MLPA assay in the molecular diagnosis of gene copy number alterations in human genetic diseases. *Int J Mol Sci*. 2012;13(3):3245–76.
- Hömig-Hölzel C, Savola S. Multiplex ligation-dependent probe amplification (MLPA) in tumor diagnostics and prognostics. *Diagnostic Mol Pathol Am J Surg Pathol part B*. 2012 Dec;21(4):189–206.
- Ameziane N, Errami A, Léveillé F, Fontaine C, de Vries Y, van Spaendonk RML, et al. Genetic subtyping of Fanconi anemia by comprehensive mutation screening. *Hum Mutat*. 2008 Jan;29(1):159–66.
- Castella M, Pujol R, Callén E, Trujillo JP, Casado JA, Gille H, et al. Origin, functional role, and clinical impact of Fanconi anemia FANCA mutations. *Blood*. 2011 Apr;117(14):3759–69.
- Kimble DC, Lach FP, Gregg SQ, Donovan FX, Flynn EK, Kamat A, et al. A comprehensive approach to identification of pathogenic FANCA variants in Fanconi anemia patients and their families. *Hum Mutat*. 2018 Feb;39(2):237–54.
- Kalb R, Neveling K, Hoehn H, Schneider H, Linka Y, Batish SD, et al. Hypomorphic mutations in the gene encoding a key Fanconi anemia protein, FANCD2, sustain a significant group of FA-D2 patients with severe phenotype. *Am J Hum Genet*. 2007 May;80(5):895–910.
- Flynn EK, Kamat A, Lach FP, Donovan FX, Kimble DC, Narisu N, et al. Comprehensive analysis of pathogenic deletion variants in Fanconi anemia genes. *Hum Mutat*. 2014 Nov;35(11):1342–53.
- Xia B, Dorsman JC, Ameziane N, de Vries Y, Roommans MA, Sheng Q, et al. Fanconi anemia is associated with a defect in the BRCA2 partner PALB2. *Nat Genet*. 2007 Feb;39(2):159–61.
- Jung M, Ramanagoudr-Bhojappa R, van Twest S, Rosti RO, Murphy V, Tan W, et al. Association of clinical severity with FANCB variant type in Fanconi anemia. *Blood*. 2020 Apr;135(18):1588–602.
- Coffa J, van den Berg J. Analysis of MLPA Data Using Novel Software Coffalyser.NET by MRC-Holland. In: Eldin AB, editor. *Modern Approaches To Quality Control* [Internet]. Rijeka: IntechOpen; 2011. Available from: <https://doi.org/10.5772/21898>
- Roy S, Coldren C, Karunamurthy A, Kip NS, Klee EW, Lincoln SE, et al. Standards and Guidelines for Validating Next-Generation Sequencing Bioinformatics Pipelines: A Joint Recommendation of the Association for Molecular Pathology and the College of American Pathologists. *J Mol Diagn*. 2018 Jan;20(1):4–27.
- Hu T, Chitnis N, Monos D, Dinh A. Next-generation sequencing technologies: An overview. *Hum Immunol*. 2021 Nov;82(11):801–11.
- Voelkerding K V, Coonrod EM, Durtschi JD, Margraf RL. Next-Generation Sequencing: Principles for Clinical Application BT - Molecular Pathology in Clinical Practice. In: Leonard DGB, editor. Cham: Springer International Publishing; 2016. p. 889–909. Available from: [https://doi.org/10.1007/978-3-319-19674-9\\_59](https://doi.org/10.1007/978-3-319-19674-9_59)

24. Qin D. Next-generation sequencing and its clinical application. *Cancer Biol Med*. 2019 Feb;16(1):4–10.
25. den Dunnen JT. Efficient variant data preparation for Human Mutation manuscripts: Variants and phenotypes. Vol. 40, Human mutation. United States; 2019. p. 1009.
26. Rizzo JM, Buck MJ. Key principles and clinical applications of “next-generation” DNA sequencing. *Cancer Prev Res (Phila)*. 2012 Jul;5(7):887–900.
27. Sroka I, Frohnmayer L, Wirkkula L. Fanconi Anemia Clinical Care Guidelines [Internet]. Fifth Edit. Vol. 1, Fanconi Anemia Research Fund, Inc. 2020. 1–276 p. Available from: <http://marefateadyan.nashriyat.ir/node/150>
28. Bogliolo M, Schuster B, Stoeper C, Derkunt B, Su Y, Raams A, et al. Mutations in ERCC4, encoding the DNA-repair endonuclease XPF, cause Fanconi anemia. *Am J Hum Genet*. 2013 May;92(5):800–6.
29. Hira A, Yoshida K, Sato K, Okuno Y, Shiraishi Y, Chiba K, et al. Mutations in the gene encoding the E2 conjugating enzyme UBE2T cause Fanconi anemia. *Am J Hum Genet*. 2015 Jun;96(6):1001–7.
30. Shamseldin HE, Elfaki M, Alkuraya FS. Exome sequencing reveals a novel Fanconi group defined by XRCC2 mutation. *J Med Genet*. 2012 Mar;49(3):184–6.
31. Rehml HL, Bale SJ, Bayrak-Toydemir P, Berg JS, Brown KK, Deignan JL, et al. ACMG clinical laboratory standards for next-generation sequencing. *Genet Med Off J Am Coll Med Genet*. 2013 Sep;15(9):733–47.
