

# **Protocolo de desarrollo y producción de paneles génicos de apoyo al diagnóstico**

## ÍNDICE

<b>CAPÍTULO</b>	<b>PÁGINA</b>
1. Objeto.....	3
2. Definiciones .....	3
3. Proceso de producción.....	4
4. Tiempo de respuesta en la preparación de los INFORMES .....	6
5. Características de los paneles .....	6
6. Anexos .....	6
Anexo I: Muestra requerida	
Anexo II: Envío	
Anexo III: Documento de solicitud de análisis	
Anexo IV: Preparación de librerías con enriquecimiento de las zonas a secuenciar	
Anexo V: Generación de clusters en la flow cell y secuenciación en el Miseq de Illumina	
Anexo VI: Análisis bioinformático	
Anexo VII: Características del informe final	

## 1. Objeto

Este documento tiene por objeto describir el protocolo general seguido en nuestra unidad para la ejecución de análisis genéticos basados en tecnología de secuenciación masiva.

## 2. Definiciones

### a. Secuenciación masiva

Secuenciación de ADN realizada en plataformas de segunda generación, capaces de generar a la vez cientos de miles de reacciones de secuencia en paralelo (alto rendimiento) gracias a la inmovilización de las reacciones en una superficie sólida. De esta forma la cantidad de reactivos necesarios se minimiza al máximo (podrían llamarse nanoreacciones) y se abarata el coste por base leída varios órdenes de magnitud.

### b. Paneles génicos de apoyo al diagnóstico

Definición conceptual de un análisis múltiple en paralelo de múltiples genes relacionados con fenotipos parecidos o solapantes. Realizando este análisis ante la presencia de un fenotipo concreto se analizan de una vez y a un precio asequible múltiples genes.

### c. Cobertura

Número de veces que una determinada base nucleotídica es leída durante un experimento de secuenciación masiva.

### d. Informe de apoyo al diagnóstico

Informe en el que se relatan las variantes raras detectadas en un paciente después de analizar un determinado panel y se valora su posible relación con el fenotipo descrito.

### e. Tiempo de respuesta

Tiempo que se tarda entre la recepción de la muestra y el envío del informe vía correo electrónico

### 3. Proceso de producción de los paneles.

**Ciente** que desea el análisis:

- El cliente prepara una muestra de sangre o ADN que deberá cumplir las especificaciones contenidas en el Anexo I.
- El cliente envía la muestra a la unidad por sus propios medios según las instrucciones detalladas en Anexo II y III.

La **unidad** recibe una muestra de sangre/ADN para realizar un análisis por secuenciación masiva y procede de la siguiente forma:

- **Registro.** La unidad efectúa el registro de la muestra y envía confirmación electrónica de llegada al cliente. El registro de la empresa de mensajería utilizada servirá como registro de tiempo inicial para calcular el tiempo de respuesta.
- **Verificación.** Verificación de la cantidad y calidad del ADN enviado. Si la muestra no cumple criterios de calidad especificados en el Anexo II se enviará un mail a la persona de contacto del cliente y se procederá como si no hubiese sido recibida.
- **Preparación de las librerías** de ADN con captura de secuencia. El protocolo seguido para este paso se describe en el Anexo IV.
- Generación de clusters en la flow cell y **secuenciación** en el Miseq (Illumina). El protocolo seguido para este paso se describe en el Anexo V.
- **Análisis bioinformático.** Ver anexo VI.  
Incluye:  
Análisis primario: Análisis de imagen y asignación de bases  
Análisis secundario: Alineamiento, detección de variantes y anotación.

Análisis terciario: Búsqueda y análisis de las variantes más probablemente patogénicas. Análisis conjunto de variantes y fenotipo del paciente.

- **Realización del informe** final de resultados. Este informe tendrá las características que se especifican en el Anexo VII.

- Envío del informe por correo electrónico a la persona de contacto del cliente. Ese momento determinará el tiempo de respuesta. Se requerirá confirmación de recepción del mismo por parte de la persona de contacto.

**Nota:** En el caso de que alguna de las variantes detectadas por secuenciación masiva no tenga la cobertura y calidad mínima, fijada por nuestro laboratorio en el momento actual en 20X y 200 respectivamente, se procederá al estudio de la variante mediante secuenciación clásica (método Sanger) sin gasto alguno adicional para el cliente. Esta peculiaridad se notificará en el informe, pero el tiempo necesario para realizar la comprobación no se computará en el tiempo de respuesta.

