

Universidad de Oviedo

TESIS DOCTORAL

Programa de Doctorado en Biomedicina y Oncología Molecular

Diagnóstico molecular del cáncer,
las sorderas y las cegueras mediante
secuenciación genómica masiva

Autora: **Marta Diñeiro Soto**

Director: **Juan Cadiñanos Bañales**





Universidad de Oviedo

TESIS DOCTORAL

Programa de Doctorado en Biomedicina y Oncología
Molecular

**Diagnóstico molecular del cáncer, las
sorderas y las cegueras mediante
secuenciación genómica masiva**

Autora: Marta Diñeiro Soto

Director: Juan Cadiñanos Bañales

Oviedo 2020



Reconocimiento-No Comercial-Sin Obra Derivada (by-nc-nd): No se permite un uso comercial de la obra original ni la generación de obras derivadas.



Usted es libre de copiar, distribuir y comunicar públicamente la obra, bajo las condiciones siguientes:



Reconocimiento – Debe reconocer los créditos de la obra de la manera especificada por el licenciador:

Marta Diñeiro Soto. (2023) *Diagnóstico molecular del cáncer, las sorderas y las cegueras mediante secuenciación genómica masiva* Universidad de Oviedo

La autoría de cualquier artículo o texto utilizado del libro deberá ser reconocida complementariamente.



No comercial – No puede utilizar esta obra para fines comerciales.



Sin obras derivadas – No se puede alterar, transformar o generar una obra derivada a partir de esta obra.

© 2023 Universidad de Oviedo

© La autora

Algunos derechos reservados. Esta obra ha sido editada bajo una licencia Reconocimiento-No comercial-Sin Obra Derivada 4.0 Internacional de Creative Commons.

Se requiere autorización expresa de los titulares de los derechos para cualquier uso no expresamente previsto en dicha licencia. La ausencia de dicha autorización puede ser constitutiva de delito y está sujeta a responsabilidad.

Consulte las condiciones de la licencia en: <https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode.es>

Servicio de Publicaciones de la Universidad de Oviedo

ISNI: 0000 0004 8513 7929

Edificio de Servicios - Campus de Humanidades

33011 Oviedo - Asturias

985 10 95 03 / 985 10 59 56

servipub@uniovi.es

www.publicaciones.uniovi.es

ISBN: 978-84-18482-74-8

FINANCIACIÓN

Los diferentes estudios que conforman la presente tesis doctoral han contado con la financiación de la **Fundación María Cristina Masaveu Peterson** a través de su mecenazgo de las actividades de I+D+i del **IMOMA**.



FUNDACIÓN
MARÍA CRISTINA
MASAVEU
PETERSON



INSTITUTO DE MEDICINA ONCOLÓGICA
Y MOLECULAR DE ASTURIAS

El proyecto correspondiente a **ONCOgenics** también fue financiado parcialmente por **FICYT / PCTI Asturias / FEDER (IE13-121)**



FUNDACION PARA EL FOMENTO EN ASTURIAS
DE LA INVESTIGACIÓN CIENTÍFICA APLICADA
Y LA TECNOLOGIA



Plan de Ciencia, Tecnología e Innovación de Asturias

Quisiera dejar constancia de mi gratitud hacia todas las personas que han participado, de una u otra manera, en la realización de este trabajo y en mi formación científica y personal.

Al IMOMA y a la Fundación María Cristina Masaveu Peterson, por haberme dado esta oportunidad y haber sustentado económicamente mi trabajo.

A Juan, por haber confiado en mí desde el primer día y brindarme la posibilidad de aprender y trabajar al mismo tiempo. Por su entrega, compromiso, esfuerzo y búsqueda de la excelencia. Siempre estaré agradecida.

A Carlos López Otín, por haber servido de inspiración desde mis comienzos universitarios y ejercer de tutor de mi trabajo.

A Gloria Velasco, por su ayuda inestimable y su buena predisposición para facilitar todos los trámites.

A Rubén, por enseñarme, asesorarme y aclararme miles de dudas. Tus brillantes ideas nunca dejarán de sorprenderme.

A Guadalupe, por su amistad, generosidad y compañerismo. Por su optimismo constante y sus ánimos en todo momento. Sé que nuestra amistad forjada en el laboratorio perdurará para siempre.

A Raquel y Adrián, por vuestra ayuda y disponibilidad incansable. Por esas conversaciones y risas en la “pecera” que alegran la jornada. Por los momentos compartidos y por los que nos quedan por vivir.

A Noelia y Rebeca, por vuestra dedicación, flexibilidad y profesionalidad. Por todas las horas de laboratorio que hemos compartido juntas, por todas las mil y una historias que hemos vivido y que ya forman parte de mi vida. Sin vuestro trabajo este camino habría sido más difícil.

A Luis Menéndez, por estar siempre accesible, por todos los mails de dudas resueltos, por sus cachondeos y por haberme hecho este camino mucho más fácil.

A Beatriz Fernández-Vega por su cercanía, su amistad y sus palabras de cariño. Por acercarme un poquito más al campo de la oftalmología.

A María costales, por su profesionalidad y disponibilidad. Por todos nuestros mensajes y mails al respecto de casos que han conformado parte de los resultados de este trabajo.

A Eva, Claudia, Lucía... y a todos los compañeros con los que me he cruzado en este viaje, todos habéis puesto vuestro granito de arena para que este proyecto saliese adelante.

A los dos pilares que siempre han estado presentes en mi vida, mis padres. Por vuestro apoyo incondicional, por enseñarme educación, humildad y valores. Por darme todo a cambio de nada. No existe tiempo en esta vida para agradecer todo lo que habéis hecho por mí.

A mis hermanos Raquel y Héctor, por estar siempre ahí, por todos los momentos de risas y discusiones. Cada instante vivido juntos ha sido importante para llegar hasta aquí.

A Pablo, por su amor, apoyo y paciencia infinitos. Por su confianza en mi y su dedicación plena en hacerme la vida más fácil. Gracias por hacerme sentir que los problemas no son de uno u otro, sino de los dos. Eres y serás mi gran apoyo en todos los aspectos de la vida.

A mis hijos, porque sin saberlo han conseguido cogerme de la mano y lograr que no me hunda. Gracias por esas sonrisas, besos y abrazos que me han dado la felicidad en los peores momentos. Esto habría sido mucho más difícil sin vosotros.

ABREVIATURAS

ACMG	Colegio americano de genética y genómica médica
ADN	Ácido desoxirribonucleico
AMP	Asociación de patología molecular
APS	Adenosina 5'-fosfosulfato
ARN	Ácido ribonucleico
ASCO	Sociedad americana de oncología clínica
ATP	Adenosina trifosfato
CAP	Colegio de patólogos americanos
cDNA	ADN complementario
CNV	Variante de número de copias
CODEPEH	Comisión para la detección precoz de la hipoacusia
Cx	Conexina
dATP	Desoxiadenosina trifosfato
dCTP	Desoxicitidina trifosfato
ddATP	Didesoxiadenosina trifosfato
ddCTP	Didesoxicitidina trifosfato
ddGTP	Didesoxiguanidina trifosfato
ddNTP	Didesoxinucleósido trifosfato
ddTTP	Didesoxitimidina trifosfato
dGTP	Desoxiguanidina trifosfato
dNMP	Desoxinucleósido monofosfato
dNTP	Desoxinucleósido trifosfato
DOE	Departamento de Energía
DRH	Distrofia retiniana hereditaria
dTTP	Desoxitimidina trifosfato
EMA	Agencia europea del medicamento
EMQN	European molecular genetics quality network

EPR	Epitelio pigmentario de la retina
ER	Receptor de estrógenos
FDA	U.S food & drug administration
HGMD	Human gene mutation database
HNPCC	Cáncer de colon no polipósico hereditario
HNS	Hipoacusia neurosensorial
ICGC	Consortio internacional del genoma del cáncer
Indel	Inserciones/deleciones.
LHON	Neuropatía óptica hereditaria de Leber
MLCRD	Síndrome de microcefalia, linfedema y displasia coriorretiniana
MLPA	Amplificación de sondas multiplex dependiente de ligación
MMR	Missmatch repair
NCBI	Centro nacional de información biotecnológica
NCCN	National comprehensive cancer network
NHGRI	National human genome research Institute
NGS	Secuenciación de nueva generación
NIH	Institutos nacionales de la salud
OMIM	Online mendelian inheritance in man
PCR	Reacción en cadena de la polimerasa
PGH	Proyecto Genoma Humano
PPi	Pirofosfato
RP	Retinosis pigmentaria
SMRT	Secuenciación a tiempo real de una sola molécula
SNP	Polimorfismo de un solo nucleótido
SNV	Variante de un solo nucleótido
TCGA	Atlas del genoma del cáncer
VICC	Consortio de interpretación de variantes para el cáncer
WES	Secuenciación completa del exoma
WGS	Secuenciación completa del genoma

ÍNDICE

INTRODUCCIÓN	1
1 Genoma.....	5
2 Secuenciación del ADN (primera generación).....	7
3 Proyecto Genoma Humano	8
4 NGS: Secuenciación Genómica de Nueva Generación.....	11
4.1 Preparación de la librería o genoteca	11
4.2 Enriquecimiento de la librería	13
4.3 Secuenciación.....	13
4.4 Control de calidad y análisis de datos.....	18
5 Medicina de precisión	19
6 Diagnóstico genómico mediante NGS.....	20
6.1 Secuenciación genómica completa (WGS).....	21
6.2 Secuenciación del exoma (WES)	22
6.3 Paneles de genes	23
7 Paneles para el diagnóstico genómico de enfermedades de base genética	24
7.1 Cáncer	25
7.2 Sorderas.....	31
7.3 Cegueras	35
8 Interpretación clínica de los resultados de diagnóstico genético	37
8.1 Bases de datos de variantes genéticas clínicamente relevantes	38
8.2 Criterios ACMG/AMP de interpretación clínica de variantes genómicas	42
OBJETIVOS	51
RESULTADOS	55
Artículo I: A novel molecular diagnostics platform for somatic and germline precision oncology	57
Artículo II: Comprehensive genomic diagnosis of nonsyndromic and syndromic hereditary hearing loss in Spanish patients.....	83
Artículo III: Sequencing results from multiple individuals of different ethnicities strongly question the existence of the <i>KCNE1B</i> pseudogene.....	103
Artículo IV: Comprehensive genomic diagnosis of inherited retinal and optical nerve disorders reveals hidden syndromes and personalized therapeutic options.....	107
DISCUSIÓN	125

CONCLUSIONES.....	151
BIBLIOGRAFÍA	155
ANEXOS.....	173

INTRODUCCIÓN

Se denomina enfermedad de base genética a cualquier trastorno causado por una alteración en el genoma. Este tipo de enfermedades ha ido ganando importancia paulatinamente en el ámbito clínico, no sólo por su elevada incidencia (p. ej., alrededor de un 1 por ciento de los recién nacidos tienen algún tipo de anomalía genética¹), sino también por la gran diversidad de manifestaciones y complicaciones asociadas y por su carácter progresivo, que habitualmente condicionan la calidad de vida del individuo. Las mejoras en la comprensión y en el diagnóstico de muchas de estas enfermedades se deben, en gran medida, a los avances en el conocimiento de sus bases moleculares. Se estima que existen miles de enfermedades de base genética², de las que en esta Tesis Doctoral abordaremos 3 amplios grupos: el cáncer, la hipoacusia neurosensorial y las cegueras que cursan con alteraciones de la retina, el vítreo, la coroides y/o el nervio óptico.

Todos los tipos de cáncer son resultado de una transformación causada inicialmente por alteraciones en la secuencia de genes clave para el control del crecimiento y división celular. La mayoría de estas alteraciones son adquiridas, provocadas por la interacción entre los factores genéticos de cada paciente y diferentes tipos de agentes externos (tabaco, radiaciones ultravioletas, amianto, virus, etc.), pero en determinados pacientes existen mutaciones oncogénicas heredadas, es decir, transmitidas de generación en generación.

Las estadísticas actuales indican que uno de cada dos hombres y una de cada tres mujeres desarrollarán cáncer a lo largo de su vida y que el cáncer es la segunda causa de muerte en Europa³. En cuanto a la hipoacusia y la retinosis pigmentaria, afectan a 5-10 de cada 1000 y a una de cada 4000 personas en el mundo, respectivamente^{4,5}, y las mutaciones heredables en ellas son mucho más prevalentes que en el cáncer⁶⁻⁸. Por tanto, el estudio de las bases moleculares de estas tres patologías es una prioridad para el desarrollo de nuevos métodos de prevención, diagnóstico y tratamiento específicos.

Las estimaciones recientes consideran que entre el 5 % y el 10 % de los casos de cáncer⁹, alrededor del 50 % de las hipoacusias neurosensoriales¹⁰ y más del 90 %

de las retinosis pigmentarias¹¹ son enfermedades mendelianas, también llamadas monogénicas, es decir, patologías cuya causa última es la alteración heredable de un único gen en los individuos afectados. El descubrimiento de la base genética de una enfermedad mendeliana establece un vínculo causal entre el genotipo y el fenotipo, lo cual permite obtener un diagnóstico definitivo para el paciente. Por tanto, la optimización de las técnicas analíticas y de secuenciación, junto con una adecuada interpretación clínica de las variantes genéticas, representan avances necesarios para el diagnóstico genético¹².

En 2001 se publicó el primer borrador del genoma humano, obtenido mediante secuenciación Sanger¹³. Aunque, por sí mismo, este fue un hito fundamental para el avance del conocimiento humano, sus implicaciones se multiplicaron con la aparición de los métodos de secuenciación de nueva generación (NGS) en 2005¹⁴. Esta metodología ha puesto a nuestro alcance la secuenciación del genoma completo de cualquier individuo o tumor, lo que ha aumentado notablemente la cantidad de datos genómicos producidos y está facilitando una explosión en el conocimiento biomédico. Actualmente, la tecnología de secuenciación del genoma sigue desempeñando una función fundamental en el descubrimiento de nuevos paradigmas biológicos con enormes implicaciones para el manejo clínico de las enfermedades.

1 Genoma

El genoma es el material hereditario de cualquier organismo, conformado por polímeros cuyos elementos informativos son cuatro de entre cinco posibles bases nitrogenadas. Estas son la adenina, la guanina y la citosina, presentes en todos los genomas, a las que se añade el uracilo en los genomas de algunos virus (virus de ácido ribonucleico, o ARN) o la timina en el resto de organismos. Cada base nitrogenada, combinada con un azúcar (ribosa en el ARN, desoxirribosa en el ácido desoxirribonucleico, o ADN) y con grupos fosfato, se encadena con otras para formar polímeros (cadenas de ADN o de ARN) que se pueden acoplar dos a dos formando pares de cadenas complementarias, en los que la adenina siempre se empareja con la guanina, mientras que la citosina lo hace con el uracilo o con la timina, de modo que cada cadena contiene la información necesaria para sintetizar la cadena complementaria. Estas cadenas de ácidos nucleicos llevan codificadas las instrucciones genéticas necesarias para desarrollar y dirigir el funcionamiento de todo el organismo.

En el caso del ser humano, nuestro genoma está formado por una secuencia de aproximadamente 3.200.000.000 pares de bases de ADN que, además, está presente en cada célula por duplicado (una copia heredada de nuestra madre y otra de nuestro padre). Estas dos copias de nuestro genoma forman los 23 pares de cromosomas¹⁵. Cada uno de nosotros es el resultado de la interacción del ambiente con las instrucciones contenidas en nuestro ADN. Así, el genoma ha evolucionado incorporando la información necesaria para producir respuestas adecuadas a diferentes cambios en el ambiente. No obstante, los fenómenos responsables de esta evolución, que son los que generan cambios o mutaciones en el genoma, son también capaces de provocar alteraciones que dan lugar a efectos adversos y patológicos.

Las unidades funcionales fundamentales del genoma son los genes, que contienen la información necesaria para producir, de manera específica, los principales ejecutores de las funciones vitales: las proteínas. Un gen eucariota está habitualmente formado por exones, la mayoría de los cuales contienen información codificante de proteínas, y por intrones (regiones de ADN que separan los exones). Además, entre

cada dos genes hay regiones de ADN intergénico (que no forma ni intrones ni exones). Aunque sería intuitivo asumir que la mayor parte de nuestro ADN está constituido por secuencias que codifican proteínas, lo cierto es que no es así, ya que sólo aproximadamente el 1,5 % se encarga de esta función.

Se estima que hay entre 20.000 y 25.000 genes que codifican proteínas humanas, pero ese número ha ido cambiando en el tiempo y continúa actualizándose como consecuencia de los avances científicos y tecnológicos. Estos genes, con sus intrones, sus exones y sus regiones reguladoras, ocupan el 30 % del genoma. El 70 % restante corresponde a las regiones intergénicas o extragénicas. Durante muchos años, tanto ese 70 % de ADN extragénico como gran parte del genoma génico (exceptuando el 1,5 % codificante) se consideró “ADN basura”, sin una función determinada. Sin embargo, a día de hoy sabemos que, al igual que el ADN codificante, el ADN no codificante es indispensable para la función celular normal, particularmente para el control de la actividad genética, por contener secuencias que actúan como elementos reguladores (promotores, potenciadores y silenciadores de la expresión de los genes o incluso elementos estructurales como las secuencias repetitivas), que determinan cuándo y dónde se “activan” y “desactivan” los genes. Por esta razón, las alteraciones del genoma que afectan al ADN no codificante pueden tener consecuencias funcionales importantes que desencadenan enfermedades.

Las enormes implicaciones asociadas al conocimiento del genoma para entender los estados de salud y enfermedad del ser humano son la principal motivación para su análisis y estudio en profundidad. El primer paso decisivo en ese análisis fue el descubrimiento de la estructura de doble hélice de la molécula del ADN en 1953 por Francis Crick y James Watson, apoyados en los trabajos de cristalografía de rayos X de Rosalind Franklin¹⁶. Este hito marcaría un antes y un después en la historia de la biología. Desde ese momento, conocer la secuencia de los nucleótidos dentro de esa estructura de doble hélice en cada individuo ha sido un objetivo fundamental y necesario para descifrar las causas de las enfermedades genéticas¹⁷.

2 Secuenciación del ADN (primera generación)

La secuenciación de ADN se define como el proceso por el cual se determina la secuencia de nucleótidos (adenina, timina, guanina y citosina) de una molécula de ADN. La necesidad de desarrollar un método de secuenciación del ADN se hizo especialmente evidente tras la aceptación del llamado dogma central de la biología molecular, enunciado por Francis Crick y publicado en la revista *Nature* en 1970, según el cual la secuencia de nucleótidos de una molécula de ADN condiciona la secuencia de aminoácidos de las proteínas que ejecutan la actividad celular¹⁸. Esto no ocurre directamente, sino con un intermediario, el ARN, que se transcribe a partir del ADN, respetando su secuencia (salvo porque el uracilo sustituye a la timina), y que es leído por los ribosomas, encargados de transformar cada secuencia de nucleótidos codificantes en una secuencia de aminoácidos¹⁸.

Fue en la misma década de los 70 cuando Frederick Sanger, con el método enzimático de terminación de cadena (método de los didesoxinucleótidos)^{19,20}, y Walter Gilbert, mediante la fragmentación química²¹, desarrollaron los primeros procedimientos robustos para obtener secuencias de ADN. Originalmente, ambos eran métodos muy manuales, tediosos, que implicaban el uso de compuestos radioactivos y cuya capacidad de lectura era de unas decenas a unos pocos cientos de bases. El avance en las técnicas bioquímicas, ópticas y robóticas durante las décadas de los 80 y de los 90 permitió que se desarrollaran estrategias de secuenciación menos nocivas y más automatizables, basadas en el método de Sanger, gracias a la sustitución de los compuestos radiactivos por didesoxinucleótidos marcados con fluorescencia y a una mejor detección a través de la electroforesis capilar. Esto permitió mejorar y aumentar el rendimiento del proceso de secuenciación, hasta el punto en que los nuevos avances hicieron posible el desarrollo de los primeros secuenciadores automáticos de tipo CEQ (Beckman Coulter)²² o ABI Prism (Applied Biosystems)²³. Estos últimos fueron posteriormente los responsables de generar la mayor cantidad de las secuencias empleadas para completar el Proyecto Genoma Humano (PGH)¹⁷ [**Figura 1**].

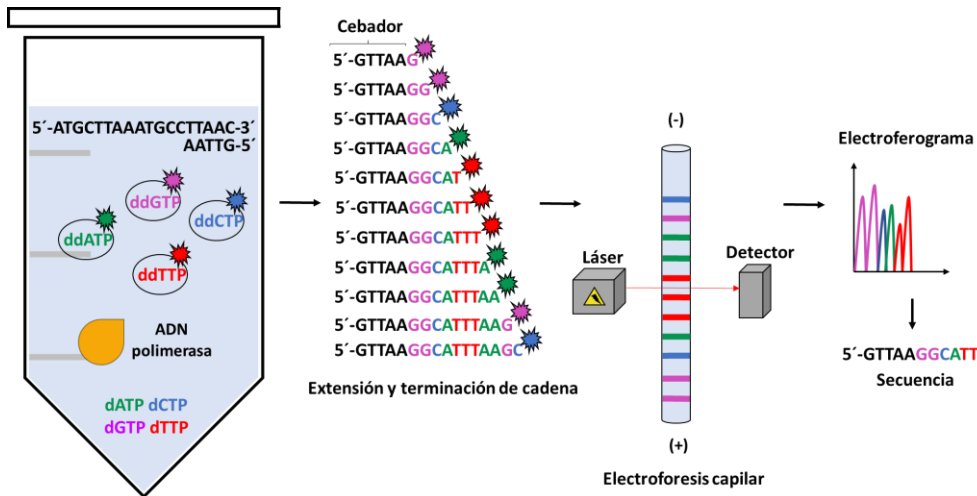


Figura 1. Secuenciación Sanger automatizada fluorescente.

Un oligonucleótido sintetizado químicamente, que forma un dúplex con el ADN monocatenario a secuenciar, es el sustrato sobre cuyo extremo 3' la ADN polimerasa incorpora los nucleótidos complementarios a la cadena a secuenciar, a partir de una mezcla de desoxinucleótidos y dideoxinucleótidos trifosforilados (dNTP y ddNTPs, respectivamente). Esta mezcla está compuesta mayoritariamente por cantidades equimoleculares entre sí de dATP, dCTP, dGTP y dTTP, así como, en mucha menor proporción, por cantidades, también equimoleculares entre sí, de ddATP, ddCTP, ddGTP y ddTTP. Estos últimos están marcados con un fluoróforo diferente en función de la base nitrogenada que los compone y, cuando se incorporan a la cadena, al tener un átomo de hidrógeno en lugar de un grupo OH unido al carbono 3' de su desoxirribosa, detienen la polimerización. De este modo, se genera una mezcla de cadenas monocatenarias fluorescentes de diversos tamaños y complementarias al ADN a secuenciar. Todas las moléculas del mismo tamaño terminan en el mismo nucleótido y, por lo tanto, producen la fluorescencia de la misma longitud de onda al ser excitadas. La separación de estas moléculas monocatenarias mediante electroforesis capilar y la detección de su fluorescencia dibuja un electroferograma en el que la sucesión de colores permite determinar la secuencia de nucleótidos complementaria al fragmento a secuenciar.

3 Proyecto Genoma Humano

Los orígenes del PGH se sitúan en una reunión organizada a mediados de la década de los 80 por Robert Sinsheimer, profesor en la Universidad de California. Desde ese momento, se desencadenaron múltiples iniciativas orientadas a descifrar la información de nuestro material genético²⁴. En el año 1986, con la intención de obtener datos para la protección del material genético de los efectos mutagénicos de la

radiación, el Departamento de Energía (*Department of Energy*, DOE) de los Estados Unidos se involucró también en este propósito. Dos años más tarde, en 1988, el Congreso de los Estados Unidos financió tanto a los Institutos Nacionales de la Salud (*National Institutes of Health*, NIH) como al DOE para que se embarcasen en una exploración más a fondo. Este acontecimiento promovió la firma de un memorando de entendimiento para "coordinar investigaciones y actividades técnicas relacionadas con el genoma humano", el primer paso para abordar lo que sería el futuro PGH²⁵.

En 1990 se presentó el Proyecto con una planificación inicial de 15 años, declarando el 1 de octubre como fecha de inicio oficial²⁶. En paralelo surgieron diferentes plataformas y avances asociados, como la publicación de BLAST, el primer algoritmo bioinformático capaz de alinear y comparar secuencias de ADN, cuya funcionalidad lo convertiría en una herramienta crítica para la ejecución del PGH²⁷. El propósito inicial era dotar al mundo de herramientas innovadoras y trascendentales para el tratamiento y prevención de enfermedades y, por ello, se marcaron una serie de objetivos enumerados en la **Tabla 1**. Con el mismo fin, en el año 1998, Craig Venter fundó la empresa *Celera Genomics*, a través de la cual puso en práctica su propuesta de secuenciar y ensamblar el genoma humano en un plazo de 3 años.

Tabla 1. Objetivos iniciales del Proyecto Genoma Humano

1/	Identificar la totalidad de los genes humanos que conformaban el ADN. (<i>En ese momento se pensaba que eran entre 50.000 y 100.000</i>)
2/	Determinar la secuencia de los 3.000 millones de nucleótidos que conforman el ADN.
3/	Crear bases de datos de acceso público con el fin de asegurar la disponibilidad de la información genómica.
4/	Desarrollar de modo rápido y eficiente tecnologías de secuenciación.
5/	Desarrollar herramientas para análisis de datos genómicos.
6/	Dirimir las cuestiones éticas, legales y sociales derivadas del Proyecto.

Previamente al abordaje del genoma humano, se llevó a cabo una serie de proyectos piloto de secuenciación Sanger en 4 organismos modelo (*Mycoplasma*

capricolum, *Escherichia coli*, *Caenorhabditis elegans* y *Saccharomyces cerevisiae*) con el fin de asentar conocimientos y desarrollar herramientas que facilitasen el análisis genómico. Así, tras el esfuerzo coordinado de cientos de investigadores en decenas de países y casi 3.000 millones de dólares¹³, el 15 de febrero de 2001 el consorcio del PGH publicó en la revista *Nature* el primer borrador del genoma humano de referencia, que abarcaba el 90 % de su secuencia. Con un día de diferencia, el 16 de febrero del mismo mes, hizo lo propio la iniciativa privada *Celera Genomics* que, tras haber secuenciado previamente por primera vez el genoma de la mosca *Drosophila melanogaster* e invertir 300 millones de dólares, publicaba el primer esbozo de la secuencia de 5 genomas de diferentes etnias humanas en la revista *Science*^{28,29}.

Al hito de contar con un primer borrador completo del genoma humano le sucedió la fase de finalización del PGH, en la que se consiguió resolver problemas que afectaban a secuencias de ADN en regiones ambiguas y dilucidar zonas que no habían sido cubiertas en la fase previa. Los resultados finales obtenidos en 2003 consiguieron determinar la secuencia correspondiente al 99 % del genoma humano con una tasa de error estimada de 1 alteración por cada 100.000 bases secuenciadas. Los planteamientos iniciales habían pronosticado un número de genes codificantes de proteínas entre 50.000 y 100.000 pero, en base a los resultados publicados en la revista *Nature* en 2004, se estimó finalmente un número de entre 20.000 y 25.000³⁰. Una de las principales conclusiones del PGH fue que la secuencia de nucleótidos tiene una similitud del 99,9 % entre dos individuos elegidos al azar, lo que significa que entre ellos tienen más de 3 millones de diferencias³¹. No obstante, como hemos mencionado anteriormente, el cambio de un único nucleótido en un gen puede ser suficiente para causar una determinada enfermedad. Por tanto, se abrió un nuevo reto: distinguir las variantes genéticas que constituyen la diversidad genética humana normal de aquellas que tienen consecuencias patológicas.

A pesar de su enorme trascendencia, la finalización del PGH no logró, de manera inmediata, el propósito inicial de descubrimiento de nuevas herramientas para combatir la enfermedad³². Sin embargo, inspiró múltiples iniciativas de adquisición de datos a gran escala, como el Proyecto Internacional HapMap en 2002, el Atlas del

genoma del Cáncer (TCGA) en 2005³³, el Consorcio Internacional del Genoma del Cáncer (ICGC) en 2008³⁴ o el Proyecto Proteoma Humano en 2011³⁵, entre otros. Asimismo, este proyecto no solo marcó el comienzo de una nueva era en la medicina, sino que también condujo a avances significativos en los tipos de tecnologías utilizados para secuenciar el ADN, fundamentales para el éxito de las iniciativas antes mencionadas.

4 NGS: Secuenciación Genómica de Nueva Generación

La disponibilidad de la información obtenida por el PGH permitió el desarrollo de la secuenciación genómica de segunda generación entre los años 2004 y 2005. Esta tecnología es capaz de determinar la secuencia de los 3.200 millones de nucleótidos de nuestro genoma a partir de millones de lecturas de longitud variable, que se ensamblan unas con otras mediante métodos bioinformáticos hasta reconstruir una secuencia genómica individual completa. Sin embargo, para esto es necesaria su comparación con un genoma de referencia, motivo por el cual no podría haberse desarrollado con éxito sin contar con los resultados del PGH. También llamada secuenciación paralela masiva por su capacidad de procesar gran cantidad de fragmentos de ADN de forma simultánea, la NGS tiene un rendimiento varios órdenes de magnitud superior a la metodología Sanger y, en consecuencia, un menor coste por cada base secuenciada.

Existen diversas técnicas de secuenciación consideradas de segunda generación, pero las fases generales del proceso son compartidas por todas ellas: 1) preparación de la “librería” de ADN (también llamada genoteca o biblioteca de ADN), 2) (opcional) enriquecimiento, 3) secuenciación genética y 4) control de calidad y análisis de datos.

4.1 Preparación de la librería o genoteca

Consiste en generar segmentos del ADN a secuenciar flanqueados por secuencias de ADN externas, llamadas adaptadores, que son requeridas para el funcionamiento de las diferentes tecnologías de secuenciación de nueva generación. Estos adaptadores se pueden añadir al menos de tres maneras distintas: mediante PCR

(añadiéndolos en el 5' de los oligonucleótidos cebadores específicos del genoma a secuenciar)³⁶, ligándolos a los extremos de fragmentos del ADN objeto de análisis³⁷ o intercalándolos en el mismo mediante transposición³⁸.

Las técnicas de PCR son las más sencillas, permiten en un solo paso añadir los adaptadores y seleccionar la región a secuenciar, pero tiene la limitación de que si la secuencia diana tiene alguna diferencia en la región de unión de los cebadores que impida su hibridación, existe la posibilidad de que no se detecten alteraciones presentes en el alelo afectado por dicha diferencia. Las técnicas de ligación requieren la fragmentación de la muestra de ADN, que puede ser mecánica (siendo la sonicación o la nebulización las metodologías más usadas para ello) o enzimática (mediante enzimas de restricción), para posteriormente poder agregar los adaptadores en ambos extremos de los fragmentos de ADN. La tercera alternativa, las técnicas de transposición o tagmentación, se basa en una fragmentación enzimática con transposasas en sitios aleatorios de la molécula, al mismo tiempo que se añaden los adaptadores. La longitud promedio del fragmento que se va a secuenciar depende directamente del ratio ADN /transposasa^{38,39}. Una vez añadidos los adaptadores, en principio se podría ya secuenciar la librería generada. En el caso de las técnicas que emplean fragmentación o tagmentación para añadir los adaptadores, esto puede permitir secuenciar directamente el ADN objeto de estudio (técnicas sin amplificación)³⁸ aunque, habitualmente, antes de secuenciar las librerías se realiza una amplificación por PCR usando cebadores complementarios a dichos adaptadores para obtener la cantidad de ADN necesaria para el proceso de secuenciación⁴⁰.

El tamaño del fragmento de ADN situado entre los adaptadores se conoce como “tamaño de inserto”, y puede variar dependiendo del objetivo del análisis. Por ejemplo, se prefieren tamaños de inserto más grandes para detectar reordenamientos genómicos; en cambio, los de menor tamaño están indicados para la detección de variantes de un solo nucleótido (SNVs). Una vez preparada la librería, esta puede ser secuenciada por igual en toda su extensión (por ejemplo, si queremos secuenciar un genoma completo) o se puede enriquecer previamente para obtener más cantidad de secuencia de determinadas regiones genómicas de interés.

4.2 Enriquecimiento de la librería

Este paso es inherente a la propia preparación de la librería cuando los adaptadores se añaden a ella directamente mediante PCR ya que, en ese caso, la secuencia de los propios cebadores se diseña para ser específica de las regiones diana⁴¹. Sin embargo, supone un procesamiento adicional cuando los adaptadores se han añadido mediante ligación o transposición y el objetivo no es secuenciar un genoma completo, sino parte de él. Los métodos más utilizados para ello se basan en la hibridación o captura con sondas (biotiniladas en solución o inmovilizadas en microarrays)⁴¹. La elección de la estrategia a seguir (captura vs PCR) a menudo está condicionada por el objetivo del análisis: en general, se prefiere la captura de secuencia para regiones genómicas grandes y PCR para regiones más pequeñas, ya que el rendimiento de la captura es más bajo cuanto menor es la región diana, mientras que su coste no se ve reducido⁴².

4.3 Secuenciación

Los procedimientos más usados en la actualidad son la secuenciación con nucleótidos terminadores reversibles fluorescentes en los equipos de tecnología Illumina (NovaSeq, HiSeq, NextSeq, MiSeq e Iseq, entre otros) y la detección de protones mediante semiconductores en los secuenciadores Ion Torrent (IonPGM, IonProton e IonS5).

En ambos casos, el primer paso consiste en la amplificación clonal de cada fragmento de la librería. Mediante sucesivas reacciones de síntesis a partir de una única molécula de ADN, se genera un conjunto de fragmentos idénticos (clones o *clusters*) que posteriormente son secuenciados. El objetivo de esta amplificación es generar una señal lo suficientemente intensa como para permitir la detección.

En el caso de Ion Torrent, la amplificación clonal ocurre mediante PCR en emulsión, es decir, cada molécula de ADN de la librería hibrida a través de sus adaptadores en microperlas sólidas de acrilamida que llevan unidas los cebadores y cada perla se incorpora en una emulsión cuidadosamente controlada, en la que cada

burbuja constituye un microrreactor con todo lo necesario para llevar a cabo la PCR [Figura 2A]. En cambio, la tecnología Illumina emplea una PCR en fase sólida, sobre una *flow-cell* recubierta de oligonucleótidos complementarios a los adaptadores, donde se lleva a cabo una "amplificación en puente", llamada así porque las cadenas de ADN replicadas tienen que arquearse para poder cebar la siguiente ronda de polimerización con los oligonucleótidos cercanos unidos a la superficie [Figura 2B]. Las moléculas de ADN se distribuyen de manera aleatoria en el soporte sólido. Por ello, con el fin de evitar tanto su saturación como su excesiva dilución, la cuantificación final de la librería debe ser precisa⁴³.

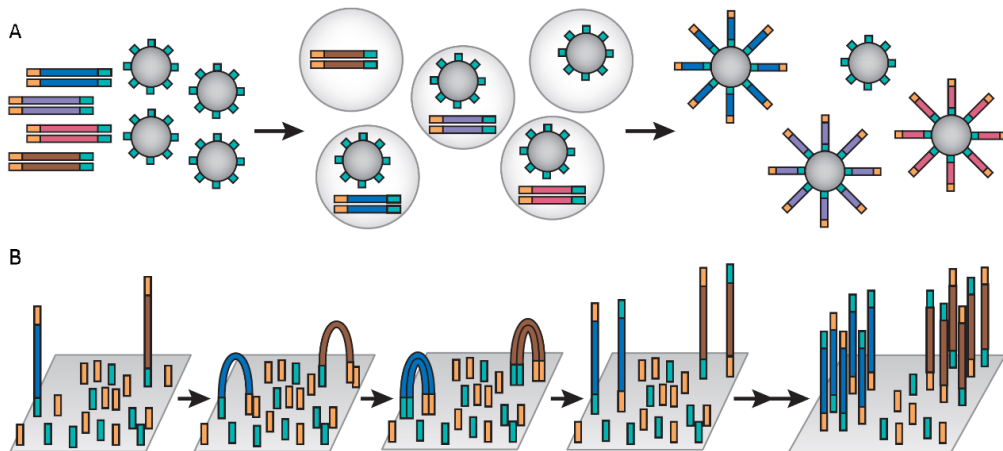


Figura 2. Amplificación clonal (adaptado de Shendure y col, 2008)⁴⁴.

(A) PCR en emulsión, Ion Torrent: PCR en emulsión propia del sistema Ion Torrent. Los fragmentos de la librería y las microperlas, recubiertas de oligonucleótidos complementarios a los adaptadores, se mezclan y diluyen en una emulsión, de modo que cada microgotita contiene idealmente un único fragmento y una única microperla. Mediante un proceso de amplificación, cada microperla queda recubierta de múltiples copias del mismo fragmento de ADN; **(B) Amplificación "en puente", Illumina:** Amplificación "en puente" propia del sistema Illumina. Los fragmentos de la librería, adecuadamente diluidos, se unen a oligonucleótidos, complementarios a sus adaptadores, que están repartidos en un soporte sólido. Al existir oligonucleótidos complementarios a ambos adaptadores del fragmento, se produce una amplificación en puente que va formando clones de moléculas agrupadas físicamente y procedentes del mismo fragmento de ADN original.

Una vez completada la amplificación clonal de la librería, el siguiente paso en un proceso de secuenciación de segunda generación es la lectura de los clones o *clusters*. La metodología usada por Ion Torrent tuvo sus orígenes en pleno desarrollo del PGH, en el año 1998, con la denominada pirosecuenciación. Esta técnica no utiliza fluorescencia ni desoxinucleótidos terminadores, sino que emplea dNTPs naturales y detecta la liberación del grupo pirofosfato, que ocurre tras la incorporación de estos nucleótidos al ADN que está siendo sintetizado por acción de la ADN polimerasa. Estos grupos pirofosfato se convierten en una señal luminosa mediante una reacción enzimática catalizada por una luciferasa [**Figura 3A**]⁴⁵.

En el año 2005, combinando este mismo principio con la PCR en emulsión, se desarrollaron las máquinas de secuenciación 454 que, capaces de producir lecturas de una longitud de alrededor de 400-500 pares de bases, permitieron la paralelización masiva de las reacciones de secuenciación, aumentando en gran medida la cantidad de ADN que se podía secuenciar en un mismo experimento. Este cambio de paradigma hizo que la secuenciación 454 pasase a considerarse, ese mismo año, la primera tecnología de NGS⁴⁶. En la misma línea, surgieron en 2010 los secuenciadores Ion Torrent, en los que la detección de luminiscencia generada a partir de grupos pirofosfato se sustituyó por la detección de protones generados también durante la polimerización del ADN, que provocan un cambio de pH en la solución. Este cambio de pH es registrado como un cambio de voltaje por un sensor de iones (semiconductor). La capacidad de esta técnica para realizar un registro directo acelera el proceso de detección en gran medida y lo hace más fiable al eliminar una reacción enzimática intermedia (la que convierte el pirofosfato en luz) [**Figura 3B**]⁴⁷.

En ambas tecnologías (454 o Ion Torrent), para conocer a qué base nucleotídica corresponde cada señal, es necesario ir incorporando por separado cada dNTP en cada ciclo de secuenciación (p.ej.: primero se añade dATP y se hace la lectura de señal, luego dCTP + lectura, después dGTP + lectura, luego dTTP + lectura y, tras todo ello, empieza el siguiente ciclo de nuevo con dATP + lectura y así sucesivamente). Por tanto, si en una posición la hebra que está sirviendo como molde de la reacción de síntesis (correspondiente a la librería que se está leyendo) tiene 4 adeninas seguidas, en ese

ciclo se incorporarán 4 adeninas. Esto impone una limitación tanto a la tecnología de 454 como a la de Ion Torrent: los polímeros de una misma base producen señales proporcionales al número de bases del polímero y, en consecuencia, resulta complicado distinguir la señal producida por seis adeninas seguidas de la que producen siete, por ejemplo⁴⁸.

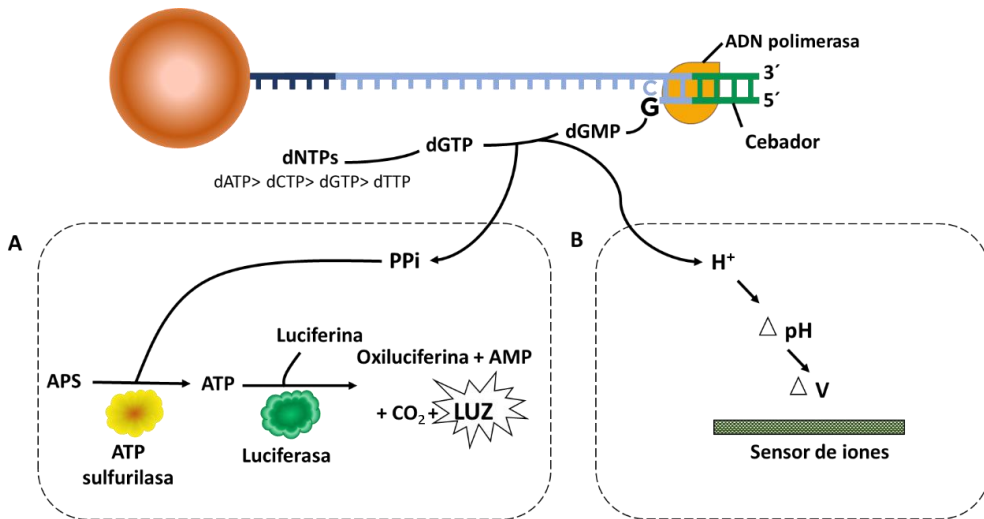


Figura 3. Pirosecuenciación y secuenciación Ion Torrent.

(A) Pirosecuenciación. El dNTP que se incorpora a la cadena en síntesis debe transformarse en dNMP, para lo cual libera un grupo pirofosfato (PPi). Posteriormente, una ATP sulfurilasa en presencia de adenosina 5'-fosfosulfato (APS) transforma el PPi en ATP, que se acoplará a la luciferina formando un complejo utilizado por una luciferasa. La intensidad de la señal luminosa producida es proporcional al número de nucleótidos incorporados. **(B) Secuenciación Ion Torrent.** La incorporación del dNMP complementario a la cadena molde implica la generación de un enlace covalente y, como consecuencia, la liberación de un protón que modifica el pH del medio. Esa variación de pH es detectada por un sensor de iones. El número de protones liberados es proporcional al número de nucleótidos incorporados.

Después del éxito del 454, surgieron varias técnicas de secuenciación paralela masiva. En retrospectiva, la más importante de ellas fue el método de secuenciación Solexa aparecido en 2006, que se basaba en la PCR en fase sólida con la "amplificación en puente" mencionada previamente [Figura 2B] combinada con una lectura basada en nucleótidos terminadores reversibles fluorescentes [Figura 4].

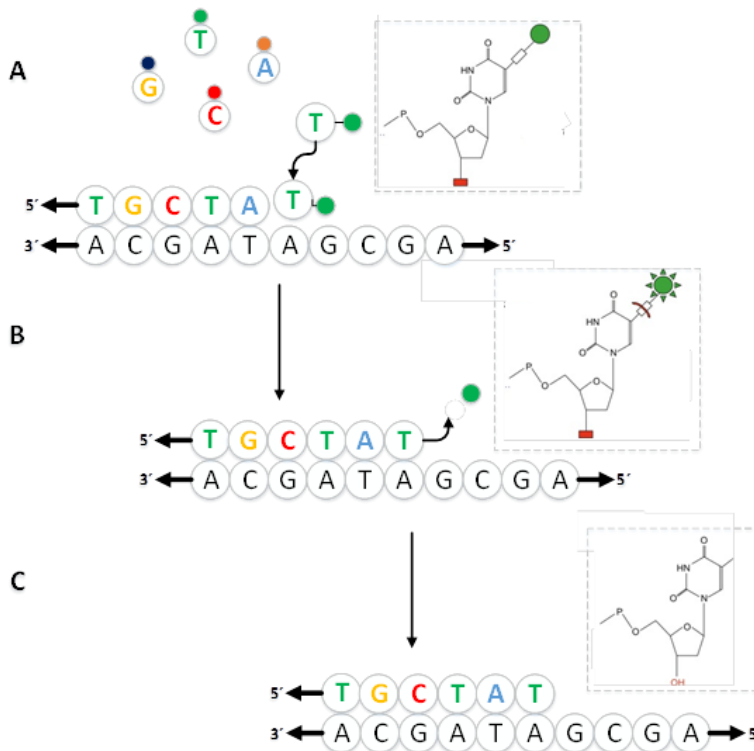


Figura 4. Secuenciación y detección Illumina.

(A) Por cada ciclo de síntesis, se añaden a la flow cell los 4 ddNTPs terminadores reversibles marcados con fluorescencia, de modo que a cada cadena en síntesis solo se puede unir uno de ellos (el complementario a la siguiente base del fragmento a secuenciar). **(B)** Después de lavar los ddNTP no incorporados, se excita la celda y una cámara recoge la emisión de fluorescencia, lo que determinará el nucleótido unido en esa posición. **(C)** Químicamente se elimina el fluoróforo y se revierte la capacidad terminadora de síntesis para comenzar un nuevo ciclo.

En este tipo de secuenciación, en cada ciclo de secuenciación solo se incorpora una base a cada molécula que está siendo sintetizada, ya que los nucleótidos fluorescentes empleados tienen bloqueado su extremo 3'. Además, cada tipo de nucleótido terminador está acoplado a un fluoróforo diferente y, en cada ciclo, se obtiene una imagen de toda la superficie de la *flow-cell*. Tras la obtención de la imagen, el agente bloqueante del extremo 3' se retira (y con él, también, el fluoróforo) antes del siguiente ciclo de secuenciación, con lo cual la reacción continúa base a base⁴⁹. Cada

imagen, por tanto, tiene señales de fluorescencia diferentes en los distintos *clusters* (en función de cuál sea la última base incorporada en ellos), de modo que la secuencia de imágenes permite reconstruir la secuencia de bases incorporadas en cada *cluster*. Esta técnica, más robusta, y capaz de distinguir mejor el tamaño de polímeros de un mismo nucleótido, es el principio de la actual plataforma de secuenciación Illumina, que es la metodología de referencia en secuenciación genómica de segunda generación⁵⁰.

4.4 Control de calidad y análisis de datos

La interpretación de los datos crudos originados por el secuenciador requiere recursos bioinformáticos capaces de convertir imágenes en datos, datos en información interpretable e información en conocimiento utilizable. En este proceso se pueden distinguir tres tipos de análisis: primario, secundario y terciario⁵¹. Cada uno de ellos aborda un paso necesario para la transformación de datos crudos en conocimiento clínicamente accionable⁵².

El análisis primario transforma los archivos de señales generados por el secuenciador en archivos de datos, los denominados FASTQ, que contienen secuencias de caracteres A, C, G y T asociadas a una puntuación que valora la calidad de las bases de cada secuencia. Sobre estos archivos FASTQ se realiza un control de calidad, en el que se filtran las lecturas de baja calidad y se eliminan las secuencias de los adaptadores^{53,54}. El análisis secundario implica el alineamiento de las lecturas con respecto a una secuencia de referencia (p. ej., el genoma humano) y la identificación de variantes entre la secuencia obtenida de la muestra objeto de análisis y esa secuencia de referencia⁵⁵. A continuación viene el análisis terciario, que consiste en la anotación, filtrado e interpretación de variantes para poder evaluar su origen, singularidad, impacto funcional y, en el caso de muestras de pacientes, sus consecuencias clínicas⁵⁶.

La NGS y sus mejoras continuas han conllevado reducciones drásticas tanto en los tiempos de respuesta de los estudios de diagnóstico genético como en los costes asociados a ellos. Con el objetivo de abaratar aún más los costes de secuenciación y de aumentar la fiabilidad de las secuencias resultantes, se han desarrollado secuenciadores de tercera y de cuarta generación. La secuenciación de tercera

generación, también conocida como secuenciación SMRT (Single Molecule Real Time), se caracteriza principalmente por la lectura de la secuencia a partir de una única molécula de ADN, y está diseñada para proporcionar lecturas más precisas a nivel individual y bastantes más largas (de hasta 30-50 kb), que faciliten el posterior ensamblado, para así, poder ampliar el abanico de aplicaciones⁴³. Por último, la secuenciación por nanoporos (considerada por algunos como de tercera generación y por otros como de cuarta generación) se basa en el paso de una única molécula de ADN a través de un poro de 1 nanómetro de diámetro^{48,57}. Al tiempo que la molécula avanza, se van realizando medidas de la corriente eléctrica que pasa a través del poro. Estos cambios son diferentes en función de la base que esté atravesando el poro en cada momento, lo que permite descifrar la secuencia del ADN. Aunque aún está en vías de optimización, se prevé que será capaz de secuenciar fragmentos de ADN de varias kilobases completos, en lugar de generar lecturas de un tamaño específico⁵⁸.

No obstante, a pesar de la existencia de estos avances metodológicos, en la práctica clínica todavía no se están utilizando tecnologías de tercera y cuarta generación, principalmente debido a sus dificultades de escalabilidad y fiabilidad. La aplicación diagnóstica de la secuenciación masiva en la actualidad se basa en tecnologías de segunda generación. El uso de estas herramientas se encuentra en constante expansión gracias al auge de la medicina de precisión: un nuevo paradigma de atención médica orientada a personalizar y optimizar la prevención, el diagnóstico y el tratamiento para cada paciente⁵⁹.

5 Medicina de precisión

La medicina de precisión, también llamada "medicina personalizada" o "medicina P4", es un prototipo emergente de atención médica que abarca la prevención, el diagnóstico y el tratamiento de enfermedades, basado en la integración de las características genómicas y moleculares, los hábitos del paciente y la información sobre su situación clínica. El objetivo inherente a esta estrategia es evitar que se produzca la enfermedad y, en el caso de que esto no sea posible, favorecer la prescripción eficiente de fármacos o hábitos saludables personalizados.

El uso de la NGS, la consecuente disponibilidad de bases de datos del genoma humano a gran escala y el avance de las herramientas bioinformáticas se consideran pilares fundamentales para el avance de este nuevo enfoque, denominado P4 por hacer referencia a las iniciales de las 4 características que lo representan⁶⁰:

- **Predictivo:** mediante la combinación de la genómica junto con otras metodologías diagnósticas, identifica riesgos de desarrollar determinadas enfermedades, incluso antes del inicio de sus síntomas.
- **Preventivo:** la detección temprana de una enfermedad abre la posibilidad de nuevas opciones de tratamiento y ayuda a las personas a tomar decisiones informadas sobre sus estilos de vida.
- **Personalizado:** el conocimiento de los procesos moleculares y celulares de cada individuo junto con los factores ambientales, contribuye a determinar la causa más probable de los síntomas, así como potenciales terapias dirigidas específicamente a corregir o contrarrestar esa causa.
- **Participativo:** el médico debe informar y hacer partícipe al paciente de sus características genómicas particulares, su estilo de vida y la interpretación de sus datos, lo que permite una toma de decisiones informada.

Hoy en día, gracias al auge de la medicina de precisión, el diagnóstico, el pronóstico y el tratamiento de gran parte de las patologías con base genética están fundamentados en la detección de alteraciones en el genoma⁶¹. Como ya hemos comentado previamente, la NGS tiene capacidad para caracterizar variaciones genéticas de una manera mucho más costo-efectiva que la secuenciación tradicional Sanger. Por ello, la explotación óptima de todos los datos obtenidos mediante esta tecnología conducirá a una comprensión integral de los eventos moleculares que se encuentran en la base del desarrollo y la evolución de la enfermedad.

6 Diagnóstico genómico mediante NGS

El desarrollo de las técnicas de secuenciación masiva ha supuesto un punto de inflexión en la comprensión de la enfermedad, abriendo una nueva era en el diagnóstico

genómico. La creciente disponibilidad de datos de variación genética propiciada por esta metodología ya ha dado como resultado el descubrimiento de un mayor número de genes asociados a trastornos hereditarios y, en consecuencia, un mejor diagnóstico de la enfermedad, que se traduce en pronóstico, manejo, vigilancia y asesoramiento genético más precisos, además de estimular la investigación de nuevas terapias⁶².

La búsqueda de la medicina personalizada ha impulsado en gran medida la incorporación progresiva de la NGS a la práctica clínica con sus diferentes aplicaciones, abarcando éstas un abanico amplio y creciente. Las principales son: la secuenciación de genoma completo (*Whole Genome Sequencing*, WGS), la secuenciación del exoma completo (*Whole Exome Sequence*, WES) y la secuenciación de paneles de genes. La selección de una u otra debe ser una decisión tomada por un asesor genético o un genetista clínico y es dependiente del objetivo del estudio, del fenotipo y de las manifestaciones del paciente.

6.1 Secuenciación genómica completa (WGS)

La secuenciación genómica completa consiste en la determinación de la secuencia del genoma completo de un individuo en un único procedimiento, llamado WGS por las siglas en inglés de *Whole Genome Sequencing*. Esta opción ofrece una visión integral del genoma, al abarcar tanto las regiones codificantes de proteínas como las no codificantes y tanto el ADN nuclear como el ADN mitocondrial. En consecuencia, permite identificar variantes genéticas que pueden estar asociadas con la enfermedad en regiones no inspeccionadas por otras metodologías.

La WGS es una prueba con una importante aplicabilidad clínica potencial y con una gran capacidad para proporcionar conocimiento sobre la diversidad humana y las enfermedades genéticas⁶³. Sin embargo, en la actualidad todavía existen ciertos desafíos que impiden su implementación en la práctica asistencial. En el año 2010, cada genoma humano individual secuenciado contenía de media aproximadamente 150.000 polimorfismos de un solo nucleótido (*single nucleotide Polymorphisms* o SNPs)⁶⁴ que no estaban presentes en las bases de datos de variación genómica. En la actualidad, con la disponibilidad de muchas más secuencias genómicas humanas individuales completas,

este número ha disminuido sensiblemente hasta un rango de entre 5.000 y 6.600, aunque la determinación de su relevancia clínica aún representa un reto muy difícil de abordar⁶⁵. Por esta razón, la complejidad del análisis de la gran cantidad de datos generados con el fin de traducir las secuencias brutas en un resultado que beneficie a los pacientes y los costes asociados a ella, constituyen los principales inconvenientes de la aplicación clínica de la WGS⁶⁶. Hasta que se superen estos obstáculos, en el ámbito de la salud esta alternativa está más orientada al campo de la investigación, con el descubrimiento de nuevos genes o variantes causales de enfermedades, así como al ensamblado de genomas aún no explorados (por ejemplo, de nuevos patógenos).

6.2 Secuenciación del exoma (WES)

La secuenciación del exoma es una técnica mediante la cual se establece la secuencia de todas las regiones codificantes de proteínas, la parte funcional mejor explorada y comprendida del genoma y que, como se ha mencionado previamente, corresponde aproximadamente con el 1,5 % del mismo. Su utilidad clínica es especialmente relevante en el diagnóstico genómico de patologías con extrema heterogeneidad, en pacientes con dos o más fenotipos no relacionados o en ausencia de características clínicas claves en el momento del estudio, recomendándose la realización de tríos padre-madre-paciente, con el fin de mejorar su rendimiento diagnóstico⁶⁷.

La secuenciación del exoma de un paciente puede revelar la expansión de los fenotipos asociados con alteraciones genéticas ya descritas en otros pacientes con patologías similares, pero aparentemente diferentes, identificar nuevos genes candidatos e incluso identificar variantes patogénicas en genes relacionados con trastornos distintos del que motivó el análisis genético⁶⁸. El principal inconveniente de esta metodología reside en la profundidad de la secuenciación, es decir, el número de veces que se lee cada una de las posiciones que queremos secuenciar. Debido al elevado número de regiones que abarca un exoma, lo más factible económicamente es utilizar una profundidad de lectura menor, lo cual repercute directamente en la capacidad de detección de algunas alteraciones, por ejemplo, mutaciones subclonales o en mosaico⁶⁹.

⁷². Este fenómeno se ve potenciado porque la distribución de la profundidad de lecturas no es homogénea, sino desigual en distintas regiones del exoma, pudiendo quedar insuficientemente cubiertas regiones genómicas relevantes para la patología que presenta el paciente.

6.3 Paneles de genes

Los estudios basados en paneles de genes consisten en un análisis específico de pequeñas secciones del genoma que corresponden a regiones codificantes y/o no codificantes en genes que están clínicamente asociados a una patología o conjunto de patologías concretas relacionadas con el fenotipo observado en el paciente. Representan la alternativa más extendida para el estudio de enfermedades genéticamente heterogéneas y con una variabilidad clínica significativa.

Esta aproximación de secuenciación dirigida está orientada a maximizar la proporción de secuencia clínicamente útil que se obtiene, facilitando su interpretación y disminuyendo el número de variantes incidentales identificadas (aquellas con potencial relevancia clínica que están asociadas a enfermedades diferentes de las que motivaron el estudio). Acotar la región secuenciada conlleva estas ventajas, pero también representa la principal de las limitaciones de esta aproximación, al estar condicionada por el conocimiento de la genética de cada patología en el momento de diseñar los reactivos necesarios para crear y enriquecer la librería (oligonucleótidos o sondas de captura). El estudio solo tendrá éxito si el gen que causa la enfermedad está incluido en el panel, lo cual limita el rendimiento diagnóstico de la prueba a medio plazo^{73,74}.

El primer paso, antes de poner a punto un ensayo de este tipo, es definir el objetivo clínico y, en consecuencia, determinar, idealmente de manera multidisciplinar, las regiones diana que queremos analizar y que son clínicamente relevantes para el diagnóstico de la patología objeto del estudio⁷⁵. El número de genes que abarca cada panel es variable y dependiente de las circunstancias y prioridades de cada laboratorio. Algunos laboratorios incluyen solo genes diana para los cuales existe suficiente literatura que justifica su relevancia diagnóstica, terapéutica o pronóstica. En cambio,

otros paneles extienden su análisis a conjuntos de genes más amplios, incluyendo genes adicionales que están siendo investigados en ensayos clínicos y/o cuya asociación con la enfermedad en la literatura científico-médica es incipiente.

Dentro de este tipo de análisis dirigido, existen dos alternativas: paneles con enriquecimiento específico, mediante sondas u oligonucleótidos diseñados en base a los genes de interés (también llamados genes diana) o “paneles virtuales”, en los que se lleva a cabo una secuenciación del exoma y se extraen únicamente los datos de las variantes que afectan a los genes diana que constituyen el panel.

Al ser una prueba cuya finalidad es el diagnóstico clínico, en cada diseño del panel se debe evaluar tanto la sensibilidad como la especificidad de la técnica para detectar alteraciones⁷⁶. En los paneles virtuales, la profundidad de lecturas en las diferentes regiones secuenciadas es habitualmente menor que en los verdaderos paneles de genes. Como se ha mencionado previamente para el exoma, este hecho, combinado con la heterogeneidad en la distribución de las lecturas, puede provocar que determinadas regiones, especialmente relevantes para una patología, no queden cubiertas por los resultados de secuenciación de un panel virtual, lo que provoca una menor sensibilidad y especificidad. Motivados por este hecho, en la presente Tesis Doctoral hemos centrado nuestro trabajo experimental en el desarrollo y aplicación de ensayos basados en paneles de captura específicos para el diagnóstico de tres enfermedades con base genética: el cáncer, la sordera y la ceguera, todas ellas con una extrema heterogeneidad genética.

7 Paneles para el diagnóstico genómico de enfermedades de base genética

La priorización implícita en el diseño de un panel permite disminuir el número de genes y variantes candidatos a evaluar, puede aumentar la probabilidad de identificar una causa genética subyacente y evita la identificación de hallazgos secundarios, irrelevantes para el diagnóstico de la patología deseada⁷⁷⁻⁷⁹. Estas mejoras han cambiado el modelo de diagnóstico molecular y, en consecuencia, el modelo de

diagnóstico clínico, especialmente en el caso de trastornos genéticamente y clínicamente heterogéneos, como los que son objeto de esta Tesis Doctoral⁸⁰.

7.1 Cáncer

El cáncer es un conjunto de enfermedades en las que una o varias células normales se transforman en células anormales que comienzan a crecer y multiplicarse de una forma descontrolada, llegando a invadir tejidos cercanos e, incluso, a diseminarse a otras partes del cuerpo a través del torrente sanguíneo y el sistema linfático⁸¹. Esa transformación celular, causada por alteraciones en la secuencia del ADN de determinados genes fundamentales para el control del crecimiento y división celular normal, es el resultado de la interacción entre los factores genéticos de cada paciente y tres categorías de agentes externos: carcinógenos físicos (radiaciones ultravioletas e ionizantes), carcinógenos químicos (humo, amianto, etc.) y carcinógenos biológicos (determinados virus, bacterias o parásitos)⁸². Nuestras células están constantemente detectando y reparando alteraciones que ocurren en nuestro genoma al vertiginoso ritmo de 4 cada 5 segundos en cada célula (0,8/seg.) pero, en algunas ocasiones, esas alteraciones evaden la maquinaria de reparación, se acumulan y pueden acabar provocando cáncer⁸³.

Las células cancerosas cuentan con múltiples mecanismos que les permiten saltarse las restricciones normales sobre el crecimiento celular, modificar el microambiente favoreciendo su propia proliferación, invadir barreras facilitando su propagación a otros órganos y evadir la vigilancia inmunológica⁸⁴. La elevada heterogeneidad molecular y celular característica de los tumores, que tiene un papel crucial en esas facultades adquiridas por el tejido canceroso, surge de la naturaleza estocástica de la evolución darwiniana, la cual cumple 3 premisas: que las características deben variar dentro de una población, que esta variación debe ser heredable de generación en generación y que debe haber competencia por la supervivencia dentro de la población⁸⁵.

Hoy en día, el cáncer es una de las principales causas de muerte en todo el mundo y se prevé que el número de muertes causadas por el mismo aumente

notablemente, debido en parte al envejecimiento de la población mundial. Esta situación exige estrategias que permitan una detección más temprana, una mejor estratificación de los tumores y el desarrollo de terapias preventivas efectivas, ya que una gran parte de los cánceres son altamente tratables y curables si se detectan en estadios tempranos (I o II)⁸⁶.

Siempre y cuando sea posible, el análisis genómico amplio de los tumores es la mejor estrategia para identificar las alteraciones genéticas responsables de la enfermedad. En 1982, mediante secuenciación Sanger, se identificaron las primeras alteraciones genéticas causantes de cáncer, entre ellas, una única variación nucleotídica que conducía a la activación del oncogén HRAS en células de carcinoma de vejiga, descrita por el equipo del español Mariano Barbacid⁸⁷. En la actualidad, la combinación entre el uso de la NGS y una atención personalizada del paciente con cáncer está permitiendo mejorar la prevención y mantener o, incluso, incrementar la calidad de vida de los pacientes durante el mayor tiempo posible⁸⁸. En el cáncer, los estudios genéticos más habituales se basan en paneles somáticos, orientados a identificar fármacos específicos para su tratamiento, y paneles germinales, cuya finalidad es detectar una predisposición heredable al desarrollo de dicha enfermedad.

7.1.1 Paneles somáticos

El concepto de mutación somática se refiere a cualquier variación genética adquiridas por una célula que no forma parte de la línea germinal, frecuentemente causada por factores ambientales, como la exposición a radiaciones o a ciertas sustancias químicas. Este tipo de alteraciones no son heredables, pero sí se acumulan en las células a lo largo de la vida de un individuo, por lo cual, aunque la mayoría de ellas no tienen un efecto notable, algunas pueden alterar funciones celulares clave y desencadenar el desarrollo de cáncer.

La secuenciación del genoma ha proporcionado una visión detallada de los procesos mutacionales y de los genes involucrados en el desarrollo del tumor. Tras la finalización del PGH, se emprendieron dos iniciativas de gran envergadura, el Atlas del Genoma del Cáncer (TCGA) y el Consorcio Internacional del Genoma del Cáncer (ICGC),

ambos con el objetivo común de documentar sistemáticamente las mutaciones somáticas que desencadenan el desarrollo de los tumores más frecuentes. El TCGA fue proyectado en 2005 con el objetivo de crear mapas multidimensionales completos que recogiesen los cambios genómicos claves en los tipos y subtipos principales de cáncer mediante la secuenciación de las regiones codificantes de proteínas de aproximadamente 11.000 genomas individuales y 33 tipos de cáncer (<https://portal.gdc.cancer.gov>)⁸⁹. Tres años más tarde, en 2008, se puso en marcha el ICGC con el cometido inicial de secuenciar las regiones codificantes de más de 20.000 genomas individuales y 22 tipos de cáncer (<https://dcc.icgc.org>)⁹⁰. Asimismo, en febrero del 2020, el Consorcio ICGC / TCGA *Pan-Cancer Analysis of Whole Genomes* (PCAWG) actualizó el estatus de ambos proyectos con la presentación de los resultados obtenidos en 2.658 genomas completos de cáncer en 38 tipos de tumores diferentes^{85,91}. Los catálogos de alteraciones genómicas que se han ido generando tras estos estudios han servido como germen de futuras investigaciones para la formulación de mejores estrategias de prevención, diagnóstico y tratamiento de cada tipo de cáncer.

Las células cancerosas no solo tienen mutaciones que resultan en la expresión desregulada de oncogenes (aceleradores de la transformación celular) y genes supresores de tumores (frenos de la formación de tumores). En muchos casos, estos cambios también alteran la expresión de cientos de genes implicados en vías de regulación de la actividad celular más sutiles. Una mejor comprensión de la etiología molecular del cáncer ha permitido el desarrollo de nuevos fármacos denominados comúnmente terapias dirigidas. La estrategia ideal contra el cáncer es aquella que restringe selectivamente la proliferación y la supervivencia de las células tumorales sin afectar a las células sanas. A diferencia de la quimioterapia convencional, citotóxica, a menudo limitada por efectos secundarios graves y, en ocasiones, potencialmente mortales, los fármacos dirigidos están diseñados para interferir específicamente las dianas moleculares que son responsables del fenotipo maligno, resultando en un aumento de su índice terapéutico y una disminución de sus efectos secundarios sistémicos. Aunque es frecuente la aparición de resistencias a este tipo de terapias dirigidas, la caracterización molecular de los mecanismos de resistencia también

permite diseñar terapias de segunda generación o terapias combinadas destinadas a prevenir la resistencia o a restablecer la respuesta⁹².

Uno de los primeros medicamentos antitumorales dirigidos fue el tamoxifeno, introducido a principios de la década de 1970 y diseñado para bloquear el crecimiento y avance de las células tumorales en el cáncer de mama receptor de estrógenos positivo (ER+). A partir de entonces, el desarrollo de nuevas tecnologías y sus aplicaciones en el campo de la genómica han permitido la identificación de numerosas vías que están desreguladas en los cánceres humanos y, en consecuencia, el descubrimiento de múltiples genes clave en el desarrollo de determinados cánceres.

Los paneles somáticos en el diagnóstico del cáncer pueden identificar de manera integral las alteraciones genéticas presentes en el tejido tumoral que se han asociado previamente con la respuesta a terapias oncológicas dirigidas. Así, en muchos casos su identificación es requerida, como biomarcador de resistencia o sensibilidad, para la aplicación de terapias específicas aprobadas por las agencias reguladoras de medicamentos (ej.: la *U.S Food & Drug Administration* (FDA) o por la Agencia Europea del Medicamento (EMA))⁹³. El impacto clínico de una variante debe determinarse según la evidencia disponible en el momento del estudio y dentro del contexto clínico del paciente. De este modo, en 2015 se consideraba que alrededor del 10 % de los pacientes con cáncer avanzado tenían una mutación accionable, es decir, tenía asociada una terapia dirigida clínicamente validada y aprobada específicamente para ese tipo de tumor⁹⁴. En la actualidad, tras la explosión de conocimiento que han proporcionado la secuenciación de miles de genomas tumorales en los últimos años y el desarrollo y aprobación de nuevos fármacos dirigidos, esa estimación alcanza aproximadamente el 25 %⁹⁵. Dadas las ventajas que puede proporcionar conocer el mecanismo subyacente de las vías biológicas en pacientes con cáncer avanzado o refractario al tratamiento, las decisiones terapéuticas se guían cada vez más por el perfil genómico del tumor que por factores convencionales, como el estadio de la enfermedad o las características histopatológicas⁹⁶.