32. McCombie WR, McPherson JD, Mardis ER. Next-Generation Sequencing Technologies. *Cold Spring Harb Perspect Med*. 2019 Nov;9(11).
33. Bao R, Huang L, Andrade J, Tan W, Kibbe WA, Jiang H, et al. Review of current methods, applications, and data management for the bioinformatics analysis of whole exome sequencing. *Cancer Inform*. 2014;13(Suppl 2):67–82.
34. DePristo MA, Banks E, Poplin R, Garimella K V, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet*. 2011 May;43(5):491–8.
35. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, et al. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinforma*. 2013;43(1110):11.10.1-11.10.33.
36. Franke KR, Crowley EL. Accelerating next generation sequencing data analysis: an evaluation of optimized best practices for Genome Analysis Toolkit algorithms. *Genomics Inform*. 2020 Mar;18(1):e10.
37. den Dunnen JT, Dalgleish R, Maglott DR, Hart RK, Greenblatt MS, McGowan-Jordan J, et al. HGVS Recommendations for the Description of Sequence Variants: 2016 Update. *Hum Mutat*. 2016 Jun;37(6):564–9.
38. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, 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 Off J Am Coll Med Genet*. 2015 May;17(5):405–24.
39. Rehder C, Bean LJH, Bick D, Chao E, Chung W, Das S, et al. Next-generation sequencing for constitutional variants in the clinical laboratory, 2021 revision: a technical standard of the American College of Medical Genetics and Genomics (ACMG). *Genet Med Off J Am Coll Med Genet*. 2021 Aug;23(8):1399–415.
40. Kumar A, Chordia N. In Silico PCR Primer Designing and Validation BT - PCR Primer Design. In: Basu C, editor. New York, NY: Springer New York; 2015. p. 143–51. Available from: [https://doi.org/10.1007/978-1-4939-2365-6\\_10](https://doi.org/10.1007/978-1-4939-2365-6_10)
41. Callén E, Casado JA, Tischkowitz MD, Bueren JA, Creus A, Marcos R, et al. A common founder mutation in FANCA underlies the world’s highest prevalence of Fanconi anemia in Gypsy families from Spain. *Blood*. 2005 Mar;105(5):1946–9.
42. Whitney MA, Jakobs P, Kaback M, Moses RE, Grompe M. The Ashkenazi Jewish Fanconi anemia mutation: incidence among patients and carrier frequency in the at-risk population. *Hum Mutat*. 1994;3(4):339–41.
43. García-de Teresa B, Frias S, Molina B, Villarreal MT, Rodríguez A, Carnevale A, et al. FANCC Dutch founder mutation in a Mennonite family from Tamaulipas, México. *Mol Genet genomic Med*. 2019 Jun;7(6):e710.
44. Reyes P, García-de Teresa B, Juárez U, Pérez-Villatoro F, Fiesco-Roa MO, Rodríguez A, et al. Fanconi Anemia Patients from an Indigenous Community in Mexico Carry a New Founder Pathogenic Variant in FANCG. *Int J Mol Sci*. 2022 Feb;23(4).
45. Kearney HM, Thorland EC, Brown KK, Quintero-Rivera F, South ST. American College of Medical Genetics standards and guidelines for interpretation and reporting of postnatal constitutional copy number variants. *Genet Med Off J Am Coll Med Genet*. 2011 Jul;13(7):680–5.
46. Rehml HL, Berg JS, Brooks LD, Bustamante CD, Evans JP, Landrum MJ, et al. ClinGen—the Clinical Genome Resource. *N Engl J Med*. 2015 Jun;372(23):2235–42.
47. Matthijs G, Souche E, Alders M, Corveleyn A, Eck S, Feenstra I, et al. Guidelines for diagnostic next-generation sequencing. *Eur J Hum Genet*. 2016 Oct;24(10):1515.
48. Smith K, Martindale J, Wallis Y, Bown N, Leo N, Creswell L, et al. General genetic laboratory reporting recommendations. *Birmingham Assoc Clin Genet Sci*. 2015;11.
49. Directors ABO. Laboratory and clinical genomic data sharing is crucial to improving genetic health care: a position statement of the American College of Medical Genetics and Genomics. *Genet Med Off J Am Coll Med Genet*. 2017 Jul;19(7):721–2.
50. Federici G, Soddu S. Variants of uncertain significance in the era of high-throughput genome sequencing: a lesson from breast and ovary cancers. *J Exp Clin Cancer Res*. 2020 Mar;39(1):46.



51. Gregory JJ, Wagner JE, Verlander PC, Levran O, Batish SD, Eide CR, et al. Somatic mosaicism in Fanconi anemia: evidence of genotypic reversion in lymphohematopoietic stem cells. *Proc Natl Acad Sci U S A*. 2001 Feb;98(5):2532–7.
52. Ikeda H, Matsushita M, Waisfisz Q, Kinoshita A, Oostra AB, Nieuwint AWM, et al. Genetic reversion in an acute myelogenous leukemia cell line from a Fanconi anemia patient with biallelic mutations in BRCA2. *Cancer Res*. 2003 May;63(10):2688–94.