- Envío por correo ordinario del informe firmado.
- En el caso de que se requiera por parte del cliente el estudio familiar de una o más variantes detectadas con un panel, se enviarán las muestras de padres y hermanos del mismo modo que la del paciente (Anexo II), junto con una nueva hoja de solicitud (Anexo III). Este estudio se realizará por secuenciación clásica (método Sanger).
- En el caso de que se requiera por parte del cliente el estudio prenatal para una familia previamente analizada por la unidad, se deberá enviar una muestra de ADN extraído de vellosidades coriónicas, junto con una nueva hoja de solicitud y consentimiento informado. Este estudio se realizará por secuenciación clásica (método Sanger). Se deberá solicitar el estudio a la UNIDAD con, como mínimo, un mes de antelación al envío de la muestra de ADN.

#### 4. Tiempo de respuesta en la preparación de los INFORMES

El tiempo de respuesta esta entre 1-3 meses.

#### 5. Características de los paneles

- La cobertura media de los paneles será de 120X. Las regiones con cobertura menor que la indicada para diagnóstico genético (20X) se informarán específicamente para cada análisis.
- La calidad de las lecturas utilizadas para la realización del informe final será de >20 Phred units (alta calidad).

#### 6. Anexos

##### 6.1. Anexo I: Muestra requerida

Tipo de muestra requerida para secuenciación masiva:

1\_ Sangre total extraída en tubo con anticoagulante EDTA (tubo malva)

2\_ADN genómico con los siguientes requisitos:

- Cantidad: debe contener >6000 ng de ADN genómico
- Calidad: sin fragmentar (una única banda sin *smear* en un gel de agarosa al 0.8%)
- Volumen : entre 50 y 200 µl

##### 6.2. Anexo II: Envío

Requisitos para envío de muestras a la UNIDAD para análisis por secuenciación masiva

1\_ Forma de envío de la muestra:

Introducir el tubo de sangre total o bien un tubo eppendorf con ADN genómico debidamente precintado con parafilm, en un sobre acolchado y enviar a temperatura ambiente. Se deberá informar por email de la salida de la/s muestra/s para estar pendientes.

El envío se realizará por correo ordinario a la siguiente dirección:

A/A Dra. Marmiesse/ Dr Cocho de Juan  
Laboratorio de Metabolopatías  
Hospital Clinico Universitario de Santiago

Rua Choupana s/n  
CP 15706  
Santiago de Compostela  
La Coruña

## 2\_Documentos deberán acompañar a la muestra:

- Informe clínico del paciente
- Hoja de solicitud debidamente cumplimentada (Anexo III)
- Consentimiento informado o bien la declaración de la existencia de este.

### 6.3. Anexo III: Documento de solicitud de análisis

Para:	Laboratorio de Metabolopatías  Unidad de Diagnóstico y Tratamiento de Errores Congénitos del Metabolismo  Complejo Hospitalario Universitario de Santiago de Compostela C/ Choupana, s/n CP: 15706 Santiago de Compostela, La Coruña
Tel/Fax	+34 981 95 01 00 / +34 981 95 01 02
E-mail:	amarmiesse@gmail.com; cochoja@yahoo.es

Fecha de solicitud:	
Médico solicitante:	
Servicio:	
Hospital:	
Dirección:	
Código Postal:	Localidad:
Teléfono:	Fax:
e-mail:	

Identificación del paciente	Sexo (V,M)	Fecha de Nacimiento	Fecha de toma de muestra	Prueba solicitada*	Tipo de muestra**

Sospecha/s diagnóstica:

\* Prueba solicitada: estudio de portador / estudio prenatal / estudio del panel génico gRX

\*\* Tipo de muestra: sangre, DNA, leucocitos.

Firma:



#### **6.4. Anexo IV: Preparación de librerías con enriquecimiento de las zonas a secuenciar**

Para enriquecer el ADN genómico del paciente en las zonas que pretendemos secuenciar se utilizarán los kits de Agilent: Sure SelectXT Reagent kit, HSQ y Sure Select Capture Library Custom 0,5-2,9 Mb (16 reacciones). Para ello se seguirá paso por paso el Protocolo SureSelectXT Target Enrichment System for Illumina Paired-End Sequencing Library.