7.1.2 Paneles germinales

Las mutaciones germinales o hereditarias son aquellas que afectan a las células reproductoras del organismo (óvulo y espermatozoide) y, por ello, se pueden transmitir de generación en generación. En algunos casos, estas alteraciones en la línea germinal son responsables de una mayor predisposición al desarrollo de una enfermedad hereditaria y su identificación es usada como criterio diagnóstico.

El cáncer es una enfermedad multifactorial causada por el efecto combinado de factores genéticos y ambientales. Las estimaciones actuales consideran que el 5-10 % de los casos de cáncer se desarrollan como consecuencia de la presencia de una mutación germinal que se transmite entre los miembros de una misma familia de acuerdo a distintos patrones de herencia⁹. Para una misma mutación, el tipo de cáncer y la edad de aparición son variables entre familias, pero también dentro de una misma familia. Es importante señalar que lo que se hereda entre generaciones es la susceptibilidad a padecer cáncer, lo cual no implica la certeza de desarrollarlo en todos los casos.

En los últimos años se han identificado múltiples genes implicados en predisposición hereditaria al desarrollo de cáncer. Estos descubrimientos han desencadenado una notable expansión de nuestro conocimiento sobre la susceptibilidad genética, su función e impacto en la biología tumoral, así como sobre las posibles intervenciones preventivas y terapéuticas basadas en ellas. Entre los tipos de cáncer hereditario más prevalentes se encuentran los causados por mutaciones en genes de susceptibilidad al cáncer de mama 1 y (*BRCA1* y *BRCA2*, respectivamente) o los implicados en la reparación de errores de emparejamiento de bases del ADN, también llamados genes MMR por las siglas de *MissMatch Repair* (*MLH1*, *MSH2*, *MSH6* y *PMS2*).

BRCA1 y *BRCA2* son genes supresores de tumores involucrados en la recombinación homóloga, una vía crítica de reparación del ADN encargada de asegurar la estabilidad del genoma⁹⁷⁻⁹⁹. La alteración patogénica en un único alelo de estos genes se considera responsable del síndrome de cáncer de mama y de ovario hereditario, que

no solo implica cáncer de mama de inicio temprano, sino también un mayor riesgo de cáncer de ovario, páncreas, estómago, laringe, trompas de Falopio y próstata. Este síndrome es responsable del 5-7 % de todos los casos de cáncer de mama. Además, las mujeres portadoras de mutaciones en estos genes tienen un 50-80 % de probabilidad de desarrollar cáncer de mama y un 30-50 % de probabilidad de desarrollar cáncer de ovario¹⁰⁰. Las mutaciones en los genes *BRCA1* y *BRCA2*, combinadas, afectan al 85 % de las familias con cuatro o más casos de cáncer de mama¹⁰¹.

Las alteraciones patogénicas en los genes *MLH1*, *MSH2*, *MSH6* y *PMS2* tienen como consecuencia defectos en la reparación de errores del emparejamiento de bases del ADN que causa síndrome de Lynch o síndrome del cáncer de colon no polipósico hereditario (HNPCC). Este síndrome es la causa más común de cáncer colorrectal y de cáncer de endometrio hereditarios. Otros cánceres extracolónicos asociados con este síndrome incluyen los de ovario, estómago, tracto urotelial e intestino delgado y, con menos frecuencia, cerebro, tracto biliar, páncreas y próstata. Aproximadamente el 70-85 % de los casos de síndrome de Lynch son causados por mutaciones en *MLH1* o *MSH2* y las variaciones patogénicas en *MSH6* y *PMS2* representan cada una entre el 10 % y el 20 % de los casos¹⁰²⁻¹⁰⁴. Asimismo, en los últimos años se ha demostrado que las deleciones germinales en el extremo 3' del gen *EPCAM*, que conducen a un silenciamiento epigenético del gen *MSH2*, están involucradas en aproximadamente el 1-3 % de los casos con síndrome de Lynch¹⁰⁵. El riesgo de cáncer colorrectal en familias con deleciones en *EPCAM* es comparable al de los casos con mutaciones *MSH2*, siendo, sin embargo, inferior el riesgo de cáncer de endometrio, lo cual ilustra la utilidad de identificar la causa genética subyacente a un fenotipo tumoral por encima de identificar exclusivamente dicho fenotipo (que en este caso es la pérdida de expresión tumoral de la proteína *MSH2* en ambos casos)¹⁰⁶. En todos estos casos, y en la mayoría de las familias con cáncer hereditario, el modo de herencia de la susceptibilidad al desarrollo de cáncer es autosómico dominante.

Más allá de los genes *BRCA1*, *BRCA2*, *MLH1*, *MSH2*, *MSH6* y *PMS2*, existen decenas de otros genes que, en conjunto, son responsables de una gran proporción de los casos de cáncer hereditario¹⁰⁷. No obstante, actualmente, todavía existe un

porcentaje muy elevado de pacientes con cáncer familiar en los que se desconoce su potencial causa genética. Por estas razones es necesario y útil identificar variantes en diferentes genes, la contribución de estas al riesgo hereditario, la combinación de alelos de susceptibilidad de baja/moderada penetrancia, o su combinación con factores ambientales. Los datos respaldan la utilidad del uso de paneles germinales para el estudio del cáncer. Por ejemplo, recientemente, Domínguez-Valentín y colaboradores, tras un estudio de 191 pacientes con cáncer familiar en los que no se habían identificado mutaciones en los genes *BRCA1*, *BRCA2*, *PTEN* o *TP53*, ni en los genes MMR, observaron que más de la mitad de las variantes patogénicas de la línea germinal identificadas afectaban a los genes *ATM* y *CHEK2*, con penetrancia moderada, genes que en muchos hospitales no son analizados de manera rutinaria¹⁰⁸. Por estas razones, los paneles multigénicos son una estrategia en expansión para el diagnóstico de la predisposición hereditaria al desarrollo de cáncer^{109,110}.

Una ventaja adicional de este tipo de estudios es la potencial capacidad que tienen para identificar portadores de mutaciones causantes de síndromes de susceptibilidad al cáncer bien conocidos en familias con fenotipos atípicos^{111,112}. Asimismo, para los pacientes con antecedentes familiares de cáncer o predisposición genética al cáncer, la detección precisa de las mutaciones responsables permite una evaluación del riesgo y una vigilancia adecuados, siendo crucial para facilitar la prevención y la detección temprana de futuros tumores tanto en los pacientes como en sus familiares y, en algunos casos, facilitando la selección de terapias dirigidas¹¹³. Por tanto, ser capaces de identificar estas alteraciones es de suma importancia para proporcionar a los pacientes un manejo clínico integral y personalizado^{114,115}.

7.2 Sorderas

La sordera es el trastorno sensorial más común que existe. En base a la localización de la lesión se puede clasificar en pérdida auditiva conductiva (oído externo o medio), neurosensorial (oído interno, nervio auditivo o vía auditiva central) o mixta, con características clínicas de las dos anteriores¹¹⁶. En esta Tesis Doctoral abordamos las pérdidas auditivas neurosensoriales o mixtas. La hipoacusia neurosensorial (HNS) representa aproximadamente el 90 % de todos los casos de

pérdida auditiva humana y se estima que el componente genético en ella es muy superior al de cualquiera de los otros dos tipos¹¹⁷.

La sordera profunda congénita afecta aproximadamente a 1 de cada 1000 nacidos vivos, mientras que 1 de cada 1000 niños adicionales sufrirá pérdida auditiva antes de convertirse en adulto¹¹⁸. El 60 % de la HNS congénita/de inicio temprano es causada por factores genéticos y, a menudo, aparece en ausencia de antecedentes familiares de sordera¹¹⁹. El resto de los casos parecen ser idiopáticos o surgir por el efecto de patógenos como la rubéola, el citomegalovirus u otras infecciones infantiles¹²⁰⁻¹²³.

De los casos con una causa genética, se considera que aproximadamente el 30 % son sindrómicos¹²²⁻¹²⁶. Hasta la fecha se han asociado más de 400 síndromes con pérdida auditiva¹²⁷. Los casos restantes corresponden a la pérdida auditiva hereditaria no sindrómica, en la cual la única consecuencia del defecto genético detectable clínicamente es la sordera. Dentro de esta categoría, en el 75-80 % de los casos la hipoacusia es heredada de forma autosómica recesiva y, aproximadamente, en el 20 % es autosómica dominante. Las formas ligadas al sexo y mitocondriales representan menos del 5 %¹²⁵. Las hipoacusias asociadas a loci genéticos se denominan con una nomenclatura basada en siglas canónicas: *DFNB* (autosómica recesiva), *DFNA* (autosómica dominante) y *DFNX* (ligada al X)¹¹⁸.

El locus genético *DFNB1*, que abarca los genes *GJB2* y *GJB6*, se considera responsable del 10-40 % de todos los casos de pérdida auditiva congénita en determinadas poblaciones, con incidencia variable dependiente de cada población^{6,123}. Codifican proteínas de membrana denominadas conexinas (Cx), Cx26 y Cx30, respectivamente, que están implicadas en la formación de conexiones intercelulares (uniones GAP) y ejercen una función crítica en la comunicación entre las células de soporte de la cóclea y la homeostasis de la perilinfa y endolinfa. Las mutaciones en el gen *GJB2* se consideran la principal causa de HNS no sindrómica autosómica recesiva en la mayoría de las partes del mundo. Entre ellas la alteración c.35delG, la más prevalente en nuestro medio, explica hasta el 70 % de los casos de hipoacusia

hereditaria en algunas poblaciones de América del norte y Europa^{6,128}. En cuanto al gen *GJB6*, situado junto a *GJB2*, se ha demostrado que sus mutaciones más prevalentes, las grandes deleciones del(*GJB6*-D13S1830) y del(*GJB6*-D13S1854), son capaces de silenciar a nivel transcripcional la expresión del gen de *GJB2* que está en disposición *cis* (en el mismo alelo), lo cual explica que, cuando una de estas deleciones se encuentra en *trans* (en alelos distintos) con una mutación patogénica en *GJB2*, el fenotipo auditivo sea similar al que se observa en pacientes homocigotos para una alteración patogénica en *GJB2*¹²⁹⁻¹³¹. La hipoacusia debida a mutaciones en el locus *DFNB1* se caracteriza por ser generalmente bilateral, de inicio prelingual, con un grado de severidad que varía de moderado a profundo y con variaciones inter e intrafamiliares, es decir, que aún con la misma mutación, puede encontrarse una evolución diferente. Ante una expresividad tan variable, los estudios genéticos se han vuelto una herramienta imprescindible en el diagnóstico de esta patología.

Actualmente, el estudio genético de *DFNB1* para pacientes con pérdida auditiva no sindrómica incluye la secuenciación *GJB2* y el análisis de las dos deleciones más comunes de *GJB6* pero, tras él, muchos casos permanecen sin diagnosticar. Esto se debe a la extrema heterogeneidad genética y fenotípica de la pérdida auditiva, con más de 400 síndromes que incluyen hipoacusia como manifestación característica y más de 100 genes asociados con HNS no sindrómica¹¹⁸. Con la tecnología NGS se ha vuelto factible y asequible secuenciar rutinariamente una gran cantidad de genes por paciente, lo cual ha facilitado el diagnóstico en gran medida^{132,133}. Tanto la guía del año 2014 para la «Evaluación clínica y el diagnóstico etiológico de la hipoacusia» del Colegio Americano de Genética Médica y Genómica (ACMG)¹¹⁸, como las recomendaciones españolas del año 2015 para el «Diagnóstico etiológico de la sordera infantil», de la Comisión para la Detección Precoz de la Hipoacusia (CODEPEH)¹³⁴, recomiendan incorporar el diagnóstico genético precozmente en los protocolos para el diagnóstico de la sordera [**Figura 5**].

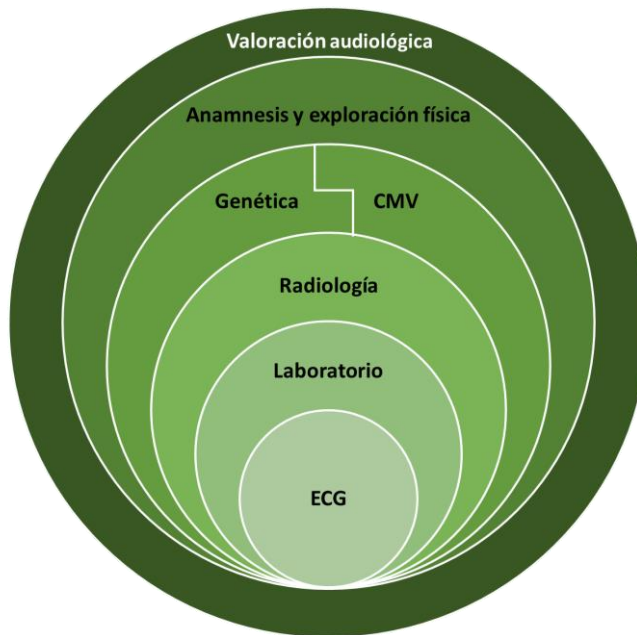


Figura 5. Secuencia recomendada para el diagnóstico etiológico por la CODEPEH.

Hacer un diagnóstico precoz de la pérdida auditiva tiene un gran impacto en el curso clínico del paciente y en su desarrollo vital, permitiendo una intervención o monitorización tempranas¹³⁵. Por este motivo, muchos países ya han instaurado programas de cribado neonatal universales de hipoacusia conforme a las recomendaciones del *Joint Committee on Infant Hearing*¹³⁶, que establece la conveniencia de que la detección no se demore más allá del primer mes de vida y que se pueda disponer de la confirmación diagnóstica en el tercer mes para asegurar que los niños reciban el tratamiento adecuado antes de los 6 meses. El principal objetivo es lograr la adquisición del lenguaje hablado y el máximo desarrollo de los niños con un déficit auditivo a todos los niveles: personal, cognitivo, educativo y social¹³⁴.

La NGS ha aumentado notablemente la eficacia tanto en la detección de variantes patogénicas en genes de sordera conocidos con fines de diagnósticos, como en la identificación de nuevos genes asociados^{137,138}.

7.3 Cegueras

La ceguera es una enfermedad de tipo sensorial que consiste en la pérdida total o parcial del sentido de la vista. Al igual que ocurre en la hipoacusia, la pérdida visual se clasifica en base a la localización anatómica de la lesión que la causa. En esta Tesis Doctoral abordamos la ceguera causada por lesiones en el vítreo, la coroides, la retina y/o el nervio óptico, enfermedades en las cuales la etiología monogénica está respaldada por una evidencia sólida. Además, se seleccionaron estas cuatro estructuras oculares porque los fenotipos que causan sus alteraciones pueden llegar a superponerse y generar dificultades a la hora de obtener un buen diagnóstico diferencial¹³⁹.

Las distrofias retinianas hereditarias (DRH) representan, con gran diferencia, la principal causa de ceguera transmisible de padres a hijos¹⁴⁰. Se trata de un conjunto de enfermedades que afectan a la retina, al vítreo y/o a la coroides, caracterizadas por una pérdida de visión lenta y progresiva, debida a la degeneración de los fotorreceptores y células epiteliales del pigmento retiniano. Individualmente, tienen una baja incidencia entre la población, pero en conjunto, alcanzan una prevalencia de aproximadamente 1 caso cada 3000 individuos¹⁴¹. Estos datos explican que se considere la principal causa de ceguera entre los adultos y la segunda más común en la infancia (la primera causa en niños son las cataratas)¹⁴².

La retinosis pigmentaria (RP) es la forma más común de degeneración retiniana hereditaria, presente en más de 1,5 millones de personas en todo el mundo, tiene una prevalencia estimada de 1 caso cada 4.000 individuos¹⁴³. Además de los efectos perjudiciales de la discapacidad visual, el 20-30 % de los pacientes con RP tienen una afección no ocular asociada, que varía desde cambios morfológicos leves hasta patologías potencialmente mortales, es decir, son casos sindrómicos. Los tipos más frecuentes de RP sindrómica son el síndrome de Usher y el síndrome de Bardet-Biedl¹⁴⁴. La forma no sindrómica es más frecuente, representando alrededor del 65 % de todos los casos^{145,146}. Dentro de esta categoría, el 30-40 % corresponden a las formas autosómicas dominantes, el 50-60 % a formas autosómicas recesivas y el 5-15 % a

formas ligadas al cromosoma X. Asimismo, también se han descrito patrones de herencia no mendelianos, entre los que se han incluido las herencias digénicas y mitocondriales, que explican una pequeña proporción de casos¹⁴⁷⁻¹⁴⁹.

Se han descrito más de 250 genes responsables de formas sindrómicas y/o no sindrómicas de DHR. La notable heterogeneidad genética característica de estas enfermedades no solo se debe a la gran cantidad de genes implicados, sino también al hecho de que las mutaciones en un gen específico pueden causar fenotipos distintos que varían ampliamente en severidad, progresión, y modo de herencia^{150,151}. Por ejemplo, las mutaciones en *ABCA4* pueden causar al menos cuatro formas diferentes de ceguera no sindrómica: retinosis pigmentaria, distrofia de conos y bastones, *fundus flavimaculatus* y enfermedad de Stargardt. Asimismo, las alteraciones patogénicas de *PRPH2* pueden ser responsables de retinosis pigmentaria tipo 7, retinosis *punctata albescens*, distrofia macular de la retina o distrofia coroidea areolar central (todas ellas con modo de herencia autosómico recesivo), amaurosis congénita de Leber (con modo de herencia autosómico dominante) y retinosis pigmentaria tipo 7 *PRPH2/ROM1* (con modo de herencia digénico recesivo). Además, aunque se han identificado distintos subtipos de DRH, existe una superposición fenotípica considerable que se hace aún más evidente con las pruebas genéticas¹⁵². Esta heterogeneidad genética y fenotípica y la existencia de patrones de herencia multigénicos y modificadores genéticos oscurecen y dificultan, en gran medida, el diagnóstico molecular de estas enfermedades. En esta situación, no son prácticos los enfoques convencionales de secuenciación Sanger. La secuenciación simultánea de todos los genes implicados se considera el mejor enfoque para el diagnóstico molecular de las DHR^{153,154}.

En ciertos tipos de ceguera hereditaria solo las pruebas moleculares proporcionan un diagnóstico preciso, por lo que las pruebas utilizadas para realizar el enfoque clínico de la enfermedad (electrorretinografía, tomografía de coherencia óptica, examen de fondo de ojo, autofluorescencia de fondo, adaptación de oscuridad o prueba de campo visual) no son suficientes para el diagnóstico¹⁵⁵. Esta situación ocurre, por ejemplo, en la neuropatía óptica hereditaria de Leber (LHON)¹⁵⁶.

Una peculiaridad que engloba a las DHR es su aparente “invisibilidad” hasta estadios avanzados de la enfermedad. Por esta razón, el diagnóstico precoz y el conocimiento de las causas que las originan son factores primordiales para reducir el impacto personal y psicológico de las mismas. Con la introducción de los paneles de genes, el diagnóstico y la atención a pacientes con enfermedades oculares han experimentado una revolución, identificando la causa genética aproximadamente en el 50 % de los casos. No obstante, las características clínicas de los individuos analizados y la existencia/ausencia de estudios moleculares previos pueden hacer variar notablemente el rendimiento diagnóstico¹⁵⁷.

Por último, en la pérdida de visión de etiología genética, el diagnóstico molecular no solo es importante para mejorar el asesoramiento de las familias y por su valor pronóstico, sino también para identificar a los pacientes candidatos a beneficiarse de los nuevos tratamientos específicos de genotipo. Estos tratamientos potenciales, principalmente de terapia génica, cuya finalidad es la prevención o la curación de la enfermedad, están siendo particularmente exitosos en oftalmología, al ser el ojo un órgano de fácil acceso, relativamente aislado y con una situación privilegiada inmunológicamente, lo cual disminuye las posibilidades de rechazo, entre otros riesgos¹⁵⁸.

8 Interpretación clínica de los resultados de diagnóstico genético

La interpretación clínica de las variantes genéticas es el proceso en el cual se evalúan y clasifican las variantes identificadas en un estudio genético en función de su relevancia médica para el paciente objeto de estudio. En la actualidad, tanto la interpretación de los resultados, como la traducción de los mismos en recomendaciones clínicas representan el principal desafío en el diagnóstico genético.

La comprensión precisa y detallada de los efectos causados por las variaciones en las diferentes regiones del genoma es imprescindible para resolver la relación variante-gen-enfermedad, el espectro de las alteraciones patogénicas en cada uno de los genes y las correspondientes consecuencias clínicas de las mismas. En los últimos

años se han desarrollado herramientas bioinformáticas y bases de datos específicas diseñadas con el fin de ayudar a comprender la importancia funcional de las variantes genéticas con respecto a sus posibles efectos sobre los genes y las enfermedades¹⁵⁹.

8.1 Bases de datos de variantes genéticas clínicamente relevantes

La interpretación de las variantes genómicas se ha convertido en el cuello de botella en los flujos de trabajo de secuenciación clínica, por lo que muchos hospitales y centros de investigación han creado de forma individual su propio repositorio para la interpretación de variantes y evaluación de la relevancia clínica¹⁶⁰. Estas bases de datos internas requieren de actualizaciones constantes, implican gran trabajo manual y carecen de consenso entre la comunidad científica.

La medicina de precisión requiere que toda esta información esté en un formato estandarizado, centralizada, de libre acceso, con capacidad de descarga masiva, consensuada y fácilmente interpretable para su aplicación en la clínica. Para satisfacer estos requisitos surgieron las primeras bases de datos públicas tanto de variantes somáticas como de variantes germinales que pueden ser consultadas por todos los especialistas para optimizar y homogeneizar la interpretación clínica.

8.1.1 Bases de datos de variantes somáticas

La oncología está cada vez más cerca de que la información molecular de cada paciente con cáncer se pueda utilizar para mejorar el diagnóstico, pronóstico y tratamiento con terapias dirigidas¹⁶¹. Para ello, las bases de datos de variantes somáticas ofrecen apoyo clínico a los oncólogos y especialistas y proporcionan información detallada basada en la evidencia disponible sobre alteraciones genéticas presentes en tumores. Aunque existen múltiples bases de datos de variantes somáticas, entre las más utilizadas se encuentran CIViC (*Clinical Interpretation of Variants in Cancer*) y OncoKB (*Precision Oncology Knowledge Base*), que anotan el efecto biológico y oncogénico, así como la importancia pronóstica y predictiva de las alteraciones moleculares somáticas para respaldar decisiones de tratamiento dirigidos¹⁶².

CIViC es el recurso ofrecido por la Universidad de Washington para la interpretación clínica de variantes en cáncer. Diseñada por un equipo interdisciplinar e internacional de expertos que colaboran de forma remota dentro de una interfaz de curación centralizada, en la actualidad, contiene información clínica de variantes en 424 genes relacionadas con terapias aprobadas por la FDA (<https://civicdb.org>)^{163,164}.

OncoKB es el repositorio web del Centro de Cáncer Memorial Sloan Kettering (*Memorial Sloan Kettering Cancer Center, MSKCC*), disponible también a través de un recurso web (<http://oncokb.org/>)¹⁶⁵. Curada por una red de investigadores clínicos y básicos del MSKCC, OncoKB contiene actualmente información detallada sobre alteraciones específicas en 675 genes de cáncer. A diferencia de CIViC, OncoKB no solo incluye información sobre el uso guiado por biomarcadores de terapias aprobadas por la FDA, sino también con respecto a las guías de la NCCN (*National Comprehensive Cancer Network*), las de la Sociedad Americana de Oncología Clínica (*ASCO, American Society of Clinical Oncology*), ClinicalTrials.gov, las recomendaciones de grupos de expertos y la literatura científica. Asimismo, aporta información sobre resultados clínicos negativos, que facilita evitar el uso de terapias que han demostrado ser ineficaces en contextos mutacionales específicos.

Basar la interpretación clínica de una variante en la información disponible en una única base de datos puede presentar desventajas. Por este motivo, se consideró una necesidad el estandarizar y coordinar los esfuerzos de curación clínico-genómica en todas las instituciones, y crear un recurso comunitario público capaz de consultar la información agregada. Así, en 2016 se fundó VICC (*The Variant Interpretation for Cancer Consortium*), un consorcio para reunir instituciones líderes que están desarrollando de manera independiente bases de datos integrales de interpretación de variantes de cáncer, entre las que se encuentran CIViC y OncoKB. Esta organización ha creado un servicio de consulta capaz de interrogar las asociaciones entre las alteraciones genéticas del cáncer y las acciones clínicas para cada tipo de cáncer, en base a la evidencia acumulada en todas las instituciones participantes en todo el mundo (cancervariants.org).

8.1.2 Bases de datos de variantes germinales

El uso generalizado de la NGS en el diagnóstico clínico para determinar las causas subyacentes de los fenotipos mendelianos resulta en la identificación continua de variantes no descritas previamente cuyo significado clínico es desconocido. Como consecuencia, la estandarización de la interpretación del significado clínico de variantes es una necesidad crítica. Se han desarrollado múltiples bases de datos de variantes germinales que facilitan la organización del conocimiento con respecto a las alteraciones genéticas y sus fenotipos asociados^{166,167}. Las 3 principales que se han consultado durante el desarrollo de este proyecto de Tesis Doctoral han sido OMIM (*Online Mendelian Inheritance in Man*), HGMD (*Human Gene Mutation Disease*) y ClinVar, que se describen brevemente a continuación:

OMIM es una base de datos que recoge la descripción de genes y fenotipos humanos y las relaciones entre ellos. Creada y editada por la Universidad Johns Hopkins con el aporte de científicos y médicos de todo el mundo, surgió en 1964 como MIM (*Mendelian Inheritance in Man*) llegando a publicar 12 ediciones impresas desde 1966 hasta 1998¹⁶⁸. Está disponible en Internet desde el año 1987 bajo la designación de "OMIM", desde el año 1995 en la World Wide Web del Centro Nacional de Información Biotecnológica (NCBI) (<http://www.ncbi.nlm.nih.gov/omim/>) y desde enero de 2011 en el sitio web oficial, OMIM.org (<http://omim.org>), con el fin de proporcionar un portal de uso y búsqueda fácil¹⁶⁹. Cada entrada de OMIM tiene un resumen completo de un fenotipo genético y/o de un gen determinado, derivando su contenido exclusivamente de la literatura biomédica publicada y siendo actualizado cada día mediante revisión por pares¹⁷⁰. En la actualidad, con más de 450 actualizaciones cada mes, contiene información de más de 16.300 genes y casi 10.000 fenotipos diferentes. Adicionalmente, OMIM ostenta el liderazgo en el nombramiento y clasificación de fenotipos genéticos, lo cual supone una diferencia muy relevante con respecto a otras bases de datos¹⁷¹.

La base de datos HGMD está diseñada para ser una recopilación de todas las mutaciones de la línea germinal publicadas en revistas científico-médicas que afectan

a los genes del ADN nuclear asociados con la enfermedad hereditaria humana¹⁷². Se estableció en 1996 para el estudio de los mecanismos mutacionales de los genes humanos asociados a enfermedades hereditarias y proporciona información con relevancia clínica^{173,174}. En los últimos años ha adquirido una mayor utilidad como repositorio de la variación genética funcional relacionada con las enfermedades causadas por alteraciones en la línea germinal, y se considera una herramienta especialmente útil para la anotación de datos de NGS¹⁷⁵. Su versión pública (<http://www.hgmd.org>) está disponible gratuitamente para usuarios registrados de instituciones académicas y organizaciones sin ánimo de lucro y se actualiza dos veces al año, pero sus actualizaciones solo incorporan información que ya ha sido publicada un mínimo de 3 años y medio atrás. En cambio, existe una versión “Professional”, de acceso para usuarios académicos, clínicos y comerciales mediante suscripción, gestionada por QIAGEN (<https://www.qiagenbioinformatics.com/>), que se actualiza trimestralmente e incluye anotaciones y características adicionales, como las coordenadas cromosómicas genómicas en base a las versiones del genoma humano hg19 y hg38, datos de conservación evolutiva y predicciones funcionales, entre otras. En el momento de escribir esta Tesis Doctoral, este recurso, con más de 24.000 nuevas mutaciones incorporadas cada año, contenía más de 282.000 alteraciones genéticas diferentes identificadas en más de 11.000 genes.

ClinVar es un recurso público, de acceso gratuito, que se puso en marcha en 2013 por el NCBI (<https://www.ncbi.nlm.nih.gov/clinvar/>), con el fin de compilar información, basada en evidencia científico-médica, al respecto de las relaciones entre variantes genómicas y fenotipos humanos¹⁷⁶. Es una base de datos diseñada para que los laboratorios clínicos, los investigadores y los paneles de expertos puedan compartir sus interpretaciones de las variantes genéticas junto a su correspondiente evidencia, permitiendo identificar fácilmente aquellas variantes que han sido revisadas por paneles de expertos. La agregación de información procedente de diversos grupos en una sola base de datos permite la comparación de interpretaciones, proporcionando transparencia en su concordancia o discordancia. En el momento presente ClinVar contiene más de 801.000 variantes genéticas con interpretación, que afectan a más de 11.000 genes. Este gran volumen de datos la ha convertido en un valioso recurso para

la comunidad genética clínica, que valora especialmente las interpretaciones consensuadas¹⁷⁷.

Todas estas bases de datos sirven de gran ayuda, pero a menudo contienen variantes incorrectamente clasificadas y/o propuestas de clasificaciones contradictorias para una misma variante¹⁵⁹. Por este motivo, en el año 2013 el NHGRI (*National Human Genome Research Institute*) fundó ClinGen, un recurso público que tiene como objetivo principal definir la relevancia clínica de genes y variantes para su uso en medicina de precisión e investigación¹⁷⁸. Conformado por diferentes grupos de trabajo de dominio clínico y paneles de expertos que interaccionan entre sí, su objetivo es abordar el conocimiento genómico de una forma multidisciplinar, optimizando el valor de la información disponible.

8.2 Criterios ACMG/AMP de interpretación clínica de variantes genómicas

La utilidad clínica de la identificación de alteraciones genómicas en un estudio genético depende de la fiabilidad de su interpretación y clasificación. Se han realizado esfuerzos significativos para desarrollar pautas que regulen la evaluación de la evidencia disponible para clasificar las variantes genómicas implicadas en las enfermedades de base genética¹⁷⁹.

Inicialmente, cada laboratorio desarrollaba sus propios métodos de evaluación de variantes. Sin embargo, en el año 2000 el grupo de trabajo del comité de prácticas de laboratorio del Colegio Americano de Genética Médica (ACMG) consiguió sentar los cimientos de las recomendaciones para la estandarización de la interpretación clínica de variantes genómicas con la descripción de las siguientes categorías: (1) descritas previamente y consideradas como causa de enfermedad, (2) no descritas previamente como patogénicas, pero que por sus características se espera que sean responsables del fenotipo, (3) no descritas previamente, pero con dudas al respecto de su patogenicidad, (4) no descritas previamente y probablemente no causantes de enfermedad y (5) descritas previamente y consideradas neutrales. Asimismo, se diseñaron estudios de seguimiento de los pacientes portadores de determinadas variantes genéticas con el

objetivo de poder evaluar sus riesgos específicos y concretar cómo se debía llevar a cabo la validación técnica y clínica de los resultados obtenidos¹⁸⁰.

El advenimiento de la NGS promovió un cambio de paradigma en las pruebas genéticas y un aumento muy notable del número de genes y variantes asociados a enfermedades genéticas. En este contexto, en el año 2013 se formó un grupo de trabajo compuesto por miembros del ACMG, la Asociación de Patología Molecular (AMP) y el Colegio de Patólogos Americanos (CAP) con el objetivo de revisar los estándares para clasificar las variantes genéticas utilizando la evidencia disponible en el momento del análisis. Así, en 2015 se publicaron las primeras guías ACMG/AMP con un sistema de puntuación para clasificar las variantes de forma consensuada y en las cuales se recomienda el uso de la terminología “patogénica”, “probablemente patogénica”, “significado incierto”, “probablemente benigna” y “benigna” para describir y clasificar las variantes identificadas en un estudio genético. Mientras que las recomendaciones previas de ACMG habían propuesto categorías interpretativas y algoritmos para la interpretación, las recomendaciones de 2015 ya utilizan criterios específicos mucho más definidos, basados no solo en la opinión de expertos, sino también en datos empíricos y cálculos probabilísticos¹⁸¹.

Las consideraciones de estas guías son únicamente aplicables para variantes con herencia monogénica, no para variantes somáticas, farmacogenómicas o de genes asociados con trastornos complejos no mendelianos multigénicos. La publicación de estos primeros estándares y pautas ACMG/AMP para la interpretación de resultados y la clasificación de variantes fue un hito para el diagnóstico molecular y la genética clínica.

Criterios ACMG/AMP

En las especificaciones iniciales se proporcionaron dos conjuntos de criterios, unos para evaluar patogenicidad y otros para evaluar benignidad, dentro de los cuales se estratificaron los criterios en varios subgrupos según el nivel de evidencia: *stand-alone* (para criterios que se valen por sí mismos para otorgarle clasificación a la

variante), *very strong*, *strong*, *moderate* o *supporting*, en orden decreciente de nivel de evidencia. En la (**Tabla 2**) y en la (**Tabla 3**) se describe y fundamenta la relevancia de los criterios de patogenicidad y benignidad, respectivamente.

Cada criterio analiza un tipo de evidencia y los resultados obtenidos se combinan de acuerdo con unas reglas de puntuación para clasificar cada variante dentro de uno de los 5 niveles: patogénica, probablemente patogénica, significado incierto, probablemente benigna o benigna. Estas reglas se aplican a todos los datos disponibles de una variante, ya sean procedentes de casos clínicos reales, de casos de investigación o de publicaciones previas. Para proporcionar cierta flexibilidad a la valoración marcada por el nivel de evidencia, algunos criterios inicialmente propuestos con un determinado nivel de evidencia se pueden considerar de otro nivel superior utilizando el juicio profesional. Esto ocurre por ejemplo con el criterio PP1 (nivel *supporting*), que se refiere a la segregación de la variante con la enfermedad en miembros afectados de una o varias familias en un gen claramente causante de la enfermedad. Este criterio puede considerarse como *moderate* o *strong* si la evidencia de cosegregación con la enfermedad es elevada¹⁸¹. Por el contrario, en situaciones en las que los datos no son tan sólidos, se puede usar el juicio profesional para degradar el nivel del criterio al que corresponda la evidencia objeto de valoración.

Tabla 2: Criterios ACMG/AMP que indican patogenicidad.

La primera letra del nombre del criterio, **P**, indica que es un criterio de patogenicidad. Las siguientes letras indican el nivel: **VS**: very strong; **S**: strong; **M**: moderate y **P**: supporting. Los números sirven para distinguir criterios del mismo nivel.

Criterio	Descripción	Fundamento
PVS1	Variante inactivadora en un gen en el cual la pérdida de función es un mecanismo conocido de enfermedad.	Se basa en las consecuencias drásticas que tiene la inactivación completa de un gen.
PS1	Variante en la cual el cambio de aminoácido es el mismo que el de una variante previamente clasificada como patogénica, pero con otro cambio de nucleótido.	Se basa en que dos secuencias de aminoácidos idénticas, aun siendo sus secuencias de nucleótidos diferentes, se comportarán de manera idéntica o muy similar.
PS2	Variante <i>de novo</i> en un paciente sin historia familiar de la enfermedad. (<i>confirmadas maternidad y paternidad</i>).	Se basa en la infrecuencia de las variantes <i>de novo</i> , lo cual las hace especialmente sospechosas cuando aparecen en un individuo afecto sin antecedentes familiares.
PS3	Estudios funcionales <i>in vitro</i> o <i>in vivo</i> sólidamente establecidos apoyan un efecto deletéreo de la variante en el gen o el producto del gen.	Se basan en la evidencia de patogenicidad derivada del análisis funcional de la variante en un contexto experimental.
PS4	La prevalencia de la variante es significativamente mayor en afectos que en controles (<i>Odds Ratio</i> >5).	Se basa en el análisis estadístico de la asociación entre la presencia de una variante y la aparición de la enfermedad.
PM1	Variante localizada en un <i>hot-spot</i> mutacional y/o un dominio funcional sólidamente establecido sin variación benigna. Aplicable solo a variantes no sinónimas (<i>missense</i>) o que no cambian la pauta de lectura (<i>inframe</i>).	Se basa en la ubicación de la variante en regiones que, según la información disponible, parecen especialmente sensibles a las alteraciones genéticas.
PM2	Variante ausente o con una frecuencia extremadamente baja en individuos control.	Se basa en la ausencia/infrecuencia de la variante en individuos sanos, lo que la hace especialmente sospechosa si aparece en un individuo afecto.
PM3	Variante descrita en homocigosis o en heterocigosis compuesta con otra alteración patogénica en el mismo gen. Aplicable solo a enfermedades recesivas.	Se basa en la probabilidad de que la variante, bien en homocigosis o en heterocigosis compuesta con otra variante, cause un fenotipo recesivo.
PM4	La longitud de la proteína cambia como resultado de una inserción/deleción en pauta en una región no repetitiva.	Se basa en el efecto potencialmente deletéreo que tienen los cambios de longitud en una proteína.

<i>Criterio</i>	<i>Descripción</i>	<i>Fundamento</i>
<i>PM5</i>	Se ha descrito otra variante no sinónima que afecta al mismo aminoácido y que ha sido considerada como patogénica.	Se basa en la probabilidad de que una variante que afecta a una posición en la que ya se han descrito variantes patogénicas sea también patogénica.
<i>PM6</i>	Variante asumida <i>de novo</i> en un paciente sin historia familiar de la enfermedad.	Similar a PS2, pero sin confirmación de paternidad y maternidad por parte de los supuestos progenitores en los que la variante está ausente.
<i>PP1</i>	Cosegregación de la variante con la enfermedad en miembros afectados de una familia/varias familias en un gen claramente causante de la enfermedad.	Se basa en la probabilidad de que la concurrencia de una variante con el fenotipo en diversos miembros de una familia sea indicativa de patogenicidad.
<i>PP2</i>	Variante no sinónima en un gen con baja tasa de variación no sinónima benigna y en el cual las variantes no sinónimas son un mecanismo común de la enfermedad.	Se basa en la probabilidad de que una variante no sinónima cualquiera pueda desencadenar una enfermedad cuando el gen al que afecta no tiene muchas variantes no sinónimas benignas.
<i>PP3</i>	La evaluación bioinformática de la variante predice un efecto funcional deletéreo.	Se basa en la capacidad de los algoritmos informáticos para predecir la patogenicidad de una variante.
<i>PP4</i>	El fenotipo del paciente se asocia de manera altamente específica con el gen.	Se basa en la probabilidad de que, al ser el gen afectado por la variante el único o uno de los únicos asociados con la enfermedad, la variante sea patogénica.
<i>PP5</i>	Una fuente reputada (LOVD/INSIGHT) ha reportado recientemente la variante como patogénica, pero la evidencia no está disponible para hacer una evaluación independiente.	Se basa en la clasificación hecha por terceros que ofrezcan fiabilidad.

Tabla 3: Criterios ACMG/AMP que indican benignidad.

La primera letra del nombre del criterio, **B**, indica que es un criterio de benignidad. Las siguientes letras indican el nivel: **A**: stand alone; **S**: strong; **M**: moderate y **P**: supporting. Los números sirven para distinguir criterios del mismo nivel.

Criterio	Descripción	Fundamento
BA1	Variante que puede ser considerada benigna por su alta frecuencia alélica.	Se basa en improbabilidad de que las variantes frecuentes en la población causen enfermedades monogénicas.
BS1	La frecuencia alélica de la variante es superior a la esperada para la enfermedad.	Similar a BA1, pero para variantes no tan frecuentes.
BS2	Variante observada en un individuo adulto sano para una enfermedad con penetrancia completa esperada a temprana edad.	Se basa en la presencia de la variante en individuos sanos como argumento a favor de su benignidad.
BS3	Estudios funcionales <i>in vitro</i> o <i>in vivo</i> sólidamente establecidos muestran un efecto no patogénico de la variante en la función de la proteína o en el <i>splicing</i> .	De modo análogo a PS3, se basa en evidencia de neutralidad derivada del análisis funcional de la variante en un contexto experimental.
BS4	Ausencia de cosegregación en miembros afectados de la familia.	Se basa en la improbabilidad de que dos individuos afectados de la misma familia tengan causas genéticas diferentes.
BP1	Variante no sinónima en un gen para el cual las variantes causantes de enfermedad conocidas son principalmente variantes truncantes.	Se basa en la improbabilidad de que una variante no truncante cause la enfermedad cuando las causas descritas provocan la inactivación del gen.
BP2	Variante observada en <i>trans</i> con una variante patogénica en un gen/enfermedad dominante con penetrancia completa u observada en <i>cis</i> con una variante patogénica para cualquier modo de herencia.	Se basa en la improbabilidad de que un mismo alelo tenga dos variantes patogénicas o de que un mismo gen con fenotipos dominantes tenga variantes patogénicas en ambos alelos.
BP3	Inserción/delección en pauta en una región repetitiva sin función conocida.	Se basa en la improbabilidad de que una inserción/delección que no modifica la pauta de lectura y afecta a una región sin función conocida cause la enfermedad.
BP4	La evaluación bioinformática de la variante no predice un efecto funcional deletéreo.	Se basa en la capacidad de los algoritmos informáticos para predecir la benignidad de una variante.
BP5	Variante identificada en un caso con una causa alternativa de enfermedad.	Se basa en la improbabilidad de que la enfermedad de un individuo tenga dos causas genéticas independientes.

Criterio	Descripción	Fundamento
BP6	Una fuente reputada ha reportado recientemente la variante como benigna, pero la evidencia no está disponible para hacer una evaluación independiente.	Se basa en la clasificación hecha por terceros que ofrezcan fiabilidad.
BP7	Variante sinónima para la que los algoritmos predictores de efecto funcional consideran que la variante no tiene efecto en el <i>splicing</i> .	Se basa en la improbabilidad de que una variante sinónima sin otros efectos conocidos sobre el gen que el cambio de su secuencia cause la enfermedad.

A pesar de estos avances, entre otros autores, Amendola y col., en 2016, tras analizar las clasificaciones hechas por diversos laboratorios al aplicar estos mismos criterios, describieron importantes inconsistencias y un sesgo hacia la sobreestimación de la patogenicidad¹⁸². En consecuencia, en los últimos años, el Grupo de Trabajo de Interpretación de Variantes de Secuencia (*Sequence Variant Interpretation, SVI*) de ClinGen se ha implicado en el refinamiento y la evolución de las directrices de la interpretación ACMG / AMP para desarrollar enfoques cuantitativos para la interpretación de variantes. Así, recientemente se han publicado recomendaciones específicas para la interpretación de las variantes de pérdida de función (criterio PVS1)¹⁸³, para la aplicación de la evidencia funcional (criterios PS3/BS3)¹⁸⁴, especificaciones para las variantes *de novo* (criterios PS2, PM6), recomendaciones para interpretar variantes que afectan a los dos alelos del mismo gen, solo aplicables a enfermedades recesivas (criterio PM3) y la eliminación de los criterios que se basan en afirmaciones que no están directamente vinculadas a evidencia científico-médica, como ocurre con los criterios PP6 y BP5¹⁸⁵. Asimismo, el Grupo de trabajo SVI está encargado de armonizar las re-especificaciones adaptadas a cada patología que marcan los paneles de expertos^{186,187}.

Como se ha descrito en esta introducción, paulatinamente, los avances técnicos, informáticos y biomédicos han puesto a nuestro alcance la realización de análisis genómicos cada vez más fiables e informativos. A lo largo de los resultados y la discusión de la presente Tesis Doctoral se ilustrará la utilidad clínica de dichos análisis

genómicos cuando se aplican a pacientes de tres patologías con un fuerte componente genético: el cáncer, las sorderas y las cegueras.

OBJETIVOS

La secuenciación masiva permite analizar ácidos nucleicos con alto rendimiento y a costes compatibles con su aplicación clínica. Estas ventajas la hacen especialmente útil para el diagnóstico de enfermedades de base genética. La utilidad de las plataformas diagnósticas basadas en NGS dependerá de su capacidad para mejorar y personalizar el manejo clínico de cada paciente.

En esta Tesis Doctoral abordamos 3 grupos de enfermedades con base genética: el cáncer, tanto el causado exclusivamente por alteraciones somáticas como aquel surgido como consecuencia de la existencia de mutaciones germinales, la hipoacusia neurosensorial, producida principalmente por lesiones que afectan al oído interno y/o a las fibras del nervio auditivo, y las cegueras que cursan con alteraciones de la retina, el vítreo, la coroides y/o el nervio óptico. En este contexto se plantearon los siguientes objetivos:

1. Desarrollo, optimización, validación y evaluación de la utilidad clínica de una plataforma de diagnóstico molecular del cáncer basada en NGS, diseñada para identificar tanto las alteraciones somáticas relacionadas con la sensibilidad o con la resistencia a terapias dirigidas aprobadas, como las mutaciones de la línea germinal que predisponen a cáncer familiar.
2. Desarrollo, optimización, validación y evaluación de la utilidad clínica de una plataforma de diagnóstico molecular de la hipoacusia hereditaria basada en NGS, capaz de detectar alteraciones genómicas en todos los genes cuyas mutaciones se han asociado con hipoacusia neurosensorial y/o mixta, sindrómica o no sindrómica.
3. Desarrollo, optimización, validación y evaluación de la utilidad clínica de una plataforma de diagnóstico molecular de la ceguera hereditaria basada en NGS, diseñada para detectar alteraciones genómicas en todos los genes cuyas mutaciones se han asociado a ceguera causada por alteraciones en la retina, vítreo, coroides y nervio óptico, tanto sindrómica como no sindrómica.

RESULTADOS

Artículo I: A novel molecular diagnostics platform for somatic and germline precision oncology

Referencia:

Rubén Cabanillas*, Marta Diñeiro*, David Castillo, Patricia C. Pruneda, Cristina Penas, Guadalupe A. Cifuentes, Álvaro de Vicente, Noelia S. Durán, Rebeca Álvarez, Gonzalo R. Ordóñez y Juan Cadiñanos. **A novel molecular diagnostics platform for somatic and germline precision oncology**. *Molecular Genetics & Genomic Medicine*. 2017; 5(4): 336–359. Doi: 10.1002/mgg3.291.

*R. C. y M. D. contribuyeron por igual en este trabajo

Factor de Impacto:

2,695, Q2 Genetics & Heredity (JCR 2017)

Resumen:

La secuenciación de nueva generación (NGS) abre nuevas opciones en la oncología clínica, desde la selección de terapias hasta el asesoramiento genético. Hemos desarrollado una nueva plataforma de diagnóstico NGS destinada a detectar (1) alteraciones genómicas somáticas asociadas con la respuesta a terapias dirigidas aprobadas contra el cáncer y (2) mutaciones de la línea germinal que predisponen a neoplasias hereditarias.

Se prepararon librerías de NGS enriquecidas mediante captura de los exones de 215 genes asociados a cáncer (97 para la selección de terapias dirigidas y 148 asociados con predisposición hereditaria al cáncer, siendo 30 de ellos informativos para ambas aplicaciones), así como en intrones seleccionados de 17 genes involucrados en reordenamientos asociados con una terapia dirigida, a partir de 39 tumores (tejidos o citologías incluidas en parafina), 36 muestras germinales (sangre) y 10 líneas celulares. El análisis de los resultados de NGS se realizó utilizando aplicaciones bioinformáticas desarrolladas específicamente.

La plataforma detectó variantes de un solo nucleótido (SNVs) e inserciones/deleciones (indels) con sensibilidad y especificidad > 99,5 % (para variantes con frecuencia alélica $\geq 0,1$), así como variantes de número de copias (CNVs) y reordenamientos. Los estudios somáticos identificaron terapias dirigidas en 35/39 tumores (89,74 %), un rendimiento diagnóstico comparable al de las principales plataformas comerciales. La detección de la mutación somática *EGFR* p.E746_S752delins en una metástasis mediastínica de un cáncer de mama provocó su reevaluación anatomopatológica, su reclasificación definitiva como cáncer de pulmón y su tratamiento con gefitinib (respuesta parcial sostenida durante 15 meses). Los estudios en las 36 muestras germinales identificaron dos mutaciones patogénicas (en *CDKN2A* y *BRCA2*). Proponemos una estrategia para la interpretación y descripción de resultados de análisis somáticos y germinales

adaptable al objetivo de la solicitud, a la disponibilidad de muestras tumorales y/o normales y al alcance del consentimiento informado.

Con una metodología adecuada, es posible trasladar a la práctica clínica los últimos avances en oncología de precisión, integrando bajo la misma plataforma la identificación de alteraciones genómicas somáticas y de la línea germinal.

Aportación personal al trabajo:

La realización de este trabajo coincidió con mi primer año en el laboratorio. En él aprendí las técnicas empleadas para la preparación de librerías y me familiaricé con la secuenciación de nueva generación. Asimismo, colaboré en el diseño del panel e hice gran parte del trabajo de interpretación de los resultados.

Molecular Genetics & Genomic Medicine

Open Access

ORIGINAL ARTICLE

A novel molecular diagnostics platform for somatic and germline precision oncology

Rubén Cabanillas^{1,*}, Marta Diñeiro^{1,*}, David Castillo², Patricia C. Pruneda², Cristina Penas¹,
 Guadalupe A. Cifuentes¹, Álvaro de Vicente¹, Noelia S. Durán¹, Rebeca Álvarez¹,
 Gonzalo R. Ordóñez² & Juan Cadiñanos¹ 

¹Instituto de Medicina Oncológica y Molecular de Asturias (IMOMA) S. A., Avda. Richard Grandío s/n, 33193 Oviedo, Spain

²Disease Research And Medicine (DREAMgenics) S. L., Vivero Empresarial de Ciencias de la Salud, C/Colegio Santo Domingo de Guzmán s/n, 33011 Oviedo, Spain

Keywords

Cancer, counseling, diagnostics, NGS, targeted-therapy

Correspondence

Juan Cadiñanos, Laboratorio de Medicina Molecular, IMOMA, S. A., Avda. Richard Grandío s/n, 33193, Oviedo, Spain.

Tel: +34985250300 Ext. 266;

Fax: +34985963710;

E-mail: jcb@imoma.es

and

Gonzalo R. Ordóñez, DREAMgenics, S. L., Vivero Empresarial de Ciencias de la Salud, C/Colegio Santo Domingo de Guzmán s/n, 33011, Oviedo, Spain. Tel: +34985088180; E-mail: gonzalo.ordonez@dreamgenics.com

Funding Information

Fundación para el Fomento en Asturias de la Investigación Científica Aplicada y la Tecnología (Grant/Award Number: '1E13-121'), Universidad de Oviedo, Instituto de Salud Carlos III (Grant/Award Number: 'PT13/0001'), Fundación María Cristina Masaveu Peterson.

Received: 15 November 2016; Revised: 13 March 2017; Accepted: 15 March 2017

Molecular Genetics & Genomic Medicine 2017; 5(4): 336–359

doi: 10.1002/mgg3.2291

*Equal contribution.

Abstract

Background

Next-generation sequencing (NGS) opens new options in clinical oncology, from therapy selection to genetic counseling. However, realization of this potential not only requires succeeding in the bioinformatics and interpretation of the results, but also in their integration into the clinical practice. We have developed a novel NGS diagnostic platform aimed at detecting (1) somatic genomic alterations associated with the response to approved targeted cancer therapies and (2) germline mutations predisposing to hereditary malignancies.

Methods

Next-generation sequencing libraries enriched in the exons of 215 cancer genes (97 for therapy selection and 148 for predisposition, with 30 informative for both applications), as well as selected introns from 17 genes involved in drug-related rearrangements, were prepared from 39 tumors (paraffin-embedded tissues/cytologies), 36 germline samples (blood) and 10 cell lines using hybrid capture. Analysis of NGS results was performed with specifically developed bioinformatics pipelines.

Results

The platform detects single-nucleotide variants (SNVs) and insertions/deletions (indels) with sensitivity and specificity >99.5% (allelic frequency ≥ 0.1), as well as copy-number variants (CNVs) and rearrangements. Somatic testing identified tailored approved targeted drugs in 35/39 tumors (89.74%), showing a diagnostic yield comparable to that of leading commercial platforms. A somatic *EGFR* p.E746_S752delinsA mutation in a mediastinal metastasis from a breast cancer prompted its anatomopathologic reassessment, its definite reclassification as a lung cancer and its treatment with gefitinib (partial response sustained for 15 months). Testing of 36 germline samples identified two pathogenic mutations (in *CDKN2A* and *BRCA2*). We propose a strategy for interpretation and reporting of results adaptable to the aim of the request, the availability of tumor and/or normal samples and the scope of the informed consent.

Conclusion

With an adequate methodology, it is possible to translate to the clinical practice the latest advances in precision oncology, integrating under the same platform the identification of somatic and germline genomic alterations.

Introduction

Precision oncology is at the frontline of personalized medicine. The use of a patient's molecular data to inform diagnosis, prognosis, treatment and prevention of cancer for that very patient is a paradigm changing approach. To drive the adoption of personalized cancer medicine, the integration of validated next-generation sequencing (NGS) diagnostics platforms into clinical practice is essential.

Targeted therapies aimed at tackling specific genomic alterations, such as the tyrosine kinase inhibitor (TKI) imatinib or the monoclonal antibody trastuzumab, were initially approved exclusively for the treatment of one type of cancer (chronic myelogenous leukemia with the *BCR-ABL* (MIM: 151410 and MIM: 200100, respectively) translocation and breast cancer with *ERBB2* (MIM: 164870) amplification, respectively) but they were also later shown to be effective on other tumor types with identical or similar alterations (gastrointestinal stromal tumors with *KIT* (MIM: 164920) or *PDGFRA* (MIM: 173490) mutations and gastric cancer with *ERBB2* amplification, respectively). This supports the application of a genomic-driven transversal view of cancer therapy. The encouraging results of molecularly driven clinical trials, such as the so called "basket trials", and case reports of exceptional responders to targeted therapies selected on the basis of somatic mutations, irrespective of primary tumor site or histology, are proving the utility of this transversal approach (Wagle *et al.* 2014; Hyman *et al.* 2015; Redig and Janne 2015).

For some cancer types, such as Non-Small Cell Lung Cancer (NSCLC), several targeted drugs acting on different mutant proteins are approved by the FDA/EMA (TKIs acting on sensitive mutant EGFR proteins or ALK inhibitors directed against the products of *ALK* (MIM: 105590) translocations) and exploring additional therapeutic options with medicines approved for different indications is recommended by the NCCN guidelines (crizotinib or cabozantinib/vandetanib for NSCLC with *ROS1* (MIM: 165020) or *RET* (MIM:164761) translocations, respectively). In such cases, probing all cancer treatment options for a given patient with conventional, "first generation" genotyping approaches (PCR, Sanger sequencing, FISH, etc.) requires analyzing multiple samples from limited specimens, with different techniques and, in some cases, by different laboratories. This approach, which leads to rising costs and increasing turnaround times, is not suitable for the current scenario of ever-growing targeted treatment opportunities (Iorio *et al.* 2016). However, NGS technologies, which allow analysis of multiple cancer biomarkers, including SNVs, indels, translocations and CNVs, on the same sample in a single assay, come as an optimal solution for this situation.

Although most actionable genomic alterations are acquired somatically, in some instances germline mutations are responsible for an inheritable increased risk for cancer development. Being able to identify these alterations is of utmost importance for providing patients with a comprehensive personalized clinical management. First, some germline alterations (such as those inactivating *BRCA1/BRCA2* (MIM: 113705 and MIM: 600185, respectively) or activating *RET*) are also linked to the response to targeted drugs. Moreover, when germline cancer predisposing mutations are identified, this facilitates prevention and early detection of future tumors both in the probands and in their families. Furthermore, analyzing matched germline DNA in parallel to the tumor DNA allows increased performance of the bioinformatics processing of NGS results. Because managing the request of a germline analysis and its clinical application may not be straightforward for oncology practitioners (as it involves specific genetic counseling, complex informed consent and, frequently, dealing with variants of uncertain clinical significance), the implementation of mechanisms to overcome these difficulties is especially welcome.

In this work, we present the development, validation and clinical application of a novel molecular diagnostic platform based on targeted NGS, specifically designed to comprehensively identify somatic alterations linked to sensitivity or resistance to approved cancer therapies and germline mutations predisposing to familial cancer. Moreover, we propose a results-processing pipeline and an interpretation algorithm designed to deal with all different request situations, from exclusively somatic testing in the absence of germline DNA to combined somatic and germline analyses.

Materials and Methods

Ethical compliance

This study was approved by the competent ethics committee (Comité Ético de Investigación Clínica Regional del Principado de Asturias).

Patients

The study was performed on a consecutive series of unselected patients attending our clinic between December 2013 and February 2017. Patients involved in the somatic analysis were stage IV cancer patients who had progressed to first-line chemotherapy. Patients involved in the germline analysis were all eligible for cancer genetic counseling based on their personal and/or familial history of cancer.

Configuration of gene panels

Somatic subpanel

Identification of cancer genes with clinically relevant somatic mutations was based on searches in the COSMIC database (Forbes et al. 2015) (<http://cancer.sanger.ac.uk/>), mycancergenome.com (<http://www.mycancergenome.com>), targetedcancercare.com (<http://www.targetedcancercare.com>), Google Scholar and PubMed. As a result of those searches, an important source of information for building our somatic database was a comprehensive publication by Dienstmann and coworkers (Dienstmann et al. 2014). Next, the preliminary gene list was further filtered considering information contained in the FDA (www.fda.gov) and EMA (www.ema.europa.eu/ema/) web-pages to identify those genes therapeutically relevant whose genomic alterations have been related to approved cancer therapies.

Germline subpanel

An initial list of cancer genes with germline alterations associated with hereditary cancer predisposition was generated through automatic searches on the HGMD database. Then, manual review of the preliminary gene list was performed based on OMIM (www.omim.org), Genetics Home Reference (ghr.nlm.nih.gov), Orphanet (www.orphanet) and other sources with the aim of filtering out of the list those genes fulfilling any of the following criteria: (1) insufficient scientific evidence or HGMD annotation error; (2) mutations assigned to the gene in HGMD correspond to a genomic region not functionally characterized and/or (3) cancer-associated alterations that may not be detected by the proposed methodology (uncharacterized gene fusions, inversions, translocations or epigenetic changes).

GenBank accession numbers

For all genes, variants were annotated according to all the GenBank RefSeq transcripts used by HGMD for that gene (HGMD transcripts). Additionally, for those genes of the somatic subpanel, variants were also annotated according to the most priority APPRIS transcript listed by Ensembl, unless this coincided with an HGMD transcript. The RefSeq transcript GenBank Accession Numbers were: *ABL1* NM_007313.2; *AKT1* NM_005163.2; *ALK* NM_004304.4; *APC* NM_000038.5; *AR* NM_000044.3; *ARA* NM_001654.4, NM_001256196.1; *ARID1A* NM_006015.4; *ATM* NM_000051.3; *ATP2A2* NM_001681.3, NM_170665.3; *ATR* NM_001184.3; *AXINI* NM_003502.3; *AXIN2* NM_004655.3; *BAP1* NM_004656.3; *BARD1* NM_000465.2; *BCR* NM_004327.3; *BLM* NM_000057.2;

BMPRI1A NM_004329.2; *BRAF* NM_004333.4; *BRCA1* NM_007294.3, NM_007300.3; *BRCA2* NM_000059.3; *BRIP1* NM_032043.2; *BTK* NM_000061.2; *BUB1B* NM_001211.5; *CBL* NM_005188.3; *CCND1* NM_053056.2; *CCND2* NM_001759.3; *CCND3* NM_001760.3; *CD274* NM_014143.3; *CDC73* NM_024529.4; *CDH1* NM_004360.3; *CDK4* NM_000075.3; *CDK6* NM_001259.6; *CDKN1A* NM_078467.2; *CDKN1B* NM_004064.3; *CDKN1C* NM_000076.2; *CDKN2A* NM_000077.4, NM_058197.4, NM_058195.3; *CDKN2B* NM_004936.3; *CDKN2C* NM_001262.2; *CEBPA* NM_004364.3; *CHEK2* NM_007194.3; *CRKL* NM_005207.3; *CSF1R* NM_005211.3; *CSF3R* NM_000760.3; *CTNNA1* NM_001903.2; *CTNNB1* NM_001904.3; *CTR9* NM_014633.3; *CTRC* NM_007272.2; *CYLD* NM_015247.2; *DAPK1* NM_004938.2; *DDB2* NM_000107.2; *DDR2* NM_006182.2; *DICER1* NM_177438.2; *DIS3L2* NM_152383.4; *DKC1* NM_001363.3; *DLEU7* NM_198989.2; *DNAAF1* NM_178452.4; *EGFR* NM_005228.3, NM_001346941.1 (*EGFRvIII*); *EGLN1* NM_022051.2; *EGLN2* NM_080732.3; *ENG* NM_000118.3; *EPAS1* NM_001430.4; *EPCAM* NM_002354.2; *EPHA2* NM_004431.3; *EPHB2* NM_017449.3; *EPHX1* NM_000120.3; *ERBB2* NM_004448.2; *ERBB3* NM_001982.3; *ERBB4* NM_005235.2; *ERCC1* NM_202001.2; *ERCC2* NM_000400.3; *ERCC3* NM_000122.1; *ERCC4* NM_005236.2; *ERCC5* NM_000123.3; *EXT1* NM_000127.2; *EXT2* NM_207122.1; *FAH* NM_000137.2; *FAM175A* NM_139076.2; *FANI* NM_014967.4; *FANCA* NM_000135.2; *FANCB* NM_001018113.1; *FANCC* NM_000136.2; *FANCD2* NM_033084.3; *FANCE* NM_021922.2; *FANCF* NM_022725.3; *FANCG* NM_004629.1; *FANCI* NM_001113378.1; *FANCL* NM_018062.3; *FANCM* NM_020937.2; *FAS* NM_000043.4; *FASLG* NM_000639.1; *FBXW7* NM_033632.3; *FGFR1* NM_023110.2; *FGFR2* NM_000141.4, NM_022970.3; *FGFR3* NM_000142.4; *FGFR4* NM_002011.3; *FH* NM_000143.3; *FLCN* NM_144997.5, NM_144606.5; *FLT3* NM_004119.2; *FRS2* NM_006654.4; *GALNT12* NM_024642.4; *GATA2* NM_032638.4; *GATA3* NM_001002295.1, NM_002051.2; *GNAI1* NM_002067.2; *GNAQ* NM_002072.3; *GNAS* NM_000516.4, NM_016592.2, NM_080425.2; *GPC3* NM_004484.3; *GREM1* NM_013372.6; *H19* NR_002196.1; *HDAC2* NM_001527.3; *HFE* NM_000410.3; *HGF* NM_000601.4; *HNF1A* NM_000545.5; *HNF1B* NM_000458.2; *HOXB13* NM_006361.5; *HOXD4* NM_014621.2; *HRAS* NM_005343.2; *IDH1* NM_005896.2; *IDH2* NM_002168.2; *IGF1R* NM_000875.3; *IGF2* NM_000612.4, NM_001007139.4, NM_001127598.1; *IL10RB* NM_000628.4; *IL7R* NM_002185.3; *INHBA* NM_002192.2; *INPP4B* NM_003866.2; *ITK* NM_005546.3; *JAK1* NM_002227.2; *JAK2* NM_004972.3; *JAK3* NM_000215.3; *KCNQ1OT1* NR_002728.3; *KDR* NM_002253.2; *KHDC3L* NM_001017361.2; *KIF1B* NM_015074.3, NM_183416.3; *KIT*

NM_000222.2; *KMT2A* NM_001197104.1; *KMT2C* NM_170606.2; *KRAS* NM_004985.3, NM_033360.2; *LZTR1* NM_006767.3; *MAP2K1* NM_002755.3; *MAP2K2* NM_030662.3; *MAPK1* NM_002745.4; *MAX* NM_002382.4; *MC1R* NM_002386.3; *MDH2* NM_005918.2; *MEN1* NM_130799.2; *MET* NM_001127500.1; *MITF* NM_000248.3, NM_006722.2; *MLH1* NM_000249.3; *MNX1* NM_005515.3; *MPL* NM_005373.2; *MRE11A* NM_005591.3; *MSH2* NM_000251.2; *MSH6* NM_000179.2; *MSR1* NM_138715.2; *MTAP* NM_002451.3; *MTOR* NM_004958.3; *MUTYH* NM_001128425.1; *MYD88* NM_002468.4; *NBN* NM_002485.4; *NF1* NM_000267.3, NM_001042492.2; *NF2* NM_000268.3; *NFKBIZ* NM_031419.3; *NHP2* NM_017838.3; *NLRP7* NM_206828.3, NM_139176.3, NM_001127255.1; *NOP10* NM_018648.3; *NOTCH1* NM_017617.3; *NOTCH2* NM_024408.3; *NRAS* NM_002524.4; *NRG1* NM_013956.3, NM_013964.3, NM_013962.2, NM_013959.3; *NSD1* NM_022455.4; *NTHL1* NM_002528.5; *NTRK1* NM_001012331.1, NM_002529.3; *NTRK3* NM_001012338.2, NM_002530.3; *OGG1* NM_002542.5; *PALB2* NM_024675.3; *PAX5* NM_016734.2; *PBRM1* NM_018313.4; *PDE11A* NM_016953.3; *PDGFB* NM_002608.2; *PDGFRA* NM_006206.4; *PDPK1* NM_002613.4; *PHOX2B* NM_003924.3; *PIF1* NM_025049.2; *PIK3CA* NM_006218.2; *PIK3CB* NM_006219.2; *PIK3R1* NM_181523.2; *PIK3R2* NM_005027.3; *PLCG2* NM_002661.3; *PMS1* NM_000534.4; *PMS2* NM_000535.5; *POLD1* NM_002691.3; *POLE* NM_006231.2; *POLH* NM_006502.2; *POT1* NM_015450.2; *PPARG* NM_015869.4; *PRF1* NM_001083116.1; *PRKARIA* NM_002734.4; *PRKCD* NM_006254.3; *PRKCH* NM_006255.3; *PRSS1* NM_002769.4; *PSMC3IP* NM_016556.3; *PTCH1* NM_000264.3, NM_001083602.1; *PTCH2* NM_003738.4; *PTEN* NM_000314.4; *PTPN11* NM_002834.3; *PTPRD* NM_002839.3; *RAC1* NM_018890.3, NM_006908.4; *RAD51C* NM_058216.2; *RAD51D* NM_002878.3; *RAF1* NM_002880.3; *RBI* NM_000321.2; *RECQL* NM_002907.3; *RECQL4* NM_004260.3; *RET* NM_020975.4; *RHBDF2* NM_024599.5; *RICTOR* NM_152756.3; *RIT1* NM_006912.5; *RNASEL* NM_021133.3; *ROS1* NM_002944.2; *RPL11* NM_000975.3; *RPL35A* NM_000996.2; *RPL5* NM_000969.3; *RPS19* NM_001022.3; *RPS24* NM_033022.3, NM_001142285.1; *RPS7* NM_001011.3; *RTEL1* NM_032957.4; *RUNX1* NM_001754.4; *SASH1* NM_015278.3; *SBDS* NM_016038.2; *SDHA* NM_004168.2; *SDHAF2* NM_017841.2; *SDHB* NM_003000.2; *SDHC* NM_003001.3, NM_001035511.1; *SDHD* NM_003002.3; *SEC23B* NM_006363.4; *SEMA4A* NM_022367.3; *SERPINA1* NM_000295.4; *SFXN4* NM_213649.1; *SH2B3* NM_005475.2; *SH2D1A* NM_002351.4; *SHOC2* NM_007373.3; *SLBP* NM_006527.2; *SLC25A13* NM_014251.2; *SLX4* NM_032444.2; *SMAD4* NM_005359.5; *SMAD9* NM_001127217.2; *SMARCA4* NM_001128849.1; *SMARCB1*

NM_003073.3; *SMO* NM_005631.4; *SOC1* NM_003745.1; *SOS1* NM_005633.3; *SOS2* NM_006939.2; *SPINK1* NM_003122.3; *SRC* NM_005417.4; *SRP72* NM_006947.3; *SRY* NM_003140.2; *STAG2* NM_001042749.1; *STIM1* NM_003156.3; *STK11* NM_000455.4; *SUFU* NM_016169.3; *TERC* NR_001566.1; *TERT* NM_198253.2; *TFE3* NM_006521.4; *TINF2* NM_001099274.1; *TMEM127* NM_017849.3; *TMPRSS2* NM_001135099.1, NM_005656.3; *TP53* NM_000546.5; *TRIM37* NM_015294.3; *TSC1* NM_000368.4; *TSC2* NM_000548.3; *UBE2T* NM_014176.3; *UNC13D* NM_199242.2; *UNC5C* NM_003728.3; *VEGFA* NM_001025366.2; *VHL* NM_000551.3; *WNT10A* NM_025216.2; *WRAP53* NM_018081.2; *WRN* NM_000553.4; *WT1* NM_024426.4; *XIAP* NM_001167.3; *XPA* NM_000380.3; *XPC* NM_004628.4; *XRCC2* NM_005431.1; *XRCC4* NM_022406.2; *YAPI* NM_001130145.2.

Iterative optimization of probe design

Ensembl Biomart was used to extract the genomic coordinates of all exons and selected introns from the different transcripts of each gene contained in any of the panels. These coordinates were merged to obtain nonredundant unique regions and those noncoding untranslated regions (UTRs) with clinical relevance were added. After this, SureDesign (Agilent, Santa Clara, CA, USA), was used to customize SureSelect enrichment probes with the following parameters: tiling density 5×, no masking and balanced boosting. Based on previously available exome sequencing data, low coverage regions contained within the genes of interest were identified and specific additional probes were designed.

For probe optimization, we calculated callabilities (percentage of bases covered with a given base quality, mapping quality and read depth) for each region after sequencing of the first 16 samples using version I (v1) of the probe design. For those regions showing callabilities below 100% at a 100× depth in a MiSeq run, additional probes were designed and the concentration of the existing ones was increased. Briefly, in the second version of the design (v2) we reduced the probe density from 5× to 3× for 2672 regions with an excess of sequencing depth based on v1 results and increased the probe density from 5× to 10× on 478 regions with poor callability, keeping the total number of probes in the same range as in v1 (within the limit of SureDesign Tier 2: 57.500 total probes). A similar approach was followed up for the design of successive versions of the probe. Furthermore, regions whose low callability was due to low mapping qualities caused by the presence of homologous regions in the reference genome were identified. All calculations for which no probe design version is specified were performed using v4.

DNA isolation and quantification

Tumor tissues/cells

Macro/microdissection of each sample was performed to increase nucleated tumor cell content to $\geq 20\%$, and $0.025\text{--}0.1\text{ mm}^3$ of tumor-enriched FFPE tissue were deparaffinized and pretreated with 1M NaSCN for 6–12 h to reverse formaldehyde-induced crosslinks (Hosein et al. 2013). After this, DNA extraction was performed using the QIAamp DNA micro kit (Qiagen, Valencia, CA, USA). *Blood*—2 mL of total peripheral blood was processed with the Flexigene DNA kit (Qiagen). *Saliva*—1 mL of saliva preserved in DanaSaliva collection kits was processed using the Danagene saliva kit (Danagen-BioTed, Badalona, Spain). All isolated DNA samples were quantified by both spectrophotometry, using NanoDrop (Thermo-Fisher Scientific, Waltham, MA, USA) (N: Nanodrop-concentration), and fluorimetry, using the Qubit[®] dsDNA HS and/or BR assay kits (Thermo-Fisher Scientific) (Q: Qubit-concentration).

Library construction and capture

50–200 ng of dsDNA (according to Qubit fluorimetric quantification) were processed following the protocol recommended by the manufacturer of the capture probes: “Low Input Sureselect^{XT} Target Enrichment System for Illumina Paired-End Sequencing Library (Agilent).”

Sequencing

Indexed libraries were sequenced using Illumina MiSeq, NextSeq500 or HiSeq1500 next-generation sequencers following the equipment's instructions. Paired-end 75–100 bp reads were obtained. MiSeq and NextSeq500 sequencing was performed at IMEGEN (Valencia, Spain), whereas HiSeq1500 sequencing was done by HealthIn-Code (A Coruña, Spain).

Processing of sequencing data and variant calling

Raw FASTQ files were first evaluated using quality control checks from FastQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>), Trimmomatic was then employed for read trimming and filtering. After removing low quality bases, adapters and other technical sequences, each library was aligned to the human reference genome (GRCh37) using BWA-mem (Li and Durbin 2010) generating sorted BAM files with SAMtools (Li et al. 2009). Reads from the same libraries were then merged, removing optical and PCR duplicates using Picard (<http://broadinstitute.github.io/picard/>). High sequencing coverage depths

allowed maintenance of sufficient independent reads at each target position after this step. Those regions with < 20 reads of base quality ≥ 10 and mapping quality ≥ 20 were considered as target-sequence gaps.

For the identification of germline point mutations, VarScan2 analysis was performed (Koboldt et al. 2012) (filtering parameters: total read depth ≥ 6 , mutated allele count ≥ 3 , variant frequency ≥ 0.1 , base quality ≥ 20 , mapping quality ≥ 30). For the identification of somatic substitutions and indels, a variation of Sidrón algorithm which has previously been described (Puente et al. 2011) was used (filtering parameters: total read depth ≥ 6 , mutated allele count ≥ 3 , variant frequency ≥ 0.025 , base quality ≥ 20 , mapping quality ≥ 30). In the latter, mutations detected in the tumor sample are interrogated in the normal sample in order to define the somatic status of each variant. In both cases the workflow performs a subsequent analysis to filter out recurrent sequencing errors, which is based on documented sequencing biases associated to NGS platforms (Kwon et al. 2013; Wall et al. 2014) as well as on the identification of specific sequencing errors found to be artifacts in previously evaluated samples. Moreover, visual inspection of the alignment is also performed as a final step for all reported variants.

The process of variant calling includes an extra evaluation layer for variants present in provided data-sources (COSMIC for somatic and HGMD for germline variants), regardless of the filtering thresholds (including the 0.025 minimum variant frequency otherwise required). Variants called according to these criteria are still evaluated by the filtering thresholds, but instead of being discarded when quality measures are not met, the variant is registered in a different output, facilitating further inspection. The objective of this added flexibility is the identification of relevant mutations that would otherwise be missed when cutting thresholds are not met. We have evidenced the benefits of this procedure by correctly identifying variants in the 2016 EMQN External Quality Assessment scheme that presented a variant frequency below 0.025.

Variants were annotated using several databases containing functional (Ensembl, CCDS, RefSeq, Pfam), populational (dbSNP, 1000 Genomes, ESP, ExAC) and disease-related (COSMIC, ICGC, HGMD professional) information, as well as 11 scores from algorithms for prediction of the impact caused by nonsynonymous variants on the structure and function of the protein (SIFT (Kumar et al. 2009), PolyPhen2 (Adzhubei et al. 2010), PROVEAN (Choi et al. 2012), Mutation Assessor (Reva et al. 2011), Mutation Taster (Schwarz et al. 2014), LRT (Chun and Fay 2009), MetaLR, MetaSVM (Dong et al. 2015), FATHMM and FATHMM-MKL (Shihab et al. 2014)), and I score (GERP++) for evolutionary conservation of the affected nucleotide (Davydov et al. 2010).

The detection of CNVs was performed by a modified version of the exome2cnv algorithm (Valdes-Mas *et al.* 2012) incorporating a combination of read depth and allelic imbalance computations for copy number assessment. For individual samples the algorithm employs a background of pooled samples processed using the same capturing protocol and sequencing technology. When paired tumor-normal samples are available, a parallel analysis is performed to reevaluate copy numbers and determine the somatic status of variants.

A customized algorithm was developed to identify intrachromosomal and interchromosomal rearrangements. For their identification, the workflow examines discordant read pairs, mapping to different chromosomes or to the same chromosome at an anomalous distance, as well as split-reads, presenting soft-clipped and/or hard-clipped bases. This procedure permitted identifying breakpoint events at nucleotide precision. These strategies have shown to be accurate for targeted gene panel-based next-generation sequencing (Abel *et al.* 2014).

Generation of reference and NGS data for evaluation of analytical performance: sensitivity, specificity, positive predictive value, and ability to detect translocations

For calculating sensitivity and specificity, 10 immortal lymphoblastoid cell lines corresponding to 10 individuals whose genomes/exomes had been sequenced by the 1000 Genomes and HapMap projects were obtained from the Coriell Institute: NA20298 (ASW), NA12872 (CEU), NA18570 (CHB), HG00320 (FIN), HG00110 (GBR), A18960 (JPT), NA19020 (LWK), NA19794 (MXL), HG00740 (PUR) and NA18486 (YRI). Cell lines were cultured according to the protocols provided by Coriell, their DNAs isolated and mixed in equimolecular amounts. An NGS library was prepared, captured using the custom probe and sequenced in a single MiSeq run. Variants were called as described in the previous section and the results compared to those expected according to the genomic information available for these cell lines.

For obtaining the positive predictive value of the test on FFPE tumor DNA, 82 SNVs and indels previously identified by the platform in two different tumors with allelic frequencies, estimated by NGS read count, ranging from 0.11 to 0.88, were evaluated by PCR and Sanger sequencing. All variants were analyzed by direct and reverse sequencing.

We used the same two tumors to calculate sensitivity and specificity of the test on DNA extracted from FFPE samples. First, we made three mixtures of pre-captured libraries prepared with DNA from each of the tumors in different proportions: 10% tumor #1 library/90% tumor #2 library; 50%

tumor #1 library/50% tumor #2 library; and 90% tumor #1 library/10% tumor #2 library. The resulting three library-mixes were captured with the probe (v4) and sequenced in three MiSeq runs, and variants were identified. To compare the results with those expected based on a technology alternative to NGS, we used GeneChip Genome Wide Human SNP Arrays (Affymetrix), hybridized at CeGeN-PRB2-ISCHII (Santiago de Compostela, Spain), with the aim of identifying positive and negative variant calls from these two patients. To maximize the number of variants, instead of focusing on somatic variants from the two tumors mentioned above, we analyzed by SNP arrays germline DNA from the patients in which the tumors originated. These germline variants should also be present in the corresponding tumor DNA (unless loss of the variant allele has happened somatically). To be able to evaluate variants at different frequencies for sensitivity calculations, we focused on SNP array-positive variants exclusively present in one of the two patients, so that all of the variants selected get diluted when the two libraries are mixed. Thus, we only considered those variants detected by SNP-arrays in one patient and not detected in the other. Variants absent from both patients were considered as negatives for specificity calculations. All expected occurrences of the same variant in any of the library mixtures were taken into account for sensitivity calculations, whereas all expected absences were considered for specificity calculations. Finally, the ability of the NGS platform to correctly call 225 positive variants with frequencies ≥ 0.05 , 165 positive variants with frequencies ≥ 0.10 and 678 negative variants, was determined.

To test the ability to detect translocations, we obtained from Horizon Discovery (Cambridge, UK) three paraffin-embedded positive control glass slides carrying *ALK* (HD319), *RET* (HD640) and *ROS1* (HD615) translocations. Considering the frequency of the translocated allele estimated by the manufacturer for each positive control and the Qubit quantification, a single DNA mixture was prepared in which each translocation was estimated to be present with an allele frequency between 0.04 and 0.05. An NGS library was prepared, captured with SureSelect using the v4 probe and sequenced in a single MiSeq run.

Use of variant databases

Somatic variants

Somatic variant analysis is based on the Knowledge Database described by Dienstmann *et al.* (2014), which is periodically updated. A team of cancer genomicists and clinicians provide expert feedback on the contents of the database, including detection of potential errors which are communicated to its curators. Additions, removals and edition of the entries of each version of the database are

identified and subjected to analysis, to detect novel associations between FDA and/or EMA approved targeted cancer drugs and gene variants. If new variants affect genes not included in the current version of the gene-panel, all exons and/or selected introns affected by the variants are added to the following version of the gene panel.

Germline variants

Germline analysis was supported by the HGMD professional and ClinVar databases. Interpretation of database information is manually reviewed in the context of patient's phenotype, after in-depth analysis of the literature.

Variant classification and reporting

Somatic variants

For the classification and reporting of therapeutically relevant variants, we followed a methodology based on that proposed by Dienstmann *et al.* (Dienstmann *et al.* 2014), but focused on variants with allele frequencies below 0.05 in all the population databases evaluated, namely dbSNP, ExAC, ESP and 1000 Genomes, and that fulfill any of the following conditions: (1) they were present as somatic mutations in the COSMIC and/or the International Cancer Genome Consortium (ICGC) databases, (2) they were predicted to have an impact on the sequence of the proteins encoded by the genes they affect. Variants were evaluated by in depth literature analysis and filtered so that only those predicted to have an effect on the function of the gene or protein which has been related to sensitivity/response or resistance/no response to an approved targeted cancer drug were reported as clinically relevant. The information contained in the report was selected to balance space limitations and clear and unambiguous communication to the clinical team with the most recent recommendations for classification and interpretation of therapeutically relevant somatic cancer variants (Ritter *et al.* 2016; Li *et al.* 2017). The level of evidence for somatic variants with therapeutic relevance is summarized in the report using the following terms: (1) "Approved/Rejected by FDA and/or EMA" in the context of a specific tumor, (2) Recorded in the "NCCN and/or CAP" guidelines, (3) "Advanced Clinical-Trial", referring to prospective clinical trials guided by genomic analyses, (4) "Preliminary Clinical-Trial", referring to non-advanced clinical trials showing preliminary evidence of efficacy or lack of it (5) "Clinical case", referring to individual cases with exceptional responses to targeted therapies in a specific genomic context and (6) "Preclinical", referring to robust but preclinical data which are being explored in

clinical trials. Somatic variants of unknown significance were also reported in the corresponding section (Fig. S1).

Germline variants

For the classification of germline variants related to hereditary cancer we defined the following three categories: (1) (Known) Pathogenic variants: those considered as disease-causing mutations (DM) by HGMD and related to hereditary cancer which, under individual up-to-date review of the literature, were considered to be supported by solid scientific evidence, and which were absent or showed germline allelic frequencies below 0.05 in all the population databases evaluated, namely dbSNP, ExAC, ESP and 1000 Genomes; (2) Likely pathogenic variants: those fulfilling the following four conditions: (a) they were not considered as DM by HGMD, (b) they were strongly predicted to produce a drastic effect on the sequence of the protein (premature stop codon, loss of initiation ATG codon, aberrant splicing or frameshift), (c) they affected genes for which germline loss of function was associated with an increased risk of developing cancer, (d) they were absent or showed germline allelic frequencies below 0.01 in any of the population databases evaluated; (3) Uncertain clinical significance variants: those fulfilling any of the following conditions: (a) they were not considered pathogenic, causal of hereditary cancer, by HGMD, nor likely pathogenic, and they showed a maximum allelic frequency below 0.01 in the considered population databases; or (b) they were considered pathogenic mutations and related to hereditary cancer (DM: disease-causing mutation) by HGMD and showed a maximum allelic frequency below 0.05 in the considered population databases, but their association with cancer was based on insufficient and/or controversial evidence; or (c) they fulfilled all the criteria of likely pathogenic variants except for their allelic frequency, which was between 0.01 and 0.05 in the population databases considered. Variants were reported as shown in Fig. S2 and S3.

Results

Definition of the ONCOgenics gene panel

Somatic gene subpanel (ONCOgenics Tumor™)

We generated a list of cancer genes whose alterations predict sensitivity or resistance to targeted drugs approved for the treatment of cancer, irrespective of its anatomical origin or histology. This panel includes all coding exons from 97 genes for detection of SNVs, indels and copy number variants (CNVs), as well as specific introns from 17 genes to detect selected

rearrangements (Table 1). Cancer genes whose somatic alterations only have diagnostic and/or prognostic clinical value (i.e., with no predictive value) are not included in the somatic subpanel.

Germline gene subpanel (ONCOgenics *Germline*[™])

We selected 148 genes with germline mutations consistently associated with an increased risk of developing cancer on a 2-tier system (Table 2). Tier-1 includes 72 genes considered as predominantly cancer-predisposing, being cancer the initial presentation or a primary clinical feature of the syndrome. Genes within Tier-1 are sub-classified based on whether there exist established guidelines/recommendations for the management of patients with germline pathogenic variants in them (Tier-1a), or not (Tier-1b). Tier-2 includes 76 genes associated with cancer in patients with overt mainly noncancerous syndromic features. Although both sets of genes are included in the panel, the latter are only considered reportable in the setting of a syndromic patient being analyzed, thus increasing the clinical sensitivity, while reducing the number of reported variants that are unrelated to the clinical manifestations of the patient.

Translational research subpanel

Sixty four additional cancer-related genes not considered to be validated targets for approved therapies or not unambiguously associated with a hereditary cancer predisposition, but showing preliminary evidence for their potential relevance in oncology, are included. Coding exons and/or relevant non-coding regions are targeted by the panel for research purposes only and their alterations are not clinically reported, but included in a separate variant list, if required (Table 3).

Improved probe performance through iterative re-design

We initially developed a single custom SureSelect probe design targeting the genes in our panel (version 1 or v1) and tested its performance by generating libraries from tumor DNA extracted from formalin-fixed paraffin-embedded (FFPE) tumor tissue and germline DNA from a single patient. MiSeq sequencing of such libraries produced total callabilities (percentages of bases covered with a given base quality, mapping quality and read depth) above 99% at depths of 10 (DP10), 20 (DP20) and 50 (DP50) reads (Table S1).

However, we observed differences in the distribution of reads throughout the target regions. Thus, we detected 173 regions for which callability was poor (i.e. <100% at DP20), whereas 2672 regions showed 100% callabilities at DP500, from a total of 6055 regions. With the aim of evening the sequencing depth throughout all target regions, we redistributed probe densities in successive versions of the design. This produced a notable improvement of the callabilities at the poor regions from 22.40% (v1) to 91.15% (v2) at DP50 (Table S2), although at the cost of a slight decrease on total callabilities at DP50 and DP100 (Table S1). Further redistribution of probes in v3 and v4 of the design allowed recovering callabilities above 99% at DP10, DP20 and DP50 in all regions (Table S3).

Sensitivity, specificity, and positive predictive value of the developed platform for point mutations and indels

We evaluated the analytical performance of the NGS platform on DNA isolated from both fresh cells and FFPE tissues. The test showed sensitivities and specificities on cell lines above 99% for the detection of the 2639 variants

Table 1. Tumor gene subpanel (v6).

Gene list for the detection of single-nucleotide variants, insertions/deletions and copy number variants (97)								
<i>ABL1</i>	<i>BRCA2</i>	<i>CDKN2B</i>	<i>ERBB4</i>	<i>GNA11</i>	<i>INPP4B</i>	<i>MPL</i>	<i>PIK3R2</i>	<i>SH2B3</i>
<i>AKT1</i>	<i>CBL</i>	<i>CDKN2C</i>	<i>FANCA</i>	<i>GNAQ</i>	<i>JAK1</i>	<i>MTOR</i>	<i>POLE</i>	<i>SMO</i>
<i>ALK</i>	<i>CCND1</i>	<i>CHEK2</i>	<i>FBXW7</i>	<i>GNA5</i>	<i>JAK2</i>	<i>MYD88</i>	<i>PRKCH</i>	<i>SOCS1</i>
<i>AR</i>	<i>CCND2</i>	<i>CSF1R</i>	<i>FGFR1</i>	<i>HDAC2</i>	<i>JAK3</i>	<i>NF1</i>	<i>PTCH1</i>	<i>SRC</i>
<i>ARAF</i>	<i>CCND3</i>	<i>CSF3R</i>	<i>FGFR2</i>	<i>HGF</i>	<i>KDR</i>	<i>NF2</i>	<i>PTEN</i>	<i>STAG2</i>
<i>ARID1A</i>	<i>CD274</i>	<i>CTNNB1</i>	<i>FGFR3</i>	<i>HRA5</i>	<i>KIT</i>	<i>NRAS</i>	<i>RAC1</i>	<i>STK11</i>
<i>ATM</i>	<i>CDK4</i>	<i>DDR2</i>	<i>FGFR4</i>	<i>IDH1</i>	<i>KRAS</i>	<i>PALB2</i>	<i>RAD51C</i>	<i>TSC1</i>
<i>ATR</i>	<i>CDK6</i>	<i>EGFR</i>	<i>FLCN</i>	<i>IDH2</i>	<i>MAP2K1</i>	<i>PDGFRA</i>	<i>RAF1</i>	<i>TSC2</i>
<i>BAP1</i>	<i>CDKN1A</i>	<i>EPHA2</i>	<i>FLT3</i>	<i>IGF1R</i>	<i>MAP2K2</i>	<i>PIK3CA</i>	<i>RET</i>	<i>VEGFA</i>
<i>BRAF</i>	<i>CDKN1B</i>	<i>ERBB2</i>	<i>FRS2</i>	<i>IGF2</i>	<i>MET</i>	<i>PIK3CB</i>	<i>RICTOR</i>	
<i>BRCA1</i>	<i>CDKN2A</i>	<i>ERBB3</i>	<i>GATA3</i>	<i>IL7R</i>	<i>MITF</i>	<i>PIK3R1</i>	<i>ROS1</i>	
Gene list for the detection of rearrangements (17)								
<i>ALK</i>	<i>BCR</i>	<i>BRAF</i>	<i>ERBB4</i>	<i>FGFR2</i>	<i>FGFR3</i>	<i>JAK2</i>	<i>MET</i>	<i>NRG1</i>
<i>NTRK1</i>	<i>PDGFB</i>	<i>PDGFRA</i>	<i>RAF1</i>	<i>RET</i>	<i>ROS1</i>	<i>TFE3</i>	<i>TMPRSS2</i>	

Table 2. Germline gene subpanel (v6). Genes in bold represent those 35 Tier 1 genes for which there exist established guidelines/recommendations for management of patients with germline mutations (Tier 1a). The remaining Tier 1 genes form Tier 1b.

Tier-1 genes (72)								
APC	BRIP1	FANCA	FANCL	MEN1	NF2	PTCH1	SDHA	SUFU
ATM	CDH1	FANCB	FANCM	MET	NTHL1	PTCH2	SDHAF2	TERT
AXIN2	CDK4	FANCC	FH	MLH1	PALB2	PTEN	SDHB	TMEM127
BAP1	CDKN1B	FANCD2	FLCN	MSH2	PDGFRA	RAD51C	SDHC	TP53
BARD1	CDKN2A	FANCE	GREM1	MSH6	PMS2	RAD51D	SDHD	TSC1
BMPR1A	CHEK2	FANCF	KIT	MUTYH	POLD1	RB1	SLX4	TSC2
BRCA1	EGFR	FANCG	MAX	NBN	POLE	RET	SMAD4	UBE2T
BRCA2	EPCAM	FANCI	MC1R	NF1	POT1	RHBDF2	STK11	VHL
Tier-2 genes (76)								
ALK	DIS3L2	FAH	KHDC3L	PHOX2B	RPL11	SEC23B	SPINK1	WT1
BLM	DKC1	FAS	KRAS	POLH	RPL35A	SERPINA1	STIM1	XIAP
BRAF	ERCC1	FASLG	LZTR1	PRF1	RPL5	SH2D1A	TERC	XPA
BUB1B	ERCC2	GPC3	MXN1	PRKAR1A	RPS19	SHOC2	TINF2	XPC
CBL	ERCC3	H19	MTAP	PRSS1	RPS24	SLC25A13	TRIM37	
CDKN1C	ERCC4	HFE	NHP2	PTPN11	RPS7	SMARCA4	UNC13D	
CYLD	ERCC5	HRAS	NLRP7	RAF1	RTEL1	SMARCB1	WNT10A	
DDIT2	EXT1	ITK	NOP10	RECQL4	RUNX1	SOS1	WRAP53	
DICER1	EXT2	KCNQ1OT1	NSD1	RIT1	SBDS	SOS2	WRN	

Table 3. Translational research subpanel (v6).

Genes (64)						
ATP2A2	DLEU7	GALNT12	KMT2C	PAX5	PTPRD	UNC5C
AXIN1	DNAAF1	GATA2	MAPK1	PBRM1	RECQL	XRCC2
BTX	EGLN1	HNF1A	MDH2	PDE11A	RNASEL	XRCC4
CDC73	EGLN2	HNF1B	MRE11A	PDPK1	SASH1	YAP1
CEBPA	ENG	HOXB13	MSR1	PIF1	SEMA4A	
CRKL	EPAS1	HOXD4	NFKBIZ	PLCG2	SFXN4	
CTNNA1	EPHB2	IL10RB	NOTCH1	PMS1	SLBP	
CTR9	EPHX1	INHBA	NOTCH2	PPARG	SMAD9	
CTRC	FAM175A	KIF1B	NTRK3 ¹	PRKCD	SRP72	
DAPK1	FAN1	KMT2A	OGG1	PSMC3IP	SRY	

¹Only specific introns to detect selected rearrangements.

(2524 SNVs and 115 indels) contained within the target region and present at frequencies ≥ 0.05 . Furthermore, sensitivity and specificity for the 1679 variants with frequencies ≥ 0.1 were both above 99.5% (Table 4). The performance of the analysis for detection of SNVs and indels on freshly isolated DNA was reproducible on DNA isolated from FFPE tumor tissues. The analytical positive predictive value was 100% based on 82 variants with frequencies 0.11–0.88 initially detected by the platform in those two tumors, all of which were later confirmed by PCR + Sanger (Table S4). Moreover, the sensitivity of the platform on three mixtures of libraries obtained from two FFPE tumors was >99% (223/225) for variants previously detected by SNP-arrays and expected to be present in the mixtures with frequencies ≥ 0.05 and 100% (165/165) for those with expected frequencies ≥ 0.1 , under >99.5% specificity conditions (Tables S5, S6).

Table 4. Sensitivity and specificity of the test on the mixture of cell lines.

Frequency	Sensitivity	Specificity
≥ 0.05	99.17% (SNV 99.30%; indel 96.60%)	>99.5%
≥ 0.10	99.66% (SNV 99.76%; indel 97.67%)	>99.5%
>0.20	99.70% (SNV 99.80%; indel 96.40%)	>99.5%

Performance at highly homologous regions

Genomic areas with high levels of homology to other parts of the genome represent challenging targets for conventional NGS: it may be very difficult or impossible to map short sequencing reads originating from those regions to the correct reference genome position. Mandelker et al. (2016) have recently generated an exome-

wide census with three different lists of problematic highly homologous regions affecting exon sequences: low-stringency, high-stringency and NGS-dead zones. Low-stringency and high-stringency zones show high degrees of homology to other regions of the genome, but with 1–5 mismatches per 250 nt window. This should facilitate their sequencing using proper mapping quality filters and taking into account the information originating from paired-end reads for mapping purposes. However, NGS-dead zones are regions ≥ 250 nucleotides long identical to at least another genomic region, and NGS approaches using inserts < 250 nt are unable to accurately sequence them.

We compared Mandelker's three lists with the target regions of our somatic and germline subpanels (v6) and *PMS2* (MIM: 600259) exon 15-3'UTR was the only NGS-dead zone in them. 8 additional *PMS2* exons were represented in the high-stringency (4 exons) or low-stringency (4 exons) lists (Table S7). Thus, we used *PMS2*, a relevant Lynch Syndrome gene for which our laboratory had previous experience in its analysis (Borrás *et al.* 2013), to test the developed NGS platform on highly homologous regions.

We focused on variants from five patients whose *PMS2* gene had been previously evaluated by Sanger sequencing after a long-range (LR) PCR strategy designed to distinguish the gene from its pseudogenes (Clendenning *et al.* 2006; Vaughn *et al.* 2010; Borrás *et al.* 2013). When the NGS results were aligned to a whole-genome reference sequence, all the variants outside exon 15-3'UTR were correctly detected. However, four false negative calls were identified, affecting two variants within the exon 15-3'UTR NGS-dead zone (NM_000535.5: c.2466T>C and c.*92dupA, with two occurrences each) (Table S8 and Fig. S4).

In the specific case of *PMS2*, on top of the homology issue, gene conversion events between its 3' end (exons 11, 12, 13, 14 and 15-3'UTR) and that of *PMS2CL* have been documented (van der Klift *et al.* 2010; Vaughn *et al.* 2010). As a result, any variant potentially present in exons 11 to 15 and detected by approaches that do not make use of LR PCR might correspond to either the *PMS2* or the *PMS2CL* locus. Accordingly, a false positive call was detected in one sample (NM_000535.5: c.2466T>C) (Table S8).

To improve the sensitivity in those *PMS2* regions, we realigned the NGS results with a reference sequence containing the *PMS2* locus only. This permitted 100% sensitivity of true *PMS2* variants (17/17), although at the expense of specificity (Table S9). Thus, although this strategy can be used on NGS-dead zones or regions affected by gene-conversion, any clinically relevant

mutation potentially identified must be confirmed by target-specific methods (i.e., LR PCR).

These results indicate that the current platform is able to accurately detect variants in highly homologous regions outside Mandelker's NGS-dead zone, as long as they are not affected by gene-conversion events (i.e., outside *PMS2*). Thus, the performance of ONCOgenics is expected to be good on all other genes from the germline and/or somatic subpanels, whereas variants from exons 11–15 of *PMS2*, detectable by aligning NGS results to a reference sequence restricted to the *PMS2* locus, must be validated by alternative, gene-specific approaches.

Ability to detect translocations

We evaluated the ability of the test to detect translocations affecting *ALK*, *ROS1* and *RET* on a mixture of DNA obtained from three paraffin-embedded positive control cell lines, which contained each translocation present in an allelic fraction between 0.04 and 0.05. Analysis by NGS using a novel algorithm developed *ad hoc* for rearrangement identification successfully detected all three translocations and mapped them to nucleotide resolution (Table S10). Notably, before data analysis, the bioinformaticians involved in this evaluation were unaware of the 5' partners of the *ROS1* and *RET* translocations, as well as of the chromosomal coordinates of all six breakpoints.

External quality assessment of analytical performance

We participated in the 2016 Oncogene Panel Testing External Quality Assessment scheme organized by the European Molecular Genetics Quality Network (EMQN) to independently assess the performance of the test. The scheme evaluated the analysis of eight established cancer genes with 30 mutations present in total in the four FPPE samples (one of which, sample #4, was mildly formalin compromised), with variant frequencies between 0.007 and 0.667 as per ddPCR quantitation (Table S7). Detection of mutations with frequencies below 0.025 in these samples was based on the calling process supporting variants present in the COSMIC database described above (see Materials and Methods).

ONCOgenics obtained the maximum score (2.00/2.00), as all present variants and no false positives were detected from all samples. Moreover, variant quantitation based on the NGS platform strongly correlated with the ddPCR variant frequencies reported by EMQN, even when considering the formalin compromised sample ($R = 0.9952$ for samples 1–3; $R = 0.9744$ for samples 1–4) (Fig. 1 and Table 5).

Determination of minimum sample requirements

According to the “Low Input Sureselect^{XT}” library preparation protocol, 200 ng of double stranded DNA (dsDNA), measured by a method able to distinguish it from single stranded or degraded DNA (such as the Qubit fluorimetric assay), is the recommended amount of starting material. This is easy to obtain from blood or saliva samples, but may be a limitation for formalin-fixed samples. Thus, based on our results with 46 formalin-fixed samples, we determined the minimum DNA amount and quality requirements for successful library preparation (Table S11).

The Qubit-concentration/NanoDrop-concentration ratio (Q/N) is an indicator of DNA purity and/or quality, as spectrophotometric (NanoDrop) measurements are based on total absorbance at 260 nm whereas fluorimetric measurements are more specific for dsDNA. According to our series, a minimum of 100 ng of DNA with a Q/N higher than 0.29 is a good indicator of a DNA sample from FFPE tissue being of enough quality for library preparation. Nonetheless, samples with lower DNA amounts or Q/N ratios should not be discarded, as they allow library preparation most of the times. Q/N ratios lower than 0.1, however, predict library preparation failure or libraries

with very low percentage of sequence on target (7/8 in our series).

Regarding the minimum tumor content of the sample, this is set at 20% because, according to Table 4, that would allow to detect, with >99.5% sensitivity and specificity, heterozygous somatic SNVs or indels present in 100% of the tumor cells (with a tumor content of 20% their allele frequencies would be ≥ 0.1).

Samples below the 20% tumor cellularity threshold can be enriched by macro/microdissection. However, we currently do not have a method for qualifying samples with Q/N lower than 0.1. In such cases, although library construction may be attempted, testing of a different sample is recommended.

Finally, in terms of tumor sample type, besides formalin-fixed, paraffin-embedded tissue obtained from excisional biopsies, we have also obtained good quality DNA and reliable results with small core biopsies and tumor cells from fine-needle aspirates.

Advantages of parallel tumor-germline analysis

Analysis of a germline sample in parallel to the tumor sample is considered desirable by the most recent standards and guidelines for the interpretation of sequence

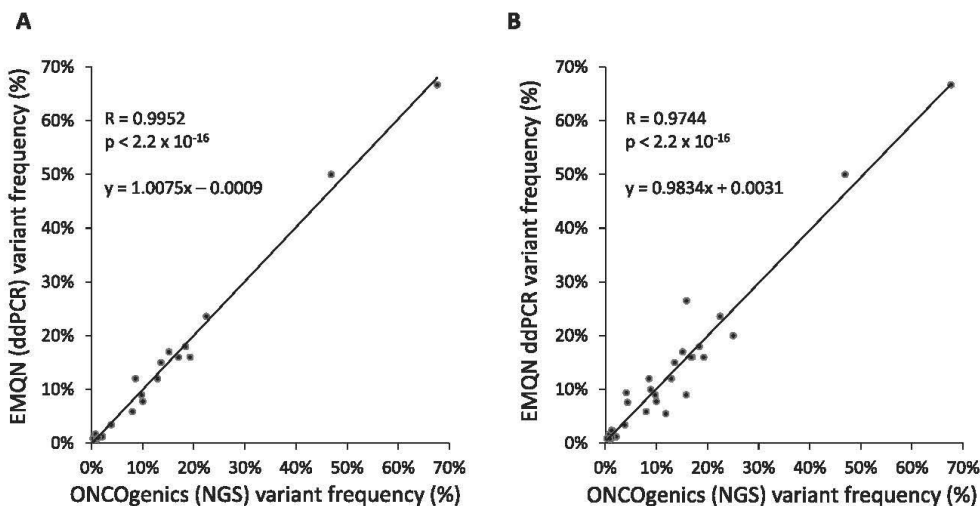


Figure 1. Correlation of variant frequencies detected by ddPCR and ONCOgenics. (A) Correlation between the frequencies of the variants present in three FFPE samples from the 2016 EQMN Oncogene Panel Testing External Quality Assessment Scheme as reported by EMQN (quantitated by ddPCR) and the frequencies of the same variants as determined by the NGS panel testing. (B) Same correlation as in (A) but after considering a fourth FFPE EMQN sample which had been mildly formalin compromised. The Pearson correlation coefficients (R), the P -values (Pearson correlation test) and the linear regression equations ($y = ax + b$) are shown.

Table 5. Performance of the platform on FFPE samples from the 2016 EMQN Oncogene Panel Testing External Quality Assessment Scheme. All variants were correctly detected. Percentages show the frequencies of the variants present as reported by EMQN (quantitated by ddPCR) and the frequencies of the same variants as determined by the developed NGS platform. Sample #4 was mildly formalin compromised. Analyses of these samples were performed with v6.

Variants present in each sample	Estimated frequencies	
	EMQN (ddPCR) (%)	NGS platform (%)
FFPE sample #1		
<i>BRAF</i> (NM_004333.4) c.1799T>A; p.V600E	16	19.24
<i>KRAS</i> (NM_004985.3) c.38G>A; p.G13D	16	16.97
<i>PIK3CA</i> (NM_006218.2) c.3140A>G; p.H1047R	17	15.12
FFPE sample #2		
<i>BRAF</i> (NM_004333.4) c.1799T>A; p.V600E	66.70	67.62
<i>EGFR</i> (NM_005228.3) c.2155G>A; p.G719S	1.10	0.72
<i>EGFR</i> (NM_005228.3) c.2235_2249del15; p.E746_A750del	0.90	0.28
<i>EGFR</i> (NM_005228.3) c.2369C>T; p.T790M	1.20	2.11
<i>EGFR</i> (NM_005228.3) c.2573T>G; p.L858R	1.30	0.95
<i>EGFR</i> (NM_005228.3) c.2582T>A; p.L861Q	0.70	0.95
<i>PIK3CA</i> (NM_006218.2) c.3140A>G; p.H1047R	50	46.87
FFPE sample #3		
<i>BRAF</i> (NM_004333.4) c.1799T>A; p.V600E	12	12.88
<i>EGFR</i> (NM_005228.3) c.2155G>A; p.G719S	23.60	22.44
<i>EGFR</i> (NM_005228.3) c.2235_2249del15; p.E746_A750del	1.70	0.72
<i>EGFR</i> (NM_005228.3) c.2369C>T; p.T790M	1.02	1.33
<i>EGFR</i> (NM_005228.3) c.2573T>G; p.L858R	3.40	3.82
<i>KIT</i> (NM_000222.2) c.2447A>T; p.D816V	9	9.75
<i>KRAS</i> (NM_004985.3) c.35G>A; p.G12D	5.90	7.97
<i>KRAS</i> (NM_004985.3) c.38G>A; p.G13D	15	13.55
<i>NRAS</i> (NM_002524.4) c.181C>A; p.Q61K	12	8.55
<i>PIK3CA</i> (NM_006218.2) c.1633G>A; p.E545K	7.80	9.98
<i>PIK3CA</i> (NM_006218.2) c.3140A>G; p.H1047R	18	18.36
FFPE sample #4 (formalin compromised)		
<i>BRAF</i> (NM_004333.4) c.1799T>A; p.V600E	9.40	4.09
<i>EGFR</i> (NM_005228.3) c.2155G>A; p.G719S	26.50	15.87
<i>EGFR</i> (NM_005228.3) c.2235_2249del15; p.E746_A750del	2.30	1.50
<i>EGFR</i> (NM_005228.3) c.2369C>T; p.T790M	0.85	0.90
<i>EGFR</i> (NM_005228.3) c.2573T>G; p.L858R	2.40	1.20
<i>KIT</i> (NM_000222.2) c.2447A>T; p.D816V	9	15.79
<i>KRAS</i> (NM_004985.3) c.35G>A; p.G12D	5.50	11.81
<i>KRAS</i> (NM_004985.3) c.38G>A; p.G13D	16	16.79
<i>NRAS</i> (NM_002524.4) c.181C>A; p.Q61K	10	8.84
<i>PIK3CA</i> (NM_006218.2) c.1633G>A; p.E545K	7.60	4.32
<i>PIK3CA</i> (NM_006218.2) c.3140A>G; p.H1047R	20	25

variants in cancer (joint consensus recommendation of the Association for Molecular Pathology, American Society of Clinical Oncology, and College of American Pathologists) (Li *et al.* 2017). The relevance of matched germline testing for somatic mutation identification is threefold. First, during the processing of sequencing results, it allows removing systematic errors happening at low frequencies, which represent considerable noise for accurate identification of subclonal or diluted somatic variants, as previously demonstrated by us and others (Puente *et al.* 2011). Second, it allows unequivocal assignment of somatic versus germline variant origin (Schrader *et al.* 2016). Third, CNV identification is also benefited from this situation: in addition to the detection of differences with read distributions from previously sequenced samples, the analysis can be complemented with the comparison between paired tumor-normal samples processed in parallel. These three advantages add up to improve the sensitivity and specificity of somatic variant identification.

To ascertain to which extent variants reported in established tumor-only analyses are of germline origin, we evaluated the germline of 6 cases from our clinic previously sequenced by Foundation One™ (Frampton *et al.* 2013). As predicted, 21 out of 67 (31%) variants reported were of germline origin, and two of them had been linked to cancer therapies (Table S12). These results are in agreement with an independent systematic evaluation of tumor-only versus paired tumor-germline analysis of a targeted panel of 111 genes in 58 tumor samples, which concluded that 31% of the variants considered as somatic mutations by the tumor-only analysis were actually of germline origin (Jones *et al.* 2015).

Report structure

Reporting of clinical NGS findings has to be succinct and aimed at transmitting specifically the relevant results.

Somatic analysis

For the somatic analysis (request specifically oriented to the identification of therapeutic options for a cancer patient) we evaluate and report the genomic alterations related with sensitivity or resistance to approved cancer drugs according to the literature (for a more detailed description see “Materials and Methods” and Fig. S1).

Germline analysis

When a germline analysis is requested for identification of a cancer predisposing hereditary genomic alteration, we report the findings in three variant categories: pathogenic, likely pathogenic and variants with uncertain

clinical significance (for a detailed description see “Materials and Methods” and Fig. S2). An “Additional comments” section summarizes pertinent negatives and includes mentions to potentially clinically relevant findings. Further detailed information is provided for each of the identified variants in the following report pages (Fig. S3).

Finally, both for the somatic and the germline analysis, information about the technical performance of the test on the sample analyzed is included in the report.

Report modalities

When ordering tumor NGS testing, oncologists face complex situations to decide whether or not they should include parallel germline NGS testing, an issue which remains unsolved and under debate nowadays (Li et al. 2017). With the aim of fitting all possibilities, the report modalities of the described platform adapt to five different situations depending on the aim of the request, availability of tumor and/or germline sample and type of informed consent (Fig. 2):

- (1) the request has a therapeutic aim (finding approved cancer drug-related genomic alterations) and only tumor sample is available;
- (2) the request has a therapeutic aim and both tumor and germline samples are provided, but no informed consent for reporting germline cancer predisposing mutations is available;
- (3) as in 2, but an informed consent is available which covers the disclosure of germline cancer predisposing mutations;
- (4) as in 3, but the request has both therapeutic and genetic counseling aims;
- (5) the request is aimed at genetic counseling, only germline sample is provided and an informed consent is available covering the disclosure of germline cancer-predisposing mutations.

The algorithms and report modalities of the platform adapt to each of these situations:

In (1), those variants in the somatic gene panel most likely to be tumor-specific are selected based on removing putative germline alterations frequently found in the population (those with variant frequency ≥ 0.05 in any of the four population databases considered: 1000 Genomes, ESP6500, ExAC and dbSNP). Genomic alterations below that threshold and related to approved cancer therapies according to the updated literature evidence are reported as clinically relevant. Additionally, if a non-subclonal alteration is found which, when present in the germline, is known to cause hereditary cancer, a warning is included in the “Additional comments” section to call the

clinician’s attention so that the pertinent measures can be adopted (such as testing for the presence of the specific alteration on germline DNA after appropriate genetic counseling consultation).

In (2), we search for genomic alterations in our somatic gene panel related to available approved therapies in the tumor sequence, as in (1), but after bioinformatics removal of germline variants. In parallel, the germline sequence is also searched for variants with potential therapeutic relevance. This is done to avoid missing them as a result of the germline subtraction performed during the tumor analysis (i.e., a potential BRCA1 loss of function mutation present both in an ovarian cancer and in the germline sequence). We report all the therapeutically relevant variants as present in the tumor and, if any of them, not subclonal, is known to cause hereditary cancer when present in the germline, we include the same warning as in (1).

In (3), we perform the same kind of analysis as in (2), but we explicitly report the presence of hereditary cancer causing variants when they are found in the germline.

In (4), an analysis similar to that done in (3) is performed, but considering both the somatic and the whole germline gene subpanels. A report with two separate sections is issued, one section for the therapeutically relevant findings and the other with known or potential germline cancer-predisposing variants (*Complete* report).

Finally, in (5) the germline panel is analyzed and the corresponding report is produced. Of note, when germline analyses are ordered, the requestors are given the option to restrict the report to Tier-1a genes: those for which there exist established guidelines/recommendations for the management of patients with germline pathogenic variants.

Clinical application

We evaluated the utility of the test in the clinical setting by applying it to 39 tumor and 36 germline samples from different patients.

Somatic findings

Analysis of the tumor samples revealed genomic alterations associated with approved cancer therapies in 89.74% of the analyzed cases (35/39) (Table 6). This yield is comparable to that reported in the literature (Zhong et al. 2015), and consistent with our previous experience with established platforms: analysis of 23 tumors from our clinic with Foundation One™ or Foundation One Heme™ had revealed approved therapies in 78% of the cases (18/23) (Table S13). Remarkably, all the somatic alterations identified by the Foundation tests associated

	1	2	3	4	5
Aim of request	Therapeutic	Therapeutic	Therapeutic	Therapeutic + Genetic counseling	Genetic counseling
Available sample	Tumor DNA	Tumor DNA	Tumor DNA	Tumor DNA	Germline DNA
Germline informed consent	Not available	Not Available	Available	Available	Available
Gene panel evaluated	ONCOgenics Tumor™ 97 genes + rearrangements in 17 genes	ONCOgenics Tumor™ 97 genes + rearrangements in 17 genes	ONCOgenics Tumor™ 97 genes + rearrangements in 17 genes	ONCOgenics Tumor™ + ONCOgenics Germline™ 97 + 148 genes + rearr. in 17 genes	ONCOgenics Germline™ 148 genes
Tumor-specific mutations filter	Population allele frequency	Germline subtraction (Inclusion of therapeutically relevant germline variants, w/o specification of their germline origin)	Germline subtraction (Inclusion of therapeutically relevant germline variants, with specification of their germline origin)	Germline subtraction	Not applicable
Putative germline variants comments	Germline testing "recommended"	Germline testing "recommended"	Germline variant "identified"	In depth germline variant analysis	In depth germline variant analysis
Report type	Tumor only	Tumor only	Tumor only	Complete (Tumor + Germline)	Germline only

Figure 2. Graphical guide for adaptation of the sample-to-report pipeline. Columns 1 to 5 correspond to the different potential situations described in the main text. The first three rows (“Aim of request”, “Available sample” and “Germline informed consent”) represent the conditions which, combined, define those 5 situations. The last four rows (“Gene panel evaluated”, “Tumor specific mutations filter”, “Putative germline variants comments” and “Report type”), represent the combination of solutions taken to adapt the sample-to-report pipeline to each of those 5 situations.

with approved targeted therapies were in genes included in our somatic gene subpanel.

Beyond its specific aims, comprehensive genomic profiling of the tumor can identify potential diagnostic errors. For instance, one of the analyzed tumors was labeled as a metastatic adenocarcinoma from a breast cancer, as it appeared as an isolated mediastinic lymph node, in a patient who had had a breast primary tumor 12 years before. However, after an *EGFR* (MIM: 131550) exon 19 indel was identified by ONCOgenics, pathology review was performed and immunohistochemistry revealed that it was a primary lung adenocarcinoma. Subsequently, the patient was treated with an EGFR tyrosin kinase inhibitor (gefitinib), obtaining a partial response, sustained to date, after 15 months on treatment.

Germline findings

Application of the test to 36 germline samples from individuals with a personal and/or family history of cancer identified a pathogenic genomic alteration in *CDKN2A* (MIM: 600160) (p.V59G) in a patient with familial melanoma, and a truncating *BRCA2* variant (p.A938Pfs*21) in a patient with familial breast/ovary cancer (Table 7). Genomic germline analysis also revealed unexpected clinically relevant findings such as the presence of a

Xeroderma pigmentosum C recessive pathogenic alteration, *XPC* (MIM: 613208) (p.L763Cfs*) in heterozygosis in a healthy patient with a family history of breast/ovary cancer and leukemia, or a mosaic pathogenic mutation in *NFI* (MIM: 162200) (*NFI* p.1679Dfs*21, allelic frequency 10.3%) in another healthy patient with a family history of gastrointestinal cancer, prostate cancer and leukemia. This mosaic mutation, affecting a polyG tract within a low-stringency highly homologous region, was validated by LR PCR followed by Sanger sequencing (Fig. S2 and Fig. S5).

We also detected several variants listed at the moment of clinical interpretation of results by HGMD as pathogenic (disease causing mutation or DM) but, in our view, with controversial or insufficient support in 41.67% of the cases (15/36; based on Tier-1 genes only), which is in agreement with previous observations (Olíson et al. 2015; Groth et al. 2016). We classified these variants as “DM-Controversial” and reported them under the “uncertain clinical significance variants” category, separated from the robust pathogenic variants.

Discussion

The relevance and number of approved targeted cancer therapies is continuously growing. Most of them are

Table 6. Clinical performance of ONCOgenics Tumor™.

Test ID	Tumor type	Actionable genomic alterations	Est. % tum. affect.	Associated drugs
ONCOE.001	Lung adenocarcinoma	<i>KRAS</i> c.35G>T; p.G12V	50 100	Everolimus, temsirolimus, trametinib
		<i>STK11</i> c.169G>T; p.E57*	50 100	Bosutinib, dasatinib, everolimus, temsirolimus, trametinib
ONCOE.002	Melanoma	<i>CDKN2A</i> c.143C>T; p.P48L	45 100	Palbociclib
ONCOE.003	Poorly diff. lung adenocarcinoma	None	NA	None
ONCOE.004	Undiff. renal carcinoma	None	NA	None
ONCOE.005	Lung adenocarcinoma	<i>EGFR</i> c.2573T>G; p.L858R	100	Afatinib, erlotinib, gefitinib
ONCOE.006	Thyroid anaplastic carcinoma	<i>PIK3CA</i> c.1624G>A; p.E542K	10 30	Everolimus, temsirolimus
		<i>NRAS</i> c.181C>A; p.Q61K	5 20	Trametinib
ONCOE.007	Laryngeal squamous cell carcinoma	<i>BAP1</i> c.1379C>G; p.S460*	95 100	Olaparib, valproic acid ¹
		<i>FGFR1</i> amplif.	NE	Pazopanib, ponatinib
ONCOE.008	Lung adenocarcinoma ²	<i>EGFR</i> c.2237_2254del; p.E746_S752delinsA	75 100	Afatinib, afatinib + cetuximab, erlotinib, gefitinib,
ONCOE.009	Paraganglioma	None	NA	None
ONCOE.010	Thyroid medullar carcinoma	<i>RET</i> c.2753T>C; p.M918T	50 100	Cabozantinib, sunitinib, vandetanib
		<i>EPHA2</i> c.1171G>A; p.G391R	15 40	Bosutinib, dasatinib, everolimus, temsirolimus
ONCOE.011	Tongue squamous cell carcinoma	<i>KRAS</i> c.35G>C; p.G12A	≤10	Everolimus/temsirolimus + trametinib, gemcitabine + trametinib, trametinib
		<i>EGFR</i> amplif.	NE	Cetuximab
ONCOE.012	Lung squamous carcinoma	<i>PTEN</i> c.802-2A>T	25 60	Everolimus, olaparib, temsirolimus
		<i>CDKN2A</i> c.71G>C; p.R24P	20 50	Palbociclib
		<i>NF1</i> c.3299_3300del; p.A110Vfs4*	15 40	Everolimus, temsirolimus, trametinib
ONCOE.013	Primary peritoneal adenocarcinoma	<i>FGFR2</i> c.758C>G; p.P253R	5 20	Pazopanib
		<i>ERBB2</i> amplif.	NE	Ado-trastuzumab, afatinib, emtansine, lapatinib, pertuzumab, trastuzumab
ONCOE.014	Liposarcoma	<i>CDKN2A</i> c.172C>T; p.R58*	40 90	Palbociclib
		<i>FRS2</i> amplif.	NE	Nintedanib, pazopanib, ponatinib
ONCOE.015	Esophageal epidermoid carcinoma	<i>CDK4</i> amplif.	NE	Palbociclib
		<i>CDKN2A</i> c.176T>G; p.V59G	45 100	Palbociclib
ONCOE.017	Lung adenocarcinoma	<i>KRAS</i> c.183A>C; p.Q61H	30 70	Everolimus, palbociclib, temsirolimus, trametinib
		<i>STK11</i> c.157_158insG; p.D53Gfs*110	40 90	Bosutinib, dasatinib, everolimus, temsirolimus, trametinib
ONCOE.018	Colorectal adenocarcinoma	<i>CD274</i> amplif.	NE	Nivolumab, pembrolizumab
		<i>KRAS</i> c.35G>A; p.G12D	45 100	Cetuximab (no response), cobimetinib, palbociclib, panitumumab (no response), trametinib
ONCOE.019	Lung acinar adenocarcinoma	<i>FRS2</i> amplif.	NE	Nintedanib, pazopanib, ponatinib
		<i>NF1</i> deletion	NE	Everolimus, temsirolimus, trametinib
		<i>CCND1</i> amplif.	NE	Palbociclib
ONCOE.020	Mediastinal choriocarcinoma	<i>CDK4</i> amplif.	NE	Palbociclib
		<i>PTEN</i> c.328C>T; p.Q110*	45 100	Everolimus, olaparib, temsirolimus
ONCOE.021	Lung adenocarcinoma	<i>KRAS</i> c.37G>T; p.G13C	40 90	Everolimus, palbociclib, temsirolimus, trametinib
		<i>STK11</i> c.290 + 1G>C	45 100	Bosutinib, dasatinib, everolimus, temsirolimus, trametinib
ONCOE.023	Lymph node metastasis of adenocarcinoma	<i>BAP1</i> deletion	100	Olaparib, platinum derivatives
		<i>BAP1</i> c.1769A>T; p.Q590L	100	
		<i>PALB2</i> deletion	100	Olaparib, platinum derivatives

(Continued)

Table 6. Continued.

Test ID	Tumor type	Actionable genomic alterations	Est. % tum. affect.	Associated drugs
	of unknown primary site	<i>PALB2</i> c.2590C>T; p.P864S <i>FBXW7</i> deletion <i>TSC1</i> deletion <i>TSC2</i> deletion <i>FLCN</i> deletion <i>PTCH1</i> deletion <i>PTCH1</i> c.505G>A; p.V169I	100 100 100 100 100 100	Taxanes (resistance) Everolimus Everolimus Everolimus Vismodegib
ONCOE.024	Lung adenocarcinoma	<i>NF1</i> c.7127-1G>T <i>NF1</i> c.7395-1G>T	30 70 30 70	Everolimus, temsirolimus, trametinib
ONCOE.025	Glioblastoma multiforme	<i>EGFRvIII</i> variant (exons 2-7 deletion) <i>CDKN2A</i> deletion <i>CDKN2B</i> deletion	NE 100 100	Erlotinib (no response) Palbociclib Palbociclib
ONCOE.026	Müllerian carcinosarcoma	None	NA	None
ONCOE.027	Lung adenocarcinoma with lepidic pattern	<i>KRAS</i> c.38G>A; p.G13D <i>KRAS</i> c.34G>T; p.G12C	15 40 5 20	Afatinib (no response), cobimetinib, erlotinib (no response), gefitinib (no response), palbociclib, trametinib
ONCOE.028	Germ cell ovarian tumor (Sertoli-Leydig cell tumor)	<i>ATM</i> c.8851-2_8851-1delAGinsTT <i>NF1</i> deletion <i>CTNNB1</i> c.110C>G; p.S37C <i>CDK6</i> amplif.	5 20 NE 5 20 NE	Olaparib Cobimetinib, everolimus, temsirolimus, trametinib Everolimus + letrozole Palbociclib
ONCOE.029	Colorectal adenocarcinoma	<i>KRAS</i> c.35G>A; p.G12D	25 60	Cetuximab (no response), cobimetinib, palbociclib, panitumumab (no response), trametinib
ONCOE.030	Colorectal adenocarcinoma	<i>KRAS</i> c.35G>T; p.G12V	30 70	Cetuximab (no response), cobimetinib, palbociclib, panitumumab (no response), trametinib
ONCOE.031	PEComa	<i>PDGFRA</i> c.2526_2537del; p.I843_D846del <i>ATR</i> deletion	20 50 NE	Imatinib, regorafenib, sunitinib Olaparib
ONCOE.032	Colorectal adenocarcinoma	<i>FLCN</i> deletion <i>PIK3R1</i> deletion	NE NE	Everolimus, temsirolimus Everolimus, temsirolimus
ONCOE.034	Small cell lung cancer	<i>PTEN</i> deletion <i>PTEN</i> c.867dupA; p.V290Sfs*8 <i>BAP1</i> deletion	100 100 NE	Everolimus, olaparib, temsirolimus Olaparib, panobinostat, valproic acid, vorinostat
ONCOE.035	Colorectal adenocarcinoma	<i>BRCA2</i> deletion <i>FLCN</i> deletion <i>PIK3R1</i> deletion <i>BRAF</i> c.1799T>A; p.V600E	NE NE NE 25 60	Nivolumab, olaparib, pembrolizumab Everolimus, temsirolimus Everolimus, temsirolimus Cetuximab (no response), dabrafenib + trametinib, panitumumab (no response), vemurafenib (no response)
ONCOE.036	Nasal teratocarcinosarcoma	<i>CCND1</i> amplification <i>CDKN2A</i> deletion <i>CDKN2B</i> deletion	NE NE NE	Palbociclib Palbociclib Palbociclib
ONCOE.039	Rectal adenocarcinoma	<i>PALB2</i> c.1547delG; p.R516Kfs*45 <i>NF1</i> c.204 + 1G>T	10 30 70 100	Olaparib Everolimus, temsirolimus, trametinib, cobimetinib
ONCOT.040	Papillary thyroid cancer	<i>BRAF</i> c.1799T>A; p.V600E <i>PIK3CA</i> c.1258T>C; p.C420R <i>PIK3CA</i> c.3145G>C; p.G1049R	45 100 25 60 30 70	Vemurafenib, dabrafenib, trametinib, cobimetinib, dabrafenib + trametinib Everolimus, temsirolimus

(Continued)

Table 6. Continued.

Test ID	Tumor type	Actionable genomic alterations	Est. % tum. affect.	Associated drugs
ONCOT.041	Adenocarcinoma of unknown primary site (sample origin: pleura)	<i>EGFR</i> c.2573T>G; p.L858R (+ mutant allele amplification)	80 100	Erlotinib, afatinib, gefitinib
ONCOT.043	Ovarian serous papillary carcinoma	<i>NF1</i> c.320delC; p.T107Rfs*58	85 100	Everolimus, temsirolimus, trametinib, cobimetinib
ONCOT.044	Mesenteric fibromatosis	<i>CTNNB1</i> c.133T>C; p.S45P	30 70	Imatinib, everolimus + exemestane
ONCOC.001_T	Pancreatic adenocarcinoma	<i>FBXW7</i> c.1013_1016delGAAAGinsAAA; p.R338Kfs*4	35 80	Everolimus, temsirolimus
		<i>FBXW7</i> c.1053G>A; p.W351*	35 80	
		<i>FBXW7</i> c.1095G>A; p.W365*	35 80	
		<i>KRAS</i> c.34G>T; p.G12C	45 100	Trametinib, cobimetinib, palbociclib

¹Valproic acid is not an approved cancer drug, but it is a widely used and easily accessible HDAC inhibitor approved for the treatment of neurological disorders.

²Initial diagnosis: breast adenocarcinoma. Definitive diagnosis: lung adenocarcinoma (after pathology review motivated by NGS results).

approved for one or a very limited number of cancer histologies. However, some have demonstrated effectiveness against different tumor types as long as they harbor specific genomic alterations. Using NGS information for the selection of targeted therapies tailored to individual cancer cases is the basis of transversal precision oncology. The approach presented herein is especially suited for stage IV patients which have progressed to standard treatment options, when non-standard therapy is considered. The platform is also adapted by design to situations in which germline predisposition to cancer development is in play, either suspected or not. The detection of germline genomic alterations that cause increased cancer risk allows to extend personalized patient management to prevention, early detection and genetic counseling.

Since insufficiently validated tests do represent a threat to patients, we have devoted a significant effort to demonstrate that the current platform is ready for clinical use. It provides reliable results on formalin-fixed paraffin-embedded tumor tissue, small core biopsies and fine-needle aspirates, with diagnostic yields comparable to those of wider platforms and within a clinically useful timeframe (3 weeks for somatic analysis, 4 weeks for germline analysis). Additionally, the proposed single-test design simplifies the laboratory workflow, facilitates scaling up, and represents a good alternative to running a series of multiple single-gene companion tests on the often scarce tumor tissue. Indeed, parallelizing the analysis of multiple markers also maximizes the rate of actionable somatic findings (89.74% in our case). All this is in agreement with recent data on the analytical and the clinical utility of cancer gene panels, predicting that 80% of patients could benefit directly from tumor sequencing (Jones et al. 2015; Zhong et al. 2015).

The high sensitivity, specificity and positive predictive value of our test are consistent with the very good performance of the bioinformatics tools on which it is based, initially developed for the Chronic Lymphocytic Leukemia (CLL) genome project within the International Cancer Genome Consortium (ICGC). These tools have been extensively validated in the literature for the identification of clinically relevant somatic and germline mutations and have contributed for the CLL project to be the first in the ICGC to reach the 500 tumor/normal genome pairs goal (Puente et al. 2011, 2015; Quesada et al. 2012; Fanjul-Fernández et al. 2013; Ramsay et al. 2013). Combining strong bioinformatics with high coverage (average of 950x), our platform is able to perform sensitive analysis of clonal heterogeneity, detecting mutant subclones which may be relevant for drug resistance.

Our analysis and reporting strategy is compatible with all options regarding sample availability, extension of informed consent and aim of request (Fig. 2). The proposed approach avoids potential problems generated by other strategies, such as missing a therapeutically actionable variant which is present both in the tumor and in the germline because of its subtraction by the bioinformatics analysis pipeline or failing to report a relevant hereditary cancer-causing germline mutation when only tumor DNA is analyzed. Moreover, comprehensive germline analysis also provides the potential for unexpected identification of mutation carriers of well-known cancer susceptibility syndromes in families with atypical phenotypes not meeting the established diagnostic criteria (LaDuca et al. 2014; Susswein et al. 2016). Future applications of NGS approaches, such as detection of cancer patients more likely to respond to immunotherapy, are

Table 7. Clinical performance of ONCOgenics *GermLine*TM.

Test ID	Pathogenic mutations	Likely pathogenic mutations	DM-Controversial mutations	Other VUS
ONCOG.001	None	None	None	<i>ATM</i> c.998C>T; p.S333F (Het.) <i>CDH1</i> c.-54G>C (Het.)
ONCOG.002	<i>CDKN2A</i> c.176T>G; p.V59G (Het.)	None	None	<i>BRCA2</i> c.4258G>T; p.D1420Y (Het.) <i>PTCH2</i> c.1073G>A; p.R358H (Het.) <i>SUFU</i> c.1018G>T; p.A340S (Het.)
ONCOG.003	None	None	None	<i>BRCA1</i> c.4883T>C; p.M1628T (Het.) <i>BRCA2</i> c.4258G>T; p.D1420Y (Het.) <i>FANCM</i> c.1667A>G; p.D556G (Het.) <i>PTCH2</i> c.1073G>A; p.R358H (Het.)
ONCOG.004	None	None	<i>BRCA2</i> c.7008-62A>G (Het.)	<i>CDH1</i> c.-54G>C (Het.) <i>SDHD</i> c.149A>G; p.H50R (Het.)
ONCOG.005	None	None	<i>FANCA</i> c.3348 + 18A>G (Het.)	None
ONCOG.006	None	None	None	<i>FANCA</i> c.932T>C; p.I311T (Het.) <i>FANCI</i> c.1573A>G; p.M525V (Het.) <i>MUTYH</i> c.1258C>A; p.L420M (Het.) <i>NF1</i> c.7259-17C>T (Het.)
ONCOG.007	None	None	<i>FANCA</i> c.3348 + 18A>G (Het.)	<i>ATM</i> c.2119T>C; p.S707P (Het.) <i>AXIN2</i> c.1685C>T; p.P562L (Het.) <i>BRCA1</i> c.2890G>A; p.G964R (Het.) <i>FANCB</i> c.2395G>A; p.A799T (Het.) <i>FANCD2</i> c.3127G>A; p.G1043S (Het.)
ONCOG.008	None	None	None	<i>ATM</i> c.998C>T; p.S333F (Het.) <i>FANCM</i> c.5177C>T; p.P1726L (Het.)
ONCOG.009	None	None	<i>MC1R</i> c.464T>C; p.I155T (Het.) <i>TSC2</i> c.5383C>T; p.R1795C (Het.)	<i>FANCD2</i> c.3275A>G; p.H1092R (Het.) <i>POLD1</i> c.189G>T; p.E63D (Het.) <i>SLX4</i> c.5501A>G; p.N1834S (Het.) <i>FANCM</i> c.3857G>T; p.S1286I (Het.)
ONCOG.010	None	None	None	<i>MSH6</i> c.*20_*24del (Het.)
ONCOG.011	None	None	<i>KIT</i> c.67 + 4G>A (Het.)	<i>PMS2</i> c.1688G>T; p.R563L (Het.) <i>SDHD</i> c.34G>A; p.G12S (Het.) <i>SUFU</i> c.1018G>T; p.A340S (Het.)
ONCOG.012	None	None	None	<i>MLH1</i> c.1852_1853delAAinsGC; p.K618A (Het.) <i>ATM</i> c.2519A>T; p.D840V (Het.) <i>FLCN</i> c.979G>A; p.A327T (Het.)
ONCOG.013	None	None	<i>ATM</i> c.1810C>T; p.P604S (Het.)	<i>ATM</i> c.6067G>A; p.G2023R (Het.)
ONCOG.014	None	None	<i>RET</i> c.1529C>T; p.A510V (Het.) <i>FANCA</i> c.3348 + 18A>G (Het.)	<i>MET</i> c.504G>T; p.E168D (Het.) <i>SDHD</i> c.34G>A; p.G12S (Het.) <i>POLE</i> c.2090C>G; p.P697R (Het.) <i>BRCA1</i> c.3083G>A; p.R1028H (Het.)
ONCOG.015	None	None	None	None
ONCOG.016	None	None	<i>ATM</i> c.1744T>C; p.F582L (Het.)	<i>BARD1</i> c.2212A>G; p.I738V (Het.) <i>BMPR1A</i> c.478A>G; p.M160V (Het.) <i>CHEK2</i> c.254C>T; p.P85L (Het.)
ONCOG.017	None	None	None	<i>FANCF</i> c.557C>T; p.A186V (Het.) <i>PALB2</i> c.2590C>T; p.P864S (Het.) <i>POLD1</i> c.2317G>A; p.A773T (Het.) <i>SUFU</i> c.1018G>T; p.A340S (Het.)
ONCOG.018	None	None	None	None
ONCOG.019	None	None	<i>FANCA</i> c.3348 + 18A>G (Het.)	<i>BRIP1</i> c.584T>C; p.L195P (Het.) <i>SLX4</i> c.710G>A; p.R237Q (Het.) <i>SLX4</i> c.4597G>C; p.A1533P (Het.)
ONCOG.020	None	None	None	<i>ATM</i> c.8560C>T; p.R2854C (Het.) <i>EPCAM</i> c.488G>A; p.R163Q (Het.)

(Continued)

Table 7. Continued.