Los pasos del proceso son:

- 1\_Fragmentación de 3 ug de ADN genómico por sonicación (Bioruptor-Diagenode).
- 2\_Purificación de la muestra fragmentada (Agencourt AMPure XP beads).
- 3\_Control de calidad de la muestra fragmentada utilizando un DNA 1000 chip en el Agilent 2100 Bioanalyzer.
- 4\_Reparación de los extremos de los fragmentos generados (T4 DNA polimerasa, Klenow, T4 PNK).
- 5\_Purificación del ADN reparado (Agencourt AMPure XP beads).
- 6\_Adición de dATP a los extremos 3' de los fragmentos de DNA (Klenow exo).
- 7\_Purificación de la muestra (Agencourt AMPure XP beads).
- 8\_Ligación de los adaptadores específicos de la plataforma de secuenciación Illumina.
- 9\_Purificación de la muestra (Agencourt AMPure XP beads).
- 10\_Control de cantidad/calidad de la librería (DNA 1000 chip- Agilent 2100 Bioanalyzer).
- 11\_Hibridación de las librerías con la solución de captura (oligonucleótidos biotinilados de captura), según protocolo SureSelect de Agilent.
- 12\_Recuperación de las regiones capturadas (DynaLink MyOne Streptavidin T1 magnetic beads).
- 13\_Purificación de la muestra (Agencourt AMPure XP beads).
- 14\_Amplificación por PCR de la librería capturada (Herculase II Fusion DNA Polymerase). En esta amplificación se incluyen los index que identificarán las muestras que van juntas en el mismo run de Miseq.
- 15\_Purificación de la muestra (Agencourt AMPure XP beads).
- 16\_Cuantificación de las librerías generadas y control de calidad (2100 Bioanalyzer High Sensitivity DNA, Agilent).
- 17\_Dilución de las librerías, mezcla de librerías que van en el mismo run y preparación para secuenciación.

## 6.5. Anexo V: Generación de clusters en la flow cell y secuenciación en el Miseq de Illumina

La secuenciación se realiza en la plataforma MiSeq de la casa Illumina. Para ello primero genera millones de clusters de secuencias de DNA (representativos de los fragmentos de DNA que constituyen cada una de las librerías) en la superficie de la flow cell. Seguidamente los clusters de DNA sirven como matrices de secuenciación y en cada ciclo hay incorporación de 1 nucleótido y registro de la base incorporada. Se utilizan los reactivos correspondiente al kit MiSeq® Reagent Kit v2 (300 cycles) de Illumina. Se realiza un run de secuenciación de 75 ciclos de "paired end", con 6 ciclos para la secuenciación de cada index

## 6.6. Anexo VI: Análisis bioinformático

### **Análisis primario y secundario**

Para el presente análisis se emplean los siguientes programas:

- FastQC v0.10.1 (<http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc/>) para control de calidad
- el software de alineamiento BWA v0.7.5a
- el software NGSrich v0.7.5 (<http://ngsrich.sourceforge.net>) como control previo a la detección de variantes
- BEDTools.2.17.0 (<http://bedtools.readthedocs.org/en/latest/#>)
- Picard 1.93 (<http://picard.sourceforge.net>) para pasos intermedios
- Samtools v0.1.19 (Li et al, 2009) para detección de variantes polimórficas
- el software Annovar Nov2011 (Wang et al 2010) para la anotación contra dbSNP y RefSeq
- otros programas desarrollados por la Unidad de Bioinformática de la unidad

El objetivo de este análisis es leer los fragmentos de 75 bases de las lecturas generadas por el secuenciador y alinearlos frente a la secuencia de referencia del genoma humano versión 'hg19'. Posteriormente, se procesarán los datos para detectar posibles variantes genéticas en las regiones de captura definidas. Para lograrlo, el análisis computacional sigue los siguientes pasos:

### **Análisis terciario de las secuencias**

Programas utilizados:

- -IGV 2.2 (visor genómico)
- -UCSC genome browser (<http://genome.ucsc.edu/>)
- -R (programa estadístico)
- -Microsoft Excel
- -SIFT, Polyphen2, Mutation Taster, Condel
- -Human Splicing Finder (<http://www.umd.be/HSF/>)