Test ID	Pathogenic mutations	Likely pathogenic mutations	DM-Controversial mutations	Other VUS
ONCOG.021	None	None	<i>RET</i> c.2531G>A p.R844Q (Het.)	<i>BARD1</i> c.2212A>G; p.I738V (Het.) <i>BRIP1</i> c.3275C>A; p.P1092Q (Het.) <i>MUTYH</i> c.950T>C; p.L317P (Het.)
ONCOG.022	None	None	None	<i>ATM</i> c.2572T>C; p.F858L (Het.) <i>FANCA</i> c.1870G>T; p.A624S (Het.) <i>FANCC</i> c.29G>A; p.C10Y (Het.) <i>PTCH2</i> c.1073G>A; p.R358H (Het.) <i>SDHD</i> c.34G>A; p.G12S (Het.)
ONCOG.023	None	<i>FANCM</i> c.4005dupA; p.V1336Sfs*8 (Het.)	None	<i>ATM</i> c.2119T>C; p.S707P (Het.) <i>BAP1</i> c.1769A>T; p.Q590L (Het.) <i>BRCA1</i> c.591C>T; p.C197C (Het.) <i>CDH1</i> c.532-18C>T (Het.) <i>FANCC</i> c.77C>T; p.S26F (Het.) <i>PALB2</i> c.2590C>T; p.P864S (Het.) <i>PTCH1</i> c.505G>A; p.V169I (Het.)
ONCOG.024	<i>NF1</i> c.2033dupC; p.I679Dfs*21 (Mosaic)	None	None	<i>FANCA</i> c.3099C>A; p.D1033E (Het.) <i>BRCA1</i> c.3929C>A; p.T1310K (Het.)
ONCOG.025	None	None	<i>BARD1</i> c.1075_1095del; p.L359_P365del (Het.) <i>cdkn2aib</i> c.187G>C; p.G63R (Het.)	<i>FANCM</i> c.3857G>T; p.S1286I (Het.) <i>MCI1R</i> c.252C>A; p.D84E (Het.) <i>PALB2</i> c.2590C>T; p.P864S (Het.) <i>PALB2</i> c.232G>A; p.V78I (Het.)
ONCOG.026	None	None	None	<i>FANCA</i> c.3099C>A; p.D1033E (Het.)
ONCOG.027	None	None	None	<i>BRCA1</i> c.3708T>G; p.N1236K (Het.) <i>FANCC</i> c.584A>T; p.D195V (Het.) <i>FANCI</i> c.3812C>T; p.S1271F (Het.) <i>FANCM</i> c.4931G>A; p.R1644Q (Het.) <i>MSH6</i> c.100G>T; p.A34S (Het.)
ONCOG.028	None	None	<i>SLX4</i> c.421G>T; p.G141W (Het.)	<i>BRCA2</i> c.68-7T>A (Het.)
ONCOG.029	None	None	None	<i>MET</i> c.504G>T; p.E168D (Het.) <i>SDHD</i> c.149A>G; p.H50R (Het.)
ONCOG.030	None	None	<i>ATM</i> c.1810C>T; p.P604S (Het.)	<i>BRCA1</i> c.4039A>G; p.R1347G (Het.) <i>TSC2</i> c.1577G>A; p.S526N (Het.)
ONCOG.031	None	None	<i>FANCA</i> c.4069_4082del; p.A1357Lfs*63 (Het.)	<i>FANCM</i> c.171G>C; p.L57F (Het.) <i>RET</i> c.166C>A; p.L56M (Het.)
ONCOG.032	None	<i>FANCI</i> c.1051_1052delAG; p.S351Ffs*2 (Het.)	None	<i>BARD1</i> c.1028 C>T; p.T343I (Het.) <i>NBN</i> c.1591A>G; p.I531V (Het.) <i>PTCH1</i> c.1306G>A; p.D426N (Het.)
ONCOG.033	<i>BRCA2</i> c.2806_2809del; p.A938Pfs*21 (Het.)	None	<i>FANCA</i> c.1874G>C; p.C625S (Het.)	None
ONCOG.034	None	None	None	<i>APC</i> c.4424C>T; p.A1475V (Het.) <i>TMEM127</i> c.221A>G; p.Y74C (Het.)
ONCOG.035	None	None	None	<i>VHL</i> c.269A>G; p.N90S <i>APC</i> c.3949G>C; p.E1317Q
ONCOC.001_G	None	None	None	<i>SDHD</i> c.34G>A; p.G12S <i>SDHD</i> c.149A>G; p.H50R

also foreseeable (Le et al. 2015; Rizvi et al. 2015; Van Allen et al. 2015; Bouffett et al. 2016).

In spite of the great potential of NGS approaches for comprehensive mutational profiling of tumor and germline samples, there remain relevant challenges for which we propose a series of solutions (Gagan and Van Allen 2015):

Selection of the gene panel and clinical implications

Our somatic approach, focused on genes whose alterations have been related to approved cancer therapies only, is aimed at making the most of the information contained in the tumor related to accessible targeted

cancer drugs. In our experience, and in agreement with previous publications, the yield of clinically relevant findings is not negatively affected by focusing on a panel of genes when gene selection is carefully performed (Zhong *et al.* 2015). On the other hand, although there exist much smaller tissue-specific panels with reduced cost compared to the presented approach, those are not suitable for centers interested in transversal precision oncology applications. We recognize that, when a promising tailored therapeutic option is found for a given cancer patient, unless it is approved for the very indication of the case, it is most of the times very hard and frequently impossible to get access to the drug, even for those approved for other indications (47, 48). However, hopefully, in the same way the design of clinical trials is changing, approval of commercialized medicines for novel cancer indications and access to off-label drugs is also likely to adapt to the enormous possibilities of transversal oncology.

Regarding the germline analysis, most panels, as ours, include both high-penetrance and moderate-penetrance genes. This increases the chances of finding causal variants, at the price of raising the number of total variants found, many of them of uncertain clinical significance (VUS) (Lincoln *et al.* 2015). Performing careful interpretation of VUS is essential to avoid inappropriate changes on the clinical management of any patient based on the findings of an incorrectly evaluated variant. Additionally, by limiting germline reports to Tier I genes, requestors have the option to restrict the analysis to those genes with the highest and clearest clinical actionability.

The use of a targeted strategy, as opposed to exome- or genome-based approaches, while allowing more comprehensive and deeper coverages of the regions of interest (including introns for rearrangement identification which, in any case, are not covered by the exome), means that the panel needs to be redesigned and re-validated as new knowledge about clinically relevant genomic alterations emerges. This is the reason why we keep in our probe design Tier-3 genes, with suspected but, so far, not validated clinical utility and, thus, more likely to be found to be clinically relevant in the near future than the genes not included in the design at all. Moreover, the combination of genes with clinically relevant somatic and/or germline alterations in our panel also reduces the need for redesign as, not rarely, a gene previously known as a target for clinically relevant somatic alterations is found to be a novel hereditary cancer gene and *vice versa*.

Target-sequence gaps

We have improved the performance and evenness of our capture-probe in a way that the fraction of insufficiently

covered target regions is minimized. Currently, the average fraction of insufficiently covered target-sequence in our somatic panel is 0.043% (ranging from 0.0002% to 0.11%; $n = 13$), with an average sequencing depth of 941-fold (ranging from 353-fold to 1630-fold). Regarding the germline panel, the average fraction of insufficiently covered target sequence is 0.016% (ranging from 0% to 0.1%; $n = 9$), with an average sequencing depth of 959-fold (ranging from 265-fold to 1534-fold). Thus, we find it unnecessary to complete the analysis with Sanger sequencing. Alternatively, we provide a detailed list of the scarce insufficiently covered positions in the report of each case (Fig. S1).

Results verification

Current recommendations advise Sanger verification of clinically relevant variants identified by NGS (Matthijs *et al.* 2016; Susswein *et al.* 2016). In the case of the current test, with sensitivities and specificities above 99.5%, we consider that verification of SNVs and indels might be skipped, specially for the somatic analysis, when the time window for clinical application of the results is very narrow. For the germline analysis, complementary approaches are likely to provide more accurate results for certain variant types (i.e., exon-focused CGH-arrays for single-exon CNV detection or LR PCR-based analyses for target genes with high degree of homology to other parts of the genome, such as *PMS2*) (Gray *et al.* 2015; Retterer *et al.* 2015).

Clinical interpretation

In our series of 36 cases analyzed by our germline sub-panel and considering only genes in Tier-1, 15 cases (41.67%) had variants that were classified as disease causing mutations by HGMD but were degraded to DM-Controversial after expert review. This is in agreement with previous reports (Olsson *et al.* 2015; Groth *et al.* 2016). We recommend performing clinical interpretation by carefully reviewing those publications supporting the HGMD and/or ClinVar classifications. When this is not conclusive, we recommend completing the analysis by searching available publications in PubMed and Google Scholar and following the recommendations provided by the ACMG/CAP guidelines for variant classification (Richards *et al.* 2015).

An additional difficulty for clinical interpretation is dealing with germline pathogenic/likely pathogenic variants in genes for which the risks associated to deleterious genetic alterations are not firmly established. We distinguish between two different situations:

- Pathogenic/Likely pathogenic variants affecting genes with associated cancer risks matching the personal/family history of the proband: the presence of the variant can be considered as an additional risk factor to tailor the proband's cancer surveillance program (which should be designed on a case by case basis, under the light of the cancer history of the proband and the family).
- Pathogenic/Likely pathogenic variants affecting genes with associated cancer risks not matching the personal/family history of the proband: the variant can be considered as a variant of uncertain significance and, thus, not taken into account.

Challenging reporting process

Our reporting strategy shows all the clinically relevant information in the first page of each report with the aim of optimizing data visualization and adapting to the situation and preferences of each patient/requestor (Figs. 2 and S1, S2). For the somatic analysis, this involves displaying the identified alterations together with their associated therapeutic options at first sight, followed by the scientific literature supporting those associations. For the germline analysis, it involves clearly identifying pathogenic, likely pathogenic and uncertain significant variants on the front page, offering a detailed description of each variant in the subsequent pages. A sufficiently explicit technical information (including the analyzed genes and the performance of the test) is also essential for the clinician to understand the relevance of both positive and negative results.

Conclusions

Clinical application of precision oncology is a reality. However, we are still taking the first steps, and difficulties, even beyond the actual NGS testing and reporting procedure, are evident. Platforms such as ONCOgenics are examples of the core building blocks that will be necessary to scale up the field, able to help clinicians, patients and families in the meanwhile of this transformation process. As proven by the genomic alterations explaining certain cases of exceptional responders, the potential for any given cancer patient to benefit from precision medicine is not just speculative or something to consider for the future, but evident and for the present.

Acknowledgments

We thank Prof. Carlos López-Otín for helpful comments. Work performed at IMOMA for this project was partially supported by grants from Fundación María Cristina

Masaveu Peterson and FICYT/PCTI Asturias/FEDER (IE13-121). Work performed at DREAMgenics was partially supported by University of Oviedo Foundation grants (D.C., P.C.P.). CeGeN-PRB2-ISCIII is supported by grant PT13/0001, ISCIII-SGEFI/FEDER, Spain.

Conflicts of Interests

The following authors are currently employed by IMOMA or DREAMgenics, the companies involved in the development and exploitation of the ONCOgenics™ platform: IMOMA: R.C. (Physician in Chief), M.D. (Clinical Molecular Geneticist), G.A.C. (Biotechnologist), N.S.D. (Lab. Technician), R.A. (Lab. Technician) and J.C. (Lab. Director); DREAMgenics: D.C. (Bioinformatician), P.C.P. (Bioinformatician) and G.R.O. (Bioinformatician, C.S.O. and C.E.O.).

References

- Abel, H. J., H. Al-Kateb, C. E. Cottrell, A. J. Bredemeyer, C. C. Pritchard, A. H. Grossmann, et al. 2014. Detection of gene rearrangements in targeted clinical next-generation sequencing. *J. Mol. Diagn.* 16:405–417.
- Adzhubei, I. A., S. Schmidt, L. Peshkin, V. E. Ramensky, A. Gerasimova, P. Bork, et al. 2010. A method and server for predicting damaging missense mutations. *Nat. Methods* 7:248–249.
- Borrás, E., M. Pineda, J. Cadiñanos, J. Del Valle, A. Brieger, I. Hinrichsen, et al. 2013. Refining the role of PMS2 in Lynch syndrome: germline mutational analysis improved by comprehensive assessment of variants. *J. Med. Genet.* 50:552–563.
- Bouffet, E., V. Larouche, B. B. Campbell, D. Merico, R. de Borja, M. Aronson, et al. 2016. Immune checkpoint inhibition for hypermutant glioblastoma multiforme resulting from germline biallelic mismatch repair deficiency. *J. Clin. Oncol.* 34:2206–2211.
- Choi, Y., G. E. Sims, S. Murphy, J. R. Miller, and A. P. Chan. 2012. Predicting the functional effect of amino acid substitutions and indels. *PLoS ONE* 7: e46688.
- Chun, S., and J. C. Fay. 2009. Identification of deleterious mutations within three human genomes. *Genome Res.* 19:1553–1561.
- Clendenning, M., H. Hampel, J. LaJeunesse, A. Lindblom, J. Lockman, M. Nilbert, et al. 2006. Long-range PCR facilitates the identification of PMS2-specific mutations. *Hum. Mutat.* 27:490–495.
- Davydov, E. V., D. L. Goode, M. Sirota, G. M. Cooper, A. Sidow, and S. Batzoglou. 2010. Identifying a high fraction of the human genome to be under selective constraint using GERP++. *PLoS Comput. Biol.* 6:e1001025.
- Dienstmann, R., F. Dong, D. Borger, D. Dias-Santagata, L. W. Ellisen, L. P. Le, et al. 2014. Standardized decision support

- in next generation sequencing reports of somatic cancer variants. *Mol. Oncol.* 8:859–873.
- Dong, C., P. Wei, X. Jian, R. Gibbs, E. Boerwinkle, K. Wang, et al. 2015. Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies. *Hum. Mol. Genet.* 24:2125–2137.
- Panjul-Fernández, M., V. Quesada, R. Cabanillas, J. Cadiñanos, T. Fontanil, A. Obaya, et al. 2013. Cell-cell adhesion genes CTNNA2 and CTNNA3 are tumour suppressors frequently mutated in laryngeal carcinomas. *Nat. Commun.* 4:2531.
- Forbes, S. A., D. Beare, P. Gunasekaran, K. Leung, N. Bindal, H. Boutselakis, et al. 2015. COSMIC: exploring the world's knowledge of somatic mutations in human cancer. *Nucleic Acids Res.* 43:D805–D811.
- Frampton, G. M., A. Fichtenholtz, G. A. Otto, K. Wang, S. R. Downing, J. He, et al. 2013. Development and validation of a clinical cancer genomic profiling test based on massively parallel DNA sequencing. *Nat. Biotechnol.* 31:1023–1031.
- Gagan, J., and E. M. Van Allen. 2015. Next-generation sequencing to guide cancer therapy. *Genome Med.* 7:80.
- Gray, P. N., C. L. Dunlop, and A. M. Elliott. 2015. Not all next generation sequencing diagnostics are created equal: understanding the nuances of solid tumor assay design for somatic mutation detection. *Cancers (Basel)* 7:1313–1332.
- Groth, K. A., M. Gaustadnes, K. Thorsen, J. R. Ostergaard, U. B. Jensen, C. H. Gravholt, et al. 2016. Difficulties in diagnosing Marfan syndrome using current FBN1 databases. *Genet. Med.* 18:98–102.
- Hosein, A. N., S. Song, A. E. McCart Reed, J. Jayanthan, L. E. Reid, J. R. Kutasovic, et al. 2013. Evaluating the repair of DNA derived from formalin-fixed paraffin-embedded tissues prior to genomic profiling by SNP-CGH analysis. *Lab. Invest.* 93:701–710.
- Hyman, D. M., I. Puzanov, V. Subbiah, J. E. Faris, I. Chau, J. Y. Blay, et al. 2015. Vemurafenib in multiple nonmelanoma cancers with BRAF V600 mutations. *N. Engl. J. Med.* 373:726–736.
- Iorio, F., T. A. Knijnenburg, D. J. Vis, G. R. Bignell, M. P. Menden, M. Schubert, et al. 2016. A landscape of pharmacogenomic interactions in cancer. *Cell* 166:740–754.
- Jones, S., V. Anagnostou, K. Lytle, S. Parpart-Li, M. Nesselbush, D. R. Riley, et al. 2015. Personalized genomic analyses for cancer mutation discovery and interpretation. *Sci. Transl. Med.* 7:283ra253.
- van der Klift, H. M., C. M. Tops, E. C. Bik, M. W. Boogaard, A. M. Borgstein, K. B. Hansson, et al. 2010. Quantification of sequence exchange events between PMS2 and PMS2CL provides a basis for improved mutation scanning of Lynch syndrome patients. *Hum. Mutat.* 31:578–587.
- Koboldt, D. C., Q. Zhang, D. E. Larson, D. Shen, M. D. McLellan, L. Lin, et al. 2012. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res.* 22:568–576.
- Kumar, P., S. Henikoff, and P. C. Ng. 2009. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat. Protoc.* 4:1073–1081.
- Kwon, S., S. Park, B. Lee, and S. Yoon. 2013. In-depth analysis of interrelation between quality scores and real errors in Illumina reads. *Conf. Proc. IEEE Eng. Med. Biol. Soc.* 2013:635–638.
- LaDuca, H., A. J. Stuenkel, J. S. Dolinsky, S. Keiles, S. Tandy, T. Pesaran, et al. 2014. Utilization of multigene panels in hereditary cancer predisposition testing: analysis of more than 2,000 patients. *Genet. Med.* 16:830–837.
- Le, D. T., J. N. Uram, H. Wang, B. R. Bartlett, H. Kemberling, A. D. Eyring, et al. 2015. PD-1 blockade in tumors with mismatch-repair deficiency. *N. Engl. J. Med.* 372:2509–2520.
- Li, H., and R. Durbin. 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 26:589–595.
- Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, et al. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics* 25:2078–2079.
- Li, M. M., M. Datto, E. J. Duncavage, S. Kulkarni, N. I. Lindeman, S. Roy, et al. 2017. Standards and guidelines for the interpretation and reporting of sequence variants in cancer: a joint consensus recommendation of the association for molecular pathology, American Society of Clinical Oncology, and College of American Pathologists. *J. Mol. Diagn.* 19:4–23.
- Lincoln, S. E., Y. Kobayashi, M. J. Anderson, S. Yang, A. J. Desmond, M. A. Mills, et al. 2015. A systematic comparison of traditional and multigene panel testing for hereditary breast and ovarian cancer genes in more than 1000 patients. *J. Mol. Diagn.* 17:533–544.
- Mandelker, D., R. J. Schmidt, A. Ankala, K. McDonald Gibson, M. Bowser, H. Sharma, et al. 2016. Navigating highly homologous genes in a molecular diagnostic setting: a resource for clinical next-generation sequencing. *Genet. Med.* 18:1282–1289.
- Matthijs, G., E. Souche, M. Alders, A. Corveleyn, S. Eck, I. Feenstra, et al. 2016. Guidelines for diagnostic next-generation sequencing. *Eur. J. Hum. Genet.* 24:2–5.
- Olsson, E., C. E. Cottrell, N. O. Davidson, C. A. Gurnett, J. W. Heusel, N. O. Stitzel, et al. 2015. Identification of medically actionable secondary findings in the 1000 genomes. *PLoS ONE* 10:e0135193.
- Puente, X. S., M. Pinyol, V. Quesada, L. Conde, G. R. Ordóñez, N. Villamor, et al. 2011. Whole-genome sequencing identifies recurrent mutations in chronic lymphocytic leukaemia. *Nature* 475:101–105.
- Puente, X. S., S. Bea, R. Valdés-Mas, N. Villamor, J. Gutiérrez-Abril, J. I. Martín-Subero, et al. 2015. Non-coding recurrent mutations in chronic lymphocytic leukaemia. *Nature* 526:519–524.

- Quesada, V., L. Conde, N. Villamor, G. R. Ordóñez, P. Jares, L. Bassaganyas, et al. 2012. Exome sequencing identifies recurrent mutations of the splicing factor SF3B1 gene in chronic lymphocytic leukemia. *Nat. Genet.* 44:47–52.
- Ramsay, A. J., V. Quesada, M. Foronda, L. Conde, A. Martínez-Trillos, N. Villamor, et al. 2013. POT1 mutations cause telomere dysfunction in chronic lymphocytic leukemia. *Nat. Genet.* 45:526–530.
- Redig, A. J., and P. A. Janne. 2015. Basket trials and the evolution of clinical trial design in an era of genomic medicine. *J. Clin. Oncol.* 33:975–977.
- Retterer, K., J. Scuffins, D. Schmidt, R. Lewis, D. Pineda-Álvarez, A. Stafford, et al. 2015. Assessing copy number from exome sequencing and exome array CGH based on CNV spectrum in a large clinical cohort. *Genet. Med.* 17:623–629.
- Reva, B., Y. Antipin, and C. Sander. 2011. Predicting the functional impact of protein mutations: application to cancer genomics. *Nucleic Acids Res.* 39:e118.
- Richards, S., N. Aziz, S. Bale, D. Bick, S. Das, J. Gastier-Foster, et al. 2015. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet. Med.* 17:405–424.
- Ritter, D. I., S. Roychowdhury, A. Roy, S. Rao, M. J. Landrum, D. Sonkin, et al. 2016. Somatic cancer variant curation and harmonization through consensus minimum variant level data. *Genome Med.* 8:117.
- Rizvi, N. A., M. D. Hellmann, A. Snyder, P. Kvistborg, V. Makarov, J. J. Havel, et al. 2015. Cancer immunology. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. *Science* 348:124–128.
- Schrader, K. A., D. T. Cheng, V. Joseph, M. Prasad, M. Walsh, A. Zehir, et al. 2016. Germline variants in targeted tumor sequencing using matched normal DNA. *JAMA Oncol.* 2:104–111.
- Schwarz, J. M., D. N. Cooper, M. Schuelke, and D. Seelow. 2014. MutationTaster2: mutation prediction for the deep-sequencing age. *Nat. Methods* 11:361–362.
- Shihab, H. A., J. Gough, M. Mort, D. N. Cooper, I. N. Day, and T. R. Gaunt. 2014. Ranking non-synonymous single nucleotide polymorphisms based on disease concepts. *Hum. Genomics* 8:11.
- Susswein, L. R., M. L. Marshall, R. Nusbaum, K. J. Vogel Postula, S. M. Weissman, L. Yackowski, et al. 2016. Pathogenic and likely pathogenic variant prevalence among the first 10,000 patients referred for next-generation cancer panel testing. *Genet. Med.* 18:823–832.
- Valdes-Mas, R., S. Bea, D. A. Puente, C. Lopez-Otin, and X. S. Puente. 2012. Estimation of copy number alterations from exome sequencing data. *PLoS ONE* 7:e51422.
- Van Allen, E. M., D. Miao, B. Schilling, S. A. Shukla, C. Blank, L. Zimmer, et al. 2015. Genomic correlates of response to CTLA-4 blockade in metastatic melanoma. *Science* 350:207–211.
- Vaughn, C. P., J. Robles, J. J. Swensen, C. E. Miller, E. Lyon, R. Mao, et al. 2010. Clinical analysis of PMS2: mutation detection and avoidance of pseudogenes. *Hum. Mutat.* 31:588–593.
- Wagle, N., B. C. Grabiner, E. M. Van Allen, E. Hodis, S. Jacobus, J. G. Supko, et al. 2014. Activating mTOR mutations in a patient with an extraordinary response on a phase I trial of everolimus and pazopanib. *Cancer Discov.* 4:546–553.
- Wall, J. D., L. E. Tang, B. Zerbe, M. N. Kvale, P. Y. Kwok, C. Schaefer, et al. 2014. Estimating genotype error rates from high-coverage next-generation sequence data. *Genome Res.* 24:1734–1739.
- Zhong, X., H. Yang, S. Zhao, Y. Shyr, and B. Li. 2015. Network-based stratification analysis of 13 major cancer types using mutations in panels of cancer genes. *BMC Genom.* 16(Suppl. 7):S7.

Supporting Information

Additional Supporting Information may be found online in the supporting information tab for this article:

Figure S1. Front page of the somatic sample report.

Figure S2. Front page of the germline sample report.

Figure S3. Detailed information on germline variants.

Figure S4. Causes of false negative calling of variants in highly homologous *PMS2* regions.

Figure S5. Sanger validation of the mosaic *NFI* c.2033dupC mutation.

Table S1. Total callabilities in ONCOgenics v1 and v2.

Table S2. Callabilities at “poor” regions in ONCOgenics v1 and v2.

Table S3. Total callabilities in ONCOgenics v3 and v4.

Table S4. Variants detected in tumor #1 (white background) and tumor #2 (gray background). All the detected variants were validated by Sanger sequencing (PPV = 100%; 82/82).

Table S5. Variants previously identified by SNP-arrays in two patients and used to calculate the sensitivity of the platform on mixtures of two tumors originated from the same patients.

Table S6. Positions previously identified as negative by SNP-arrays in two patients and used to calculate the specificity of the platform on mixtures of two tumors originated from the same patients.

Table S7. High homology regions affecting *PMS2* exons (as per Mandelker et al.).

Table S8. *PMS2* variants called after aligning ONCOgenics *Germline*TM results against the whole reference genome.

Table S9. Potential PMS2 variants called in the exon 11-3'UTR interval after aligning ONCOgenics *Germline*TM results against the PMS2 locus alone.

Table S10. ALK, ROS1 and RET translocations analyzed by ONCOgenics v4.

Table S11. Q/N ratio of libraries with DNA extracted from formaldehyde-fixed specimens.

Table S12. True somatic versus germline origin of 68 variants reported in 6 Foundation One tests.

Table S13. Somatic alterations associated with approved targeted cancer therapies reported by Foundation One/One Heme tests performed on samples from 23 patients of our clinic.

Artículo II: Comprehensive genomic diagnosis of nonsyndromic and syndromic hereditary hearing loss in Spanish patients

Referencia:

Rubén Cabanillas*, [Marta Diñeiro*](#), Guadalupe A. Cifuentes, David Castillo, Patricia C. Pruneda, Rebeca Álvarez, Noelia Sánchez-Durán, Raquel Capín, Ana Plasencia, Mónica Viejo-Díaz, Noelia García-González, Inés Hernando, José L. Llorente, Alfredo Repáraz-Andrade, Cristina Torreira-Banzas, Jordi Rosell, Nancy Govea, Justo Ramón Gómez-Martínez Faustino Núñez-Batalla, José A. Garrote, Ángel Mazón-Gutiérrez, María Costales, María Isidoro-García, Belén García-Berrocal, Gonzalo R. Ordóñez y Juan Cadiñanos. **Comprehensive genomic diagnosis of nonsyndromic and syndromic hereditary hearing loss in Spanish patients.** *BMC Medical Genomics*. 2018; 9;11(1):58. Doi: 10.1186/s12920-018-0375-5.

*R. C. y M. D. contribuyeron por igual en este trabajo

Factor de Impacto:

2,568, Q3 Genetics & Heredity (JCR 2018)

Resumen:

La hipoacusia neurosensorial (HNS) es la discapacidad sensorial más frecuente. La secuenciación de nueva generación (NGS) se ha convertido en la aproximación estándar para el diagnóstico etiológico de la HNS de inicio temprano. Sin embargo, la selección precisa de regiones genómicas diana (panel de genes/exoma/genoma), el rendimiento analítico y la interpretación clínica de las variantes siguen representando los principales impedimentos para su implementación clínica.

Hemos desarrollado un nuevo panel NGS con 199 genes asociados a HNS no sindrómica y/o sindrómica y hemos evaluado su sensibilidad y su especificidad analítica sobre 1.624 variantes de un solo nucleótido (SNVs) e indels conocidas presentes en una mezcla de ADN genómico de 10 líneas celulares linfoblastoides previamente caracterizadas. Asimismo, hemos analizado el ADN genómico de 50 pacientes españoles con HNS presuntamente hereditaria no causada por mutaciones en *GJB2/GJB6*, *OTOF* ni *MT-RNR1*.

La sensibilidad analítica de la prueba para detectar SNVs e indels en la mezcla de ADN de las líneas celulares fue > 99,5 %, con una especificidad > 99,9 %. El rendimiento diagnóstico en los pacientes con HNS fue del 42 % (21/50): 47,6 % (10/21) con patrón de herencia autosómico recesivo (con mutaciones en los genes *BSND*, *CDH23*, *MYO15A*, *STRC* [n = 2], *USH2A* [n = 3], *RDX* y *SLC26A4*); 38,1 % (8/21) con modo de herencia autosómico dominante (con mutaciones en genes *ACTG1* [n = 3; 2 *de novo*], *CHD7*, *GATA3* [*de novo*], *MITF*, *P2RX2* y *SOX10*) y 14,3 % (3/21) con herencia ligada al cromosoma X (con mutaciones en los genes *COL4A5* [*de novo*], *POU3F4* y *PRPS1*). El 46,9 % de las variantes causantes de hipoacusia (15/32) no

estaban presentes en las bases de datos. El 28,6 % de los casos diagnosticados genéticamente (6/21) tenían síndromes previamente no detectados (Barakat, Usher tipo 2A [n = 3] y Waardenburg [n = 2]). El 19 % de los diagnósticos genéticos (4/21) fueron atribuibles a grandes deleciones/duplicaciones (delección de *STRC* [n = 2]; duplicación parcial de *CDH23*; delección del exón 2 de *RDX*).

En la era de la medicina de precisión, es imperativo obtener un diagnóstico etiológico de la HNS. Con estos resultados hemos contribuido a demostrar que, con la metodología adecuada, la NGS puede trasladarse a la práctica clínica, impulsando el rendimiento del diagnóstico genético de la HNS al 50-60 % (considerando las alteraciones en *GJB2/GJB6*), mejorando la precisión diagnóstica/pronóstica, refinando el asesoramiento genético y reproductivo y revelando síndromes relevantes no diagnosticados clínicamente.

Aportación personal al trabajo:

En este trabajo participé activamente en el diseño del panel y en sus periódicas revisiones. Además de realizar gran parte del trabajo experimental y de interpretación de datos, contribuí a la redacción del manuscrito y a la elaboración de la mayoría de las tablas.

RESEARCH ARTICLE

Open Access



Comprehensive genomic diagnosis of non-syndromic and syndromic hereditary hearing loss in Spanish patients

Rubén Cabanillas^{1*}, Marta Diñeiro^{1†}, Guadalupe A. Cifuentes¹, David Castillo², Patricia C. Pruneda², Rebeca Álvarez¹, Noelia Sánchez-Durán¹, Raquel Capín¹, Ana Plasencia³, Mónica Viejo-Díaz³, Noelia García-González³, Inés Hernando³, José L. Llorente³, Alfredo Repáraz-Andrade⁴, Cristina Torreira-Banzas⁴, Jordi Rosell⁵, Nancy Govea⁵, Justo Ramón Gómez-Martínez³, Faustino Núñez-Batalla³, José A. Garrote⁶, Ángel Mazón-Gutiérrez⁷, María Costales^{3,7}, María Isidoro-García⁸, Belén García-Berrocal⁸, Gonzalo R. Ordóñez² and Juan Cadiñanos¹

Abstract

Background: Sensorineural hearing loss (SNHL) is the most common sensory impairment. Comprehensive next-generation sequencing (NGS) has become the standard for the etiological diagnosis of early-onset SNHL. However, accurate selection of target genomic regions (gene panel/exome/genome), analytical performance and variant interpretation remain relevant difficulties for its clinical implementation.

Methods: We developed a novel NGS panel with 199 genes associated with non-syndromic and/or syndromic SNHL. We evaluated the analytical sensitivity and specificity of the panel on 1624 known single nucleotide variants (SNVs) and indels on a mixture of genomic DNA from 10 previously characterized lymphoblastoid cell lines, and analyzed 50 Spanish patients with presumed hereditary SNHL not caused by *GJB2/GJB6*, *OTOF* nor *MT-RNR1* mutations.

Results: The analytical sensitivity of the test to detect SNVs and indels on the DNA mixture from the cell lines was > 99.5%, with a specificity > 99.9%. The diagnostic yield on the SNHL patients was 42% (21/50): 47.6% (10/21) with autosomal recessive inheritance pattern (*BSND*, *CDH23*, *MYO15A*, *STRC* [*n* = 2], *USH2A* [*n* = 3], *RDX*, *SLC26A4*); 38.1% (8/21) autosomal dominant (*ACTG1* [*n* = 3; 2 de novo], *CHD7*, *GATA3* [de novo], *MTF*, *P2RX2*, *SOX10*), and 14.3% (3/21) X-linked (*COL4A5* [de novo], *POU3F4*, *PRPS1*). 46.9% of causative variants (15/32) were not in the databases. 28.6% of genetically diagnosed cases (6/21) had previously undetected syndromes (Barakat, Usher type 2A [*n* = 3] and Waardenburg [*n* = 2]). 19% of genetic diagnoses (4/21) were attributable to large deletions/duplications (*STRC* deletion [*n* = 2]; partial *CDH23* duplication; *RDX* exon 2 deletion).

Conclusions: In the era of precision medicine, obtaining an etiologic diagnosis of SNHL is imperative. Here, we contribute to show that, with the right methodology, NGS can be transferred to the clinical practice, boosting the yield of SNHL genetic diagnosis to 50–60% (including *GJB2/GJB6* alterations), improving diagnostic/prognostic accuracy, refining genetic and reproductive counseling and revealing clinically relevant undiagnosed syndromes.

Keywords: Hereditary, Hearing loss, Precision, Diagnostics, NGS, Gene panel

* Correspondence: rcabanillas@imoma.es; jcb@imoma.es

†Rubén Cabanillas and Marta Diñeiro contributed equally to this work.

¹Instituto de Medicina Oncológica y Molecular de Asturias (IMOMA) S. A., Avda. Richard Grandío s/n, 33193 Oviedo, Spain

Full list of author information is available at the end of the article



© The Author(s). 2018 **Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated.

Background

Congenital profound deafness affects ~ 1 in 1000 live births and an additional 1 in 1000 children will suffer from hearing loss (HL) before becoming adult [1]. Up to 60% of congenital/early-onset sensorineural HL (SNHL) is caused by genetic factors and often appears in the absence of a family history for deafness. Although alterations in the *GJB2* and *GJB6* genes (*DNFB1* locus) account for a large proportion of cases in different populations (10–40%) [2, 3], many cases remain undiagnosed after *GJB2/GJB6* testing. This is not surprising given the extreme genetic and phenotypic heterogeneity of HL, with more than 400 syndromes that include HL as a feature and more than 100 genes associated with nonsyndromic SNHL [1]. With next-generation sequencing (NGS) technology, it has become feasible and affordable to routinely sequence a large number of genes per patient [4]. Therefore, genetic diagnosis of SNHL has evolved from single-mutation Sanger sequencing to comprehensive multi-gene testing, and NGS has become the new standard of care [5].

Accordingly, once a case of newborn SNHL is confirmed, testing for congenital cytomegalovirus infection and NGS are recommended [6]. Obtaining a SNHL genetic diagnosis has a number of advantages for patients and parents [7]: it provides information about genetic heritability; it helps diagnosing or excluding syndromic causes of HL to better define medical and educational needs; it can also provide information about the evolution of the HL and/or of its associated syndromic features, improving prognostic accuracy; and it can prevent other unnecessary and costly testing [5]. Furthermore, a genetic diagnosis can contribute to prevent triggers such as aminoglycosides in mitochondrial mutation carriers [8], or even to improve treatment selection, including future mutation-driven clinical trials [9, 10].

In order to implement targeted NGS into the clinical practice, there is an urgent need to solve a number of issues such as the selection of the most efficient gene panel, the achievement of high analytical specificity and sensitivity, and the establishment of pipelines able to unambiguously define the clinical impact of genetic variants [11]. Therefore, the aims of this study are: (1) to present the development and validation of a NGS-based approach for the genetic diagnosis of patients with hereditary syndromic and non-syndromic SNHL; (2) to pinpoint and resolve the main problems associated with the introduction of targeted NGS into routine deafness diagnostics; (3) to evaluate the panel's performance and diagnostic yield; and (4) to initiate a comprehensive catalogue of the Spanish genome-wide SNHL variation spectrum.

Methods

Purpose of test

The aim of the performed test (OTOgenics™) was to detect the molecular basis of individual clinical diagnoses

of sensorineural or mixed hearing loss after non-genetic causes had been explored and not identified.

Design of panel content: Rationale for inclusion of specific genes

Genes associated with prelingual, postlingual and adult-onset sensorineural or mixed HL, either symmetric or asymmetric, irrespective of the pattern of inheritance, and including both syndromic and non-syndromic forms, were considered. To generate a preliminary gene list, the professional version of the Human Gene Mutation Database (HGMD) was queried to identify genes associated to HL, using as search keywords a list of phenotypes potentially related to hearing defects (Additional file 1). The resulting gene list was manually curated by analysis of the literature and information available in the databases (HGMD, OMIM, PubMed, GeneReviews, and the Hereditary Hearing Loss Homepage; last accessed 19/09/2017) to identify those fulfilling the following criteria: I) the gene had been associated to sensorineural and/or mixed HL phenotypes (as opposed to exclusively conductive HL), II) there existed published evidence supporting the gene-phenotype association in at least two independent families and III) at least one of the existing publications demonstrated convincing cosegregation of phenotype with gene variants. Based on the curation results, a tiered classification system was devised as previously proposed by Abou Tayoun et al. [11]. Genes with strong/moderate association with HL (fulfilling criteria I, II and III described above, and corresponding mainly to Evidence level 3 according to Abou Tayoun et al. [11]) formed tier 1, while genes with weak/preliminary association (fulfilling criterion I, but not criteria II and/or III, and corresponding mainly to Evidence level 2 according to Abou Tayoun et al. [11]) were grouped in tier 2. The panel evolved with revision of newly published literature, yielding versions v1, v2 and v3 (Additional file 2, Additional file 3 and Table 1, respectively). v1–2 were used during the research and development phase of the study, whereas v3 was considered the first clinical-grade version of the panel. 32 cases were analyzed with v1, 13 with v2 and 5 with v3 (Additional file 4).

Sample types

4 ml of peripheral blood in conventional EDTA-tubes or ≥ 200 ng of germline genomic DNA (quantitated by a fluorimetric method) were required per patient.

Sample preparation and evaluation of genomic DNA integrity

Germline genomic DNA was isolated as previously described [12] and calculation of its DNA integrity number (DIN) was performed using the Genomic DNA ScreenTape Assay on a TapeStation 4200 system (Agilent Technologies,

Table 1 Tier 1 and tier 2 genes included in the OTogenics™ panel (v3)

Tier 1 (154 genes) genes with strong/moderate association with SNHL*											
ABHD12	B5ND	CLRN1	DIABLO	MARVELD2	MYO3A	P2RX2	PTPN11	SLC32A3	TNME		
NM_001042472.2	NM_057176.2	NM_174878.2	NM_019887.5	NM_001080432.2	NM_017433.4	NM_174873.2	NM_00284.3	NM_033409.3	NM_147196.2		
ACTB	CABP2	COCH	DIAPH1	GATA3	MYO6	PAX3	PTPRQ	SLITRK6	TMPRSS3		
NM_001101.3	NM_016366.2	NM_004086.2	NM_005219.4	NM_00102295.1	NM_004999.3	NM_181457.3	NM_00145026.1	NM_032229.2	NM_034022.2		
ACTG1	CACNA1D	COL2A1	DNM1T1	GIPC3	KARS	MIR96	MYO7A	SMPX	TPRN		
NM_001614.3	NM_000720.3	NM_001844.4	NM_00130623.1	NM_133261.2	NM_001130089.1	NR_029512.1	NM_000260.3	NM_001453.2	NM_00128228.2		
ADGRV1	CDC450	COL4A3	ECF1S1	GJB2	KCNE1	MTF	MYO15A	SNK2	TRIOBP		
NM_032119.3	NM_178355.2	NM_000091.4	NM_004092.3	NM_000404.5	NM_000219.5	NM_000248.3	NM_016239.3	NM_000306.3	NM_001039141.2		
AFMT1	CDH23	COL4A4	EDN3	GJB3	KCNJ10	MSRB3	NARS2	SOX10	TSPPEAR		
NM_004208.3	NM_022124.5	NM_000092.4	NM_207034.2	NM_024009.2	NM_002241.4	NM_198080.3	NM_024678.5	NM_001799.3	NM_144891.2		
ALMS1	CEACAM16	COL4A5	EDNRB	GJB6	KCNQ1	MT-CO1	NDP	SERAC1	USH1C		
NM_015120.4	NM_001039213.3	NM_00495.4	NM_000115.3	NM_006783.4	NM_000218.2	NC_012920.1	NM_000266.3	NM_032861.3	NM_005709.3		
ANKK	CHD7	COL9A1	EPS8L2	GPSM2	KCNQ4	MT-RNR1	NLRP3	SERPINB6	USH1G		
NM_094027.4	NM_017780.3	NM_001851.4	NM_022772.3	NM_013296.4	NM_004700.3	NC_012920.1	NM_004895.4	NM_153700.2	NM_173477.4		
AP15T	CIB2	COL11A1	ESPN	GRHL2	LARS2	MT-TH	OPA1	SK1	USH2A		
NM_001283.3	NM_006383.3	NM_001854.3	NM_031475.2	NM_024915.3	NM_015340.3	NC_012920.1	NM_01560.2	NM_005982.3	NM_206933.2		
ATP1A3	CISD2	COL11A2	ESRRB	GRXCR1	LHPPL5	MT-TK	OSBP2	SLC17A8	WFS1		
NM_152296.4	NM_00100888.4	NM_080680.2	NM_004652.3	NM_001080476.2	NM_182548.3	NC_012920.1	NM_144468.2	NM_19319.2	NM_006005.3		
ATP6V1B1	CLCNKA	DCAF17	EYA1	HARS2	LHX3	MT-TL1	OTDA	SLC19A2	WHRN		
NM_001692.3	NM_004070.3	NM_025000.3	NM_000503.5	NM_012208.3	NM_014564.3	NC_012920.1	NM_146672.3	NM_005422.2	NM_015404.3		
BCAP31	CLCNKB	DDX11	EYA4	HGF	LOXHD1	MT-TS1	OTOF	SLC26A4	XYLT2		
NM_001139441.1	NM_000085.4	NM_030653.3	NM_004100.4	NM_000601.4	NM_144612.6	NC_012920.1	NM_194248.2	NM_004411.1	NM_022167.3		
BCS1L	CLDN14	DFNA5	FGF3	HOKA1	LRP2	MYH9	OTOG	SLC33A1			
NM_004328.4	NM_144492.2	NM_004403.2	NM_005247.2	NM_005522.4	NM_004525.2	NM_002473.5	NM_001277269.1	NM_04733.3			
BRAF	CLPP	DFNB59	FGFR3	HOBX1	LRTOMT	MYH14	OTDGL	TMC1			
NM_004333.4	NM_08012.2	NM_001042702.3	NM_000142.4	NM_002144.3	NM_00145308.4	NM_024729.3	NM_173591.3	NM_024531.4	NM_138691.2		
Tier 2 (45 genes) genes with weak/preliminary association with SNHL*											
ADCY1	COL4A6	CRYM	ELMO3	FGFR1	GTF2IRD1	ITILG	MT-CO3	TK2	TPK3		
NM_001116.2	NM_001847.3	NM_001888.4	NM_03213.4	NM_03110.2	NM_016238.2	NM_000899.4	NC_012920.1	NM_175975.4	NM_004614.4	NM_03722.4	
ATP2B2	COL9A2	DCDC2	EP58	FGFR2	HMX2	MAF	MT-TA	SLC4A11	TMEM132E		
NM_001683.3	NM_001852.3	NM_016556.4	NM_00447.5	NM_000141.4	NM_00519.1	NM_005360.4	NC_012920.1	NM_032084.3	NM_001304581.1		
ATP6V1B2	COL9A3	DIAPH3	FAM65B	FOXJ1	HMX3	HARS2	MT-TE	SLC9A1	TMPRSS5		
NM_001693.3	NM_001863.3	NM_001042517.1	NM_014722.3	NM_012188.4	NM_01105574.1	NC_012920.1	NC_012920.1	NM_033109.4	NM_030770.3		
BDF1	CO6	DSPP	FBLN1	GRKCR2	HOWER2	HCMD2	MT-TS2	SLC28A5	TNC		
NM_016429.2	NM_184276.2	NM_014208.3	NM_006486.2	NM_001080516.1	NM_004689.3	NM_004526.3	NC_012920.1	NM_19899.2	NM_002160.3		

*Sensoneutral hearing loss

Santa Clara, CA, USA), following the manufacturer's instructions.

Library preparation, target enrichment and sequencing

Library preparation was performed on genomic DNA physically sheared by ultrasonication on a Covaris S2 instrument (Covaris, MA, USA). For library construction and gene target enrichment by hybrid capture, the SureSelectXT protocol was followed, as previously described [12]. This approach has a series of advantages over other library construction and target enrichment methods. Thus, libraries from randomly fragmented DNA show higher complexities than PCR-based ones, enabling the identification and removal of PCR duplicates (important for the accurate identification of low frequency variants present in mosaic patients) [13]. Additionally, capture probes, although laborious to use, are more tolerant to mismatches than PCR primers, circumventing issues of allelic dropout (underrepresentation or absence of an existing allele in the library) caused by polymorphisms in the hybridization sequence that can be observed in amplification-based assays [13]. Finally, capturing and sequencing at least all coding exons of every targeted gene, and not just hotspots, facilitates creating background references for CNV calling. Sequencing was performed on a NextSeq500 sequencer (Illumina, CA, USA), following the manufacturers specifications. The optimized NGS diagnostic pipeline (OTOgenics™) targets the coding exons and intron-exon junctions of 199 genes (v3) (Table 1).

Bioinformatics for variant identification and annotation

NGS results were processed using the bioinformatics software HD Genome One (DREAMgenics, Oviedo, Spain), certified with IVD/CE-marking. The pipeline has been adapted from the one previously described as part of the ONCOgenics NGS platform [12], the performance of which has been externally evaluated through participation in the Oncogene Panel Testing schemes organized by the European Molecular Genetics Quality Network (EMQN), obtaining satisfactory results (maximum genotyping score) for three consecutive years (2015, 2016 and 2017). The analysis workflow was as follows:

FASTQ read generation, alignment and duplicate removal

FASTQ reads were generated from base call files (BCL) produced by the Illumina NextSeq500 sequencing platform using the bcl2fastq v2.19 Conversion Software (https://support.illumina.com/sequencing/sequencing_software/bcl2fastq-conversion-software.html). Raw FASTQ files were evaluated using quality control checks from FastQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) and Trimmomatic was employed to remove low quality bases, adapters and other technical sequences. Each FASTQ file was aligned to the human reference genome (GRCh37/

hg19 before 2017; GRCh38/hg38 afterwards) using BWA-mem [14] generating sorted BAM files with SAMtools [15]. Reads from the same libraries were then merged and optical and PCR duplicates were removed using Picard (<http://broadinstitute.github.io/picard/>).

SNV/Indel identification

SNVs and indels were identified using a variation of Sidrón algorithm, previously described [16], with the following parameters: total read depth ≥ 6 , mutated allele count ≥ 3 , variant frequency ≥ 0.1 , base quality ≥ 20 , mapping quality ≥ 30 . Stricter criteria (total read depth ≥ 10 , mutated allele count ≥ 4) were applied before the selection of reportable variants. Manual inspection was then carried out to discard false positives and avoid missing true variants not meeting those criteria (i.e long indels with underestimated frequencies).

CNV identification

The detection of CNVs was performed with an adapted version of the exome2cnv algorithm, incorporating a combination of read depth and allelic imbalance computations for copy number assessment. The algorithm employs a background of pooled samples processed using the same capturing protocol and sequencing technology [12, 17]. For increased sensitivity in the detection of large homozygous deletions, genomic regions with no sequencing coverage in an individual sample, but showing proper coverages in the remaining samples, were identified.

Variant annotation

Variants were annotated using several databases containing functional (Ensembl, CCDS, RefSeq, Pfam), population (dbSNP, 1000 Genomes, ESP6500, ExAC) and disease-related (Clinvar, HGMD professional) information, as well as 12 scores from algorithms for prediction of the impact caused by nonsynonymous variants on the structure and function of the protein (SIFT [18], PolyPhen2 [19], PROVEAN [20], Mutation Assessor [21], Mutation Taster [22], LRT [23], MetaLR, MetaSVM [24], FATHMM, FATHMM-MKL [25] and M-CAP [26]), and 1 score (GERP++) for evolutionary conservation of the affected nucleotide [27].

Analytical sensitivity and specificity

The analytical sensitivity and specificity of our panel to detect SNVs/indels was calculated using the v3 probe set to evaluate 1624 total variants (1503 SNVs + 121 indels) with allelic frequency ≥ 0.1 , following a procedure similar to that previously described [12]. Briefly, 10 immortal lymphoblastoid cell lines, corresponding to 10 individuals whose genomes/exomes had been sequenced by the 1000 Genomes and HapMap projects, were obtained from the Coriell Institute: NA20298 (ASW), NA12872 (CEU), NA18570 (CHB), HG00320 (FIN), HG00110 (GBR), A18960 (JPT), NA19020

(LWK), NA19794 (MXL), HG00740 (PUR) and NA18486 (YRI). Cell lines were cultured according to the protocols provided by Coriell, their DNAs isolated and mixed in equimolar amounts. An NGS library was prepared, captured using the custom probe and sequenced in 20% of a NextSeq500 MidOutput run (2 × 75 sequencing cycles). Variants were called as described in the previous section and the results compared to those expected according to the genomic information available for these cell lines.

Interfering highly homologous regions

Regions with < 100% callability at DP20 (i.e. with less than 100% of the target bases covered by ≥ 20 reads) in > 50% of v3 cases were considered as potential conflictive regions (Additional file 5). Those showing high homology to at least another region of the GRCh38 human reference genome are listed in Additional file 6. Realignment of the NGS results with reference sequences containing only the panel genes affected by these conflictive regions was performed, as previously described for the *PMS2* gene [12], followed by validation, using gene-specific analyses (i.e. Long-Range PCR followed by Sanger sequencing) of putative pathogenic/likely pathogenic SNV and indel variants [28].

Variant filtering, interpretation, classification and diagnostic yield

Database resources to evaluate gene variants included HGMD professional, OMIM, PubMed, dbSNP, 1000 Genomes Project, ESP, ExAC, and ClinVar. A minor allele frequency (MAF) cut-off of 5% was applied to variants considered as pathogenic (DM) by HGMD or as pathogenic/likely pathogenic by ClinVar, as well as to those variants predicted to create a null allele (nonsense, frameshift causing premature STOP codons, canonical splicing-site disruption, ATG-loss and complete exon deletions/duplications). A MAF cut-off of 1%, as suggested by Shearer et al. [29], was applied to all other variants predicted to affect the sequence/expression of the protein (for protein-coding genes) or of the RNA (for non-coding genes).

Clinical classification of all variants from v1 and v2 cases was performed as described [12]. For v3 cases, only variants that could potentially explain the SNHL phenotype of the probands, based on zygosity of the variant, presence of additional variants on the same gene and mode of inheritance of the audiological phenotypes associated to the gene, were further considered. After that, variants were clinically classified according to the American College of Medical Genetics and Genomics (ACMG) guidelines as pathogenic (class 5), likely pathogenic (class 4), uncertain significance (class 3), likely benign (class 2) and benign (class 1) [30]. Class 3–5 variants were thoroughly curated searching the literature and databases for clinically relevant data. Class 3–5 variants were reported for tier 1 genes. To reduce the interpretation burden of tier 2 genes, only class 4–5 variants affecting

genes whose associated putative phenotype matched the patient's phenotype were evaluated and reported.

Diagnostic yield (generally described as the likelihood that a test or procedure will provide the information needed to establish a diagnosis) was defined as the percentage of tested patients with pathogenic/likely pathogenic variants capable of explaining their HL phenotype.

Variant validation

Pathogenic/likely pathogenic variants considered responsible for SNHL were validated by approaches alternative to NGS. SNVs/indels were validated by PCR + Sanger sequencing. For validation and breakpoint identification of the partial *CDH23* duplication (exons 11–15) and the *RDX* exon 2 deletion multiple PCR reactions were performed. *STRC* CNVs were validated with a qPCR assay able to distinguish *STRC* from the *pSTRC* pseudogene [28] as well as by MLPA (MRC-Holland, Amsterdam, The Netherlands; cat.# P64-DIS). Primers used in validation PCRs are described in Additional file 7). Segregation was determined in 8/8 (100%) relatives with available biospecimens.

Reportable range

Reportable variants had to be supported by ≥ 4 independent reads, with a total read depth ≥ 10, belonging to genes from Tier 1 (consistently associated) and showing allelic frequencies in the sample ≥ 0.1. Variants considered responsible or likely responsible for the phenotype of the patient additionally needed to have been validated by a method alternative to NGS to be considered reportable. Occasionally, a variant from a Tier 2 gene fulfilling all other criteria could be reportable, as long as the phenotype of the patient was compatible with the phenotype considered in the existing publications supporting the gene-phenotype association of the Tier 2 gene.

Reference range

Only variants consistent with the mode of inheritance of the auditory phenotypes associated with the gene they affect (for instance, biallelic variants on a gene with a recessive phenotype) were evaluated according to ACMG/AMP guidelines and their resulting clinical classification was reported [30]. For those classified as pathogenic, likely pathogenic or VUS, additional information supporting their clinical classification was provided. Those classified as benign or likely benign were considered to lay within the reference range of results and, thus, no further details about them were provided. The remaining variants (not consistent with the mode of inheritance of the auditory phenotypes associated with the genes they affect) were included in the reports for informative purposes but were not considered responsible for the patient's phenotype.

Sample tracking

A series of 6 SNPs on Tier 1 genes, with population MAFs between 0.463 and 0.483, were selected for sample tracking: rs10864198 on *USH2A* (MAF = 0.4531; ExAC); rs7598901 on *ALMS1* (MAF = 0.4736, 1000 Genomes); rs2228557 on *COL4A4* (MAF = 0.4657; ExAC), rs7624750 on *OPA1* (MAF = 0.4683; ExAC), rs734312 on *WFS1* (MAF = 0.4633; ExAC), and rs2438349 on *ADGRV1* (MAF = 0.4830; ExAC). The genotypes identified by the NGS pipeline were compared to those obtained from the corresponding TaqMan qPCR genotyping assays (cat #: C_31803731_10, C_29307975_10, C_11523965_10, C_2715859_10, C_2401729_1 and C_16236492_10, respectively; Applied Biosystems, CA, USA) run on a 7900HT Fast Real-Time PCR System (Applied Biosystems, CA, USA). All samples showed coincident genotypes for all SNPs on both platforms.

Patient population

Between September 2014 and March 2017, 50 consecutive patients (21 male, 29 female) with syndromic/non-syndromic SNHL were selected after excluding non-genetic causes and causative variants in the *DENB1* (*GJB2/GJB6*), *OTOF* and *MT-RNR1* loci, considered the most frequent causes of hereditary deafness in Spain [7, 31]. Consent was obtained from patients or their parents. The study was approved by the Comité de Ética de Investigación del Principado de Asturias (research project #75/14). The ages at SNHL onset ranged between 0 to 47 years (median: 12 years). 20 cases (40%) were congenital. To identify syndromic SNHL, a clinical geneticist evaluated the patients. 2/50 patients were diagnosed (pre-test) of Alport and CHARGE syndromes. Other 3 patients presented with potentially syndromic complications, without fulfilling criteria for known syndromes.

Results

Panel validation

Performance of targeted NGS

Mean coverage of tier 1 genes was 445× for v1, 515× for v2 and 1121× for v3, and 98.87, 99.56 and 99.95% of their target bases were covered by 20 or more reads, respectively (these calculations exclude the *STRC* and *OTOA* genes due to their high homology to other genomic regions). The minimum, average and maximum coverage (average read depth of all target bases of the gene) and callabilities (% of the target bases of the gene with minimum read depths of 10, 20, 50 and 100 reads per each target base of the gene) for every tier 1 and tier 2 gene on samples analysed with OTogenics v3 is shown in Additional file 5. In v3 cases, regions from tier 1 genes with less than 100% coverage with a minimum of 20 reads (DP20) and specific positions within those regions affected by such limitation were included in each individual patient's report.

Analytical sensitivity and specificity

Prior to its use in the diagnostic setting, the clinical version of the panel (v3) was evaluated for sensitivity and specificity on a genotyped mixture of 10 lymphoblastoid cell lines. 1617/1624 variants with allele frequency ≥ 0.1 were detected (1497/1503 SNVs and 120/121 indels), yielding a sensitivity of 0.9957 ($> 99.5\%$). Additionally, 1,034,047/1,034,817 true negative positions of the target region were called by the platform as not bearing SNVs or indels, representing a specificity of 0.9992 ($> 99.9\%$).

Orthogonal validation of sequencing results

All variants considered responsible for the SNHL phenotypes of the probands (Table 2) were successfully validated by approaches alternative to NGS. These included 25 instances of SNVs or indels (validated by PCR and Sanger sequencing) and 4 CNVs: 1 heterozygous partial duplication of *CDH23* (exons 11–15), 1 homozygous deletion of *RDX* exon 2 (both validated by breakpoint-specific PCR) and 2 homozygous *STRC* whole gene deletions (validated by qPCR and MLPA). Apart from these, 4 samples had heterozygous *STRC* CNVs (3 deletions and 1 duplication) all of which were validated by qPCR and MLPA (Additional file 8). These results indicate that our CNV calling procedure is highly specific.

Performance at interfering highly homologous regions

Genomic regions with high sequence homology cause misalignment of sequencing data and represent a major challenge for short-read NGS technologies. Out of the 199 genes included in the v3 panel, *STRC*, *OTOA*, *ESPN* and *KCNE1* contain a total of 22 interfering highly homologous regions (as defined in the Methods section), most of which overlap with those previously identified by Mandelker et al. [32] (Additional file 6). To avoid missing clinically relevant variants present in those target regions, the panel NGS reads from all samples were realigned to reference sequences containing only the *STRC*, *OTOA*, *ESPN* and *KCNE1* loci, as previously described by us for the *PMS2* gene in a cancer panel [12]. This approach revealed that all samples might potentially carry a pathogenic variant in the *STRC* gene: c.4057C > T, p.Gln1353* (coincident with the reference sequence for exon 20 from the *pSTRC* pseudogene). To unequivocally discriminate the origin of this variant, LR-PCR specific for the *STRC* gene followed by Sanger sequencing was performed as described [28]. This approach discarded the potential genic origin of the variant in 28/50 samples; not enough DNA was available from 1 sample and no LR-PCR product was obtained from 21. Of note, the average genomic DNA Integrity Number (DIN) of the 28 samples with successful LR-PCR was significantly higher than that of the remaining 21 samples (8.71 vs 7.21; p -value = 1.8×10^{-4} ; Student's T test), suggesting that DNA degradation precluded LR-PCR. Alternative approaches would be

needed to discard or confirm the genic origin of the variant in those 21 samples.

Analysis of causative variants and diagnostic yield

Of 50 cases with severe-to-profound SNHL not caused by *GJB2/GJB6*, *OTOF* or *MT-RNR1* mutations, a genetic justification for their HL phenotype was found in 21 (42%) after identifying 31 pathogenic/likely pathogenic variants in 16 genes: *ACTG1*, *BSND*, *CDH23*, *CHD7*, *COL4A5*, *GATA3*, *MITE*, *MYO15A*, *P2RX2*, *POU3F4*, *PRPS1*, *RDX*, *SLC26A4*, *SOX10*, *STRC* and *USH2A* (Tables 2 and 3). Three more cases had recessive variants of uncertain significance in homozygosis (affecting the *LOXHD1* and *SLC26A4* genes) or in hemizygosis (affecting the *OTOA* gene: 1 SNV + heterozygous whole-gene *OTOA* deletion), which were suspicious of pathogenicity, but did not fulfill ACMG criteria and, thus, were not counted nor reported as positives (Table 4). Had they been counted, the diagnostic yield would have been 48% (24/50).

In our cohort, 47.6% (10/21) of the unambiguously molecularly diagnosed patients had autosomal recessive (AR) inheritance patterns, 38.1% (8/21) autosomal dominant (AD), and 14.3% (3/21) were X-linked (Table 2). The molecular basis of deafness was found in 44.4% (20/45) of the cases with symmetric SNHL, whereas only 1 of 5 cases with asymmetric SNHL was genetically diagnosed (Waardenburg syndrome caused by a *MITF* mutation) (Tables 2 and 3).

The most common SNHL causative genes in our pre-screened population were *ACTG1* (3 patients), *USH2A* (3 patients) and *STRC* (2 patients). Interestingly, 2 of 3 pathogenic variants in *ACTG1* were de novo, as well as 1 *GATA3* and 1 *COL4A5* pathogenic variants.

CNV analysis identified causative variants in 4 of the 21 molecularly diagnosed patients (19%): 2 with a homozygous complete *STRC* deletion, 1 with a previously unreported partial *CDH23* duplication (exons 11–15) in compound heterozygosity with a second pathogenic variant (missense), and 1 with a homozygous *RDX* exon 2 deletion. One of the patients with a homozygous causative *STRC* deletion was also a carrier of a heterozygous substitution in *TECTA*, previously reported in the literature as a dominant pathogenic variant (c.3107G > A; p.Cys1036Tyr) [33]. However, reevaluation of this variant according to ACMG guidelines reclassified it as a variant of uncertain significance (it only fulfilled ACMG pathogenicity criteria PP1 and PP3).

In total, 451 variants, of which 406 were unique, in 121 distinct genes were identified in the full cohort of 50 patients: 394 variants in 97 genes were identified in tier 1. Tier 2 added 57 variants that contributed to the overall interpretation burden. No tier 2 variant was considered responsible for the SNHL phenotype (Fig. 1).

199/394 (50.5%) and 45/57 (78.9%) of the identified tier 1 and tier 2 variants, respectively, were absent from the

databases (HGMD professional and ClinVar). Fifteen of them (all from tier 1), were classified as pathogenic or likely pathogenic and responsible for the SNHL phenotype of the patient (Fig. 1). Globally, those 15 variants were considered responsible for the SNHL phenotype in 13 cases. As a result, 61.9% of the genetically diagnosed cases (13/21) were explained by variants not described in the databases (Fig. 2 and Table 2). Moreover, of 25 non-redundant variants classified as pathogenic (DM) by HGMD for hearing-related phenotypes, after looking for plausible published support in the literature, solid evidence could only be found for 13 of them (52%) (Additional file 9). Of the 12 variants considered DM by HGMD but, in our view, without enough supporting evidence, 3 (25%) were also considered as Pathogenic/Likely pathogenic by at least one ClinVar submitter (Additional file 9). To deal with these limitations, an average of 40 min of expert review was dedicated per variant. With an average of 9 variants per case, this represents 360 min (6 h) per case. These results highlight the importance of manual interpretation and curation for clinical classification of variants, even for those considered as (potentially) disease causing by reputable databases.

Increase of clinical sensitivity by analysis of syndromic genes on apparently non-syndromic SNHL

28.6% of the genetically diagnosed cases (6/21) had a previously unrecognized syndrome: Barakat (1 patient), Usher type 2A (3 patients) and Waardenburg (2 patients) (Table 2). These unexpected syndromic findings not only increased the diagnostic yield, but they provided diagnostics of utmost clinical relevance. Additionally, 6 patients carried pathogenic variants in genes associated with syndromic and non-syndromic conditions (Table 2): 1 had variants associated with Pendred syndrome and *DFNB4*, 1 with Bartter syndrome type IV and *DFNB73*, 1 with Usher syndrome 1D and *DFNB12* and 3 with Baraitser-Winter type 2 and *DFNA20/DFNA26*. These patients will need close follow-up in case syndromic features develop.

Discussion

Hearing loss is one of the most genetically heterogeneous disorders known. 60% of cases are believed to be of genetic origin and 30% of them syndromic [34]. Due to its high diagnostic yield [35, 36], the newest ACMG guidelines include NGS testing in the standard SNHL diagnostic algorithm [1], whereas the use of non-genetic tests should be considered case-by-case, usually as a complement to genetic testing. However, except for the preeminent relevance of *GJB2* mutations, little is known about the frequency of SNHL variants in Europeans [9, 37].

Our results contribute to define the mutation spectra in the Spanish population, underlining the SNHL genetic heterogeneity, as the causative variants of 21 patients affected 16 different genes. The genes most commonly altered in

Table 2 Clinical and genetic characteristics of cases with causative mutations

Case ID	Pre-test phenotype	Pre-test suspected inheritance pattern	Time of deafness onset	Gene	Allele variants	Variant zygosity	ACMG ^a classification ³⁰	Fulfilled pathogenicity criteria ³⁰	Present in HGMD and/or ClinVar ^b	Gene-associated phenotypes	Inheritance patterns of gene-associated phenotypes	Hidden syndrome
OTO.008	Bilateral non-syndromic sensorineural deafness	AR	Congenital	MYO15A	c.8050 T>C p.(Y172684His) c.8968-1G>T	Heterozygous	Likely pathogenic	PM2, PM3, PP1, PP3 PM1, PM2, PP1	Yes No	Non-syndromic sensorineural deafness (DFNB3)	AR	No
OTO.001	Bilateral non-syndromic sensorineural deafness	AR	Congenital	STRC	Whole-gene deletion	Homozygous	Pathogenic	PM1, PM2, PM3	Yes	Non-syndromic sensorineural deafness (DFNB16)	AR	No
OTO.033	Bilateral non-syndromic sensorineural deafness	AR	Childhood	STRC	Whole-gene deletion	Homozygous	Pathogenic	PM1, PM2, PM3	Yes	Non-syndromic sensorineural deafness (DFNB16)	AR	No
OTO.050	Bilateral non-syndromic sensorineural deafness	AD / AR	Childhood	RDX	Exon 2 deletion	Homozygous	Likely pathogenic	PM1, PM2	No	Non-syndromic sensorineural deafness (DFNB24)	AR	No
OTO.009	Bilateral non-syndromic sensorineural deafness	AR	Congenital	BSND	c.23G>A p.(Arg8Gln)	Homozygous	Likely pathogenic	PM1, PM2, PM5, PP3	No	Non-syndromic sensorineural deafness (DFNB73) / Bartter syndrome type IV	AR	Potential
OTO.006	Bilateral non-syndromic sensorineural deafness	AR	Childhood	SLC26A4	c.412G>T p.(Val138Phe) c.1370A>T p.(Asn457Ile)	Heterozygous	Pathogenic	PM1, PM2, PM3, PP3 PM1, PM2, PM3, PP3	Yes Yes	Non-syndromic sensorineural deafness (DFNB4) / Pendred syndrome	AR	Potential
OTO.018	Bilateral non-syndromic sensorineural deafness	AR	Congenital	CDH23	c.4488G>C p.(Gln496His) Duplication of exons 11–15	Heterozygous	Pathogenic	PM1, PM2, PM3, PP1, PP3 PM2, PM3, PM4	Yes No	Non-syndromic sensorineural deafness (DFNB12) / Usher syndrome type 1D	AR	Potential
OTO.004	Bilateral non-syndromic sensorineural deafness	AR	Childhood	USH2A	c.11864G>A p.(Trp3955*)	Homozygous	Pathogenic	PM1, PM2, PM3	Yes	Usher syndrome type 2A	AR	Yes
OTO.005	Bilateral non-syndromic sensorineural deafness	AR	Congenital	USH2A	c.1724G>A p.(Cys575Yr)	Homozygous	Likely pathogenic	PM1, PM2, PM3, PP3	Yes	Usher syndrome type 2A	AR	Yes
OTO.014	Bilateral non-syndromic sensorineural deafness	AR	Congenital	USH2A	c.1724G>A p.(Cys575Yr) c.1841–2A>G	Heterozygous	Likely pathogenic	PM1, PM2, PM3, PP3 PM1, PM2, PM3, PP3	Yes Yes	Usher syndrome type 2A	AR	Yes

Table 2 Clinical and genetic characteristics of cases with causative mutations (Continued)

Case ID	Pre-test phenotype	Pre-test suspected inheritance pattern	Time of deafness onset	Gene	Allele variants	Variant zygosity	ACMG ^a classification ³⁰	Fulfilled ACMG ^a pathogenicity criteria ³⁰	Present in HGMD and/or ClinVar ^b	Gene-associated phenotypes	Inheritance patterns of gene-associated phenotypes	Hidden syndrome
OTO.003	Bilateral non-syndromic sensorineural deafness	AD	Childhood	P2RX2	c.178G>T p.(Val60Leu)	Heterozygous	Likely pathogenic	PS3, PM2, PP1	Yes	Non-syndromic sensorineural deafness (DFNA41)	AD	No
OTO.043	Bilateral non-syndromic sensorineural deafness	AD	Congenital	ACTG1	c.434C>G p.(Ser145Cys)	Heterozygous	Likely pathogenic	PM1, PM2, PP2, PP3	No	Non-syndromic sensorineural deafness (DFNA20/DFNA26) / Baraitser-Winter syndrome type 2	AD	Potential
OTO.023	Bilateral non-syndromic sensorineural deafness	AR	Childhood	ACTG1	c.548G>A p.(Arg183Gln)	Heterozygous	Likely pathogenic	PS2, PP(PM2), PP2, PP3	No	Non-syndromic sensorineural deafness (DFNA20/DFNA26) / Baraitser-Winter syndrome type 2	AD (de novo)	Potential
OTO.041	Bilateral non-syndromic sensorineural deafness	AR	Childhood	ACTG1	c.848 T>C p.(Met283Thr)	Heterozygous	Likely pathogenic	PS2, PM2, PP2	No	Non-syndromic sensorineural deafness (DFNA20/DFNA26) / Baraitser-Winter syndrome type 2	AD (de novo)	Potential
OTO.011	Unilateral non-syndromic sensorineural deafness	AD	Childhood	MITF	c.909G>A p.(Thr303Thr)	Heterozygous	Likely pathogenic	PS3, PM2, PP1	Yes	Waardenburg syndrome type 2A	AD	Yes
OTO.051	Bilateral non-syndromic sensorineural deafness	AR	Congenital	SOX10	c.135_154del p.(Ser45Argfs*15)	Heterozygous	Likely pathogenic	PM1, PM2, PP3	No	Waardenburg syndrome type 2E	AD	Yes
OTO.019	Bilateral non-syndromic sensorineural deafness	AR	Congenital	GATA3	c.1018A>C p.(Asn340His)	Heterozygous	Likely pathogenic	PM1, PM2, PM6, PP2, PP3	No	Barakat syndrome	AD (de novo)	Yes
OTO.010	CHARGE syndrome	AD	Congenital	CHD7	c.235A>T p.(Lys79*)	Heterozygous	Pathogenic	PM1, PM2, PP4	No	CHARGE syndrome	AD	No
OTO.015	Bilateral non-syndromic sensorineural deafness	AR	Childhood	POU3F4	c.692C>T p.(Thr231Ile)	Hemizygous (male)	Likely pathogenic	PM1, PM2, PP2, PP3	No	Non-syndromic sensorineural deafness (DFNX2/DFN3)	XR	No

Table 2 Clinical and genetic characteristics of cases with causative mutations (Continued)

Case ID	Pre-test phenotype	Pre-test suspected inheritance pattern	Time of deafness onset	Gene	Allele variants	Variant zygosity	ACMG ^a classification ³⁰	Fulfilled ACMG ^a pathogenicity criteria ³⁰	Present in HGMD and/or ClinVar ^b	Gene-associated phenotypes	Inheritance patterns of gene-associated phenotypes	Hidden syndrome
OTO.016	Bilateral non-syndromic sensorineural deafness	AD	Childhood	PRPS1	c.826C>T p.(Pro276Ser)	Hemizygous (male)	Likely pathogenic	PM1, PM2, PP2, PP3	No	Non-syndromic sensorineural deafness (DFNX1)	XD	No
OTO.007	Alport syndrome	AR	Childhood	COL4A5	c.3525_3529dup p.(Pro1177Leufs*124)	Hemizygous (male)	Pathogenic	PM51, PM2, PM6	No	Alport syndrome	XD (de novo)	No

GenBank Accession and version numbers of the genes listed in the table: ACTG1 (NM_001614.3), BSND (NM_057176.2), CDH23 (NM_022124.5), CHD7 (NM_017780.3), COL4A5 (NM_000495.4), GATA3 (NM_001002295.1), MYO15A (NM_016239.3), P2RX2 (NM_174873.2), POU3F4 (NM_000307.4), PRPF51 (NM_002764.3), RDX (NM_002906.3), SLC26A4 (NM_000441.1), SOX10 (NM_000941.3), STRC (NM_153706.2), USH2A (NM_206933.2)

^aAmerican College of Medical Genetics and Genomics

^bYes: variants present in HGMD and/or ClinVar at the moment of clinical interpretation of the case; No: variants absent from both HGMD and ClinVar at the moment of clinical interpretation of the case

our pre-screened population were *ACTG1* ($n = 3$), *USH2A* ($n = 3$) and *STRC* ($n = 3$). Although variants in *USH2A* and *STRC* are often reported as common causes of SNHL [28, 37–40], the identification of *ACTG1* as the most frequent causative gene in our cohort is surprising.

ACTG1 variants are responsible for DFNA20/DFNA26 and type 2 Baraitser-Winter syndrome. None of our 3 cases has syndromic features to date, and all of them had early-onset profound SNHL, expanding the phenotypic spectrum of *ACTG1*, usually associated to post-lingual and progressive SNHL [41, 42], to prelocutive SNHL. Since none of the 3 causative variants had been described and 2 of them were de novo, a targeted hot-spot mutation assay or an AR oriented gene panel (the 2 de novo mutations took place in patients without familial background, simulating a recessive pattern) would have missed them. Therefore, the prevalence of *ACTG1* pathogenic variants could be higher, and its expression pattern more variable, than previously thought [43].

Our 38.1% rate of syndromic SNHL (8/21, including 2 syndromes diagnosed before and 6 after NGS genetic testing) is within expected rates. In contrast, our 38% incidence of AD and 14% of X-linked SNHL are higher than expected [44]. This might reflect the consequences of pre-screening, which excluded the most common AR (*GJB2/GJB6* and *OTOF*) and mitochondrial (*MT-RNR1*) mutations [2, 31, 45]. However, since in our patients 50% (4/8) of causative dominant variants were de novo (2 in *ACTG1*, 1 in *GATA3*; and 1 in *COL4A5*, Table 2) it might also be the consequence of using an unbiased NGS panel, able to identify unexpected de novo variants. Despite the limited size of our cohort, our de novo detection rate is strikingly similar to that reported recently in whole-exome sequencing (WES) studies for different clinical indications (37–68%) [35, 46].

A technical difficulty encountered for the implementation of a clinical-grade test was the presence of highly homologous pseudogene background for some of the target genes included in the panel (Additional file 6), especially *STRC* and *OTOA*. The measures proposed to deal with this problem (gene-restrictive realignment of sequencing results and validation of putative causative variants by gene-specific methods) should reduce misdiagnosis. Moreover, the *STRC* gene, one of the largest contributors to AR SNHL [28, 47, 48], is also a common site for large deletions [28], and CNVs can be refractory to general NGS approaches. As displayed in our population, where 19% of cases (4/21) were justified by CNVs, large genetic rearrangements are increasingly recognized as a common cause of genetic hearing loss, accounting for 13–19% of all causative variants [5, 48–50]. Therefore, CNV analysis should be a requirement for all patients undergoing genetic testing for SNHL. In this regard, our 100% validation rate of NGS-detected *STRC* CNVs by qPCR and MLPA is encouraging (Additional file 8).

Diagnostic rates of up to 60% are expected in patients with suspected AR congenital deafness. This percentage strongly declines for AD hearing loss, especially with the increase in the age of onset [5, 48]. In our series, as in most of published studies [5, 48, 51], prior to comprehensive genetic testing, patients were prescreened for common deafness mutations (in our cohort, *GJB2/GJB6*, *OTOF* and *MT-RNR1*, selected for their high prevalence in Spain [7, 31]). Mutations in the *GJB2* gene are among the most frequent causes for congenital hearing loss. The prevalence of its biallelic pathogenic mutations among non-syndromic SNHL cases ranges geographically from 0% to over 50% [3, 52–54]. Recent analyses show a worldwide and European prevalence of around 13%, increasing in <5 year-old patients [2, 3]. In our laboratory, *GJB2/GJB6*, *OTOF* and *MT-RNR1* prescreening of 180 patients identified the cause of deafness in 34 (18.9%) (unpublished results). This figure, combined with the 42–48% diagnostic rate of our panel in pre-screened patients (48% considering as causative the highly suspicious variants of Table 4), allows us to estimate that combining prescreening with our panel will lead to a diagnosis in about 53–58% of patients.

Our 42–48% detection rate is slightly higher than the average reported with NGS-panels: 41% (10–83%) for a mix of pre-screened and not pre-screened patients [5, 9, 28, 48, 51]. Proper target region coverage and bioinformatics approaches shouldn't be underestimated for maximizing clinical sensitivity. Additionally, the inclusion of syndromic genes, revealing 'hidden syndromes', increased the diagnostic yield. The 6 a priori clinically unrecognized syndromes in our cohort diagnosed after genetic testing (Table 2), representing 28.6% of the genetically diagnosed cases, are a proof of concept of how NGS is changing medicine. In fact, undiagnosed syndromes in families with apparently non-syndromic SNHL are increasingly reported [55–58], expanding the phenotypes associated with SNHL-syndromes [35]. Moreover, 6 patients in our series had pathogenic variations in genes associated with both syndromic and non-syndromic HL: *CDH23* (Usher syndrome 1D and DFNB12), *ACTG1* (Baraitser-Winter syndrome type 2 and DFNA20/DFNA26), *BSND* (Bartter syndrome type IV and DFNB73) or *SLC26A4* (Pendred syndrome and DFNB4) (Table 2). Close follow-up of these patients is mandatory, since syndromic features may develop.

The clinical interpretation of genomic findings is a cornerstone of NGS diagnostic pipelines. Beyond deafness, a recent study indicated that as many as 30% of all disease-causing genetic variants cited in the literature may have been misinterpreted [59]. In our cohort, manual interpretation of variants required an average of 6 h/case, dedicated to in-depth review of the databases and scientific literature, under the perspective of the patient's phenotype and family history, which is imperative for accurate variant interpretation [11, 60].

Table 3 Clinical characteristics of cases without causative mutations

Case ID	Phenotype	Suspected inheritance pattern	Time of deafness onset
OTO.017	Bilateral non-syndromic sensorineural deafness	AR	Congenital
OTO.021	Bilateral sensorineural deafness. Nystagmus, strabismus, delay in psychomotor development and autism spectrum disorder	AR	Congenital
OTO.024	Bilateral non-syndromic sensorineural deafness	AR	Childhood
OTO.025	Bilateral non-syndromic sensorineural deafness	AR	Childhood
OTO.026	Unilateral non-syndromic sensorineural deafness	AR	Childhood
OTO.027	Bilateral non-syndromic sensorineural deafness	AR	Congenital
OTO.028	Bilateral non-syndromic sensorineural deafness	AR	Childhood
OTO.029	Unilateral non-syndromic sensorineural deafness	AR	Congenital
OTO.030	Unilateral sensorineural deafness. Connective tissue problems, digestive problems, urinary reflux and knee hypermobility	AR	Childhood
OTO.032	Unilateral non-syndromic sensorineural deafness	AR	Congenital
OTO.034	Bilateral non-syndromic sensorineural deafness	AR	Childhood
OTO.035	Bilateral non-syndromic sensorineural deafness	AR	Childhood
OTO.038	Bilateral non-syndromic sensorineural deafness	AR	Unknown
OTO.040	Bilateral non-syndromic sensorineural deafness	AR	Childhood
OTO.042	Bilateral non-syndromic sensorineural deafness	AR	Congenital
OTO.044	Bilateral non-syndromic sensorineural deafness	AR	Congenital
OTO.045	Bilateral non-syndromic sensorineural deafness	AR	Congenital
OTO.046	Bilateral non-syndromic sensorineural deafness	AR	Childhood
OTO.049	Bilateral non-syndromic sensorineural deafness	AR	Adulthood
OTO.052	Bilateral sensorineural deafness. Lobe of the auricular pavilion with grooves. Polysyndactyly in hands and feet. Hypospadias	AR	Congenital
OTO.053	Bilateral non-syndromic sensorineural deafness	AR	Childhood
OTO.036	Bilateral non-syndromic sensorineural deafness	AR/AD	Childhood
OTO.039	Bilateral non-syndromic sensorineural deafness	AR/AD	Adulthood
OTO.020	Bilateral non-syndromic sensorineural deafness	AD	Childhood
OTO.022	Bilateral non-syndromic sensorineural deafness	AD	Congenital
OTO.031	Bilateral non-syndromic sensorineural deafness	AD	Childhood
OTO.037	Bilateral non-syndromic sensorineural deafness	AD	Adulthood
OTO.047	Bilateral non-syndromic sensorineural deafness	AD	Childhood
OTO.048	Bilateral non-syndromic sensorineural deafness	AD	Childhood

To date, several studies using NGS for genetic diagnosis of deafness have been published, involving either gene panels or WES [5, 61]. When WES is ordered, sequenced regions not only include genes of interest (“targeted disease-specific panels” such as the one presented in this paper), but also all exons of all genes in the genome. Although WES avoids the need for specific gene panel enrichment, a literature-based selection of the genes involved in the pathology is anyway required for results interpretation. WES increases the requirements for sequencing resources, complicates the analysis and normally provides insufficient coverage of key target regions [29]. Moreover, WES carries

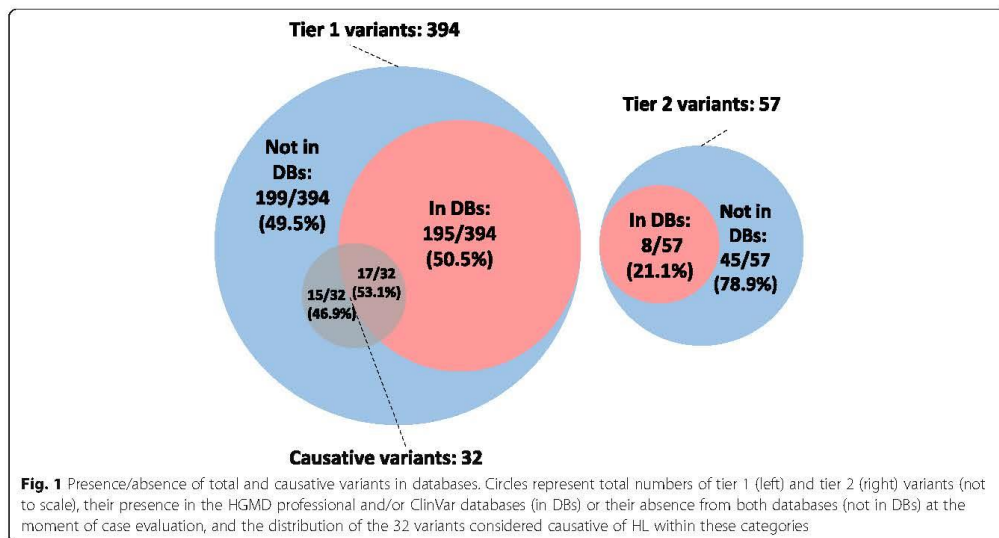
increased chance of secondary findings (variants identified in genes unrelated to the primary medical reason for testing [62]), which introduce noise into the genetic counseling procedure. A comparison of a disease-focused panel versus WES for inherited eye diseases found improved accuracy and performance of the disease-specific panel, a finding that can be translated to hearing loss panels [63]. For these reasons, disease-focused genetic tests have become the standard when evaluating hearing loss [64]. However, WES does have an advantage: the ability to identify alterations in genes not definitely associated with the disease yet. To minimize this disadvantage, a tiered approach was implemented: tier

Table 4 Clinical and genetic characteristics of cases with suspicious VUS*

Case ID	Pre-test phenotype	Pre-test suspected inheritance pattern	Time of deafness onset	Gene	Allele variants	Variant zygosity	ACMG [†] classification ³⁰	Fulfilled ACMG [†] pathogenicity criteria ³⁰	Gene-associated phenotype	Inheritance patterns of gene-associated phenotypes	Hidden syndrome
OTO.028	Bilateral non-syndromic sensorineural deafness	AR	Childhood	OTOA	Whole-gene deletion	Heterozygous	Pathogenic (likely pathogenic over/under pathogenic)	PM1, PM3	Non-syndromic sensorineural deafness (DFNB22)	AR	No
OTO.044	Bilateral non-syndromic sensorineural deafness	AR	Congenital	LOXHD1	c.1282G > T (p.Val428Phe) c.3571A > G (p.Thr1191Ala)	Hemizygous	VUS [#]	PM2, PM3, PP2			No
OTO.045	Bilateral non-syndromic sensorineural deafness	AR	Congenital	SLC26A4	c.695 T > G (p.Leu232Arg)	Homozygous	VUS [#]	PM1, PM2, PP3	Non-syndromic sensorineural deafness (DFNB4)/ Pendred syndrome	AR	Potential

*Variant/s of unknown significance

[†]American College of Medical Genetics and Genomics

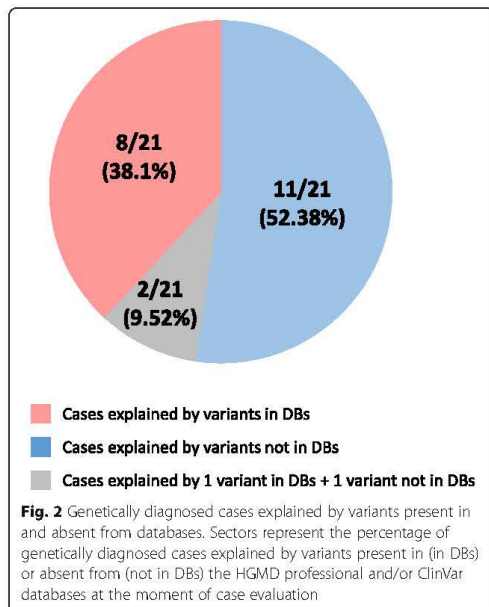


1 includes all genes consistently associated with SNHL, whereas tier 2 includes genes without sufficient clinical validity to be included in clinical testing. Non-systematic reporting of tier 2 genes reduces the uncertainty and simplifies the genetic counseling procedure. However,

meanwhile, it facilitates fast pipeline incorporation of clinically validated genes, as soon as confirmatory discoveries are published.

Conclusions

Our results underscore the importance of a comprehensive approach with careful gene selection to the genetic diagnosis of SNHL. Here, we contribute to show that, with the right methodology, NGS can be transferred to the clinical practice, boosting the yield of SNHL genetic diagnosis to 50–60% (including *GJB2/GJB6* alterations), improving diagnostic/prognostic accuracy, refining genetic and reproductive counseling and revealing clinically relevant undiagnosed syndromes. Lowering cost and increasing quality of WES and whole-genome sequencing (WGS) will probably prompt substitution of physical gene panels by non-targeted approaches. However, WES/WGS results are likely to be filtered through *in-silico* gene panels, based on a meticulously curated gene selection, such as the gene-set of the current panel. Thus, the methodology implemented on the present study is expected to be useful in the years to come. Since comprehensive genetic testing using NGS should be the standard of care for genetic evaluation of patients with SNHL, hereditary deafness should become a paradigm on the raising field of precision medicine. In this context, we expect that the use of the current platform, or others developed on the knowledge presented herein, will help to bring to the clinical arena the advantages of predictive and preventive SNHL genetic testing.



Additional files

- Additional file 1:** List of phenotypes potentially related to hearing defects used as keywords for initial query on HGMD. (XLSX 16 kb)
- Additional file 2:** Tier 1 and tier 2 genes included in v1 of the panel. (XLSX 13 kb)
- Additional file 3:** Tier 1 and tier 2 genes included in v2 of the panel. (XLSX 13 kb)
- Additional file 4:** Cases analyzed with each panel version. (XLSX 10 kb)
- Additional file 5:** Coverages and callabilities for every tier 1 and tier 2 gene on v3 samples. (XLSX 31 kb)
- Additional file 6:** Highly homologous conflictive regions. (XLSX 11 kb)
- Additional file 7:** Primers used to validate causative (pathogenic and likely pathogenic) variants. (XLSX 13 kb)
- Additional file 8:** Samples with STRC CNVs. (XLSX 9 kb)
- Additional file 9:** Variants identified in our cohort classified as pathogenic (DM) by HGMD for hearing-related phenotypes. (XLSX 11 kb)

Abbreviations

ACMG: American college of medical genetics and genomics; AD: Autosomal dominant; AR: Autosomal recessive; CNVs: Copy number variations; DIN: DNA integrity number; HL: Hearing loss; NGS: Next-generation sequencing; SNHL: Sensorineural hearing loss; SNVs: Single nucleotide variants; WES: Whole-exome sequencing; WGS: Whole-genome sequencing

Acknowledgements

We are grateful to all participating patients and their families.

Funding

Work performed at IMOMA for this project was partially supported by a grant from Fundación María Cristina Masaveu Peterson. Work performed at DREAMgenics was partially supported by University of Oviedo Foundation grants (D.C., P.C.P.).

Availability of data and materials

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Authors' contributions

RuC participated in study design, interpretation of results and writing of the manuscript; MD performed most of the experimental work and contributed to interpretation of results; GAC contributed to experimental work and results interpretation; DG, PC and GRO, performed bioinformatic analyses; RA and NSD contributed to experimental work; RaC contributed to results interpretation; AP, MVD, NGG, IH, JLL, ARA, CTB, JR, NG, JRGM, FNB, JAG, AMG, MC, MIG and BGB provided patient samples and associated clinical data; JC participated in study design, interpretation of results and writing of the manuscript; all authors commented on the manuscript and approved the submitted version.

Ethics approval and consent to participate

The study was approved by the Comité de Ética de Investigación del Principado de Asturias (research project #75/14). Consent was obtained from all patients or their parents.

Consent for publication

Not applicable.

Competing interests

The following authors are currently employed by IMOMA or DREAMgenics, the companies involved in the development and exploitation of the OTogenics™ platform: IMOMA: Ru.C. (Physician in Chief), M.D. (Clinical Molecular Geneticist), G.A.C. (Biotechnologist), N.S.D. (Lab. Technician), R.A. (Lab. Technician), Ra.C. (Molecular Biologist) and J.C. (Scientific Director); DREAMgenics: D.C. (Bioinformatician), P.C.P. (Bioinformatician) and G.R.O. (Bioinformatician, C.S.O. and C.E.O.). G.R.O. is a shareholder of DREAMgenics. The other authors declare no conflict of interest.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

¹Instituto de Medicina Oncológica y Molecular de Asturias (IMOMA) S. A, Avda. Richard Grandío s/n, 33193 Oviedo, Spain. ²Disease Research And Medicine (DREAMgenics) S. L., Oviedo, Spain. ³Hospital Universitario Central de Asturias, Oviedo, Spain. ⁴Hospital Álvaro Cunqueiro, Vigo, Spain. ⁵Hospital Universitario Son Espases, Palma de Mallorca, Spain. ⁶Hospital Universitario Río Hortega, Valladolid, Spain. ⁷Hospital Universitario Marqués de Valdecilla, Santander, Spain. ⁸Instituto de Investigación Biomédica de Salamanca, Salamanca, Spain.

Received: 20 March 2018 Accepted: 14 June 2018

Published online: 09 July 2018

References

- Alford RL, Amos KS, Fox M, Lin JW, Palmer CG, Pandya A, Rehm HL, Robin NH, Scott DA, Yoshinaga-Itano C, et al. American College of Medical Genetics and Genomics guideline for the clinical evaluation and etiologic diagnosis of hearing loss. *Genet Med*. 2014;16(4):347–55.
- Burke WF, Warnecke A, Schoner-Heinisch A, Lesinski-Schiedat A, Maier H, Lenarz T. Prevalence and audiological profiles of GJB2 mutations in a large collective of hearing impaired patients. *Hear Res*. 2016;333:77–86.
- Chan DK, Chang KW. GJB2-associated hearing loss: systematic review of worldwide prevalence, genotype, and auditory phenotype. *Laryngoscope*. 2014;124(2):E34–53.
- Sabatini LM, Mathews C, Ptak D, Doshi S, Tynan K, Hegde MR, Burke TL, Bossler AD. Genomic sequencing procedure microcosting analysis and health economic cost-impact analysis: A Report of the Association for Molecular Pathology. *J Mol Diagn*. 2016;18(3):319–28.
- Shearer AE, Smith RJ. Massively parallel sequencing for genetic diagnosis of hearing loss: the new standard of care. *Otolaryngol Head Neck Surg*. 2015;153(2):175–82.
- Nunez-Batalla F, Jaudenes-Casaubon C, Sequi-Canet JM, Vivanco-Allende A, Zubizaray-Ugarteche J, Cabanillas-Farpon R. Aetiological diagnosis of child deafness: CODEPEH recommendations. *Acta Otorrinolaringol Esp*. 2017;68(1):43–55.
- Cabanillas Farpón R, Cadiñanos Bañales J. Hereditary hearing loss: genetic counselling. *Acta Otorrinolaringol Esp*. 2012;63(3):218–29.
- Rahman S, Ecob R, Costello H, Sweeney MG, Duncan AJ, Pearce K, Strachan D, Forge A, Davis A, Bitner-Glindzic M. Hearing in 44-45 year olds with m. 1555A>G, a genetic mutation predisposing to aminoglycoside-induced deafness: a population based cohort study. *BMJ Open*. 2012;2:e000411.
- Sommen M, Schrauwen I, Vandeweyer G, Boeckx N, Corneveaux JJ, van den Ende J, Boudewyns A, De Leenheer E, Janssens S, Claes K, et al. DNA diagnostics of hereditary hearing loss: a targeted resequencing approach combined with a mutation classification system. *Hum Mutat*. 2016;37(8):812–9.
- Geleoc GS, Holt JR. Sound strategies for hearing restoration. *Science*. 2014;344(6184):1241062.
- Abou Tayoun AN, Al Turki SH, Oza AM, Bowser MJ, Hernandez AL, Funke BH, Rehm HL, Amr SS. Improving hearing loss gene testing: a systematic review of gene evidence toward more efficient next-generation sequencing-based diagnostic testing and interpretation. *Genet Med*. 2016;18(6):545–53.
- Cabanillas R, Diñeiro M, Castillo D, Pruneda PC, Penas C, Cifuentes GA, de Vicente A, Durán NS, Álvarez R, Ordóñez GR, et al. A novel molecular diagnostics platform for somatic and germline precision oncology. *Mol Genet Genomic Med*. 2017;5(4):336–59.
- Jennings LJ, Arcila ME, Corless C, Karmel-Reid S, Lubin IM, Pfeifer J, Temple-Smolkin RL, Voelkerding KV, Nikiforova MN. Guidelines for validation of next-generation sequencing-based oncology panels: A Joint Consensus Recommendation of the Association for Molecular Pathology and College of American Pathologists. *J Mol Diagn*. 2017;19(3):341–65.
- Li H, Durbin R. Fast and accurate long-read alignment with burrows-wheeler transform. *Bioinformatics*. 2010;26(5):589–95.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. Genome project data processing 5. The sequence alignment/map format and SAMtools. *Bioinformatics*. 2009;25(16):2078–9.

16. Puente XS, Pinyol M, Quesada V, Conde L, Ordóñez GR, Villamor N, Escaramís G, Jares P, Bea S, González-Díaz M, et al. Whole-genome sequencing identifies recurrent mutations in chronic lymphocytic leukaemia. *Nature*. 2011;475(7354):101–5.
17. Valdes-Mas R, Bea S, Puente DA, Lopez-Otin C, Puente XS. Estimation of copy number alterations from exome sequencing data. *PLoS One*. 2012;7(12):e51422.
18. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc*. 2009;4(7):1073–81.
19. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. A method and server for predicting damaging missense mutations. *Nat Methods*. 2010;7(4):248–9.
20. Choi Y, Sims GE, Murphy S, Miller JR, Chan AP. Predicting the functional effect of amino acid substitutions and indels. *PLoS One*. 2012;7(10):e46688.
21. Reva B, Antipin Y, Sander C. Predicting the functional impact of protein mutations: application to cancer genomics. *Nucleic Acids Res*. 2011;39(17):e118.
22. Schwarz JM, Cooper DN, Schuelke M, Seelow D. MutationTaster2: mutation prediction for the deep-sequencing age. *Nat Methods*. 2014;11(4):361–2.
23. Chun S, Fay JC. Identification of deleterious mutations within three human genomes. *Genome Res*. 2009;19(9):1553–61.
24. Dong C, Wei P, Jian X, Gibbs R, Boerwinkle E, Wang K, Lu X. Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies. *Hum Mol Genet*. 2015;24(8):2125–37.
25. Shihab HA, Gough J, Mort M, Cooper DN, Day IN, Gaunt TR. Ranking nonsynonymous single nucleotide polymorphisms based on disease concepts. *Hum Genomics*. 2014;8:11.
26. Jagadeesh KA, Wenger AM, Berger MJ, Güturu H, Stenson PD, Cooper DN, Bernstein JA, Bejerano G. M-CAP eliminates a majority of variants of uncertain significance in clinical exomes at high sensitivity. *Nat Genet*. 2016;48(12):1581–6.
27. Daydov EV, Goode DL, Sirota M, Cooper GM, Sidow A, Batzoglu S. Identifying a high fraction of the human genome to be under selective constraint using GERP++. *PLoS Comput Biol*. 2010;6(12):e1001025.
28. Mandelker D, Amr SS, Pugh T, Gowrisankar S, Shakhbatyan R, Duffy E, Bowser M, Harrison B, Lafferty K, Mahanta L, et al. Comprehensive diagnostic testing for stereocilin: an approach for analyzing medically important genes with high homology. *J Mol Diagn*. 2014;16(6):639–47.
29. Shearer AE, Black-Ziegelbein EA, Hildebrand MS, Epsteiner RW, Ravi H, Joshi S, Guiffre AC, Sloan CM, Happe S, Howard SD, et al. Advancing genetic testing for deafness with genomic technology. *J Med Genet*. 2013;50(9):e27–34.
30. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E, Spector E, 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*. 2015;17(5):405–24.
31. Gallo-Teran J, Morales-Angulo C, Rodríguez-Ballesteros M, Moreno-Pelayo MA, del Castillo I, Moreno F. Prevalence of the 35delG mutation in the GJB2 gene, del (GJB6-D1351830) in the GJB6 gene, Q829X in the OTOF gene and A1555G in the mitochondrial 12S rRNA gene in subjects with non-syndromic sensorineural hearing impairment of congenital/childhood onset. *Acta Otorinolaringol Esp*. 2005;56(10):463–8.
32. Mandelker D, Schmidt RJ, Ankala A, McDonald Gibson K, Bowser M, Sharma H, Duffy E, Hegde M, Santani A, Lebo M, et al. Navigating highly homologous genes in a molecular diagnostic setting: a resource for clinical next-generation sequencing. *Genet Med*. 2016;18(12):1282–9.
33. Hildebrand MS, Morin M, Meyer NC, Mayo F, Modamio-Hoybjør S, Mencia A, Olavarrieta L, Morales-Angulo C, Nishimura CJ, Workman H, et al. DFNA8/12 caused by TECTA mutations is the most identified subtype of nonsyndromic autosomal dominant hearing loss. *Hum Mutat*. 2011;32(7):825–34.
34. Toumpas CJ, Clark J, Harris A, Beswick R, Nourse CB. Congenital cytomegalovirus infection is a significant cause of moderate to profound sensorineural hearing loss in Queensland children. *J Paediatr Child Health*. 2015;51(5):541–4.
35. Retterer K, Jussola J, Cho MT, Vitazka P, Millan F, Gibellini F, Vertino-Bell A, Smaoui N, Neidich J, Monaghan KG, et al. Clinical application of whole-exome sequencing across clinical indications. *Genet Med*. 2016;18(7):696–704.
36. Millan F, Cho MT, Retterer K, Monaghan KG, Bai R, Vitazka P, Everman DB, Smith B, Angle B, Roberts V, et al. Whole exome sequencing reveals de novo pathogenic variants in KAT6A as a cause of a neurodevelopmental disorder. *Am J Med Genet A*. 2016;170(7):1791–8.
37. Hilgert N, Smith RJ, Van Camp G. Forty-six genes causing nonsyndromic hearing impairment: which ones should be analyzed in DNA diagnostics? *Mutat Res*. 2009;681(2–3):189–96.
38. Reardon W, Coffey R, Phelps PD, Luxon LM, Stephens D, Kendall-Taylor P, Britton KE, Grossman A, Trembath R. Pendred syndrome—100 years of underascertainment? *QJM*. 1997;90(7):443–7.
39. Bonnet C, Riahi Z, Chantot-Bastarud S, Smaghe L, Letexier M, Marcaillou C, Lefevre GM, Hardelin JP, El-Amraoui A, Singh-Estivalet A, et al. An innovative strategy for the molecular diagnosis of usher syndrome identifies causal biallelic mutations in 93% of European patients. *Eur J Hum Genet*. 2016;24(12):1730–8.
40. Yoshimura H, Miyagawa M, Kumakawa K, Nishio SY, Usami S. Frequency of usher syndrome type 1 in deaf children by massively parallel DNA sequencing. *J Hum Genet*. 2016;61(5):419–22.
41. Kemerley A, Sloan C, Pfeiffer W, Smith R, Drack A. A novel mutation in ACTG1 causing Baraitser-winter syndrome with extremely variable expressivity in three generations. *Ophthalmic Genet*. 2017;38(2):152–6.
42. Yuan Y, Gao X, Huang B, Lu J, Wang G, Lin X, Qu Y, Dai P. Phenotypic heterogeneity in a DFNA20/26 family segregating a novel ACTG1 mutation. *BMC Genet*. 2016;17:33.
43. Morin M, Bryan KE, Mayo-Merino F, Goodyear R, Mencia A, Modamio-Hoybjør S, del Castillo I, Cabalka JM, Richardson G, Moreno F, et al. In vivo and in vitro effects of two novel gamma-actin (ACTG1) mutations that cause DFNA20/26 hearing impairment. *Hum Mol Genet*. 2009;18(16):3075–89.
44. Sommen M, Wuys W, Van Camp G. Molecular diagnostics for hereditary hearing loss in children. *Expert Rev Mol Diagn*. 2017;17(8):751–60.
45. Miglioni V, Modamio-Hoybjør S, Moreno-Pelayo MA, Rodríguez-Ballesteros M, Villamar M, Tellería D, Menéndez I, Moreno F, Del Castillo I. Q829X, a novel mutation in the gene encoding otoferlin (OTOF), is frequently found in Spanish patients with prelingual non-syndromic hearing loss. *J Med Genet*. 2002;39(7):502–6.
46. Posey JE, Harel T, Liu P, Rosenfeld JA, James RA, Coban Akdemir ZH, Walkiewicz M, Bi W, Xiao R, Ding Y, et al. Resolution of disease phenotypes resulting from multilocus genomic variation. *N Engl J Med*. 2017;376(1):21–31.
47. Francey LJ, Conlin LK, Kadesch HE, Clark D, Berodini D, Sun Y, Glessner J, Hakonarson H, Jalas C, Landau C, et al. Genome-wide SNP genotyping identifies the Stereocilin (STRC) gene as a major contributor to pediatric bilateral sensorineural hearing impairment. *Am J Med Genet A*. 2012;158A(2):298–308.
48. Zazo Seco C, Wespord M, Feenstra I, Pfundt R, Hehir-Kwa JY, Lelieveld SH, Castelein S, Gillissen C, de Wijs IJ, Admiraal RJ, et al. The diagnostic yield of whole-exome sequencing targeting a gene panel for hearing impairment in the Netherlands. *Eur J Hum Genet*. 2017;25(3):308–14.
49. Shearer AE, Kolbe DL, Azaiiz H, Sloan CM, Fries KL, Weaver AE, Clark ET, Nishimura CJ, Black-Ziegelbein EA, Smith RJ. Copy number variants are a common cause of non-syndromic hearing loss. *Genome Med*. 2014;6(5):37.
50. Ji H, Lu J, Wang J, Li H, Lin X. Combined examination of sequence and copy number variations in human deafness genes improves diagnosis for cases of genetic deafness. *BMC Ear Nose Throat Disord*. 2014;14:9.
51. Likar T, Hasanahodzic M, Teran N, Maver A, Peterlin B, Writzl K. Diagnostic outcomes of exome sequencing in patients with syndromic or non-syndromic hearing loss. *PLoS One*. 2018;13(1):e0188578.
52. Gasparini P, Rabionet R, Barbujani G, Melchionda S, Petersen M, Brøndum-Nielsen K, Metspalu A, Otrmaa E, Pisano M, Fortina P, et al. High carrier frequency of the 35delG deafness mutation in European populations. Genetic analysis consortium of GJB2 35delG. *Eur J Hum Genet*. 2000;8(1):19–23.
53. Estvill X, Fortina P, Surrey S, Rabionet R, Melchionda S, D'Aguma L, Mansfield E, Rappaport E, Govea N, Milla M, et al. Connexin-26 mutations in sporadic and inherited sensorineural deafness. *Lancet*. 1998;351(9100):394–8.
54. Kenneson A, Van Naarden Braun K, Boyle C. GJB2 (connexin 26) variants and nonsyndromic sensorineural hearing loss: a HuGE review. *Genet Med*. 2002;4(4):258–74.
55. Wei X, Sun Y, Xie J, Shi Q, Qu N, Yang G, Cai J, Yang Y, Liang Y, Wang W, et al. Next-generation sequencing identifies a novel compound heterozygous mutation in MYO7A in a Chinese patient with usher syndrome 1B. *Clin Chim Acta*. 2012;413(23–24):1866–71.
56. Behar DM, Davidov B, Brownstein Z, Ben-Yosef T, Avraham KB, Shohat M. The many faces of sensorineural hearing loss: one founder and two novel mutations affecting one family of mixed Jewish ancestry. *Genet Test Mol Biomarkers*. 2014;18(2):123–6.

57. Lu Y, Zhou X, Jin Z, Cheng J, Shen W, Ji F, Liu L, Zhang X, Zhang M, Cao Y, et al. Resolving the genetic heterogeneity of prelingual hearing loss within one family: performance comparison and application of two targeted next generation sequencing approaches. *J Hum Genet.* 2014;59(11):599–607.
58. Qing J, Yan D, Zhou Y, Liu Q, Wu W, Xiao Z, Liu Y, Liu J, Du L, Xie D, et al. Whole-exome sequencing to decipher the genetic heterogeneity of hearing loss in a Chinese family with deaf by deaf mating. *PLoS One.* 2014;9(10):e109178.
59. Boycott KM, Vanstone MR, Bulman DE, MacKenzie AE. Rare-disease genetics in the era of next-generation sequencing: discovery to translation. *Nat Rev Genet.* 2013;14(10):681–91.
60. Maxwell KN, Hart SN, Vijai J, Schrader KA, Slavin TP, Thomas T, Wubbenhorst B, Ravichandran V, Moore RM, Hu C, et al. Evaluation of ACMG-guideline-based variant classification of Cancer susceptibility and non-Cancer-associated genes in families affected by breast Cancer. *Am J Hum Genet.* 2016;98(5):801–17.
61. Bademci G, Foster J 2nd, Mahdieh N, Bonyadi M, Durman D, Cengiz FB, Menendez I, Diaz-Horta O, Shirikavand A, Zeinali S, et al. Comprehensive analysis via exome sequencing uncovers genetic etiology in autosomal recessive nonsyndromic deafness in a large multiethnic cohort. *Genet Med.* 2016;18(4):364–71.
62. Kalia SS, Adelman K, Bale SJ, Chung WK, Eng C, Evans JP, Herman GE, Hufnagel SB, Klein TE, Korf BR, et al. Recommendations for reporting of secondary findings in clinical exome and genome sequencing, 2016 update (ACMG SF v2.0): a policy statement of the American College of Medical Genetics and Genomics. *Genet Med.* 2017;19(2):249–55.
63. Consugar MB, Navaro-Gomez D, Place EM, Bujakowska KM, Sousa ME, Fonseca-Kelly ZD, Taub DG, Janessian M, Wang DY, Au ED, et al. Panel-based genetic diagnostic testing for inherited eye diseases is highly accurate and reproducible, and more sensitive for variant detection, than exome sequencing. *Genet Med.* 2015;17(4):253–61.
64. Rehm HL. Disease-targeted sequencing: a cornerstone in the clinic. *Nat Rev Genet.* 2013;14(4):295–300.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions



Artículo III: Sequencing results from multiple individuals of different ethnicities strongly question the existence of the *KCNE1B* pseudogene

Referencia:

Marta Diñeiro, Guadalupe A. Cifuentes, Raquel Capín, Adrián Santiago, Andrea Otero, David Castillo, Patricia C. Pruneda, Gonzalo R. Ordóñez, Rubén Cabanillas and Juan Cadiñanos. **Sequencing results from multiple individuals of different ethnicities strongly question the existence of the *KCNE1B* pseudogene.** *Eur J Hum Genet.* 2020; 28(4): 401–402. Doi: 10.1038/s41431-019-0502-6.

Factor de Impacto:

3,65, Q2 Genetics & Heredity (JCR 2018, último disponible)

Resumen:

En la publicación “The potential presence of the highly similar paralogue gene *KCNE1B* blurs the genetic basis of *KCNE1*-LQTS patients”, Pantou y col. discutían las dificultades para la detección de variantes patogénicas en *KCNE1* que se derivan de la existencia de *KCNE1B*, un pseudogén del gen *KCNE1* anotado por primera vez en la versión GRCh38 del genoma humano. En su trabajo, además, aportaban indicios de que *KCNE1B* podía ser el resultado de un artefacto durante el ensamblaje y anotación de GRCh38, pero concluían que eran necesarios estudios adicionales para confirmar esta posibilidad. Puesto que *KCNE1* es un gen asociado a hipoacusia sindrómica (sus mutaciones causan el Síndrome de Jervell y Lange-Nielsen tipo 2) y, por tanto, forma parte del panel OTOgenics™, hemos recopilado evidencia basada en las secuencias de más de 70 individuos de diferentes orígenes étnicos que apoya consistentemente la posibilidad de que el supuesto pseudogén *KCNE1B*, ausente en la versión GRCh37 del genoma humano, sea en realidad un artefacto introducido en la versión GRCh38 del genoma humano.

Aportación personal al trabajo:

Este trabajo, relacionado con el proyecto de diagnóstico molecular de la sordera hereditaria, al ser *KCNE1* uno de los genes incluidos en el panel OTOgenics™, participé en el análisis de los datos y colaboré en la redacción del texto.



Sequencing results from multiple individuals of different ethnicities strongly question the existence of the *KCNE1B* pseudogene

Marta Diñeiro¹ · Guadalupe A. Cifuentes¹ · Raquel Capín¹ · Adrián Santiago¹ · Andrea Otero¹ · David Castillo² · Patricia C. Pruneda² · Gonzalo R. Ordóñez² · Rubén Cabanillas¹ · Juan Cadiñanos¹

Received: 2 August 2019 / Accepted: 20 August 2019 / Published online: 16 September 2019
© The Author(s) 2019. This article is published with open access

To the Editor:

In regard to the Viewpoint by Pantou et al. “The potential presence of the highly similar paralogue gene *KCNE1B* blurs the genetic basis of *KCNE1*-LQTS patients” [1], we have collected the following evidence in our laboratory, strongly supporting the possibility that the pseudogene *KCNE1B*, absent from the GRCh37 version of the human genome, is an artifact introduced in GRCh38:

We have evaluated the genotype of the rs1805127 SNP on 58 deaf patients analyzed with the OTogenics Next Generation Sequencing (NGS) capture panel [2]. This SNP was selected because the reference nucleotide for *KCNE1* (T) is different from the reference nucleotide for the paralogue position of the *KCNE1B* pseudogene (C), according to GRCh38 (Supplementary Fig. 1). Considering together all reads potentially originating from *KCNE1* or *KCNE1B*, all patients show variant reads/total reads ratios of 0, 0.5 or 1, as it would be expected for the existence of a single locus with two alleles, instead of two paralogue loci with four alleles. If *KCNE1B* was not an artifact, since the sequences surrounding rs1805127 are identical between *KCNE1* and the corresponding paralogue *KCNE1B* region, reads originating from all four alleles would be mapped together and *KCNE1* heterozygous patients with a ratio of 0.25 should be identified (Supplementary Table 1).

Supplementary information The online version of this article (<https://doi.org/10.1038/s41431-019-0502-6>) contains supplementary material, which is available to authorized users.

✉ Juan Cadiñanos
jcb@imoma.es

¹ Instituto de Medicina Oncológica y Molecular de Asturias (IMOMA) S. A., Oviedo, Spain

² DREAMgenics S. L., Oviedo, Spain

We have evaluated a daughter-father-mother trio in which the father and the daughter are heterozygous for the rs74315445 *KCNE1* SNV, while the mother is wild type. PCR was performed using primers that, according to GRCh38, should not be able to distinguish between *KCNE1* and *KCNE1B* (Supplementary Fig. 1). Sanger sequencing of the obtained PCR products shows heterozygous peaks with identical heights for both alleles from the father’s and the daughter’s DNAs, which is not compatible with a variant affecting only one of four alleles. Although the father could be homozygous for the variant in the *KCNE1* locus (which could explain the similar height of both peaks in his sample if *KCNE1B* was real), the mother is wild type and, therefore, the daughter could not possibly inherit two variant alleles from her parents (Supplementary Fig. 2).

We have PCR amplified the regions containing the positions for rs1805127 and rs1805128 SNPs on DNA from lymphoblastoid cell lines from ten individuals representing ten different ethnic origins from four continents in the same conditions unable to distinguish *KCNE1* from *KCNE1B*. In all cases, Sanger sequencing of the PCR product shows clear wild type or heterozygous genotypes for rs1805127 (with peaks of identical heights for the alternative and reference *KCNE1* alleles in heterozygotes) and homozygous *KCNE1* reference (C/C) genotypes for rs1805128, with no sign of the reference *KCNE1B* (T) allele (Supplementary Table 2 and Supplementary Fig. 3).

In our opinion, these data, together with those presented by Pantou et al., strongly suggest that *KCNE1B* is an artifact that should be reviewed and curated. This would reduce the probabilities of false negative results caused by misalignment of variant reads originating from the true and clinically relevant *KCNE1* locus to the *KCNE1B* region of the GRCh38 reference genome.

Acknowledgements Work performed at IMOMA has been supported by Fundación María Cristina Masaveu Peterson.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Pantou MP, Gourzi P, Degiannis D. The potential presence of the highly similar paralogue gene KCNE1B blurs the genetic basis of KCNE1-LQTS patients. *Eur J Hum Genet.* 2019;27:1175–7.
2. Cabanillas R, Diñeiro M, Cifuentes GA, Castillo D, Pruneda PC, Álvarez R, et al. Comprehensive genomic diagnosis of non-syndromic and syndromic hereditary hearing loss in Spanish patients. *BMC Med Genom.* 2018;11:58.

Artículo IV: Comprehensive genomic diagnosis of inherited retinal and optical nerve disorders reveals hidden syndromes and personalized therapeutic options

Referencia:

Marta Diñeiro, Raquel Capín, Guadalupe A. Cifuentes, Beatriz Fernández-Vega, Eva Villota, Andrea Otero, Adrián Santiago, Patricia C. Pruneda, David Castillo, Mónica Viejo-Díaz, Inés Hernando, Noelia S. Durán, Rebeca Álvarez, Claudia G. Lago, Gonzalo R. Ordóñez, Álvaro Fernández-Vega, Rubén Cabanillas y Juan Cadiñanos. **Comprehensive genomic diagnosis of inherited retinal and optical nerve disorders reveals hidden syndromes and personalized therapeutic options.** *Acta Ophthalmol.* 2020; 1 jun. Doi: 10.1111/aos.14479. Online ahead of print.

Factor de Impacto:

2,153, Q1 Ophthalmology (JCR 2018, último disponible)

Resumen:

En la era de la medicina de precisión, la caracterización genómica de los pacientes con ceguera es esencial. Por este motivo, hemos evaluado las consecuencias del análisis genético sobre el diagnóstico etiológico de la pérdida de visión potencialmente hereditaria y su impacto en el manejo clínico de los pacientes. Para ello, hemos analizado a 100 pacientes españoles no sindrómicos y sindrómicos con un diagnóstico clínico de ceguera causada por alteraciones en la retina, la coroides, el vítreo y/o el nervio óptico. El estudio se basó en la aplicación de un panel de secuenciación de nueva generación (NGS) (OFTALMOgenics™), desarrollado y validado dentro de este trabajo, que incluye hasta 362 genes previamente asociados con estas patologías.

Hemos identificado la causa genética de la ceguera en el 45 % de los pacientes (45/100). El 28,9 % de los casos diagnosticados genéticamente (13/45) fueron sindrómicos y, de ellos, en el 30,8 % (4/13), antes de las pruebas genéticas, las características extraoftalmológicas se habían pasado por alto y/o no se habían relacionado con la discapacidad visual, incluyendo casos con síndrome de Mainzer-Saldino, Bardet-Biedl, mucopolisidosis y MLCRD. En dos casos adicionales se había propuesto ceguera sindrómica previamente, pero sin diagnóstico específico, y un paciente con síndrome de Heimler había sido diagnosticado erróneamente como síndrome de Usher antes del estudio. En el 33,3 % de los pacientes diagnosticados genéticamente (15/45) se identificaron variantes causantes de ceguera en genes para los que existen ensayos clínicos que están evaluando el potencial curativo de la terapia génica.


El estudio genético proporcionó información clínicamente relevante en una gran proporción de pacientes, identificando potenciales opciones terapéuticas o

síndromes previamente no diagnosticados en el 42,2 % de los casos diagnosticados genéticamente (19/45).

Aportación personal al trabajo

Mi contribución a este trabajo consistió en la realización de gran parte del trabajo experimental y en el análisis e interpretación de los resultados obtenidos. Además, participé en la redacción del manuscrito.

Comprehensive genomic diagnosis of inherited retinal and optical nerve disorders reveals hidden syndromes and personalized therapeutic options

Marta Diñeiro,¹ Raquel Capín,¹ Guadalupe Á. Cifuentes,¹ Beatriz Fernández-Vega,² Eva Villota,² Andrea Otero,¹ Adrián Santiago,¹ Patricia C. Pruneda,³ David Castillo,³ Mónica Viejo-Díaz,⁴ Inés Hernando,⁴ Noelia S. Durán,¹ Rebeca Álvarez,¹ Claudia G. Lago,¹ Gonzalo R. Ordóñez,³ Álvaro Fernández-Vega,² Rubén Cabanillas¹ and Juan Cadiñanos¹ 

¹Instituto de Medicina Oncológica y Molecular de Asturias (IMOMA) S. A., Oviedo, Spain

²Instituto Oftalmológico Fernández-Vega (IOFV), Oviedo, Spain

³Disease Research And Medicine (DREAMgenics) S. L., Oviedo, Spain

⁴Hospital Universitario Central de Asturias, Oviedo, Spain

ABSTRACT.

Purpose: In the era of precision medicine, genomic characterization of blind patients is critical. Here, we evaluate the effects of comprehensive genetic analysis on the etiologic diagnosis of potentially hereditary vision loss and its impact on clinical management.

Methods: We studied 100 non-syndromic and syndromic Spanish patients with a clinical diagnosis of blindness caused by alterations on the retina, choroid, vitreous and/or optic nerve. We used a next-generation sequencing (NGS) panel (OFTALMOgenics™), developed and validated within this study, including up to 362 genes previously associated with these conditions.

Results: We identified the genetic cause of blindness in 45% of patients (45/100). A total of 28.9% of genetically diagnosed cases (13/45) were syndromic and, of those, in 30.8% (4/13) extraophthalmic features had been overlooked and/or not related to visual impairment before genetic testing, including cases with Mainzer-Saldino, Bardet-Biedl, mucopolisidosis and MLCRD syndromes. In two additional cases—syndromic blindness had been proposed before, but not specifically diagnosed, and one patient with Heimler syndrome had been misdiagnosed as an Usher case before testing. 33.3% of the genetically diagnosed patients (15/45) had causative variants in genes targeted by clinical trials exploring the curative potential of gene therapy approaches. **Conclusion:** Comprehensive genomic testing provided clinically relevant insights in a large proportion of blind patients, identifying potential therapeutic opportunities or previously undiagnosed syndromes in 42.2% of the genetically diagnosed cases (19/45).

Key words: genomic diagnostics – hereditary – inherited retinal dystrophies – next-generation sequencing – panel sequencing – precision ophthalmology – vision loss

Acta Ophthalmol.

© 2020 The Authors. Acta Ophthalmologica published by John Wiley & Sons Ltd on behalf of Acta Ophthalmologica Scandinavica Foundation.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

doi: 10.1111/aos.14479

Introduction

The approval of Luxturna[®], the first gene therapy for an inherited condition to reach the clinic in the USA and Europe, has made reality the promise of a cure for genetic blindness (Russell et al. 2017). Although only patients with visual impairment caused by a defective *RPE65* gene can benefit from this therapy, advanced clinical trials involving several forms of blindness caused by mutations in other genes are under way. In them, patient's eligibility depends on the fact that his/her disease-causing gene matches the target of the particular therapy. Thus, by holding the key to therapeutic intervention, molecular diagnostics of genetic blindness has reached the top of clinical applicability.

Inherited retinal dystrophies (IRDs), characterized by the degeneration of photoreceptor and retinal pigment epithelial cells, are the main cause of genetic blindness. With a worldwide prevalence of about 1/4000, retinitis pigmentosa (RP) is the commonest IRD (Verbakel et al. 2018). Besides the handicapping effects of visual impairment, between 20% and 30% of patients with RP have an associated non-ocular condition, ranging from mild morphologic changes to life-threatening pathologies. With more

than 250 genes described as responsible for syndromic and/or non-syndromic forms of the disease, the genetic heterogeneity of IRDs is overwhelming. Phenotypic heterogeneity is also characteristic of many IRD genes (Costa et al. 2017). As a result, the clinical testing approaches used for IRD diagnosis (electroretinography, optical coherence tomography, funduscopy examination, fundus autofluorescence, dark adaptation or visual field testing) are not usually able to predict the causative gene in a given patient (Hafler 2017).

Although the features mentioned above have historically hampered the clinical genetics of IRD, molecular diagnostics has demonstrated its utility for multiple aspects of patient management long before the availability of gene therapies. Thus, in paediatric patients, genetic diagnosis may dictate visual prognosis and guide educational and/or supportive plans for the proband (Jauregui et al. 2018). Correct IRD molecular diagnosis can also uncover associations with syndromes that manifest with disabling systemic disease, leading to early interventions that facilitate improved outcomes (Werdich et al. 2014). This is especially relevant, because many syndromic patients present initially to the ophthalmologist long before they are seen by the paediatrician with systemic symptoms (Sadagopan 2017).

Therefore, since 2012, the American Academy of Ophthalmology Task Force recommends genetic testing when clinical findings suggest that a known Mendelian disorder may affect the patient (Stone et al. 2012). However, comprehensive genetic testing of IRDs is not feasible by traditional Sanger sequencing techniques, as these are not suited for analysing hundreds of genes per patient. With the advent of next-generation sequencing (NGS), extensive IRD molecular diagnosis has become possible, even for patients whose clinical diagnosis and inheritance pattern are uncertain after ophthalmologic evaluation (Lee & Garg 2015).

Here we: (1) present the development and validation of an NGS-based panel for the comprehensive genomic diagnosis of blind patients with IRDs and/or optical nerve disorders; (2) evaluate its clinical performance and diagnostic yield in a series of 100

Spanish patients; and (3) illustrate, with real-life data, how the availability of a molecular diagnosis frequently reveals unnoticed syndromes/phenotypic expansions and opens the door to personalized therapeutic opportunities.

Material

Purpose of test

The performed test (OFTALMOgenics™) was aimed at detecting the molecular aetiology of individual clinical diagnoses of syndromic/non-syndromic IRDs and/or optic nerve disorders.

Panel content design: criteria for inclusion of specific genes

Genes associated with blindness caused by alterations in the vitreous, choroid, retina and optic nerve, including both syndromic and non-syndromic forms, were considered.

First, the Human Gene Mutation Database (HGMD) was queried using a list of phenotypes potentially related to ophthalmologic defects (Table S1) to generate a preliminary gene list. This was manually curated to identify genes fulfilling the following criteria: (I) gene defects cause alterations in any of the four ocular structures previously mentioned; (II) published evidence supports the gene-phenotype association in at least two independent families; and (III) at least one publication demonstrates convincing cosegregation of phenotype with gene variants. Then, a tiered classification system was devised. Genes with strong/moderate association with blindness (criteria I, II and III described above) formed tier 1, while genes with weak/preliminary association (criterion I, but not criteria II and/or III) were grouped in tier 2. The panel evolved with revision of newly published literature, yielding versions v1, v2, v3 and v4 (Tables S2–S4 and Table 1, respectively). v1 and v2 were used in the research and development (R&D) phase of the study, while v3 was used both during the R&D phase and as the first version of the clinical panel. v4 is the current optimized clinical version. Each patient was tested with only one of the four different versions: 20 cases were analysed with v1, 19 with v2, 55 with v3 (16

during the R&D phase and 39 in the clinical phase) and 6 with v4 (Table S5).

Sample types

Four millilitre of peripheral blood in conventional EDTA-tubes, 1.5 ml of saliva in Danagen saliva collection containers (Danagen-Biotec, S.L., Badalona, Spain) or ≥ 200 ng of germline genomic DNA (quantitated by a fluorimetric method) were required per patient.

Library preparation, target enrichment and sequencing

Library preparation was carried out as previously described for the OTOgenics NGS platform (Cabanillas et al. 2018), except for library capture, that was performed with the OFTALMOgenics™ probes. The OFTALMOgenics™ NGS pipeline targeted the coding exons and intron–exon junctions of 278 genes (v1) (Table S2), 290 genes (v2) (Table S3), 297 genes (v3) (Table S4) or 362 genes (v4) (Table 1).

Bioinformatics for variant identification and annotation

Next-generation sequencing results were processed using the bioinformatics software Genome One Core (DREAMgenics, Oviedo, Spain), certified with CE/IVD-marking. The pipeline has been adapted from those previously described as part of the ONCOgenics and OTOgenics NGS platforms (Cabanillas et al. 2017; Cabanillas et al. 2018). The workflow of bioinformatics analysis that includes the FASTQ read generation, alignment, duplicate removal, variant identification, filtering and annotation has been previously described (Cabanillas et al. 2018).

Copy-number variants (CNVs) were systematically assessed as described in Cabanillas et al. (2018). Detection was performed with an adapted version of the exome2cnv algorithm, incorporating a combination of read depth and allelic imbalance computations. The algorithm employs a background of pooled samples processed using the same capturing protocol and sequencing technology (Valdes-Mas et al. 2012; Cabanillas et al. 2017). To improve sensitivity for large homozygous

Table 1. (Continued)

Tier 2 (86 genes), genes with weak/preliminary association with Inherited Retinal and Optic Nerve Disorders	
<i>AP2B2</i>	
NM_004644.4	
<i>AQP4</i>	
NM_001650.6	
<i>ARL3</i>	
NM_004311.3	
<i>BBIP1</i>	
NM_001193506.1	
<i>BRAT1</i>	
NM_152743.3	
<i>CSAR2</i>	
NM_018485.2	
<i>CI2B</i>	
NM_006383.3	
<i>COL4A6</i>	
NM_001847.3	
<i>CODR2</i>	
NM_001852.3	
<i>EPRS</i>	
NM_00446.2	
<i>ESPN</i>	
NM_031475.2	
<i>FDX2</i>	
NM_001031734.3	
<i>GRID2</i>	
NM_001310.3	
<i>GUCAL1B</i>	
NM_002098.5	
<i>HIKESHI</i>	
NM_016401.3	
<i>IFTB1</i>	
NM_014055.3	
<i>ITM2B</i>	
NM_021969.4	
<i>KCTD7</i>	
NM_153033.4	
<i>MRP34</i>	
NM_023936.1	
<i>MT-TP</i>	
NC_012920.1	
<i>MT-TS2</i>	
NC_012920.1	
<i>MRP34</i>	
NM_023936.1	
<i>MT-TP</i>	
NC_012920.1	
<i>MT-TS2</i>	
NC_012920.1	
<i>PITPNM3</i>	
NM_031220.3	
<i>PMP71</i>	
NM_033109.4	
<i>POGZ</i>	
NM_015100.3	
<i>RP9</i>	
NM_020288.1	
<i>SDHA</i>	
NM_004168.3	
<i>SLC25A1</i>	
NM_005984.4	
<i>TUBA8</i>	
NM_018943.2	
<i>TUBB4B</i>	
NM_006088.5	
<i>TUBGCP4</i>	
NM_014444.4	
<i>ZNF513</i>	
NM_144611.5	

deletions, genomic regions with no sequencing coverage in an individual sample, but showing proper coverages in the remaining samples, were identified (i.e. homoz. BBS4 exon 1 deletion).

Estimation of analytical sensitivity, specificity and accuracy

The methodology for calculation of analytical sensitivity and specificity for SNVs and Indels has been described (Cabanillas et al. 2018). Accuracy was estimated based on the same dataset and it was expressed as the average ± standard deviation of the absolute value of the difference between the expected and the observed allelic frequencies of each variant.

Multiplex Ligation-dependent Probe Amplification (MLPA)

Multiplex Ligation-dependent Probe Amplification analysis was performed using MRC-Holland (Amsterdam, The Netherlands) probe sets targeting five genes *EYS* (cat. # P328), *PRPF31*, *RHO* and *RPE65* (cat. # P235) and *USH2A* (cat. #'s P361 and P362) on 31 patients without causative variants, following the manufacturer's instructions.

Variant interpretation, classification and diagnostic yield

Clinical classification of variants from v1 and v2 cases was performed as described (Cabanillas et al. 2017). For v3 and v4 cases, variants that could potentially explain the vision loss phenotype of the probands, based on zygosity of the variant, presence of additional variants and mode of inheritance, were further considered. After that, variants were clinically classified according to the American College of Medical Genetics and Genomics (ACMG) guidelines as described (Richards et al. 2015; Cabanillas et al. 2018). Diagnostic yield was defined as the percentage of tested patients with pathogenic/likely pathogenic variants capable of explaining their IRD phenotype.

Variant validation

Pathogenic/likely pathogenic SNVs and indels were validated by PCR + Sanger sequencing. For

validation and identification of the breakpoints flanking the BBS4 exon 1 deletion multiple PCR reactions were performed. Primers used in validation PCRs are described in Table S6.

Patient population

Between September 2014 and May 2019, 100 consecutive patients (47 male, 53 female) with syndromic/non-syndromic IRDs were selected after excluding non-genetic causes. Consent was obtained from patients or their parents. The study was approved by the Comité de Ética de la Investigación del Principado de Asturias (research project #74/14).

Results

Performance of targeted NGS

Mean coverage of tier 1 genes was 453x for v1, 383x for v2, 347x for the 16 R&D phase cases performed with v3, 910x for the 39 clinical cases performed with v3 and 1056x for v4. 99.08%, 99.58%, 98.39%, 99.86% and 99.74% of their target bases were covered by ≥20 reads, respectively. The minimum, average and maximum coverage (average read depth of all target bases of the gene) and callabilities (% of the target bases of the gene with minimum read depths of 10, 20, 50 and 100 reads per each target base of the gene) for every tier 1 and tier 2 gene on samples analysed with OPTALMOgenics™ v4 is shown in Table S7. In clinical cases, regions from tier 1 genes with less than 100% coverage ≥20 reads (DP20) and specific positions within those regions affected by such limitation were included in each individual patient's report.

Analytical sensitivity, specificity and accuracy

A genotyped mixture of 10 lymphoblastoid cell lines was evaluated to determine the analytical sensitivities, specificities and accuracies of the clinical versions of the panel (v3 and v4). For v3, 1277/1286 variants with expected allelic frequency ≥0.1 (1208/1216 SNVs and 69/70 indels) were detected, what gives a sensitivity >0.993 (>99.3%). The v3 version of the platform did not call any SNV or indel in 936684 of the 937048 positions

Table 2. Clinical and genetic characteristics of cases with causative mutations.