Bases de datos consultadas:

- OMIM: online mendelian inheritance in man (<http://www.omim.org/>)
- Chip bioinformatic tools (<http://snpper.chip.org/>)

- Orphanet  
([http://www.orpha.net/consor/cgi-bin/OC\\_Exp.php?lng=ES&Expert=53347](http://www.orpha.net/consor/cgi-bin/OC_Exp.php?lng=ES&Expert=53347))
- Gene Cards (<http://www.genecards.org/>)
- HGMD: human gene mutation database  
(<http://www.hgmd.org/>)
- LSVD: locus specific variant databases  
(<http://www.hgvs.org/dblist/glsdb.htm>)
- dbSNP: single nucleotide polymorphism database  
(<http://www.ncbi.nlm.nih.gov/SNP/>)

## 6.7. Anexo VII: Características del informe final.

El informe final redactado a partir de los resultados de los análisis anteriores incluirá los siguientes apartados:

Datos del paciente, (nombre y fecha de nacimiento).

Datos del hospital, servicio y profesional peticionario.

Tipo de muestra recibida.

Fecha de recepción de la muestra.

Sospecha/s diagnóstica/s.

Análisis: breve descripción del análisis que se realiza a la muestra.

Antecedentes: breve resumen de la historia clínica del paciente, incluyendo antecedentes familiares.

Resultados:

Se comunicarán:

1\_Variantes detectadas con frecuencia  $<0.01$  (variantes raras), es decir, con probabilidad de ser patogénicas. Para cada variante se hará la descripción del gen en el que se encuentra, así como la descripción de la patología asociada a mutaciones en ese gen concreto y su forma de herencia más común.

2\_Las variantes con frecuencia  $\geq 0.01$  (polimorfismos) se proporcionarán a petición en los genes que crea conveniente.

Conclusión:

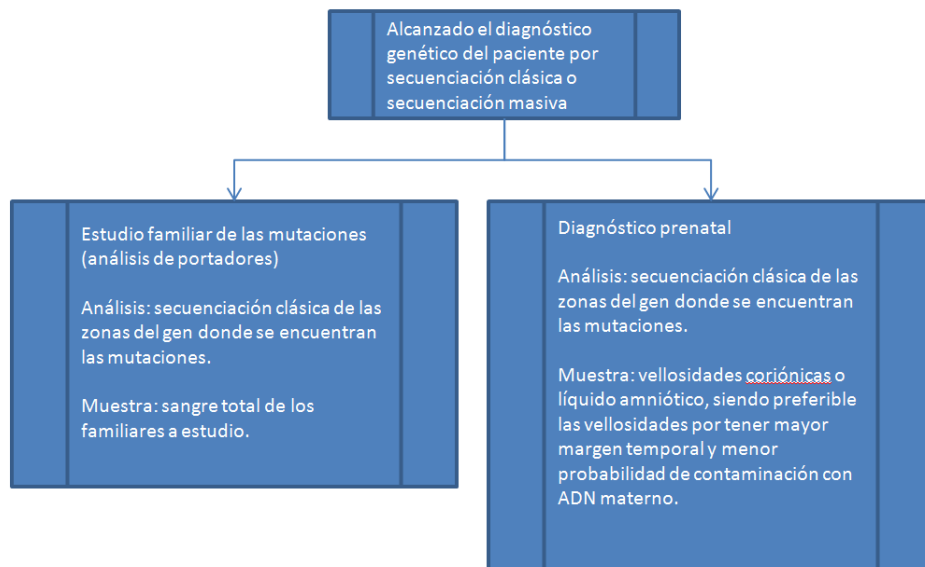
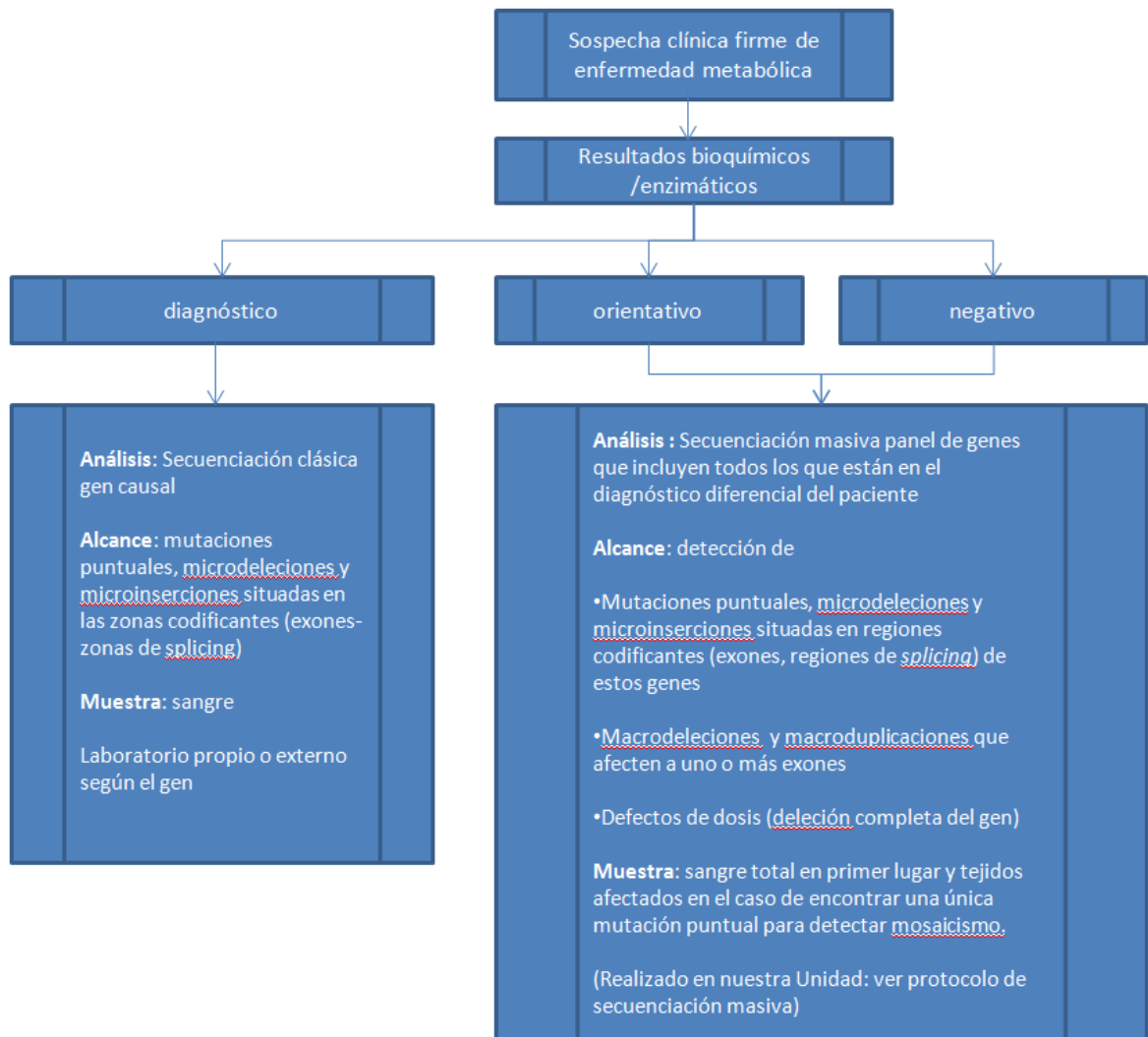
Apartado en el que se discutirán los resultados obtenidos y la probabilidad de que éstos se encuentren detrás del proceso patológico del paciente, según literatura previa, y consulta de bases de datos adecuadas.

Incluye:

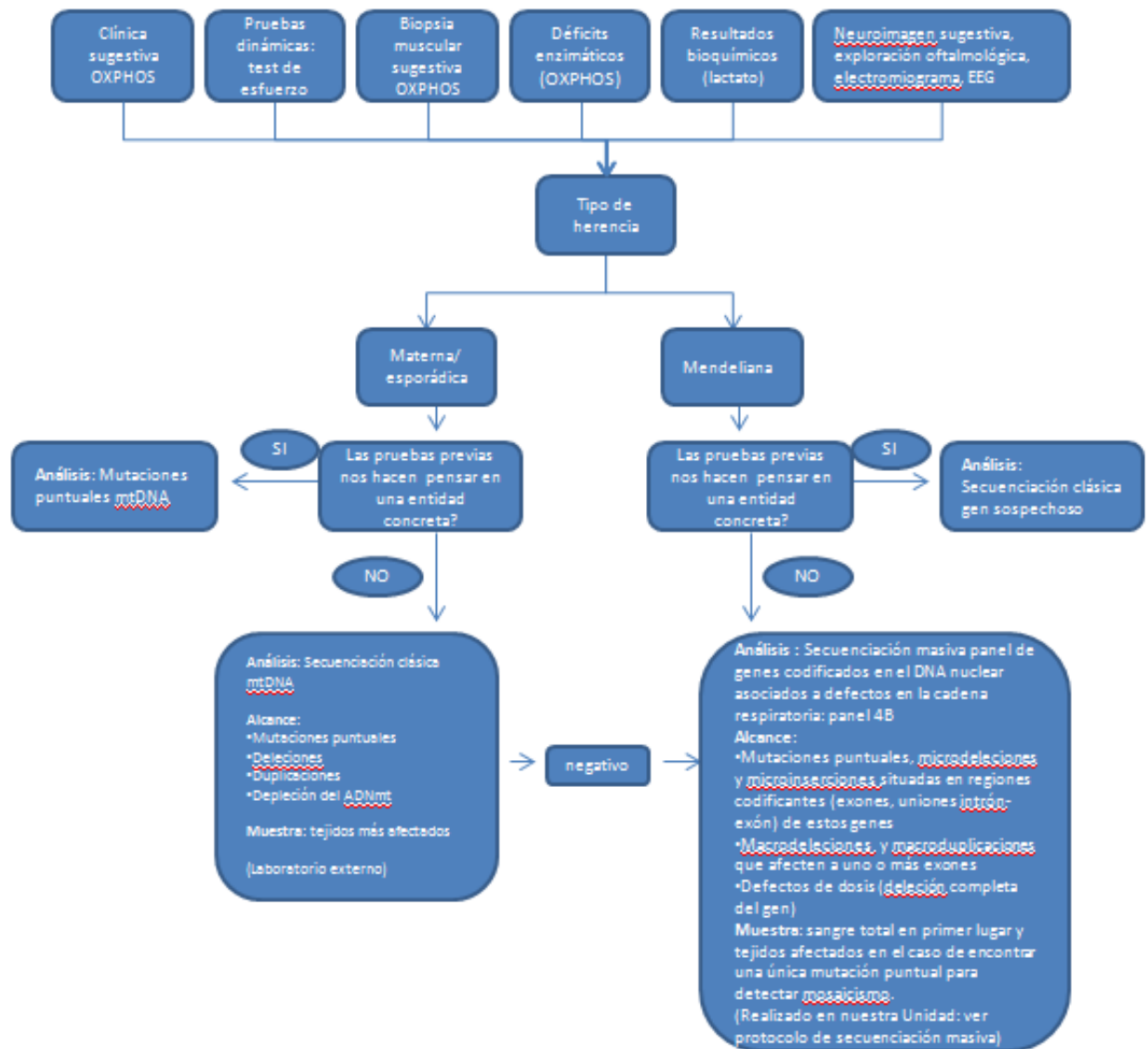
- La búsqueda de las variantes raras en la Human Gene Mutation Database (HGMD) y en Locus Specific DataBase (LSDB) para comprobar si han sido asociadas a patología previamente. En caso positivo se adjuntarán las referencias bibliográficas.
- En caso de mutaciones *missense* nuevas, la predicción *in silico* de su patogenicidad según SIFT, Polyphen2, Mutation Taster, y Condel.
- La búsqueda de la variante en la base de datos del proyecto 1000 genomas y en la base de datos dbSNP, así como la frecuencia del alelo menor si existe

- Bibliografía relevante en relación con el resultado obtenido.
- En los casos en los que no se haya podido discernir la causa molecular subyacente a la patología del paciente se incluirá un anexo a modo de tabla especificando las regiones no cubiertas (cobertura <5) y las cubiertas con cobertura entre 5 y 20X, para cada gen incluido en el panel a estudio.

## Protocolo general análisis genético ECM



## Protocolo análisis genético ante sospecha de enfermedad mitocondrial





## Análisis de portadores y diagnóstico prenatal