Case ID	Pretest phenotype	Previous inheritance genetic study pattern	Pretest inheritance pattern	Age at diagnosis	Gene	Allele variant	Variant zygosity	ACMG classification	Classification HGMD/ClinVar	Fulfilled ACMG pathogenicity criteria	Gene-associated phenotype	Cause-associated phenotype, hidden syndrome inheritance or non-syndromic
OFTALMO.001	Retinitis pigmentosa	Not reported	AR ¹	Old age (61 years)	<i>DHSD5</i>	c.124A>G, p.(Lys42Gln)	Hom.	P	DM/P	PM3_VS, PPI_S, PM2, PM3_P	Retinitis pigmentosa type 39	Non-syndromic
OFTALMO.002	Retinitis pigmentosa	Not reported	AD/AR	Youth (26 years)	<i>ABCA4</i>	c.888C>T, p.(Arg290Trp)	Hom.	LP	DM/Absent	PM3_P, PM2_S, PM2	Stargardt disease type 1 Fundus flavimaculatus Retinitis pigmentosa type 19	Non-syndromic
OFTALMO.003	Retinitis pigmentosa	Not reported	AD/AR	Youth (26 years)	<i>ETYS</i>	c.592S>A>G	Hom.	P	DM/P	PVS1, PM2, PM3, PPI	Cone-rod dystrophy type 3 Retinitis pigmentosa type 25	Non-syndromic
OFTALMO.004	Retinitis pigmentosa	Not reported	AD/AR	Childhood (9 years)	<i>RPI</i>	c.162S>C>G, p.(Ser542*)	Hom.	LP	DM/Absent	PM2, PM3_VS	Retinitis pigmentosa type 1	Non-syndromic
OFTALMO.005	Retinitis pigmentosa	Not reported	AR ¹	Youth (24 years)	<i>CNGBI</i>	c.2957A>T, p.(Asn986Ile)	Hom.	P	DM/P	PM1, PM2, PM3_VS, PPI	Retinitis pigmentosa type 45	Non-syndromic
OFTALMO.006	Retinitis pigmentosa	Not reported	AR ¹	Youth (13 years)	<i>IFT140</i>	c.874G>A, p.(Val292Met)	Hom.	P	DM/P	PS1_P, PM2, PM3_VS, PPI	Retinitis pigmentosa type 89 Menzel-Saldino syndrome	Non-syndromic
OFTALMO.007	Usher syndrome	Not reported	AR ¹	Youth (10 years)	<i>USH2A</i>	c.9304C>T, p.(Gln3102*)	Hom.	P	Absent/P	PVS1, PM2, PM3_P	Jeune syndrome Lowe congenital amaurosis Usher syndrome type 2A	Pretest syndromic
OFTALMO.008	Retinitis pigmentosa	Not reported	XD/ XR	Youth (28 years)	<i>CEM</i>	c.1514delT, p.(Leu507Phe*4)	Hom.	LP	Absent/Absent	PVS1, PM2	Chorioideremia	Non-syndromic
OFTALMO.009	Retinitis pigmentosa	Not reported	AR	Youth (21 years)	<i>IFT140</i>	c.874G>A, p.(Val292Met)	Hom.	P	DM/P	PS1_P, PM2, PM3_VS, PPI	Retinitis pigmentosa type 89 Menzel-Saldino syndrome	Hidden syndromic
OFTALMO.010	Retinitis pigmentosa	Not reported	AR	Youth (18 years)	<i>PROM1</i>	c.1414delC, p.(Asp472Gln*18)	Hom.	P	Absent/Absent	PVS1, PM2, PM3_P	Jeune syndrome Lowe congenital amaurosis Retinitis pigmentosa type 41	Non-syndromic
OFTALMO.011	Retinitis pigmentosa	Yes	AD/AR	Adolescent (44 years)	<i>FAM161A</i>	c.1355_1356delCA, p.(Thr452Ser*3)	Hom.	P	DM/P	PVS1, PM2, PM3_VS, PPI_S	Retinitis pigmentosa type 28	Non-syndromic
OFTALMO.012	Retinitis pigmentosa	Not reported	AD ¹	Adolescent (31 years)	<i>RHO</i>	c.173C>G, p.(Thr58Arg)	Hom.	LP	DM/P	PM2, PPI_S, PPI_3	Retinitis pigmentosa type 4	Non-syndromic
OFTALMO.013	Retinitis pigmentosa	Not reported	AR ¹	Adolescent (37 years)	<i>CNGBI</i>	c.2957A>T, p.(Asn986Ile)	Hom.	P	DM/P	PM1, PM2, PM3_VS, PPI	Retinitis pigmentosa type 45	Non-syndromic
OFTALMO.014	Retinitis pigmentosa	Not reported	AR ¹	Youth (18 years)	<i>CNGBI</i>	c.2957A>T, p.(Asn986Ile)	Hom.	P	DM/P	PM1, PM2, PM3_VS, PPI	Retinitis pigmentosa type 45	Non-syndromic
OFTALMO.015	Retinitis pigmentosa	Not reported	AD/AR	Adolescent (40 years)	<i>BBS1</i>	c.1169T>G, p.(Met390Arg)	Hom.	P	DM/P	PS3, PM3_VS, PPI, BSL_P	Bardet-Biedl syndrome type 1	Hidden syndromic
OFTALMO.016	Retinitis pigmentosa	Not reported	AD/AR	Childhood (8 years)	<i>MCO1N1</i>	c.878L>G>A	Hom.	P	Absent/Absent	PVS1, PM2, PM3_P	Maculoidrosis type IV	Hidden syndromic
OFTALMO.017	Usher syndrome	Yes	AR ¹	Childhood (10 years)	<i>MTO7A</i>	c.602SdelG, p.(Ala2009Pro*6432)	Hom.	P	DM/P	PM1, PM2, PM3_VS, Usher syndrome, type 1B	Usher syndrome, type 1B	Pretest syndromic
OFTALMO.018	Usher syndrome	Not reported	AR	Adolescent (31 years)	<i>MTO7A</i>	c.397C>T, p.(His1313Yp)	Hom.	LP	DM/ILP	PM1, PM2, PM3, PM5, PPI_3	Usher syndrome, type 1B	Pretest syndromic
OFTALMO.019	Stargardt disease	Not reported	AR	Youth (16 years)	<i>ABCA4</i>	c.6449G>A, p.(Cys2130Urp)	Hom.	P	DM/ILP	PM1, PM2, PM3_VS, PM5, PPI_3	Stargardt disease type 1 Fundus flavimaculatus Retinitis pigmentosa type 19 Cone-rod dystrophy type 3	Non-syndromic
OFTALMO.020		Not reported	XD/ XR		<i>CEM</i>	c.320G>T, p.(Gln14*)	Hom.	LP	Absent/Absent	PVS1, PM2	Chorioideremia	Non-syndromic

Table 2. (Continued)

Case ID	Pretest phenotype	Previous genetic study pattern	Pretest inheritance pattern	Age at diagnosis	Gene	Allele variant	Variant zygosity	ACMG classification	Classificación HGMD/ClinVar	Fulfilled ACMG pathogenicity criteria	Gene-associated phenotype	Gene-associated phenotype	Gene-associated phenotype	Pretest syndrome, phenotype hidden syndrome inheritance or non-syndromic
	Retinitis pigmentosa		Youth (15 years)											
OF7ALMO.021	Maculopathy	Not reported	AR ¹	Youth (18 years)	<i>ABCA4</i>	c.904_508delGTTGCCA(TCTGGCTG, p.(Val1682_Val1684del) c.3943C>T, p.(Gln1315*)	Het. Het. P	P	DM/np DM/Absent	PM2, PM3, NS, PM4 PVS1, PM2, PM3, S	Stargardt disease type 1 Fundus flavimaculatus Retinitis pigmentosa type 19	AR	Non-syndromic	
OF7ALMO.022	Retinitis pigmentosa	Not reported	AR	Adulthood (27 years)	<i>OMT</i>	c.416T>G, p.(Met139Asn)	Hom.	LP	Absent/Absent	PM1, PM2, PM3, P, PF3	Gyrate atrophy of choroid and retina with or without ophthalmia	AR	Non-syndromic	
OF7ALMO.024	Retinitis pigmentosa	Yes	AD/AR	Old age (57 years)	<i>PRPH2</i>	c.660_565del, p.(Pro221_Cys222del)	Het.	LP	Absent/Absent	PM1, PM2, PM4	Macular dystrophy Retinitis pigmentosa type 7 Retinitis punctata albescens Choroidal dystrophy, central areolar type 2	AD	Non-syndromic	
OF7ALMO.029	Retinitis pigmentosa	Not reported	XR ¹	NA	<i>RPGR (ORF15)</i>	c.2284delG, p.(Gln762Lys*53)	Hom.	LP	Absent/Absent	PVS1, PM2	Retinitis pigmentosa type 3 XR Cone-rod dystrophy type 1	XR	Non-syndromic	
OF7ALMO.031	Retinitis pigmentosa	Not reported	AD/AR	Youth (14 years)	<i>RPI</i>	c.162C>G, p.(Ser54*) c.227T>C, p.(Leu70Pro) c.356G>A, p.(Gly119Asp)	Het. Het. LP LP LP		DM/Absent DM/Absent DM/Absent	PM2, PM3, S PM1, PM2, PM3, S, PPI	Retinitis pigmentosa type 1 AR Retinitis pigmentosa type 26 Cone-rod dystrophy Bardet-Biedl syndrome	AR	Non-syndromic	
OF7ALMO.033	Retinitis pigmentosa	Not reported	AD/AR	Adulthood (39 years)	<i>CEKZL</i>		Hom.	LP	Absent/Absent	PVS1, PM2, PM3, P		AR	Non-syndromic	
OF7ALMO.041	Malformative syndrome, suspicion of Bardet-Biedl syndrome.	Yes	AR	Childhood (10 years)	<i>BBS4</i>	c.2274_23 + 930del	Hom.	LP	Absent/Absent	PVS1, PM2, PM3, P		AR	Pretest syndrome	
OF7ALMO.046	Retinitis pigmentosa, mild mental retardation, patent foramen ovale, polydactyly in hands and feet, hypoplasia, renal dysplasia, obesity	Not reported	AR ¹	Youth (18 years)	<i>ARL6</i>	c.36_38dupCCTG, p.(Lys49Profs*15) c.699G>A, p.(Gly316Arg)	Het. Het. P LP		Absent/Absent Absent/Absent	PVS1, PM2, PM3, S, PPI PM1, PM2, NS, S, PPI, PF3	Retinitis pigmentosa type 33	AR	Non-syndromic	
OF7ALMO.047	Retinitis pigmentosa	Not reported	AD/AR	Childhood (4 years)	<i>IMPDH1</i>	c.93C>G, p.(Asp311Gln)	Het.	LP	Absent/Absent	PVS2, M, PM1, PM2, PVS1, PM2, PM3, P, PPI, PF3	Retinitis pigmentosa type 10	AD	Non-syndromic	
OF7ALMO.052	Retinitis pigmentosa	Not reported	AD/AR	Youth (29 years)	<i>USH2A</i>	c.930(C>T, p.(Gln310P*) c.2274G>T, p.(Cys159Phe)	Het. Het. P LP LP		DM/P Absent/P	PVS1, PM2, PM3, P, PPI, PF3, PVS1, PM2, PM3, S, PPI, PF3	Retinitis pigmentosa type 39	AR	Non-syndromic	
OF7ALMO.054	Usher syndrome	Not reported	AR	Adulthood (49 years)	<i>PEX1</i>	c.2077T>C, p.(Leu1098Pro) c.1548delT, p.(Leu517Cys*2)	Het. Het. LP LP		DM/Absent Absent/Absent	PM1, PM2, PM3, P, PPI, PM1, PM2	Hemiker syndrome	AR	Pretest syndrome	
OF7ALMO.060	Complex syndrome of the optic nerve, Retinitis pigmentosa.	Not reported	AD/AR	Congenital (at birth)	<i>KIF11</i>	c.2548_23A>G	Het.	P	Absent/Absent	PVS1, PM2, PF3	MLCRD syndrome	AD	Höllera syndrome	

Table 2. (Continued)

Case ID	Pretest phenotype	Previous genetic study	Pretest inheritance pattern	Age at diagnosis	Gene	Allele variant	Variant zygosity	ACMG classification	Classification HGMD/ClinVar	Fulfilled ACMG pathogenicity criteria	Gene-associated phenotype	Gen-associated phenotype	Gen-associated inheritance or non-syndromic pattern
	Macular atrophy, Mental retardation												
OFTALMO.063	Macular atrophy with perimacular 'flecks' compatible with Stargardt's disease	Not reported	AD/AR	Childhood	<i>ABCA4</i>	c.5819T>C, p.(Leu1940Pro) c.3386G>T, p.(GArg129Leu)	Het. Het. P	P	DM(P/DM)/LP	PM2, PM3, VS, PFI, PFI3, PFI4, PM1, PM2, PM3, S, PM5, PFI3	Stargardt disease type 1 Fundus flavimaculatus	AR	Non-syndromic
OFTALMO.068	Retinitis pigmentosa. Atrophy of the optic nerve. Deafness.	Not reported	AD/AR ¹	Childhood (8 years)	<i>USH2A</i>	c.2276G>T, p.(Cys539Phe) c.120954dC, p.(Cys403I*)	Het. Het. LP	LP	DM(P/DM)/Absent	PM1, PM3, S, PFI, PFI3, BSI, LP, PVS1, PM2	Usher syndrome type 2A	AR	Pretest syndrome
OFTALMO.072	Retinitis pigmentosa. Atrophy of the optic nerve. Deafness.	Not reported	AR ¹	Adulthood (40 years)	<i>ABCA4</i>	c.5882G>A, p.(Gly1961Glu)	Hom.	P	DM/P	PFI3, P, PFI4, PM1, VS, PFI3, BA1	Stargardt disease type 1 Fundus flavimaculatus Retinitis pigmentosa type 1 Pre-cone-rod dystrophy type 3	AR	Pretest syndrome
OFTALMO.076	Macular atrophy in pattern with flecks	Not reported	AR ¹	Old age (67 years)	<i>ABCA4</i>	c.3056C>T, p.(Thr1018Met) c.6718A>G, p.(Thr2240Asp)	Het. Het. LP	LP	DM(P/DM)/Absent	PM1, PM2, PM3, PFI3, PM2, PM3, S, PFI3	Stargardt disease type 1 Fundus flavimaculatus	AR	Non-syndromic
OFTALMO.077	Retinitis pigmentosa with macular edema. Mean myopia.	Not reported	AR	Youth (13 years)	<i>USH2A</i>	c.8917, 8918delCT, p.(Leu2973IleAsp8779) c.9119C>A, p.(Trp3040P*)	Het. Het. P	P	DM/Absent DM/Absent	PVS1, PM2, PM3, S, PVS1, PM2, PM3, PFI	Usher syndrome type 2A	AR	Pretest syndrome
OFTALMO.081	Bilateral moderate non-progressive hearing loss	Not reported	AD/AR	Old age (66 years)	<i>PRPH2</i>	c.422delA, p.(Tyr141His)	Het.	LP	Absent/Absent	PVS1, PM2	Macular dystrophy Retinitis pigmentosa type 7 Retinitis punctata albescens Chorioidal dystrophy, central areolar type 2	AD	Non-syndromic
OFTALMO.082	Infectious chorioiditis (25 years). Macular degeneration	Not reported	AR ¹	Old age (63 years)	<i>PRPH2</i>	c.421T>C, p.(Tyr141His)	Het.	LP	DM/Absent	PM1, PM2, PM5, PFI, S, PFI3	Macular dystrophy Retinitis pigmentosa type 7 Retinitis punctata albescens Chorioidal dystrophy, central areolar type 2	AD	Non-syndromic
OFTALMO.087	Nystagmus and visual deficit	Yes	AR ¹	Congenital (3 months)	<i>CEP290</i>	c.381, 387delTAGA, p.(Asp128Glu8734) c.7341, 7344dupACTT, p.(Ser2449Thr878)	Het. Het. P	P	DM/P Absent/Absent	PVS1, PM1, S, BA1, PVS1, PM2, PM3	Leber congenital amaurosis type 10 Senior-Loken syndrome type 6	AR	Non-syndromic
OFTALMO.093	Bilateral atrophy, diabetes and hypertension	Not reported	AD ¹	Adulthood (40 years)	<i>PRPH2</i>	c.537G>A, p.(Tyr179Y)	Het.	LP	Absent/Absent	PVS1, PM2	Macular dystrophy Retinitis pigmentosa type 7 Retinitis punctata albescens Chorioidal dystrophy, central areolar type 2	AD	Non-syndromic
OFTALMO.094	Ball's eye maculopathy.	Not reported	AD/AR ¹	Youth (12 years)	<i>ABCA4</i>	c.982G>T, p.(Gln228*) c.3882G>A, p.(Gly1961Glu)	Het. Het. P	P	DM/Absent DM/P		Stargardt disease type 1 Fundus flavimaculatus	AR	Non-syndromic

Table 2. (Continued)

Case ID	Preset phenotype	Previous genetic study	Preset inheritance pattern	Age at diagnosis	Gene	Allele variant	Variant zygosity	ACMG classification	Clasificación HGMD/ClinVar	Fulfilled ACMG pathogenicity criteria	Gene-associated phenotype	Gene-associated phenotype inheritance pattern
	Possible macular dystrophy									PVS1, PM2, PM3 PS1, P, PS4, PM3, VS, PFS, BA1		
OFTALMO.095	Macular atrophy with perimacular flecks	Not reported	AD/AR	Childhood (13 years)	<i>ABCA4</i>	c.338G>T, p.(Arg129Leu) c.606c2>A>G	Het, Het, P LP	DM/LP Absent/Absent	DM/LP Absent/Absent	PM1, PM2, PM3, PFS, PVS1, VS, PFS, BA1	Stargardt disease type 1 Fundus flavimaculatus	AR Non-syndromic
OFTALMO.097	Retinitis pigmentosa, diabetic retinopathy, deafness, epilepsy	Not reported	AD/AR	Childhood (8 years)	<i>PRPS1</i>	c.292G>A, p.(Asp98Asn)	Het, LP	LP	Absent/Absent	PM1, PM2, P066 P, PFS	Syndromic retinitis pigmentosa (ocular asymmetry, hearing loss and neurological disorders)	XD
OFTALMO.098	Oxoid and macular atrophy	Not reported	AD†	Youth (13 years)	<i>OPAI</i>	c.184T + 1G>T	Het, LP	LP	DM/LB	PVS1, PM2	Optic atrophy 1 Optic atrophy plus syndrome	AD Non-syndromic

AD = autosomal dominant; AR = autosomal recessive; Hem = Hemizygous; Het = Heterozygous; Hom = Homozygous; LB = Likely benign; LP = Likely pathogenic; NA = not available; np = not provided; P = Pathogenic; XD = X-linked dominant.
 † Family history of eye disease.

of the probe design region known to not be affected by these type of variants, yielding a specificity of 0.9996. (936684/937048; >99.95%). The accuracy of v3 for determining the allelic frequencies of SNVs and indels with expected allelic frequencies ≥ 0.1 was 0.03 ± 0.06 (Table S8). The sensitivity, specificity and accuracy of v4, calculated as described above for v3, were 0.9917, 0.9995 and 0.033 ± 0.055 , respectively (Table S9).

Orthogonal validation of sequencing results

All variants considered responsible for the IRDs phenotypes of the probands (Table 2) were successfully validated by approaches alternative to NGS. These included 48 instances of SNVs or indels (validated by PCR and Sanger sequencing) and 1 CNV: 1 homozygous deletion of *BBS4* exon 1 (validated by breakpoint-specific PCR).

Performance at challenging regions: RPGR ORF15

Certain regions in the genome represent a major challenge for short-reading NGS technologies. Of the genes included in the OFTALMOgenics designs, *RPGR* poses the highest difficulties. *RPGR* encodes several isoforms, but only isoform C (NM_001034853), also known as *RPGR ORF15*, is highly expressed in the retina and, consequently, involved in the pathogenesis of retinitis pigmentosa (Vervoort et al. 2000). Several regions of *RPGR ORF15* bear low sequence complexity, causing a decrease in base quality at the end of NGS reads. This, combined with the highly repetitive nature of the *RPGR ORF15* reference sequence, leads to low mapping quality, limiting NGS-based *RPGR ORF15* analysis. As a result, many laboratories analyse *RPGR ORF15* using traditional Sanger sequencing (Chiang et al. 2015). Being aware of these difficulties, and as we observed that 92.7% of R&D cases (51/55) showed regions with less than 100% of their bases covered by ≥ 20 reads, we increased sequencing coverage and dedicated a larger number of capture probes to this conflictive region in the design of the panel, reaching 100% of the *RPGR* target region covered by 20 or more reads in v4 cases

(Table S10). Of note, one of the cases in our series had a causative variant detected by the NGS pipeline in *RPGR* affecting *ORF15* (NM_001034853.1: c.2284delG; p.(Glu762Lysfs*53); corresponding with c.1905+379delG according to the main NM_000328.2 *RPGR* RefSeq), demonstrating the ability of the test to identify these challenging mutations.

Analysis of causative variants and diagnostic yield

Of 100 cases with low vision caused by alterations in the vitreous, choroid, retina and/or optic nerve, a genetic cause for their vision loss phenotype was found in 45 (45%) after identifying 49 different pathogenic/likely pathogenic variants in 26 genes (Fig. 1, Table 2 and Table S6). Causative mutations were found in 31.2% (5/16) of cases in which the existence of a previous genetic study, with negative results, had been reported before OFTALMOgenics testing (Table 2 and Table S11). No causative mutation could be identified in 55 patients (55%) (Table S11). These are likely to be a pool of patients without a genetic aetiology, with mutations in genes not included in the panel or with mutations in genes included in the panel but not detectable by the methodology used. The commonest causative genes in our population were *ABCA4* (8 patients), *PRPH2* (4 patients) and *USH2A* (4 patients). 1 pathogenic variant in *IMPDH1* was *de novo* (after confirmation of maternity and paternity). The post-test inheritance patterns of the molecularly diagnosed patients were autosomal recessive (AR) in 75.5% (34/45) of cases, autosomal dominant (AD) in 17.8% (8/45), and X-linked in 6.5% (3/46) (Fig. 2 and Table 2).

CNV analysis identified a causative variant in 1 of the 45 molecularly diagnosed patients (2.2%): a homozygous *BBS4* exon 1 deletion in a patient with suspected Bardet-Biedl syndrome (retinitis pigmentosa, mild mental retardation, polydactyly in hands and feet, genital hypoplasia, renal dysplasia, obesity and hypertension). As no heterozygous CNVs were detected by our NGS platform, we analysed by MLPA a group of five genes included in our panel and previously described to be affected by pathogenic deletions/duplications (including *USH2A*, *EYS*,

PRPF31, *RHO* and *RPE65*) on a subset of 31 cases without a genetic diagnosis from our series. This analysis identified no homozygous or heterozygous alterations (Table S13).

Two of the 49 causative variants (*ABCA4* c.5882G>A and *CEP290* c.384_387delTAGA) are present in one control population each with allele frequencies above 0.005, the threshold we used for the application of the BA1 (benign-standalone) ACMG criterion to recessive variants: 0.005054 (613/121302) in ExAC for *ABCA4* c.5882G>A and 0.01074 (122/11360) in GO-ESP for *CEP290* c.384_387delTAGA. Their respective GnomAD frequencies are 0.00350 (110/31396) and 0.00003 (1/31362), below the BA1 threshold and compatible with disease incidence in the general population. Moreover, the 0.01074 (122/11360) frequency reported for *CEP290* c.384_387delTAGA in GO-ESP does not match those of the same variant in other databases mentioned by dbSNP, all with higher numbers of alleles analysed [0.000056 (14/249068) in GnomAD_exome, 0.000104 (13/125568) in TOPMED and 0.00074 (58/78700) in PAGE_STUDY] and is suspiciously high for a truncating variant in a gene in which inactivating mutations have consistently been shown to be pathogenic. Taking these data altogether and considering that, otherwise, both of these variants meet ACMG criteria for pathogenicity, we propose that they should be part of the exclusion list for application of the BA1 criterion, to avoid false negatives caused by automatic filtering based on allele frequency.

In total, 1081 variants in tier 1 genes and 119 variants in tier 2 genes were identified and evaluated in the full cohort of 100 patients. 476/1081 (44%) and 81/119 (68.1%) of the identified tier 1 and tier 2 variants, respectively, were absent from the HGMD professional and ClinVar databases (hereafter 'clinical databases'). No tier 2 variant was considered responsible for the IRD phenotype. Eighteen of the variants absent from the clinical databases (all from tier 1) were classified as pathogenic or likely pathogenic and responsible for the IRD. Globally, those 18 variants were considered responsible for the IRD phenotype in 18 cases. As a result, 40% of the genetically diagnosed cases (18/45)

were explained by variants not described in the clinical databases (Table 2). In addition, 31 non-redundant variants were identified that explained positive cases: 30 of these were classified as pathogenic (DM) by HGMD and/ or pathogenic/ likely pathogenic by ClinVar for ophthalmological phenotypes, and 1 (*ABCA4* c.6718A>G, p.(Thr2240Ala)) absent in ClinVar and 'questionable' (DM?) in HGMD for a phenotype of retinal dystrophy (Table 2). We found 363 variants of uncertain significance (VUS) that appeared a total of 412 instances considering all patients (Table S12). These results highlight the importance of manual interpretation and curation for the clinical classification of variants.

Increased clinical sensitivity and specificity by jointly analysing non-syndromic and syndromic genes

For 7 of the 45 patients in which a genetic diagnosis was obtained, the phenotypes associated with the genes affected by causative variants either did not match the pretest clinical features/diagnoses or corresponded to syndromes not previously diagnosed in the proband's personal and familial medical records (hereafter 'hidden syndromes'; Table 2):

Two patients, male and female, both with a pretest diagnosis of non-syndromic RP, shared the same homozygous mutation in *IFT140* (c.874G>A, p.(Val292Met)). The *IFT140* gene had been associated with recessive syndromic (Mainzer-Saldino and Jeune syndromes) (Perrault et al. 2012; Schmidts et al. 2013) and non-syndromic phenotypes (RP80 and Leber Congenital Amaurosis) (Xu et al. 2015). Post-test clinical assessment of the male patient revealed renal hypertension and two collapsed vertebrae, both signs compatible with Mainzer-Saldino syndrome (RP, renal and skeletal dysplasia). Thus, a hidden syndrome explained the phenotype. The clinical diagnosis of the female remained as non-syndromic RP post-test. Interestingly, in our two cases, the same variant (*IFT140* c.874G>A; p.(Val292-Met)) was associated with a syndromic and a non-syndromic phenotype.

A male patient with a pretest diagnosis of RP was affected by a homozygous pathogenic variant in *BBS1*

(c.1169T>G, p.(Met390Arg)). *BBS1* alterations have been associated with recessive non-syndromic (Estrada-Cuzcano et al. 2012; Sharon & Banin 2015) and syndromic RP (Bardet-Biedl syndrome -RP, diabetes-linked obesity, polydactyly, kidney abnormalities and intellectual disability-). Upon post-test reevaluation of the case, the patient was found to have pancreas divisum and unilateral postaxial polydactyly emerging from the fifth toe (right foot). However, no intellectual disability, nor renal or hormonal alterations were detected.

A female patient with a pretest diagnosis of RP had a novel canonical splicing mutation in *MCOLN1* (c.878-1G>A) in homozygosity. *MCOLN1* inactivation has only been associated with type 4 mucopolidiosis, characterized by neurologic and ophthalmologic abnormalities (Boudewyn & Walkley 2019). Post-test clinical review of the case revealed that the patient had psychomotor delay, intellectual disability and a corneal lesion with microcystic epithelial oedema, suggestive of corneal dystrophy.

A male with nystagmus, macular atrophy, typical RP, retinal ischaemia, complete optic nerve atrophy and mental retardation that had not been clinically matched to any specific syndrome pretest, had a heterozygous pathogenic *KIF11* variant (c.2548-2A>G). This gene has been associated with exudative familial vitreoretinopathy and with microcephaly with or without choroidretinopathy, lymphedema or mental retardation (MLCRD), both dominant phenotypes (Robitaille et al. 2014). Post-test reevaluation of the case confirmed that the phenotype matched MLCRD syndrome.

A male patient with a clinical diagnosis of Usher syndrome was found to be affected by two likely pathogenic compound heterozygous variants affecting the *PEX1* gene (c.1548delT, p.(Leu517Cysfs*2) and c.3077T>C, p.(Leu1026 Pro)). The *PEX1* gene has been associated with peroxisome biogenesis disorders types 1A and 1B, as well as with Heimler syndrome, all with recessive inheritance (Ratbi et al. 2015). Heimler syndrome combines the hearing loss and retinal dystrophy phenotypes typical of Usher syndrome with enamel hypoplasia of the secondary dentition and nail abnormalities, which may be overlooked. Therefore, genetic

assessment increased the specificity of the diagnosis.

Finally, a female patient with hearing loss, diabetes and RP without a specific syndromic diagnosis pretest was affected by a missense, likely pathogenic, heterozygous *PRPS1* variant (c.292G>A, p.(Asp98Asn)). Pathogenic variants in this gene have been associated with three X-linked conditions involving blindness: Charcot-Marie-Tooth disease type 5, Arts syndrome and syndromic retinitis pigmentosa (ocular asymmetry, hearing loss and neurological alterations). The latter has only been described in females with missense variants and shows a dominant mode of inheritance (Fiorentino et al. 2018).

Aside from these seven cases, another female patient with a pretest diagnosis of non-syndromic RP and a blind sister had two compound heterozygous variants affecting *ARL6* [p.(Lys14Profs*15), pathogenic and p.(Gly167Arg), likely pathogenic]. This gene has been associated both with non-syndromic RP and with Bardet-Biedl syndrome. Clinical review of the family post-test revealed that, while the patient had no extraocular manifestations, her sister was born with an extra toe that was later removed by surgery. Although the sister has not been genetically analysed so far, this likely represents another case of a single genotype causing syndromic and non-syndromic hereditary blindness, in this case within the same family.

In total, 28.9% of the genetically diagnosed patients (13/45) had syndromic visual impairment (Table 2). Of those, 30.8% (4/13) had a previously unrecognized (hidden) syndrome: Mainzer-Saldino (1 patient), Bardet-Biedl type 1 (1 patient), mucopolidiosis type IV (1 patient) and MLCRD (1 patient). Likewise, 23.1% (3/135) had different syndromic manifestations, but they did not have a specific clinically diagnosed syndrome before the test was performed: Usher type 2A (2 patients) and syndromic retinitis pigmentosa caused by mutations in the *PRPS1* gene (1 patient). Additionally, one patient with Heimler syndrome according to the test results had previously been imprecisely diagnosed as an Usher case based on clinical features (2.2%; 1/45) (Table 2). In contrast, of the 55 patients without a genetic

diagnosis, only 5 (9.1%) had potentially syndromic features (Table S11). These findings demonstrate that comprehensive genomic diagnosis not only increases diagnostic performance (sensitivity and specificity), but also provides information with high clinical relevance.

Genomic diagnosis identifies potential gene therapy options for a considerable proportion of IRD patients

Thirteen gene therapy clinical trials aimed at compensating genetic defects responsible for hereditary blindness of patients genetically diagnosed by our platform were open as of November 2019 or had been open for a period of time during this study (Table 3). Of the genes targeted by these clinical trials, *ABCA4* is the one altered by variants responsible for blindness in most patients of our series (8 patients), followed by *USH2A* (2 patients), *CHM* (2 patients), *MYO7A* (2 patients) and *RPGR* (1 patient) (Table 3). In total, 33.3% of the genetically diagnosed patients (15/45) and 15% of the genetically evaluated patients (15/100) could have potentially benefited from therapeutic approaches with curative potential in this time window (Table 3).

Discussion

Out of the different outputs of this work, the most appealing one is probably the confirmation of the utility of comprehensive genomic diagnosis of blind patients for the identification of potential therapeutic options matching their genetic profile in a relatively large proportion of them. Nonetheless, the advantages of testing hundreds of blindness genes in parallel reach far beyond therapeutic guidance. Thus, in the diagnostic setting, because the complex genetic heterogeneity of retinal dystrophies makes very hard predicting the genetic alteration based on the clinical characterization of the patient only, and mutation of a single gene may cause a broad range of different clinical manifestations, comprehensive testing is particularly useful (Hafler 2017).

In fact, the ability to perform phenotype-agnostic testing is providing a growing body of evidence in support of the concept of phenotypic expansion,

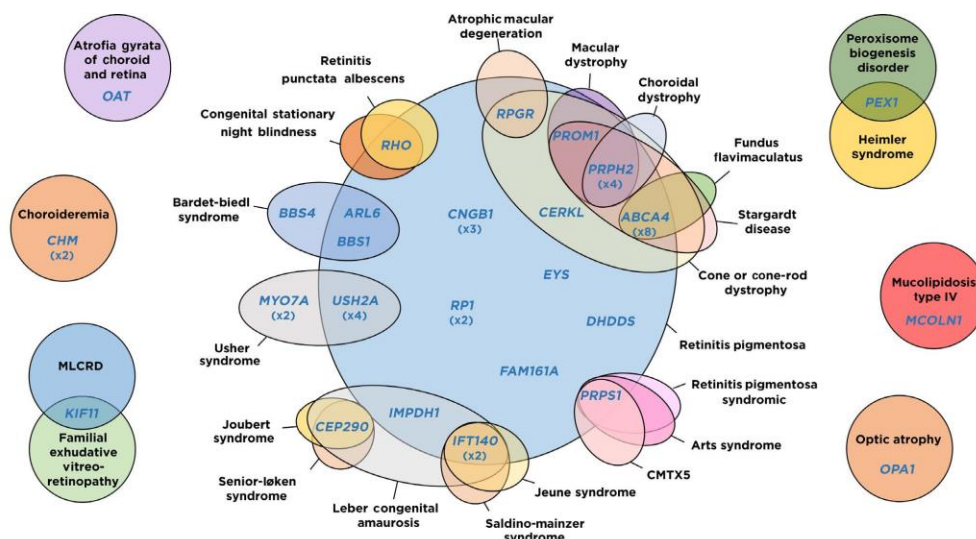


Fig. 1. Genes considered responsible for hereditary visual impairment in our cohort and their associated phenotypes. Ellipses represent phenotypes (black characters) involving vision loss caused by alterations in the retina, vitreous, choroid and/or optic nerve and consistently associated to defects in the genes (blue characters) they contain. For genes with causative alterations affecting more than one patient in our cohort, the number of affected patients is shown in brackets below. Phenotype acronyms: CMTX5: Charcot-Marie-Tooth disease, X-linked recessive, 5; MLCRD: microcephaly, lymphoedema, chorioretinal dysplasia syndrome.

by means of the identification of disease-causing mutations in genes not previously related to the proband's clinical phenotype (Bowne et al. 2011; Nishiguchi et al. 2014; Consugar et al. 2015). This approach can also tackle the diagnostic dilemmas appearing when phenotypic variability is observed within the same family, a common situation in inherited retinal dystrophies (Consugar et al. 2015). Moreover, as phenotypes evolve over time, comprehensive testing reduces the possibility that the causative gene is not evaluated because the proband's symptoms do not match the phenotype associated to it at the moment of testing. Finally, as visual impairment may be the presenting symptom of a systemic syndrome, an accurate molecular diagnosis may be critical for identifying a so far not revealed and wider health problem. Apart from its informative value with regard to prognosis, because some of these syndromes can cause significant morbidity, prompt identification and referral for systemic assessment and monitoring may secure early intervention and a better outcome (Werdich et al. 2014).

A total of 15.6% (7/45) of the patients for which a genetic diagnosis was obtained in our study corresponded to cases whose clinical diagnoses pretest did not match the phenotypes associated with the affected genes at the time of reporting the results to the requesting clinician, involving the following genes: *BBS1*, *IFT140* (2x), *KIF11*, *MCOLN1*, *PEX1* and *PRPS1*. Post-test reassessment revealed that four of these cases represented hidden syndromes [cases with *BBS1*, *IFT140* (1 of the 2 cases), *KIF11* and *MCOLN1* causative variants]. Of the remaining three, one of them is a case of phenotypic expansion (the second *IFT140* case), other had been diagnosed as syndromic pre-test, but without a specific diagnosis (the *PRPS1* case) and the remaining one corresponded to misdiagnosis of a Heimler syndrome (caused by a *PEX1* variant) as an Usher syndrome. These results reinforce the utility of using phenotype-agnostic panels for the genomic diagnosis of vision loss.

The two *IFT140* cases from our series, caused by the same mutation, showed very different phenotypes: the

male had a phenotype matching Mainzer-Saldino syndrome whereas the female had non-syndromic RP. Similarly, the non-syndromic patient with two causative variants affecting *ARL6* had a sister (not genetically tested) with polydactyly, which is strongly suggestive of Bardet-Biedl syndrome. These are clear examples of phenotypic heterogeneity that further reinforce the difficulty to predict the affected genes from clinical assessment and, as a result, the convenience of using comprehensive gene panels.

So far, the clinical application of NGS to the molecular diagnosis of hereditary blindness has followed two main paths: exome sequencing and panel sequencing. While exome sequencing has the advantage of being able to identify variants in any gene (thus not depending on gene selection and being able to discover novel genes not previously related to vision loss), this advantage is actually more useful in the research setting than in the clinical setting, where a solid molecular diagnosis, based on existing published gene-phenotype associations and able to guide patient management, is

Table 3. Gene therapy clinical trials aimed at correcting defects responsible for hereditary blindness diagnosed in this study. Trials currently open or open at some point between the start of the study and November 2019 are shown.

OFTALMO ID	Date of report	Sex	Gene	Causative mutations	Zigosity	Clinical trial title	Web links to trial details	Phase
OFTALMO.002	26/02/2016	F	ABCA4	c.886C>T, p.(Arg290Trp)	Homozygous	A Phase I/IIa Study of SAR422439 in Patients With Stargardt's Macular Degeneration	https://www.clinicaltrialsregister.eu/ctr-search/html/2019-023111-34/IT/	1, 2a
OFTALMO.019	06/04/2016	F		c.6449 G>A, p.(Cys21397Trp)	Homozygous			
OFTALMO.021	21/04/2016	M		c.5944_5988delCTTGGCAATCTGGCGTG, p.(Val1682_Val1686del) c.3943C>T, p.(Gln1315*)	Compound heterozygous			
OFTALMO.063	31/01/2017	M		c.3819T>C, p.(Leu1940Pro>G) c.3386G>T, p.(Arg1129Leu)	Compound heterozygous			
OFTALMO.072	19/05/2017	F		c.583AG>A, p.(Gly1961Glu)	Homozygous			
OFTALMO.076	24/08/2017	M		c.3056C>T, p.(Thr1019Met) c.2628G>A, p.(Thr2240Ala)	Compound heterozygous			
OFTALMO.084	28/03/2019	M		c.952C>T, p.(Gln328*) c.3832G>A, p.(Gly1961Glu)	Compound heterozygous			
OFTALMO.095	08/04/2019	F		c.3386G>T, p.(Arg1129Leu) c.6006>2A>G	Compound heterozygous			
OFTALMO.008	07/03/2016	M		c.1314delT, p.(Leu505Trp)fs*4)		Safety and Dose Escalation Study of AAV2$>v$CHM in Subjects With CHM (Choroideremia) Gene Mutations Long-term Safety and Efficacy Followup of AAV2-RP1 for the Treatment of Choroideremia (SOLSTICE)	https://clinicaltrials.gov/ct2/show/NCT02341807	1, 2
OFTALMO.020	20/04/2016	M	CHM	c.340C>T, p.(Gln114*)	Hemizygous	A Safety Study of Retinal Gene Therapy for Choroideremia (GEMINI)	https://clinicaltrials.gov/ct2/show/NCT03584465	?
OFTALMO.017	31/03/2016	M	MYO7A	c.60234delG, p.(Ala2009Pro)fs*32)	Homozygous	Choroideremia Gene Therapy Clinical Trial	https://clinicaltrials.gov/ct2/show/NCT02553135	2
OFTALMO.018	04/04/2016	F		c.397C>T, p.(His131Trp)	Homozygous	An Open Label Clinical Trial of Retinal Gene Therapy for Choroideremia	https://clinicaltrials.gov/ct2/show/NCT02077361	1, 2
OFTALMO.029	27/07/2016	M	RPGR	c.2384delG, p.(Cys762Lys)fs*33)	Hemizygous	Efficacy and Safety of AAV2-RP1 for the Treatment of Choroideremia (STAR)	https://clinicaltrials.gov/ct2/show/NCT03496012	3
OFTALMO.052	02/11/2016	F	USH2A	c.9304C>T, p.(Gln3102*) c.2276G>T, p.(Cys759Phe)	Compound heterozygous	Study of SAR421869 in Patients With Retinitis Pigmentosa Associated With Usher Syndrome Type IB Gene Therapy for X-linked Retinitis Pigmentosa (XLRP)	https://clinicaltrials.gov/ct2/show/NCT01505962	1, 2a
OFTALMO.068	18/05/2017	F		c.2276G>T, p.(Cys759Phe) c.12093delC, p.(Tyr1031*)	Compound heterozygous	Retinitis Pigmentosa GTPase Regulator (RPGR) A Clinical Trial of Retinal Gene Therapy for X-linked Retinitis Pigmentosa (XIRIUS)	https://clinicaltrials.gov/ct2/show/NCT0316113	1, 2, 3
						Safety and Efficacy of rAAV2YF-GRK1-RPGR in Subjects With X-linked Retinitis Pigmentosa Caused by RPGR- Gene Therapy Trial for People with Retinitis Pigmentosa (progressive reduction in vision due to a gene defect on USH2A Gene (Stellar))	https://clinicaltrials.gov/ct2/show/NCT03165660	1, 2
						Study to Evaluate Safety and Tolerability of QR-421a in Subjects With RP Due to Mutations in Exon 13 of the USH2A Gene (Stellar)	https://g4bclinicaltrialsdata.com/html/GCT0202016-003967-21	1, 2
							https://clinicaltrials.gov/ct2/show/NCT03780257	1, 2

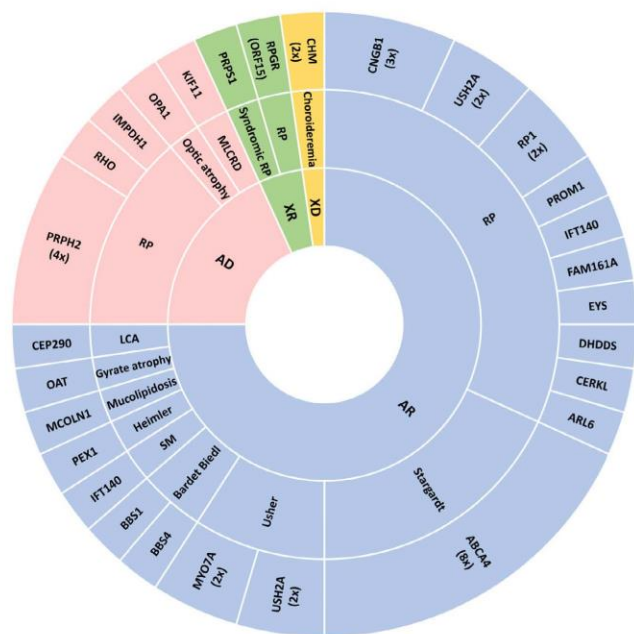


Fig. 2. Modes of inheritance of the genetic conditions considered responsible for hereditary vision loss in our cohort. From outer to inner rings, affected genes, associated phenotypes and modes of inheritance are shown. The area of each sector is proportional to the number of cases represented within it. For genes with causative alterations affecting more than one patient, the number of affected patients is shown in brackets. AD: autosomal dominant; AR: autosomal recessive; X-linked recessive, 5; LCA: Leber congenital amaurosis; MLCRD: microcephaly, lymphoedema, chorioretinal dysplasia syndrome; RP: retinitis pigmentosa; SM: Saldino–Mainzer syndrome; XD: X-linked dominant; XR: X-linked recessive.

required. In fact, this advantage is deemed to be less and less relevant with time, as a recent study shows that the discovery of novel genes associated to recessive IRD approaches saturation (Patel et al. 2018). As a result, up-to-date panels including both genes with strong/moderate association with inherited blindness and recently discovered candidate genes, as the one presented herein, should minimize this theoretical advantage of exome sequencing.

Consugar et al. concluded that focusing the sequencing and interpretation efforts in a panel with a limited subset of genes consistently associated with blindness phenotypes outperforms exome-based sequencing strategies (Consugar et al. 2015). This outperformance was based on improved detection rates (attributable to a probe design better tailored to genes associated with IRD, as probes for approx.

10% of the regions targeted by their panel were missing from commercially available exome capture kits) and reduced turnaround times. Time reduction does not only apply to laboratory geneticists, but also to clinicians, as exome sequencing demands from them additional time for the informed consent process (including discussions about incidental secondary findings), for interpreting laboratory reports and for reviewing results (many of which refer to conditions they are likely not familiar with) with patients (Biesecker & Green 2014; Gillespie et al. 2014). A more recent study, aimed at evaluating the performance of exome sequencing in the clinical context, concludes that the current standard of 120x average read depth may be insufficient for consistent breadth of coverage across all relevant genes (Kong et al. 2018).

Our test includes genes with all modes of inheritance in a single gene

panel. An advantage of this is the possibility of detecting dominant *de novo* or incomplete penetrance mutations in families with apparent recessive patterns of disease transmission. In total, in our series, 6 of the 45 cases (13.3%) with a genetic diagnosis have causative variants in genes associated with dominant phenotypes and, however, do not have a family history of IRDs (Table 2). Of note, a recessive IRD panel would not have detected these dominant causative variants.

As any diagnostic technique, panel sequencing has its difficulties and limitations. The ability of our panel to detect SNVs and indels with high analytical sensitivity and specificity has been validated in this study. However, certain variant types, such as trinucleotide expansions and CNVs, as well as those affecting highly homologous or repetitive regions, can be missed. Nonetheless, these limitations are not exclusive of panel sequencing, but extensive to other NGS approaches, such as exome or, to a lower extent, genome sequencing (Bamshad et al. 2011; Schrijver et al. 2012). One relevant IRD gene affected by such limitations is *RPGR*, specifically in its *ORF15* variant isoform (Audo et al. 2012). Although we initially observed difficulties in the capture, sequencing and/or alignment of sequences from this region, optimization of our capture probe has allowed us to overcome them in the current clinical version of the panel (v4) (Table S10).

The ability of our methodology to detect CNVs from panel data has previously been evaluated for a panel of 199 deafness-related genes (Cabanillas et al. 2018). In that case, our NGS pipeline was able to correctly detect all homozygous and heterozygous whole gene CNVs previously found by qPCR and MLPA affecting *STRC* (a particularly challenging gene because of its highly homologous *STRCP1* pseudogene). However, detection of partial gene deletions/duplications by NGS is even more challenging. The lack of homozygous or heterozygous deletions/duplications in the five genes we evaluated by MLPA in a subset of 31 patients without a genetic diagnosis from our current blindness series is encouraging. However, we do not rule out the possibility that these and other patients from this study without a

genetic diagnosis have causative variants in genes that were investigated, either because of the aforementioned limitations or because the regions affected by the variants were not included in the design and, thus, were not captured and sequenced.

It is also possible that some of the patients from the current study tested with the earlier versions of the probe have causative variants in genes included in the later versions. These are not likely to represent a large fraction of the cases, though, as all causative variants in our series affected genes present in all panel versions. Nonetheless, a limitation of the panel approach is that, if no causative variants are detected in the patient and a new disease-associated gene is discovered, that gene cannot be retrospectively evaluated from the available sequence. As mentioned before, exome data could provide this information with no extra sequencing.

The gene most frequently found to be responsible for the visual impairment phenotype in our series is *ABCA4* (8/46 cases, 17.4%) followed by *USH2A* and *PRPH2* (4/46 cases, 8.7%, each), *CNGB1* (3/46, 6.5%) and *CHM*, *IFT140*, *MYO7A* and *RP1* (2/46, 4.3%, each). The remaining 18 cases (40%) were each caused by variants in one of 18 different genes (Figs 1 and 2). The high prevalence of *ABCA4*, *USH2A* and *CHM* variants is in agreement with a recent and large retrospective analysis done by Motta et al. that evaluated 549 patients with IRD genetic testing in Brazil (Motta et al. 2018). Of the other four genes highlighted by Motta and collaborators as very prevalent disease-causing genes (*CEP290*, *CRB1*, *RPGR* and *CHM*), three of them (*CEP290*, *RPGR* and *CHM*) are also affected by causative mutations in our series, although not recurrently (Figs 1 and 2).

Of the 100 cases evaluated in our series, 15 (15%) could potentially benefit from clinical trials aimed at correcting their genetic defects, including those affected by mutations in *ABCA4* (8 cases), *USH2A* (2 cases), *CHM* (2 cases), *MYO7A* (2 cases) and *RPGR* (1 case). Thus, our results suggest that potential eligibility from a gene-corrective clinical trial with curative purposes is not a rare possibility for IRD patients. Moreover, this data, combined with the four cases in which a

hidden syndrome was revealed, indicate that the proportion of genetically diagnosed patients in which comprehensive genetic testing identified clinically relevant insights was at least 42.2% (19/45).

Genetic diagnosis is the key eligibility criterion for clinical trials and treatments based on gene therapy. With the advent of CRISPR/Cas9 technologies the possibility to correct not just the mutated genes but also their specific pathogenic variants has become a reality. This is the case of the *CEP290* c.2991 + 1655A>G intronic mutation, targeted by the EDIT-101 investigational new drug for type 10 Leber congenital amaurosis (Maeder et al. 2019). In this scenario, comprehensive genomic characterization should soon be a must in the clinical management of blind patients. Hopefully, a substantial proportion of the ongoing clinical trials will boost the approval of more therapies that, combined with appropriate genomic testing, will provide life-changing options for those affected by such highly disabling conditions.

References

- Audo I, Bujakowska KM, Leveillard T et al. (2012): Development and application of a next-generation-sequencing (NGS) approach to detect known and novel gene defects underlying retinal diseases. *Orphanet J Rare Dis* 7: 8.
- Bamshad MJ, Ng SB, Bigham AW, Tabor HK, Emond MJ, Nickerson DA & Shendure J (2011): Exome sequencing as a tool for Mendelian disease gene discovery. *Nat Rev Genet* 12: 745–755.
- Biesecker LG & Green RC (2014): Diagnostic clinical genome and exome sequencing. *N Engl J Med* 370: 2418–2425.
- Boudewyn LC & Walkley SU (2019): Current concepts in the neuropathogenesis of mucopolysaccharidosis type IV. *J Neurochem* 148: 669–689.
- Bowne SJ, Humphries MM, Sullivan LS et al. (2011): A dominant mutation in RPE65 identified by whole-exome sequencing causes retinitis pigmentosa with choroidal involvement. *Eur J Hum Genet* 19: 1074–1081.
- Cabanillas R, Dineiro M, Castillo D et al. (2017): A novel molecular diagnostics platform for somatic and germline precision oncology. *Mol Genet Genomic Med* 5: 336–359.
- Cabanillas R, Dineiro M, Cifuentes GA et al. (2018): Comprehensive genomic diagnosis of non-syndromic and syndromic hereditary hearing loss in Spanish patients. *BMC Med Genomics* 11: 58.
- Chiang JP, Lamey T, McLaren T, Thompson JA, Montgomery H & De Roach J (2015): Progress and prospects of next-generation sequencing testing for inherited retinal dystrophy. *Expert Rev Mol Diagn* 15: 1269–1275.
- Consugar MB, Navarro-Gomez D, Place EM et al. (2015): Panel-based genetic diagnostic testing for inherited eye diseases is highly accurate and reproducible, and more sensitive for variant detection, than exome sequencing. *Genet Med* 17: 253–261.
- Costa KA, Salles MV, Whitebirch C, Chiang J & Sallum JMF (2017): Gene panel sequencing in Brazilian patients with retinitis pigmentosa. *Int J Retina Vitreous* 3: 33.
- Estrada-Cuzcano A, Koenekoop RK, Senchal A et al. (2012): *BBS1* mutations in a wide spectrum of phenotypes ranging from nonsyndromic retinitis pigmentosa to Bardet-Biedl syndrome. *Arch Ophthalmol* 130: 1425–1432.
- Fiorentino A, Fujinami K, Arno G et al. (2018): Missense variants in the X-linked gene *PRPS1* cause retinal degeneration in females. *Hum Mutat* 39: 80–91.
- Gillespie RL, Hall G & Black GC (2014): Genetic testing for inherited ocular disease: delivering on the promise at last? *Clin Exp Ophthalmol* 42: 65–77.
- Hafler BP (2017): Clinical progress in inherited retinal degenerations: gene therapy clinical trials and advances in genetic sequencing. *Retina* 37: 417–423.
- Jauregui R, Cho GY, Takahashi VKL, Takiuti JT, Bassuk AG, Mahajan VB & Tsang SH (2018): Caring for hereditary childhood retinal blindness. *Asia Pac J Ophthalmol (Phila)* 7: 183–191.
- Kong SW, Lee IH, Liu X, Hirschhorn JN & Mandl KD (2018): Measuring coverage and accuracy of whole-exome sequencing in clinical context. *Genet Med* 20: 1617–1626.
- Lee K & Garg S (2015): Navigating the current landscape of clinical genetic testing for inherited retinal dystrophies. *Genet Med* 17: 245–252.
- Maeder ML, Stefanidakis M, Wilson CJ et al. (2019): Development of a gene-editing approach to restore vision loss in Leber congenital amaurosis type 10. *Nat Med* 25: 229–233.
- Motta FL, Martin RP, Filippelli-Silva R, Salles MV & Sallum JMF (2018): Relative frequency of inherited retinal dystrophies in Brazil. *Sci Rep* 8: 15939.
- Nishiguchi KM, Avila-Fernandez A, van Huet RA et al. (2014): Exome sequencing extends the phenotypic spectrum for *ABHD12* mutations: from syndromic to nonsyndromic retinal degeneration. *Ophthalmology* 121: 1620–1627.
- Patel N, Alkuraya H, Alzahrani SS et al. (2018): Mutations in known disease genes account for the majority of autosomal recessive retinal dystrophies. *Clin Genet* 94: 554–563.

- Perrault I, Saunier S, Hanein S et al. (2012): Mainzer-Saldino syndrome is a ciliopathy caused by IFT140 mutations. *Am J Hum Genet* **90**: 864–870.
- Ratbi I, Falkenberg KD, Sommen M et al. (2015): Heimler syndrome is caused by hypomorphic mutations in the peroxisome-biogenesis genes PEX1 and PEX6. *Am J Hum Genet* **97**: 535–545.
- Richards S, Aziz N, Bale S et al. (2015): Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med* **17**: 405–424.
- Robitaille JM, Gillett RM, LeBlanc MA et al. (2014): Phenotypic overlap between familial exudative vitreoretinopathy and microcephaly, lymphedema, and chorioretinal dysplasia caused by KIF11 mutations. *JAMA Ophthalmol* **132**: 1393–1399.
- Russell S, Bennett J, Wellman JA et al. (2017): Efficacy and safety of voretigene neparovovec (AAV2-hRPE65v2) in patients with RPE65-mediated inherited retinal dystrophy: a randomised, controlled, open-label, phase 3 trial. *Lancet* **390**: 849–860.
- Sadagopan KA (2017): Practical approach to syndromic pediatric retinal dystrophies. *Curr Opin Ophthalmol* **28**: 416–429.
- Schmidts M, Frank V, Eisenberger T et al. (2013): Combined NGS approaches identify mutations in the intraflagellar transport gene IFT140 in skeletal ciliopathies with early progressive kidney disease. *Hum Mutat* **34**: 714–724.
- Schrijver I, Aziz N, Farkas DH et al. (2012): Opportunities and challenges associated with clinical diagnostic genome sequencing: a report of the Association for Molecular Pathology. *J Mol Diagn* **14**: 525–540.
- Sharon D & Banin E (2015): Nonsyndromic retinitis pigmentosa is highly prevalent in the Jerusalem region with a high frequency of founder mutations. *Mol Vis* **21**: 783–792.
- Stone EM, Aldave AJ, Drack AV, Maccumber MW, Sheffield VC, Traboulsi E & Weleber RG (2012): Recommendations for genetic testing of inherited eye diseases: report of the American Academy of Ophthalmology task force on genetic testing. *Ophthalmology* **119**: 2408–2410.
- Valdes-Mas R, Bea S, Puente DA, Lopez-Otin C & Puente XS (2012): Estimation of copy number alterations from exome sequencing data. *PLoS ONE* **7**: e51422.
- Verbakel SK, van Huet RAC, Boon CJF et al. (2018): Non-syndromic retinitis pigmentosa. *Prog Retin Eye Res* **66**: 157–186.
- Vervoort R, Lennon A, Bird AC et al. (2000): Mutational hot spot within a new RPGR exon in X-linked retinitis pigmentosa. *Nat Genet* **25**: 462–466.
- Werdich XQ, Place EM & Pierce EA (2014): Systemic diseases associated with retinal dystrophies. *Semin Ophthalmol* **29**: 319–328.
- Xu M, Yang L, Wang F et al. (2015): Mutations in human IFT140 cause non-syndromic retinal degeneration. *Hum Genet* **134**: 1069–1078.
- study design, interpretation of results and writing of the manuscript. All authors commented on the manuscript and approved the submitted version.
- Data Availability Statement: The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.
- [Correction added on 10 June 2020, after first publication: the acknowledgements section has been updated with funding information in this version.]

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Table S1. List of phenotypes potentially related to visual impairment used as keywords for initial query on HGMD professional.

Table S2. Tier 1 and tier 2 genes included in v1 of the panel.

Table S3. Tier 1 and tier 2 genes included in v2 of the panel.

Table S4. Tier 1 and tier 2 genes included in v3 of the panel.

Table S5. Cases analyzed with each version of the panel.

Table S6. Primers used to validate causative variants (pathogenic and likely pathogenic variants considered responsible for the ophthalmologic phenotype).

Table S7. Coverages[¥] and callabilities[#] for every tier 1 and tier 2 gene on v4 samples ($n = 6$).

Table S8. SNV and indel sensitivity, specificity and accuracy of the OFTALMOgenics v3 platform.

Table S9. SNV and indel sensitivity, specificity and accuracy of the OFTALMOgenics v4 platform.

Table S10. RPGR target regions with <100% bases covered by ≥ 10 reads or ≥ 20 reads in R&D, clinical v3 and clinical v4 cases.

Table S11. Clinical characteristics of cases without causative mutations.

Table S12. Variants of uncertain significance identified in this study.

Table S13. Results from MLPA analysis on 31 cases from this study with no causative variants identified.

Received on December 20th, 2019.

Accepted on April 29th, 2020.

Correspondence:

Juan Cadiñanos, PhD and Ruben Cabanillas, MD, PhD
IMOMA
Avda. Richard Grandío s/n, 33193
Oviedo
Spain
Tel: +34985250300 Ext. 266 (JC)
Tel: +34985250300 Ext. 700 (RC)
Fax: +34985272166
Emails: jcb@imoma.es (JC);
rcabanillas@imoma.es (RC)

Acknowledgements: We are grateful to all participating patients and their families. Work performed at IMOMA has been supported by Fundación María Cristina Masaveu Peterson.

Conflict of interest: The following authors are or have been employed by IMOMA, the owner of the OFTALMOgenics™ platform: MD (Clinical Molecular Geneticist), RaC (Molecular Biologist), GAC (Biotechnologist), AO (Genetic Counsellor), AS (Bioinformatician), NSD (Lab. Technician), RA (Lab Technician), RuC (Director of the Area of Precision Medicine) and JC (Director of the Laboratory of Molecular Medicine). The authors declare no other conflicts of interests.

Ethics approval and consent to participate: The study was approved by the Comité de Ética de Investigación del Principado de Asturias (research project #74/14). Consent was obtained from all patients or their parents.

Author's contributions: MD performed most of the experimental work and contributed to results analysis and interpretation and writing of the manuscript. RaC contributed to results interpretation and data analysis. GAC contributed to experimental work and results interpretation. BF-V, EV, AO, MV-D, IH and AF-V provided patient samples and associated clinical data. AS contributed to results analysis. PCP, DC and GRO performed bioinformatics analyses. NSD, RA and CGL contributed to experimental work. RuC and JC participated in

DISCUSIÓN

La medicina de precisión se basa en la integración de los aspectos genéticos y ambientales de la enfermedad para individualizar la prevención, el pronóstico, el diagnóstico y el tratamiento, ponderando especialmente todos los datos moleculares disponibles para la toma de decisiones en la práctica clínica.

En los últimos años, los estudios moleculares convencionales, consistentes habitualmente en el análisis de un único gen, están siendo reemplazados por la evaluación simultánea de decenas/cientos de genes implicados en una patología. Mediante esta aproximación, basada en la secuenciación de nueva generación (NGS), la obtención de información molecular es más rápida y costo-efectiva que el análisis por separado de cada uno de los genes diana, permitiendo obtener rendimientos diagnósticos muy superiores a los alcanzables mediante las metodologías convencionales.

Diagnóstico genómico del cáncer (Artículo I)

El aprovechamiento de los datos moleculares de cada paciente para afinar la prevención, el diagnóstico, pronóstico y/o tratamiento del cáncer ha conseguido impulsar un cambio de paradigma en el manejo clínico oncológico. La NGS ha proporcionado nuevas alternativas a la oncología clínica, desde la selección de terapias dirigidas contra las alteraciones que impulsan el desarrollo de un tumor hasta el asesoramiento genético en casos de cáncer familiar. Por este motivo, se considera esencial la integración de plataformas de diagnóstico NGS validadas en la práctica clínica oncológica.

El planteamiento de partida de este proyecto consideró que analizar al mismo tiempo las alteraciones genéticas somáticas, exclusivas del tumor, y las alteraciones germinales, presentes en todas las células del individuo, es lo ideal desde un punto de vista científico-académico y, además, permitiría al oncólogo disponer de toda la información molecular clínicamente relevante a la hora de tomar decisiones fundamentales para combatir la enfermedad.

El artículo I de la presente Tesis Doctoral describe el desarrollo de una plataforma de diagnóstico NGS destinada a detectar (1) alteraciones genómicas somáticas asociadas con la respuesta a terapias dirigidas aprobadas contra el cáncer y (2) mutaciones de la línea germinal que predisponen a neoplasias hereditarias. El panel descrito, en su primera versión clínica (v6), constaba de un total de 215 genes asociados a cáncer: 97 para la selección de terapias dirigidas y 148 asociados con predisposición hereditaria al cáncer (con 30 genes informativos para ambas aplicaciones). El panel incluye intrones seleccionados de 17 genes afectados por reordenamientos asociados con al menos una terapia dirigida. Su aplicación está indicada para pacientes con enfermedad en estadio IV que han progresado a opciones de tratamiento estándar y/o para aquellos con una potencial predisposición hereditaria al desarrollo de cáncer. El análisis de 39 muestras tumorales y 36 muestras germinales permitió identificar terapias dirigidas potencialmente asociadas a las alteraciones presentes en el tumor en un 89,74 % de los casos (35/39), así como alteraciones germinales asociadas con cáncer hereditario en un 5,5 % de los casos (2/36).

El panel está siendo objeto de revisión y optimización continua. El rediseño de sondas ha permitido obtener valores de *callability* a DP20 (porcentaje de nucleótidos diana que tienen 20 lecturas o más) superiores al 99,5 % en todos los casos. Detecta variantes de un solo nucleótido (SNV) e inserciones/delecciones (indels) con sensibilidad y especificidad superiores al 99,5 % (para variantes con frecuencia alélica $\geq 0,1$), así como variantes de número de copia (CNV) y reordenamientos. Los programas de control externo de calidad a los que se ha sometido el panel anualmente desde 2016, organizados por la Red Europea de Calidad de Genética Molecular (*European Molecular Genetics Quality Network, o EMQN*), han corroborado sistemáticamente su idoneidad para el diagnóstico molecular del cáncer.

Además del propio diseño químico del panel, el análisis de los resultados de secuenciación NGS también ha sido desarrollado y mejorado para superar las dificultades que entraña esta metodología. Por ejemplo, el genoma de referencia contiene diversas regiones que presentan altos niveles de homología entre sí y la

correcta identificación de variantes genéticas en estas regiones representa un desafío para las metodologías de NGS más utilizadas en la práctica clínica, basadas en lecturas individuales de 75-300 nucleótidos. En el año 2016, Mandelker y col. realizaron un inventario de este tipo de regiones¹⁸⁸. De entre los genes incluidos en el panel ONCOgenics™, la única región genómica que fue considerada altamente problemática en ese trabajo está localizada en el gen *PMS2*¹⁸⁸. Este gen tiene un pseudogén cercano denominado *PMS2CL*, cuyas secuencias codificantes de los exones 11 al 15 presentan un 98 % de identidad de secuencia con *PMS2*. Esto complica el alineamiento preciso de las lecturas generadas por el secuenciador y, en consecuencia, la identificación de variantes en esos exones¹⁸⁹.

Durante la evaluación del panel ONCOgenics™, realizamos el análisis de 5 pacientes en los que el gen *PMS2* había sido previamente evaluado por PCR de largo alcance, capaz de amplificar específicamente *PMS2* (distinguiéndolo de *PMS2CL*) seguida de secuenciación de Sanger¹⁹⁰⁻¹⁹². Los resultados obtenidos indicaron que ONCOgenics™ es capaz de detectar correctamente las variantes que no afectan los exones 11-15 de *PMS2*, mientras que puede detectar con una sensibilidad del 100 % las variantes que afectan a esos exones más problemáticos. Sin embargo, al igual que sucede con otras técnicas de diagnóstico, la aproximación seguida para lograr la máxima sensibilidad compromete la especificidad. Por esta razón, las variantes identificadas que afectan potencialmente a los exones 11-15 de *PMS2* deben ser validadas mediante aproximaciones alternativas específicas que permitan distinguir entre el gen y su pseudogén.

Detectar y reportar en paralelo tanto alteraciones somáticas asociadas con la respuesta a terapias contra el cáncer como alteraciones germinales que predisponen a neoplasias hereditarias tiene ventajas para el paciente y para su manejo clínico. Por un lado, evita ignorar la presencia de mutaciones germinales relevantes que causan cáncer hereditario, algo que puede ocurrir con relativa frecuencia si solo se analiza el ADN tumoral. Asimismo, un adecuado planteamiento bioinformático impide la substracción no supervisada de variantes germinales terapéuticamente accionables presentes tanto en el tumor como en la línea germinal. La inclusión del análisis de la línea germinal en

cada caso también brinda la posibilidad de identificar portadores no sospechados de mutaciones que causan síndromes de susceptibilidad al cáncer bien conocidos en familias con fenotipos atípicos, que no cumplen con los criterios de diagnóstico establecidos^{110,193}. Asimismo, la identificación fidedigna de variantes somáticas (al distinguirlas mejor de las germinales), puede permitir una predicción más precisa de la probabilidad de respuesta a la inmunoterapia¹⁹⁴⁻¹⁹⁸.

Más allá de sus objetivos principales (identificación de biomarcadores terapéuticos), el perfil genómico completo del tumor puede identificar posibles errores en el diagnóstico clínico. Por ejemplo, uno de los tumores analizados por la plataforma ONCOgenics™ había sido inicialmente clasificado como un adenocarcinoma metastásico de un cáncer de mama, ya que se manifestó como un ganglio linfático mediastínico aislado que fue sometido exclusivamente a análisis citológico y había sido identificado en una paciente que había tenido un tumor primario de mama 12 años antes. Sin embargo, la identificación por ONCOgenics™ de la indel p.E746_S752delins, localizada en el exón 19 de *EGFR* y característica de adenocarcinomas pulmonares, motivó la reevaluación anatomopatológica e inmunohistoquímica del tumor y su reclasificación definitiva como adenocarcinoma primario de pulmón. En consecuencia, posteriormente, la paciente fue tratada con un inhibidor de la tirosina quinasa de *EGFR* (gefitinib), obteniendo una respuesta parcial al tratamiento, que se sostuvo durante 21 meses. De igual modo, el análisis genómico de la línea germinal también reveló hallazgos clínicamente relevantes e inesperados, como la presencia de la alteración patogénica en mosaico *NF1* p.I679Dfs*21 (frecuencia alélica 10,3 %) en una paciente sana con antecedentes familiares de cáncer gastrointestinal, cáncer de próstata y leucemia.

Los reordenamientos cromosómicos desempeñan un papel importante en el proceso de tumorigénesis, representando aproximadamente el 20 % de la morbilidad en el cáncer humano¹⁹⁹. En los últimos años se están desarrollando cada vez más fármacos cuyas dianas son genes de fusión producidos como consecuencia de translocaciones. En algunos casos, como en los que implican a las quinasas de receptores neurotróficos (genes *NTRK1*, *NTRK2* y *NTRK3*), los intrones implicados en

esas translocaciones son tan extensas que requerirían una gran cantidad de sondas para lograr capturarlos por completo y obtener una calidad de secuencia adecuada como para identificar los puntos de corte²⁰⁰. Una alternativa para mejorar el rendimiento diagnóstico en esas regiones complejas es añadir al análisis de ADN el análisis de ARN dirigido, orientado a secuenciar los cDNAs que codifican las proteínas de fusión que implican a esos genes²⁰¹.

Las recomendaciones actuales aconsejan la verificación mediante PCR y secuenciación Sanger (para CNVs e indels) u otras metodologías, como el MLPA (para grandes deleciones/duplicaciones), de las variantes clínicamente relevantes identificadas por NGS^{193,202}. Sin embargo, previsiblemente la NGS se llegará a establecer como técnica *gold-standard* y, en consecuencia, no será necesaria la validación de resultados detectados por secuenciación masiva mediante metodologías alternativas. De hecho, en análisis somáticos no es práctico hacer dicha validación porque, en muchos casos, las frecuencias alélicas de las variantes detectadas están por debajo del límite de detección de la secuenciación Sanger y de otras tecnologías convencionales y porque, además, el tiempo de respuesta es lo suficientemente crítico como para no justificar un retraso en la entrega de resultados debido a la validación. En esos casos, una vez establecido que un determinado panel somático tiene sensibilidad y especificidad adecuadas, ya no es necesario hacer la validación de cada variante. En el caso de los paneles germinales, en cambio, debido a que las variantes tienen frecuencias de 0,5 o 1, a que el tiempo de respuesta no es tan apremiante y a que las consecuencias de los resultados no solo afectan al probando, sino también a sus familiares, realizar la validación mediante metodologías alternativas aún es recomendable. Además, ese proceso de validación de la variante en el probando permite la verificación del test necesario para la realización de estudios genéticos dirigidos en sus familiares.

La interpretación clínica de los resultados genómicos es uno de los desafíos más importantes para lograr la generalización de la aplicación clínica de las tecnologías NGS. Las variantes somáticas identificadas en un tumor funcionan como una firma molecular que puede proporcionar información en cuanto al diagnóstico, pronóstico o una posible terapia dirigida. En los últimos años, con la publicación de un número

creciente de proyectos de secuenciación del genoma de múltiples tumores, se ha generado una gran cantidad de información genómica consolidada en un número creciente de bases de datos públicas que sirven como repositorios para facilitar la interpretación clínica de estas alteraciones⁹³. En cuanto a las variantes germinales, su interpretación está estandarizada según las recomendaciones proporcionadas por las directrices ACMG/AMP publicadas en 2015, cuya aplicación aún requiere de una importante intervención humana que debe incluir la revisión cuidadosa de las publicaciones que respaldan las clasificaciones en ClinVar, HGMD, Pubmed y/o Google Scholar.

El informe es una parte esencial de cualquier prueba de laboratorio y debe contener toda la información necesaria de manera clara, sencilla y concisa para que el especialista que solicita el estudio interprete los resultados sin apenas dificultad. Una representación incompleta o poco clara de los datos puede conducir a errores clínicos y al manejo incorrecto del paciente. Por este motivo, y debido a que las implicaciones de los estudios somáticos difieren de las de los estudios germinales, la separación de los informes de ambos análisis es la alternativa más adecuada para simplificar los informes y conseguir una mejor comprensión que redundará en un mejor manejo clínico del paciente con cáncer.

A pesar del gran potencial de los enfoques de NGS en oncología, siguen existiendo desafíos en el análisis integral y exhaustivo del perfil mutacional de muestras tumorales y germinales. Por ejemplo, en muchos casos, la inclusión de un número elevado de genes en el análisis somático desencadena la identificación de variantes genómicas que, a pesar de haber sido asociadas con terapias que están aprobadas para el tratamiento de algún tipo de cáncer en la literatura científico-médica, no lo están para el tipo de cáncer que presenta el paciente y solo son accesibles a través de ensayos clínicos. Este fue el caso en 34 de los 35 casos en los que se identificaron alteraciones asociadas a potenciales terapias. Además, cuando el número de genes diana del análisis somático es muy elevado, para conseguir las profundidades de secuencia necesarias para detectar con confianza mutaciones somáticas subclonales o presentes en muestras con alto contenido de ADN normal, la cantidad de secuencia

necesaria se dispara, lo que incrementa los costes. Asimismo, aunque, como se muestra en la figura 2 del artículo I, es posible establecer una sistemática para manejar las distintas combinaciones de indicación de estudios moleculares, tipos de muestras disponibles y alcance de los consentimientos informados a la hora de plantear un análisis conjunto tumoral/germinal, esta aproximación no deja de representar una mayor dificultad para el oncólogo que el planteamiento de cada tipo de análisis por separado.

Siendo conscientes de estas limitaciones, desde la publicación del artículo I de la presente Tesis Doctoral se ha desdoblado y simplificado la plataforma ONCOgenics™ para ajustarla a las necesidades diversas de oncólogos y genetistas. Así, se han desarrollado dos nuevos paneles: por un lado un panel tumoral, denominado ONCOgenics™ *Tumor*^{Focus}, que analiza variantes puntuales, CNVs y reordenamientos en 19 genes seleccionados porque sus alteraciones son reconocidas por las agencias reguladoras de los medicamentos (EMA y FDA) como biomarcadores de respuesta a fármacos antitumorales y porque las terapias a las que se asocian son habitualmente accesibles [**Anexo A**]; por el otro, un panel germinal, denominado ONCOgenics™ *Germline*^{Focus}, con 45 genes asociados consistentemente con la predisposición a neoplasias hereditarias y seleccionados en base a la utilidad clínica de la detección de sus alteraciones [**Anexo B**]. La selección de genes a analizar en este último panel está abierta, llegando a ser posible el análisis desde 1 a los 148 genes del panel original, incluyendo preselecciones de mini-paneles virtuales en función de los antecedentes personales y familiares de cáncer del paciente.

Idealmente, con el paso del tiempo, se producirán avances tanto en los aspectos tecnológicos (ej. abaratamiento de los costes de secuenciación que permita obtener secuencias genómicas completas de muestras tumorales a grandes profundidades) como en la interpretación bioinformática y clínica de los resultados, que permitan realizar análisis genómicos integrales del tumor y de la línea germinal en paralelo en cada caso, con el objetivo de que todos los pacientes oncológicos puedan beneficiarse de las enormes implicaciones que la información genética tiene para el manejo de su enfermedad.

Las enfermedades que afectan a los órganos de los sentidos se denominan enfermedades neurosensoriales y, en una proporción muy superior a la del cáncer, tienen un componente hereditario determinante en su desarrollo, es decir, se deben a alteraciones genéticas que pueden transmitirse de padres a hijos. Además de la evidente utilidad de la identificación de esas mutaciones para aportar un adecuado asesoramiento reproductivo tanto a los individuos afectados como a sus familiares, la identificación de la causa última de la enfermedad tiene consecuencias relevantes para el manejo clínico de una gran proporción de los pacientes diagnosticados genéticamente. En la presente Tesis Doctoral hemos desarrollado y aplicado dos nuevas plataformas de análisis NGS para el diagnóstico genómico de la sordera y la ceguera, cuyos resultados, presentados en los artículos **II**, **III** y **IV** y discutidos a continuación, demuestran la enorme utilidad clínica de estas herramientas.

Diagnóstico genómico de la hipoacusia hereditaria (Artículos II y III)

La pérdida auditiva es uno de los trastornos hereditarios conocidos más heterogéneos genéticamente. El elevado número de genes involucrados en esta patología representa un desafío para la identificación de alteraciones genómicas causales, necesaria para lograr su diagnóstico molecular²⁰³. Debido al gran rendimiento diagnóstico de la NGS en esta área, tanto las guías del ACMG específicas para la evaluación clínica y el diagnóstico etiológico de la hipoacusia, como las recomendaciones españolas de la CODEPEH la incluyen en el algoritmo de diagnóstico estándar de la hipoacusia neurosensorial (HNS)^{118 134}.

En el artículo **II** de la presente Tesis Doctoral se ha desarrollado un panel que, en su primera versión clínica (v3), constaba de 154 genes consistentemente asociados a sordera hereditaria neurosensorial y/o mixta, síndrómica y/o no síndrómica, y 45 genes candidatos. 76 de estos genes se habían asociado únicamente a formas no síndrómicas de sordera, 88 se habían relacionado exclusivamente con sordera en el contexto de trastornos síndrómicos y 35 estaban asociados tanto con formas síndrómicas de la enfermedad como con hipoacusia no síndrómica. El panel es revisado

y optimizado de manera continua (la versión más reciente es la v5, que incluye 262 genes [*Anexo C*]).

Una característica fundamental de cualquier panel diagnóstico es que sea capaz de obtener secuencia de la suficiente calidad en la máxima cantidad posible de nucleótidos diana. El rediseño de sondas ha permitido obtener valores de *callability* a DP20 superiores al 99,5 % de manera rutinaria. Gracias a esto, la sensibilidad analítica de la prueba para detectar SNVs e indels con frecuencias alélicas superiores a 0,1 fue mayor del 99,5 %, con una especificidad por encima del 99,9 % en la v3. También se ha demostrado la capacidad del panel para detectar específicamente CNVs y alteraciones en genes con pseudogenes, entre los cuales se encuentran *STRC* y *OTOA*. En relación con otros genes del panel OTOgenics™ con secuencias con muy alto grado de homología en otras regiones del genoma de referencia, en el artículo III de la presente Tesis Doctoral, tras el análisis de más de 70 individuos de diferentes orígenes étnicos, hemos obtenido evidencias que apoyan consistentemente la posibilidad de que el pseudogén *KCNE1B*, ausente en la versión GRCh37 del genoma humano, sea en realidad un artefacto introducido en la versión GRCh38.

En el artículo II se analizaron 50 pacientes con HNS presuntamente hereditaria no causada por mutaciones en *GJB2/GJB6*, *OTOF* ni *MT-RNR1*. El rendimiento diagnóstico obtenido en esta cohorte de pacientes pre-cribados fue del 42 % (21/50). Teniendo en cuenta los fenotipos de hipoacusia hereditaria asociados a los genes considerados responsables de sordera en esta cohorte, un 47,6 % (10/21) mostraron patrón de herencia autosómico recesivo, un 38,1 % (8/21) autosómico dominante y un 14,3 % (3/21) ligado al cromosoma X. El 28,6 % de los casos diagnosticados genéticamente (6/21) tenían síndromes previamente no detectados y el 19 % (4/21) fueron atribuibles a grandes deleciones/duplicaciones [*Figura 6*].

Nuestros resultados han contribuido a definir el espectro de alteraciones genéticas responsables de la hipoacusia en la población española, resaltando la heterogeneidad genética característica de la pérdida auditiva neurosensorial. Identificamos variantes genéticas consideradas responsables de hipoacusia en 16

genes diferentes que explicaban el fenotipo de 21 pacientes. Los genes alterados más habitualmente fueron *ACTG1* (n = 3), *USH2A* (n = 3) y *STRC* (n = 3). Aunque, según la literatura, tras *GJB2*, los genes más frecuentemente alterados en HNS hereditaria son *USH2A* y *STRC*, en nuestra cohorte de 50 pacientes *ACTG1* explica la sordera en el mismo número de casos que cada uno de estos (n=3)^{204,205}.



Figura 6. Modos de herencia y carácter sindrómico/no sindrómico de las patologías genéticas consideradas responsables de la pérdida auditiva hereditaria en nuestra cohorte de pacientes sin mutaciones en *GJB2/GJB6*, *OTOF* ni *MT-RNR1*.

De los anillos externos a los internos, se muestran los genes afectados, los fenotipos asociados y sus modos de herencia. El área de cada sector es proporcional al número de casos representados dentro de él. Para los genes con alteraciones causales que afectan a más de un paciente, el número de pacientes afectados se muestra entre paréntesis. AD: autosómico dominante; AR: autosómico recesivo; XD: dominante ligado al X; XR: recesivo ligado al X; HNS: hipoacusia neurosensorial. Usher, Waardenburg, Barakat, CHARGE y Alport representan los síndromes homónimos.

Las variantes de *ACTG1* son responsables de DFNA20/DFNA26 y del síndrome de Baraitser-Winter tipo 2. Hasta el momento, ninguno de nuestros pacientes con mutaciones en *ACTG1* ha manifestado características sindrómicas y todos mostraron HNS profunda de inicio temprano (prelocutiva), lo cual expande el espectro fenotípico de *ACTG1*, cuyas mutaciones se asocian generalmente a HNS postlingual y progresiva^{206,207}. Ninguna de las 3 variantes de *ACTG1* consideradas responsables de sordera hereditaria en nuestra cohorte inicial se había descrito previamente y 2 de ellas eran *de novo*. Por tanto, la detección de estas variantes se habría visto comprometida si se hubieran utilizado metodologías como la detección de mutaciones frecuentes previamente descritas (panel *hot-spot*) o como el análisis exclusivo de genes asociados a sordera recesiva (que, por definición, dejaría de detectar mutaciones dominantes surgidas *de novo*, ya que la ausencia de historia de sordera en las familias de los pacientes en los que ocurren induce a pensar que su modo de herencia es recesivo). Por estas razones, la prevalencia real de variantes patogénicas en *ACTG1* podría ser más alta y su patrón de expresión más variable de lo que se pensaba previamente²⁰⁸.

La proporción de HNS sindrómica en nuestra cohorte es del 38.1 % (8/21), lo cual está dentro de las tasas esperadas. En cambio, el 38 % de casos con modo de herencia AD y el 14 % con HNS ligada al cromosoma X son valores superiores a lo descrito previamente²⁰⁹. Esto podría ser consecuencia de que los pacientes analizados habían sido cribados previamente para excluir las causas AR más frecuentes (especialmente las mutaciones en *GJB2/GJB6*, pero también las que afectan a *OTOF*) y las alteraciones mitocondriales en *MT-RNR1*^{67,210}. El 50 % de las variantes patogénicas dominantes detectadas (4/8) surgieron *de novo*, es decir, estaban presentes en pacientes sin antecedentes familiares de hipoacusia (2 en *ACTG1*, 1 en *GATA3* y 1 en *COL4A5*) (**Tabla 4**). A pesar del tamaño limitado de nuestra cohorte, nuestra tasa de detección *de novo* se encuentra dentro del rango observado en estudios recientes de secuenciación de exoma completo (WES) realizados sobre pacientes con distintas patologías hereditarias (37-68 %) ^{67,211}.

Tabla 4: Variantes de novo

ID caso	Gen	Variante	Clasificación ACMG	Fenotipos audiológicos asociado al gen	Fenotipos audiológicos observados en pacientes (pre-test)
OTO.007	COL4A5	c.3525_3529dup; p.(Pro1177Leufs*124)	Patogénica	Síndrome de Alport	Síndrome de Alport
OTO.019	GATA3	c.1018A>C; p.(Asn340His)	Probablemente patogénica	Síndrome de Barakat	HNS bilateral no sindrómica
OTO.023	ACTG1	c.548G>A; p.(Arg183Gln)	Probablemente patogénica	Hipoacusia no sindrómica (DFNA20/DFNA26) / Síndrome de Baraitser-Winter	HNS bilateral no sindrómica
OTO.041	ACTG1	c.848T>C; p.(Met283Thr)	Probablemente patogénica	Hipoacusia no sindrómica (DFNA20/DFNA26) / Síndrome de Baraitser-Winter	HNS bilateral no sindrómica

El rendimiento diagnóstico descrito para técnicas NGS en pacientes con sospecha de sordera congénita AR se aproxima al 60 %¹³³. Como se ha mencionado previamente, en nuestra serie, en consonancia con las principales recomendaciones del manejo clínico del paciente con hipoacusia, los casos evaluados con el panel NGS habían sido preseleccionados descartando aquellos con mutaciones en los genes previamente descritos como responsables mayoritarios de HNS hereditaria en nuestra población (*GJB2/GJB6*, *OTOF* y *MT-RNR1*^{135,212}). Durante el proceso de preselección de pacientes para el análisis NGS, la evaluación de estos genes tuvo un rendimiento diagnóstico del 18,9 % (datos no publicados). Esta cifra, combinada con una tasa de diagnóstico estimada en el 42-48 % para nuestro panel en pacientes pre-cribados (el valor de 48 % resulta de considerar como potenciales causantes de hipoacusia 3 variantes sospechosas que en el momento de la publicación del artículo **II** no cumplían suficientes criterios de patogenicidad: 2 SNVs en homocigosis en los genes *LOXHD1* y *SLC26A4* y 1 SNV en hemizigosis que afectaba el gen *OTOA*), sugiere que la aplicación de nuestro panel a una población no cribada podría identificar la causa de la hipoacusia en el 53-58 % de los pacientes. Este valor es comparable al 56 % descrito recientemente por

Costales y colaboradores en otro trabajo en el que participó nuestro laboratorio y que consideraba 27 pacientes no cribados previamente y analizados mediante el panel OTOgenics™ [*Anexo D*]²¹³.

La tasa de detección del 42-48 % se encuentra en el promedio reportado para otros paneles NGS aplicados a pacientes con hipoacusia: 41 % (10-83 %), un rango amplio obtenido a partir de estudios que incluyen tanto casos previamente cribados como no cribados^{133,214}. Uno de los motivos que contribuyen a ese rendimiento es la inclusión de genes sindrómicos, los cuales permiten revelar "síndromes ocultos", es decir, síndromes que no habían sido diagnosticados clínicamente antes de la realización del estudio genético.

La sordera es una manifestación presente en 300-400 síndromes diferentes, algunos de ellos muy poco frecuentes, en los que la gravedad de la audición varía²¹⁵. La extrema heterogeneidad genética y la expresividad variable de estos fenotipos sindrómicos representan grandes desafíos para su evaluación clínica, lo cual dificulta la identificación de un síndrome sin un estudio genético que lo soporte^{216,217}. En nuestra cohorte, hemos identificado síndromes ocultos en 6 pacientes (Waardenburg tipo 2A (x1), Waardenburg tipo 2E (x1), Usher tipo 2A (x3) y Barakat (x1)), lo cual representa el 28,6 % (6/21) de los casos diagnosticados genéticamente. Además, en otros 6 individuos se identificaron variantes patogénicas en genes asociados tanto a fenotipos sindrómicos como no sindrómicos: 1x *CDH23* (síndrome de Usher 1D y DFNB12), 3x *ACTG1* (síndrome de Baraitser-Winter tipo 2 y DFNA20 / DFNA26), 1x *BSND* (tipo de síndrome de Bartter IV y DFNB73) o 1x *SLC26A4* (síndrome de Pendred y DFNB4).

En el supuesto de que en estos últimos pacientes se confirmara la presencia de fenotipos sindrómicos, la proporción de pacientes con síndromes ocultos podría ascender hasta el 57,1 % (12/21). En estos casos, un seguimiento cercano resulta fundamental para identificar potenciales alteraciones extra-audiológicas, que se podrían desarrollar en cualquier etapa de la vida. Por lo tanto, el diagnóstico temprano

y preciso de la pérdida auditiva sindrómica puede aportar información muy relevante, y no disponible de otro modo, de cara al pronóstico y del manejo clínico del paciente.

La interpretación clínica de los hallazgos genómicos es un componente crítico para un diagnóstico genético preciso. Las discrepancias en la interpretación y clasificación de variantes pueden tener serias implicaciones para la atención al paciente²¹⁸. Un estudio relativamente reciente indicó que hasta el 30 % de todas las variantes genéticas causantes de enfermedades descritas en la literatura podrían haberse interpretado erróneamente²¹⁹. Los resultados obtenidos en la cohorte de hipoacusia hereditaria, incluida en la presente Tesis Doctoral, fueron interpretados manualmente en base a las directrices ACMG/AMP publicadas en 2015, una serie de instrucciones generales, aplicables a cualquier patología hereditaria mendeliana, que tienen como objetivo la clasificación de las variantes genéticas utilizando la evidencia disponible en el momento del análisis, bajo la perspectiva del fenotipo del paciente y de sus antecedentes familiares¹⁸¹.

En 2018 se publicaron las “Especificaciones de las directrices ACMG/AMP para la interpretación de variantes genéticas asociadas a hipoacusia”. En estas especificaciones, elaboradas por el Panel de Expertos en Hipoacusia Hereditaria de ClinGen, de los 28 criterios ACMG/AMP publicados en 2015 se eliminaron 4, se detallaron 21 y se mantuvieron sin cambios 3²²⁰. La revisión de las 3 variantes de significado incierto identificadas en casos de nuestra cohorte considerados “no resueltos” y mencionados con anterioridad, para cuya evaluación, en el momento del estudio genético, no se disponía de información suficiente, nos ha permitido reclasificar como probablemente patogénica la variante *LOXHD1* c.3571A>G, p.Thr1191Ala, identificada en nuestra cohorte inicial en un paciente con hipoacusia neurosensorial y que posteriormente hemos vuelto a detectar en homocigosis en otra paciente no relacionada, con un fenotipo similar. Incluyendo este caso, el rendimiento diagnóstico del panel OTOgenics™ ascendería al 44 % en la cohorte inicial y al 55 % en pacientes no cribados. Esto demuestra que la estandarización, el refinamiento y la adaptación de los criterios ACMG/AMP de interpretación de variantes genéticas a cada patología aumenta la resolución de la clasificación de variantes, lo cual, en última instancia,

debería conducir a una mayor tasa de diagnóstico y, en definitiva, a una mejor atención a los pacientes.

En la actualidad se están aplicando principalmente dos tipos de test basados en NGS para el diagnóstico genómico de las enfermedades hereditarias: los paneles de genes y la secuenciación de exoma completo. En el caso de la sordera hereditaria, la secuenciación del exoma completo, al abarcar un número de genes diana casi 100 veces superior al de los paneles (considerando aprox. 20.000 vs. aprox. 200), requeriría recursos de secuenciación mucho mayores para obtener coberturas similares en las regiones genómicas de interés. En la práctica, la cobertura de los exomas no alcanza a la de los paneles, lo cual puede provocar coberturas insuficientes en regiones clave y, en consecuencia, la no detección de variantes clínicamente relevantes. Además, la mayor cantidad de secuencia requerida para la secuenciación de exomas encarece los costes de secuenciación y complica el análisis bioinformático²²¹. Por último, el análisis de exoma completo genera un mayor número de hallazgos secundarios (aquellos que causan patologías no relacionadas con el fenotipo objeto de estudio) que dificultan el procedimiento de asesoramiento genético²²².

No obstante, la secuenciación del exoma completo tiene una ventaja con respecto a la secuenciación de paneles: la capacidad de identificar alteraciones en genes que, en el momento del análisis, no se han asociado con la enfermedad. Estas alteraciones podrían servir para explicar el fenotipo del paciente si, en estudios posteriores, se demuestra la asociación consistente del gen afectado con el fenotipo objeto de estudio. Para minimizar esta desventaja, durante el desarrollo de los paneles OTOgenics™ implementamos un enfoque con dos grupos de genes: genes consistentemente asociados con HNS (*tier 1*) y genes candidatos (*tier 2*). Las variantes identificadas en los genes candidatos no se incluyen en el informe sistemáticamente, sino solo en aquellos casos en los que, en base a la evidencia disponible y a la coincidencia entre el fenotipo del paciente y el asociado al gen afectado, podrían explicar la hipoacusia. Esta solución de compromiso tiene como objetivo maximizar el rendimiento diagnóstico del panel sin complicar el procedimiento de asesoramiento genético.

Los resultados obtenidos en nuestra cohorte subrayan la importancia de un enfoque integral con una cuidadosa selección de genes para el diagnóstico genético de la HNS. De este modo, la NGS puede transferirse a la práctica clínica, aumentando el rendimiento del diagnóstico genético de la hipoacusia al 50–60 % (incluyendo las alteraciones *GJB2/GJB6*), mejorando la precisión diagnóstica/pronóstica, refinando el asesoramiento genético y reproductivo y revelando síndromes no diagnosticados clínicamente con implicaciones relevantes para los pacientes.

Diagnóstico genómico de la ceguera hereditaria (Artículo IV)

De una manera similar a lo observado en la hipoacusia, en el caso de la ceguera hereditaria la compleja heterogeneidad genética de la enfermedad hace muy difícil predecir la alteración responsable del fenotipo del paciente basándose únicamente en su caracterización clínica. Asimismo, las mutaciones de un solo gen pueden causar un amplio abanico de manifestaciones diferentes. Por estas razones, los análisis que abordan el estudio de un gran número de genes en paralelo son especialmente útiles para el diagnóstico genómico de la ceguera hereditaria¹⁵⁵.

En cuanto al manejo clínico de la enfermedad, la amplia superposición genética entre la retinosis pigmentaria (RP) y otros tipos de distrofias retinianas hace que su tratamiento sea extremadamente complicado y, aunque se han probado aproximaciones terapéuticas utilizando diversos agentes farmacológicos como factores neurotróficos, antioxidantes o agentes antiapoptóticos, la mayoría no están dirigidos a la causa de la RP y, por tanto, no se ha conseguido demostrar su eficacia clínica²²³. Sin embargo, con la irrupción de la terapia génica, enfocada directamente a compensar o corregir la deficiencia genética responsable de la enfermedad, en los últimos años se ha producido un cambio de paradigma que hace imprescindible disponer de un diagnóstico genético preciso.

En el artículo **IV** de la presente Tesis Doctoral se ha desarrollado un panel que, en su primera versión clínica (v3) incluía 255 genes consistentemente asociados a ceguera hereditaria y 42 genes candidatos. 130 de estos genes se habían relacionado

exclusivamente a formas no sindrómicas de ceguera, 135 se habían asociado con ceguera sindrómica y 32 estaban implicados tanto en formas sindrómicas de la enfermedad como en ceguera no sindrómica. Al igual que el resto de los paneles desarrollados en nuestro laboratorio, este también es revisado y optimizado de manera continua (en la actualidad se está empleando la v5, con 369 genes [**Anexo E**] y el rediseño de sondas ha permitido obtener valores de *callability* a DP20 superiores al 99,5 %, mejorando el rendimiento en regiones en el genoma que, por sus características, representan un desafío para la NGS. Esta situación ocurre con el gen del panel *RPGR*, en concreto con su isoforma *ORF15*, que por la baja complejidad de la secuencia y por su naturaleza altamente repetitiva plantea dificultades en la secuenciación. Con el incremento del número de sondas de captura dedicadas a esta región, hemos conseguido aumentar la calidad y cantidad de datos aprovechables de secuenciación de este gen. Al igual que en los paneles ONCOgenics™ y OTOgenics™, la sensibilidad y la especificidad de la plataforma para detectar SNVs e indels fueron muy elevadas (>99,5 % y >99,9 %, respectivamente).

Utilizando el panel desarrollado, se analizaron 100 pacientes con ceguera presuntamente hereditaria causada por alteraciones en las siguientes estructuras oculares: vítreo, coroides, retina y/nervio óptico. El rendimiento diagnóstico obtenido fue del 45 % (45/100). Teniendo en cuenta los fenotipos de ceguera hereditaria asociados a los genes considerados responsables de ceguera en esta cohorte, un 73,3 % (33/45) mostraron patrones de herencia autosómicos recesivos, un 17,8 % (8/45) autosómicos dominantes y un 8,9 % (4/45) ligados al cromosoma X. El 28,9 % de los casos diagnosticados genéticamente (13/45) fueron sindrómicos y, de ellos, en el 30,8 % (4/13) las características extraoftalmológicas se habían pasado por alto y/o no se habían relacionado con la discapacidad visual antes de las pruebas genéticas, incluidos casos con síndrome de Mainzer-Saldino, Bardet-Biedl, mucopolidosis y MLCRD [**Figura 7**].



Figura 7. Modos de herencia y carácter sindrómico/no sindrómico de las patologías genéticas consideradas responsables de la pérdida de visión hereditaria en nuestra cohorte. De los anillos externos a los internos, se muestran los genes afectados, los fenotipos asociados y los modos de herencia. El área de cada sector es proporcional al número de casos representados dentro de él. Para los genes con alteraciones causales que afectan a más de un paciente, el número de pacientes afectados se muestra entre paréntesis. ACL: amaurosis congénita de Leber; AD: autosómico dominante; AR: autosómico recesivo; MLCRD: síndrome de microcefalia, linfedema y displasia coriorretiniana; RP: retinosis pigmentaria; SM; síndrome de Saldino-Mainzer; XD: dominante ligado al X; XR: recesivo ligado al X.

Los resultados obtenidos han contribuido a definir el espectro de alteraciones genéticas responsables de la ceguera hereditaria en la población española, resaltando

su heterogeneidad genética y su expresividad variable. Se identificaron variantes genéticas consideradas responsables de la ceguera en 27 genes diferentes que explicaban el fenotipo de 45 pacientes. Los genes alterados más habitualmente fueron *ABCA4* (n = 8), *PRPH2* (n = 4), *USH2A* (n = 4), y *CNGB1* (n = 3).

Las variantes consideradas responsables de la ceguera fueron SNVs o indels en 44/45 casos diagnosticados genéticamente. En el caso restante, la plataforma identificó una delección en homocigosis del exón 1 de *BBS4*. Con el fin de evaluar en más profundidad el rendimiento de la plataforma para la detección de CNVs, se llevó a cabo un análisis MLPA (*Multiplex Ligation-dependent Probe Amplification*) de 5 genes en los que se han descrito delecciones/duplicaciones causantes de ceguera hereditaria con una frecuencia relativamente elevada (*USH2A*, *EYS*, *PRPF31*, *RHO* y *RPE65*) en 31 pacientes en los que el panel NGS no había identificado la causa genética de la ceguera. Este análisis no identificó delecciones/duplicaciones adicionales, lo cual sugiere que la potencial no detección de CNVs no es una causa relevante de pérdida de rendimiento diagnóstico del panel en nuestra cohorte.

El 15,6 % (7/45) de los pacientes para los que se obtuvo un diagnóstico genético en nuestro estudio correspondió a casos cuyos diagnósticos clínicos previos al estudio mediante el panel no coincidían con los fenotipos asociados a los genes afectados en el momento en el que se notificaron los resultados al médico solicitante, esta situación afectó a los genes: *BBS1*, *IFT140* (2x), *KIF11*, *MCOLN1*, *PEX1* y *PRPS1*. La reevaluación clínica de los pacientes posterior al estudio reveló que 4 de estos casos representaban síndromes ocultos (casos con variantes patogénicas en *BBS1*, *IFT140* (1 de los 2 casos), *KIF11* y *MCOLN1*). De los tres restantes uno de ellos era un caso de expansión fenotípica (el segundo caso con alteraciones en *IFT140*), otro había sido diagnosticado de ceguera sindrómica, pero sin filiar (caso con mutación patogénica en *PRPS1*) y el restante correspondió a un síndrome de Heimler (causado por mutación en *PEX1*) que había sido diagnosticado clínicamente como síndrome de Usher. Estos resultados refuerzan la utilidad del uso de paneles independientes del fenotipo para el diagnóstico genómico de la pérdida de visión.

Los dos casos de nuestra serie con alteración en el gen *IFT140* eran portadores de la misma variante patogénica, pero mostraron fenotipos muy diferentes: uno de ellos era un hombre cuyas manifestaciones coincidían con las del síndrome de Saldino-Mainzer y el otro era una mujer con RP no sindrómica. En una situación similar se encontraba una paciente afectada de RP no sindrómica portadora de dos variantes en heterocigosis compuesta en el gen *ARL6*, ya que tenía una hermana (sin estudio genético) con polidactilia, lo que sugiere fuertemente un diagnóstico de síndrome de Bardet-Biedl. Estos son ejemplos claros de heterogeneidad fenotípica que refuerzan la dificultad de predecir los genes afectados a partir de la evaluación clínica y, como resultado, la conveniencia de utilizar paneles genéticos amplios para el diagnóstico de la ceguera hereditaria.

El panel desarrollado permite analizar en un solo test todos los genes asociados en la literatura con el fenotipo de ceguera hereditaria tanto sindrómica y no sindrómica, como con cualquier modo de herencia. Una ventaja de este abordaje es la posibilidad de detectar variantes *de novo* o con penetrancia incompleta en familias con patrones de transmisión de la enfermedad aparentemente recesivos. Esto ocurre en uno de nuestros pacientes con retinosis pigmentaria no sindrómica, en el cual se identificó la variante patogénica *IMPDH1* (c.933C>G, p.(Asp311Glu)) y se confirmó que había surgido *de novo* (ausente en los padres; paternidad y maternidad confirmadas). En nuestra cohorte, también había cuatro pacientes con variantes heterocigotas dominantes en *PRPH2* consideradas responsables de su pérdida de visión. Dos de estos pacientes, uno heterocigoto para la variante *PRPH2* c.660_665del y el otro para la variante *PRPH2* p.Tyr141Serfs*12, no presentaban antecedentes familiares de enfermedad oftalmológica, lo que es consistente con la penetrancia incompleta característica de los fenotipos asociados a este gen o, alternativamente, podría deberse a la ocurrencia *de novo* de las variantes (en estos casos no disponíamos de muestra de progenitores para poder hacer esa comprobación). En conjunto, en nuestra serie, 6 de los 45 casos (13,3 %) con un diagnóstico genético tenían variantes causales en genes asociados con fenotipos dominantes y, sin embargo, no tenían antecedentes familiares de distrofias hereditarias de retina (DHR). Cabe destacar que un panel recesivo no habría detectado estas variantes dominantes.

Hasta ahora, al igual que ocurre en otras patologías con las mismas características, la aplicación clínica de la NGS al diagnóstico molecular de ceguera hereditaria ha seguido dos caminos principales: la secuenciación de exoma y la secuenciación mediante paneles de genes. En una comparación entre la aplicación un panel y la utilización de WES para el diagnóstico de la ceguera hereditaria, Consugar y col., basándose en las tasas de detección y en los tiempos de respuesta de ambas metodologías, concluyeron que centrar los esfuerzos de secuenciación e interpretación en un panel de genes consistentemente asociados a un fenotipo de ceguera superaba a las estrategias de secuenciación basadas en exoma⁸⁰. Un estudio más reciente, dirigido a evaluar el rendimiento de la secuenciación de exomas en el contexto clínico, concluyó que el estándar actual de profundidad de lectura promedio en exomas, de 120x, puede ser insuficiente para obtener la cobertura necesaria en todos los genes relevantes para el diagnóstico²²⁴. En la práctica, la cobertura de los exomas no alcanza a la de los paneles, lo cual provoca coberturas insuficientes en regiones clave y, en consecuencia, la no detección de variantes clínicamente relevantes. En nuestra cohorte, por ejemplo, se identificaron dos variantes en heterocigosis compuesta en el gen *CEP290* en un paciente en el que un análisis de exoma previo, realizado en otro laboratorio, solamente había detectado una de las dos alteraciones. La profundidad media de lecturas del gen *CEP290* en nuestro panel en este estudio fue cercana a 1000x, muy superior a la profundidad media habitual de los análisis basados en secuenciación del exoma.

El gen responsable del fenotipo de discapacidad visual en un mayor número de casos de nuestra cohorte fue *ABCA4* (8/46 casos, 17,4 %) seguido de *USH2A* y *PRPH2* (4/46 casos, 8,7 %, cada uno), *CNGB1* (3/46, 6,5 %), y *CHM*, *IFT140*, *MYO7A* y *RP1* (2/46, 4,3 %, cada uno). En los 18 casos restantes (40 %) se identificaron variantes causales en 18 genes diferentes (*ARL6*, *BBS1*, *BBS4*, *CEP290*, *CERKL*, *DHDDS*, *EYS*, *FAM161A*, *IMPDH1*, *KIF11*, *MCOLN1*, *OAT*, *OPA1*, *PEX1*, *PROM1*, *PRPS1*, *RHO* y *RPGR*) [Figura 7]. La alta prevalencia de las variantes en *ABCA4*, *USH2A* y *CHM* coincide con lo publicado en un análisis retrospectivo reciente realizado por Motta et al. en el que se evaluaron genéticamente 549 pacientes brasileños con DHR²²⁵. De los otros cuatro genes considerados más frecuentemente mutados (*CEP290*, *CRB1*, *RPGR* y *CHM*), en tres de ellos (*CEP290*, *RPGR* y *CHM*) también identificamos alteraciones patogénicas en nuestra

serie, pero en un número muy inferior de pacientes. Esta ausencia de recurrencia detectada en nuestro estudio es probablemente debida al tamaño de la muestra analizada (100 casos totales, con 45 positivos). Tres de los otros cuatro genes prevalentes de nuestro estudio (*CNGB1*, *PRPH2* y *MYO7A*) causan enfermedades en múltiples pacientes del estudio de Motta y col. Asimismo, 7 de los 18 genes cuyas mutaciones son responsables de la enfermedad de un único caso (*BBS1*, *CEP290*, *CERKL*, *PROM1*, *PRPS1*, *RHO* y *RPGR*) también son la causa genética de la DHR en su serie²²⁵.

La terapia génica retiniana es una alternativa muy prometedora en el tratamiento de diferentes formas de ceguera hereditaria, principalmente debido a que la barrera hemato-ocular hace que la retina sea un tejido inmunológicamente privilegiado, en el que con pequeñas cantidades del vector viral se consigue una respuesta terapéutica potente. Asimismo, la naturaleza hermética del ojo reduce el riesgo de diseminación generalizada del vector administrado localmente y, por tanto, los efectos sistémicos no deseados²²⁶⁻²²⁸. De los 100 pacientes evaluados en nuestra serie, el 15 % (15/100) podrían beneficiarse potencialmente de ensayos clínicos destinados a corregir sus defectos genéticos, entre los cuales se encuentran los afectados por mutaciones en *ABCA4* (8 casos), *USH2A* (2 casos), *CHM* (2 casos), *MYO7A* (2 casos) y *RPGR* (1 caso). Nuestros resultados sugieren que la identificación de un ensayo clínico potencial basado en terapia génica con fines curativos es una posibilidad relativamente probable para los pacientes con DHR. Además, estos datos, combinados con los 4 casos en los que se reveló un síndrome oculto, indican que la proporción de pacientes diagnosticados genéticamente en los que el análisis genómico exhaustivo identificó información clínicamente relevante fue de al menos el 42,2 % (19/45).

Disponer de un diagnóstico genético es determinante en la elección de ensayos clínicos y tratamientos basados en la terapia génica. Esta alternativa es óptima para DHRs recesivas. Sin embargo, para las DHRs dominantes, la terapia de suplementación génica no siempre es pertinente, porque la haploinsuficiencia no es la única causa de la enfermedad. Los mecanismos que causan la enfermedad en ese caso a menudo son de ganancia de función o de dominancia negativa, por lo que requieren enfoques terapéuticos alternativos, como la tecnología CRISPR/Cas9, que ya está siendo

evaluada en ensayos clínicos²²⁹. Esta técnica de edición del genoma proporciona la posibilidad de corregir y/o revertir el efecto de variantes patogénicas específicas y ha aumentado la esperanza de obtener tratamientos curativos en una proporción cada vez mayor de pacientes con ceguera hereditaria. Es el caso del fármaco EDIT-101, cuya capacidad para revertir el efecto de la variante patogénica intrónica CEP290 c.2991 + 1655A>G, causante de amaurosis congénita de Leber tipo 10, está siendo evaluada²³⁰.

Al igual que lo descrito en los artículos **I** y **II** para el cáncer y la sordera, nuestros resultados demuestran que la caracterización genómica es imprescindible para un correcto manejo clínico de la ceguera hereditaria. Además de los beneficios derivados de un correcto diagnóstico, las opciones terapéuticas ya disponibles y los múltiples ensayos clínicos que están siendo desarrollados con finalidad curativa representan opciones reales para cambiar la vida de las personas con ceguera hereditaria, una de las patologías más incapacitantes que, hasta la fecha, la medicina no había sido capaz de corregir.

CONCLUSIONES

1. Se ha desarrollado un panel NGS (ONCOgenics™) que permite analizar, en un mismo test, tanto genes cuyas alteraciones se han asociado a terapias aprobadas para el tratamiento del cáncer como aquellos implicados en cáncer hereditario.
2. La aplicación del panel ONCOgenics™ contribuye a mejorar el manejo clínico del paciente con cáncer y de sus familiares, pues puede proporcionar información útil para su tratamiento, para su seguimiento, para la prevención de segundos tumores primarios y para el asesoramiento genético a sus familiares.
3. Se ha desarrollado un panel NGS (OTOgenics™) que permite analizar, en un único test, los genes asociados a sordera neurosensorial y mixta, sindrómica y no sindrómica.
4. La aplicación del panel OTOgenics™ a 50 pacientes españoles con sospecha de hipoacusia hereditaria y sin mutaciones en *GJB2*, *GJB6*, *OTOF* y *MT-RNR1* ha identificado la causa genética de la sordera en el 42 % de los casos.
5. Se estima que el rendimiento diagnóstico del panel OTOgenics™ en pacientes de nuestra población con sospecha de hipoacusia hereditaria no cribados previamente sería del 50-60 %.
6. La aplicación del panel OTOgenics™ ha detectado síndromes ocultos en el 29 % de los casos diagnosticados genéticamente y mutaciones *de novo* en el 37,5 % de los casos con modo de herencia autosómico dominante.
7. Se ha desarrollado un panel NGS (OFTALMOgenics™) que permite analizar, en un solo test, los genes asociados a ceguera hereditaria, tanto sindrómica como no sindrómica, causada por alteraciones en el vítreo, la coroides, la retina y el nervio óptico.

8. El análisis de 100 pacientes españoles con ceguera mediante el panel OFTALMOgenics™ ha identificado la causa genética de la enfermedad en el 45 % de los casos.
9. Entre los casos diagnosticados genéticamente mediante OFTALMOgenics™, en el 33,3 % la ceguera está causada por mutaciones en genes que son objeto de ensayos clínicos de terapia génica con potencial curativo, mientras que el 8,9 % de los casos presentan síndromes ocultos.
10. Los paneles ONCOgenics™, OTOgenics™ y OFTALMOgenics™ tienen sensibilidades y especificidades analíticas compatibles con su aplicación clínica.
11. Los resultados obtenidos con los paneles OTOgenics™ y OFTALMOgenics™ y, específicamente, su capacidad para detectar tanto mutaciones *de novo* en familias con herencia pseudorecesiva, como síndromes ocultos, demuestran la utilidad de analizar, en un mismo test, todos los genes asociados con estas patologías, independientemente del carácter sindrómico o no sindrómico y de los modos de herencia de los fenotipos a los que se asocian.

BIBLIOGRAFÍA

1. **Yang, Y. et al.** Clinical whole-exome sequencing for the diagnosis of mendelian disorders. *N Engl J Med* **369**, 1502-11 (2013).
2. **Prohaska, A. et al.** Human Disease Variation in the Light of Population Genomics. *Cell* **177**, 115-131 (2019).
3. **Galceran, J. et al.** Cancer incidence in Spain, 2015. *Clin Transl Oncol* **19**, 799-825 (2017).
4. **Ramsey, T., Svider, P.F. & Folbe, A.J.** Health Burden and Socioeconomic Disparities From Hearing Loss: A Global Perspective. *Otol Neurotol* **39**, 12-16 (2018).
5. **Corton, M. et al.** Identification of the Photoreceptor Transcriptional Co-Repressor SAMD11 as Novel Cause of Autosomal Recessive Retinitis Pigmentosa. *Sci Rep* **6**, 35370 (2016).
6. **Chan, D.K. & Chang, K.W.** GJB2-associated hearing loss: systematic review of worldwide prevalence, genotype, and auditory phenotype. *Laryngoscope* **124**, E34-53 (2014).
7. **Martin-Merida, I. et al.** Toward the Mutational Landscape of Autosomal Dominant Retinitis Pigmentosa: A Comprehensive Analysis of 258 Spanish Families. *Invest Ophthalmol Vis Sci* **59**, 2345-2354 (2018).
8. **Mersch, J. et al.** Prevalence of Variant Reclassification Following Hereditary Cancer Genetic Testing. *JAMA* **320**, 1266-1274 (2018).
9. **Ferlay, J. et al.** Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *Int J Cancer* **136**, E359-86 (2015).
10. **McDermott, J.H. et al.** Diagnosing and Preventing Hearing Loss in the Genomic Age. *Trends Hear* **23**, 2331216519878983 (2019).
11. **Narayan, D.S., Wood, J.P., Chidlow, G. & Casson, R.J.** A review of the mechanisms of cone degeneration in retinitis pigmentosa. *Acta Ophthalmol* **94**, 748-754 (2016).
12. **Chong, J.X. et al.** The Genetic Basis of Mendelian Phenotypes: Discoveries, Challenges, and Opportunities. *Am J Hum Genet* **97**, 199-215 (2015).
13. **Lander, E.S. et al.** Initial sequencing and analysis of the human genome. *Nature* **409**, 860-921 (2001).
14. **Mardis, E.R.** DNA sequencing technologies: 2006-2016. *Nat Protoc* **12**, 213-218 (2017).
15. **Maxwell, O.C., I.; Ifeanyi, A.** Numbers in Life: A Statistical Genetic Approach. *Scientific Review* **5**, 8 (2019).
16. **Watson, J.D. & Crick, F.H.** The structure of DNA. *Cold Spring Harb Symp Quant Biol* **18**, 123-31 (1953).
17. **Heather, J.M. & Chain, B.** The sequence of sequencers: The history of sequencing DNA. *Genomics* **107**, 1-8 (2016).

18. **Crick, F.** Central dogma of molecular biology. *Nature* **227**, 561-3 (1970).
19. **Sanger, F. & Coulson, A.R.** A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *J Mol Biol* **94**, 441-8 (1975).
20. **Sanger, F., Nicklen, S. & Coulson, A.R.** DNA sequencing with chain-terminating inhibitors. 1977. *Biotechnology* **24**, 104-8 (1992).
21. **Maxam, A.M. & Gilbert, W.** A new method for sequencing DNA. *Proc Natl Acad Sci U S A* **74**, 560-4 (1977).
22. **Zimmermann, J., Voss, H., Schwager, C., Stegemann, J. & Ansorge, W.** Automated Sanger dideoxy sequencing reaction protocol. *FEBS Lett* **233**, 432-6 (1988).
23. **Smith, L.M. et al.** Fluorescence detection in automated DNA sequence analysis. *Nature* **321**, 674-9 (1986).
24. **Roberts, L., Davenport, R.J., Pennisi, E. & Marshall, E.** A history of the Human Genome Project. *Science* **291**, 1195 (2001).
25. **Roberts, L.** NIH and DOE draft genome pact. *Science* **241**, 1596 (1988).
26. **Watson, J.D.** The human genome project: past, present, and future. *Science* **248**, 44-9 (1990).
27. **Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J.** Basic local alignment search tool. *J Mol Biol* **215**, 403-10 (1990).
28. **Adams, M.D. et al.** The genome sequence of *Drosophila melanogaster*. *Science* **287**, 2185-95 (2000).
29. **Venter, J.C. et al.** The sequence of the human genome. *Science* **291**, 1304-51 (2001).
30. **International Human Genome Sequencing, C.** Finishing the euchromatic sequence of the human genome. *Nature* **431**, 931-45 (2004).
31. **Kwok, P.Y., Deng, Q., Zakeri, H., Taylor, S.L. & Nickerson, D.A.** Increasing the information content of STS-based genome maps: identifying polymorphisms in mapped STSs. *Genomics* **31**, 123-6 (1996).
32. **Hood, L. & Galas, D.** The digital code of DNA. *Nature* **421**, 444-8 (2003).
33. **Collins, F.S. & Barker, A.D.** Mapping the cancer genome. Pinpointing the genes involved in cancer will help chart a new course across the complex landscape of human malignancies. *Sci Am* **296**, 50-7 (2007).
34. **International Cancer Genome, C. et al.** International network of cancer genome projects. *Nature* **464**, 993-8 (2010).
35. **Hood, L. & Rowen, L.** The Human Genome Project: big science transforms biology and medicine. *Genome Med* **5**, 79 (2013).

36. **Hernan, I. et al.** Detection of genomic variations in BRCA1 and BRCA2 genes by long-range PCR and next-generation sequencing. *J Mol Diagn* **14**, 286-93 (2012).
37. **Borgstrom, E., Lundin, S. & Lundeberg, J.** Large scale library generation for high throughput sequencing. *PLoS One* **6**, e19119 (2011).
38. **Picelli, S. et al.** Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. *Genome Res* **24**, 2033-40 (2014).
39. **Heavens, D., Accinelli, G.G., Clavijo, B. & Clark, M.D.** A method to simultaneously construct up to 12 differently sized Illumina Nextera long mate pair libraries with reduced DNA input, time, and cost. *Biotechniques* **59**, 42-5 (2015).
40. **Knierim, E., Lucke, B., Schwarz, J.M., Schuelke, M. & Seelow, D.** Systematic comparison of three methods for fragmentation of long-range PCR products for next generation sequencing. *PLoS One* **6**, e28240 (2011).
41. **Turner, E.H., Ng, S.B., Nickerson, D.A. & Shendure, J.** Methods for genomic partitioning. *Annu Rev Genomics Hum Genet* **10**, 263-84 (2009).
42. **Samorodnitsky, E. et al.** Evaluation of Hybridization Capture Versus Amplicon-Based Methods for Whole-Exome Sequencing. *Hum Mutat* **36**, 903-14 (2015).
43. **Slatko, B.E., Gardner, A.F. & Ausubel, F.M.** Overview of Next-Generation Sequencing Technologies. *Curr Protoc Mol Biol* **122**, e59 (2018).
44. **Shendure, J. & Ji, H.** Next-generation DNA sequencing. *Nat Biotechnol* **26**, 1135-45 (2008).
45. **Ronaghi, M., Uhlen, M. & Nyren, P.** A sequencing method based on real-time pyrophosphate. *Science* **281**, 363, 365 (1998).
46. **Margulies, M. et al.** Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**, 376-80 (2005).
47. **Berglund, E.C., Kiialainen, A. & Syvanen, A.C.** Next-generation sequencing technologies and applications for human genetic history and forensics. *Investig Genet* **2**, 23 (2011).
48. **Lowden, M.J. & Henry, K.A.** Oxford nanopore sequencing enables rapid discovery of single-domain antibodies from phage display libraries. *Biotechniques* **65**, 351-356 (2018).
49. **Voelkerding, K.V., Dames, S.A. & Durtschi, J.D.** Next-generation sequencing: from basic research to diagnostics. *Clin Chem* **55**, 641-58 (2009).
50. **Chen, F. et al.** The history and advances of reversible terminators used in new generations of sequencing technology. *Genomics Proteomics Bioinformatics* **11**, 34-40 (2013).
51. **Moorthie, S., Hall, A. & Wright, C.F.** Informatics and clinical genome sequencing: opening the black box. *Genet Med* **15**, 165-71 (2013).

52. **Oliver, G.R., Hart, S.N. & Klee, E.W.** Bioinformatics for clinical next generation sequencing. *Clin Chem* **61**, 124-35 (2015).
53. **Celesti, F., Celesti, A., Wan, J. & Villari, M.** Why Deep Learning Is Changing the Way to Approach NGS Data Processing: A Review. *IEEE Rev Biomed Eng* **11**, 68-76 (2018).
54. **Brouwer, R.W., van den Hout, M.C., Grosveld, F.G. & van Ijcken, W.F.** NARWHAL, a primary analysis pipeline for NGS data. *Bioinformatics* **28**, 284-5 (2012).
55. **Flicek, P. & Birney, E.** Sense from sequence reads: methods for alignment and assembly. *Nat Methods* **6**, S6-S12 (2009).
56. **Kuhlenbaumer, G., Hullmann, J. & Appenzeller, S.** Novel genomic techniques open new avenues in the analysis of monogenic disorders. *Hum Mutat* **32**, 144-51 (2011).
57. **Feng, Y., Zhang, Y., Ying, C., Wang, D. & Du, C.** Nanopore-based fourth-generation DNA sequencing technology. *Genomics Proteomics Bioinformatics* **13**, 4-16 (2015).
58. **Bowden, R. et al.** Sequencing of human genomes with nanopore technology. *Nat Commun* **10**, 1869 (2019).
59. **Jauk, F.** Next-Generation Sequencing (NGS): basic concepts and applications. *Hematología* **23**, 13 (2019).
60. **Morganti, S. et al.** Complexity of genome sequencing and reporting: Next generation sequencing (NGS) technologies and implementation of precision medicine in real life. *Crit Rev Oncol Hematol* **133**, 171-182 (2019).
61. **Xue, Y. & Wilcox, W.R.** Changing paradigm of cancer therapy: precision medicine by next-generation sequencing. *Cancer Biol Med* **13**, 12-8 (2016).
62. **Wright, C.F., FitzPatrick, D.R. & Firth, H.V.** Paediatric genomics: diagnosing rare disease in children. *Nat Rev Genet* **19**, 253-268 (2018).
63. **Ormond, K.E. et al.** Challenges in the clinical application of whole-genome sequencing. *Lancet* **375**, 1749-51 (2010).
64. **Pelak, K. et al.** The characterization of twenty sequenced human genomes. *PLoS Genet* **6**, e1001111 (2010).
65. **Zhernakova, D.V. et al.** Genome-wide sequence analyses of ethnic populations across Russia. *Genomics* **112**, 442-458 (2020).
66. **Chrystoja, C.C. & Diamandis, E.P.** Whole genome sequencing as a diagnostic test: challenges and opportunities. *Clin Chem* **60**, 724-33 (2014).
67. **Retterer, K. et al.** Clinical application of whole-exome sequencing across clinical indications. *Genet Med* **18**, 696-704 (2016).

68. **Fichna, J.P. et al.** Whole-exome sequencing identifies novel pathogenic mutations and putative phenotype-influencing variants in Polish limb-girdle muscular dystrophy patients. *Hum Genomics* **12**, 34 (2018).
69. **Peveling-Oberhag, J. et al.** Whole exome sequencing of microdissected splenic marginal zone lymphoma: a study to discover novel tumor-specific mutations. *BMC Cancer* **15**, 773 (2015).
70. **Biesecker, L.G., Shianna, K.V. & Mullikin, J.C.** Exome sequencing: the expert view. *Genome Biol* **12**, 128 (2011).
71. **Brittain, H.K., Scott, R. & Thomas, E.** The rise of the genome and personalised medicine. *Clin Med (Lond)* **17**, 545-551 (2017).
72. **Lee, H. et al.** Clinical exome sequencing for genetic identification of rare Mendelian disorders. *JAMA* **312**, 1880-7 (2014).
73. **Sun, Y. et al.** Next-generation diagnostics: gene panel, exome, or whole genome? *Hum Mutat* **36**, 648-55 (2015).
74. **Xue, Y., Ankala, A., Wilcox, W.R. & Hegde, M.R.** Solving the molecular diagnostic testing conundrum for Mendelian disorders in the era of next-generation sequencing: single-gene, gene panel, or exome/genome sequencing. *Genet Med* **17**, 444-51 (2015).
75. **Matthijs, G. et al.** Guidelines for diagnostic next-generation sequencing. *Eur J Hum Genet* **24**, 2-5 (2016).
76. **Lim, E.C. et al.** Next-generation sequencing using a pre-designed gene panel for the molecular diagnosis of congenital disorders in pediatric patients. *Hum Genomics* **9**, 33 (2015).
77. **Martin, A.R. et al.** PanelApp crowdsources expert knowledge to establish consensus diagnostic gene panels. *Nat Genet* **51**, 1560-1565 (2019).
78. **Klein, C.J. & Foroud, T.M.** Neurology Individualized Medicine: When to Use Next-Generation Sequencing Panels. *Mayo Clin Proc* **92**, 292-305 (2017).
79. **Bevilacqua, J. et al.** Clinical utility of a 377 gene custom next-generation sequencing epilepsy panel. *J Genet* **96**, 681-685 (2017).
80. **Consugar, M.B. et al.** Panel-based genetic diagnostic testing for inherited eye diseases is highly accurate and reproducible, and more sensitive for variant detection, than exome sequencing. *Genet Med* **17**, 253-261 (2015).
81. **Stratton, M.R., Campbell, P.J. & Futreal, P.A.** The cancer genome. *Nature* **458**, 719-24 (2009).
82. **Boiocchi, M.** The genetic nature of cancer. *Eur J Cancer Prev* **5**, 370-2 (1996).
83. **Lindahl, T. & Barnes, D.E.** Repair of endogenous DNA damage. *Cold Spring Harb Symp Quant Biol* **65**, 127-33 (2000).
84. **Hanahan, D. & Weinberg, R.A.** Hallmarks of cancer: the next generation. *Cell* **144**, 646-74 (2011).

85. **Consortium, I.T.P.-C.A.o.W.G.** Pan-cancer analysis of whole genomes. *Nature* **578**, 82-93 (2020).
86. **Roy, P.S. & Saikia, B.J.** Cancer and cure: A critical analysis. *Indian J Cancer* **53**, 441-442 (2016).
87. **Reddy, E.P., Reynolds, R.K., Santos, E. & Barbacid, M.** A point mutation is responsible for the acquisition of transforming properties by the T24 human bladder carcinoma oncogene. *Nature* **300**, 149-52 (1982).
88. **Blaschke, S.** The role of nature in cancer patients' lives: a systematic review and qualitative meta-synthesis. *BMC Cancer* **17**, 370 (2017).
89. **Tomczak, K., Czerwinska, P. & Wiznerowicz, M.** The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. *Contemp Oncol (Pozn)* **19**, A68-77 (2015).
90. **Alexandrov, L.B. et al.** Signatures of mutational processes in human cancer. *Nature* **500**, 415-21 (2013).
91. **Gerstung, M. et al.** The evolutionary history of 2,658 cancers. *Nature* **578**, 122-128 (2020).
92. **Stegmeier, F., Warmuth, M., Sellers, W.R. & Dorsch, M.** Targeted cancer therapies in the twenty-first century: lessons from imatinib. *Clin Pharmacol Ther* **87**, 543-52 (2010).
93. **Li, M.M. et al.** Standards and Guidelines for the Interpretation and Reporting of Sequence Variants in Cancer: A Joint Consensus Recommendation of the Association for Molecular Pathology, American Society of Clinical Oncology, and College of American Pathologists. *J Mol Diagn* **19**, 4-23 (2017).
94. **Dienstmann, R., Jang, I.S., Bot, B., Friend, S. & Guinney, J.** Database of genomic biomarkers for cancer drugs and clinical targetability in solid tumors. *Cancer Discov* **5**, 118-23 (2015).
95. **Flaherty, K.T. et al.** The molecular analysisi for therapy choice (NCI-MATCH) trial: lessons for genomic trial design. *J Natl Cancer Inst* (2020).
96. **Gyorffy, B. et al.** Multigene prognostic tests in breast cancer: past, present, future. *Breast Cancer Res* **17**, 11 (2015).
97. **Snouwaert, J.N. et al.** BRCA1 deficient embryonic stem cells display a decreased homologous recombination frequency and an increased frequency of non-homologous recombination that is corrected by expression of a brca1 transgene. *Oncogene* **18**, 7900-7 (1999).
98. **Moynahan, M.E., Pierce, A.J. & Jasin, M.** BRCA2 is required for homology-directed repair of chromosomal breaks. *Mol Cell* **7**, 263-72 (2001).
99. **Pfeffer, C.M., Ho, B.N. & Singh, A.T.K.** The Evolution, Functions and Applications of the Breast Cancer Genes BRCA1 and BRCA2. *Cancer Genomics Proteomics* **14**, 293-298 (2017).

100. **Roy, R., Chun, J. & Powell, S.N.** BRCA1 and BRCA2: different roles in a common pathway of genome protection. *Nat Rev Cancer* **12**, 68-78 (2011).
101. **Ford, D. et al.** Genetic heterogeneity and penetrance analysis of the BRCA1 and BRCA2 genes in breast cancer families. The Breast Cancer Linkage Consortium. *Am J Hum Genet* **62**, 676-89 (1998).
102. **Hampel, H. et al.** Screening for the Lynch syndrome (hereditary nonpolyposis colorectal cancer). *N Engl J Med* **352**, 1851-60 (2005).
103. **Barnetson, R.A. et al.** Identification and survival of carriers of mutations in DNA mismatch-repair genes in colon cancer. *N Engl J Med* **354**, 2751-63 (2006).
104. **Engel, C. et al.** Risks of less common cancers in proven mutation carriers with lynch syndrome. *J Clin Oncol* **30**, 4409-15 (2012).
105. **Kuiper, R.P. et al.** Recurrence and variability of germline EPCAM deletions in Lynch syndrome. *Hum Mutat* **32**, 407-14 (2011).
106. **Tutlewska, K., Lubinski, J. & Kurzawski, G.** Germline deletions in the EPCAM gene as a cause of Lynch syndrome - literature review. *Hered Cancer Clin Pract* **11**, 9 (2013).
107. **Hoang, L.N. & Gilks, B.C.** Hereditary Breast and Ovarian Cancer Syndrome: Moving Beyond BRCA1 and BRCA2. *Adv Anat Pathol* **25**, 85-95 (2018).
108. **Dominguez-Valentin, M. et al.** Results of multigene panel testing in familial cancer cases without genetic cause demonstrated by single gene testing. *Sci Rep* **9**, 18555 (2019).
109. **Selkirk, C.G. et al.** Cancer genetic testing panels for inherited cancer susceptibility: the clinical experience of a large adult genetics practice. *Fam Cancer* **13**, 527-36 (2014).
110. **LaDuca, H. et al.** Utilization of multigene panels in hereditary cancer predisposition testing: analysis of more than 2,000 patients. *Genet Med* **16**, 830-7 (2014).
111. **Pritchard, C.C. et al.** ColoSeq provides comprehensive lynch and polyposis syndrome mutational analysis using massively parallel sequencing. *J Mol Diagn* **14**, 357-66 (2012).
112. **Walsh, T. et al.** Mutations in 12 genes for inherited ovarian, fallopian tube, and peritoneal carcinoma identified by massively parallel sequencing. *Proc Natl Acad Sci U S A* **108**, 18032-7 (2011).
113. **Lin, P.C. et al.** Germline susceptibility variants impact clinical outcome and therapeutic strategies for stage III colorectal cancer. *Sci Rep* **9**, 3931 (2019).
114. **Shieh, Y. et al.** Population-based screening for cancer: hope and hype. *Nat Rev Clin Oncol* **13**, 550-65 (2016).
115. **Gupta, S. et al.** NCCN Guidelines Insights: Genetic/Familial High-Risk Assessment: Colorectal, Version 2.2019. *J Natl Compr Canc Netw* **17**, 1032-1041 (2019).

116. **Kochhar, A., Hildebrand, M.S. & Smith, R.J.** Clinical aspects of hereditary hearing loss. *Genet Med* **9**, 393-408 (2007).
117. **Yuan, Y. et al.** Comprehensive genetic testing of Chinese SNHL patients and variants interpretation using ACMG guidelines and ethnically matched normal controls. *Eur J Hum Genet* **28**, 231-243 (2020).
118. **Alford, R.L. et al.** American College of Medical Genetics and Genomics guideline for the clinical evaluation and etiologic diagnosis of hearing loss. *Genet Med* **16**, 347-55 (2014).
119. **Withrow, K.A., Burton, S., Arnos, K.S., Kalfoglou, A. & Pandya, A.** Consumer motivations for pursuing genetic testing and their preferences for the provision of genetic services for hearing loss. *J Genet Couns* **17**, 252-60 (2008).
120. **Marazita, M.L. et al.** Genetic epidemiological studies of early-onset deafness in the U.S. school-age population. *Am J Med Genet* **46**, 486-91 (1993).
121. **Tekin, M., Arnos, K.S. & Pandya, A.** Advances in hereditary deafness. *Lancet* **358**, 1082-90 (2001).
122. **Toumpas, C.J., Clark, J., Harris, A., Beswick, R. & Nourse, C.B.** Congenital cytomegalovirus infection is a significant cause of moderate to profound sensorineural hearing loss in Queensland children. *J Paediatr Child Health* **51**, 541-544 (2015).
123. **Burke, W.F. et al.** Prevalence and audiological profiles of GJB2 mutations in a large collective of hearing impaired patients. *Hear Res* **333**, 77-86 (2016).
124. **Morton, N.E., Shields, D.C. & Collins, A.** Genetic epidemiology of complex phenotypes. *Ann Hum Genet* **55**, 301-14 (1991).
125. **Rennels, M. & Pickering, L.K.** Sensorineural hearing loss in children. *Lancet* **365**, 2085-6 (2005).
126. **Tekin, M. & Arici, Z.S.** Genetic epidemiological studies of congenital/prelingual deafness in Turkey: population structure and mating type are major determinants of mutation identification. *Am J Med Genet A* **143A**, 1583-91 (2007).
127. **Gettelfinger, J.D. & Dahl, J.P.** Syndromic Hearing Loss: A Brief Review of Common Presentations and Genetics. *J Pediatr Genet* **7**, 1-8 (2018).
128. **del Castillo, F.J. & del Castillo, I.** The DFNB1 subtype of autosomal recessive non-syndromic hearing impairment. *Front Biosci (Landmark Ed)* **16**, 3252-74 (2011).
129. **Rodriguez-Paris, J., Tamayo, M.L., Gelvez, N. & Schrijver, I.** Allele-specific impairment of GJB2 expression by GJB6 deletion del(GJB6-D13S1854). *PLoS One* **6**, e21665 (2011).
130. **Rodriguez-Paris, J. & Schrijver, I.** The digenic hypothesis unraveled: the GJB6 del(GJB6-D13S1830) mutation causes allele-specific loss of GJB2 expression in cis. *Biochem Biophys Res Commun* **389**, 354-9 (2009).

131. **Moisan, S., Le Nabec, A., Quillevere, A., Le Marechal, C. & Ferec, C.** Characterization of GJB2 cis-regulatory elements in the DFNB1 locus. *Hum Genet* **138**, 1275-1286 (2019).
132. **Sabatini, L.M. et al.** Genomic Sequencing Procedure Microcosting Analysis and Health Economic Cost-Impact Analysis: A Report of the Association for Molecular Pathology. *J Mol Diagn* **18**, 319-328 (2016).
133. **Shearer, A.E. & Smith, R.J.** Massively Parallel Sequencing for Genetic Diagnosis of Hearing Loss: The New Standard of Care. *Otolaryngol Head Neck Surg* **153**, 175-82 (2015).
134. **Nunez-Batalla, F. et al.** Aetiological diagnosis of child deafness: CODEPEH recommendations. *Acta Otorrinolaringol Esp* **68**, 43-55 (2017).
135. **Cabanillas Farpon, R. & Cadinanos Banales, J.** [Hereditary hearing loss: genetic counselling]. *Acta Otorrinolaringol Esp* **63**, 218-29 (2012).
136. **American Academy of Pediatrics, J.C.o.I.H.** Year 2007 position statement: Principles and guidelines for early hearing detection and intervention programs. *Pediatrics* **120**, 898-921 (2007).
137. **Lin, X. et al.** Applications of targeted gene capture and next-generation sequencing technologies in studies of human deafness and other genetic disabilities. *Hear Res* **288**, 67-76 (2012).
138. **Sloan-Heggen, C.M. et al.** Comprehensive genetic testing in the clinical evaluation of 1119 patients with hearing loss. *Hum Genet* **135**, 441-450 (2016).
139. **Daiger, S.P., Sullivan, L.S. & Bowne, S.J.** Genes and mutations causing retinitis pigmentosa. *Clin Genet* **84**, 132-41 (2013).
140. **Solebo AL, et al.** Epidemiology of blindness in children. *Arch Dis Child* **102**, 995 (2017).
141. **Broadgate, S., Yu, J., Downes, S.M. & Halford, S.** Unravelling the genetics of inherited retinal dystrophies: Past, present and future. *Prog Retin Eye Res* **59**, 53-96 (2017).
142. **Liew, G., Michaelides, M. & Bunce, C.** A comparison of the causes of blindness certifications in England and Wales in working age adults (16-64 years), 1999-2000 with 2009-2010. *BMJ Open* **4**, e004015 (2014).
143. **Verbakel, S.K. et al.** Non-syndromic retinitis pigmentosa. *Prog Retin Eye Res* **66**, 157-186 (2018).
144. **Ferrari, S. et al.** Retinitis pigmentosa: genes and disease mechanisms. *Curr Genomics* **12**, 238-49 (2011).
145. **Daiger, S.P., Bowne, S.J. & Sullivan, L.S.** Perspective on genes and mutations causing retinitis pigmentosa. *Arch Ophthalmol* **125**, 151-8 (2007).
146. **Haim, M.** Epidemiology of retinitis pigmentosa in Denmark. *Acta Ophthalmol Scand Suppl*, 1-34 (2002).

147. **Hartong, D.T., Berson, E.L. & Dryja, T.P.** Retinitis pigmentosa. *Lancet* **368**, 1795-809 (2006).
148. **Mansergh, F.C. et al.** Retinitis pigmentosa and progressive sensorineural hearing loss caused by a C12258A mutation in the mitochondrial MTT2 gene. *Am J Hum Genet* **64**, 971-85 (1999).
149. **Kajiwara, K., Berson, E.L. & Dryja, T.P.** Digenic retinitis pigmentosa due to mutations at the unlinked peripherin/RDS and ROM1 loci. *Science* **264**, 1604-8 (1994).
150. **Costa, K.A., Salles, M.V., Whitebirch, C., Chiang, J. & Sallum, J.M.F.** Gene panel sequencing in Brazilian patients with retinitis pigmentosa. *Int J Retina Vitreous* **3**, 33 (2017).
151. **Karali, M. & Banfi, S.** Inherited Retinal Dystrophies: the role of gene expression regulators. *Int J Biochem Cell Biol* **61**, 115-9 (2015).
152. **Berger, W., Kloeckener-Gruissem, B. & Neidhardt, J.** The molecular basis of human retinal and vitreoretinal diseases. *Prog Retin Eye Res* **29**, 335-75 (2010).
153. **Chacon-Camacho, O.F., Garcia-Montano, L.A. & Zenteno, J.C.** The clinical implications of molecular monitoring and analyses of inherited retinal diseases. *Expert Rev Mol Diagn* **17**, 1009-1021 (2017).
154. **Chiang, J.P. et al.** Progress and prospects of next-generation sequencing testing for inherited retinal dystrophy. *Expert Rev Mol Diagn* **15**, 1269-75 (2015).
155. **Hafler, B.P.** Clinical Progress in Inherited Retinal Degenerations: Gene Therapy Clinical Trials and Advances in Genetic Sequencing. *Retina* **37**, 417-423 (2017).
156. **Verma, I.C., Paliwal, P. & Singh, K.** Genetic Testing in Pediatric Ophthalmology. *Indian J Pediatr* **85**, 228-236 (2018).
157. **Jespersgaard, C. et al.** Molecular genetic analysis using targeted NGS analysis of 677 individuals with retinal dystrophy. *Sci Rep* **9**, 1219 (2019).
158. **Campbell, J.P., McFarland, T.J. & Stout, J.T.** Ocular Gene Therapy. *Dev Ophthalmol* **55**, 317-21 (2016).
159. **Li, Q. & Wang, K.** InterVar: Clinical Interpretation of Genetic Variants by the 2015 ACMG-AMP Guidelines. *Am J Hum Genet* **100**, 267-280 (2017).
160. **Good, B.M., Ainscough, B.J., McMichael, J.F., Su, A.I. & Griffith, O.L.** Organizing knowledge to enable personalization of medicine in cancer. *Genome Biol* **15**, 438 (2014).
161. **Lever, J. et al.** Text-mining clinically relevant cancer biomarkers for curation into the CIViC database. *Genome Med* **11**, 78 (2019).
162. **Saadeh, C., Bright, D. & Rustem, D.** Precision Medicine in Oncology Pharmacy Practice. *Acta Med Acad* **48**, 90-104 (2019).

163. **Griffith, M. et al.** CIViC is a community knowledgebase for expert crowdsourcing the clinical interpretation of variants in cancer. *Nat Genet* **49**, 170-174 (2017).
164. **Barnell, E.K. et al.** Open-Sourced CIViC Annotation Pipeline to Identify and Annotate Clinically Relevant Variants Using Single-Molecule Molecular Inversion Probes. *JCO Clin Cancer Inform* **3**, 1-12 (2019).
165. **Chakravarty, D. et al.** OncoKB: A Precision Oncology Knowledge Base. *JCO Precis Oncol* **2017**(2017).
166. **Landrum, M.J. et al.** ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res* **46**, D1062-D1067 (2018).
167. **Amberger, J.S., Bocchini, C.A., Schiettecatte, F., Scott, A.F. & Hamosh, A.** OMIM.org: Online Mendelian Inheritance in Man (OMIM(R)), an online catalog of human genes and genetic disorders. *Nucleic Acids Res* **43**, D789-98 (2015).
168. **McKusick, V.A.** Mendelian Inheritance in Man and its online version, OMIM. *Am J Hum Genet* **80**, 588-604 (2007).
169. **Amberger, J.S. & Hamosh, A.** Searching Online Mendelian Inheritance in Man (OMIM): A Knowledgebase of Human Genes and Genetic Phenotypes. *Curr Protoc Bioinformatics* **58**, 1 2 1-1 2 12 (2017).
170. **Amberger, J., Bocchini, C.A., Scott, A.F. & Hamosh, A.** McKusick's Online Mendelian Inheritance in Man (OMIM). *Nucleic Acids Res* **37**, D793-6 (2009).
171. **Hamosh, A., Scott, A.F., Amberger, J.S., Bocchini, C.A. & McKusick, V.A.** Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Res* **33**, D514-7 (2005).
172. **Stenson, P.D. et al.** The Human Gene Mutation Database (HGMD) and its exploitation in the fields of personalized genomics and molecular evolution. *Curr Protoc Bioinformatics* **Chapter 1**, Unit1 13 (2012).
173. **Stenson, P.D. et al.** The Human Gene Mutation Database: building a comprehensive mutation repository for clinical and molecular genetics, diagnostic testing and personalized genomic medicine. *Hum Genet* **133**, 1-9 (2014).
174. **Cooper, D.N. et al.** Genes, mutations, and human inherited disease at the dawn of the age of personalized genomics. *Hum Mutat* **31**, 631-55 (2010).
175. **Stenson, P.D. et al.** The Human Gene Mutation Database: towards a comprehensive repository of inherited mutation data for medical research, genetic diagnosis and next-generation sequencing studies. *Hum Genet* **136**, 665-677 (2017).
176. **Landrum, M.J. et al.** ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res* **42**, D980-5 (2014).
177. **Landrum, M.J. & Kattman, B.L.** ClinVar at five years: Delivering on the promise. *Hum Mutat* **39**, 1623-1630 (2018).

178. **Rehm, H.L. et al.** ClinGen--the Clinical Genome Resource. *N Engl J Med* **372**, 2235-42 (2015).
179. **Strande, N.T., Brnich, S.E., Roman, T.S. & Berg, J.S.** Navigating the nuances of clinical sequence variant interpretation in Mendelian disease. *Genet Med* **20**, 918-926 (2018).
180. **Kazazian HH, B.C., Seltzer WK.** ACMG recommendations for standards for interpretation of sequence variations. *Genet Med* **2**, 2 (2000).
181. **Richards, S. et al.** Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med* **17**, 405-24 (2015).
182. **Amendola, L.M. et al.** Performance of ACMG-AMP Variant-Interpretation Guidelines among Nine Laboratories in the Clinical Sequencing Exploratory Research Consortium. *Am J Hum Genet* **98**, 1067-1076 (2016).
183. **Abou Tayoun, A.N. et al.** Recommendations for interpreting the loss of function PVS1 ACMG/AMP variant criterion. *Hum Mutat* **39**, 1517-1524 (2018).
184. **Brnich, S.E. et al.** Recommendations for application of the functional evidence PS3/BS3 criterion using the ACMG/AMP sequence variant interpretation framework. *Genome Med* **12**, 3 (2019).
185. **Biesecker, L.G., Harrison, S.M. & ClinGen Sequence Variant Interpretation Working, G.** The ACMG/AMP reputable source criteria for the interpretation of sequence variants. *Genet Med* **20**, 1687-1688 (2018).
186. **Gelb, B.D. et al.** ClinGen's RASopathy Expert Panel consensus methods for variant interpretation. *Genet Med* **20**, 1334-1345 (2018).
187. **Kelly, M.A. et al.** Adaptation and validation of the ACMG/AMP variant classification framework for MYH7-associated inherited cardiomyopathies: recommendations by ClinGen's Inherited Cardiomyopathy Expert Panel. *Genet Med* **20**, 351-359 (2018).
188. **Mandelker, D. et al.** Navigating highly homologous genes in a molecular diagnostic setting: a resource for clinical next-generation sequencing. *Genet Med* **18**, 1282-1289 (2016).
189. **De Vos, M., Hayward, B.E., Picton, S., Sheridan, E. & Bonthron, D.T.** Novel PMS2 pseudogenes can conceal recessive mutations causing a distinctive childhood cancer syndrome. *Am J Hum Genet* **74**, 954-64 (2004).
190. **Clendenning, M. et al.** Long-range PCR facilitates the identification of PMS2-specific mutations. *Hum Mutat* **27**, 490-5 (2006).
191. **Borras, E. et al.** Refining the role of PMS2 in Lynch syndrome: germline mutational analysis improved by comprehensive assessment of variants. *J Med Genet* **50**, 552-63 (2013).

192. **Vaughn, C.P. et al.** Clinical analysis of PMS2: mutation detection and avoidance of pseudogenes. *Hum Mutat* **31**, 588-93 (2010).
193. **Susswein, L.R. et al.** Pathogenic and likely pathogenic variant prevalence among the first 10,000 patients referred for next-generation cancer panel testing. *Genet Med* **18**, 823-32 (2016).
194. **Le, D.T. et al.** PD-1 Blockade in Tumors with Mismatch-Repair Deficiency. *N Engl J Med* **372**, 2509-20 (2015).
195. **Rizvi, N.A. et al.** Cancer immunology. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. *Science* **348**, 124-8 (2015).
196. **Van Allen, E.M. et al.** Genomic correlates of response to CTLA-4 blockade in metastatic melanoma. *Science* **350**, 207-211 (2015).
197. **Bouffet, E. et al.** Immune Checkpoint Inhibition for Hypermutant Glioblastoma Multiforme Resulting From Germline Biallelic Mismatch Repair Deficiency. *J Clin Oncol* **34**, 2206-11 (2016).
198. **Fumet, J.D., Truntzer, C., Yarchoan, M. & Ghiringhelli, F.** Tumour mutational burden as a biomarker for immunotherapy: Current data and emerging concepts. *Eur J Cancer* **131**, 40-50 (2020).
199. **Mitelman, F., Johansson, B. & Mertens, F.** The impact of translocations and gene fusions on cancer causation. *Nat Rev Cancer* **7**, 233-45 (2007).
200. **Zito Marino, F. et al.** NTRK Fusions, from the Diagnostic Algorithm to Innovative Treatment in the Era of Precision Medicine. *Int J Mol Sci* **21**(2020).
201. **Heyer, E.E. et al.** Diagnosis of fusion genes using targeted RNA sequencing. *Nat Commun* **10**, 1388 (2019).
202. **Matthijs, G. et al.** Guidelines for diagnostic next-generation sequencing. *Eur J Hum Genet* **24**, 1515 (2016).
203. **Askew, C. et al.** Tmc gene therapy restores auditory function in deaf mice. *Sci Transl Med* **7**, 295ra108 (2015).
204. **Mandelker, D. et al.** Comprehensive diagnostic testing for stereocilin: an approach for analyzing medically important genes with high homology. *J Mol Diagn* **16**, 639-47 (2014).
205. **Bonnet, C. et al.** An innovative strategy for the molecular diagnosis of Usher syndrome identifies causal biallelic mutations in 93% of European patients. *Eur J Hum Genet* **24**, 1730-1738 (2016).
206. **Kemerley, A., Sloan, C., Pfeifer, W., Smith, R. & Drack, A.** A novel mutation in ACTG1 causing Baraitser-Winter syndrome with extremely variable expressivity in three generations. *Ophthalmic Genet* **38**, 152-156 (2017).
207. **Yuan, Y. et al.** Phenotypic Heterogeneity in a DFNA20/26 family segregating a novel ACTG1 mutation. *BMC Genet* **17**, 33 (2016).

208. **Morin, M. et al.** In vivo and in vitro effects of two novel gamma-actin (ACTG1) mutations that cause DFNA20/26 hearing impairment. *Hum Mol Genet* **18**, 3075-89 (2009).
209. **Sommen, M., Wuyts, W. & Van Camp, G.** Molecular diagnostics for hereditary hearing loss in children. *Expert Rev Mol Diagn* **17**, 751-760 (2017).
210. **Migliosi, V. et al.** Q829X, a novel mutation in the gene encoding otoferlin (OTOF), is frequently found in Spanish patients with prelingual non-syndromic hearing loss. *J Med Genet* **39**, 502-6 (2002).
211. **Posey, J.E. et al.** Resolution of Disease Phenotypes Resulting from Multilocus Genomic Variation. *N Engl J Med* **376**, 21-31 (2017).
212. **Gallo-Teran, J. et al.** [Prevalence of the 35delG mutation in the GJB2 gene, del (GJB6-D13S1830) in the GJB6 gene, Q829X in the OTOF gene and A1555G in the mitochondrial 12S rRNA gene in subjects with non-syndromic sensorineural hearing impairment of congenital/childhood onset]. *Acta Otorrinolaringol Esp* **56**, 463-8 (2005).
213. **Costales, M. et al.** Clinical utility of next-generation sequencing in the aetiological diagnosis of sensorineural hearing loss in a Childhood Hearing Loss Unit. *Acta Otorrinolaringol Esp* **71**, 166-174 (2020).
214. **Rahman, S. et al.** Hearing in 44-45 year olds with m.1555A>G, a genetic mutation predisposing to aminoglycoside-induced deafness: a population based cohort study. *BMJ Open* **2**, e000411 (2012).
215. **Likar, T. et al.** Diagnostic outcomes of exome sequencing in patients with syndromic or non-syndromic hearing loss. *PLoS One* **13**, e0188578 (2018).
216. **Neveling, K. et al.** A post-hoc comparison of the utility of sanger sequencing and exome sequencing for the diagnosis of heterogeneous diseases. *Hum Mutat* **34**, 1721-6 (2013).
217. **Yang, T., Wei, X., Chai, Y., Li, L. & Wu, H.** Genetic etiology study of the non-syndromic deafness in Chinese Hans by targeted next-generation sequencing. *Orphanet J Rare Dis* **8**, 85 (2013).
218. **Amendola, L.M. et al.** Performance of ACMG-AMP Variant-Interpretation Guidelines among Nine Laboratories in the Clinical Sequencing Exploratory Research Consortium. *Am J Hum Genet* **99**, 247 (2016).
219. **Boycott, K.M., Vanstone, M.R., Bulman, D.E. & MacKenzie, A.E.** Rare-disease genetics in the era of next-generation sequencing: discovery to translation. *Nat Rev Genet* **14**, 681-91 (2013).
220. **Oza, A.M. et al.** Expert specification of the ACMG/AMP variant interpretation guidelines for genetic hearing loss. *Hum Mutat* **39**, 1593-1613 (2018).
221. **Shearer, A.E. et al.** Advancing genetic testing for deafness with genomic technology. *J Med Genet* **50**, 627-34 (2013).

222. **Kalia, S.S. et al.** Recommendations for reporting of secondary findings in clinical exome and genome sequencing, 2016 update (ACMG SF v2.0): a policy statement of the American College of Medical Genetics and Genomics. *Genet Med* **19**, 249-255 (2017).
223. **Dias, M.F. et al.** Molecular genetics and emerging therapies for retinitis pigmentosa: Basic research and clinical perspectives. *Prog Retin Eye Res* **63**, 107-131 (2018).
224. **Kong, S.W., Lee, I.H., Liu, X., Hirschhorn, J.N. & Mandl, K.D.** Measuring coverage and accuracy of whole-exome sequencing in clinical context. *Genet Med* **20**, 1617-1626 (2018).
225. **Motta, F.L., Martin, R.P., Filippelli-Silva, R., Salles, M.V. & Sallum, J.M.F.** Relative frequency of inherited retinal dystrophies in Brazil. *Sci Rep* **8**, 15939 (2018).
226. **Gupta, P.R. & Huckfeldt, R.M.** Gene therapy for inherited retinal degenerations: initial successes and future challenges. *J Neural Eng* **14**, 051002 (2017).
227. **Campa, C., Gallenga, C.E., Bolletta, E. & Perri, P.** The Role of Gene Therapy in the Treatment of Retinal Diseases: A Review. *Curr Gene Ther* **17**, 194-213 (2017).
228. **Oner, A.** Recent Advancements in Gene Therapy for Hereditary Retinal Dystrophies. *Turk J Ophthalmol* **47**, 338-343 (2017).
229. **Benati, D., Patrizi, C. & Recchia, A.** Gene editing prospects for treating inherited retinal diseases. *J Med Genet* **57**, 437-444 (2020).
230. **Burnight, E.R. et al.** CRISPR-Cas9 genome engineering: Treating inherited retinal degeneration. *Prog Retin Eye Res* **65**, 28-49 (2018).

ANEXOS

Anexo A: Genes analizados en ONCOgenics™ TumorFocus

GENES (16 genes)							
<i>ALK</i>	<i>BRAF</i>	<i>BRCA1</i>	<i>BRCA2</i>	<i>CDK4</i>	<i>EGFR</i>	<i>ERBB2</i>	<i>KIT</i>
<i>KRAS</i>	<i>MET</i>	<i>NRAS</i>	<i>PDGFRA</i>	<i>PIK3CA</i>	<i>TP53*</i>	<i>TSC1</i>	<i>TSC2</i>
REORDENAMIENTOS (4 genes)							
<i>ALK</i>	<i>PDGFB</i>	<i>RET</i>	<i>ROS1</i>				

*El gen *TP53* no se ha asociado a terapias dirigidas aprobadas para el tratamiento del cáncer, pero su análisis puede ofrecer información genómica complementaria.

Anexo B: Genes analizados en ONCOgenics™ Germline^{Focus}

ONCOgenics™ Germline^{Focus} Core (45 genes)								
APC	ATM	BAP1	BARD1	BMPR1A	BRCA1	BRCA2	BRIP1	CDH1
CDK4	CDKN2A	CHEK2	EPCAM (deleciones)	FH	FLCN	GREM1 (duplicaciones)	MAX	MET
MLH1	MSH2	MSH6	MUTYH	NBN	NF1	PALB2	PMS2	POLD1
POLE	PRSS1	PTEN	RAD51C	RAD51D	RET	SDHA	SDHAF2	SDHB
SDHC	SDHD	SMAD4	STK11	TMEM127	TP53	TSC1	TSC2	VHL
ONCOgenics™ Germline^{Focus} Mama-Ovario-Próstata (21 genes)								
ATM	BARD1	BRCA1	BRCA2	BRIP1	CDH1	CDKN2A	CHEK2	EPCAM (deleciones)
MLH1	MSH2	MSH6	NBN	NF1	PALB2	PMS2	PTEN	RAD51C
RAD51D	STK11	TP53						
ONCOgenics™ Germline^{Focus} Páncreas-Melanoma-HNSCC* (18 genes)								
APC	ATM	BAP1	BMPR1A	BRCA1	BRCA2	CDK4	CDKN2A	EPCAM (deleciones)
MLH1	MSH2	MSH6	PALB2	PMS2	PRSS1	SMAD4	STK11	TP53
ONCOgenics™ Germline^{Focus} Colon-Gástrico (17 genes)								
APC	BMPR1A	CDH1	CHEK2	EPCAM (deleciones)	GREM1 (duplicaciones)	MLH1	MSH2	MSH6
MUTYH	PMS2	POLD1	POLE	PTEN	SMAD4	STK11	TP53	
ONCOgenics™ Germline^{Focus} Lynch (5 genes)								
EPCAM (deleciones)	MLH1	MSH2	MSH6	PMS2				
ONCOgenics™ Germline^{Focus} Paraganglioma-Feocromocitoma-Renal (15 genes)								
FH	FLCN	MAX	MET	NF1	RET	SDHA	SDHAF2	SDHB
SDHC	SDHD	TMEM127	TSC1	TSC2	VHL			

*HNSCC: *Head and neck squamous cell carcinoma* (Carcinoma de células escamosas de cabeza y cuello)

Anexo C: Genes incluidos en la v5 del panel de hipoacusia hereditaria OTOgenics™

CONSISTENTEMENTE ASOCIADOS - SINDRÓMICOS (86 genes)

ABHD12	ACTB	ALMS1	AMMECR1	ANKH	AP1B1	AP1S1	ATP1A3	ATP6V0A4	ATP6V1B1	BCAP31	BCS1L	BRAF	CACNA1D
CEP78	CHD7	CISD2	CLCNKA	CLCNKB	CLPP	CLRN1	COL2A1	COL4A3	COL4A4	COL4A5	COL9A1	COL11A1	COLEC11
DCAF17	DDX11	DNMT1	ECHS1	EDN3	EDNRB	EYA1	FDXR	FGF3	FGFR3	FTO	GATA3	HARS1	HARS2
HNRNPk	HOXA1	HOXB1	HSD17B4	KCNE1	KCNJ10	KCNQ1	KDM6A	KMT2D	LARS2	LHX3	LRP2	MASP1	MITF
MT-TK	NDP	PAX3	PEX1	PEX2	PEX26	PEX3	PEX5	PEX6	PISD	POGZ	PTPN11	RAF1	RMND1
SALL1	SERAC1	SLC19A2	SLC33A1	SLC52A2	SLC52A3	SLITRK6	SNAI2	SOX10	SPATA5	TIMM8A	TRPV4	TWNK	USH1G
USH2A	XYLT2												

CANDIDATOS - SINDRÓMICOS (48 genes)

AP3D1	ATP6V1B2	BMP4	CASK	CDK10	CDK8	COL9A2	COQ6	DPF2	ELOVL1	ERAL1	EXOSC2	FAM149B1	FBLN1
FGFR1	FGFR2	GPC4	GSTP1	GTF2IRD1	LOXL3	MAF	MAFB	MARS2	MRPS2	MT-CO3	NDUFA13	NFIX	NTRK3
PANX1	PIK3R4	PLOD3	PMP22	POLD1	PSIP1	PTPRD	SCHIP1	SEMA3E	SIX5	SLC4A11	SLC9A1	SMARCB1	TBL1XR1
TK2	TMEM67	TP63	TUBB4B	TWIST1	UNC45A								

CONSISTENTEMENTE ASOCIADOS - NO SINDRÓMICOS (61 genes)

CABP2	CCDC50	CDH23	CEACAM16	CLDN14	COCH	DIABLO	DIAPH1	EPS8L2	ESPN	ESRRB	GIPC3	GJB3	GJB6
GRHL2	GRXCR1	GSDME	HGF	ILDR1	KCNQ4	LHFPL5	LMX1A	LOXHD1	LRTOMT	MARVELD2	MIR96	MPZL2	MSRB3
MT-RNR1	MYH14	MYO15A	MYO3A	OPA1	OSBPL2	OTOA	OTOF	OTOG	OTOGL	P2RX2	PCDH15	PDZD7	PJKV
PLS1	POU3F4	POU4F3	PTPRQ	RDX	SERPINB6	SLC17A8	SMPX	STRC	SYNE4	TECTA	TJP2	TMC1	TMEM132E
TMIE	TMPRSS3	TPRN	TRIOBP	TSPEAR									

CANDIDATOS - NO SINDRÓMICOS (35 genes)

ADCY1	ATP2B2	BDP1	CCS	CD151	CD164	CDC14A	CLIC5	COL4A6	COL9A3	CRYM	DCDC2	DIAPH3	DSPP
ELMOD3	EPS8	GRXCR2	HMX2	HMX3	HOMER2	MCM2	MT-TA	MT-TE	PNPT1	PPIP5K2	REST	RIPOR2	ROR1
S1PR2	SLC22A4	SLC26A5	SLC44A4	TMPRSS5	TNC	WBP2							

CONSISTENTEMENTE ASOCIADOS - SINDRÓMICOS/NO SINDRÓMICOS (26 genes)

ACTG1	ADGRV1	AIFM1	BSND	CIB2	COL11A2	EYA4	GJB2	GPSM2	KARS	MT-CO1	MT-TH	MT-TL1	MT-TS1
MYH9	MYO6	MYO7A	NARS2	NLRP3	PRPS1	SIX1	SLC26A4	TBC1D24	USH1C	WFS1	WHRN		
MYH9	MYO6	MYO7A	NARS2	NLRP3	PRPS1	SIX1	SLC26A4	TBC1D24	USH1C	WFS1	WHRN		

CANDIDATOS - SINDRÓMICOS/NO SINDRÓMICOS (6 genes)

FOX11	KITLG	MT-TS2	PTPN1	RAI1	YWHAH								
-------	-------	--------	-------	------	-------	--	--	--	--	--	--	--	--

Página | 179

Anexo D: Artículo adicional Acta de Otorrinolaringológica Española

Model
OTORRI-955; No. of Pages 9

ARTICLE IN PRESS

Acta Otorrinolaringol Esp. 2019;xxx(xx):xxx-xxx



Acta Otorrinolaringológica Española

www.elsevier.es/otorrino



ARTÍCULO ORIGINAL

Utilidad clínica de la secuenciación de nueva generación en el diagnóstico etiológico de la hipoacusia neurosensorial en una Unidad de Hipoacusia Infantil

María Costales^{a,*}, Marta Diñeiro^b, Guadalupe A. Cifuentes^b, Raquel Capín^b,
Andrea Otero^b, Mónica Viejo-Díaz^c, Ana Plasencia^c, Faustino Núñez^a,
Justo Ramón Gómez^a, José Luis Llorente^a, Juan Cadiñanos^b y Rubén Cabanillas^b

^a Servicio de Otorrinolaringología, Hospital Universitario Central de Asturias, Oviedo, España

^b Instituto Medicina Oncológica Molecular de Asturias (IMOMA), Oviedo, España

^c Servicio de Genética, Hospital Universitario Central de Asturias, Oviedo, España

Recibido el 18 de febrero de 2019; aceptado el 30 de mayo de 2019

PALABRAS CLAVE

Hipoacusia
neurosensorial;
Secuenciación de
ADN;
Niños

Resumen

Introducción: La hipoacusia neurosensorial (HNS) es el déficit sensorial más prevalente en nuestro medio. La secuenciación genómica de nueva generación (NGS) permite obtener un diagnóstico etiológico en un alto porcentaje de pacientes. Nuestro estudio piloto muestra los resultados de la aplicación sistemática de la NGS en una Unidad de Hipoacusia Infantil, así como sus implicaciones en el manejo clínico de los pacientes y sus familiares.

Material y método: Se incluyó a 27 pacientes diagnosticados de HNS entre 2014 y 2017 en los que se descartó una causa ambiental. El test genético consistió en un panel de genes analizados mediante NGS (panel OTOgenics™). Este panel ha sido diseñado para incluir genes asociados con hipoacusia neurosensorial o mixta, de inicio precoz o tardío, síndromica y no síndromica, independientemente de su patrón de herencia.

Resultados: Se obtuvo un diagnóstico genético en el 56% (15/27) de los pacientes (62% en el caso de las HNS bilaterales); 5/27 (19%) presentaron variantes patogénicas en el gen *GJB2* y el resto variantes patogénicas o probablemente patogénicas en otros genes asociados con HNS aislada (*PRX2*, *TECTA* y *STRC*), con HNS síndromicas (*CHD7*, *GATA3*, *COL4A5*, *MITF* y *SOX10*) o con HNS síndromicas y no síndromicas (*BSND*, *ACTG1* y *CDH23*).

Discusión: El diagnóstico etiológico de la HNS supone un desafío en la práctica clínica. Nuestra serie demuestra que es posible implementar el diagnóstico genético en la rutina asistencial y que esta información tiene implicaciones pronósticas y terapéuticas.

© 2019 Sociedad Española de Otorrinolaringología y Cirugía de Cabeza y Cuello. Publicado por Elsevier España, S.L.U. Todos los derechos reservados.

* Autor para correspondencia.

Correo electrónico: cost2@hotmail.com (M. Costales).

<https://doi.org/10.1016/j.otorri.2019.05.005>

0001-6519/© 2019 Sociedad Española de Otorrinolaringología y Cirugía de Cabeza y Cuello. Publicado por Elsevier España, S.L.U. Todos los derechos reservados.

Cómo citar este artículo: Costales M, et al. Utilidad clínica de la secuenciación de nueva generación en el diagnóstico etiológico de la hipoacusia neurosensorial en una Unidad de Hipoacusia Infantil. Acta Otorrinolaringol Esp. 2019. <https://doi.org/10.1016/j.otorri.2019.05.005>

Descargado para Anonymous User (n/a) en Gobierno del Principado de Asturias Consejería de Sanidad de ClínicaKey.es por Elsevier en marzo 09, 2020. Para uso personal exclusivamente. No se permiten otros usos sin autorización. Copyright ©2020. Elsevier Inc. Todos los derechos reservados.

KEYWORDS

Sensorineural hearing loss;
DNA sequencing;
Children

Clinical utility of next-generation sequencing in the aetiological diagnosis of sensorineural hearing loss in a Childhood Hearing Loss Unit**Abstract**

Introduction: Sensorineural hearing loss (SNL) is the most prevalent sensory deficit in our environment. Next generation genomic sequencing (NGS) enables an aetiological diagnosis in a high percentage of patients. Our pilot study shows the results of the systematic application of NGS in a Childhood Hearing Loss Unit, as well as its implications for the clinical management of patients and their families.

Material and method: We included 27 patients diagnosed with SNL between 2014 and 2017, in which an environmental cause was ruled out. The genetic test consisted of a panel of genes analyzed by NGS (OTOgenics™ panel). This panel has been designed to include genes associated with sensorineural or mixed hearing loss, early onset or late, syndromic and non-syndromic, regardless of their inheritance pattern.

Results: A genetic diagnosis was obtained in 56% (15/27) of the patients (62% in the case of bilateral SNL). Of the patients, 5/27 (19%) presented pathogenic variants in the *GJB2* gene and the rest pathogenic and / or probably pathogenic variants in other genes associated with isolated SNL (*PRX2*, *TECTA* and *STRC*), with syndromic SNL (*CHD7*, *GATA3*, *COL4A5*, *MITF* and *SOX10*) or with syndromic and non-syndromic SNL (*BSND*, *ACTG1* and *CDH23*).

Discussion: The aetiological diagnosis of SNL is a challenge in clinical practice. Our series demonstrates that it is possible to implement genetic diagnosis in the care routine and that this information has prognostic and therapeutic implications.

© 2019 Sociedad Española de Otorrinolaringología y Cirugía de Cabeza y Cuello. Published by Elsevier España, S.L.U. All rights reserved.

Introducción

La hipoacusia afecta aproximadamente a un 30% de la población en algún momento de sus vidas¹. La hipoacusia neurosensorial (HNS) de inicio precoz es una de las enfermedades hereditarias más frecuentes en nuestro medio². Aproximadamente, 2/3 de las HNS infantiles tienen una etiología genética, habitualmente no sindrómica y sin antecedentes familiares. En la era de la medicina de precisión, intentar obtener un diagnóstico etiológico de la HNS de inicio precoz es imperativo, siendo los estudios genéticos la prueba con mayor rendimiento diagnóstico^{2,3}. La prevalencia de la enfermedad y la implantación de los programas de detección precoz de la hipoacusia, unidos al desarrollo exponencial experimentado en los últimos años por las técnicas de secuenciación del ADN, convierten a la otorrinolaringología en una especialidad con potencial para liderar esta nueva forma de ejercer la medicina: preventiva, predictiva, personalizada y participativa.

Hacer un diagnóstico precoz de la pérdida auditiva tiene un gran impacto en el curso clínico del paciente y en su desarrollo vital, permitiendo una intervención o monitorización tempranas⁴. Obtener un diagnóstico etiológico tiene múltiples ventajas: 1) puede proporcionar información sobre el pronóstico de la hipoacusia, individualizando así el seguimiento; 2) permite identificar síndromes ocultos, anticipándose a sus manifestaciones, bien previniéndolas o bien paliando sus consecuencias, mejorando la calidad de vida de los pacientes; 3) permite evitar la realización de pruebas diagnósticas innecesarias, conteniendo el gasto médico y minimizando los riesgos para el paciente;

4) proporciona información sobre las diferentes opciones reproductivas, todas ellas con el objetivo común de evitar transmitir la enfermedad a futuros descendientes^{5,7}, y 5) en determinados casos, fundamentalmente sindrómicos, facilita la selección del tratamiento más adecuado. Asimismo, en un futuro cercano, un diagnóstico genético será imprescindible para lograr la inclusión de pacientes en ensayos clínicos de terapia génica, farmacológica o celular, dirigidos a corregir su anomalía molecular específica⁸.

El diagnóstico genético de la HNS ha supuesto un gran reto, dada la extrema heterogeneidad genética de la patología⁹⁻¹¹. Las nuevas tecnologías de secuenciación del ADN (*next generation sequencing* [NGS]), que permiten analizar simultáneamente, en plazos y costes compatibles con la práctica asistencial, cientos de genes, se han convertido en el nuevo estándar para obtener un diagnóstico etiológico de la hipoacusia hereditaria^{12,13}.

Tanto la guía del año 2014 para la «Evaluación clínica y el diagnóstico etiológico de la hipoacusia» del Colegio Americano de Genética Médica y Genómica (ACMG)², como las recomendaciones españolas del año 2015 para el «Diagnóstico etiológico de la sordera infantil», de la Comisión para la Detección Precoz de la Hipoacusia (CODEPEH)¹⁴, recomiendan incorporar el diagnóstico genético precozmente en los protocolos para el diagnóstico de la sordera. Este consistirá o bien en un estudio secuencial en el que se analicen en primer lugar los genes *GJB2* y *GJB6*, y, en los casos negativos, en un segundo tiempo, un panel de genes mediante NGS, o bien directamente en el análisis de un panel de genes en los casos en los que se considere indicado.

Cómo citar este artículo: Costales M, et al. Utilidad clínica de la secuenciación de nueva generación en el diagnóstico etiológico de la hipoacusia neurosensorial en una Unidad de Hipoacusia Infantil. Acta Otorrinolaringol Esp. 2019. <https://doi.org/10.1016/j.otorri.2019.05.005>

El propósito de este artículo es mostrar los primeros resultados de la implantación en la práctica clínica de la NGS en una Unidad de Hipoacusia Infantil, haciendo especial hincapié en su potencial utilidad para el manejo de los pacientes.

Material y método

Pacientes

Se incluyó a 27 pacientes diagnosticados de HNS o hipoacusia mixta en la Unidad de Hipoacusia Infantil del Hospital Universitario Central de Asturias entre los años 2014 y 2017. En el 56% (15/27) de los pacientes la HNS se diagnosticó tras no haber pasado el cribado neonatal. El 44% restante (12/27) se remitió posteriormente a la Unidad, por detectarse la hipoacusia a una edad más avanzada o por haber nacido en otra comunidad autónoma. En ningún paciente, se identificó una causa ambiental que pudiera justificar su pérdida auditiva. Todos ellos fueron evaluados por un especialista en genética clínica, ante la sospecha de una etiología hereditaria.

Para etiquetar la pérdida auditiva se siguió la clasificación de la ACMG².

Este estudio fue aprobado por el Comité de Ética de Investigación del Principado de Asturias (proyecto de investigación #75/14).

Panel genético

Tras obtener 4ml de sangre periférica en tubos EDTA y extraer el ADN genómico, se procedió a secuenciar un panel de genes mediante NGS (panel OTOgenics™, IMOMA, España). Este panel ha sido diseñado para incluir genes asociados con HNS o hipoacusia mixta, de inicio precoz o

tardío, síndromica o no síndromica, independientemente de su patrón de herencia. La metodología utilizada en el panel ha sido publicada previamente por Cabanillas et al.¹⁵. La primera versión empleada del panel incluía 165 genes: 76 genes implicados en HNS síndromicas, 62 genes en HNS no síndromicas y 27 genes relacionados con ambos tipos. La segunda versión incluía 184 genes: 84 asociados con HNS síndromica, 73 relacionados con HNS no síndromica y 27 genes descritos en ambos tipos (tabla 1).

La interpretación clínica y clasificación final de las variantes genéticas se obtuvo tras el análisis tanto de las bases de datos (HGMD profesional [Qjagen], ClinVar [<https://www.ncbi.nlm.nih.gov/clinvar/>], OMIM [<https://www.omim.org/>], PubMed [<https://www.ncbi.nlm.nih.gov/pubmed/>], proyecto 1000 Genomes [<https://www.genome.gov/27528684/1000-genomes-project>], ESP [<https://esp.gs.washington.edu/drupal/>] y ExAC [<http://exac.broadinstitute.org/about/>]) como de la bibliografía disponible en cada caso. Las variantes se clasificaron según los criterios de la ACMG² como patogénica (P) (clase 5), probablemente patogénica (PP) (clase 4), significado incierto (clase 3), probablemente benigna (clase 2) y benigna (clase 1).

Resultados

Características clínicas y audiológicas de los pacientes

El 74% (20/27) de las HNS se clasificaron como congénitas. La media de edad al diagnóstico fue de 2,6 años, con un rango de entre los 0 y los 10 años de edad.

Los detalles clínicos relevantes se detallan en la tabla 2; 19/27 (70%) pacientes no presentaban antecedentes

Tabla 1 Genes incluidos en el panel versión 2

Genes consistentemente asociados en la literatura científico-médica con hipoacusia neurosensorial o mixta

ABHD12	BSND	CLRN1	DIAPH1	GJB3	KCNJ10	MSRB3	NARS2	PEX1	SERAC1	STRC	USH1G
ACTB	CABP2	COCH	DNMT1	GJB6	KCNQ1	MT-CO1	NDP	PEX2	SERPINB6	SYNE4	USH2A
ACTG1	CACNA1D	COL2A1	EDN3	GPSM2	KCNQ4	MT-RNR1	NLRP3	PEX3	SIX1	TBC1D24	WFS1
ADGRV1	CCDC50	COL4A3	EDNRB	GRHL2	LARS2	MT-TH	OPA1	PEX5	SLC17A8	TECTA	
AIFM1	CDH23	COL4A4	ESPN	GRXCR1	LHFPL5	MT-TK	OSBPL2	PEX26	SLC19A2	TIMM8A	
ALMS1	CEACAM16	COL4A5	ESRRB	HARS2	LHX3	MT-TL1	OTOA	POU3F4	SLC26A4	TJP2	
ANKH	CHD7	COL9A1	EYA1	HGF	LOXHD1	MT-TS1	OTOF	POU4F3	SLC33A1	TMC1	
AP1S1	CIB2	COL11A1	EYA4	HOXA1	LRP2	MYH9	OTOG	PRPS1	SLC52A2	TMIE	
ATP1A3	CISD2	COL11A2	FGF3	HOXB1	LRTOMT	MYH14	OTOGL	PTPN11	SLC52A3	TMPRSS3	
ATP6V1B1	CLCNKA	DDX11	FGFR3	HSD17B4	MARVELD2	MYO3A	P2RX2	PTPRQ	SLITRK6	TPRN	
BCAP31	CLCNKB	DFNA5	GATA3	ILDR1	MASP1	MYO6	PAX3	RAF1	SMPX	TRIOBP	
BCS1L	CLDN14	DFNB59	GIPC3	KARS	MIR96	MYO7A	PCDH15	RDX	SNAI2	TSPEAR	
BRAF	CLPP	DIABLO	GJB2	KCNE1	MITF	MYO15A	PDZD7	RMND1	SOX10	USH1C	

Genes candidatos con evidencia preliminar a favor de su asociación con hipoacusia neurosensorial o mixta

ADCY1	COL4A6	CRYM	ELMOD3	FGFR1	HMX2	MT-TA	SEMA3E	SLC4A11	SLC26A5	TMEM132E	TNC
ATP2B2	COL9A2	DCDC2	EPS8	FGFR2	HMX3	MT-TE	SIX5	SLC9A1	TK2	TMPRSS5	TP63
ATP6V1B2	COL9A3	DIAPH3	FAM65B	FOX11	HOMER2	MT-TS2					
BDP1	COQ6	DSPP	FBLN1	GRXCR2	MT-CO3	PNPT1					

Resaltados los genes que mostraron variantes patogénicas/probablemente patogénicas en nuestros pacientes.

Cómo citar este artículo: Costales M, et al. Utilidad clínica de la secuenciación de nueva generación en el diagnóstico etiológico de la hipoacusia neurosensorial en una Unidad de Hipoacusia Infantil. Acta Otorrinolaringol Esp. 2019. <https://doi.org/10.1016/j.otorri.2019.05.005>

Tabla 2 Características de los pacientes

Casos	AF hipoacusia	Otros AF de interés	FR perinatales	Malformaciones otológicas	Problemas asociados	Tipo hipoacusia	Resultado estudio genético (gen responsable)
1	No	No	No	No	No	HNS bilateral simétrica profunda prelocutiva	Positivo (GJB2 / GJB6)
2	No	No	No	Dilatación conductos auditivos internos	Hipertelorismo	HNS bilateral simétrica profunda prelocutiva	Positivo (SOX10)
3	No	No	No	No	No	HNS bilateral simétrica profunda prelocutiva	Positivo (CDH23)
4	No	No	No	No	No	HNS bilateral simétrica severa prelocutiva	Positivo (GATA3)
5	No	No	No	No	No	Hipoacusia mixta bilateral simétrica	Positivo (ACTG1)
6	Bisabuela paterna	Padre y abuela paterna β-talasemia <i>minor</i>	No	Pabellones despegados	Testes atróficos, hipogonadismo, micropene, hipoplasia escrota, coartación aórtica, subclavia aberrante	severa prelocutiva HNS bilateral simétrica severa prelocutiva	Positivo (CHD7)
7	Varios mal documentados	No	No	No	No	HNS bilateral simétrica profunda prelocutiva	Positivo (GJB2)
8	No	No	No	No	No	HNS bilateral simétrica profunda prelocutiva	Positivo (GJB2)
9	No valorable	No valorable	No valorable	No	Lenticono bilateral, catarata en ojo derecho, glomeruloesclerosis	HNS bilateral simétrica moderada poslocutiva	Positivo (COL4A5)
10	No	No	No	No	Pie zambo	HNS bilateral simétrica profunda prelocutiva	Positivo (BSND)
11	Abuelo materno, tía abuela materna	Madre y abuelo de pelo blanco	No	No	No	HNS unilateral profunda prelocutiva	Positivo (MITF)
12	No	No	No	No	No	HNS bilateral asimétrica moderada-severa prelocutiva	Positivo (GJB2)
13	Bisabuelo materno	No	No	No	No	HNS bilateral simétrica moderada poslocutiva	Positivo (P2RX2)
14	No	Tío materno diabetes mellitus tipo 1	No	No	Diabetes mellitus pendiente de filiar	HNS bilateral simétrica leve-moderada poslocutiva	Positivo (GJB2)
15	Hermano, tío materno, prima paterna	No	No	No	No	HNS bilateral simétrica moderada poslocutiva	Positivo (STRC/TECTA)

Cómo citar este artículo: Costales M, et al. Utilidad clínica de la secuenciación de nueva generación en el diagnóstico etiológico de la hipoacusia neurosensorial en una Unidad de Hipoacusia Infantil. Acta Otorrinolaringol Esp. 2019. <https://doi.org/10.1016/j.otorri.2019.05.005>

Tabla 2 (continuación)

Casos	AF hipoacusia	Otros AF de interés	FR perinatales	Malformaciones otológicas	Problemas asociados	Tipo hipoacusia	Resultado estudio genético (gen responsable)
16	No	No	No	No	No	HNS bilateral simétrica moderada prelocutiva	Negativo
17	No	No	No	No	No	HNS unilateral profunda prelocutiva	Negativo
18	Abuelo y tío maternos	Abortos frecuentes en la familia materna	No	No	Clinodactilia 5.º dedo	HNS bilateral simétrica leve- moderada poslocutiva	Negativo
19	No	No	No	No	No	HNS unilateral severa prelocutiva	Negativo
20	Hermano y bisabuela materna	No	No	No	No	HNS unilateral profunda prelocutiva	Negativo
21	No	No	No	No	Déficit factor VII coagulación	HNS bilateral asimétrica moderada/severa poslocutiva	Negativo
22	No	Alteración en el factor II de la coagulación en la familia materna	No	No	Estrabismo bilateral, alternante, convergente con instigmo ocasional	HNS bilateral simétrica severa prelocutiva	Negativo
23	Madre	No	No	No	No	HNS unilateral moderada poslocutiva	Negativo
24	No	No	No	No	No	HNS bilateral simétrica moderada prelocutiva	Negativo
25	Tía paterna	No	Sufrimiento fetal. Ictericia	Dilatación del acueducto vestibular	No	HNS bilateral asimétrica profunda/leve prelocutiva	Negativo
26	No	No	No	Cóclea: partición incompleta tipo III y dilatación conductos auditivos internos	No	HNS bilateral simétrica profunda prelocutiva	Negativo
27	No	Padre: polidactilia postaxial en pies y una mano y polidactilia en 3.ª y 4.º dedos de la mano contralateral. Abuelo paterno: hipospadias	No	Lóbulos orejas surcado. Cóclea: cavidad quística única bilateral con displasia semicircular asociada	Poli y sindactilias. Hipospadias. Foramen ovale permeable	HNS bilateral simétrica profunda prelocutiva	Negativo

AF: antecedentes familiares; FR: factores de riesgo; HNS: hipoacusia neurosensorial.

Cómo citar este artículo: Costales M, et al. Utilidad clínica de la secuenciación de nueva generación en el diagnóstico etiológico de la hipoacusia neurosensorial en una Unidad de Hipoacusia Infantil. Acta Otorrinolaringol Esp. 2019. <https://doi.org/10.1016/j.otorri.2019.05.005>

Descargado para Anonymous User (n/a) en Gobierno del Principado de Asturias Consejería de Sanidad de ClinicalKey.es por Elsevier en marzo 09, 2020. Para uso personal exclusivamente. No se permiten otros usos sin autorización. Copyright ©2020. Elsevier Inc. Todos los derechos reservados.

familiares de hipoacusia. Entre los 8/27 (30%) que sí los presentaban, en 1 fue de primer grado, en 2 de segundo grado, en 4 de tercer grado y uno tenía varios familiares afectos mal definidos en la historia clínica.

Veintidós de veintisiete (81%) casos no presentaban antecedentes familiares de otro tipo; 5/27 (19%) tenían antecedentes familiares de otra enfermedad relevante.

Según la clasificación de la ACMG²:

- 19/27 (70%) HNS bilaterales simétricas, 3/27 (11%) bilaterales asimétricas y 5/27 (19%) unilaterales.
- 19/27 (70%) HNS prelocutivas, 1/27 (4%) hipoacusia mixta prelocutiva; 7/27 (26%) HNS poslocutivas.
- Dentro de las simétricas: 16/27 (59%) severas o profundas; 3/27 (11%) leve-moderada/moderada. Asimétricas: 2/27 (7%) moderadas en un oído y severas en el oído contralateral y 1/27 (4%) leve en un oído y profunda en el contralateral.
- Hasta el momento, salvo las detectadas poslocutivamente, y que, por tanto, han variado, el resto se mantienen estables.

En la evolución: 9/27 (33%) pacientes implantados; 12/27 (44%) con prótesis auditivas adaptadas; 5/27 (19%) en seguimiento; 1/27 (4%) no consta porque le controlan en otra comunidad autónoma. La decisión de implantar uni o bilateralmente se tomó de manera individualizada en cada caso, en función de las indicaciones del momento y de las características del paciente. Las características clínicas de los pacientes se encuentran en la [tabla 2](#).

Variantes genéticas patogénicas y probablemente patogénicas identificadas

Se obtuvo un diagnóstico genético en el 56% (15/27) de los pacientes. En concreto, en el 62% de las HNS bilaterales (13/21) y en el 40% de las unilaterales (2/5). El 19% (5/27) presentó variantes P en el *locus DFNB1* (genes *GJB2* y *GJB6*) y el resto variantes P o PP en otros genes: *PRX2* y *STRC* asociados con HNS no síndrómica; *CHD7*, *GATA3*, *COL4A5*, *MITF* y *SOX10* asociados con hipoacusias síndrómicas, y *BSND*, *ACTG1* y *CDH23* asociados con ambas.

El 67% (10/15) de las variantes encontradas presentaron un patrón de herencia autosómico recesivo (AR) y el 33% (5/15) un patrón autosómico dominante (AD). Se hallaron variantes consideradas causantes del fenotipo de hipoacusia en 14/19 pacientes con hipoacusia bilateral y en 1/5 con hipoacusia unilateral (mutación en *MITF* causante del síndrome de Waardenburg). Además, se identificaron 3 variantes P/PP *de novo*: *ACTG1* (no descrita previamente en la literatura), *GATA3* y *COL4A5*.

Correlaciones genotipo-fenotipo

Clasificamos nuestros resultados en función de la alteración genética identificada y su fenotipo en 3 grupos ([tabla 3](#)):

- 1) HNS síndrómicas sospechadas pretest: 7% (2/27) de los casos diagnosticados genéticamente (*probandos 6 y 9*).
- 2) HNS síndrómicas no sospechadas pretest («síndromes ocultos»): 11% (3/27) de los casos diagnosticados

genéticamente encajan en esta clasificación (*probandos 2, 4 y 11*), pues las alteraciones genéticas identificadas solo se asocian con HNS síndrómicas. Adicionalmente, otro 11% (*probandos 3, 5 y 10*) presentan alteraciones genéticas que se pueden asociar tanto a HNS síndrómicas, como no síndrómicas. La evolución clínica de los pacientes, siendo conscientes de las potenciales complicaciones que podrían desarrollar, nos permitirá clasificarlos en uno u otro grupo en el futuro. En todos ellos, la hipoacusia fue la única manifestación clínica identificada pretest.

- 3) HNS aisladas no síndrómicas: 26% (7/27) de los casos diagnosticados (*probandos 1, 7, 8, 12, 13, 14 y 15*).

Discusión

Alrededor de los 4 años de edad, la plasticidad neuronal está prácticamente finalizada. De ahí la importancia de una estimulación lo más precoz posible para conseguir un adecuado desarrollo tanto cognitivo, como de la comunicación oral¹⁴. Obtener un diagnóstico etiológico resulta sumamente importante, pues tiene múltiples ventajas para el paciente y sus familiares¹⁶. Los resultados de nuestro estudio piloto no solo demuestran que actualmente la NGS permite integrar el diagnóstico genético en la práctica clínica, si no que proporcionan ejemplos de los beneficios derivados de este diagnóstico. Asimismo, nuestra serie, a pesar de su tamaño limitado, refleja con fidelidad las características genéticas de las hipoacusias hereditarias (heterogeneidad genética, expresividad variable, penetrancia incompleta), características que, clásicamente, han supuesto un reto para las tecnologías de secuenciación convencionales (secuenciación Sanger). Más aún, nuestros resultados nos proporcionan una primera aproximación a la «epidemiología molecular» de nuestra población con HNS de inicio precoz, información que resulta de vital importancia para su prevención, diagnóstico y tratamiento.

El rendimiento diagnóstico del panel OTOgenics™ en nuestra cohorte ha sido del 56% (15/27). Los 12 casos potencialmente genéticos en los que no se ha identificado ninguna variante responsable de la hipoacusia siguen a día de hoy sin diagnóstico etiológico. Esto podría deberse a una causa ambiental no identificada o a las limitaciones propias de la tecnología empleada (grandes deleciones/duplicaciones, reordenamientos, variantes en regiones genómicas no implicadas hasta la fecha en HNS, etc.). El rendimiento diagnóstico se incrementa cuando evaluamos exclusivamente los casos con HNS bilateral, alcanzado el 59% (13/22).

En nuestra serie, el tipo de herencia difiere de lo descrito en la literatura, identificando un 67% de casos con herencia AR y un 33% herencia AD. Puesto que más de la mitad (3/5) de los casos AD fueron causados por variantes *de novo* (*GATA3*, *ACTG1* y *COL4A5*), el hecho de haber utilizado un panel agnóstico capaz de identificar variantes *de novo* inesperadas justifica estos resultados¹⁵. Esta información resulta de vital importancia, pues en ausencia de antecedentes familiares de hipoacusia *a priori* se podría haber sospechado un patrón de herencia AR y, sin embargo, se debe tener en cuenta que en nuestro medio las mutaciones *de novo* son una causa relativamente frecuente de HNS.

Cómo citar este artículo: Costales M, et al. Utilidad clínica de la secuenciación de nueva generación en el diagnóstico etiológico de la hipoacusia neurosensorial en una Unidad de Hipoacusia Infantil. Acta Otorrinolaringol Esp. 2019. <https://doi.org/10.1016/j.otorri.2019.05.005>

Tabla 3 Correlaciones genotipo-fenotipo

Edad diagnóstica de la HNS	Fenotipo pretest	Variantes genéticas P y PP identificadas	Fenotipos asociados a las variantes genéticas identificadas	Patrón de herencia	Manejo audiológico
1	Congénita HNS bilateral simétrica profunda aislada	Heterocigoto compuesto <i>GJB2</i> (c.35delG;p.Gly12Valfs*2) y <i>GJB6</i> (c.-237135..-5094del; del(GJB6-D1351854)) <i>SOX10</i> (c.135..154del;p.Ser45Argfs*15) en heterocigosis	HNS no sindrómica (DFNB1)	AR	IC unilateral
2	Congénita HNS bilateral simétrica profunda con hiperteiorismo	<i>CDH23</i> (c.4488G>C;p.Gln1496His) en heterocigosis y exones 11, 12, 13, 13 y 15 duplicados en <i>CDH23</i> <i>GATA3</i> (c.1018A>C;p.N340H) en heterocigosis <i>ACTG1</i> (c.548G>A;p.R183Q) en heterocigosis <i>GHD7</i> (c.235A>T;p.Lys79*) en heterocigosis <i>GJB2</i> (c.35delG;p.G12Vfs*2) en homocigosis <i>GJB2</i> (c.35delG;p.G12Vfs*2) en homocigosis <i>KCNE1</i> (p.Asp76Asn) en heterocigosis <i>COL4A5</i> (c.3525..3529dupTGGAC;p.P1177Lfs*124) en hemocigosis	Síndrome de Waardenburg tipo 2B Síndrome de Waardenburg tipo 4C Síndrome PCWH (o síndrome de Waardenburg tipo 4 «plus») HNS no sindrómica (DFNB12) Síndrome de Usher tipo 1D (USH1D) Síndrome de Barakat HNS no sindrómica (DFNA20/26) Síndrome de Baritser-Winster Síndrome CHARGE	AD AR AD AD AR AR AD AR AD AR	IC unilateral IC bilateral PA bilateral PA bilateral PA bilateral IC bilateral IC unilateral+PA PA bilateral
3	Congénita HNS bilateral simétrica profunda aislada	<i>BSND</i> (c.23G>A;p.R8Q) en homocigosis <i>MITF</i> (c.909G>A;p.T303T) en heterocigosis <i>GJB2</i> (c.35delG;p.G12Vfs*2) en homocigosis <i>P2RX2</i> (c.178G>T;p.Val60Leu) en heterocigosis Heterocigoto compuesto <i>GJB2</i> (c.35delG;p.G12Vfs*2) y <i>GJB2</i> (c.101T>C; p.M34T)	HNS no sindrómica (DFNB73) Síndrome de Bartter tipo iv Síndrome de Waardenburg tipo 2A HNS no sindrómica (DFNB1) HNS no sindrómica (DFNA41) HNS no sindrómica (DFNB1) HNS no sindrómica (DFNA 8/12)	AR AR AD AR AR AR AD	IC unilateral+PA Control PA bilateral PA bilateral Control PA bilateral
4	Congénita HNS bilateral simétrica severa	Delección de todos los exones del gen <i>STRC</i> en homocigosis (la delección incluye también al gen <i>CATSPEAR2</i>) TECTA en heterocigosis	HNS no sindrómica (DFNB16) HNS no sindrómica (DFNA 8/12)	AR AD	PA bilateral AD

AD: autosómico dominante; AR: autosómico recesivo; HNS: hipoacusia neurosensorial; IC: implante coclear; P: patogénica; PA: prótesis auditiva; PP: probablemente patogénica.

Cómo citar este artículo: Costales M, et al. Utilidad clínica de la secuenciación de nueva generación en el diagnóstico etiológico de la hipoacusia neurosensorial en una Unidad de Hipoacusia Infantil. Acta Otorrinolaringol Esp. 2019. <https://doi.org/10.1016/j.otorri.2019.05.005>

Una de las cuestiones fundamentales a la hora de planificar una estrategia para afrontar el diagnóstico etiológico de la HNS es conocer el porcentaje de casos que pueden ser explicados por mutaciones en los genes que conforman el locus *DFNB1* (*GJB2* y *GJB6*). Dependiendo de esta frecuencia, se justificará o no la realización de un cribado previo al empleo de tecnología NGS, mediante el análisis de estos genes con métodos convencionales (secuenciación Sanger de *GJB2* y PCR específicas de grandes deleciones frecuentes de *GJB6*). Variantes P en este locus pueden ser responsables de hasta el 50% de los casos de HNS de inicio precoz. Sin embargo, su prevalencia es muy variable según la población, oscilando entre el 0 y el 60%. La variante P *GJB2* c.35delG es la causa más frecuente de HNS autosómica recesiva en la población caucásica^{17,18}. En nuestra población tenemos una frecuencia del 18,5% de casos explicados por mutaciones en el locus *DFNB1*, lo que, *a priori*, justificaría realizar un pretest previo al análisis mediante un panel de genes¹⁴.

Una de las características de las HNS hereditarias que añade dificultad a su diagnóstico y tratamiento es la expresividad variable de las mismas^{2,4,6,17}. En nuestra serie encontramos varios ejemplos que reflejan esta característica, empezando por el gen *GJB2*. Se han descrito distintos grados de hipoacusia aún con un misma alteración genética y determinadas alteraciones de este gen también pueden ser causantes de DFNA3A (HNS con herencia AD) y de síndromes de mayor o menor gravedad^{18,19}.

Además de su utilidad para el asesoramiento genético y la planificación familiar, en ocasiones los beneficios del estudio genético van más allá, proporcionándonos información inesperada (hallazgos secundarios). Así, el probando 8 presentó una variante P en heterocigosis en el gen *KCNE1* heredada por vía paterna, que se ha asociado a síndrome de QT largo. Los portadores silenciosos de mutaciones P presentan un riesgo de eventos cardíacos. Debe considerarse el tratamiento con betabloqueantes, evitar fármacos que prolonguen el intervalo QT y corregir posibles anomalías electrolíticas⁶.

Otro ejemplo de medicina predictiva son los síndromes ocultos. Se identificaron 3 síndromes ocultos (probandos 2, 4 y 11, genes *SOX10*, *GATA3* y *MITF*) y otras 3 alteraciones genéticas que se podrían asociar tanto a HNS sindrómicas, como a no sindrómicas (probandos 3, 5 y 10, genes *CDH23*, *ACTG1* y *BSND*), condicionando todos estos diagnósticos genéticos el manejo clínico de estos pacientes.

Una cuestión que debemos tener en cuenta a la hora de evaluar los resultados de un estudio genético basado en NGS es la relativa a las limitaciones inherentes a esta tecnología y si en el test realizado se han tomado las precauciones necesarias para minimizarlas. Este es el caso del gen *STRC*, identificado como responsable de la HNS del probando 15. La similitud entre *STRC* y *STRCP1* (su pseudogen) es excepcional, ambos son homólogos en el 98,9% (en exones e intrones) e idénticos en > 99% (únicamente secuencia codificante), por este motivo, el análisis de este gen representa un gran desafío para las metodologías estándar de NGS. Las dificultades metodológicas necesarias para evitar la interferencia que pueden generar los pseudogenes podrían justificar la hipótesis actual de que *STRC* es uno de los genes más contribuyentes a la HNS autosómica recesiva²⁰.

Finalmente, nuestros resultados ponen de manifiesto que los estudios genéticos pueden aportar información

clínicamente relevante también en el caso de HNS unilaterales, si bien en nuestro estudio piloto su rendimiento es inferior al obtenido en las HNS bilaterales (40% versus 62%). La HNS unilateral ha cobrado mayor importancia desde la aprobación del *screening* neonatal de sordera, ya que antes era un problema infradiagnosticado. Se cree que tiene una prevalencia de 1,7/1.000. Hoy en día se sabe que la pérdida de binauralidad puede tener implicaciones negativas en el desarrollo del lenguaje y de la expresión, en el comportamiento y en los méritos académicos. Frecuentemente, las HNS unilaterales se han asociado a malformaciones del oído interno, por lo que se recomienda siempre realizar una prueba de imagen, así como un estudio genético²¹⁻²³.

El estudio genético que realizamos en este trabajo piloto supuso un coste aproximado inferior a los 2.000 € por paciente, incluida la parte técnica más la interpretación y el informe de los resultados. Aproximadamente se dedican unas 10 h a la elaboración del informe que incluye: la interpretación de los datos y la comparación de los mismos con los datos previos publicados en las bases de datos. La secuenciación genómica es una técnica rentable en términos de coste-beneficio en la población pediátrica, ya que ha incrementado la tasa de diagnóstico a un 16-79% y ha disminuido el coste un 11-64% respecto a la vía diagnóstica estándar²⁴.

Nuestro estudio piloto no solo demuestra que es posible y útil la implementación del diagnóstico genético en la práctica clínica, sino también que realizar una primera aproximación agnóstica y sistemática a las bases genéticas de nuestra población de pacientes con HNS reporta ventajas clínicamente relevantes. Estos resultados deberán ser validados y ampliados en cohortes más numerosas en las que el objetivo al que debemos aspirar es alcanzar un diagnóstico etiológico en el 100% de los pacientes con HNS de inicio precoz.

Conclusiones

- La secuenciación de un panel de genes mediante tecnología NGS ha permitido establecer un diagnóstico genético en el 56% de los casos evaluados (59% en el caso de HNS bilaterales), confirmando los test genéticos como la prueba con un mayor rendimiento diagnóstico en la HNS de inicio precoz.
- En nuestra población, el porcentaje de variantes P en los genes *GJB2* y *GJB6* se acerca al que justificaría realizar un cribado previo a la secuenciación NGS mediante el análisis de estos genes con tecnología convencional.
- La HNS de causa genética identificada en nuestra muestra cumple con las características propias de las sorderas hereditarias (heterogeneidad genética, expresividad variable y penetrancia incompleta).

Conflicto de intereses

Los autores no presentan conflicto de intereses.

Agradecimientos

El trabajo desarrollado en el IMOMA ha sido financiado por la Fundación María Cristina Masaveu Peterson (España).

Cómo citar este artículo: Costales M, et al. Utilidad clínica de la secuenciación de nueva generación en el diagnóstico etiológico de la hipoacusia neurosensorial en una Unidad de Hipoacusia Infantil. Acta Otorrinolaringol Esp. 2019. <https://doi.org/10.1016/j.otorri.2019.05.005>

Bibliografía

- Smith RJ, Bale JF Jr, White KR. Sensorineural hearing loss in children. *Lancet*. 2005;365:879–90.
- Alford RL, Amos KS, Fox M, Lin JW, Palmer CG, Pandya A, et al. American College of Medical Genetics and Genomics guideline for the clinical evaluation and etiologic diagnosis of hearing loss. *Genet Med*. 2014;16:347–55.
- Shearer AE, deLuca AP, Hildebrand MS, Taylor KR, Gurrola J 2nd, Scherer S, et al. Comprehensive genetic testing for hereditary hearing loss using massively parallel sequencing. *Proc Natl Acad Sci USA*. 2010;107:21104–9.
- Cabanillas Farpón R, Cadiñanos Bañales J. Hereditary hearing loss: Genetic counselling. *Acta Otorrinolaringol Esp*. 2012;63:218–29.
- Tekin D, Yan D, Bademci G, Feng Y, Guo S, Foster J 2nd, et al. A next-generation sequencing gene panel (MiamiOtoGenes) for comprehensive analysis of deafness genes. *Hear Res*. 2016;333:179–84.
- Tekin D, Tutar E, Ozturkmen Akay H, Blanton S, Foster J 2nd, Tekin M. Comprehensive genetic testing can save lives in hereditary hearing loss. *Clin Genet*. 2015;87:190–1.
- Gu X, Guo L, Ji H, Sun S, Chai R, Wang L, et al. Genetic testing for sporadic hearing loss using targeted massively parallel sequencing identifies 10 novel mutations. *Clin Genet*. 2015;87:588–93.
- Chien WW, Monzack EL, McDougald DS, Cunningham LL. Gene therapy for sensorineural hearing loss. *Ear Hear*. 2015;36:1–7.
- Atik T, Bademci G, Diaz-Horta O, Blanton SH, Tekin M. Whole-exome sequencing and its impact in hereditary hearing loss. *Genet Res (Camb)*. 2015;97:e4.
- Bademci G, Foster J 2nd, Mahdieh N, Bonyadi M, Duman D, Cengiz FB, et al. Comprehensive analysis via exome sequencing uncovers genetic etiology in autosomal recessive nonsyndromic deafness in a large multiethnic cohort. *Genet Med*. 2016;18:364–71.
- Lin JW, Chowdhury N, Mody A, Tonini R, Emery C, Haymond J, et al. Comprehensive diagnostic battery for evaluating sensorineural hearing loss in children. *Otol Neurotol*. 2011;32:259–64.
- Shearer AE, Smith RJ. Massively parallel sequencing for genetic diagnosis of hearing loss: The new standard of care. *Otolaryngol Head Neck Surg*. 2015;153:175, 182.
- Sloan-Heggen CM, Smith RJ. Navigating genetic diagnostics in patients with hearing loss. *Curr Opin Pediatr*. 2016;28:705–12.
- Núñez-Batalla F, Jáudenes-Casaubón C, Sequí-Canet JM, Vivanco-Allende A, Zubizaray-Ugarteche J, Cabanillas-Farpón R. Aetiological diagnosis of child deafness: CODEPEH recommendations. *Acta Otorrinolaringol Esp*. 2017;68:43–55.
- Cabanillas R, Diñeiro M, Cifuentes GA, Castillo D, Prunedada PC, Álvarez R, et al. Comprehensive genomic diagnosis of non-syndromic and syndromic hereditary hearing loss in Spanish patients. *BMC Med Genomics*. 2018;11:58. <http://dx.doi.org/10.1186/s12920-018-0375-5>.
- Sakuma N, Moteki H, Takahashi M, Nishio SY, Arai Y, Yamashita Y, et al. An effective screening strategy for deafness in combination with a next-generation sequencing platform: A consecutive analysis. *J Hum Genet*. 2016;61:253–61.
- Burke WF, Warnecke A, Schoner-Heinisch A, Lesinski-Schiedat A, Maier MH, Lenarz T. Prevalence and audiological profiles of GJB2 mutations in a 630 large collective of hearing impaired patients. *Hear Res*. 2016;333:77–86.
- Gualandi F, Ravani A, Berto A, Sensi A, Trabanelli C, Falciano F, et al. Exploring the clinical and epidemiological complexity of GJB2-linked deafness. *Am J Med Genet*. 2002;112:38–45.
- Gallo-Terán J, Morales-Angulo C, Rodríguez-Ballesteros M, Moreno-Pelayo MA, del Castillo I, Moreno F. Prevalence of the 35delG mutation in the GJB2 gene, del (GJB6-D1351830) in the GJB6 gene, Q829X in the OTOF gene and A1555G in the mitochondrial 12S rRNA gene in subjects with non-syndromic sensorineural hearing impairment of congenital/childhood onset. *Acta Otorrinolaringol Esp*. 2005;56:463–8.
- Vona B, Hofrichter MA, Neuner C, Schröder J, Gehrig A, Hennenmann JB, et al. DFNB16 is a frequent cause of congenital hearing impairment: Implementation of STRC mutation analysis in routine diagnostics. *Clin Genet*. 2015;87:49–55.
- Paul A, Martin S, Parodi M, Rouillon I, Guertain J, Pingault V, et al. Unilateral sensorineural hearing loss: Medical context and etiology. *Audiol Neurootol*. 2017;22:83–8.
- Núñez-Batalla F, Jáudenes-Casaubón C, Sequí-Canet JM, Vivanco-Allende A, Zubizaray-Ugarteche J. Early diagnosis and treatment of unilateral or asymmetrical hearing loss in children: CODEPEH recommendations. *Acta Otorrinolaringol Esp*. 2018. <http://dx.doi.org/10.1016/j.otorri.2018.09.004>, piiS0001-6519(18)30178-X.
- Lin PH, Hsu CJ, Lin YH, Lin YH, Lee HY, Wu CC, et al. Etiologic and audiologic characteristics of patients with pediatric-onset unilateral and asymmetric sensorineural hearing loss. *JAMA Otolaryngol Head Neck Surg*. 2017;143:912–9.
- Alam K, Schofield D. Economic evaluation of genomic sequencing in the paediatric population: A critical review. *Eur J Hum Genet*. 2018;26:1241–7.

Cómo citar este artículo: Costales M, et al. Utilidad clínica de la secuenciación de nueva generación en el diagnóstico etiológico de la hipocausia neurosensorial en una Unidad de Hipocausia Infantil. *Acta Otorrinolaringol Esp*. 2019. <https://doi.org/10.1016/j.otorri.2019.05.005>

Anexo E: Genes incluidos en la v5 del panel de ceguera hereditaria OFTALMOgenics™

CONSISTENTEMENTE ASOCIADOS - SINDRÓMICOS (129 genes)

ABCC6	ABCD1	ABHD12	ADAMTS18	ADGRV1	ALMS1	ALPK1	ATP1A3	ATP8A2	BBS10	BBS12	BBS4	BBS5	BBS7	BBS9	C12orf65
CC2D2A	CDH23	CDH3	CEP78	CEP164	CEP41	CISD2	CLN3	CLN5	CLN6	CLN8	CNNM4	COL11A1	COL18A1	COL4A3	COL4A4
COL4A5	COL9A1	CP	CPLANE1	CSPP1	CTC1	DNAJC5	ECHS1	EXOSC2	FDXR	FGF8	FGFR1	FKRP	FLVCR1	GALC	GGCX
GJA1	GNPTG	GRN	HARS1	HESX1	HMX1	IDS	IDUA	INVS	KIAA0586	LAMA1	LOXL3	LZTFL1	MCOLN1	MECR	MFN2
MKKS	MKS1	MT-ATP6	MT-ND1	MT-ND4	MT-ND6	MTPP	MVK	MYO7A	NFIX	NOD2	NPHP1	NPHP3	NPHP4	NR2F1	P3H2
PANK2	PAX2	PCDH15	PEX1	PEX6	PEX10	PEX12	PEX13	PEX2	PEX26	PEX3	PEX5	PEX7	PGK1	PHYH	PNPLA6
PPT1	PRPS1	RNU4ATAC	RPGRIP1L	SCAPER	SDCCAG8	SLC19A2	SLC25A46	SPG11	SPG7	SRD5A3	SSBP1	TBCE	TCTN2	TIMM8A	TMEM138
TMEM216	TMEM231	TMEM237	TMEM67	TPP1	TRIM32	TTC21B	TTPA	TTR	USH1C	USH1G	VCAN	VPS13B	WDR19	WDR34	WFS1
WHRN															

CANDIDATOS - SINDRÓMICOS (62 genes)

ADAMTSL1	AFG3L2	AP3B2	AQP4	ARMC9	ARSG	ATAD3A	ATP13A2	B3GALNT2	BBIP1	BMP4	BRAT1	C5AR2	CASK	CDK10	CIB2
COL4A6	COL9A2	COL9A3	DNAJC17	DTHD1	EMC1	EPRS	ESPN	FDX2	FIBP	GALNS	GNS	GRID2	HIKESHI	IFT74	IFT80
KARS	KCTD7	LAMA5	LRRC32	MAG	MFF	MIR204	MRPS34	MT-TP	MT-TS2	NDUFA13	PDZD7	PIK3R4	PNPT1	POGZ	PSIP1
RAB11B	RDH11	SDHA	SLC25A1	SLC52A2	SON	TINF2	TUBA8	TUBB4B	TUBGCP4	TUBGCP6	WDPCP	WDR60	ZNF423		

CONSISTENTEMENTE ASOCIADOS - NO SINDRÓMICOS (117 genes)

ABCA4	ADAM9	AGBL5	AIPL1	ARHGEF18	ARL2BP	ATF6	ATOX7	BEST1	C1QTNF5	PCARE	CA4	CABP4	CACNA1F	CACNA2D4	CAPN5
CDHR1	CERKL	CHM	CNGA1	CNGA3	CNGB1	CNGB3	CRB1	CRX	CYP4V2	DHDDS	DHX38	DRAM2	EFEMP1	ELOVL4	EYS
FAM161A	FZD4	GNAT1	GNAT2	GPR179	GRK1	GRM6	GUCA1A	GUCY2D	IDH3B	IMPDH1	IMPG1	IMPG2	KCNJ13	KCNV2	KIZ
LCA5	LRAT	LRIT3	LRP5	MAK	MAPKAP3	MERTK	NEUROD1	NMNAT1	NRL	NYX	OAT	OPA3	OTX2	PAX6	PDE6A
PDE6B	PDE6C	PDE6G	PDE6H	PLK4	POC1B	PRCD	PROM1	PRPF3	PRPF31	PRPF4	PRPF6	PRPF8	PRPH2	RAB28	RAX2
RBP3	RBP4	RD3	RDH12	RDH5	REEP6	RGR	RGS9	RGS9BP	RHO	RIMS1	RLBP1	ROM1	RP1	RP1L1	RP2
RPE65	RPGR	RPGRIP1	RS1	RTN4IP1	SAG	SEMA4A	SIX6	SLC24A1	SLC7A14	SNRNP200	SPATA7	TIMP3	TMEM126A	TOPORS	TREX1
TRPM1	TSPAN12	TTL5	TULP1	ZNF408											

CANDIDATOS - NO SINDRÓMICOS (23 genes)

ACBD5	AHR	AQP4	CDH16	CFH	CTNNA1	FSCN2	GDF6	GNB3	GUCA1B	IFT43	IFT88	ILK	ITM2B	KIAA1549	NEK2
NXNL1	PITPNM3	PRDM13	RP9	SPP2	UNC119	ZNF513									

CONSISTENTEMENTE ASOCIADOS - SINDRÓMICOS/NO SINDRÓMICOS (35 genes)

ACO2	AH1	ARL6	BBS1	BBS2	CFAP410	C8orf37	CEP250	CEP290	CLRN1	COL2A1	CTNNA1	CWC27	HGSNAT	HK1	IFT140
IFT172	INPP5E	IQCB1	JAG1	KIF11	KLHL7	MFRP	MFSB8	NDP	NDUFS2	NR2E3	OFD1	OPA1	PCYT1A	POMGNT1	RCBTB1
SLC38A8	TTC8	USH2A													

CANDIDATOS - SINDRÓMICOS/NO SINDRÓMICOS (3 genes)

ADIPOR1	IFT81	TEAD1													
---------	-------	-------	--	--	--	--	--	--	--	--	--	--	--	--	--




FUNDACIÓN
M^a CRISTINA MASAVEU PETERSON

IMOMA
INSTITUTO DE MEDICINA ONCOLÓGICA
Y MOLECULAR DE ASTURIAS